Seoul Bike Sharing Demand Analysis

Written Report

Carlo Alberto Sartori

Project Notebook:

See section headings in "LISA Project Notebook.ipynb"

1. Dataset Overview

- Source: Seoul Bike Sharing Demand dataset (2017–2018) with 8760 hourly records.
- Target variable: "Rented Bike Count".
- Features: hourly weather conditions, holiday flag, season, functioning day.

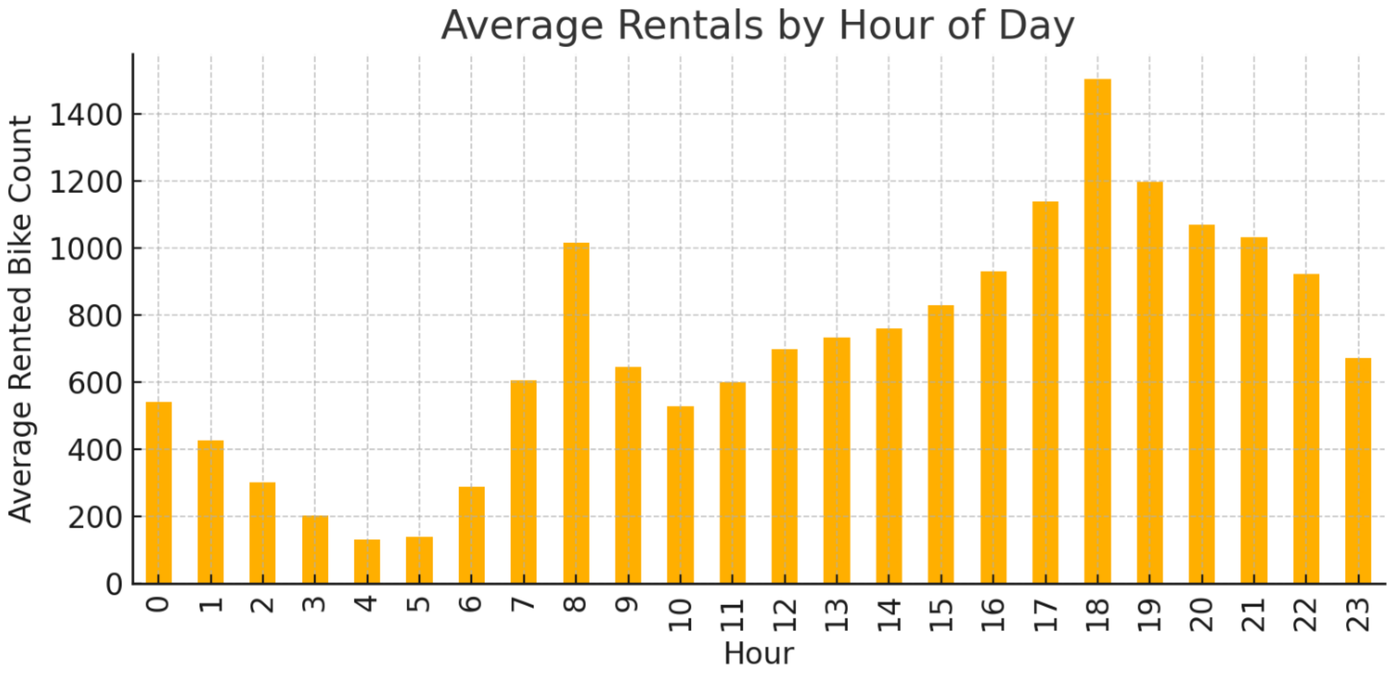
2. Methodology

Section 2 of the notebook ("Initial Setup and Import of Libraries") establishes the working environment. Section 3 ("Data loading and cleaning") handles missing-value checks and type conversion. Exploratory analysis (Section 4) studies temporal patterns and weather relations, while Section 5 transforms categorical variables via one-hot encoding before model training.

We split the data 80/20 and fit a Random Forest regressor (Section 6), selected after grid search hyper-parameter tuning (Section 6.3). Metrics include RMSE and R^2 on the held-out set.

