Student Attendance Analytics Project Summary

Overview

This report summarizes the end-to-end data analytics workflow performed on the student attendance dataset. The project includes loading and exploring the dataset in Python, cleaning and transforming the data, running SQL queries in PostgreSQL, and building an interactive Power BI dashboard to visualize attendance insights.

Dataset

The dataset includes daily attendance information for multiple schools, containing the following fields:

- School DBN

- Date

- Enrolled

- Absent

- Present

- Released

The dataset was loaded using Python, processed, and later used for SQL and Power BI analysis.

Tools Used

- Python (Pandas, NumPy, Matplotlib)

- PostgreSQL & SQL

- Power BI

- Jupyter Notebook

- Excel

Project Steps

1. Data Loading:

  - Imported data using Pandas.

  - Verified column types and structure.

2. Exploratory Data Analysis (EDA):

- Summary statistics and distributions.
- Trends in daily attendance.
- Visual insights (Present vs Absent).

3. Data Cleaning:

- Fixed missing data.
- Standardized date formats.
- Created new calculated fields such as attendance_rate.

4. SQL Analysis (PostgreSQL):

- Loaded cleaned dataset into PostgreSQL.
- Executed SQL queries including:
  - Total absences by school
  - Attendance rate trends
- -Find average daily attendance (Present) for each school.
- -List days where more than 100 students were absent.
- -Count number of days each school reported attendance.
- -Find the percentage of students released each day.
- -Rank days by attendance rate for each school.

5. Power BI Dashboard:

- Developed interactive visualizations:
  - -Top 5 Sum of Present Students by School.
  - -Top 5 Sum of Released Students by School.
  - Line charts for attendance trends.

- Cards for attendance metrics.

- School comparison visuals.

- Present vs. Absence Scatter Plot for Comparison Visuals.

-

Dashboard Summary

The Power BI dashboard provides:

- Attendance trends across time

- School-level performance comparisons

- Key metrics for decision-making

- Absence distribution patterns

Key Insights

- Several schools show consistently strong attendance.

- Certain dates show spikes in absenteeism.

- Daily attendance rate highlights patterns that can support strategic interventions.

- Visualizations allow stakeholders to compare attendance across schools effectively.

How to Run the Project

1. Clone the repository.

2. Install dependencies with: pip install -r requirements.txt

3. Run the Jupyter Notebook to explore and clean the dataset.

4. Import cleaned data into PostgreSQL and run SQL queries.

5. Open the Power BI .pbix file and refresh all data connections.