

UNIVERSIDAD NACIONAL DE CAAGUAZU
FACULTAD DE CIENCIAS Y TECNOLOGÍAS
CARRERA DE INGENIERÍA EN ELECTRICIDAD



ELABORACION DE ALTERNATIVAS DE SISTEMAS DE
CONTROL DE PERDIDAS NO-TECNICAS EN EL SISTEMA
DE DISTRIBUCION DE LA ANDE EN EL DEPARTAMENTO
DE CAAGUAZU.

Ever Ovelar Aquino
Carlos Anibal Casco Galeano

Coronel Oviedo - Paraguay
Año 2023

UNIVERSIDAD NACIONAL DE CAAGUAZU
FACULTAD DE CIENCIAS Y TECNOLOGIAS
CARRERA DE INGENIERÍA EN ELECTRICIDAD

Elaboración de alternativas de sistemas de control de pérdidas no-técnicas en el sistema de distribución de la ANDE en el departamento de Caaguazú.

Elaborado por:

Ever Ovelar Aquino
Carlos Anibal Casco Galeano

Tutor:

Moises Britez

Trabajo presentado a la Facultad de Ciencias y Tecnología de la Universidad Nacional de Caaguazú, como requisito para la obtención del título de Ingeniero en Electricidad

Coronel Oviedo - Paraguay
Año 2023

Página de aprobación

Trabajo de fin de grado para la obtención del Título de Ingeniero en Electricidad aprobado en representación de la Facultad Ciencias y Tecnologías de la Universidad Nacional de Caaguazú, por el Tribunal Examinador constituido por los siguientes profesores.

Prof. Ing.

Prof. Ing.

Prof. Ing.

Acta Nro.: -----

Fecha: -----

Calificación: -----

Dedicatoria

A la memoria de mi madre Amalia Galeano.

Carlos Casco

A mi familia por ser mi principal pilar de apoyo y motivación para llegar al objetivo.

Ever Ovelar

Agradecimientos

A mi padre por apoyarme y darme la oportunidad de ser un profesional.

A mi esposa Tania y mis hijos Carlos Ian y Ámbar por ser la motivación para cumplir mi objetivo de ser ingeniero.

A los directivos de la Facultad de Ciencias y Tecnologías por brindarme el apoyo necesario para este logro.

A mis compañeros y amigos con quienes hemos transitado estos años de duración de la carrera.

Al ing Moises Britez por el acompañamiento como tutor de este proyecto.

Carlos Casco

Primeramente a Dios, quien me ha guiado y brindado la sabiduría para culminar esta carrera.

A mis padres y hermanos por brindarme el apoyo incondicional durante el proceso de formación.

A mi tutor el Ing. Moises Britez, por sus conocimientos brindados, su tiempo, y por sobre todo su buena predisposición, lo cual posibilitó la culminación satisfactoria de este trabajo final de grado.

A la facultad de Ciencias y Tecnologías por brindarme la oportunidad de formarme como profesional. Y a todas las personas que de una u otra forma me apoyaron en la realización de este trabajo.

Ever Ovelar

Resumen

Este proyecto final de grado tiene como objetivo principal la elaboración de alternativas capaces de controlar y mitigar las pérdidas no técnicas de la empresa ANDE, esto se refiere a fraudes cometidos por usuarios de la empresa, consecuentemente significan pérdidas monetarias por energía no facturada. Se consolidó una base de datos con registro históricos de consumo, potencia de contrato y tarifa de clientes en la ciudad de Coronel Oviedo, identificando usuarios intervenidos por fraude se logró procesar dichos datos en un entorno de aprendizaje supervisado, encontrando 13 % de clientes sospechosos en relación a los datos inspeccionados, de este agrupamiento se halló 8,5 % de usuarios con registro de fraude. Otra alternativa elaborada consiste en la aplicación de métodos y herramientas de medición en línea y local de puestos de distribución, difiriendo estos en costos de aplicación y operación.

Palabras claves: Pérdidas no técnicas, Robo de energía, Minería de datos, fraude, monitoreo.

Abstract

This final degree project has as its main objective the elaboration of alternatives capable of controlling and mitigating the non-technical losses of the ANDE company, this refers to fraud committed by users of the company, consequently they mean monetary losses for unbilled energy. A database was consolidated with historical records of consumption, contract power and customer rates in the city of Coronel Oviedo, identifying users intervened for fraud, it was possible to process said data in a supervised learning environment, finding 13% of suspicious customers in relation to the inspected data, 8.5% of users with a record of fraud were found from this grouping. Another elaborated alternative consists of the application of on-line and off-line measurement methods and tools for distribution transformers, these differing in application and operation costs.

Key words: Non-technical loss, Data minning, Energy theft, fraud, control.

Índice general

1. Introducción	1
I Revisión Bibliográfica	3
2. Pérdidas negras	5
2.1. Pérdidas negras, características	5
2.2. Naturaleza y origen de las pérdidas no técnicas	7
2.3. Dispositivos de medición eléctrica	7
2.3.1. Medidores electrónicos	10
2.4. Conceptos básicos empleados en el estudio de pérdidas no técnicas	11
2.4.1. Carga instalada	11
2.4.2. Capacidad instalada	11
2.4.3. Demanda	12
2.4.4. Curva de carga diaria	12
2.4.5. Demanda promedio	13
2.4.6. Factor de demanda	13

2.4.7. Factor de utilización	14
2.4.8. Factor de potencia	14
2.4.9. Factor de carga	14
2.4.10. Factor de pérdidas	14
2.4.11. Pérdidas de potencia y energía	15
2.4.12. Balance energético	15
2.5. Pérdidas no técnicas	17
2.5.1. Fallas administrativas	17
2.5.2. Fraude y hurto	18
2.6. Métodos de reducción de pérdidas no técnicas	20
2.6.1. Monitoreo de grupos de transformadores	21
3. Minería de datos	23
3.1. Modelos	24
3.2. Aprendizaje de máquina	24
II Resumen Ejecutivo	27
4. Resumen Ejecutivo	29
4.1. Descripción del trabajo	29
4.1.1. Métodos y Técnicas utilizadas	29
4.2. Justificación	30
4.3. Finalidad del proyecto	31

4.4. Metas	31
4.5. Objetivos	31
4.5.1. Objetivo general	31
4.5.2. Objetivos específicos	31
4.6. Beneficiarios	32
4.7. Producto	32
4.8. Especificaciones de actividades y tareas realizadas	32
4.9. Recursos necesarios	33
4.9.1. Recursos humanos	33
4.9.2. Recursos materiales	33
4.10. Factibilidad económica	33
4.10.1. Beneficios	34
4.10.2. Costos	35
4.10.3. Análisis financiero	37

III Ingeniería de Diseño 39

5. Análisis de datos 41

5.1. Relevamiento de datos	41
5.2. Limpieza de datos	43
5.3. Análisis exploratorio	44

6. Monitoreo de transformadores y balance energético 51

6.1. Alimentador Coronel Oviedo 1 (COV-1)	51
6.2. Determinación de pérdidas técnicas	52
6.3. Determinación de pérdidas no técnicas	55
6.4. Control de pérdidas	56
6.4.1. Control de suministros clandestinos	56
6.4.2. Medición off-line a puestos de distribución	57
6.4.3. Medición on-line de puestos de distribución	63
IV	65
7. Resultados y Discusión	67
8. Conclusiones	69
9. Recomendaciones	71
Apéndice	71
A. Ingeniería de diseño	73
A.1. Análisis de datos	73
A.1.1. Limpieza de datos	73
A.1.2. Exploración de datos y búsqueda de sospechosos	76
A.1.3. Agrupación de usuarios	123
A.2. Monitoreo de transformadores y balance energético	131
<hr/>	
<i>E. Ovelar, C. Casco</i>	X

A.2.1. Cálculo de constantes de modelo estadísticos y pérdidas técnicas: .	131
A.2.2. Modelo DigSilent	132
A.2.3. Medidores en PD	133
A.2.4. Cálculo probabilístico	134
B. Resumen ejecutivo	137
B.1. Cálculo de beneficios	137
B.2. Cálculo de costos	137
B.3. Análisis financiero	140
Referencias	142

Lista de abreviaturas

ANDE: Administración Nacional de Electricidad.

OPEN: Sistema de gestión de cobranzas y trabajos comerciales de la ANDE.

ES-COV: Estación Coronel Oviedo.

PD: Puesto de distribución.

TC: Transformador de corriente.

TP: Transformador de potencia.

kWh: Kilowatt-hora

kVA: Kilovolt-amper.

kVAr: Kilovolt-amper reactivo.

Clyfsa: Compañía de Luz y Fuerza S.A.

AP: Alumbrado público.

BT: Baja tensión.

MT: Media tensión.

GSM: Sistema global para comunicaciones móviles.

HHP: Modalidad de contrato de la empresa ANDE

VAN: Valor actual neto

TIR: Tasa interna de retorno

Glosario de términos

Factura: Documento con respectivo monto a abonar por parte del cliente.

DigSilent: Programa para simulaciones eléctricas.

Machine Learning: Técnicas de procesamiento de datos.

On-line: Transferencia de datos por medio de una red.

Off-line: Adquisición de datos de manera local.

Índice de tablas

4.1. Posible aumento de recaudación	34
4.2. Costos de análisis e inspección de sospechosos	35
4.3. Costo operativos del primer mes	35
4.4. Costo de implementación de medición off-line	36
4.5. Costo de implementación de la medición on-line	37
5.1. Categorías de clientes según la segmentación encontrada	44
5.2. Selección de clientes analizados	49
6.1. Cálculo de pérdidas no técnicas.	55
6.2. Resumen de pérdidas	56
6.3. Valores comerciales de transformadores de corriente. Fuente: Catalogo de TC's de ABB.	58
6.4. Listado de materiales necesarios para realizar monitoreo off-line de puestos de distribución de 100 KVA.	59
6.5. Listado de materiales para esquema de medición de puesto de distribución de 25 KVA	61

A.1. Alternativa, verificación aleatoria de usuarios	134
A.2. Alternativa, medición off-line	134
A.3. Alternativa, medición on-line	135
B.1. Costos de materiales para implementación de medidor totalizador para PD de 100 KVA.	138
B.2. Costo de mano de obra de montaje de totalizador.	138
B.3. Costos totales de implementación de medición totalizadora ara PD de 100 KVA.	138
B.4. Costos de materiales para implementación de medidor totalizador para PD de 25 KVA	139
B.5. Costo de mano de obra de montaje de totalizador para PD 25 kVA	139
B.6. Costos totales de implementación de medición totalizadora para PD de 25 kVA	139
B.7. Rentabilidad A1	140
B.8. Rentabilidad A2	140
B.9. Rentabilidad A3	141

Índice de figuras

2.1. Medidor electromecánico y sus partes	9
2.2. Procesamiento de la señal digital (tiempo discreto)	11
2.3. Cálculo de pérdidas de un grupo de transformadores mediante el uso de módulos totalizadores	22
3.1. División de los métodos en aprendizaje de maquina	25
5.1. Datos en bruto del sistema de gestión de facturas ANDE	42
5.2. Datos limpios y procesados	44
5.3. Consumos de clientes en general (kWh)	45
5.4. Grafica de cajas, relación media con contrato	46
5.5. Histograma a) Set de datos completo b) Sin valores atípicos	47
5.6. Regresión lineal y clientes sospechosos	48
5.7. Registro histórico de clientes sospechosos	48
6.1. Zona de influencia del alimentador COV-1. Obtenido de Administración Nacional de Electricidad – Dpto de Operaciones 2022.	53

6.2. Ubicación geográfica de puestos de distribución con sus respectivos códigos de PD. Obtenido de Administración Nacional de Electricidad – Dpto de Operaciones 2022.	53
6.3. Resultados obtenidos tras flujo de carga en DigSilent	54
6.4. Tablero de medición de puesto de distribución	60
6.5. Diseño de montaje de gabinete de medición de puestos de distribución públicos.	62
6.6. Sistema de transferencia de datos	64
A.1. Regresión y agrupación de clientes con contrato de 13200 kW en el estrato 410.	123
A.2. Regresión y agrupación de clientes con contrato de 6600 kW en el estrato 410.	123
A.3. Regresión y agrupación de clientes con contrato de 7040 kW en el estrato 142.	124
A.4. Regresión y agrupación de clientes con contrato de 19800 kW en el estrato 142.	124
A.5. Regresión y agrupación de clientes con contrato de 13200 kW en el estrato 142.	125
A.6. Regresión y agrupación de clientes con contrato de 21120 kW en el estrato 142.	125
A.7. Regresión y agrupación de clientes con contrato de 16500 kW en el estrato 142.	126
A.8. Regresión y agrupación de clientes con contrato de 26400 kW en el estrato 142.	126

A.9. Consumo histórico de clientes sospechosos con contrato de 13200 kW en el estrato 410.	127
A.10. Consumo histórico de clientes sospechosos con contrato de 6600 kW en el estrato 410.	127
A.11. Consumo histórico de clientes sospechosos con contrato de 7040 kW en el estrato 142.	127
A.12. Consumo histórico de clientes sospechosos con contrato de 19800 kW en el estrato 142.	128
A.13. Consumo histórico de clientes sospechosos con contrato de 13200 kW en el estrato 142.	128
A.14. Consumo histórico de clientes sospechosos con contrato de 21120 kW en el estrato 142.	128
A.15. Consumo histórico de clientes sospechosos con contrato de 16500 kW en el estrato 142.	129
A.16. Consumo histórico de clientes sospechosos con contrato de 26400 kW en el estrato 142.	129
A.17. Std vs mean de clientes con fraudes intervenidos.	130
A.18. Consumo histórico de clientes con fraudes intervenidos.	130
A.19. Esquema modelado en DigSilent	132
A.20. Esquema de mando para conexión del medidor	133

Capítulo 1

Introducción

ANDE es la empresa encargada de comercializar la energía eléctrica dentro del territorio paraguayo, en este proceso se encarga del suministro, medición y facturación del servicio. Las pérdidas no técnicas, como su nombre lo indica se refiere a la energía no facturada pero consumida por el usuario final, también se le denomina pérdidas negras. Según la propia empresa en el departamento de Caaguazú estas pérdidas llegan hasta 3,22 % de la energía total facturada, las causas que contribuyen a que las pérdidas no técnicas se encuentren en los niveles mencionados, según investigaciones realizadas por la propia ANDE corresponden a: i) Cantidad considerable de medidores electromecánicos en el parque de baja tensión. ii) Falencias en la calidad de medición en media tensión. iii) Conexión directa de familias en territorios sociales y conexiones fraudulentas de usuarios.

Esta problemática motiva a la elaboración de este trabajo, cuyo objetivo principal es elaborar alternativas de sistemas de control de pérdidas negras, a fin de mitigar las mismas. Con la aplicación de estas alternativas la empresa mejoraría su flujo de ingreso reflejándose en los índices de recaudación.

Para este proyecto se toma como muestra la población de la ciudad de Coronel Oviedo, departamento de Caaguazú, por la variedad y cantidad de clientes que posee. En la elaboración del análisis de datos se filtran dichos usuarios disminuyendo el alcance a dos estratos diferenciados de consumidores.

En el capítulo 2 de este trabajo se revisan las principales definiciones y bibliografías sobre pérdidas negras o no técnicas, como así también los diferentes equipos de medición utilizados por la empresa.

En el capítulo 3 se exploran conceptos de minería de datos a fin de aplicarlos en el set de datos obtenidos.

En el capítulo 5 se procede al análisis de datos, como una alternativa a proponer, primeramente se recabaron los mismos, se limpiaron y filtraron, para posteriormente gracias al procesamiento y relacionamiento de antecedentes considerar grupos sospechosos.

Finalmente en el capítulo 6 se elabora la alternativa de mediciones de transformadores, de manera on-line y off-line, se confeccionaron lista de materiales para llevar a cabo estos trabajos, además una distribución probabilista como criterio de implementación de estas opciones.

Parte I

Revisión Bibliográfica

Capítulo 2

Pérdidas negras

2.1. Pérdidas negras, características

Durante todo el proceso de producción, transporte y venta de la energía eléctrica, no es posible entregar toda la energía disponible a los consumidores. Principalmente por los fenómenos físicos que se presentan en la conducción y transformación de la energía eléctrica conocidas como pérdidas técnicas, las mismas que son inevitables, sin embargo, es posible reducirlas a valores aceptables. Las empresas distribuidoras deben llevar registros precisos de la energía suministrada a los usuarios, con el fin de poder cobrar toda la energía consumida. Sin embargo, debido a errores de diversa índole, se presentan inexactitudes en el registro de los consumos, representando pérdidas para la institución. Por otra parte, es muy difícil que la empresa logre recaudar el pago de toda la energía que fue registrada. La diferencia entre la energía que fue entregada a los usuarios y la que se logra facturar su pago se denomina pérdidas no técnicas [11].

Las pérdidas no técnicas globales se calculan como la diferencia entre las pérdidas totales de un sistema eléctrico y las pérdidas técnicas estimadas para el mismo. Dentro de la comercialización son varias las actividades que contribuyen al porcentaje de pérdidas negras de energía y en consecuencia al total; siendo necesario identificar las mismas y a los usuarios por el destino y la magnitud de la energía consumida.

La realidad en los últimos tiempos ha puesto en evidencia que la des-inversión en los sistemas de distribución y comercialización de la energía eléctrica no solo conduce a un deterioro en la calidad del servicio que se presta, sino que es uno de los factores contribuyentes al incremento de las pérdidas, tanto las técnicas como las no técnicas.

Los proyectos de control y reducción de pérdidas en los sistemas de distribución y comercialización libera equipos de generación y transmisión, incidiendo en los niveles de inversión futura, como en el valor de los costos marginales y de las tarifas basadas en ellos. Además, constituye la base del planeamiento, diseño y operación de los sistemas eléctricos. La implementación y desarrollo de programas de control y reducción de pérdidas negras depende de las condiciones y características propias de cada sistema [2].

Como medida efectiva para la detección de usuarios infractores y para el control de pérdidas negras, constituye la investigación completa de las instalaciones de medida y acometidas. La investigación según la base del Banco Mundial persigue dos objetivos:

1. Identificar las instalaciones con medidas defectuosas.
2. Servir como datos muestrales para la estimación de pérdidas debidas al fraude, donde se recomienda usar una muestra estratificada.

A diferencia de las pérdidas técnicas, la reducción de las pérdidas no técnicas produce un incremento de los ingresos (por ventas) sin afectar sustancialmente la producción en la generación. Sin embargo, existe una relación indirecta, ya que la reducción puede provocar una reducción en los consumos de ciertos clientes, produciendo una reducción en la generación. Según estimaciones del Banco Mundial, se ha demostrado que al aplicar mediciones correctas, en ciertos estratos estos tienden a reducir sus consumos hasta en un 30 [10].

2.2. Naturaleza y origen de las pérdidas no técnicas

Las pérdidas no técnicas son por naturaleza, pérdidas íntimamente vinculadas con la calidad de la gestión entre los clientes y la empresa, el origen de estas pérdidas se da en cada una de las etapas que normalmente se siguen para dar servicio al abonado y que a continuación se describen:

- **Identificar:** Es decir se debe conocer los datos técnicos, administrativos y comerciales característicos de cada cliente pero en ciertos casos los datos del mismo son erróneos y no se encuentran bien identificados originándose con ellos pérdidas no técnicas, como por ejemplo error en la tarifa.
- **Medir:** El consumo de cada cliente debe ser registrado sin error pero se pueden tener medidores en fraude, defectuosos, estimaciones de consumo erróneo etc., los cuales también originan pérdidas no técnicas.
- **Facturar:** Con las mediciones que se registran del cliente se procede a la facturación de acuerdo al contrato establecido, pero por razones de datos erróneos, lentitud e irregularidad en la edición y cobro de la factura se originan estos tipos de pérdidas.
- **Cobrar:** Se debería recaudar en el plazo más corto posible la suma debida por los clientes, aquí las pérdidas no técnicas tienen su origen en la falta de pago por parte del cliente.

2.3. Dispositivos de medición eléctrica

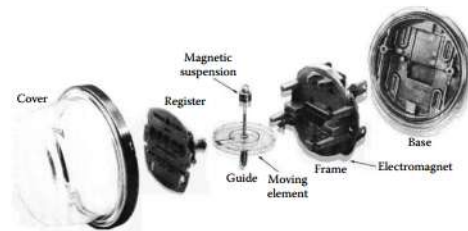
Un medidor eléctrico es el dispositivo utilizado para medir la electricidad vendida por la compañía eléctrica. Se ocupa de dos cantidades básicas: energía y potencia. La energía es equivalente al trabajo, la potencia es la tasa de hacer este trabajo. La potencia aplicada (o consumida) durante cualquier período de tiempo es energía. En otras palabras, la potencia integrada en el tiempo es energía, la unidad básica de potencia es el watt, la unidad eléctrica básica de la energía es watt-hora.

Un medidor de watt-hora se utiliza para medir la energía eléctrica entregada a un usuario residencial, comercial o clientes industriales, también se utiliza para medir la energía eléctrica que pasa por varias partes de sistemas de generación, transmisión y distribución. Un vatímetro mide la tasa de energía, es decir, potencia (en vatios por hora o simplemente vatios). Los contadores eléctricos pueden ser de dos tipos: los contadores electromecánicos y los electrónicos (o también llamados medidores digitales). La Figura 2.1a muestra un medidor monofásico (electromecánico); La Figura 2.1b muestra sus partes básicas, las cuales son; visor, registrador, parte móvil, el marco y la base. La Figura 2.1c muestra un diagrama de un motor típico y un sistema de retardo magnético para un medidor monofásico, el sistema de retardo magnético hace que el disco del rotor establezca, en combinación con el estator, la velocidad a la que girará el eje para determinar una condición de carga determinada.

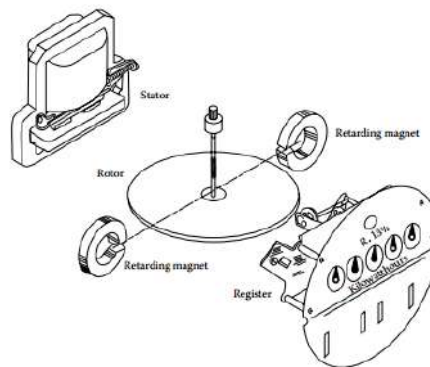
El estator o bobina de potencial, es un elemento electromecánico, que se conecta en paralelo a las líneas de la fuente, donde a su vez puede ser medido el voltaje aplicado, esta bobina de potencial es enrollada en un núcleo de hierro laminado por n vueltas de un calibre de alambre muy fino, de forma adecuada para suministrar la distribución requerida del flujo magnético. El rotor o disco, constituye la parte rotatoria del motor que opera el tren de engranajes. Está hecha de una pieza delgada de aluminio, cuando el electroimán actúa sobre el disco para hacerlo dar vueltas el tornillo sin fin se pone en movimiento con la primera rueda dentada del juego de engranajes de la esfera y así se registra el número de revoluciones del disco. El imán permanente tiene la función de regular la velocidad del disco, lo que es necesario para impedir que gire muy rápidamente. En los medidores antiguos de marcas Sangamo, Westinghouse etc. los imanes permanentes son ajustables para calibrar su acción reguladora, es decir la velocidad del disco. El registro ciclo métrico de lectura o esfera, consta de una serie de engranajes que contabilizan el número de revoluciones del disco para registrarlos en KWH de acuerdo a la relación entre el primer y último engranaje guardando estrecha relación con la constante de disco. En el registro ciclo métrico es fácil la toma de lectura por que consta de tambores con números que avanzan de acuerdo a la relación de vueltas del registro; existen registradores que poseen manecillas de reloj que para tomar la lectura se debe considerar una técnica de



(a) Medidor electromecánico



(b) Partes del medidor electromecánico



(c) Diagrama del mecanismo del medidor

Figura 2.1: Medidor electromecánico y sus partes

lectura respecto a la posición de sus manecillas; además que las manecillas de las unidades, decenas, centenas, etc giran en direcciones opuestas [4].