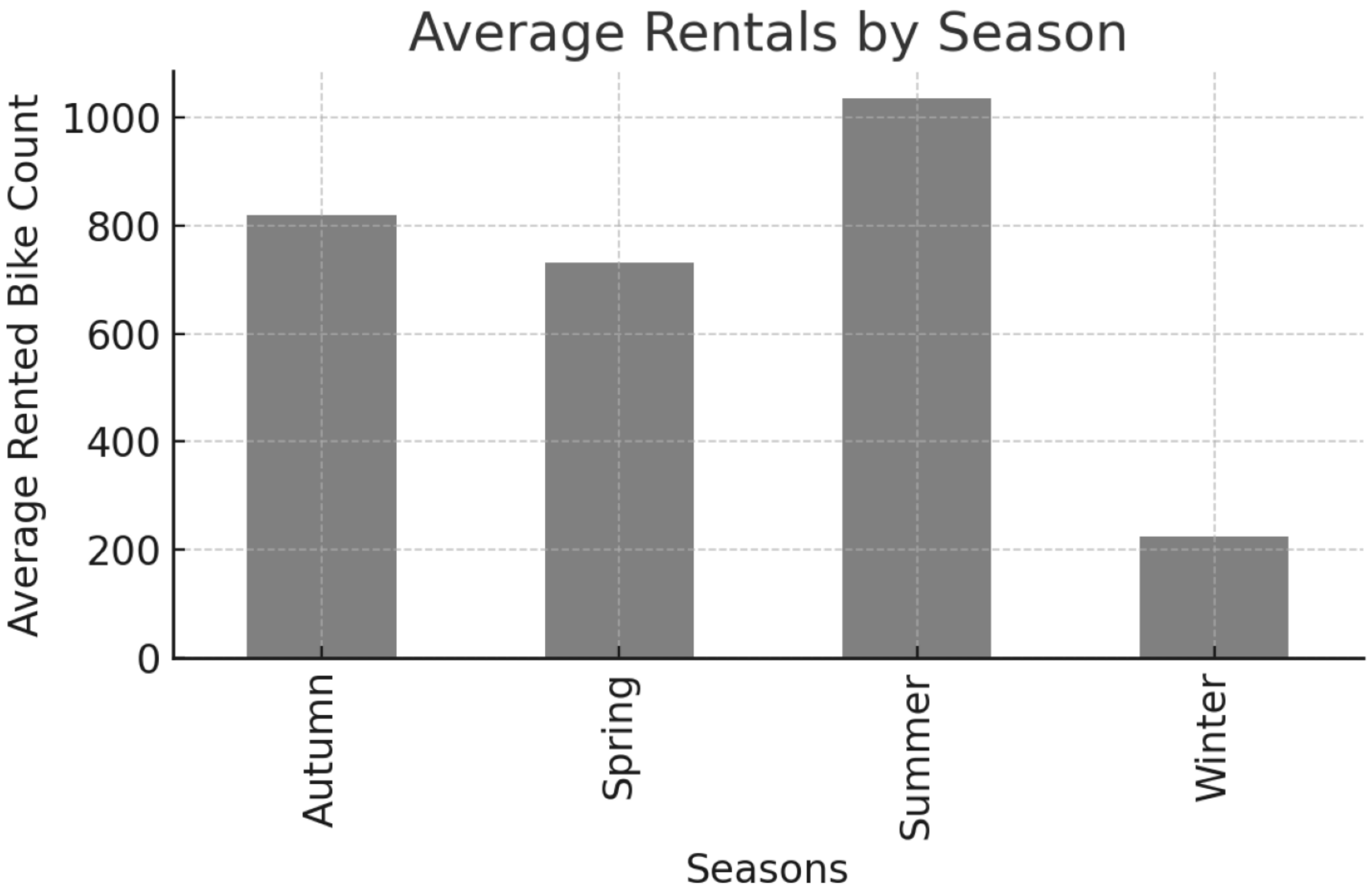
3. Exploratory Data Analysis

3.1 Average rentals by hour (Section 4)



Peak demand occurs at 18:00, coinciding with the evening commute. Overnight demand drops below 150 rentals between 03:00-05:00.

