**INFERA DATA SCIENCE INTERNSHIP**

**Daily Progress Report**

**Date:** 12th June 2025  
**Intern Name:** Aniruddh Vijayvargia  
**Project:** Multi-Dataset Analysis – Week 1 Assignment  
**Report Day:** Day 2

**1. Executive Summary**

On Day 2, I focused on advanced **data preprocessing**, **feature engineering**, and **exploratory data analysis (EDA)** for both the **climate** and **pedestrian traffic** datasets. This included transforming raw data into structured formats, deriving meaningful features, and generating analytical visualizations. The decision to treat climate and traffic datasets separately allowed for more precise insights in their respective domains.

**2. Tasks Completed Today**

* Cleaned and preprocessed climate data covering Jan–Mar, May–Jul, Oct–Dec 2024 & Jan 2025.
* Transformed traffic dataset from wide to long format for over 100 Melbourne pedestrian sensors.
* Engineered new features (e.g., temperature range, rain category, time periods).
* Created and interpreted 12+ visualizations showing climate patterns and urban mobility trends.
* Exported cleaned datasets and summary statistics for modeling in Day 3.

**3. Technical Work Details**

**Climate Dataset**

* Imputation using median and forward fill for missing meteorological values.
* Created:
* Temperature range (max - min)
* Rain category (none, light, moderate, heavy)
* Season classification (Summer, Autumn, Winter, Spring)
* Visualizations:
* Temperature histograms, monthly trends
* Rainfall category pie charts
* Seasonal boxplots

**Traffic Dataset**

* Converted hourly wide-format to long-format with 2,448+ time-stamped entries.
* Derived fields:
  + **Sensor category** (Transport, Retail, etc.)
  + **Time period classification** (Morning, Daytime, Evening)
  + **Daily summaries** (total, average, peak, std deviation)
* Visualizations:
  + Peak hour line plots
  + Location-wise bar graphs
  + Heatmaps (hour vs location type)

**4. Key Learnings & Insights**

**Climate**

* Peak average temperatures in January, lowest in July.
* 47% of days were rainy; heavy rain was rare (4%).
* Comfort index suggests Autumn & Spring are most pleasant.