Los medidores utilizados para medir la energía eléctrica que pasa por varias partes de los sistemas de generación, transmisión y distribución son necesarios para medir grandes cantidades de energía a voltajes relativamente altos. Para esas aplicaciones, se desarrollan medidores clasificados para transformadores, se utilizan junto con los transformadores de instrumentos estándar, es decir, los transformadores de corriente (CTs) y transformadores de potencial (PTs). Estos transformadores reducen la tensión y la corriente a valores adecuados para medidores de baja tensión y baja corriente. La muestra un medidor electromecánico conectado a un circuito de alto voltaje a través de CTs y PTs.

Un medidor de demanda es básicamente un medidor de vatios con un elemento de tiempo agregado, las funciones del medidor como integrador y suma los kilovatios de energía utilizados en un cierto intervalo de tiempo, por ejemplo, 15, 30 o 60 min. Por lo tanto, el medidor de demanda indica energía por intervalo de tiempo, o potencia promedio, que se expresa en kilovatios [7].

2.3.1. Medidores electrónicos

Las empresas de servicios públicos han comenzado a utilizar nuevos medidores electrónicos (o digitales) con registros de demanda programable (PDR) desde la década de 1980. Tales medidores electrónicos no tienen partes móviles en su lugar, se construyen alrededor de circuitos integrados de gran escala, componentes de estado sólido y lógica digital. El funcionamiento de un medidor electrónico es muy diferente a los medidores electromecánicos, los circuitos electrónicos muestrean el voltaje y formas de onda actuales durante cada ciclo eléctrico y las convierte en cantidades digitales. Los circuitos luego manipulan estos valores para determinar numerosos parámetros eléctricos, como kW, kWh, kvar, kvarh, PF, kVA, corriente rms y voltaje rms. Un PDR también puede medir la demanda, mientras que un registro tradicional mide solo la cantidad de electricidad utilizada en un mes, un perfil de demanda muestra la cantidad de electricidad que un cliente usa en un mes.

El medidor electrónico digital se basa en un proceso llamado conversión de señales analógicas a digitales. Esto se efectúa usando un aparato llamado Convertidor Analógico/Digital (Analogic to digital converter). El ADC toma muestras de la señal analógica en distancias discretas de tiempo. Las señales de tiempo discreto son a su vez convertidas a valores numéricos por el ADC. Una vez obtenido el formato digital, los circuitos digitales, por ejemplo, los microprocesadores, se puede, en forma fácil y confiable, procesar estas señales. La Figura 2.2 muestra una representación gráfica de este procesamiento digital.

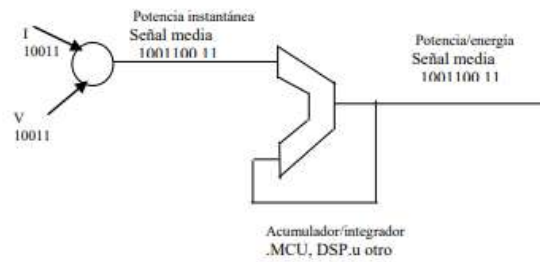


Figura 2.2: Procesamiento de la señal digital (tiempo discreto)

2.4. Conceptos básicos empleados en el estudio de pérdidas no técnicas

2.4.1. Carga instalada

Es la suma de todas las potencias nominales de las cargas conectadas a un sistema o a parte de él, se expresa generalmente en kVA, MVA, kW o MW.

$$CI = \sum \text{Potencias nominales de las cargas} \quad (2.1)$$

La carga instalada se refiere a la totalidad del equipo eléctrico instalado en cierta ubicación. Tal es el caso de una planta que tiene 10 motores de 100 kW que cuenta con una carga instalada de 1,000 kW.

2.4.2. Capacidad instalada

Es la suma de todas las potencias nominales de los equipos instalados (transformadores, generadores) a líneas que suministran la potencia eléctrica a las cargas conectadas. Se la conoce también como capacidad nominal del sistema.

$$C_{INS} = \sum \text{Potencia nominal de los equipos} \quad (2.2)$$

2.4.3. Demanda

Es la cantidad de potencia que un consumidor usa en cualquier momento, la misma que normalmente es variable con el tiempo.

La demanda de una instalación eléctrica en los terminales receptores, es el valor medio de la cantidad de potencia tomada en un determinado intervalo de tiempo. El intervalo de tiempo durante el cual se toma el valor medio de potencia se denomina intervalo de demanda. La demanda se puede expresar en kVA, kW, kVAR, A, etc.

La demanda es el consumo instantáneo de potencia de un usuario de energía eléctrica. A pesar de que existen diversas cantidades que pudieran reflejar una demanda, tales como la potencia activa (kW), potencia reactiva (kVAr) y potencia aparente (kVA), lo más común es encontrar la demanda expresada en kilowatts (kW). Se puede hablar de la demanda como el consumo de una máquina eléctrica (por ejemplo, un motor) o como la suma de varias demandas individuales (por ejemplo, en una planta se tiene la suma de las demandas individuales de los motores, lámparas y demás máquinas). Es importante considerar que muchas cargas variarán no solo a lo largo del día, sino que el patrón de consumo cambiará a través de las distintas estaciones del año [5].

Demanda máxima

Corresponde a la mayor demanda que se presenta en el sistema en un periodo de trabajo previamente establecido. En demanda máxima es donde se presenta la máxima caída de tensión en el sistema y las mayores pérdidas de potencia kW, KVAR y energía kWh.

2.4.4. Curva de carga diaria

La variación de la demanda en el tiempo para una carga dada origina el ciclo de carga la cual se conoce como curva de carga. La curva de carga diaria está formada por los promedios de los valores registrados en intervalos de una hora para cada hora del día. Las curvas de carga diaria dan una indicación de las características de la carga en el sistema,

sean estas predominantemente residenciales, comerciales o industriales y de la forma que estas se combinan para producir el pico de demanda.

Su análisis permite conocer la forma en que han venido variando las cargas durante un determinado período y constituyen una base para determinar las tendencias predominantes de las cargas del sistema [3].

2.4.5. Demanda promedio

Se define como la relación entre el consumo de energía del usuario durante un intervalo dado y el intervalo mismo.

$$P_m = \frac{\text{Energía consumida en el tiempo } (t) \text{ en kWh}}{\text{T en h}} \quad (2.3)$$

2.4.6. Factor de demanda

A la razón entre la demanda máxima y la carga total instalada en un intervalo de tiempo se la conoce como factor de demanda.

Por lo general este factor es menor a la unidad, siendo igual a la unidad sólo cuando en el intervalo considerado, todos los aparatos conectados al sistema estén absorbiendo sus potencias nominales, lo cual es muy improbable.

$$F_D = \frac{D_{max}}{CI} \leq 1 \quad (2.4)$$

El factor de demanda indica el grado al cual la carga total instalada se opera simultáneamente.

2.4.7. Factor de utilización

El factor de utilización en un sistema eléctrico en un intervalo de tiempo, es la razón entre la demanda máxima y la capacidad instalada del sistema.

$$F_U = \frac{D_{max}}{C_{INS}} \quad (2.5)$$

Este factor da una indicación de la fracción de la capacidad del sistema que se está utilizando durante el pico de carga en el intervalo considerado.

2.4.8. Factor de potencia

Es la relación entre la potencia activa y la potencia aparente determinada en el sistema o en uno de sus componentes.

$$F_{pot} = \cos \phi = \frac{P}{S} \quad (2.6)$$

2.4.9. Factor de carga

Se lo obtiene de la relación entre la demanda promedio en un intervalo de tiempo establecido y la demanda máxima en el mismo intervalo de tiempo.

$$F_C = \frac{P_m}{D_{max}} = \frac{\text{Energía absorbida en el intervalo de tiempo } t}{D_{max} \cdot t} \quad (2.7)$$

Este factor indica el grado al cual el pico de la carga es sostenido durante el período.

2.4.10. Factor de pérdidas

El factor de pérdidas está determinado por la razón entre las pérdidas promedio de potencia y las máximas pérdidas de potencia en un intervalo especificado.

$$F_p = \frac{P_{pmedia}}{P_{pmax}} \quad (2.8)$$

Con este factor se pueden determinar las pérdidas de energía en un periodo de tiempo de la siguiente manera:

$$F_p = \frac{\text{kWh de pérdidas en el periodo de tiempo } t}{\text{kW máximos de pérdidas} * \text{número de horas del período}} \quad (2.9)$$

2.4.11. Pérdidas de potencia y energía

Las pérdidas de potencia se presentan en cada uno de los componentes del sistema en funcionamiento, la mayoría de estas debido al paso de corriente a través de ellos, las cuales consideradas dentro de un periodo de tiempo se transforman en pérdidas de energía

Porcentaje de pérdida de energía

Las pérdidas de energía se las expresa como un porcentaje respecto de la energía suministrada a un sistema o parte del mismo

$$\% \text{Pérdidas Energía} = \frac{\text{Pérdidas de Energía}}{\text{Energía suministrada}} * 100 \quad (2.10)$$

2.4.12. Balance energético

El balance energético se lo realiza con la finalidad de conocer como se encuentra el uso de la energía dentro del sistema de la empresa. Se lo expresa con la siguiente expresión:

$$E_s = E_R + E_{AP} + E_{Ptot} \quad (2.11)$$

Donde:

E_s = Energía suministrada

E_R = Energía Registrada

E_{AP} = Energía Consumida en Alumbrado público

E_{Ptot} = Energía de pérdidas totales

Siendo:

$$E_{Ptot} = E_{PT} + E_{PNT} \quad (2.12)$$

E_{PT} = Energía de pérdidas técnicas

E_{PNT} = Energía de pérdidas no técnicas

La energía suministrada es la que entra al sistema de la empresa en un determinado período de tiempo para poder satisfacer la demanda energética de todos sus clientes, esta energía es medida en las subestaciones para verificar el total de la energía recibida por la empresa.

La energía registrada es la suma entre la energía que se factura a todos los clientes registrados en el sistema comercial de la empresa y la energía para el consumo propio de la misma.

La energía utilizada para iluminación pública se la estima a partir del inventario de todas las luminarias existentes en el sistema de la empresa, conociendo de esta manera la carga total instalada para este servicio y al suponer 12 horas de funcionamiento continuo, se determina la energía consumida por todo el sistema de alumbrado público.

Dentro de las pérdidas totales de energía se puede diferenciar entre pérdidas técnicas y pérdidas no técnicas. Las pérdidas técnicas se producen debido al paso de la energía en conductores y transformadores mientras que las pérdidas no técnicas se deben a la energía utilizada ilegalmente.

2.5. Pérdidas no técnicas

Se las obtiene como la diferencia entre las pérdidas totales y las pérdidas técnicas del sistema, estas pérdidas a diferencia de las pérdidas técnicas es energía que se utiliza por algún usuario registrado o no en el sistema comercial de la empresa distribuidora la misma que no recibe la respectiva retribución por el servicio prestado. Se puede decir que las pérdidas no técnicas de energía se deben a varios factores, unos ocasionados por parte de la empresa como problemas técnicos, comerciales, administrativos y otros que afectan el comportamiento normal de los clientes como son los factores económicos, culturales o sociales. Las pérdidas de energía no técnicas se las puede dividir entre pérdidas de energía gestionables y no gestionables. Las pérdidas de energía gestionables son aquellas que se dan dentro de zonas consideradas normales donde la empresa puede realizar el control de una manera segura, en tanto que las pérdidas de energía no gestionables tienen lugar en zonas conflictivas con graves problemas de orden público, por lo cual para realizar acciones de control la empresa necesita ayuda de la fuerza pública. Según las causas que las originan las pérdidas no técnicas se clasifican en:

- Fallas Administrativas
- Fraude y Hurto

2.5.1. Fallas administrativas

Estas pérdidas son originadas debido a errores en los diferentes procesos de gestión comercial y administrativa de la empresa; entre los cuales se puede mencionar:

- Errores en el proceso de ingreso de nuevos clientes

Se pueden producir cuando los contratistas encargados de instalar los medidores de los nuevos clientes, no entregan a tiempo la información al departamento comercial de la empresa, por lo que estos nuevos usuarios no son registrados de manera oportuna en el sistema de comercialización, consumiendo energía sin tener que pagar por ella. Así mismo

se pueden cometer errores por parte del personal encargado de realizar el ingreso de datos de los clientes, afectando de esta manera la calidad de la información registrada en el sistema comercial de la empresa.

- Errores en el proceso de lectura

El proceso de lectura es uno de los procesos más críticos, que de no realizarse de manera efectiva, podría generar pérdidas de energía considerables. Los errores se producen por la toma incorrecta del valor registrado en el medidor de los clientes por parte de los lectores encargados de anotar los consumos, lo cual causa una facturación mayor o menor de la energía realmente consumida. Otra causa de errores en la lectura de las mediciones se debe al propio estado en el que se encuentra el equipo de medición, pudiendo éste estar obsoleto, registrando consumos no reales.

- Errores en el proceso de facturación

El correcto desarrollo de este proceso conjuntamente con lecturas confiables garantiza el cobro real de la energía consumida por los clientes. Una práctica común en las distribuidoras es realizar la estimación de consumos mediante un factor que pondera la medición del medidor de un usuario para poder registrar su consumo. Esto no es recomendable, ya que se basa en supuestos, produciendo una medición menor o mayor que el consumo real afectando los datos para su posterior facturación. Durante el proceso de facturación también se pueden presentar irregularidades por parte del personal encargado de realizar el mismo, ya que pueden alterar los parámetros de facturación llegando a favorecer a personas o entidades conocidas.

2.5.2. Fraude y hurto

Existen diversas maneras mediante las cuales los usuarios, clientes o no de la empresa, pueden usar la energía de manera ilícita, interviniendo las redes del sistema de distribución o alterando su equipo de medida.

Entre las prácticas de hurto más comunes tenemos:

- Energía sin legalizar

Son aquellas conexiones que se derivan del sistema de distribución sin autorización por parte de la empresa distribuidora por lo cual estas conexiones clandestinas no poseen ningún tipo de medición. Es común encontrar este tipo de conexiones para actividades relacionadas a la construcción como pueden ser soldadoras, concretadoras, etc., también por parte de vendedores ambulantes, en fiestas populares, en asentamientos irregulares o invasiones, etc. Este tipo de pérdidas es muy difícil de combatir ya que son por lo general conexiones temporales, además el infractor no posee vínculo comercial con la empresa por lo que es complicado recuperar el valor de la energía consumida. Una de las mejores alternativas para combatir este tipo de pérdidas es a través de la construcción de redes lo menos vulnerables posible.

- Línea Directa

Se denomina línea directa a las conexiones ilegales que se derivan desde el sistema de distribución hacia los usuarios, los mismos que se encuentran registrados en el sistema de la empresa y que bien pueden estar al día o atrasados en el pago del servicio. En este caso a diferencia de la energía sin legalizar es menos difícil realizar el control, ya que se puede llegar a identificar a los infractores, pudiendo de esta manera multar y sancionar a los mismos, con lo cual se puede recuperar el valor de la energía consumida ilegalmente. Las pérdidas de energía no técnicas también se deben a fraudes ocasionados por los clientes de la empresa, entre los cuales tenemos:

Adulteración del equipo de medición: Son todas las maniobras irregulares realizadas por personal ajeno a la empresa en el equipo de medida de los clientes, buscando de esta manera tener subfacturación respecto del valor real del consumo de energía. Para el caso de medidores de inducción, estas maniobras ocasionan movimiento más lento del disco, logrando registrar menos energía de la verdaderamente consumida. Se pueden detectar las

adulteraciones internas del equipo de medida mediante la revisión de los sellos de seguridad, los cuales pueden estar rotos o adulterados, de igual manera revisando la carcasa, ya que otra práctica común para acceder a la parte interna del equipo de medida es mediante la perforación de la misma.

Fase interrumpida: Este fraude se lo puede encontrar para el caso de medidores con más de dos hilos, el cual consiste en puentear una de las fases, lo que produce una detención total o parcial del mecanismo de registro de los medidores. Este fraude al igual que la adulteración del equipo de medida conduce a una subfacturación.

Adulteración de la señal medida: Los clientes que tienen grandes consumos como son industrias, comercios, viviendas lujosas, etc., requieren de medición indirecta es decir la utilización de transformadores de potencial y transformadores de corriente cuyo objetivo es adaptar las señales a los equipos de medición. El fraude se produce al alterar las conexiones en los transformadores de potencial y de corriente, con lo cual se adulteran las señales de tensión y corriente que son enviadas al medidor de energía, produciendo subfacturación al igual que los fraudes descritos anteriormente [3].

2.6. Métodos de reducción de pérdidas no técnicas

En la presente sección se mencionarán algunos métodos de reducción de pérdidas no técnicas, los cuales son: Reingeniería de procesos para reducir las pérdidas debidas a los procesos administrativos.

Revisión completa de los abonados ubicados en barrios o zonas no marginales, siguiendo rutas preestablecidas para evitar pérdidas por conexiones ilegales, fraude y conexiones clandestinas.

Instalación de medidores en cajas anti-hurtó con el objetivo de evitar que sean intervenidos.

Colocación de sello de plástico, difícil de violar, numerado donde conste la sigla de la empresa sobre relieve para permitir un control inmediato sobre sus abonados.

Instalación de conductores antihurto para evitar las conexiones clandestinas. Dicho conductor está compuesto por un alma de aluminio, una capa de aislante plástico que la rodea, una cubierta conductora de cobre que recubre a la primera capa aislante y otra capa aislante de plástico que se encuentra al exterior del conductor.

Planes especiales de facilidad de pago de deudas acumuladas, condonación de deudas, etc.

Asesorar en el uso racional de la energía a fin de lograr que el cliente una vez ingresado modere sus consumos, evitando así abonar abultadas facturas con las posibles consecuencias de suspensiones de servicios, avisos de corte, etc [7].

Evaluación de tecnologías para mejorar el sistema de medición de la energía [13].

2.6.1. Monitoreo de grupos de transformadores

Consiste en el análisis y seguimiento de los resultados obtenidos mediante el uso de totalizadores de energía colocados en los grupos de transformadores, comparando la energía distribuida por éstos, con la facturación de los clientes asociados a dichos grupos. La lectura de los totalizadores está incluida dentro de los ciclos de lectura de los clientes asociados a ellos. Esto se hace con la finalidad de que el período en el cual es distribuida la energía desde el grupo de transformadores, coincida con el lapso de facturación de los clientes asociados a dicho grupo, y así obtener valores precisos de pérdidas. Empleando este módulo, mensualmente, también se obtienen reportes de las pérdidas de energía (expresados en kWh y porcentajes) de cada uno de los grupos de transformadores controlados. Además, permite organizar las pérdidas de cada grupo de transformadores por circuito y zona geográfica [1].

Con base en los reportes por grupo de transformadores, se establecen aquellos con mayores pérdidas de energía para emitir la orden de revisión en campo 2.3.

Nro Medidor:	4010107	Banco Transformador:	E2DCUA81
Medidor Totalizador		Medidores Asociados	
Lectura anterior:	2602	Fecha lectura anterior:	21/03/2005
Lectura actual:	2811	Fecha lectura actual:	21/04/2005
Tipo Banco Transf.:	3 x 37,5	Relación TC:	300
Consumo (Kwh/Día):	404	Factor Multiplicador:	60
Carga Alumbrado-Otros (Kwh/Día):	25	Despacho de Energía (Kwh/Día):	379
Promedio Total Kwh/Día Clientes:	200	Promedio Total Kwh/Mes Clientes:	6 000
Pérdida Total Kwh/Día:	179	Pérdida Total Mensual:	5 370
Porcentaje Pérdida:	47,23	Media de Consumo Mensual:	667

Figura 2.3: Cálculo de pérdidas de un grupo de transformadores mediante el uso de módulos totalizadores

Capítulo 3

Minería de datos

Es parte del área de inteligencia de negocio, da un paso más y apoya no solo a entender que pasó, como está y por qué, sino que responde a preguntas como qué pasará y por qué. Debido a la creciente importancia que se ha dado a la minería de datos, esta ha sido señalada por el Instituto Tecnológico de Massachussets (MIT, 2001) como una de las diez tecnologías emergentes más importantes del siglo 21 que cambiará el sentido de la investigación en el mundo. En efecto, una realidad actual de la minería de datos es su papel de apoyo para explorar, analizar, comprender y aplicar el conocimiento adquirido de grandes volúmenes de datos, así como identificación de tendencias que sirven de ayuda para la toma de decisiones.

La minería de datos es una combinación de bases de datos y tecnologías de inteligencia artificial, también lo definen como el proceso de planteamiento de búsqueda y extracción de patrones, a menudo desconocido de grandes cantidades de datos mediante comparación de patrones y otras técnicas de razonamiento, a esta definición apoya la empresa KASPERU (2010) indicando que la Minería de Datos es el proceso de descubrir conocimiento desde los datos, mediante un proceso de extracción no trivial de información implícita, previamente desconocida y potencialmente útil. También indica que es el conjunto de técnicas para el análisis de los datos y el descubrimiento de patrones escondidos y concluye indicando que es el conocimiento representado mediante patrones o modelos [6].

3.1. Modelos

Según Kasperu (2010), un modelo es un intento de entender algún aspecto de la realidad. También indica es un Conocimiento, un modelo. Que intenta representar algún aspecto del mundo y explicar su comportamiento, que permite pasar de la observación a la teoría; se construyen para ser transmitido.

Modelos y tareas de la Minería de Datos. Los modelos empleados, en la minería de datos, son el descriptivo y el predictivo:

En el modelo descriptivo, se identifican patrones que describen los datos mediante tarea, agrupamiento y reglas de asociación. Mediante este modelo se identifican patrones que explican o resumen el conjunto de datos, siendo estos útiles para explorar las propiedades de los datos examinados.

En el modelo predictivo se pretende estimar valores futuros o desconocidos de variables de interés.

3.2. Aprendizaje de máquina

El aprendizaje de máquina es un subcampo general de la inteligencia artificial (IA) que se enfoca en el diseño y desarrollo de algoritmos y técnicas, que permiten a los computadores aprender y mejorar su desempeño en el tiempo, basado en información que proviene de bases de datos.

El enfoque principal en la investigación del aprendizaje de maquina es extraer información de los datos automáticamente, usando métodos estadísticos y computacionales y su vez generar modelos que se basan en reglas o patrones. Por lo tanto, el aprendizaje de máquina no solo se relaciona con la minería de datos, reconocimiento de patrones o estadísticas, sino que también se considera como una parte de la informática teórica [9].