3.2 Average rentals by season (Section 4)



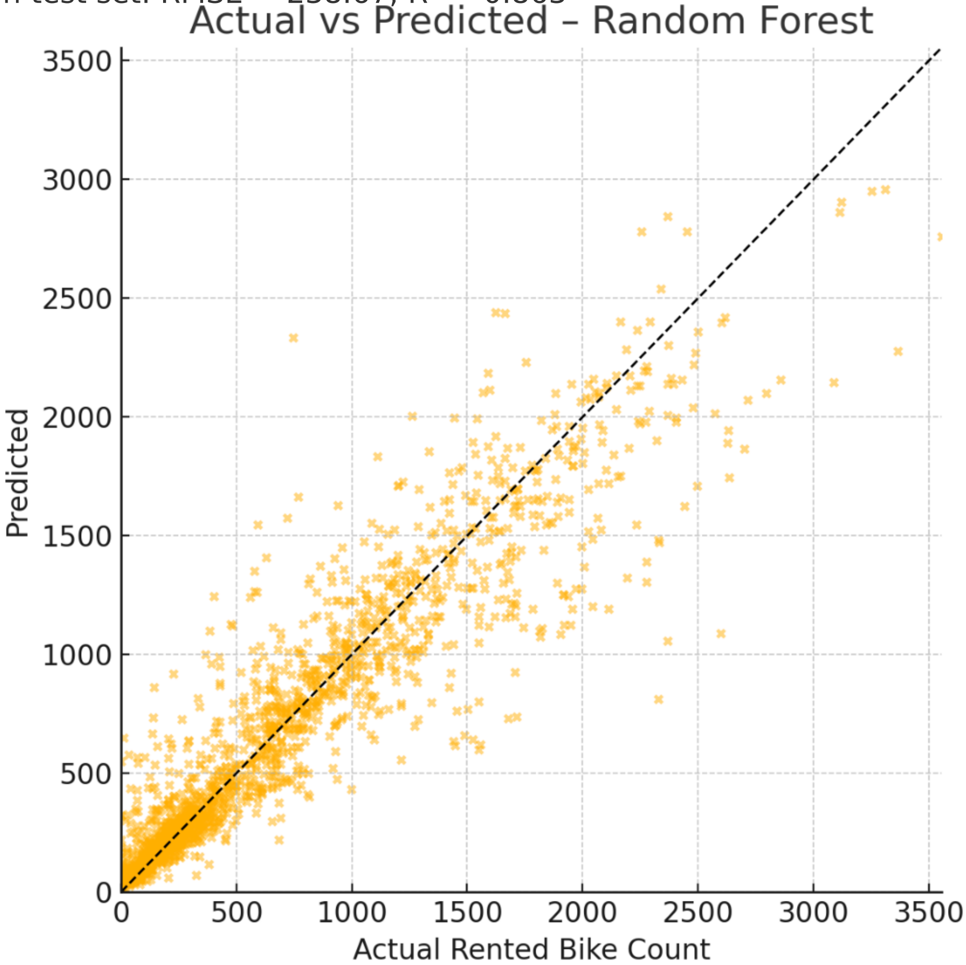
Summer leads with >1 k average hourly rentals, while Winter demand plummets to ~225.

Weather-dependent usage pattern confirms the outdoor nature of bike sharing.

4. Predictive Modeling (Section 6)

- Algorithm: Random Forest Regressor (100 trees, tuned).

- Performance on test set: RMSE = 238.67, R^2 = 0.863



Most points cluster near the diagonal—model captures variability well, yet underestimates extreme peaks (>2500 rentals). Further feature engineering may address this tail.

5. Conclusions & Future Work (Section 7)

- Demand exhibits pronounced diurnal and seasonal patterns, peaking on summer evenings.
- Weather variables (temperature, solar radiation) and temporal features are strong predictors.
- Random Forest achieves $R^2 \approx 0.86$, indicating solid predictive power with limited tuning.
- Model struggles on extreme outliers—investigate gradient boosting or quantile regression.
- Integrate real-time weather forecasts and special events calendar for deployment.
- Consider feature selection to reduce complexity and improve model interpretability.