**Etapas del Proyecto de Análisis Tarifario – Dashboard Interactivo**

**1. Planteamiento del Problema**

**Mi punto de partida fue hacerme una pregunta:  
¿Cómo varían los costos del servicio de acueducto y alcantarillado en los diferentes estratos y municipios, y qué implicaciones tiene esto para las comunidades?**

Decidí trabajar con una base de datos pública de tarifas por servicios de acueducto y aguas residuales (EPM) desde 2017 hasta 2024. Contenía múltiples variables como:

* Año, mes, municipio, estrato
* Tipo de servicio (acueducto o alcantarillado)
* Sector (residencial, comercial, industrial, oficial exento)
* Cargos: fijo, por consumo mínimo, por consumo máximo, reconexión, suspensión, etc.

Desde un principio, quise que los datos hablaran por sí solos, pero sin perder de vista que un buen análisis te dice algo a través de los datos, pero me tope algunos desafíos por el camino y decisiones que tuve que realizar para priorizar la brevedad y la buena lectura de los datos; Adelante les podre compartir el paso a paso de este proyecto personal.

**2. Proceso ETL (Extracción, Transformación y Carga)**

Este fue uno de los desafíos más importantes, ya que el conjunto de datos:

* Tenía más de 40 mil filas con distintos esquemas de cobro según el sector una cantidad de datos gigantesca y con mucho de donde tomar.
* Mezclaba cargos que aparecían como rangos (mínimo y máximo) en estratos residenciales, pero como valores únicos en los demás sectores en este al darme cuenta de las diferencias de mínimo y máximo en los sectores logre averiguar que a los estratos del 1 al 3 pueden recibir bonos que les ayudan al momento de pagar las tarifas, un dato bastante importante.
* Incluía meses como texto (enero, febrero…), lo cual impedía un orden cronológico directo y también tuve en claro desde este momento que mezclar todos los meses a lo largo de tantos años podría dificultar el storytelling de la grafica.

**Transformaciones aplicadas:**

* Normalicé los nombres de municipios, estratos y meses.
* Creé una columna auxiliar de "mes numérico" para ordenar correctamente debido a que el programa los organizaba por defecto alfabéticamente, este cambio me ayudo a darle un orden cronológico.
* Separé los datos por tipo de servicio en este caso me centré en el del acueducto por estrato para permitir comparaciones claras.
* Traté los valores nulos y establecí filtros para facilitar el análisis por año, mes y lugar algo fundamental.
* La base de datos contaba con un error en su base de datos que constaba que el municipio de la estrella tenia datos divididos en la variable “la estrella’’ y otro que decía ‘’Estrella’’ me encargue de corregir este error y combinar los datos

Este trabajo de limpieza tomó tiempo, pero fue crucial para que el análisis no solo para que no fuera confuso sino que también para su uso correcto.

**3. Foco del Análisis: Estratos y Municipio a lo Largo del Tiempo**

**Fase 1: Delimitación del análisis**

**Decisión**: Excluir el servicio de alcantarillado y los sectores no residenciales a pesar de que se encuentren en la base de datos (comercial, industrial, oficial, exento).  
**Razonamiento:**

* Los sectores no residenciales no presentan distinción por estratos ni rangos mínimos/máximos, lo que limita el análisis comparativo.
* El servicio de alcantarillado tiene una estructura tarifaria más homogénea y menos expresiva visualmente.
* En cambio, el acueducto residencial sí permite observar diferencias reales de cobro que afectan a hogares con diferentes capacidades económicas.
* Pienso que los sectores que excluí se diferencian tantos de los de los estratos que no me imagino a una persona sin experiencia en filtración de datos complicándose para ver cuanto paga aproximadamente en su estrato.
* Decidí separar el alcantarillado porque siento que a pesar de que son ramas parecidas son cobros distintos por mucho que los mismos los realice la misma EPM, puede confundir a las personas y dificulta la visualización de los datos.

*‘’He decidido enfocar este proyecto para que cualquiera pueda visualizarlo’’*

**Fase 2: Enfoque social**

**Decisión:** Analizar exclusivamente los estratos socioeconómicos residenciales.  
**Motivación:**

* Representan directamente a la ciudadanía, especialmente a hogares vulnerables.
* Los estratos 1 al 3 muestran variaciones significativas entre el cargo mínimo y el máximo, lo que permite analizar subsidios o posibles injusticias tarifarias.
* El análisis deja de ser puramente técnico para plantear preguntas sociales:  
  ¿Pagamos lo justo según dónde vivimos y nuestro nivel económico?
* Los estratos 1 al 3 según el municipio reciben variaciones de costos demasiado notorias lo cual llama la atención rápidamente y quiero que se note fácilmente para cualquiera.

**Fase 3: Análisis territorial**

**Decisión:** Incluir la opción de una comparativa entre municipios.  
**Motivación:**

* El mismo estrato no paga lo mismo en Medellín que en Bello o Itagüí.
* Esto revela desigualdades en la política tarifaria que no se entienden si se analiza el dato de forma global.
* Al hacer comparaciones cruzadas por municipio y estrato, se hace visible si hay coherencia o disparidades significativas entre territorios.
* Esto ayuda a ver las desigualdades no solo del pago de acueducto si no de diferencias de condiciones en lo que serian los estratos del 1 al 3 que son los que reciben bonos.

