PROYECTO INGENIERIA DE DATOS - BOOTCAMP IAN SAURA

EJERCICIO 1:

Empresa ficticia: **In-nova Tech**

- In-nova Tech es una empresa e-commerce de productos tecnológicos que vende artículos de tecnología (notebooks, auriculares, monitores y accesorios) .
  Los pedidos llegan desde la app web o mobile, y el equipo de sistemas construyó un pipeline de datos para analizar las ventas y comportamiento de los clientes.

1) **Fuentes de Datos (Dummy Json)**
DummyJSON es la librería a utilizar para la ingesta de datos
La misma contiene datos de un E-commerce y es una librería abierta de datos para lectura (GET) que se comunica mediante APIs REST.

Las APIs que se van a utllizar para este proyecto son la siguientes:

- DummyJSON API - /carts - Carrito de Ventas
- DummyJSON API - /products - Catalogo de Productos
- DummyJSON API - /users - Clientes

2) **Transformación**
En esa etapa lo que se hara es la limpieza, validación de los datos para convertirlos en información útil.

Tecnologias a utilizar:

- **Python + Pandas:** Extracción de APIs, limpieza y Carga.
- **SQL (PostgreSQL):** Validaciones, Joins, Agregaciones y construcciones de Vistas.
- A futuro estaría bueno aplicar la Tecnología **DBT** para gestionar las transformaciones y modelado directamente desde el DW, con control de Versionado, documentación y Validaciones.

3)**Almacenamiento (Storage)**
Se guardara los Datos Extraídos, de forma cruda o procesada, para observar el historial y organizar la información.

Tecnologias a utilizar:

- **AIVEN** (PostgreSQL en la nube): es una Base de Datos administrada en la Nube.
  Se almacena los datos de forma estructurada y permite realizar las consultas eficientemente con SQL.

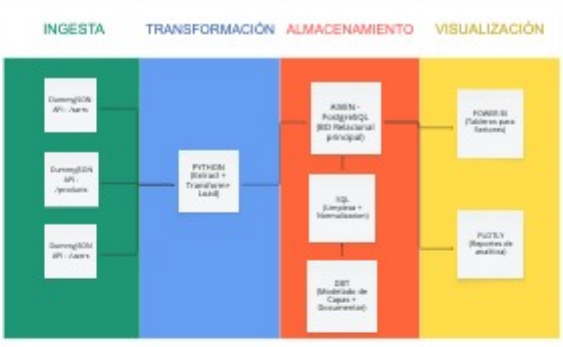
Organización de Capas:

- Bronze: Datos Crudos, obtenidos directamente de las APIs.
- Silver: Datos Transformados, validados y normalizados.
- Gold: Datos listos para Dashboard o Reportes del Negocio.

4) **Visualización**
Etapa final del proceso donde se mostraran los Datos ya procesados a través de Dashboards o reportes para facilitaciones las decisiones,

Herramientas a utilizar:

- **Power Bi:** Dashboard para diferentes Areas.
- **Plotly - Dash:** Visualizaciones personalizadas para el armado de Reportes.



EJERCICIO 2:

Breve Descripción de 6 Herramientas a utilizar

Nombre	¿Para que Sirve?	Etapa del Pipeline	Nivel de Complejidad
Pandas	Libreria de Python para Manipular los Datos: limpieza, filtros, agrupaciones, joines, entre otras.	Transformacion	Básico
Request	Libreria de Python para conectarse a las APIs y Obtener los datos externos	Ingesta	Básico
PostgreSQL (Aiven)	BD Relacional en la Nube, utilizada para almacenar y consultar los datos	Almacenamiento	Intermedio
Power BI	Herramienta utilizada para la visualizacion de los Datos (Creacion de Dashboard	Visualización	Intermedio
SQL (PostgreSQL)	Lenguaje utilizado para transformar, unir o agregar datos dentro del DW.	Transformación	Intermedio
Cron	Sistema para la automatización de Tareas (Scripts)	Orquestacion	Básico / Intermedio

EJERCICIO 3

Diferencias Entre las distintas profesiones de Data.

Nombre	Tareas Principales	Herramientas Utilizadas	Nivel de Programaciones	Ejemplo de entregable
Data Enginner	Diseña y Mantiene Piplines de ETL, asegura la Calidad de los Datos y crear modelos Escalables	Python, SQL, DBT, Cron, Airflow, Docker	Alto	Base de Datos Limpia, normalizada y Automatizada
Data Analyst	Exploracion de datos, generar Insights, construir dashboard o reportes	Excel, SQL, Power BI, Tableau, Plotly	Intermedio	Reporte de Ventas Mensual con Visualizaciones
Data Scientist	Desarollar Modelos predictivos y ML. Experimentas con funciones Estadisticas	Python, R, Sickit-learn, Jupyter	Alto	Modelo de Prediccion de Ventas (Forecast)
Analytics Enginner	Crear Modelos transformados dentro del DW, documentar y Testear	Dbt, SQL, BigQuery	Intermedio	Vistas limpias de una tabla en concretos (Ventas) con documentacion en Dbt

Con respecto al rol que mas me interesa, la verdad todavía no tengo un favorito definido dentro del mundo de datos.

Vengo del análisis, que es un área que me gusta mucho, pero a medida que fui aprendiendo más sobre la ingeniería de datos, me fui dando cuenta de algo clave: **sin datos bien organizados, limpios y estructurados, es muy difícil sacar conclusiones útiles**.

Me está interesando cada vez más entender cómo se construyen los pipelines que hacen posible que los datos estén listos para analizar.
Me gusta la idea de armar una base sólida sobre la que pueda trabajar con confianza y sabiendo que lo que se muestra / presenta es informacion confiable.

Creo que, para seguir avanzando, necesito fortalecer mis conocimientos en automatización, SQL, programación y buenas prácticas de modelado.
Siento que este camino me va a dar una visión más completa de todo el ecosistema de datos, y eso me entusiasma.