En años recientes, el aprendizaje de maquina ha sido aplicado a una amplia variedad de aplicaciones, las cuales se incluyen: procesamiento natural del lenguaje, motores de

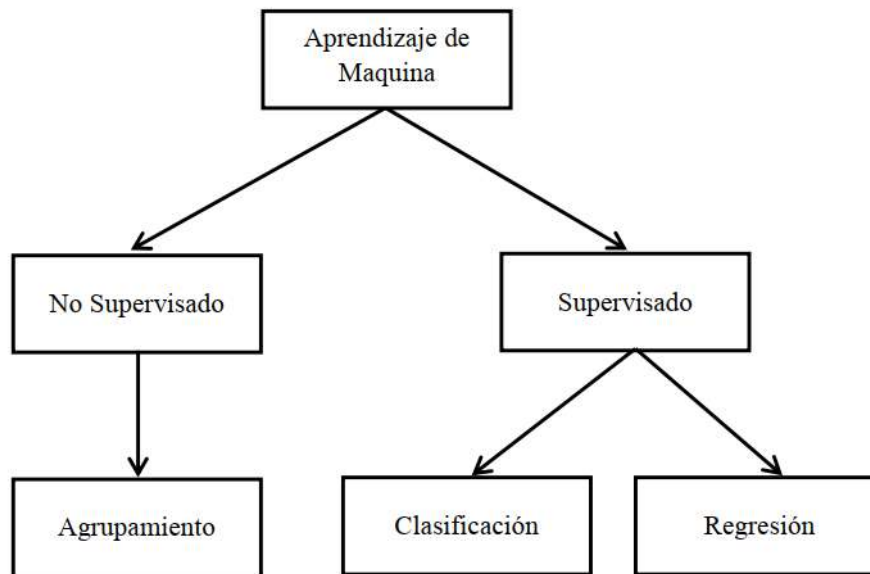


Figura 3.1: División de los métodos en aprendizaje de maquina

búsqueda ,diagnósticos médicos, microinformática, detección del fraude en tarjetas de crédito, clasificación de secuencias de ADN, reconocimiento de rostros y movimiento de autómatas (robots). Los algoritmos utilizados en aprendizaje de maquina se organizan de acuerdo al tipo de entradas disponibles o al resultado que se desea. Adicionalmente los métodos utilizados se diferencian uno del otro según como se use el conocimiento obtenido, ya sea para agrupar, clasificar o realizar algún tipo de regresión figura 3.1.

Parte II

Resumen Ejecutivo

Capítulo 4

Resumen Ejecutivo

4.1. Descripción del trabajo

En la presente investigación se realiza un análisis intensivo de datos, específicamente de las facturaciones de la empresa de concesión de energía eléctrica ANDE, por otra parte una forma de atacar el hurto de energía, se plantea un método de trabajo para la administración de lecturas y utilización de equipos amigables a fin de llevar un balance energético por zona de distribución.

4.1.1. Métodos y Técnicas utilizadas

Se adopta una metodología con enfoque descriptivo pues se describen los procesos que se contemplan en la empresa sobre las facturaciones de la energía, se utilizó un método numérico para la realización del análisis exploratorio de los datos.

Se recabaron datos de facturación de energía de la empresa, específicamente un histórico de 37 meses, específicamente clientes de la ciudad de Coronel Oviedo, se escogió esta muestra del departamento de Caaguazú, estos datos se limpiaron de forma a quitar los valores inexistentes o nulos, en todo caso fueron reemplazados por valores medios. Posteriormente se procedió a la evaluación del set de datos, utilizando la herramienta de programación

Python, tanto para la limpieza como el análisis exploratorio.

Se escogió como muestra un alimentador de la ciudad de Coronel Oviedo, el COV-1, sobre este alimentador se procedió a evaluar un balance energético del mismo, realizando mediciones individuales de todos los clientes conectados, para finalmente proponer un proyecto de mejora del sistema de lecturas, a fin de disminuir las pérdidas no técnicas.

4.2. Justificación

La empresa ANDE gestiona la energía a nivel país y esta es entregada a los usuarios en distintos niveles de tensión, en esta interacción se invierten recursos, los cuales se ahorrarían al disminuir las pérdidas, traduciéndose en beneficios financieros para la empresa. Es considerado que la reducción de pérdidas de energía es un impulsor de la Red Eléctrica Inteligente, que tiene relación con otros factores de gran relevancia tales como la continuidad y calidad del servicio, la seguridad en el suministro y la disminución del impacto en el medio ambiente.

Con la realización de este trabajo, ANDE obtendrá como beneficio directo, reducir los índices de pérdidas no-técnicas en el departamento de Caaguazú, disminuyendo así los casos en donde se utiliza de manera fraudulenta la energía, además la política de la empresa en cuanto a la distribución y comercialización de la energía mejorará pues se establecerán sistemas de monitoreo y control del flujo de la energía. Por otra parte, se logrará el fortalecimiento de la imagen institucional como empresa eficiente y el fomento al ahorro de energía. El usuario también obtendrá beneficios pues la empresa proveerá una mejor atención en cuanto a las actividades relacionadas con la toma de lecturas, corte, re-conexión y atención de reclamos.

Con la conclusión de este trabajo se logran métodos y modelos para el análisis de datos de facturación de clientes, esto puede ser implementado en otras ciudades, inclusive en todo el país.

4.3. Finalidad del proyecto

Con el desarrollo de este Proyecto se busca elaborar sistemas para el control de pérdidas no técnicas de la ANDE.

Con metodologías que impliquen bajo costo para la empresa, por esa razón se decidió enfocar el trabajo en el análisis de datos y metodologías de trabajo del personal de la empresa encargada de las lecturas y las mediciones.

4.4. Metas

Analizar al menos dos alternativas de sistemas de control de pérdidas no técnicas de la ANDE.

4.5. Objetivos

4.5.1. Objetivo general

Elaborar alternativas de sistemas de control de pérdidas no-técnicas en el sistema de distribución de la ANDE en el Departamento de Caaguazú.

4.5.2. Objetivos específicos

- Identificar y analizar las prácticas fraudulentas realizadas por los usuarios de la red eléctrica de la ANDE.
- Seleccionar tecnologías para control y detección de pérdidas no técnicas.
- Proyectar la adecuación del sistema de distribución a las tecnologías seleccionadas.
- Evaluar técnica y económicamente las alternativas propuestas.

4.6. Beneficiarios

El beneficiario directo de esta investigación es la empresa proveedora de energía eléctrica ANDE. Indirectamente los usuarios, pues mejorando las recaudaciones de la empresa esta puede proveer un mejor servicio a su clientela.

4.7. Producto

Metodologías de trabajos y sistemas abocados al control de pérdidas no técnicas.

4.8. Especificaciones de actividades y tareas realizadas

En este Proyecto Final de Grado se llevaron a cabo las siguientes actividades:

- Revisión de materiales bibliográficos, artículos y trabajos científicos para la selección de materiales a estudiar.
- Relevamiento de datos de facturaciones de energía de la ciudad de Coronel Oviedo.
- Limpieza de los datos relevados.
- Análisis exploratorio de los datos, a fin de encontrar clientes sospechosos de fraude.
- Verificación de clientes sospechosos.
- Realización de una metodología de trabajo e implementación de herramientas para la mitigación de fraudes.
- Ponderación financiera de las alternativas.

4.9. Recursos necesarios

4.9.1. Recursos humanos

Para la realización de esta investigación fue de vital importancia la colaboración de asesores científicos y técnicos calificados, tales como:

- Personal especializado en análisis extensivo de grandes datos.
- Personal especializado en elaboración de proyectos.

4.9.2. Recursos materiales

Para la ejecución del presente proyecto final de grado fueron necesarios los siguientes materiales:

- Biblioteca virtual.
- Normas y especificaciones técnicas.
- Computadora.
- Sistema Open.

4.10. Factibilidad económica

Se realizó una evaluación financiera de las alternativas elaboradas en este proyecto, calculando los costos operativos y materiales.

4.10.1. Beneficios

Los principales beneficios de este proyecto es la reducción de perdidas no-técnicas, una de estas alternativas es el análisis de facturaciones, en este aspecto se estima una reducción de clientes fraudulentos en 8,5 % como mínimo, no se puede aplicar un relacionamiento de manera lineal a las facturaciones de la ciudad, sin embargo, se puede estimar un beneficio monetario suponiendo el mínimo aumento de recaudación, por ejemplo un aumento del 2 %, seria de 1.545.167.465 Gs. (año 2021) el beneficio si se logra aumentar la recaudación en la ciudad de Coronel Oviedo en el porcentaje mencionado, lo cual es muy plausible, en la tabla 4.1 se aprecian los diferentes escenarios posibles.

Aumento de recaudación	Porcentaje
386.291.866	0.50 %
772.583.732	1 %
1.545.167.465	2 %
2.317.751.197	3 %

Tabla 4.1: Posible aumento de recaudación

En cuanto a la alternativa de implementación de métodos y herramientas de mitigación de pérdidas no técnicas, se puede estimar una reducción considerable de pérdidas no técnicas.

Como se tomo el alimentador COV-1 para este efecto se consideran los beneficios a obtener sobre el mismo. Para ello se calcula el promedio de la energía no factura y poder estimar lo que representa en guaraníes (4.1).

$$\bar{E} = \frac{\sum_{i=1}^n Ep_i}{n} \quad (4.1)$$

Considerando que la mayoría de los usuarios están conectados en baja tensión y corresponde a tarifa residencial con una faja de consumo de 0 – 500 kWh se estima un monto de 405 guaraníes por kilovatio-hora, se obtiene un beneficio por mes de **8.295.774 Gs.** apéndice B.1.

4.10.2. Costos

En la alternativa de análisis de datos, los costos son por demasía menores a los posibles beneficios, se pueden considerar como gastos operativos lo expresado en la tabla 4.2.

Concepto	Costo
Analista de datos (15 días)	5,000,000
Inspección de sospechosos	16,450,000
Total	21,450,000

Tabla 4.2: Costos de análisis e inspección de sospechosos

Costos reducción de pérdidas por control de suministros clandestinos:

Para determinar los costos de este apartado, el estudio se centra en los costos operativos, ya que para este método no se requiere de una inversión inicial.

Objeto de Gasto	Monto unitario	Cantidad	Subtotal
Salario de Personal	532,000	2	1,064,000
Movilidad	1,000,000	1	1,000,000
Equipos & herramientas	1,300,000	1	1,300,000
Gastos administrativos & imprevistos	1,925,400	1	1,925,400
Aspectos legales & misceláneos	1,500,000	1	1,500,000
Totales			6,789,400

Tabla 4.3: Costo operativos del primer mes

En la tabla 4.3, se puede observar los costos que conlleva la operativa de intervención de suministros clandestinos en el primer mes, teniendo en cuenta que para esta operativa será necesaria una cuadrilla con dos funcionarios, un vehículo para movilidad, equipos de protección personal y colectiva, herramientas entre las cuales se puede citar escalera, pinzas, multímetros, etc. También se introduce gastos legales como intervenciones, notificaciones, inicio de expedientes judiciales, etc. Y los gastos administrativos que conlleva cualquier tipo de trabajo. Ahora bien, a partir del segundo mes los gastos se reducen pues la operativa en si es para la revisión de re-incidentes, descartando la adquisición de herramientas y reduciendo los gastos legales a un 20 %. De esta manera se puede considerar del segundo al doceavo mes, costo de 3,757,400 Gs por mes, totalizando **51,878,200 Gs** al año de gastos.

Costos de reducción de pérdidas por medio de medición off-line

Para determinar los costos de este modelo para reducción de pérdidas es necesario costear los materiales necesarios para realizar la medición en PD's (totalizadores), la mano de obra requerida para el montaje de los mismo y por ultimo la operativa de costos para intervención, que se estima un 25 % del costo total de reducción de pérdidas por intervención de suministros clandestinos.

Una vez revisado los valores del apéndice B.2, se resume los costos totales de la implementación de medición off-line en la tabla 4.4.

Descripción	Costo unitario	Cantidad	Subtotales
Medición totalizador PD de 25 KVA	2.871.022	3	8.613.066
Medición totalizador PD de 100 KVA	3.350.930	3	10.052.790
Totales			18.665.856

Tabla 4.4: Costo de implementación de medición off-line

Se puede observar en la tabla 4.4, los costos de implementación o inversión requeridos para este método de reducción de pérdidas, si bien es efectivo para combatir la reducción de pérdidas no técnicas no lo es para obtener valores de curva de demanda, además los datos obtenidos deben ser provistos por un lector de distribución. También hay que considerar al momento de analizar la rentabilidad de la operativa, los costos de intervención que representaría un 25 % de la primera metodología, ya que al obtener datos se vuelve más eficiente las intervenciones.

Costos de reducción de pérdidas por medición on-line

Al igual que el método off-line es necesario costear los montos para la inversión, que de manera similar se tiene en cuenta los materiales y aparte de ello los costos de licencia para la adquisición de datos.

La diferencia en la implementación consiste en la utilización de módulos de comunicación GSM-LTE, por lo que la varianza de precios se dará en este ítem del planteamiento anterior. Entonces, la diferencia de precio en los equipos de medición de acuerdo a la

capacidad del PD se muestran en la tabla 4.5.

Descripción	Costo unitario	Cantidad	Subtotales
Medición totalizador PD de 25 KVA	4.171.022	3	12.513.066
Medición totalizador PD de 100 KVA	4.747.280	3	14.241.840
Totales			26.754.906

Tabla 4.5: Costo de implementación de la medición on-line

4.10.3. Análisis financiero

Una vez completado y estimado los costos y beneficios, se procedió a realizar el análisis de rentabilidad de las opciones de sistemas de control de pérdidas no técnicas planteadas en este trabajo.

En cuanto al control por medio de revisiones aleatorias de clientes se obtiene lo siguiente; en cuanto al VAN se estima en 131.422.276 Gs, el TIR 87,6 % y la relación beneficio-costos 1,506.

En la opción de medición off-line se tiene como VAN 150.796.829 Gs, el TIR de 83,84 % y la relación beneficio-costos 1,627.

La opción de medición on-line llega a un VAN de 88.607.644 Gs, el TIR de 46,24 % y la relación beneficio-costos 1,293. Véase apéndice B.3.

Parte III

Ingeniería de Diseño

Capítulo 5

Análisis de datos

En este capítulo se desarrolla un análisis de datos de facturaciones, en cuanto a la muestra, del departamento de Caaguazú se utilizaron los datos de Coronel Oviedo.

Se relevaron los datos, se limpiaron de acuerdo a criterios establecidos y posteriormente se procedió a realizar un análisis exploratorio buscando clientes sospechosos.

5.1. Relevamiento de datos

La empresa ANDE cuenta con varias gerencias y direcciones encargadas de las funciones principales de la compañía en generación, transmisión, distribución y comercialización de energía eléctrica. La gerencia comercial se encarga del ultimo punto mencionado, en sus diferentes departamentos se tiene el departamento de estudios comerciales en donde se manejan las bases de datos sobre consumos, facturaciones y características de los clientes.

Se consiguieron los datos de la ciudad Coronel Oviedo, desde 01-01-2019 hasta 01-01-2022, total de 37 meses. 34667 clientes, de esta manera se genero una tabla de 42 columnas por 34667 filas, en esta tabla se agregaron los campos de identificación de cliente, el máximo, el mínimo, media y dispersión de cada conjunto.

Como se puede ver en la figura 5.1, se extrajeron del sistema de gestión OPEN, los

	A	B	C	D	E	F
1	NIS_R	CONSUMO_	CONSUMI	COD_T	F_FAC	PC
2		29,000	135	142	03/02/2020	2200
3		14,000	97	142	02/03/2020	2200
4		23,580	103	142	01/04/2020	2200
5		23,195	101	142	04/05/2020	2200
6		68,278	214	142	01/06/2020	2200
7		84,258	267	142	01/07/2020	2200
8		198,319	438	142	03/08/2020	2200
9		172,556	380	142	01/09/2020	2200
10		238,019	508	142	01/10/2020	2200
11		103,000	243	142	02/11/2020	2200
12		35,000	77	142	01/12/2020	2200
13		22,000	57	142	04/01/2021	2200
14		51,000	104	142	01/02/2021	2200
15		41,000	92	142	02/03/2021	2200
16		52,000	118	142	05/04/2021	2200
17		44,000	101	142	03/05/2021	2200
18		40,000	90	142	01/06/2021	2200
19		51,000	117	142	01/07/2021	2200
20		38,000	83	142	02/08/2021	2200
21		38,000	83	142	01/09/2021	2200
22		46,000	106	142	01/10/2021	2200
23		54,000	124	142	01/11/2021	2200
24		63,000	149	142	01/12/2021	2200
25		57,000	133	142	03/01/2022	2200
26		39,000	195	141	03/02/2020	2200
27		5,000	51	141	02/03/2020	2200
28		5,195	54	141	01/04/2020	2200
29		83,212	276	141	04/05/2020	2200
30		20,014	104	141	01/06/2020	2200
31		24,825	129	141	01/07/2020	2200
32		22,515	117	141	03/08/2020	2200
33		21,554	112	141	01/09/2020	2200

Figura 5.1: Datos en bruto del sistema de gestión de facturas ANDE

clientes pertenecientes a Coronel Oviedo, con si respectivo NIS (identificador), consumo en moneda local, consumo en kWh, código del tipo de cliente, fecha de facturación y potencia contratada. Se consolido en una sola base de datos en el rango de fecha mencionado para su mejor manipulación en análisis.

5.2. Limpieza de datos

Una vez recabados los datos, tiene valores faltantes, se procedió a la limpieza de datos, el cual consistió en lo siguiente:

- En caso de que el cliente no sobrepase 4 facturaciones faltantes o de valor igual a cero, se reemplazó con la media móvil de las facturaciones adyacentes.
- Se descartó al cliente con faltantes mayores a 4 registros de facturaciones.

Se utilizó para este caso particular el entorno de Python 3.10.4. en el apéndice A.1.1 se revisa el código utilizado para limpieza de datos.

De esta manera aplicando el primer filtro mencionado, el dataset quedo con un rango de 16114 históricos de clientes, reduciéndose en 53,5 % el dataset original.

Una vez descartado las entradas mencionadas, se reemplazo los valores faltantes y ceros con media de 8 periodos adyacentes. En la figura 5.2, se observa el dataset posterior a esta limpieza.

En cuanto a la categoría de clientes la segmentación se aprecia claramente (tabla 5.1). Esta categorización se estipula en [2]. La mayor cantidad usuarios se concentran en la categoría social, residencial e industrial.

En la figura 5.3, se aprecian los consumos de todos estos clientes, como se esperaba existe una concentración por debajo de los 5000 kWh, cabe destacar que la media de todos los consumos mostrados es de 627.46 kWh.

Luego se procedió a revisar, cada estrato de consumo, y en cada uno se verificaron las

	NIS	COD_TAR	POT	1-2019	2-2019	3-2019	4-2019	5-2019	6-2019	7-2019	...	4-2021	5-2021	6-2021	7-2021	8-2021	9-2021	10-2021	11-2021	12-2021	1-2022
0		142	2200.0	99.000	121.000	131.000	110.0	316.0	89.0	96.0	...	118.0	101.0	90.0	117.0	83.0	83.0	106.0	124.0	149.0	133.0
1		141	2200.0	194.000	200.000	162.000	166.0	159.0	120.0	113.0	...	248.0	202.0	198.0	246.0	245.0	254.0	279.0	236.0	248.0	355.0
2		142	2200.0	458.000	514.000	212.000	11.0	183.0	186.0	139.0	...	109.0	242.0	301.0	298.0	298.0	267.0	267.0	266.0	272.0	286.0
3		141	3300.0	462.000	452.000	160.000	117.0	128.0	143.0	146.0	...	318.0	173.0	228.0	177.0	203.0	259.0	321.0	21.0	24.0	45.0
4		142	2200.0	225.000	233.000	366.000	366.0	311.0	331.0	294.0	...	162.0	136.0	205.0	233.0	234.0	222.0	254.0	246.0	166.0	172.0
...	
16109		142	3300.0	73.750	85.625	110.875	62.0	189.0	146.0	89.0	...	132.0	96.0	110.0	103.0	133.0	122.0	165.0	150.0	189.0	219.0
16110		142	3300.0	45.250	52.125	60.250	43.0	89.0	63.0	85.0	...	239.0	124.0	95.0	106.0	118.0	128.0	175.0	156.0	248.0	320.0
16111		142	3520.0	66.750	77.250	98.125	157.0	111.0	52.0	119.0	...	140.0	83.0	74.0	82.0	70.0	107.0	116.0	134.0	213.0	309.0
16112		142	7040.0	76.875	95.500	120.875	73.0	196.0	109.0	118.0	...	290.0	207.0	157.0	131.0	246.0	114.0	246.0	246.0	323.0	415.0
16113		142	5500.0	147.750	174.375	199.250	74.0	487.0	357.0	173.0	...	171.0	180.0	264.0	374.0	256.0	195.0	156.0	139.0	222.0	310.0

Figura 5.2: Datos limpios y procesados

Categoría	Cantidad
142	10593
141	3204
410	2049
413	134
846	104
833	11
412	9
832	8
732	1
372	1

Tabla 5.1: Categorías de clientes según la segmentación encontrada

potencias contratadas, por ejemplo, la categoría 410 (Otro grupo de consumo, comercial), existen contratos desde 2200 W hasta 43810 W, a modo de reducir el análisis se filtraron los datos a este grupo dentro del contrato de 19800 W (3x30 A) con 171 registros históricos.

5.3. Análisis exploratorio

A fin de mostrar la metodología utilizada se procesaron los datos filtrados en la sección anterior, dentro del estrato comercial, específicamente los usuarios con una llave limitadora de 3x30 A, 171 registros fueron considerados, además se generaron nuevos campos:

- Media de consumo por registro.
- Mínimo consumo por registro.

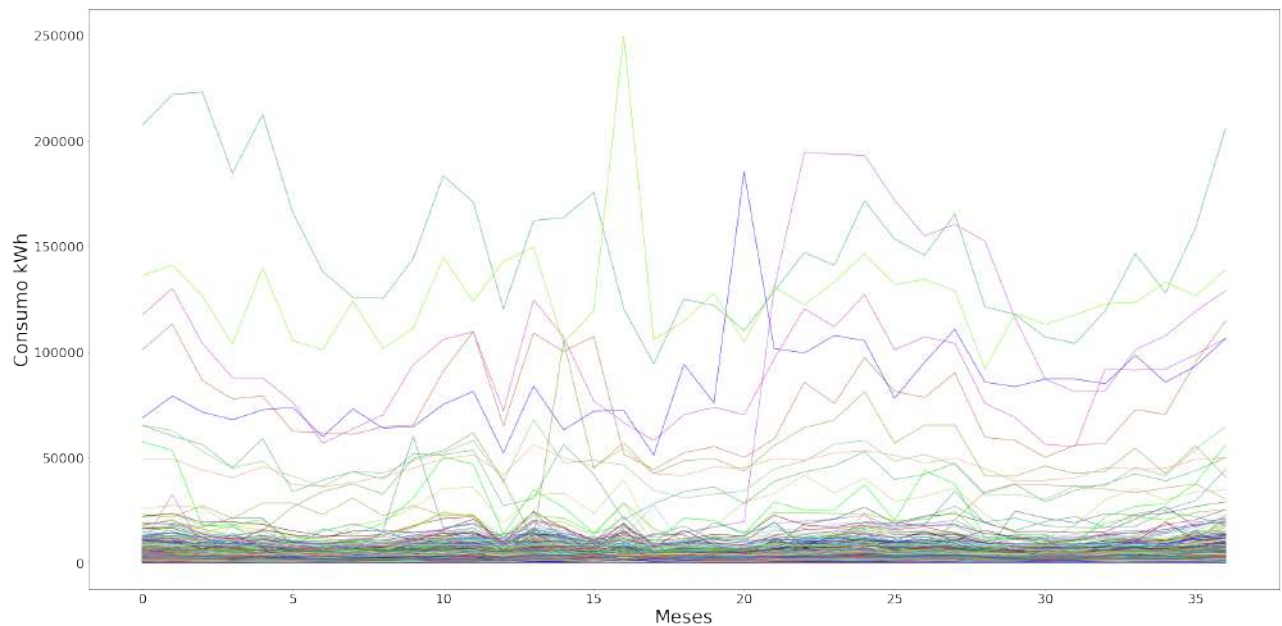


Figura 5.3: Consumos de clientes en general (kWh)

- Máximo consumo por registro.
- Desviación standart por registro.
- Relación de media con el contrato por registro.

