Report for Modern ELT Pipeline

Theme Digital Health: Telemedicine adoption, mental health tech, wearable devices

This project explores Digital Health through its examination of telemedicine adoption and mental health technologies and wearable health devices. The aim is to create a data engineering pipeline which could process multiple data sources to track the development of digital healthcare solutions.

This project uses three different data sources:

- A public CDC API dataset related to telemedicine usage
- Google Trends time-series data for relevant keywords
- A structured local CSV dataset downloaded from Kaggle\

The pipeline followed an ELT approach which involved extraction, storing raw data files and cleaning and transforming data before generating insights and visual analysis.

Key Findings:

- Telemedicine interest increased significantly especially during the COVID-19 period, which shows rapid adoption of remote healthcare services.
- People maintained high search activity for mental health technology which shows their growing acceptance of digital mental health solutions.
- The API data required flattening because it contained nested JSON structures.
- The majority of datasets appeared clean but the process required standardization of date formats and categorical values.

Challenges Encountered:

- Identifying the correct API endpoint from the CDC platform.
- Handling semi-structured JSON data and converting it into usable tabular form.
- Standardizing date formats and categorical values across datasets.
- Ensuring proper organization of raw and cleaned data for reproducibility.