

Índice

1. Entendimiento del negocio	2
1.1. Objetivos del negocio	2
1.2. Evaluación de la situación	3
1.3. Objetivos desde la analítica	6
1.4. Plan de acción	6
2. Entendimiento de los datos	7
2.1. Recolección	7
2.2. Descripción	7
2.3. Exploración	7
2.4. Calidad	8
3. Preparación de datos	8
4. Modelado	8
5. Evaluación	8
6. Despliegue	8

Proyecto Consumo de Energía del Metro de Medellín

4 de diciembre de 2025

1. Entendimiento del negocio

1.1. Objetivos del negocio

El Metro de Medellín es el sistema de transporte masivo más importante de la ciudad y un referente nacional en gestión operativa y sostenibilidad urbana. En su compromiso por mejorar la eficiencia energética, actualmente la empresa avanza hacia la acreditación ISO 50001, un estándar que exige un control más riguroso, trazable y eficiente del uso de la energía.

Sin embargo, el proceso de seguimiento al consumo y al costo energético sigue siendo completamente manual, lo que dificulta el análisis y limita la capacidad de respuesta. Cada mes, el Metro recibe aproximadamente 817 facturas de energía, provenientes de diversas fuentes de consumo distribuidas en tres grandes áreas: negocios (locales), infraestructura y sistemas de transporte, y soporte administrativo (+600). Hoy en día, un único departamento se encarga de recibir las facturas físicas, digitalizarlas y realizar un análisis descriptivo básico, lo que genera cuellos de botella y poca capacidad de profundización analítica.

Este reto se vuelve aún más relevante considerando que, para 2027, el Metro incorporará una nueva flota de trenes, lo que incrementará significativamente la demanda energética del sistema. Este contexto resalta la necesidad urgente de modernizar el flujo completo de gestión energética y contar con herramientas que permitan procesar, analizar y planificar el recurso eléctrico de manera más eficiente.

A partir de lo discutido con la empresa, se establecieron los siguientes objetivos:

- Automatizar el procesamiento de datos energéticos al transformar facturas electrónicas en formato digital a datos estructurados.
- Visualizar el consumo o costo de energía mediante diferentes métricas de negocio para identificar estaciones y plantas de suministro con mayor impacto.
- Pronosticar el consumo o costo energético agregado del Metro con datos históricos

Las metas para el metro son transformar y guardar la información de las facturas en la nube, preferiblemente en AWS, analizar informes gráficos y métricas del negocio para conocer tendencias y realizar proyecciones.

1.2. Evaluación de la situación

El equipo del proyecto está conformado por los miembros del semillero y por el personal del Metro, principalmente Lina Parra, del área de Investigación, Desarrollo e Innovación.

Como fuente principal de datos, el Metro comparte las facturas en formato PDF de todas las estaciones de energía y, en algunos casos, archivos en Excel con datos ya estructurados. A medida que estos documentos se transforman y limpian, la información se incorpora a la laguna de datos para su uso.

El proceso ETL de las facturas puede desarrollarse utilizando servicios de AWS o mediante código propio. Para la visualización de resultados, se espera trabajar con Amazon QuickSight o, en su defecto, Power BI. Finalmente, el entorno principal de desarrollo es Jupyter Notebook para analítica de datos y documentar el flujo de trabajo, utilizando Python 3 y sus librerías especializadas.

Requisitos	Suposiciones	Restricciones
<ul style="list-style-type: none">■ Disponibilidad de datos históricos completos■ Herramientas de pago para almacenamiento en la nube	<ul style="list-style-type: none">■ Las facturas reflejan el consumo energético real de las plantas■ Los datos de las diferentes facturas presentan unidades comunes■ No hay factores externos en las operaciones de las plantas que puedan afectar el consumo energético	<ul style="list-style-type: none">■ Las facturas solo reflejan la ubicación de la planta eléctrica, mas no las estaciones que alimenta■ No se consideran otras características de cada servicio: cantidad de trenes, buses, teleféricos y tranvías; la eficiencia kV/km según medio de transporte.■ Se desprecia la naturaleza humana en la operación de los medios de transporte.

Tabla 1: Resumen de los requisitos, suposiciones y restricciones del proyecto

Riesgo	Contingencia
Cambios en el formato de factura	Diseñar un pipeline flexible de ETL con validaciones automáticas
Factores externos: sequías → generación eléctrica; desastres naturales o humanos → reducción consumo	Añadir variables indicadoras que reflejen comportamientos imprevistos
Factores antropológicos (ferias, conciertos, demanda social)	Incorporar indicadores de eventos especiales en el modelado
Errores de medición en la factura electrónica	Comparar valores de consumo en la factura con los contadores que tiene el Metro
Daños en las plantas eléctricas que alimentan a las estaciones	Identificar las líneas afectadas y evaluar la opción de trabajar con un subconjunto que tenga datos completos. En caso extremo, deshabilitar en el proyecto las líneas dañadas
Mala especificación de los modelos necesarios para cumplir los supuestos	Comparar diferentes modelos que puedan cumplir con los supuestos

Tabla 2: Riesgos más probables y contingencias posibles para mantener a flote el proyecto

Tabla 3: Terminología usada en las facturas y en el negocio, donde \$ está en COP

Variable	Descripción	Formato	Tipo de dato
Periodo inicio	Fecha de inicio del periodo facturado	AAAA-MM-DD	Fecha
Periodo fin	Fecha de fin del periodo facturado	AAAA-MM-DD	Fecha
Días facturados	Número de días incluidos en el periodo de facturación	días	Entero
Consumo kwh	Energía importada/consumida en el periodo	kWh	Numérica
Excedentes kwh	Energía inyectada a la red por autogeneración	kWh	Numérica
Saldo neto kwh	Consumo neto (consumo – excedentes)	kWh	Numérica
Promedio 6m kwh	Promedio mensual de consumo de los últimos 6 meses	kWh	Numérica