La maquina utilizada para dicho proceso tiene las siguientes especificaciones: AMD ryzen 7 3.5 GHZ y 16 GB de ram.

Una vez obtenido estos campos, se procedió a ilustrar un gráfico de cajas sobre el campo que relaciona la media con el contrato. Se observa una cantidad considerable de datos atípicos figura 5.4, por lo que se retiraron del set de datos dichos valores.

En la figura 5.5, se observan los histogramas relacionales entre la cantidad en el rango con la relación a media, quitando los valores atípicos la distribución de datos mejora por la concentración y forma tipo gaussiana del histograma.

En este caso se reduce la cantidad de datos a 142 registros históricos, se crearon dos vectores de igual dimensión representando la desviación standart y la relación media-contrato, considerando estos como cartesianos se elaboró una regresión lineal en busca de sospechosos, utilizando la librería Sklearn (herramienta de machine learning) se crea

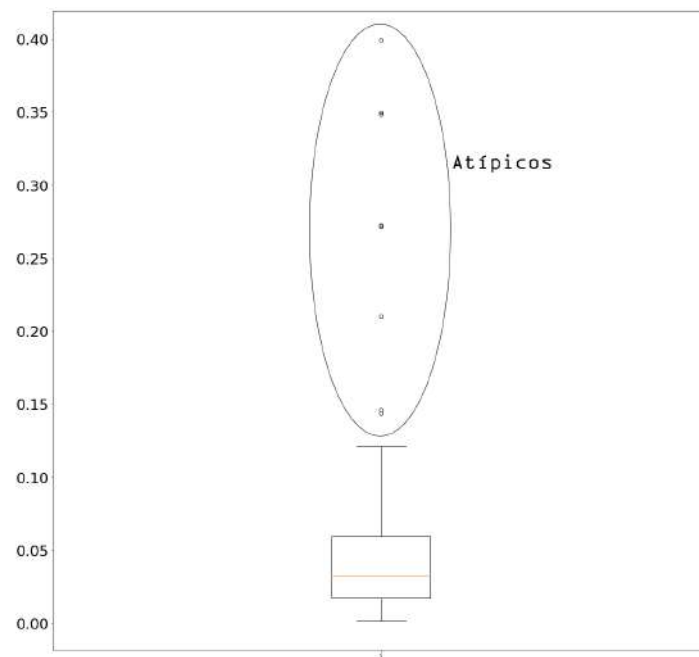
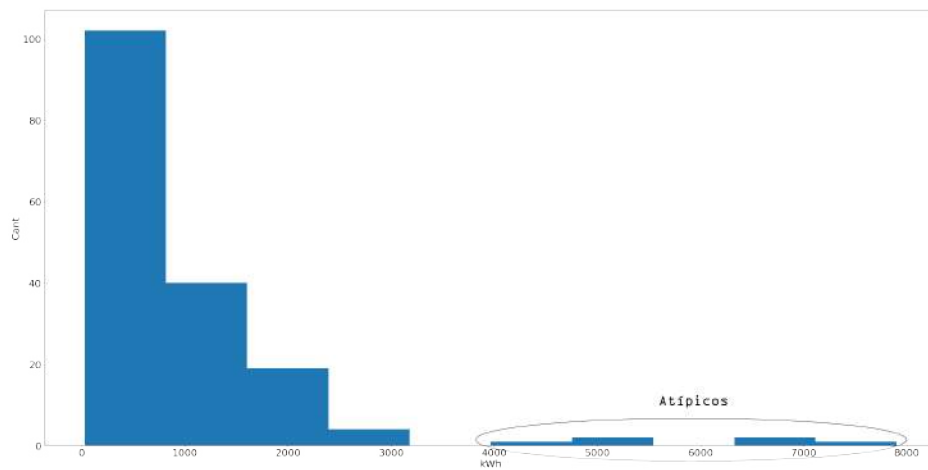


Figura 5.4: Grafica de cajas, relación media con contrato

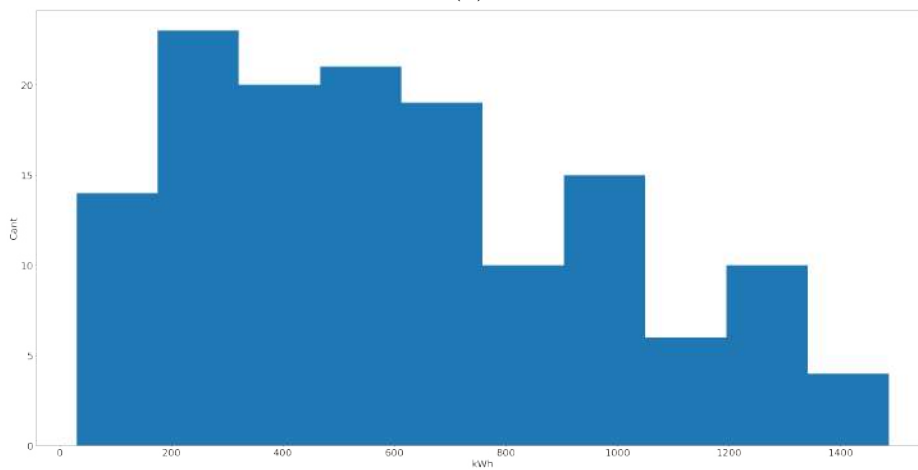
esta línea en donde la misma representa el promedio de desviación y media-contrato. Entonces se genera la siguiente interrogante ¿Quiénes podrían ser usuarios sospechosos? según un escrutinio a clientes intervenidos por fraude hallados en el sistema de gestión de la empresa, en sus históricos facturados se observaron en la relación media-contrato valores bajos y en cuanto a la dispersión este es alta por lo que se agrupa un conjunto que supera la desviación en 250 kWh y la relación media-contrato por debajo de la regresión (figura 5.6).

En cuanto a los usuarios fraudulentos hallados en el sistema, estos apenas representan el 1,88% de toda la base de datos de la ciudad de Coronel Oviedo, este valor se redujo aún más, pues se compararon con la base de datos filtrada para poder generar los puntos en la representación dispersión y media relacional, en esta comparación se encontró baja densidad en el grupo seleccionado en este trabajo como sospechosos, por lo que el universo de búsqueda se vuelve menor, énfasis de este estudio (figura A.17).

Se reduce la lista a 26 usuarios, de esta lista según revisión en el sistema de gestión de la ANDE, se encontraron 46% de ellos con incidencias, esto implica cambio de medidor, medidores averiados o revisiones del mismo que llevaron a precinta por ocasional evento) y dos de ellos con fraude e intervención.



(a)



(b)

Figura 5.5: Histograma a) Set de datos completo b) Sin valores atípicos

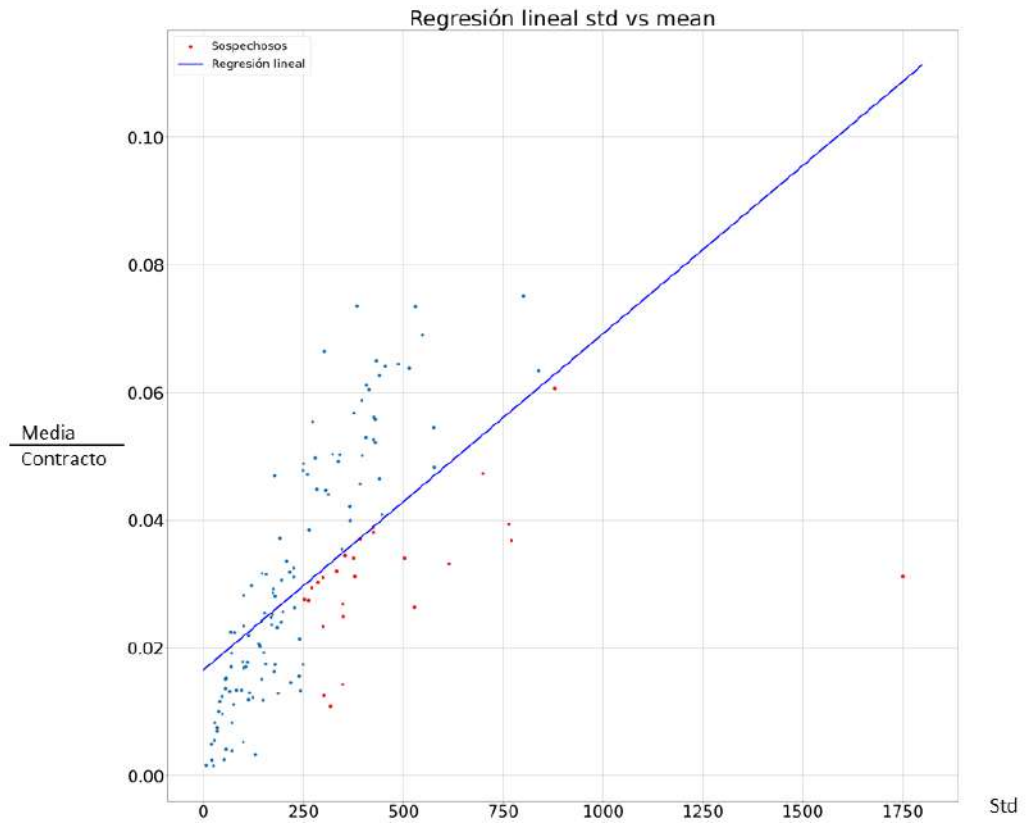


Figura 5.6: Regresión lineal y clientes sospechosos

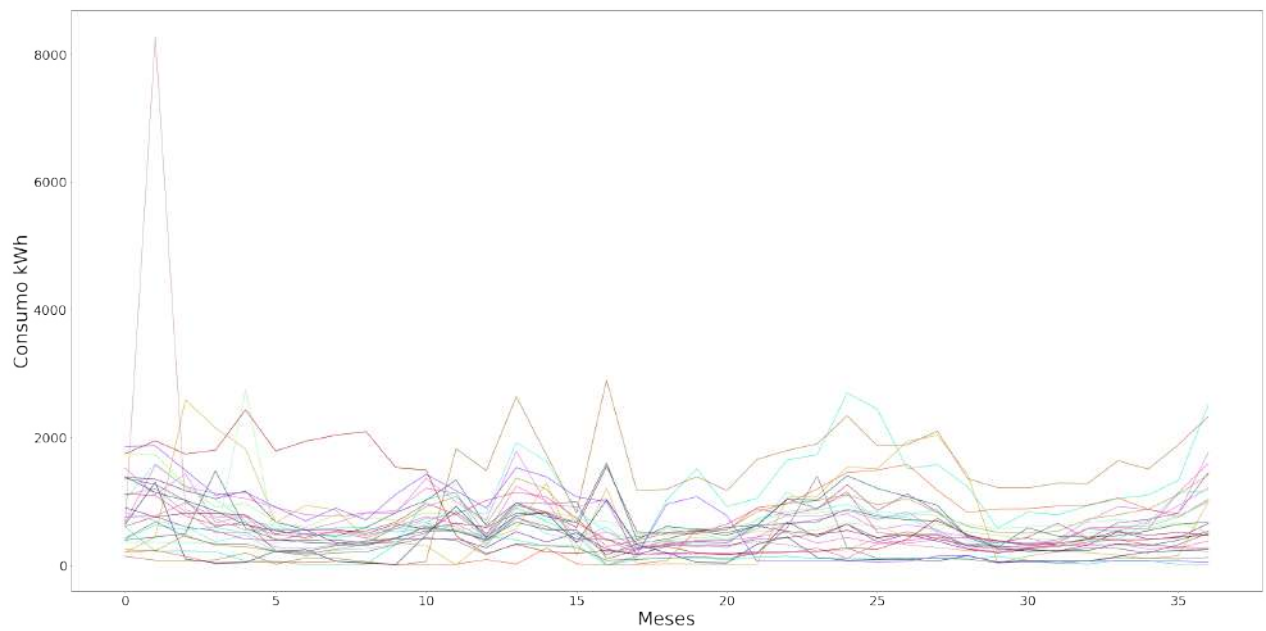


Figura 5.7: Registro histórico de clientes sospechosos

Se repitió este análisis para una segmentación seleccionada por ser de mayor relevancia para el estudio específicamente de mayor consumo y mayor flujo de pago (tabla 5.2).

El segmento analizado el cual se mostró en esta sección es la primera fila de la tabla 5.2, en apéndice A.1.2 se pueden verificar cada revisión de segmento. En todos los casos se aplicó la secuencia de códigos para reducir los sospechosos de fraude, posteriormente se revisó en cada segmento las incidencias y las intervenciones por presunto fraude. No se presentan las verificaciones en el sistema de la empresa para resguardar datos de los clientes.

Categoría	Contrato (kW)	Cant. Clientes	Cant. Sospechosos
410	19800	171	26
410	13200	154	24
410	6600	276	19
142	7040	420	22
142	19800	407	64
142	13200	339	60
142	21120	331	60
142	16500	218	20
142	26400	146	34
Total clientes explorados:		2462	329

Tabla 5.2: Selección de clientes analizados

La categoría 410 corresponde al estrato comercial y otros, mientras el identificador 142 corresponde al estrato residencial, utilizando el algoritmo de exploración se logrará reducir la inspección de clientes a 13% del total. En este estudio se realiza un escrutinio de este grupo reducido dentro del sistema de gestión OPEN, buscando intervenciones e incidencias de estos usuarios.

Finalmente se encontraron los siguientes resultados en el resto de clientes sospechosos:

- En la categoría 410, con potencia de contrato de 13200 kW, se encontraron de los 24 sospechosos, 5 con verificación por medidor tumbado, 2 intervenciones por fraude, 1 precintado de medidor.
- En la categoría 410, con potencia de contrato de 6600 kW, se encontraron de los 19

sospechosos, 5 usuarios con incidencias en donde se verificaron por personal de la empresa y se realizaron precintado de medidor, 1 intervención por fraude.

- En la categoría 142, con potencia de contrato de 7040 kW, se encontraron de los 22 sospechosos, 6 medidores tumbados (no gira con carga), 1 medidor obsoleto y 1 fraude.
- En la categoría 142, con contrato de 19800 kW, se encontraron de los 64 sospechosos, 18 con incidencias en donde se han precintado sus medidores, 5 fraudes.
- En la categoría 142, con contrato de 13200 kW, se encontraron de los 60 sospechosos, 18 con incidencias.
- En la categoría 142, con contrato de 21120 kW, de los 60 sospechosos con incidencias 21 usuarios y 1 fraude.
- En la categoría 142, con contrato de 16500 kW, de los 20 usuarios sospechosos, 5 con incidencias y 2 fraudes.
- En la categoría 142, con contrato de 26400 kW, de los 34 usuarios sospechosos, 12 con incidencias y 2 fraudulentos.

En el caso del medidor tumbado se considera como fraude, pues es una forma de no registrar la medición a facturar, este es un hecho común entre los usuarios que aun cuentan con el medidor de disco giratorio.

Estos valores, 28 usuarios fraudulentos, es solo de los trabajos realizados actualmente por la ANDE, donde según entrevistas realizadas, los usuarios sospechosos son proporcionados por los propios lectores, o por los recorridos del departamento de intervenciones de la empresa.

Estos usuarios pueden ser geo-referenciados por la empresa a fin de mejorar los recorridos realizados y optimizar de esta forma la utilización de recursos. Al realizar las verificaciones in situ, el índice de acierto del método aplicado puede mejorar hasta un 16 % [12].

Capítulo 6

Monitoreo de transformadores y balance energético

En este capítulo se desarrolla una metodología a fin de proyectar un sistema para monitorear los consumos por transformadores de distribución, esto con el objeto de contrastar las medidas obtenidas con las lecturas realizadas a los medidores de los clientes conectados a la red de la empresa.

Se utilizaron en este proyecto software y hardware en versión demo para las simulaciones preliminares de las alternativas propuestas a la problemática planteada.

Se llevaron a cabo simulaciones de tipo experimental, utilizando equipos de medición de alta precisión.

6.1. Alimentador Coronel Oviedo 1 (COV-1)

En este apartado se definen y cuantifican las pérdidas tanto técnicas, como las no técnicas enfatizando estas. Para ello se toma como población toda la zona de influencia del alimentador Coronel Oviedo 1, desde aquí más denominado COV-1, cuyos datos fueron obtenidos en campo y también provistos por la ANDE.

El alimentador mencionado cuenta con las siguientes características:

- 1.628 m de recorrido.
- Salida subterránea de 240 mm² (40m).
- Troncal de 150 mm² (900m).
- Derivación 35mm² (688m).
- Puestos de distribución.
- 3 puestos de distribución de 100 KVA.
- 2 puestos de distribución de 25 KVA.
- 1 puesto de distribución para iluminación pública de 25 KVA (87 Equipos).
- 1 puesto de distribución exclusivo de 45 KVA.
- Usuarios.
- 109 usuarios en BT.
- 1 usuario en MT.
- 24 usuarios en conexión clandestina.

En la figura 6.1 se puede observar el recorrido del alimentador en estudio, que parte desde la estación Coronel Oviedo.

6.2. Determinación de pérdidas técnicas

Para determinar las pérdidas técnicas el método utilizado es el método de flujo de potencia, el cual arroja valores de corriente, potencia, tensión y otros parámetros en todo el sistema o red de estudio y en distintos escenarios de operación posibles, las cuales para el estudio de consideraron máxima, media y mínima. En cuanto al modelo se elaboro con los datos de carga del COV-1, se puede apreciar en la figura A.19.

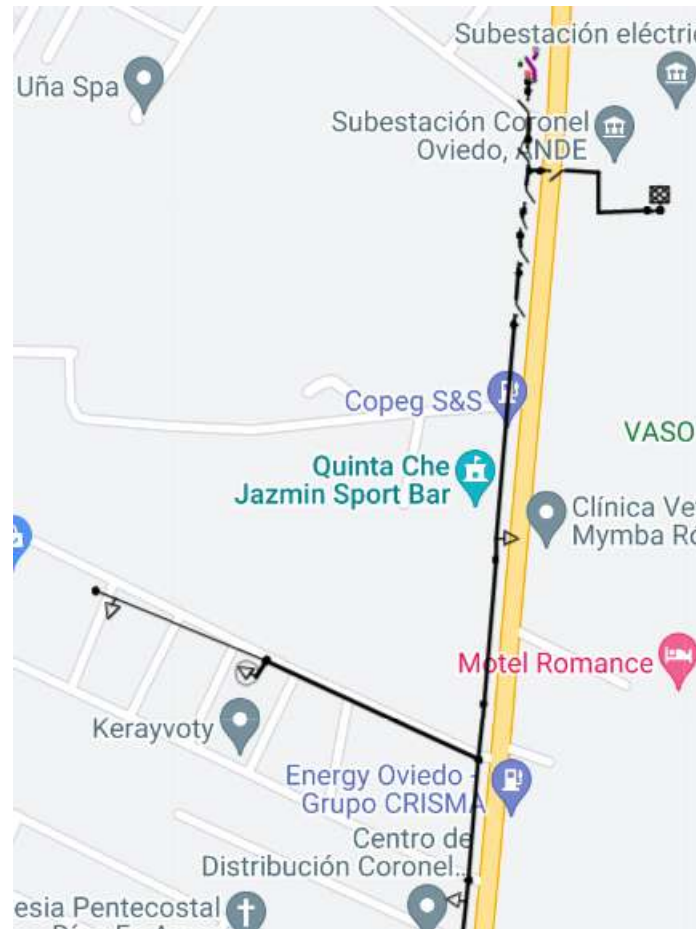


Figura 6.1: Zona de influencia del alimentador COV-1. Obtenido de Administración Nacional de Electricidad – Dpto de Operaciones 2022.



Figura 6.2: Ubicación geográfica de puestos de distribución con sus respectivos códigos de PD. Obtenido de Administración Nacional de Electricidad – Dpto de Operaciones 2022.

En la figura 6.3 se puede observar el reporte de pérdida de potencia el cual brinda la información para calcular la energía perdida durante el lapso de un mes teniendo en cuenta el ciclo de medición del equipo instalado en la cabecera del alimentador COV-1. Actualmente se encuentra instalado un equipo totalizador sobre la salida del alimentador COV-1, este brinda la medición de energía mensualmente.

					DIGSILENT PowerFactory 15.1.7		Project: Date: 12/26/2022		
Load Flow Calculation					Complete System Report: Voltage Profiles, Grid Interchange				
AC Load Flow, unbalanced, 3-phase (ABC)					Automatic Model Adaptation for Convergence				
Automatic Tap Adjust of Transformers					Max. Acceptable Load Flow Error for				
Consider Reactive Power Limits					Nodes				
					Model Equations				
					1,00 kVA				
					0,10 %				
Total System Summary					Study Case: Study Case			Annex: / 4	
Generation	Motor Load	Load	Compensation	External Infeed	Inter Area Flow	Total Losses	Load Losses	NoLoad Losses	
[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	[MW]/ [Mvar]	
\Admin\COV1\Network Model\Network Data\Grid									
0,00	0,00	0,23	0,00	0,24	0,00	0,01	0,00	0,01	
0,00	0,00	0,09	0,00	0,10	0,00	0,01	0,00	0,00	
Total:									
0,00	0,00	0,23	0,00	0,24	0,01		0,00	0,01	
0,00	0,00	0,09	0,00	0,10	0,01		0,00	0,00	

Figura 6.3: Resultados obtenidos tras flujo de carga en DigSilent

Resumiendo, el reporte proporciona un valor de pérdidas de 10 kW para condición de máxima demanda, dichas pérdidas son debidas al efecto joule en la distribución MT, pérdidas por transformación, armónicos presentes en la red, factor de potencia bajo, etc. Con el dato obtenido se procede a cuantificar la energía perdida durante el lapso de un mes por medio de la ecuación 6.1.

$$P_{tec} = (C_1 * h_1 + C_2 * h_2 + C_3 * h_3) * P_p * D \quad (6.1)$$

Donde:

P_p : Potencia pérdida en el sistema. (Kilovatios). h_1, h_2, h_3 : Horas correspondientes a mínima, media y máxima demanda. (Horas). D : Días correspondientes a un ciclo de facturación. C_1, C_2, C_3 : Constantes de modelos estadísticos.

En la sección A.2.1 del apéndice, se realizan los cálculos de estas pérdidas siendo de 1263,6 kWh-mes.

6.3. Determinación de pérdidas no técnicas

Las pérdidas no técnicas se calculan mediante la obtención de las pérdidas totales del sistema, dichas pérdidas no técnicas corresponden a errores en la medición, facturación y por «malas prácticas» de los consumidores como lo son las instalaciones no autorizadas, las alteraciones en los medidores de luz, siendo algunos de los más comunes.

Meses	Energía saliente ES_COV1 (KWh)	Energía facturada (KWh)	Energía estimada utilizada en APs (kwh)	Pérdidas técnicas (kwh)	Energía no facturada KWh	% perdidas técnicas + no técnicas
Mayo	60050.91	33609	4740	1263	20438.91	36.14 %
Junio	60385.45	38259	4740	1263	16123.45	28.79 %
Julio	64901.82	34011	4740	1263	24887.82	40.29 %

Tabla 6.1: Cálculo de pérdidas no técnicas.