**Fase 4: Análisis temporal**

**Decisión:** Incorporar la dimensión mensual y anual para entender variaciones.  
**Motivación:**

* Permite detectar incrementos estacionales o patrones que llamen la atención en los tiempos más recientes.
* Aunque agrupar año y mes fue algo complejo visualmente, decidí centrar el análisis en todos los meses de 2024 ofreció equilibrio entre detalle y legibilidad.
* Un tiempo fijo añade brevedad narrativa al análisis, ayudando a contextualizar decisiones tarifarias.
* Decidí llevar el análisis de forma concurrente a la actualidad, no me pareció pertinente meter todos los años de la base de datos para no colapsar visualmente si no que por el contrario hacerlos del año anterior.

**Fase 5: Enfoque visual y claridad narrativa**

**Decisión:** Usar gráficas diferenciadas para cargos fijos, cargos minímo/máximo y reconexión/suspensión.  
**Motivación visual y comunicativa:**

* El rango de cobro mínimo/máximo en acueducto ofrece mayor expresividad visual, especialmente en estratos bajos y es fundamental para el analisis ya que nos permite observar las diferencias entre cobros.
* Mostrar todos los datos en una sola gráfica generaba ruido y confusión, por lo que se priorizó la claridad en cada pieza visual.
* El storytelling visual se construyó con criterio: cada gráfico responde a una pregunta y cada tarjeta resume una carga económica.

**4. Diseño del Dashboard y Elección de Visualizaciones**

Elegí Power BI como herramienta visual por su capacidad de ofrecer segmentadores interactivos y visualizaciones limpias. Cada gráfico se eligió con cabeza:

* **Gráfica de líneas** para visualizar cómo varía el cargo por consumo mínimo y máximo a lo largo del tiempo en los estratos 1 al 6. Este tipo de visual resalta diferencias sutiles o tendencias por municipio y estrato.
* **Tarjetas** para mostrar totales como costos de reconexión o suspensión, destacando la carga económica adicional para quienes no pueden pagar a tiempo.
* **Segmentadores**:
  + Un rango deslizante por año (slicer) + una lista de meses, para permitir comparaciones flexibles y ordenadas.
  + Segmentadores **por municipio y estrato**, permitiendo al usuario explorar variaciones locales.
* **Gráfica de pastel** para mostrar el promedio del cargo fijo por estrato, visualizando inequidades de forma directa y comprensible.

Todo esto lo diseñé con foco en la claridad y la narrativa.

**5. Lo Que Sigue: Llevarlo a la Web ----- En proceso!**

La siguiente etapa del proyecto será crear una versión web del dashboard, utilizando herramientas como:

* Flask para realizar el backend.
* Streamlit para replicar la visualización interactiva.
* Conexión a una base de datos y realizar pruebas que permitan corroborar y confirmar que se puede actualizar la gráfica a través de la actualización de la base de datos y verse reflejado en la web
* Exportación automática de estadísticas claves en Excel o PDF, útil para tomadores de decisiones o entidades de control Como podrían ser los datos que he puesto en las distintas tarjetas en el dashboard (Un ejemplo serian los promedios).

Además, quiero que el usuario también interactue con los respectivos filtros personalizados para que se pueda seleccionar su municipio y estrato, y obtener un resumen tarifario de su realidad cosa que podría.

**Preguntas:**

**¿Para ti que es este proyecto?**

Este proyecto es más que un análisis de tarifas. Es un ejercicio de justicia de datos.  
Quise que cualquier persona, incluso sin conocimientos técnicos, pueda ver en gráficos lo que muchas veces se esconde en tablas densas o documentos PDF.

Para mí, trabajar con datos es ayudar a que la información sea accesible, comprensible y útil. Y este es solo el primer paso.

**¿Por qué esta base de datos?**

Para este proyecto me mantuve unas horas navegando en distintas bases de datos públicas de mi país Colombia, estaba enfocándome en un problema de ámbito social y algo que pueda impactar a las personas independientemente de su situación; La mayoría no me llamaba la atención y se sentían incompletas teniendo una o dos tablas de las cuales no podría hacer un análisis completo mostrando todos mis conocimientos sobre base de datos, limpieza y análisis de estos

**¿Qué fue lo más difícil del proyecto?**  
Sin duda, encontrar una visualización que contara bien la historia sin volverse confusa. A veces los datos no se ven bien todos juntos, otras veces se ven demasiado bien y no reflejan lo que pasa al filtrar. Me costó mucho balancear eso y decidir qué mostrar. También fue un reto no estresarme cuando sentía que nada se veía bien.

**¿Qué aprendiste haciendo esto?**  
Aprendí que a veces no necesitas hacer modelos complejos para contar algo importante. Lo más valioso fue aprender a elegir el enfoque, limpiar bien los datos, y contar algo claro con ellos. También aprendí a tener paciencia con los errores y a confiar más en mi criterio.

**¿Qué te gustaría mejorar?**  
Me gustaría llevar esto a la web para que más personas puedan explorarlo sin necesidad de tener Power BI. También me gustaría automatizar la limpieza de datos para cuando la base se actualice, y hacer más comparaciones con otras regiones del país.

**¿Qué sigue después de esto?**  
Seguir aprendiendo. Quiero practicar más SQL y Python para automatizar análisis y limpieza. También quiero hacer más proyectos sociales con datos públicos y armar un portafolio donde se note mi estilo: análisis con sentido humano, no solo números.

**¿Qué parte te gustó más?**  
Ver cómo todo empieza a tener sentido. Cuando una gráfica por fin muestra lo que tú ya intuías, o cuando descubres una diferencia fuerte entre municipios y dices “esto nadie lo está contando”. Me gusta cuando los datos se vuelven voz.