Bike Sharing Demand Capstone

End-to-End Analysis, Modeling, and Shiny Dashboard

Omar Essam

2025-10-30

Executive Summary

- Objective: build a robust, reproducible workflow to assess bike-sharing demand across cities, leveraging weather forecasts and historical usage.
- Approach: collect data (OpenWeather + public datasets), clean and integrate; perform EDA (SQL + visual); train and refine linear models; deliver an interactive Shiny dashboard.
- Results: best regression model achieves strong fit with interpretable coefficients; dashboard provides real-time insights and forecast-driven demand estimates.
- Impact: enables operations teams to anticipate demand swings, optimize bike availability, and plan maintenance windows.

Introduction

- Motivation: weather significantly influences bike demand (temperature, humidity, wind, precipitation).
- Data sources: OpenWeather API, global cities metadata, recorded bike usage (Seoul + systems overview).
- Tools: R (tidyverse, modeling, Shiny), SQL for EDA, and reproducible reporting.

Methodology Overview

- Data collection: scripted downloads and API calls; tracked outputs under `project5-capstone/output/`.
- Cleaning & integration: consistent types, missing values handled, feature engineering for model inputs.
- Modeling: baseline linear models $\rightarrow$ refined models; diagnostics and model comparison to select the best.
- Visualization & reporting: ggplot2 for EDA; beamer PDF for submission; Shiny for interactive exploration.

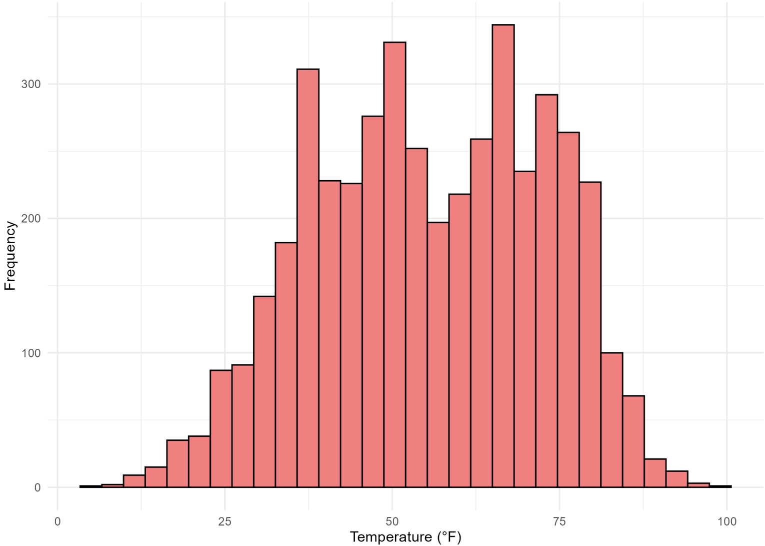
EDA with SQL

- Used SQLite queries to explore distributions, correlations, and station-level patterns.
- Example pattern: identify top cities/days by predicted demand and weather conditions.

```sql
-- Example SQL snippet used during EDA (illustrative)
SELECT city, date, AVG(temperature) AS avg_temp, AVG(humidity)
FROM weather_by_city
GROUP BY city, date
HAVING COUNT(*) > 12
ORDER BY avg_temp DESC
LIMIT 10;
```

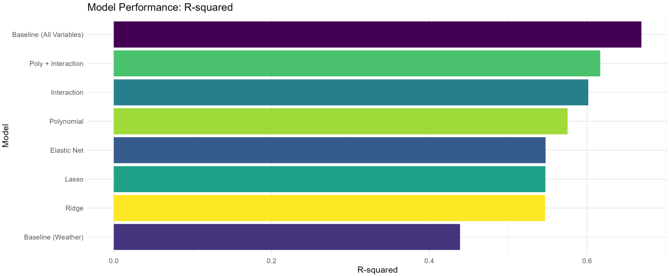
EDA with Visualization

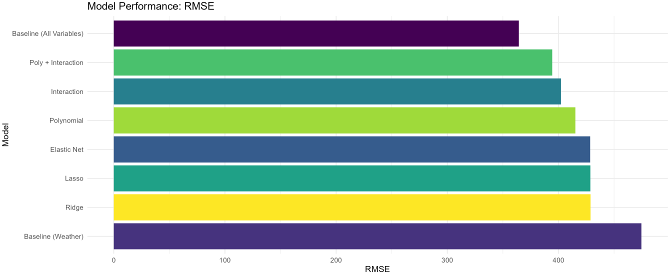
Distribution of Dry Bulb Temperature



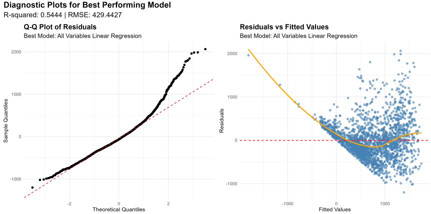
Distribution of Wind Speed

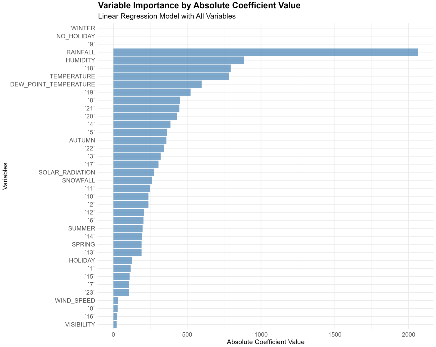
Predictive Analysis: Model Comparison





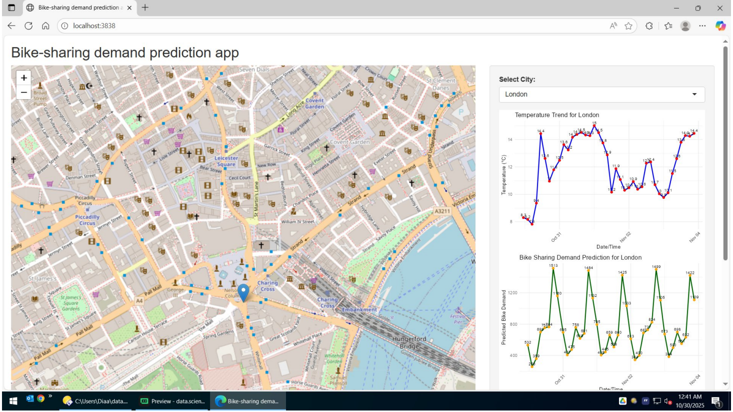
Predictive Analysis: Diagnostics & Coefficients





Shiny Dashboard (Module 5)

- Interactive map with city markers and popups.
- City selector drives trend plots for temperature and demand prediction.
- Clickable prediction trend returns point-level details.



Conclusions

- Weather signals are strong predictors of bike demand; temperature and humidity dominate.
- The refined linear model offers interpretable insights with actionable coefficients.
- The Shiny dashboard operationalizes predictions and supports real-time decision-making.
- Future work: add seasonality, holiday effects, and ML models (e.g., GBMs).

Appendix: R Code Snippet

```
## # Refined Linear Regression Models for Bike Sharing Demand
## # This script builds advanced regression models with polyno
##
## # Load required libraries
## library(tidymodels)
## library(tidyverse)
## library(glmnet)
## library(patchwork)
##
## # Load the bike sharing dataset
## cat("Loading bike sharing dataset...\n")
## bike_data <- read_csv("data/raw_seoul_bike_sharing.csv")
##
## # Remove Date and FUNCTIONING_DAY columns as specified
## bike_data <- bike_data %>%
##   select(-Date, -FUNCTIONING_DAY)
##
```

Appendix: Additional Figures (Optional)