Como se puede observar en la tabla 6.1, se tiene la energía entregada por el alimentador que es medida en la cabecera del COV-1, la energía facturada el cual se obtiene de los usuarios que son suministrados por el alimentador en estudio, la energía estimada de los alumbrados públicos que no cuentan con equipos de medición, las pérdidas técnicas resultantes del cálculo de la sección A.2.1, la energía no facturada que es la diferencia entre la energía saliente entre la facturada, estimada en AP's y pérdidas técnicas. El porcentaje de pérdidas totales el cual se halla por la ecuación

$$\%P_{tec} + P_{notec} = (E_S - E_f - E_{AP})/E_S \quad (6.2)$$

Donde:

E_S : es la energía saliente.

E_f : Es la energía facturada.

E_{AP} : es la energía estimada por uso de alumbrado público.

Mes	% Pérdidas técnicas	% Pérdidas no técnicas	Pérdidas totales en el sistema
Mayo	2,1	34,04	36,14
Junio	2,1	25,69	28,79
Julio	2,0	38,29	40,29

Tabla 6.2: Resumen de pérdidas

En la tabla 6.2 se puede observar que el valor de las pérdidas no técnicas es de mayor porcentaje respecto a las pérdidas técnicas, estas ultimas se distinguen como valores normales y acordes a la extensión y zona de influencia del COV-1.

6.4. Control de pérdidas

Para poder reducir las pérdidas se elaboran las siguientes alternativas, teniendo en cuenta que no se necesita tomar acciones sobre las pérdidas técnicas, en cambio si sobre las no técnicas.

6.4.1. Control de suministros clandestinos

El método consiste en obtener mayor control sobre los usuarios con conexiones clandestinas, mediante recorridos periódicos de las zonas donde se crea existe dicha actividad. Para llevar acabo esta operativa es necesario personal encargado del control, movilidad para el recorrido, de tal forma a generar de manera rápida una intervención por la vía correspondiente. Para esta operativa es necesaria:

- Móvil institucional.
- Personal capacitado.
- Otros gastos correspondientes a la operativa.

Esta forma operativa es una de las utilizadas actualmente por la ANDE, no obstante, aún no se ha obtenido resultados satisfactorios por las siguientes razones:

- Falta de acompañamiento para toma de acciones legales.
- Personal con falta de motivación para realizar trabajos de reducción de pérdidas.
- Cuestiones sociales.

La probabilidad de éxito de este método se cuantifica de la siguiente manera considerando que al menos el 95 % tiene una lectura correcta y que realicen el pago correspondiente. Teniendo como dato, 134 usuarios totales, 24 con conexiones directas.

$$P(X \geq 127) = \sum_k^{134} \binom{134}{k} 0,2^x (1 - 0,2)^{(134-x)}$$

$$P(X \geq 127) = 0,0001$$

El cual da un 0,00001 % de probabilidad de éxito.

6.4.2. Medición off-line a puestos de distribución

Este método consiste en instalar medidores totalizadores a la salida de los transformadores de Distribución, el cual es integrada a uno de los itinerarios de los lectores de distribución. Desde este punto se calcula de manera similar al apartado anterior tomando como energía saliente la suministrada por el medidor totalizador del PD y contrastar dicha lectura con la medición de energía ligado al puesto de distribución correspondiente.

Para realizar este esquema de medición es necesario contar con los materiales para la construcción del gabinete de medición del PD. Primeramente, hay que partir de la corriente nominal del transformador para dimensionar el transformador de corriente a instalar y conectar al medidor de energía, para ello se utiliza la expresión.

Primario .../ 5A	Modelo	Código	VA clase		
			0,5	1	3
50	CT20	41399	-	-	3
75	CT20	41400	-	2	3,5
100	CT20	41404	1,5	2,5	3,75
150	CT20	41406	2,5	3,5	5
200	CT30	41412	3,5	5	7,5
250	CT30	41414	5	7,5	10
300	CT30	41416	5	7,5	10
400	CT30	41418	5	7,5	10
500	CT50	41422	7,5	10	20
600	CT50	41424	10	15	25
800	CT50	41426	15	20	30
1000	CT50	41428	15	20	30

Tabla 6.3: Valores comerciales de transformadores de corriente. Fuente: Catalogo de TC's de ABB.

$$I = \frac{S}{(\sqrt{3} \cdot V)} \quad (6.3)$$

Donde:

I : es la corriente en amperios.

S : es la potencia nominal del transformador.

V : es el voltaje de línea del transformador en el secundario.

Entonces:

$$I = (100KVA)/(\sqrt{3} \cdot 0,4KV) = 151A$$

Por lo cual el transformador de corriente adoptado debe soportar al menos 151 amperios de corriente, no obstante, la situación in situ indica sobrecarga sobre algunos transformadores conllevando a la operación con corrientes mayores a la nominal, esto debido a motivos de falta de ajuste en la protección, etc. Entonces por razones de seguridad se adopta el inmediato superior el cual se presenta en la tabla 6.3.

En la tabla 6.4, se aprecia la lista materiales para el caso del transformador de 100 kVA trifásico uno de los mas comunes en la ciudad de Coronel Oviedo.

DESCRIPCIÓN	CANTIDAD
Conector a compresión AL 35/70 mm2	3
Medidor digital	1
Caja metálica 65x40x15cm. Para intemperie	1
Mano Francesa BT	2
Cable unifilar tipo comando y control 10x2,5	1,3
Cable NYY 70 mm2	9
TC 200/5	3
PG de 21	6
Varilla roscable 1/2 pulgada	0,5
Varilla roscable 3/8	0,1
Tuerca 3/8 x 3	4
Bulón Inox de 3/8X1”	2
Conector a compresión 2/8	4
Conector dentado 10-95	4
Conector dentado doble 25- 95/ 25-95	3
Arandela lisa redonda 3/8 Inoxidable	2
Terminal pre-aislado pino tipo Horquilla 2,5mm2	9
Abrazadera plástica tipo cintillo de150x3,6 MM	30
Abrazadera plástica tipo cintillo de 30 cm	10
Arandela lisa 1/2 pulgada	4
Tuerca ½	4
Perno con tuerca 3/16x3/4	7
Arandela lisa redonda 3/8 galvanizado	4
Terminal pre-aislado pino plano de 4-6mm	3
Arandela lisa redonda de 3/16”	7
Bornera tipo DIN de 10mm2	1

Tabla 6.4: Listado de materiales necesarios para realizar monitoreo off-line de puestos de distribución de 100 KVA.

En la figura 6.4, se observa la propuesta de gabinete a instalar en el PD (Fuente: CLYFSA, diseño de medición de puestos de distribución, gentileza), este debe cumplir con las norma IEC-61439 [8].

Para el caso de los transformadores monofásicos se dimensiona la corriente nominal y en base a ello se elabora el listado de materiales.

$$I = \frac{S}{V} \tag{6.4}$$

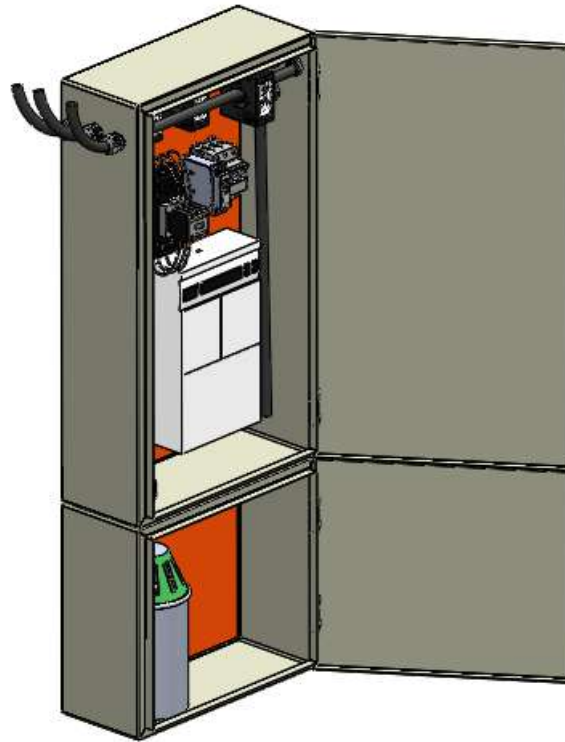


Figura 6.4: Tablero de medición de puesto de distribución

Entonces:

$$I = (25KVA)/(0,23KV) = 108A \quad (6.5)$$

Por lo tanto el TC adoptado será de 150/5 A. El diseño constructivo del gabinete de medición es similar al de la figura 6.4, pero el listado de materiales difiere y se presenta en la tabla 6.5.

Para el montaje del mismo se propone la configuración mostrada en el modelo de la figura 6.5 el cual indica cómo debe ir montado el gabinete de medición, la misma debe ir a una altura prudencial para evitar el hurto u accidente de tipo eléctrico (Fuente: CLYFSA, diseño de medición de puestos de distribución, gentileza). El esquema de conexión del TC se muestra en la figura A.20.

El grado de éxito de esta propuesta es mayor a la del apartado 6.4.1 debido a que se tendrá medición mes a mes de una zona de influencia menor, también se podrá realizar balance

DESCRIPCIÓN	CANTIDAD
Conector a compresión AL 25/70 mm ²	3
Medidor digital monofásico	1
Caja metálica 65x40x15cm. Para intemperie	1
Mano Francesa BT	2
Cable unifilar tipo comando y control 10x2,5	1,3
Cable NYY 70 mm ²	9
TC 150/5	1
PG de 21	6
Varilla roscable 1/2 pulgada	0,5
Varilla roscable 3/8	0,1
Tuerca 3/8 x 3	4
Bulón Inox de 3/8X1"	2
Conector a compresión 2/8	4
Conector dentado 10-95	4
Conector dentado doble 25- 95/ 25-95	3
Arandela lisa redonda 3/8 Inoxidable	2
Terminal pre-aislado pino tipo Horquilla 2,5mm ²	9
Abrazadera plástica tipo cintillo de 150x3,6 MM	30
Abrazadera plástica tipo cintillo de 30 cms	10
Arandela lisa 1/2 pulgada	4
Tuerca 1/2	4
Perno con tuerca 3/16x3/4	7
Arandela lisa redonda 3/8 galvanizado	4
Terminal pre-aislado pino plano de 4-6mm	3
Arandela lisa redonda de 3/16"	7
Bornera tipo DIN de 10mm ²	1

Tabla 6.5: Listado de materiales para esquema de medición de puesto de distribución de 25 KVA

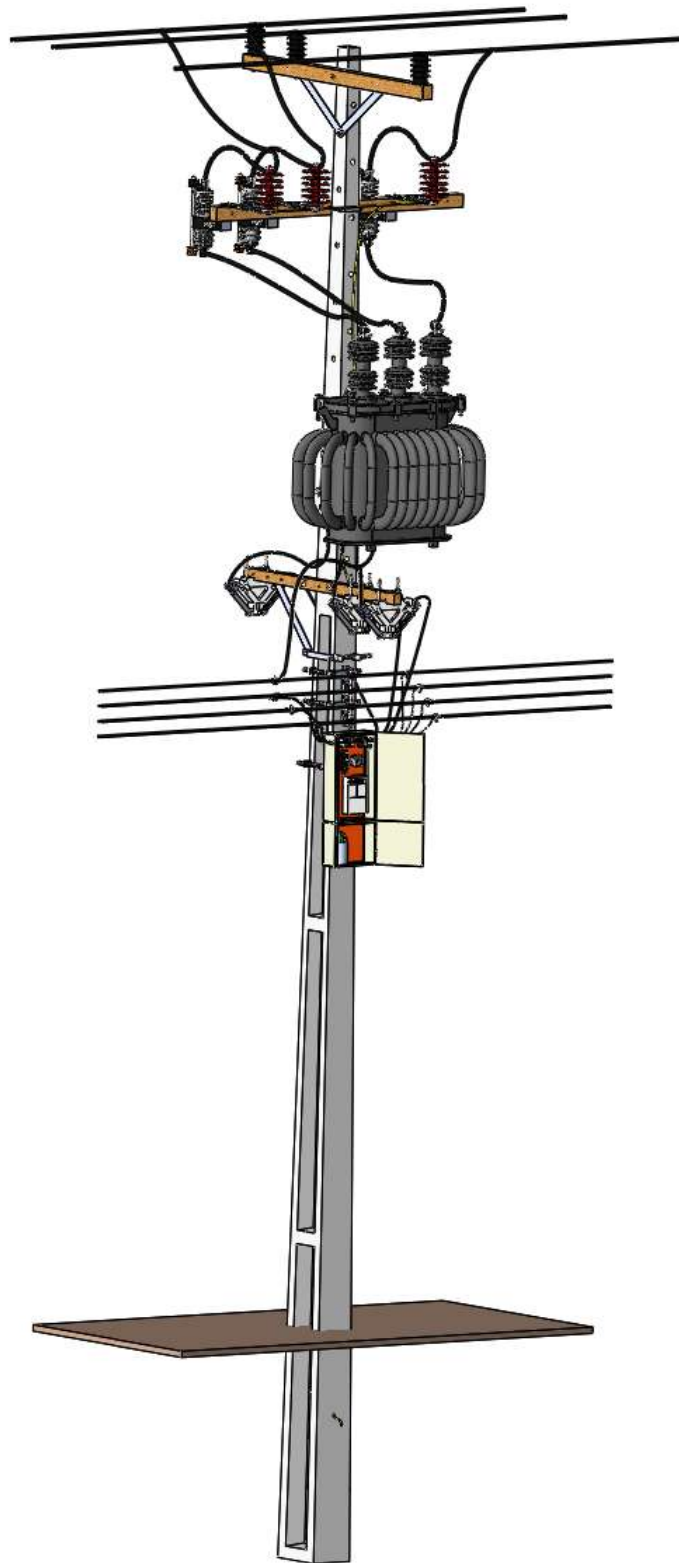


Figura 6.5: Diseño de montaje de gabinete de medición de puestos de distribución públicos.

energético por zona de PD, obteniendo resultados de las pérdidas en menor tiempo, obtener la zona a atacar para reducir las pérdidas, en especial las no técnicas, con valores más elevados. Este método también contribuye a reducir la huella de carbono siendo más amigable con el medio ambiente y dando mayor eficiencia a la operativa de intervención de conexiones clandestinas. La probabilidad de éxito esperada de este método es de 44 %.

La desventaja que presenta este método es la necesidad de modificar, coincidir o reestructurar los itinerarios de lectores de tal forma a que coincida los ciclos de facturación del usuario, PD y cabecera del alimentador, esto para afinar los cálculos de pérdidas no técnicas.

6.4.3. Medición on-line de puestos de distribución

Consiste básicamente en las mismas implementaciones que en el apartado 6.4.2, la diferencia en este método consiste en que la medición es realizada por medidores inteligentes, el cual transfiere los datos de medición a un servidor que a su vez facilita los datos en una interfaz para su análisis. También requiere conectividad por parte de los medidores para poder realizar la transferencia de datos, hoy día empresas como ANDE y CLYFSA están optando por la conectividad por modulo GSM debido a que la tasa de transferencia de datos requerida no es elevada y resulta económico comparada con otros medios como por ejemplo la fibra óptica, en la conectividad GSM se inserta un chip de una compañía telefónica con una tarifa fija para la transferencia de datos.

Como ya se mencionó, los materiales son similares al apartado 6.4.2, exceptuando el medidor, el cual debe ser un medidor inteligente, aparte de ello se debe adquirir la licencia de un software de gestión de los datos para que sea amigable, fácil de usar y disponer de los datos. Este método ya esta siendo aplicado en la medición de cabecera de los alimentadores en particular aplicado al COV1 el cual es el alimentador en estudio.

En la figura 6.6, se muestra el esquema de cómo se transfieren los datos (Fuente: ISKRA, fabricante de medidores inteligentes).

Para las simulaciones en el tiempo puede ser utilizado otros softwares de simulación, como

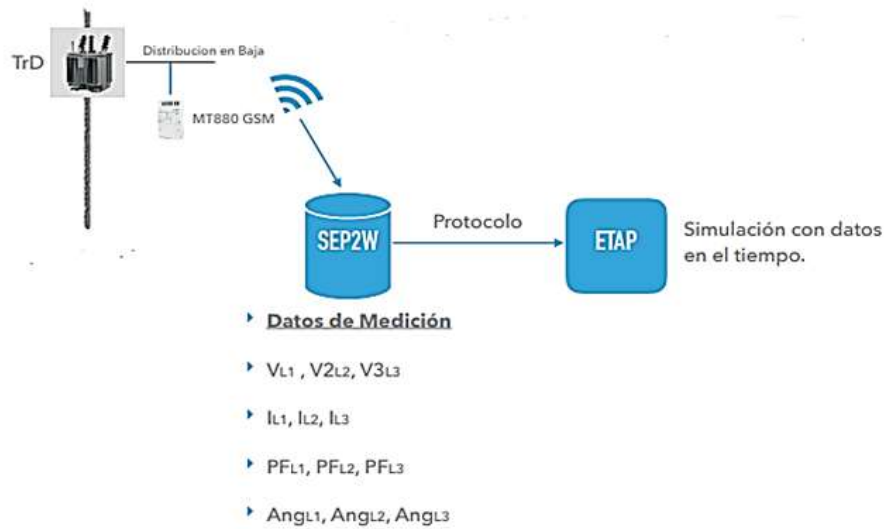


Figura 6.6: Sistema de transferencia de datos

el DiGSILENT, CYMDIST, ETAP, entre otros, en los cuales se puede realizar el modelado de red y simular el comportamiento de los mismos.

La probabilidad de éxito es de un 60 a 65 % ya que, al tener medición en tiempo real de los puestos de distribución, poder contrastar con la medición de cabecera, y tener el balance energético por zona de PD, permite tomar acciones de manera rápida y eficiente, para tomar un plan de acción respecto a las pérdidas. No es necesario introducir el medidor al itinerario de un lector ya que los datos pueden ser adquiridos de manera remota.

Parte IV

Capítulo 7

Resultados y Discusión

El análisis de datos comenzó con la limpieza y filtrado de los mismos, se reemplazaron datos faltantes por media móvil y se eliminaron los registros con mucha falta de lecturas. Esto conllevó a la reducción de la base de datos quedando en 38,2% del original.

Se utilizaron herramientas para el procesamiento de datos históricos en busca de clientes sospechosos en selectos estratos de consumo, se procesaron 2462 registros, de los cuales 329 fueron considerados como sospechosos posterior a correr el algoritmo elaborado de aprendizaje supervisado.

De estos usuarios sospechosos se verificó en el sistema de gestión de la ANDE, se encontraron 28 usuarios con registro de intervención por fraude, representado el 8,5% de los usuarios considerados sospechosos.

El estrato de consumo con mayor cantidad de fraudes es el 142 (residencial) con potencia contratada de 7040 kW (8 casos de fraude), seguido del estrato 410 (comercial) con potencia contratada de 13200 kW con 7 usuarios fraudulentos, esto según registros de la empresa. sin embargo importante aumentar los estudios sobre estos segmentos de clientes.

Se consultó al sistema de gestión sobre las intervenciones por fraude realizados en Coronel Oviedo, estos representan apenas el 1,8% de todos los usuarios de la ciudad. Esto fue una limitante para la aplicación de otros métodos de aprendizaje de maquina.

Se estima una reducción de 75 % la utilización de recursos y gastos para trabajo de campo en intervenciones por fraude, gracias a la reducción del universo de clientes sospechosos. Esta alternativa de análisis de datos se vuelve viable si la recaudación aumenta como mínimo en 0,04 %.

En cuanto a la alternativa de monitoreo de transformadores y balance energético para control de pérdidas no técnicas se plantearon 3 (tres) opciones principales.

Se utilizó software de simulación para estimar las pérdidas técnicas del alimentador seleccionado para estudio, de lo cual se obtuvo como pérdida técnica un 2,1 % de la total suministrada en uno de los meses evaluados y 33 % de pérdidas no técnicas, por lo que se vuelve preponderante para mejorar la situación comercial en cuanto a la venta de energía se refiere.

En cuanto al estudio de probabilidad de éxito de estas alternativas de control, se tiene muy bajo porcentaje para el método actualmente utilizado por ANDE, 44 % para la medición off-line de transformadores y 60 % para la medición on-line de zonas de PD.

La medición off-line de zonas de PD, dentro de un análisis financiero, se considera como la mejor opción teniendo un criterio de beneficio-costos, el cual llega a 1,62 de índice. Sin embargo la medición on-line solo difiere con el anterior en 0,33 puntos.

Capítulo 8

Conclusiones

Se identificaron los usuarios con registros de fraudes en el sistema de gestión OPEN, esto fue utilizado para realizar un algoritmo de minería de datos encontrándose sospechosos de fraudes.

En una segunda alternativa se elaboraron las principales directrices para aplicar metodologías y herramientas de control de suministro de energía conllevando a la detección temprana de pérdidas no técnicas.

Se proyectaron las configuraciones necesarias, materiales a utilizar y dimensionamiento de equipos para los sistemas de control de suministro de energía.

Finalmente se evaluaron las alternativas en un marco financiero.

Capítulo 9

Recomendaciones

Realizando todos estos estudios se proponen las siguientes recomendaciones:

- Mejorar el análisis de datos utilizando otras técnicas de Machine Learning, una vez que se tengan más datos de usuarios fraudulentos y mejorar con ello el aprendizaje automático.
- Aplicar el sistema de control con medición de transformadores on-line por su mayor probabilidad de llevar a cabo y su poca diferencia de costo con la medición off-line.
- Desarrollar nuevos itinerarios y rutas de lectura teniendo en cuenta como referencia las zonas de PD.
- Continuar la investigación y enfocarlo en la mitigación de conexiones clandestinas en zonas sociales.