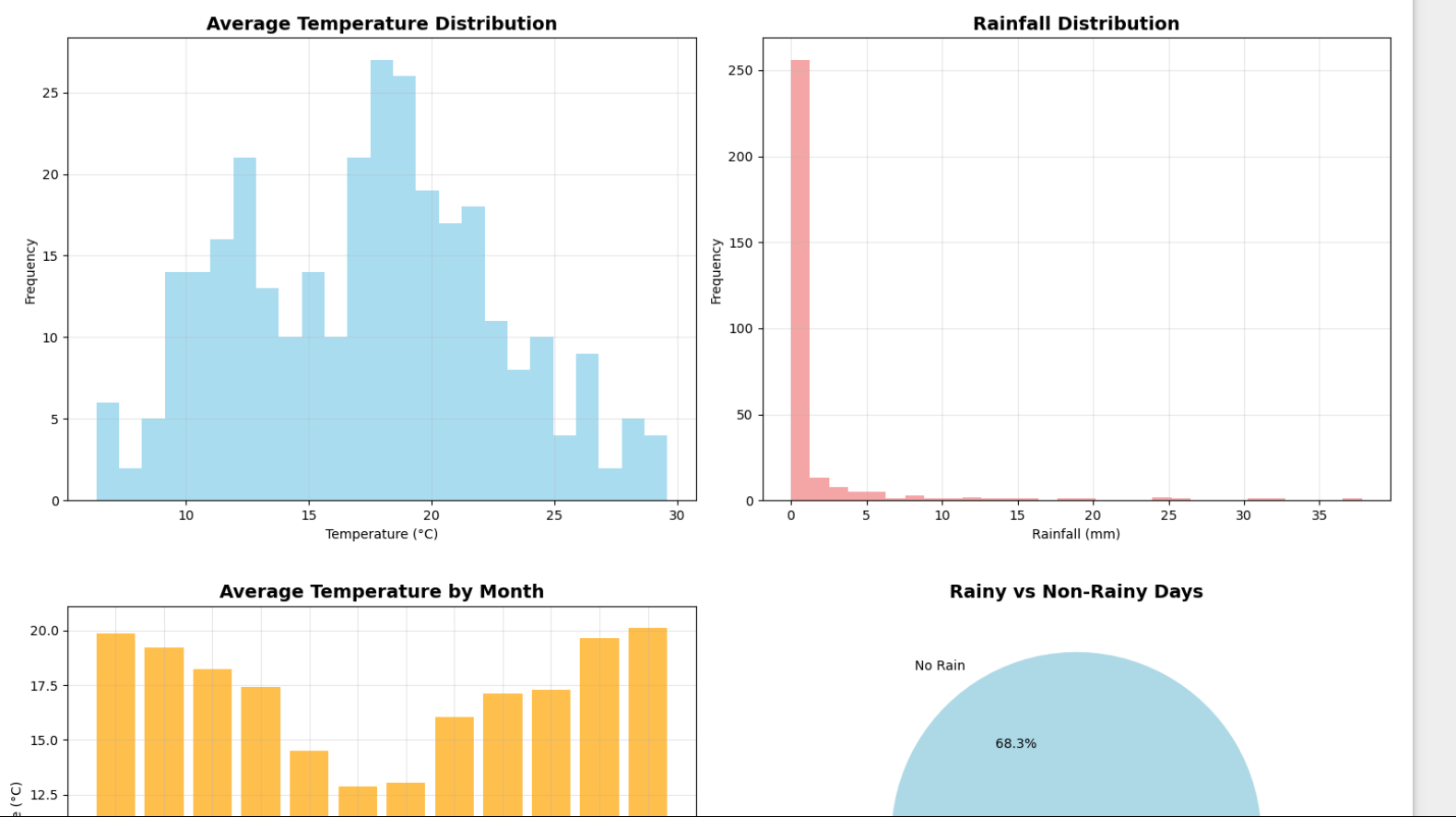
**Traffic**

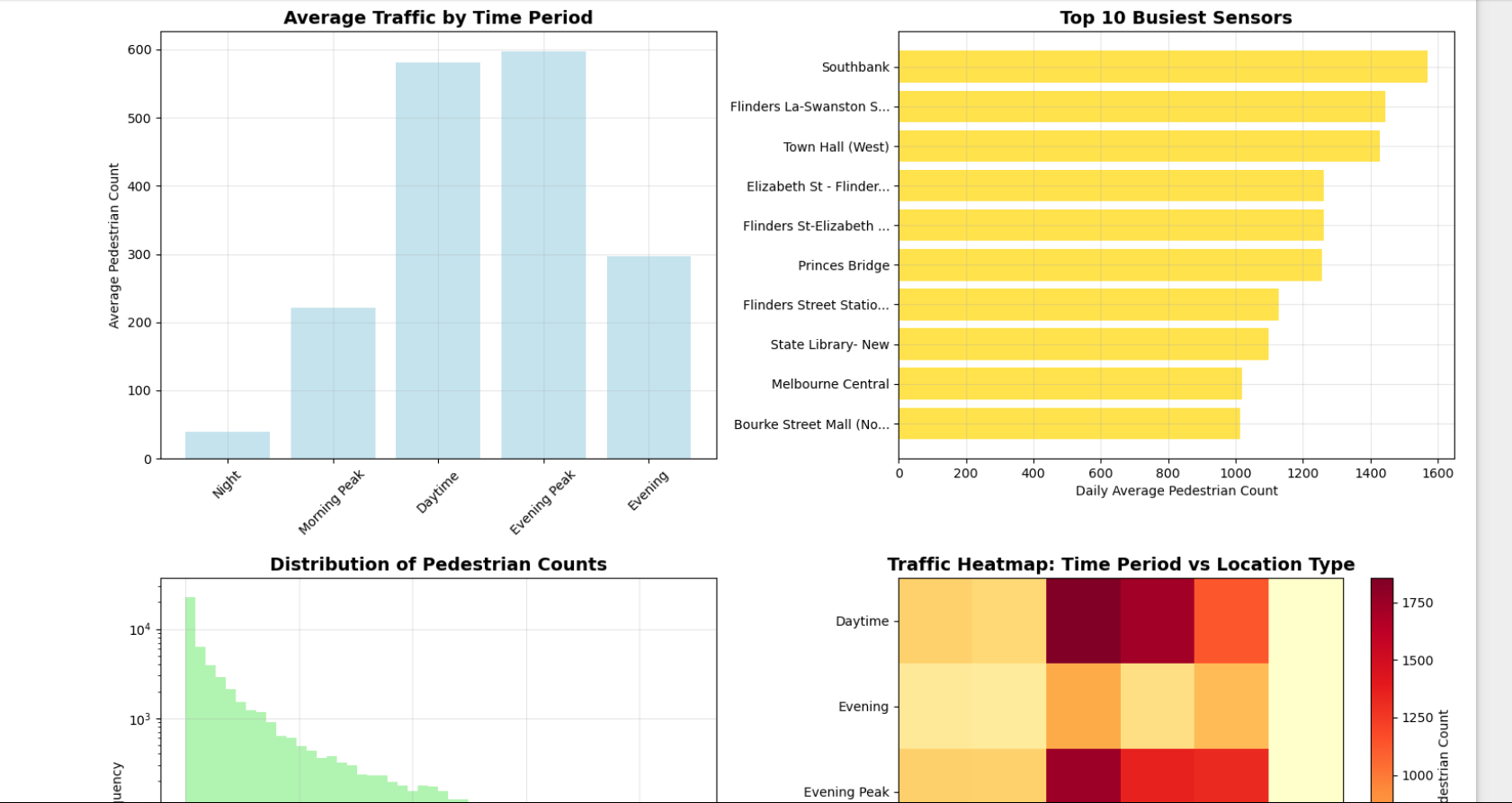
* Peak pedestrian movement at **1:00 PM**.
* Transport Hubs receive **3x more footfall** than other areas.
* Evening Peak (5–7 PM) consistently shows higher activity than mornings.