Continúa en la siguiente página

Variable	Descripción	Formato	Tipo de dato
Precio unitario energía	Precio unitario de energía consumida	\$/kWh	Numérica
Precio unitario excedente	Precio unitario pagado por excedentes	\$/kWh	Numérica
Costo generación unit	Costo unitario de generación	\$/kWh	Numérica
Costo transmisión unit	Costo unitario de transmisión	\$/kWh	Numérica
Costo distribución unit	Costo unitario de distribución	\$/kWh	Numérica
Costo comercialización unit	Costo unitario de comercialización	\$/kWh	Numérica
Costo perdidas unit	Costo unitario por pérdidas	\$/kWh	Numérica
Costo restricciones unit	Costo unitario por restricciones	\$/kWh	Numérica
Contribución activa pct	Porcentaje de contribución activa	%	Numérica
Costo energía total	Valor total facturado por energía	\$	Numérica
Costo excedentes total	Valor total pagado por excedentes (negativo si descuenta)	\$	Numérica
Costo contribución total	Valor total de la contribución activa	\$	Numérica
Otras entidades total	Valor cobrado por otros conceptos (ej. alumbrado público)	\$	Numérica
Capacidad instalada kw	Potencia instalada del autogenerador	kW	Numérica
Nivel tensión volt	Nivel de tensión del suministro	Voltios	Entero

Continúa en la siguiente página

Variable	Descripción	Formato	Tipo de dato
Categoría servicio	Tipo de usuario (Industrial, Comercial, Residencial, etc.)	Texto	Entrada
Plan tarifario	Nombre del plan de energía	Texto	Entrada
Total a pagar	Valor total de la factura (incluye otros cobros)	\$	Numérica

Costos:

- El proyecto no cuenta con financiación externa; los recursos provienen del semillero.
- El único costo significativo es el uso de servicios AWS, estimado hasta en \$USD 25 mensuales según la volumetría y el tiempo de uso.

Beneficios:

- Generación de nuevo conocimiento en metodologías y herramientas matemáticas, fortaleciendo la formación de los estudiantes.
- Entrega al Metro de bases sólidas para avanzar hacia una solución integral de gestión y planificación energética.
- Identificación de periodos de alto consumo para optimizar recursos.
- Predicción de consumo de energía a corto plazo.
- Detección de consumos anómalos o facturas irregulares, mejorando el control y la prevención.
- Simulación de escenarios futuros, como impactos de cambios en tarifas o estrategias de reducción de consumo.

1.3. Objetivos desde la analítica

Realizar pronósticos fuera de muestra para el año 2025 del consumo de energía en las principales estaciones de energía, utilizando modelos de predicción basados en datos mensuales entre 2019 y 2024, para mejorar la planificación del recurso eléctrico.

El éxito de la predicción se evaluará validando el modelo y verificando que el error de pronóstico se mantenga dentro del umbral definido por la empresa.

1.4. Plan de acción

En la hoja Metro del archivo de Excel del repositorio del semillero, se encuentra la programación semana a semana de las etapas del proyecto y las reuniones con el personal del metro.

Por otro lado, preparar archivos estructurados donde las fechas actúen como índice, las direcciones de facturación correspondan a las columnas y cada uno contenga únicamente una de las variables clave del estudio: consumo de energía, costo total y costo por kWh

Se consideran métodos estadísticos tradicionales, técnicas de machine learning y modelos basados en redes neuronales. Cada grupo deberá entrenar y validar sus modelos, aplicar las métricas de evaluación correspondientes y resumir los resultados obtenidos.

2. Entendimiento de los datos

2.1. Recolección

Mencionar la fuente en la que salen los datos y cómo es la metodología para medir estos datos.

Campo	Descripción	Valores
direccion	Dirección donde se presta el servicio de EPM (ej. Cl 18 Sur Cr 39 A - 235)	Texto
periodo_consumo	Periodo de facturación correspondiente al consumo (ej. 01 nov al 30 nov)	Texto
consumo_kwh	Cantidad de energía consumida en kilovatios-hora (kWh) durante el periodo de consumo	Numérica
valor_total_pagar	Valor monetario total en COP de la factura a pagar en el periodo (suma de todos los servicios facturados)	Numérica
tarifa_kwh	Precio en COP por un kwh consumido, esto lo establece la entidad prestadora de energía	Numérica
archivo	Nombre del archivo de origen (PDF de la factura) del cual se extrajo la información	Texto

Tabla 4: Diccionario de datos empleados

2.2. Descripción

Aquí se menciona el formato de cada variable, la cantidad de registros no vacíos, los identificadores únicos en cada campo y otras posibles descripciones que permitan cumplir los requerimientos del proyecto. Recomendación: utilizar el .info y .describe de pandas

2.3. Exploración

Tablas resumen como dashboards, graficas de series de tiempo y un mapa que ubique los puntos de facturación.

2.4. Calidad

Nos basamos en 6 dimensiones ampliamente analizadas en la literatura (Collibra, [2023](#)).

3. Preparación de datos

4. Modelado

5. Evaluación

6. Despliegue

Referencias

Collibra (2023). *The 6 data quality dimensions with examples*. URL: <https://www.collibra.com/blog/the-6-dimensions-of-data-quality> (visitado 12-02-2025).