Apéndice A

Ingeniería de diseño

A.1. Análisis de datos

A.1.1. Limpieza de datos

Para la limpieza de datos se utilizó el siguiente código:

```
# %%  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import statistics  
  
# %%  
from google.colab import drive  
drive.mount('/content/drive')  
  
# %%  
#df = pd.read_excel("drive/MyDrive/Tesis_Perdidas/DATOS/int2/19 al 21 datos.xlsx")  
df= pd.read_excel("19 al 21 datos.xlsx")
```

```
# %%  
df  
  
# %%  
#Seleccionar clientes en cierto tipo de tarifa  
df= df.loc[(df.COD_TAR ==141) | (df.COD_TAR ==142)]  
  
# %%  
len(df)  
  
# %%  
df = df.replace(0, np.nan)  
  
# %%  
quitar = []  
for i in range(len(df)):  
if df.iloc[i].isna().sum()>4:  
quitar.append(False)  
else:  
quitar.append(True)  
  
# %%  
df_limpieza_1 = df[quitar]  
  
# %%  
df_limpieza_1 = df_limpieza_1.reset_index(drop=True)  
  
# %%
```

```
row1, col1 = df_limpieza_1.shape

# %%
df_limpieza_1 = df_limpieza_1.replace(np.nan, 0)

# %%
for x in range(col1):
    for y in range(row1):
        if df_limpieza_1.iloc[y,x]==0:
            med = []
            for z in range(8):
                if x < 30:
                    med.append(df_limpieza_1.iloc[y,x+z])
                else:
                    med.append(df_limpieza_1.iloc[y,x-z])
            med_val = statistics.mean(med)
            df_limpieza_1.iloc[y,x]=med_val

# %%
df_limpieza_1

# %%
dgh = df_limpieza_1["COD_TAR"].value_counts()
print(dgh)
```

En cuanto a la agrupación de clientes por sus contratos, se utilizaron los siguientes códigos:

```
df_410_6600 = df_410[df_410["POT"]==6600]
```

```
df_410_19800 = df_410[df_410["POT"]==19800]
df_410_13200 = df_410[df_410["POT"]==13200]
df_142_7040 = df_142[df_142["POT"]==7040]
df_142_19800 = df_142[df_142["POT"] == 19800]
df_142_13200 = df_142[df_142["POT"] == 13200]
df_142_21120 = df_142[df_142["POT"] == 21120]
df_142_16500 = df_142[df_142["POT"] == 16500]
df_142_26400 = df_142[df_142["POT"] == 26400]

# %%
mean_410_19800 = df_410_19800.iloc[:,3:].mean().mean()
aux = []
for x in range(len(df_410_19800)):
    aux_in = []
    for y in range(col1-3):
        if df_410_19800.iloc[x,y+3] < mean_410_19800:
            aux_in.append(1)
        else:
            aux_in.append(0)
    if sum(aux_in)>12:
        aux.append(True)
    else:
        aux.append(False)
```

A.1.2. Exploración de datos y búsqueda de sospechosos

A continuación se muestra la secuencia de código utilizado para el filtrado y búsqueda de clientes sospechosos por estrato seleccionado:

```
plt.subplots(figsize=(40,20))
```

```
for x in range(len(df_410_19800[aux])):
    med = []
    for y in range(col1-3):
        med.append(df_410_19800[aux].iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
        plt.plot(med, color=clor, linewidth=1)
        plt.yticks(fontsize=25)
        plt.xticks(fontsize=25)
        plt.xlabel("Meses", fontsize=35)
        plt.ylabel("Consumo kWh", fontsize=35)
        plt.show()

# %%
x_1 = np.full_like(df_410_19800.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_410_19800.iloc[:,44].values)
plt.yticks(fontsize=25)
#plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
#plt.grid()
plt.show()

# %%
df_410_19800_f1 = df_410_19800[df_410_19800["mean_sc"]<0.08]

# %%
df_410_19800_f1= df_410_19800_f1.reset_index(drop=True)
```

```
df_410_19800_f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_410_19800_f1)):
    med = []
    for y in range(41-3):
        med.append(df_410_19800_f1.iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
        plt.plot(med, color=clor, linewidth=1)
        plt.yticks(fontsize=25)
        plt.xticks(fontsize=25)
        plt.xlabel("Meses", fontsize=35)
        plt.ylabel("Consumo kWh", fontsize=35)
    plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_410_19800.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_410_19800_f1.iloc[:,40].values)
plt.yticks(fontsize=25)
```

```
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
#plt.grid()
plt.show()

# %%
df_410_19800_f1

# %%
x_mean = df_410_19800_f1.iloc[:,43].values
y_mean = df_410_19800_f1.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
#plt.xticks(fontsize=25)
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_))
```

```
# %%
def f(x_):
    y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
    return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title("Regresión lineal std vs mean", fontsize=45)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_410_19800_f1[df_410_19800_f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
    else:
        med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
```

```
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title("Regresión lineal std vs mean", fontsize=45)
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
    med = []
    for y in range(40-3):
        med.append(sos_aux.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=35)
    plt.ylabel("Consumo kWh", fontsize=35)
    plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_410_13200)):
    med = []
```

```
for y in range(col1-3):
med.append(df_410_13200.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=25)
plt.ylabel("Consumo kWh", fontsize=25)
plt.title(f"Consumo del segmento {df_410_13200.iloc[1,1]}, potencia de contrato: {roun
plt.show()

# %%
x_1 = np.full_like(df_410_13200.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_410_13200.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_410_13200.iloc[1,1]}, potencia de contra
#plt.grid()
plt.show()

# %%
df_410_13200_f1 = df_410_13200[df_410_13200["mean_sc"]<0.15]

# %%
df_410_13200_f1= df_410_13200_f1.reset_index(drop=True)
df_410_13200_f1
```

```
# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_410_13200_f1)):
    med = []
    for y in range(41-3):
        med.append(df_410_13200_f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_410_13200.iloc[1,1]}, potencia de contrato: {roun
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_410_13200.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_410_13200.iloc[1,1]}, potencia de contra
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_410_13200_f1.iloc[:,40].values)
```

```
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_410_13200.iloc[1,1]}, potencia de contrato: {roun
#plt.grid()
plt.show()

# %%
x_mean = df_410_13200_f1.iloc[:,43].values
y_mean = df_410_13200_f1.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_410_13200.iloc[1,1]}, potencia de contrato: {roun
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_))
```

```
# %%
def f(x_):
    y_ = regresion_lineal.coef*x_ + regresion_lineal.intercept_
    return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_410_13200.iloc[1,1]}, potencia de contrata")
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_410_13200_f1[df_410_13200_f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
    else:
        med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux
```

```
# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sospechosos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_410_13200.iloc[1,1]}, potencia de contrato")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
    med = []
    for y in range(40-3):
        med.append(sos_aux.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=35)
    plt.ylabel("Consumo kWh", fontsize=35)
    plt.title(f"Consumo de sospechosos {df_410_13200.iloc[1,1]}, potencia de contrato: {df_410_13200.iloc[1,2]}")
    plt.show()
    sos_aux.to_excel(f"{df_410_13200.iloc[1,1]}_contrato_{round(df_410_13200.iloc[1,2])}.xlsx")

# %%
plt.subplots(figsize=(40,20))
```

```
for x in range(len(df_410_6600)):
    med = []
    for y in range(col1-3):
        med.append(df_410_6600.iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
        plt.plot(med, color=clor, linewidth=1)
        plt.yticks(fontsize=25)
        plt.xticks(fontsize=25)
        plt.xlabel("Meses", fontsize=25)
        plt.ylabel("Consumo kWh", fontsize=25)
        plt.title(f"Consumo del segmento {df_410_6600.iloc[1,1]}, potencia de contrato: {round}
        plt.show()

# %%
x_1 = np.full_like(df_410_6600.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_410_6600.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_410_6600.iloc[1,1]}, potencia de contrato
#plt.grid()
plt.show()

# %%
df_410_6600 = df_410_6600[df_410_6600["mean_sc"]<0.15]
```

```
# %%
df_410_6600= df_410_6600.reset_index(drop=True)
df_410_6600

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_410_6600)):
    med = []
    for y in range(41-3):
        med.append(df_410_6600.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_410_6600.iloc[1,1]}, potencia de contrato: {round}
    plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_410_6600.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_410_6600.iloc[1,1]}, potencia de contrato
#plt.grid()
plt.show()
```

```
# %%  
plt.subplots(figsize=(40,20))  
plt.hist( df_410_6600.iloc[:,40].values)  
plt.yticks(fontsize=25)  
plt.xticks(fontsize=25)  
plt.ylabel("Cant", fontsize=25)  
plt.xlabel("kWh", fontsize=25)  
plt.title(f"Histograma sin datos atípicos {df_410_6600.iloc[1,1]}, potencia de contrato")  
#plt.grid()  
plt.show()
```

```
# %%  
x_mean = df_410_6600.iloc[:,43].values  
y_mean = df_410_6600.iloc[:,44].values
```

```
# %%  
plt.subplots(figsize=(40,20))  
plt.scatter(x_mean, y_mean)  
plt.yticks(fontsize=25)  
plt.xticks(fontsize=25)  
plt.title(f"Distribución puntual {df_410_6600.iloc[1,1]}, potencia de contrato: {round(x_mean[0],2)}")  
plt.ylabel("mean", fontsize=25)  
plt.xlabel("std", fontsize=25)  
plt.grid()  
plt.show()
```

```
# %%  
from sklearn.linear_model import LinearRegression  
regresion_lineal = LinearRegression()
```

```
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_))

# %%
def f(x_):
    y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
    return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_410_6600.iloc[1,1]}, potencia de contrat")
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_410_6600[df_410_6600["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
    else:
        med.append(False)
sos_aux = sos_aux[med]
```

```
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_410_6600.iloc[1,1]}, potencia de contrato")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35)
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
    med = []
    for y in range(40-3):
        med.append(sos_aux.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=35)
    plt.ylabel("Consumo kWh", fontsize=35)
    plt.title(f"Consumo de sospechosos {df_410_6600.iloc[1,1]}, potencia de contrato: {ro}")
    plt.show()
sos_aux.to_excel(f"{df_410_6600.iloc[1,1]}_contrato_{round(df_410_6600.iloc[1,2])}_so")
```

```
# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_7040)):
    med = []
    for y in range(col1-3):
        med.append(df_142_7040.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_7040.iloc[1,1]}, potencia de contrato: {round}
    plt.show()

# %%
x_1 = np.full_like(df_142_7040.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_7040.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_7040.iloc[1,1]}, potencia de contrato
#plt.grid()
plt.show()
```

```
# %%
df_142_7040_f1 = df_142_7040[df_142_7040["mean_sc"]<0.11]

# %%
df_142_7040_f1= df_142_7040_f1.reset_index(drop=True)
df_142_7040_f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_7040_f1)):
    med = []
    for y in range(41-3):
        med.append(df_142_7040_f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_7040.iloc[1,1]}, potencia de contrato: {round}
    plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_7040.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_142_7040.iloc[1,1]}, potencia de contrato
```

```
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_7040_f1.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_142_7040.iloc[1,1]}, potencia de contrato")
#plt.grid()
plt.show()

# %%
x_mean = df_142_7040_f1.iloc[:,43].values
y_mean = df_142_7040_f1.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_142_7040.iloc[1,1]}, potencia de contrato: {round}")
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()
```

```
# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_))

# %%
def f(x_):
    y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
    return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_7040.iloc[1,1]}, potencia de contrat")
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_142_7040_f1[df_142_7040_f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
```

```
else:
med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_7040.iloc[1,1]}, potencia de contrato")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
med = []
for y in range(40-3):
med.append(sos_aux.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=35)
plt.ylabel("Consumo kWh", fontsize=35)
```

```
plt.title(f"Consumo de sospechosos {df_142_7040.iloc[1,1]}, potencia de contrato: {rou
plt.show()
sos_aux.to_excel(f"{df_142_7040.iloc[1,1]}_contrato_{round(df_142_7040.iloc[1,2])}_so

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_19800)):
    med = []
    for y in range(col1-3):
        med.append(df_142_19800.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_19800.iloc[1,1]}, potencia de contrato: {rou
    plt.show()

# %%
x_1 = np.full_like(df_142_19800.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_19800.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_19800.iloc[1,1]}, potencia de contr
```

```
#plt.grid()
plt.show()

# %%
df_142_19800 = df_142_19800[df_142_19800["mean_sc"]<0.15]

# %%
df_142_19800= df_142_19800.reset_index(drop=True)
df_142_19800

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_19800)):
    med = []
    for y in range(41-3):
        med.append(df_142_19800.iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
        plt.plot(med, color=clor, linewidth=1)
        plt.yticks(fontsize=25)
        plt.xticks(fontsize=25)
        plt.xlabel("Meses", fontsize=25)
        plt.ylabel("Consumo kWh", fontsize=25)
        plt.title(f"Consumo del segmento {df_142_19800.iloc[1,1]}, potencia de contrato: {roun
    plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_19800.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
```

```
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_142_19800.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_19800.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_142_19800.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
x_mean = df_142_19800.iloc[:,43].values
y_mean = df_142_19800.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_142_19800.iloc[1,1]}, potencia de contrato: {rou
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
```

```
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_))

# %%
def f(x_):
    y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
    return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_19800.iloc[1,1]}, potencia de contractura")
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_142_19800[df_142_19800["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
```

```
for i in range(len(sos_aux)):
if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
med.append(True)
else:
med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_19800.iloc[1,1]}, potencia de contra")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
med = []
for y in range(40-3):
med.append(sos_aux.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
```

```
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=35)
plt.ylabel("Consumo kWh", fontsize=35)
plt.title(f"Consumo de sospechosos {df_142_19800.iloc[1,1]}, potencia de contrato: {r
plt.show()
sos_aux.to_excel(f"{df_142_19800.iloc[1,1]}_contrato_{round(df_142_19800.iloc[1,2])}_s

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_13200)):
    med = []
    for y in range(col1-3):
        med.append(df_142_13200.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_13200.iloc[1,1]}, potencia de contrato: {rou
    plt.show()

# %%
x_1 = np.full_like(df_142_13200.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_13200.iloc[:,44].values)
plt.yticks(fontsize=25)
```

```
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_13200.iloc[1,1]}, potencia de contrato: {roun
#plt.grid()
plt.show()

# %%
df_142_13200f1 = df_142_13200[df_142_13200["mean_sc"]<0.15]

# %%
df_142_13200f1= df_142_13200f1.reset_index(drop=True)
df_142_13200f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_13200f1)):
    med = []
    for y in range(41-3):
        med.append(df_142_13200f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_13200.iloc[1,1]}, potencia de contrato: {roun
    plt.show()

# %%
plt.subplots(figsize=(40,20))
```

```
plt.hist( df_142_13200.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_142_13200.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_13200f1.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_142_13200.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
x_mean = df_142_13200f1.iloc[:,43].values
y_mean = df_142_13200f1.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
```

```
plt.title(f"Distribución puntual {df_142_13200.iloc[1,1]}, potencia de contrato: {roun
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept

# %%
def f(x_):
y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_13200.iloc[1,1]}, potencia de contrat
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
```

```
sos_aux = df_142_13200f1[df_142_13200f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
med.append(True)
else:
med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_13200.iloc[1,1]}, potencia de contrato")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
med = []
for y in range(40-3):
med.append(sos_aux.iloc[x,y+3])
```

```
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=35)
plt.ylabel("Consumo kWh", fontsize=35)
plt.title(f"Consumo de sospechosos {df_142_13200.iloc[1,1]}, potencia de contrato: {rou
plt.show()
sos_aux.to_excel(f"{df_142_13200.iloc[1,1]}_contrato_{round(df_142_13200.iloc[1,2])}_s

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_21120)):
med = []
for y in range(col1-3):
med.append(df_142_21120.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=25)
plt.ylabel("Consumo kWh", fontsize=25)
plt.title(f"Consumo del segmento {df_142_21120.iloc[1,1]}, potencia de contrato: {rou
plt.show()

# %%
x_1 = np.full_like(df_142_21120.iloc[:,1],1)
x_1.shape

# %%
```

```
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_21120.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_21120.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
df_142_21120f1 = df_142_21120[df_142_21120["mean_sc"]<0.08]

# %%
df_142_21120f1= df_142_21120f1.reset_index(drop=True)
df_142_21120f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_21120f1)):
    med = []
    for y in range(41-3):
        med.append(df_142_21120f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_21120.iloc[1,1]}, potencia de contrato: {roun
plt.show()
```

```
# %%  
plt.subplots(figsize=(40,20))  
plt.hist( df_142_21120.iloc[:,40].values)  
plt.yticks(fontsize=25)  
plt.xticks(fontsize=25)  
plt.ylabel("Cant", fontsize=25)  
plt.xlabel("kWh", fontsize=25)  
plt.title(f"Histograma con datos atípicos {df_142_21120.iloc[1,1]}, potencia de contra  
#plt.grid()  
plt.show()
```

```
# %%  
plt.subplots(figsize=(40,20))  
plt.hist( df_142_21120f1.iloc[:,40].values)  
plt.yticks(fontsize=25)  
plt.xticks(fontsize=25)  
plt.ylabel("Cant", fontsize=25)  
plt.xlabel("kWh", fontsize=25)  
plt.title(f"Histograma sin datos atípicos {df_142_21120.iloc[1,1]}, potencia de contra  
#plt.grid()  
plt.show()
```

```
# %%  
x_mean = df_142_21120f1.iloc[:,43].values  
y_mean = df_142_21120f1.iloc[:,44].values
```

```
# %%  
plt.subplots(figsize=(40,20))
```

```
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_142_21120.iloc[1,1]}, potencia de contrato: {roun
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept

# %%
def f(x_):
y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
return y_
x_ = np.arange((1800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_21120.iloc[1,1]}, potencia de contrat
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
```

```
plt.show()

# %%
sos_aux = df_142_21120f1[df_142_21120f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
    else:
        med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_21120.iloc[1,1]}, potencia de contrato")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
```

```
med = []
for y in range(40-3):
med.append(sos_aux.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=35)
plt.ylabel("Consumo kWh", fontsize=35)
plt.title(f"Consumo de sospechosos {df_142_21120.iloc[1,1]}, potencia de contrato: {r
plt.show()
sos_aux.to_excel(f"{df_142_21120.iloc[1,1]}_contrato_{round(df_142_21120.iloc[1,2])}_s

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_16500)):
med = []
for y in range(col1-3):
med.append(df_142_16500.iloc[x,y+3])
clor = (np.random.random(), np.random.random(), np.random.random())
plt.plot(med, color=clor, linewidth=1)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=25)
plt.ylabel("Consumo kWh", fontsize=25)
plt.title(f"Consumo del segmento {df_142_16500.iloc[1,1]}, potencia de contrato: {roun
plt.show()

# %%
x_1 = np.full_like(df_142_16500.iloc[:,1],1)
```

```
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_16500.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_16500.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
df_142_16500f1 = df_142_16500[df_142_16500["mean_sc"]<0.08]

# %%
df_142_16500f1= df_142_16500f1.reset_index(drop=True)
df_142_16500f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_16500f1)):
    med = []
    for y in range(41-3):
        med.append(df_142_16500f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
```

```
plt.ylabel("Consumo kWh", fontsize=25)
plt.title(f"Consumo del segmento {df_142_16500.iloc[1,1]}, potencia de contrato: {roun
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_16500.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_142_16500.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_16500f1.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_142_16500.iloc[1,1]}, potencia de contr
#plt.grid()
plt.show()

# %%
x_mean = df_142_16500f1.iloc[:,43].values
y_mean = df_142_16500f1.iloc[:,44].values
```

```
# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_142_16500.iloc[1,1]}, potencia de contrato: {roun
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept

# %%
def f(x_):
y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
return y_
x_ = np.arange((2800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_16500.iloc[1,1]}, potencia de contrato
```

```
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_142_16500f1[df_142_16500f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
    if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
        med.append(True)
    else:
        med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos')
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_16500.iloc[1,1]}, potencia de contra")
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
```

```
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
    med = []
    for y in range(40-3):
        med.append(sos_aux.iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=35)
    plt.ylabel("Consumo kWh", fontsize=35)
    plt.title(f"Consumo de sospechosos {df_142_16500.iloc[1,1]}, potencia de contrato: {r}
    plt.show()
    sos_aux.to_excel(f"{df_142_16500.iloc[1,1]}_contrato_{round(df_142_16500.iloc[1,2])}.s

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_26400)):
    med = []
    for y in range(col1-3):
        med.append(df_142_26400.iloc[x,y+3])
        clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=25)
    plt.ylabel("Consumo kWh", fontsize=25)
    plt.title(f"Consumo del segmento {df_142_26400.iloc[1,1]}, potencia de contrato: {roun
    plt.show()
```

```
# %%
x_1 = np.full_like(df_142_26400.iloc[:,1],1)
x_1.shape

# %%
plt.subplots(figsize=(20,20))
plt.boxplot( df_142_26400.iloc[:,44].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("kWh", fontsize=25)
plt.title(f"Gráfico de cajas del segmento {df_142_26400.iloc[1,1]}, potencia de contra")
#plt.grid()
plt.show()

# %%
df_142_26400f1 = df_142_26400[df_142_26400["mean_sc"]<0.08]

# %%
df_142_26400f1= df_142_26400f1.reset_index(drop=True)
df_142_26400f1

# %%
plt.subplots(figsize=(40,20))

for x in range(len(df_142_26400f1)):
    med = []
    for y in range(41-3):
        med.append(df_142_26400f1.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
```

```
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.xlabel("Meses", fontsize=25)
plt.ylabel("Consumo kWh", fontsize=25)
plt.title(f"Consumo del segmento {df_142_26400.iloc[1,1]}, potencia de contrato: {roun
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_26400.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma con datos atípicos {df_142_26400.iloc[1,1]}, potencia de contra
#plt.grid()
plt.show()

# %%
plt.subplots(figsize=(40,20))
plt.hist( df_142_26400f1.iloc[:,40].values)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.ylabel("Cant", fontsize=25)
plt.xlabel("kWh", fontsize=25)
plt.title(f"Histograma sin datos atípicos {df_142_26400.iloc[1,1]}, potencia de contra
#plt.grid()
plt.show()
```

```
# %%
x_mean = df_142_26400f1.iloc[:,43].values
y_mean = df_142_26400f1.iloc[:,44].values

# %%
plt.subplots(figsize=(40,20))
plt.scatter(x_mean, y_mean)
plt.yticks(fontsize=25)
plt.xticks(fontsize=25)
plt.title(f"Distribución puntual {df_142_26400.iloc[1,1]}, potencia de contrato: {roun
plt.ylabel("mean", fontsize=25)
plt.xlabel("std", fontsize=25)
plt.grid()
plt.show()

# %%
from sklearn.linear_model import LinearRegression
regresion_lineal = LinearRegression()
regresion_lineal.fit(x_mean.reshape(-1,1), y_mean)

print('w = ' + str(regresion_lineal.coef_) + ', b = ' + str(regresion_lineal.intercept_

# %%
def f(x_):
y_ = regresion_lineal.coef_*x_ + regresion_lineal.intercept_
return y_
x_ = np.arange((2800))
y = f(x_)

# %%
plt.subplots(figsize=(50,20))
```

```
plt.scatter(x_mean, y_mean)
plt.plot(x_, y, color='blue', linewidth=3)
plt.title(f"Regresión lineal std vs mean {df_142_26400.iloc[1,1]}, potencia de contra
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
plt.show()

# %%
sos_aux = df_142_26400f1[df_142_26400f1["std"]>250]
sos_aux = sos_aux.reset_index(drop=True)
med= []
for i in range(len(sos_aux)):
if y[round(sos_aux.iloc[i,43])] > sos_aux.iloc[i,44] :
med.append(True)
else:
med.append(False)
sos_aux = sos_aux[med]
sos_aux = sos_aux.reset_index(drop=True)
sos_aux

# %%
plt.subplots(figsize=(30,30))
plt.scatter(x_mean, y_mean)
plt.scatter(sos_aux["std"].values, sos_aux["mean_sc"].values, color='red', label='Sos
plt.plot(x_, y, color='blue', linewidth=3, label='Regresión lineal')
plt.title(f"Regresión lineal std vs mean {df_142_26400.iloc[1,1]}, potencia de contra
plt.legend(loc="upper left", fontsize=25)
plt.grid()
plt.yticks(fontsize=35)
plt.xticks(fontsize=35 )
```

```
plt.show()

# %%
plt.subplots(figsize=(40,20))

for x in range(len(sos_aux)):
    med = []
    for y in range(40-3):
        med.append(sos_aux.iloc[x,y+3])
    clor = (np.random.random(), np.random.random(), np.random.random())
    plt.plot(med, color=clor, linewidth=1)
    plt.yticks(fontsize=25)
    plt.xticks(fontsize=25)
    plt.xlabel("Meses", fontsize=35)
    plt.ylabel("Consumo kWh", fontsize=35)
    plt.title(f"Consumo de sospechosos {df_142_26400.iloc[1,1]}, potencia de contrato: {r}
    plt.show()
    sos_aux.to_excel(f"{df_142_26400.iloc[1,1]}_contrato_{round(df_142_26400.iloc[1,2])}_s
```

A.1.3. Agrupación de usuarios

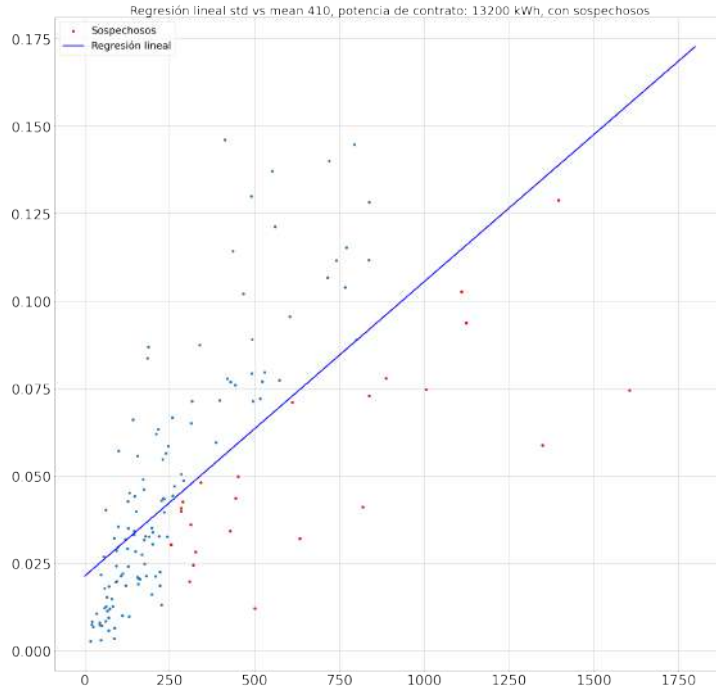


Figura A.1: Regresión y agrupación de clientes con contrato de 13200 kW en el estrato 410.