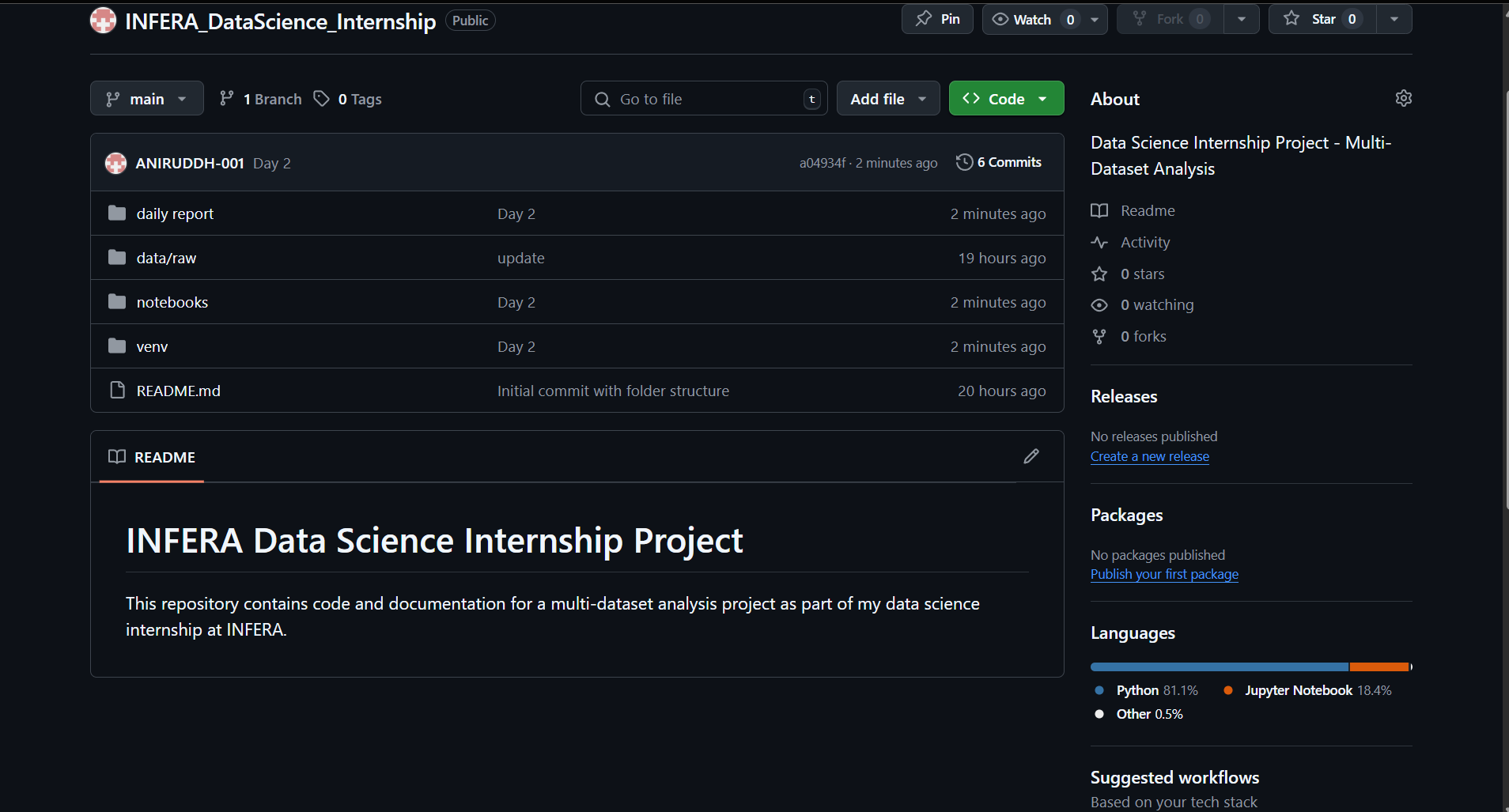
**5. Challenges Encountered**

* Imputing missing values in critical columns (evaporation, sunshine) while preserving natural patterns.
* Standardizing inconsistent datetime formats across multiple CSVs.
* Transforming and validating wide-format traffic data without losing time-alignment.
* Ensuring visualizations remain interpretable for large data volumes.

**7. Screenshots & Evidence**

****

****

****

**8. Tomorrow’s Action Plan**

* **ML modeling** for both datasets:
* Time-series or regression model for climate predictions.
* Classification/regression for traffic flow forecasting.
* Train-test splitting, metric evaluation (MAE, RMSE, Accuracy, ROC).
* Model interpretation and stakeholder insight generation.