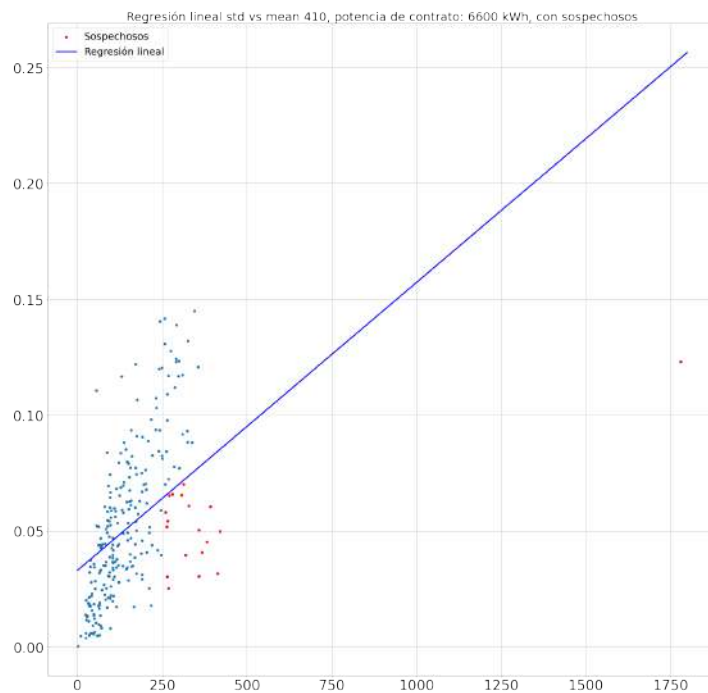


Figura A.2: Regresión y agrupación de clientes con contrato de 6600 kW en el estrato 410.

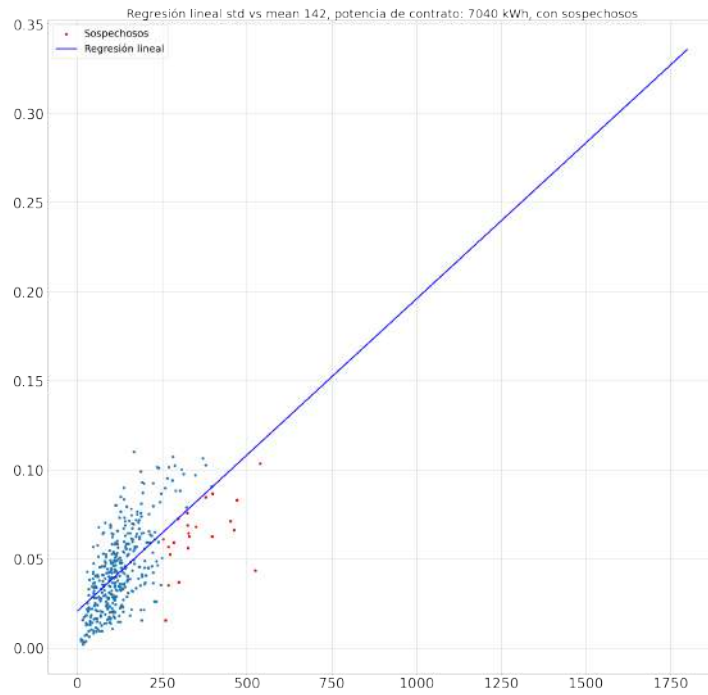


Figura A.3: Regresión y agrupación de clientes con contrato de 7040 kW en el estrato 142.

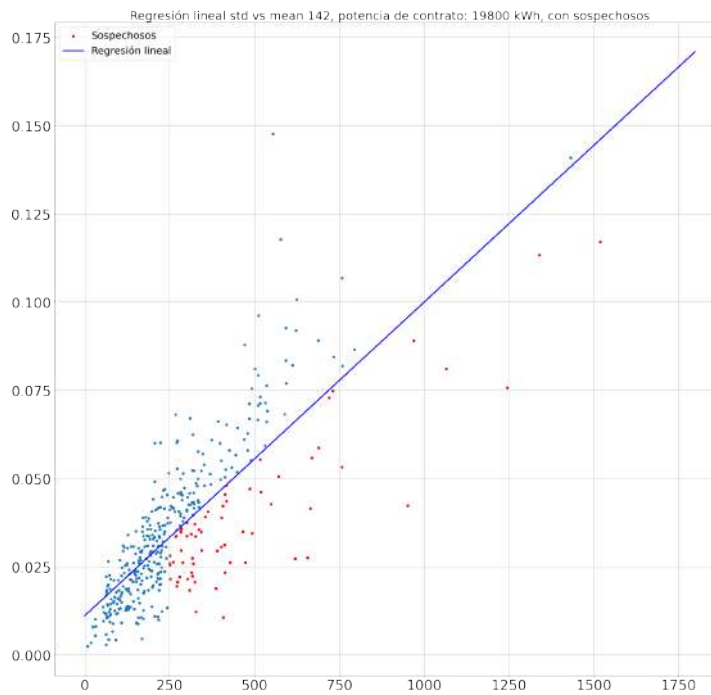


Figura A.4: Regresión y agrupación de clientes con contrato de 19800 kW en el estrato 142.

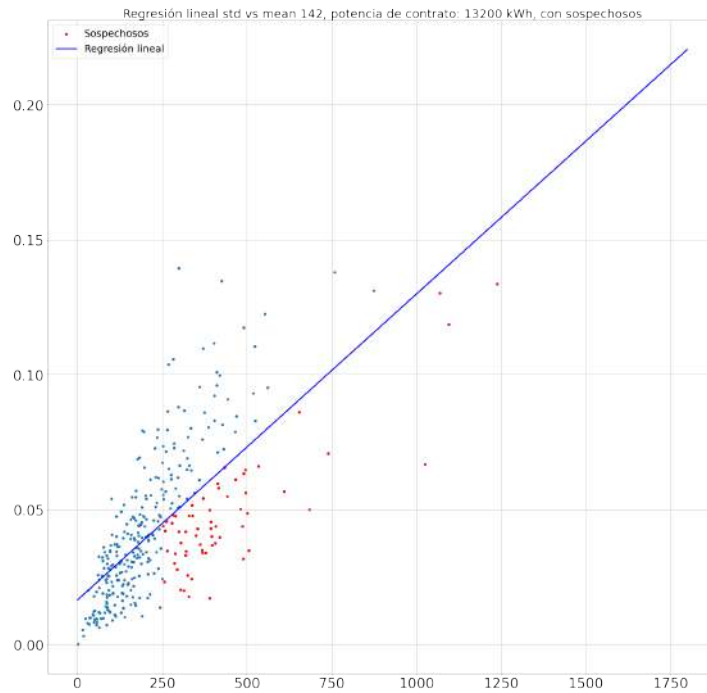


Figura A.5: Regresión y agrupación de clientes con contrato de 13200 kW en el estrato 142.

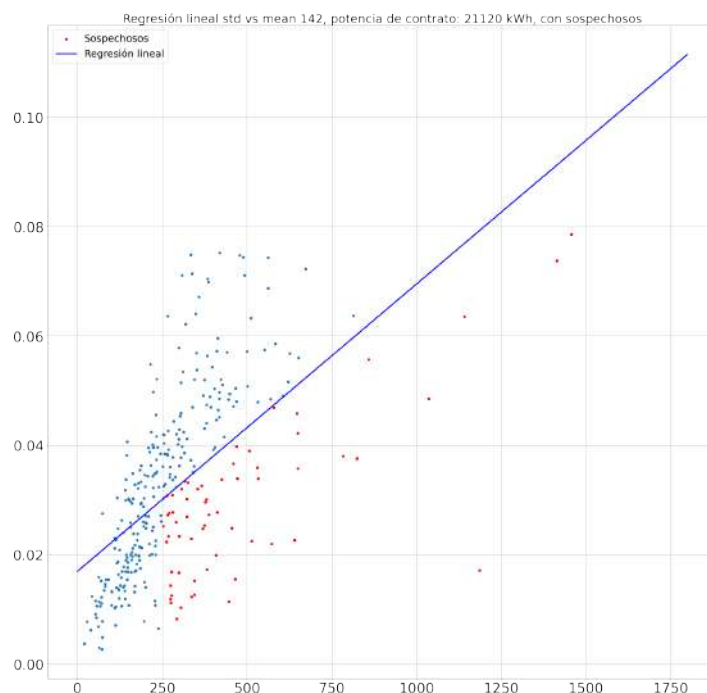


Figura A.6: Regresión y agrupación de clientes con contrato de 21120 kW en el estrato 142.

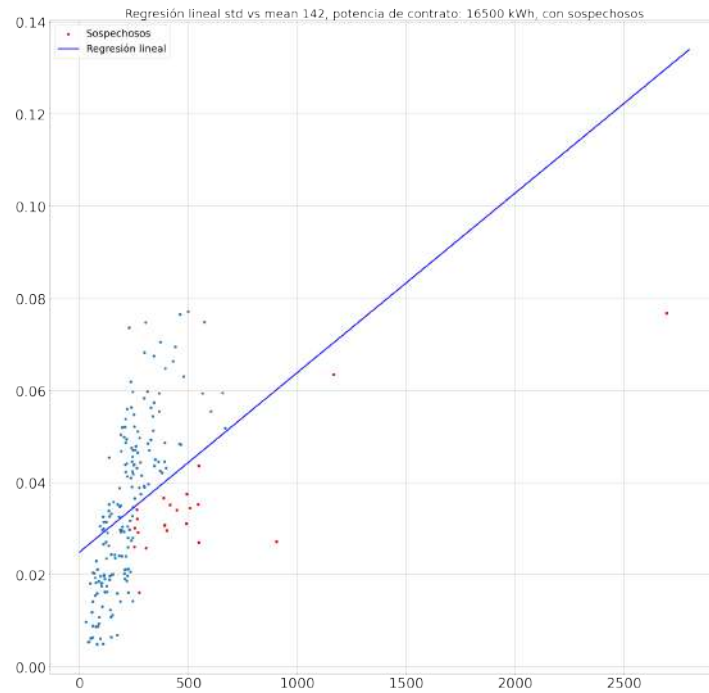


Figura A.7: Regresión y agrupación de clientes con contrato de 16500 kW en el estrato 142.

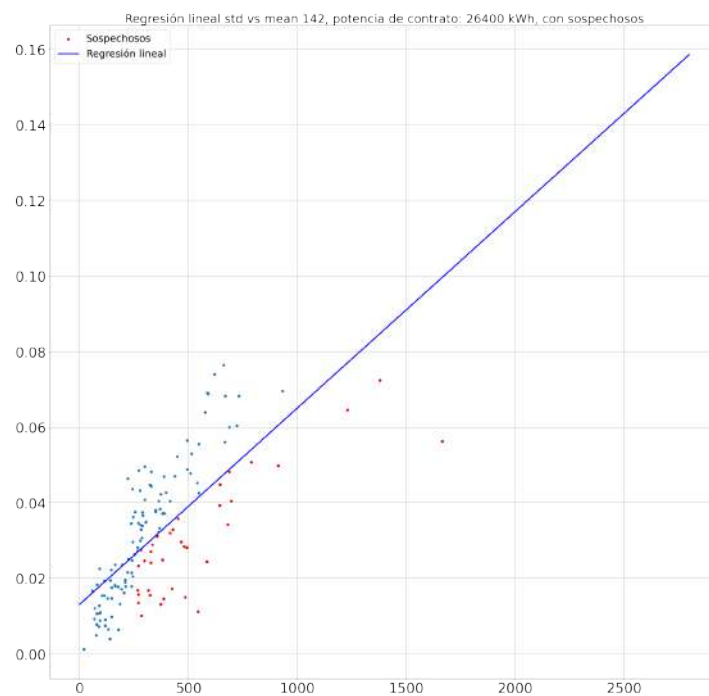


Figura A.8: Regresión y agrupación de clientes con contrato de 26400 kW en el estrato 142.

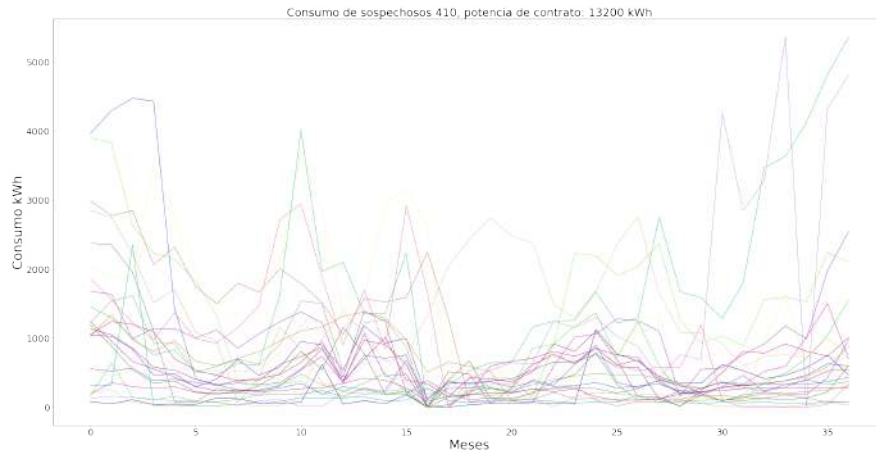


Figura A.9: Consumo histórico de clientes sospechosos con contrato de 13200 kW en el estrato 410.

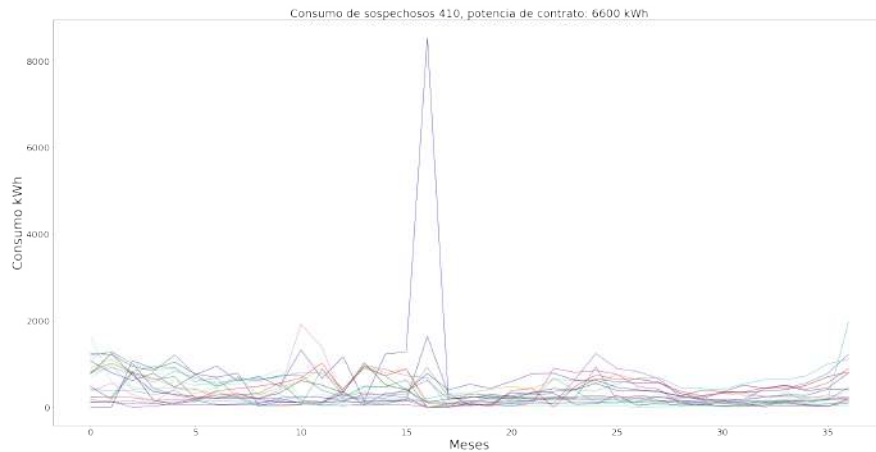


Figura A.10: Consumo histórico de clientes sospechosos con contrato de 6600 kW en el estrato 410.

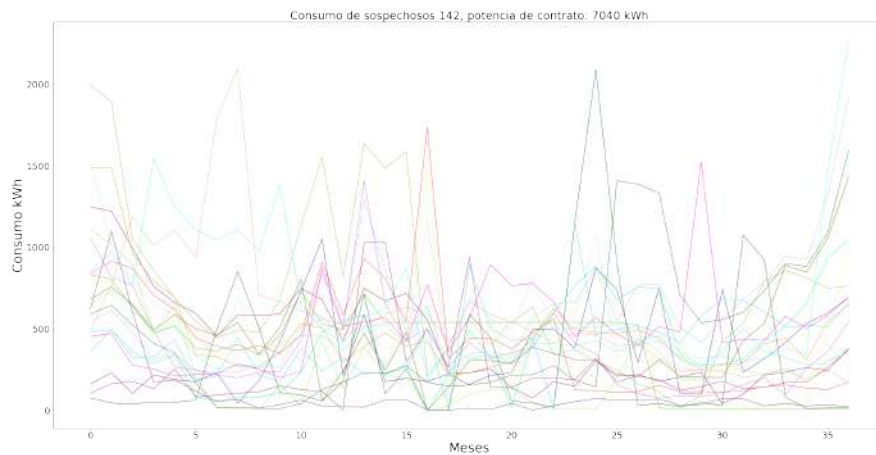


Figura A.11: Consumo histórico de clientes sospechosos con contrato de 7040 kW en el estrato 142.

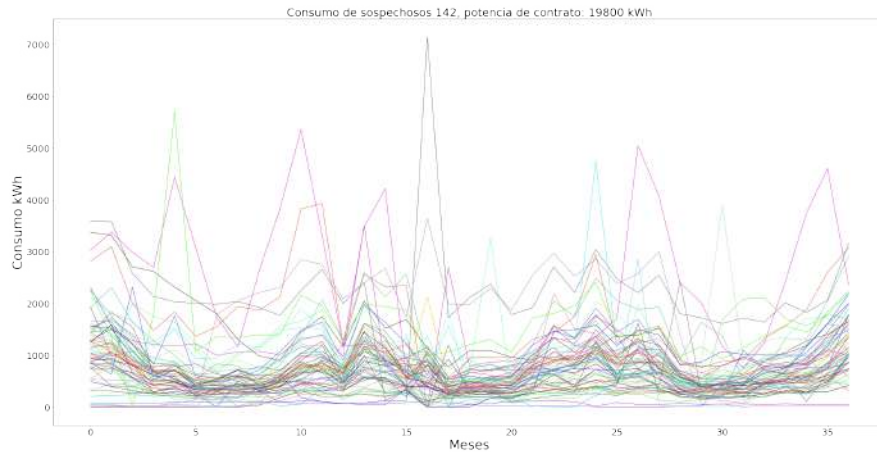


Figura A.12: Consumo histórico de clientes sospechosos con contrato de 19800 kW en el estrato 142.

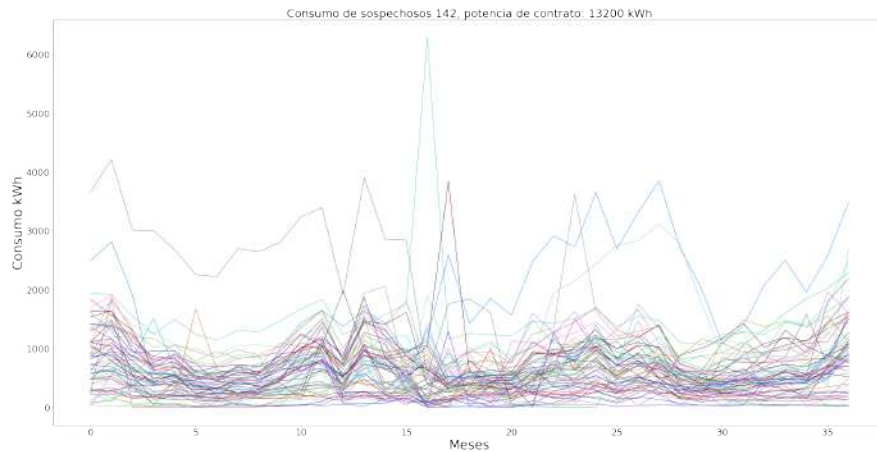


Figura A.13: Consumo histórico de clientes sospechosos con contrato de 13200 kW en el estrato 142.

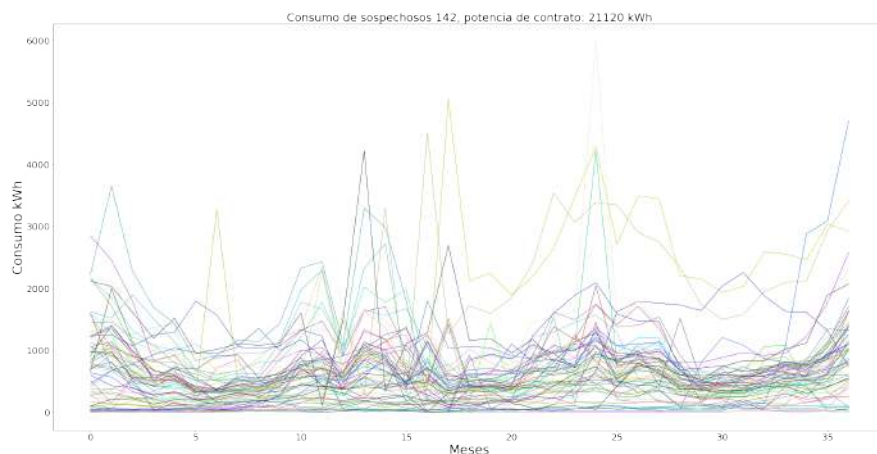


Figura A.14: Consumo histórico de clientes sospechosos con contrato de 21120 kW en el estrato 142.

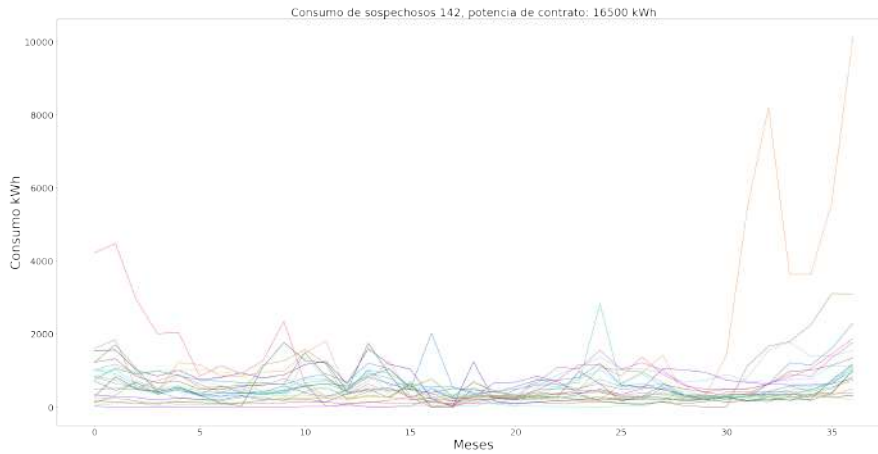


Figura A.15: Consumo histórico de clientes sospechosos con contrato de 16500 kW en el estrato 142.

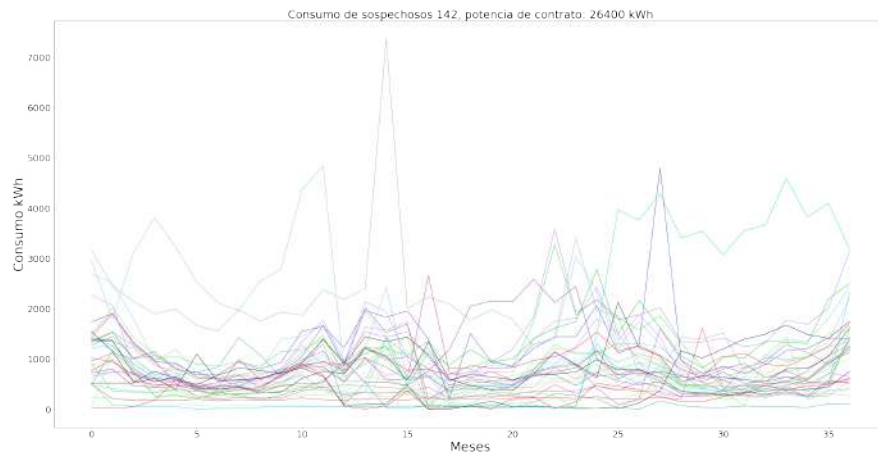


Figura A.16: Consumo histórico de clientes sospechosos con contrato de 26400 kW en el estrato 142.

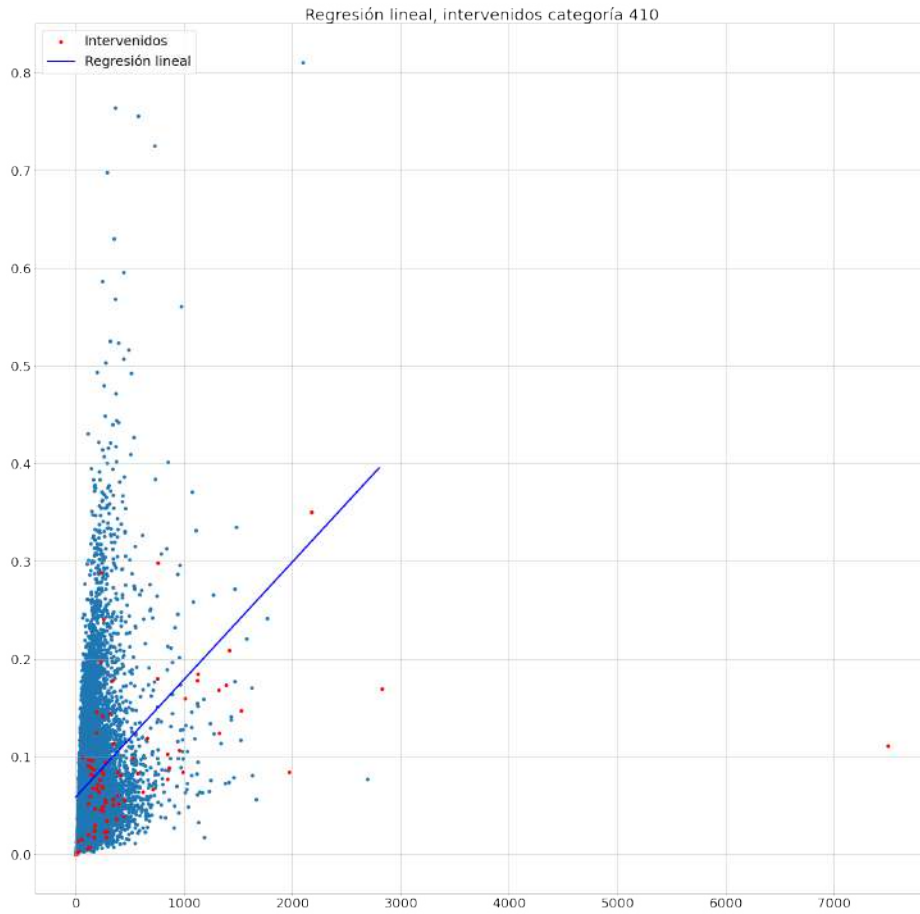


Figura A.17: Std vs mean de clientes con fraudes intervenidos.

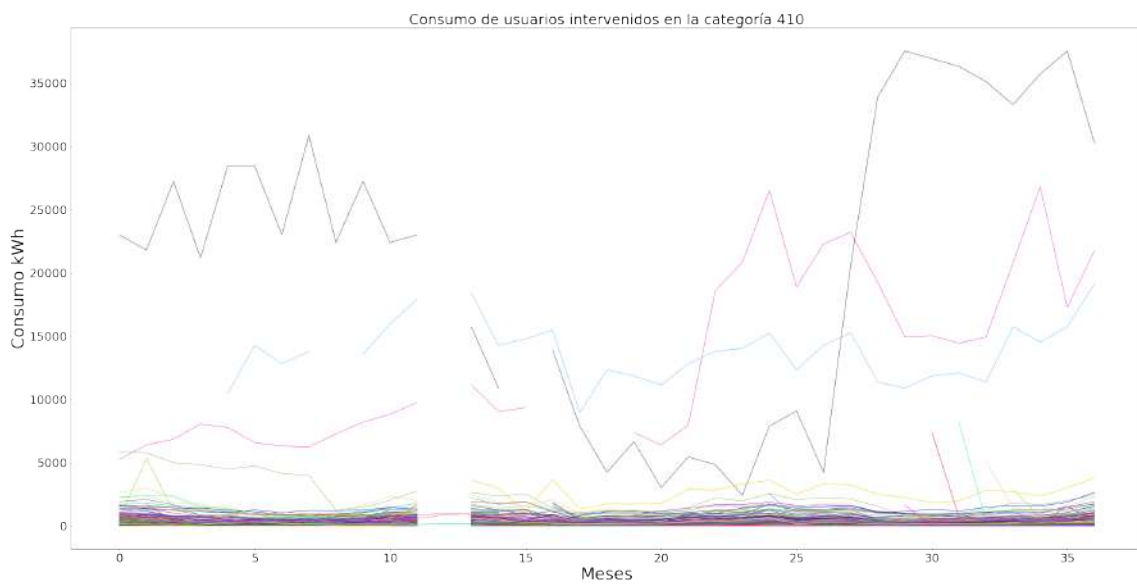


Figura A.18: Consumo histórico de clientes con fraudes intervenidos.

A.2. Monitoreo de transformadores y balance energético

A.2.1. Cálculo de constantes de modelo estadísticos y pérdidas técnicas:

$$C_1 = 12/24 \cdot 0,3 = 0,15$$

$$C_2 = 8/24 \cdot 0,6 = 0,218$$

$$C_3 = 4/24 \cdot 1 = 0,167$$

$$P_{tec} = (0,15 * 12 + 0,218 * 8 + 0,167 * 4)(10 * 30) = 1263,6kWh - mes$$

A.2.2. Modelo DigSilent

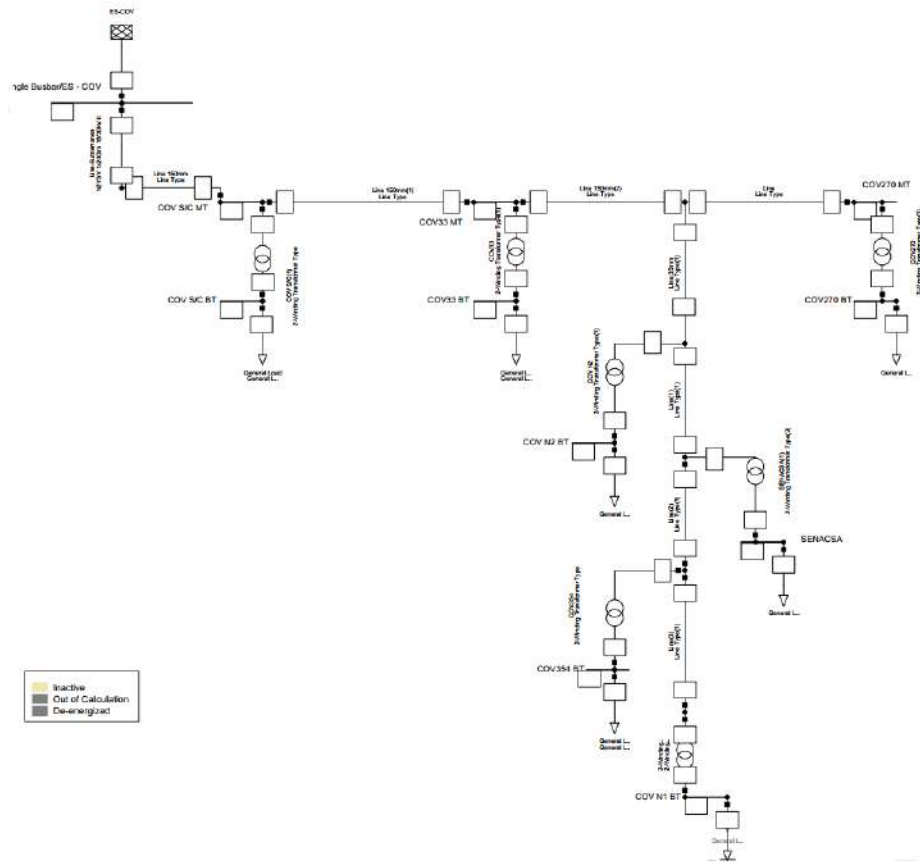


Figura A.19: Esquema modelado en DigSilent

Esquema de Mando

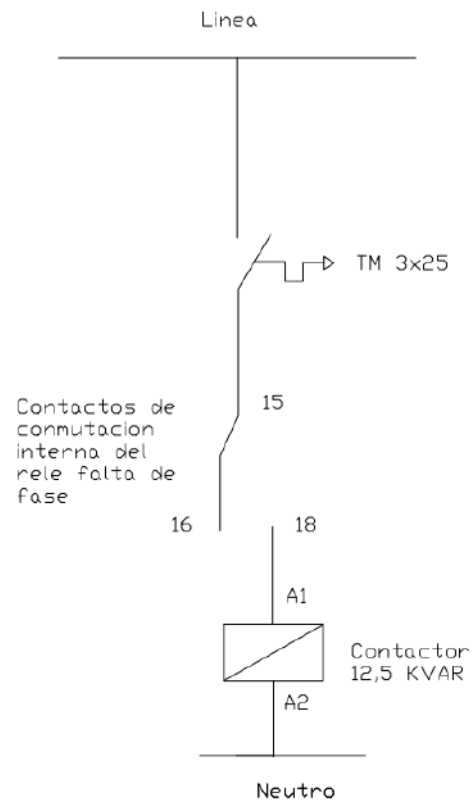


Figura A.20: Esquema de mando para conexión del medidor

A.2.3. Medidores en PD

A.2.4. Cálculo probabilístico

	n	134	p	0.5	R
k	127	1.3125E+11	5.87747E-39	0.007813	6.02691E-30
k	128	7177979809	2.93874E-39	0.015625	3.29597E-31
k	129	333859526	1.46937E-39	0.03125	1.53301E-32
k	130	12840751	7.34684E-40	0.0625	5.89618E-34
k	131	392084	3.67342E-40	0.125	1.80036E-35
k	132	8911	1.83671E-40	0.25	4.09173E-37
k	133	134	9.18355E-41	0.5	6.15298E-39
k	134	1	4.59177E-41	1	4.59177E-41
					0.0000 %

Tabla A.1: Alternativa, verificación aleatoria de usuarios

	n	134	p	0.94	R
k	127	1.3125E+11	0.00038658	2.7994E-09	0.14203989
k	128	7177979809	0.00036338	4.6656E-08	0.12169563
k	129	333859526	0.00034158	7.776E-07	0.08867744
k	130	12840751	0.00032109	0.00001296	0.05343384
k	131	392084	0.00030182	0.000216	0.02556123
k	132	8911	0.00028371	0.0036	0.00910135
k	133	134	0.00026669	0.06	0.00214418
k	134	1	0.00025069	1	0.00025069
					44 %

Tabla A.2: Alternativa, medición off-line

	n	134	p	0.95	R
k	127	1.3125E+11	0.00148217	7.8125E-10	0.15198547
k	128	7177979809	0.00140806	1.5625E-08	0.1579224
k	129	333859526	0.00133766	3.125E-07	0.13955933
k	130	12840751	0.00127078	6.25E-06	0.10198566
k	131	392084	0.00120724	0.000125	0.05916726
k	132	8911	0.00114687	0.0025	0.0255495
k	133	134	0.00108953	0.05	0.00729986
k	134	1	0.00103505	1	0.00103505

64 %

Tabla A.3: Alternativa, medición on-line

Apéndice B

Resumen ejecutivo

B.1. Cálculo de beneficios

Cálculo del promedio de energía no facturada en los meses de estudio:

$$E = \frac{20438,91 + 16123,45 + 24887,82}{3} = 20483,39 \quad kWh - mes$$

Cálculo monetario de la recaudación si se eliminará el 100 % de las pérdidas no-técnicas:

$$ENF_{GS} = 20483,39 \quad kWh - mes \times 405 \frac{gs}{kWh} = 8295774,3 \frac{GS}{mes}$$

B.2. Cálculo de costos

El monto considerado de HHP fue de 300.000 gs, costo superior al de HHP4 de la ANDE e inferior a HHPO.

DESCRIPCIÓN	CANTIDAD	COSTO UNITARIO	COSTO TOTAL
Conector a compresión AL 35/70 mm2	3	8.409	25.227
Medidor digital	1	700	700
Caja metálica 65x40x15cm. Para intemperie	1	186.364	186.364
Mano Francesa BT	2	9.091	18.182
Cable unifilar tipo comando y control 10x2,5	1.3	23.115	30.05
Cable NYY 70 mm2	9	46.409	417.681
TC 200/5	3	132.329	396.987
PG de 21	6	2.727	16.362
Varilla roscable 1/2 pulgada	0.5	8.03	4.015
Varilla roscable 3/8	0.1	4.388	439
Tuerca 3/8 x 3	4	118	472
Bulón Inox de 3/8X1"	2	2.05	4.1
Conector a compresión 2/8	4	1.549	6.196
Conector dentado 10-95	4	4.545	18.18
Conector dentado doble 25- 95/ 25-95	3	15.273	45.819
Arandela lisa redonda 3/8 Inoxidable	2	167	334
Terminal pre-aislado pino tipo Horquilla 2,5mm2	9	273	2.457
Abrazadera plástica tipo cintillo de150x3,6 MM	30	118	3.54
Abrazadera plástica tipo cintillo de 30 cms	10	325	3.25
Arandela lisa 1/2 pulgada	4	364	1.456
Tuerca 1/2	4	222	888
Perno con tuerca 3/16x3/4	7	118	826
Arandela lisa redonda 3/8 galvanizado	4	227	908
Terminal pre-aislado pino plano de 4-6mm	3	273	819
Arandela lisa redonda de 3/16"	7	22	154
Bornera tipo DIN de 10mm2	1	2.224	2.224
TOTAL			1.886.930 Gs

Tabla B.1: Costos de materiales para implementación de medidor totalizador para PD de 100 KVA.

DESCRIPCION DE COSTO DE MANO DE OBRA EN HHP				
HHP		DESCRIPCION	CANTIDAD	Monto
0,1800	1,440	Montaje de PG	8	432.000
0,1800	0,540	Colocación de TC	3	162.000
0,5000	0,500	Conexionado de medidor	1	150.000
0,5000	0,500	Colocación de soporte para caja	1	150.000
19,000	1,900	Montaje de caja totalizador por PD	1	570.000
Totales				1.464.000

Tabla B.2: Costo de mano de obra de montaje de totalizador.

Costo de materiales	1.886.930
Costo de la mano de obra	1.464.000
TOTAL GENERAL GS.	3.350.930

Tabla B.3: Costos totales de implementación de medición totalizadora ara PD de 100 KVA.

DESCRIPCION	CANTIDAD	COSTO UNITARIO	COSTO TOTAL
Conector a compresión AL 25/70 mm ²	3	8.409	25.227
Medidor digital monofásico	1	550	550
Caja metálica 65x40x15cm. Para intemperie	1	186.364	186.364
Mano Francesa BT	2	9.091	18.182
Cable unifilar tipo comando y control 10x2,5	1.3	23.115	30.05
Cable NYY 25 mm ²	9	22.1	198.9
TC 150/5	1	122.86	122.86
PG de 21	6	2.727	16.362
Varilla roscable 1/2 pulgada	0.5	8.03	4.015
Varilla roscable 3/8	0.1	4.388	439
Tuerca 3/8 x 3	4	118	472
Bulón Inox de 3/8X1"	2	2.05	4.1
Conector a compresión 2/8	4	1.549	6.196
Conector dentado 10-95	4	4.545	18.18
Conector dentado doble 25- 95/ 25-95	3	15.273	45.819
Arandela lisa redonda 3/8 Inoxidable	2	167	334
Terminal pre-aislado pino tipo Horquilla 2,5mm ²	9	273	2.457
Abrazadera plástica tipo cintillo de 150x3,6 MM	30	118	3.54
Abrazadera plástica tipo cintillo de 30 cms	10	325	3.25
Arandela lisa 1/2 pulgada	4	364	1.456
Tuerca 1/2	4	222	888
Perno con tuerca 3/16x3/4	7	118	826
Arandela lisa redonda 3/8 galvanizado	4	227	908
Terminal pre-aislado pino plano de 4-6mm	3	273	819
Arandela lisa redonda de 3/16"	7	22	154
Bornera tipo DIN de 10mm ²	1	2.224	2.224
TOTAL			1.407.022 Gs

Tabla B.4: Costos de materiales para implementación de medidor totalizador para PD de 25 KVA

DESCRIPCION DE COSTO DE MANO DE OBRA EN HHP				
HHP	DESCRIPCION	CANTIDAD	Monto	
0,1800	1,440	Montaje de PG	8	432.000
0,1800	0,540	Colocación de TC	3	162.000
0,5000	0,500	Conexión de medidor	1	150.000
0,5000	0,500	Colocación de soporte para caja	1	150.000
19,000	1,900	Montaje de caja totalizador por PD	1	570.000
Totales				1.464.000

Tabla B.5: Costo de mano de obra de montaje de totalizador para PD 25 kVA

Costo de materiales	1.407.022
Costo de la mano de obra	1.464.000
TOTAL GENERAL GS.	2.871.022

Tabla B.6: Costos totales de implementación de medición totalizadora para PD de 25 kVA

B.3. Análisis financiero

En esta sección se realizan los cálculos de rentabilidad de los proyectos de control de pérdidas no-técnicas.

Tasa de utilización 10%

AÑO	INGRESOS	COSTOS	FLUJO DE EFECTIVO	TASA DE TASA $(1+t)^{-n}$	INGRESOS AC-TUALIZADOS	EGRESOS AC-TUALIZADOS
0	-	51.878.200	-51.878.200	1,000000000	-	51.878.200
1	99.549.292	52.915.764	46.633.528	0,909090909	90.499.356	48.105.240
2	101.540.277	53.974.079	47.566.198	0,826446281	83.917.585	44.606.677
3	103.571.083	55.053.561	48.517.522	0,751314801	77.814.488	41.362.555
4	105.642.505	56.154.632	49.487.873	0,683013455	72.155.252	38.354.369
5	107.755.355	57.277.725	50.477.630	0,620921323	66.907.597	35.564.961
TOTAL	518.058.511	327.253.961	190.804.550		391.294.278	259.872.002

VAN	131.422.276
TIR	87,62%
B/C	1,506

Tabla B.7: Rentabilidad A1

Tasa de utilización 10%

AÑO	INGRESOS	COSTOS	FLUJO DE EFECTIVO	TASA DE TASA $(1+t)^{-n}$	INGRESOS AC-TUALIZADOS	EGRESOS AC-TUALIZADOS
0	-	62.950.056	-62.950.056	1,000000000	-	62.950.056
1	99.549.292	45.169.884	54.379.408	0,909090909	90.499.356	41.063.531
2	101.540.277	46.073.282	55.466.996	0,826446281	83.917.585	38.077.092
3	103.571.083	46.994.747	56.576.336	0,751314801	77.814.488	35.307.849
4	105.642.505	47.934.642	57.707.862	0,683013455	72.155.252	32.740.006
5	107.755.355	48.893.335	58.862.020	0,620921323	66.907.597	30.358.914
TOTAL	518.058.511	298.015.946	220.042.565		391.294.278	240.497.448

VAN	150.796.829
TIR	83,84%
B/C	1,627

Tabla B.8: Rentabilidad A2

Tasa de utilización 10%

ANO	INGRESOS	COSTOS	FLUJO DE EFECTIVO	TASA $(1+t)^{-n}$	INGRESOS AC-TUALIZADOS	EGRESOS AC-TUALIZADOS
0	-	81.839.106	-81.839.106	1,000000000	-	81.839.106
1	99.549.292	56.185.884	43.363.408	0,909090909	90.499.356	51.078.076
2	101.540.277	57.309.602	44.230.676	0,826446281	83.917.585	47.363.307
3	103.571.083	58.455.794	45.115.289	0,751314801	77.814.488	43.918.703
4	105.642.505	59.624.910	46.017.595	0,683013455	72.155.252	40.724.616
5	107.755.355	60.817.408	46.937.947	0,620921323	66.907.597	37.762.825
TOTAL	518.058.511	374.232.703	143.825.809		391.294.278	302.686.633

VAN	88.607.644
TIR	46,24%
B/C	1,293

Tabla B.9: Rentabilidad A3

Bibliografía

- [1] R. A. O. ACOSTA y L. E. F. LÓPEZ, “Comisión de integración energética regional comité nacional venezolano,” *CNV*, 2015.
- [2] ANDE, “Pliego de tarifas n 21,” 2019.
- [3] E. R. Castillo Abad y A. F. Pérez Ayala, “Diseño de una metodología para la reducción y control de pérdidas de energía utilizando índices sectorizados de pérdidas en alimentadores primarios para enel sa regional el oro,” *ESP*, 2012.
- [4] G. P. Cifuentes Casquete y A. A. García Torres, “Diseño e implementación de un equipo de contrastación de medidores monofásicos.” *B.S. thesis*, 2010.
- [5] P. De Jarmy Villarreal y P. D. J. Villarreal, “Demanda, demanda máxima y carga instalada,” *Sistemas de distribución*, 2018.
- [6] J. D. Flores Coaguila, “Propuesta de modelo de detección de fraudes de energía eléctrica en clientes residenciales de lima metropolitana aplicando minería de datos,” *renati.sunedu.gob.pe*, 2014.
- [7] T. Gonen, *Electrical Power Transmission System Engineering*. Apple Academic Press, 2014.
- [8] LEGRAND, *TABLEROS A NORMA*, 2020.
- [9] T. M. Mitchell, *Machine Learning (1 ed.)*. McGraw-Hill, Inc., 1997.

- [10] B. Mundial, “Programa de las naciones unidas para el desarrollo (pdnu)” programa de mejoramiento de la eficiencia del sector eléctrico,” *Reducción de pérdidas en la distribución de la electricidad Ecuador*, 1992.
- [11] J. M. O. Osorio, *ESTIMACION DE PERDIDAS TECNICAS EN EL SISTEMA DE LA EMPRESA ELECTRICA MUNICIPAL DE SAN PEDRO SACATEPEQUEZ, SAN MARCOS*. Universidad de San Carlos de Guatemala, 2006.
- [12] R. D. T. RAMIREZ, “Metodología para la deteccion de perdidas no tecnicas en sistemas de distribucion utilizando metodos de mineria de datos,” 2014.
- [13] W. O. Sánchez Sánchez, K. E. Morales Soriano, *et al.*, “Identificación y control de pérdidas de energía en el sistema de distribución secundaria,” B.S. thesis, Instituto de Información Científica y Tecnológica Cuba, 2000.