Anticipating Shared Bicycle Demand

Decision Support Systems

14th of June 2025

Nuno Pinto

Information System Management student, Universidade Autónoma de Lisboa

Palácio Dos Condes Do Redondo, R. de Santa Marta 56, 1169-023 Lisboa

***Abstract* — This report details the development of a Decision Support System (DSS) focused on the analysis and forecasting of demand for shared bicycle systems in the city of Seoul. Adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, the project covers a complete data science cycle, from understanding the business problem to implementing an interactive dashboard. Using the R programming language and its "Tidyverse" ecosystem packages, tasks such as collection, cleaning, exploratory analysis, and predictive modeling were performed. The result is an "Elastic Net" regression model capable of predicting the hourly count of rentals and a DSS prototype in "R Shiny" that translates data into actionable insights for urban mobility management. This document explores the fusion of DSS theory, data quality, and modeling techniques with their practical application.**

Keywords — Introduction, CRISP-DM Based Methodology, Problem Understanding and Data Preparation, Data Preparation and Quality Assurance, Modeling, Evaluation, Deployment, Exploratory Data Analysis (EDA) Results, Conclusion.

I. INTRODUCTION

Decision Support Systems (DSS) represent a class of information systems that assist in the decision-making activities of organizations. In the context of urban mobility, the efficient management of resources, such as shared bicycle fleets, critically depends on the ability to anticipate demand. Operational decisions, like rebalancing bicycles between stations or scheduling maintenance, can be optimized through the analysis of historical data and the construction of robust predictive models.

This project addresses this challenge by applying the CRISP-DM methodology, an iterative and structured process for data mining projects. The CRISP-DM (Cross-Industry Standard Process for Data Mining) covers a complete data science cycle, from understanding the business problem to implementing an interactive dashboard. The central goal is to develop a DSS for the city of **Seoul**, which is notable for the extensive use of its shared bicycle system and the availability of public data.

The **R programming language** was chosen as the primary technological tool due to its strong capabilities for data manipulation, visualization, and statistical modeling, as highlighted in the fundamentals of R programming (Decision Support Systems UAL, Lecture 07, 2025). The project culminates in the creation of a predictive model and an interactive dashboard, materializing the concept of a DSS that transforms raw data into actionable intelligence.

II. METHODOLOGY BASED ON CRISP-DM

The project structure strictly followed the phases defined by the CRISP-DM model, ensuring a systematic and goal-focused approach.

A. Problem Understanding and Data Preparation

The first phase, **Problem Understanding**, focused on defining the main objective: to predict the hourly count of rented bicycles to support operational decisions. In the subsequent **Data Understanding** phase, data collection from heterogeneous sources began. The technique of **web scraping** was used with the `rvest` package to extract a list of bike-sharing systems from Wikipedia, and calls were made to the Open Weather API with the `httr` package to obtain meteorological data.

B. Data Preparation and Quality Assurance

The **Data Preparation** phase is often the most time-consuming in a data science project and is fundamental to its success. This stage focused on improving **data quality**, a concept that includes multiple dimensions such as accuracy, completeness, consistency, and timeliness. Columns were renamed to a standardized format and translated to Portuguese. Categorical variables, such as seasons, were recoded and transformed into factors with explicit levels.

Feature engineering was applied to create predictor variables with greater informational power, such as the day of the week and an interaction term between temperature and the hour of the day.

C. Modeling

In the **Modeling** phase, the objective was to select and apply appropriate techniques to build a predictive model. The problem was framed as a supervised learning regression task. The choice fell on an **"Elastic Net" regression model**, implemented through the glmnet package. This technique is a form of regularized linear regression that combines L1 ("Lasso") and L2 ("Ridge") penalties, being effective in managing collinearity and selecting variables. A critical methodological aspect was the data split. Instead of random sampling, a

D. Evaluation

The performance of the glmnet model was measured on the test subset using two metrics:

1) **Root Mean Square Error (RMSE)**
2) **Coefficient of Determination (R^2)**

The joint use of RMSE and R^2 provides a balanced view: absolute precision (RMSE) and explanatory power (R^2). In addition to quantitative evaluation, a qualitative assessment was performed by visualizing the predicted versus actual values.

E. Deployment

The final phase, **Deployment**, consists of integrating the model into a production environment so it can be used by decision-makers. In this project, this phase was materialized by building a prototype interactive dashboard with the **"R Shiny"** package. This web application represents the final delivery of the DSS, transforming the complex analysis and predictive model into an easy-to-use tool.

III. RESULTS OF EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is an investigative process that uses statistical summaries and graphical tools to understand data, discover patterns, detect anomalies, and formulate hypotheses. Visualizations were generated with the `ggplot2` package.

The distribution of the target variable (Fig. 1) shows a **right-skewed distribution**, indicating that most hours have a relatively low count of rentals, but with a long tail of high values representing demand peaks.

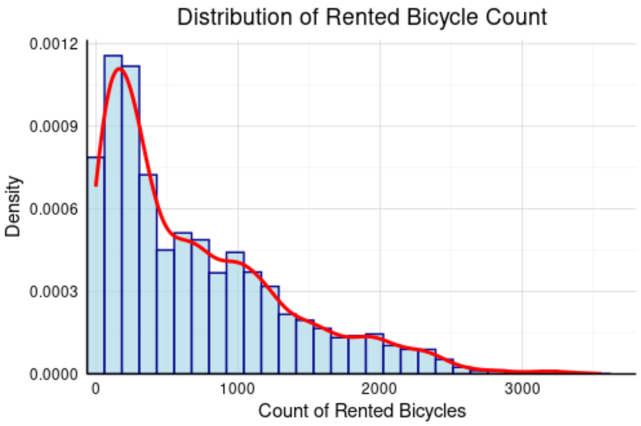

Fig. 1. Distribution of Rented Bicycle Count. The figure demonstrates a right-skewed distribution.

Identifying temporal patterns, such as seasonality (Fig. 2), is crucial for time-series modeling. The data reveals a clear seasonal pattern: demand is minimal in the winter months, grows during the spring, peaks in the summer, and decreases in the autumn.

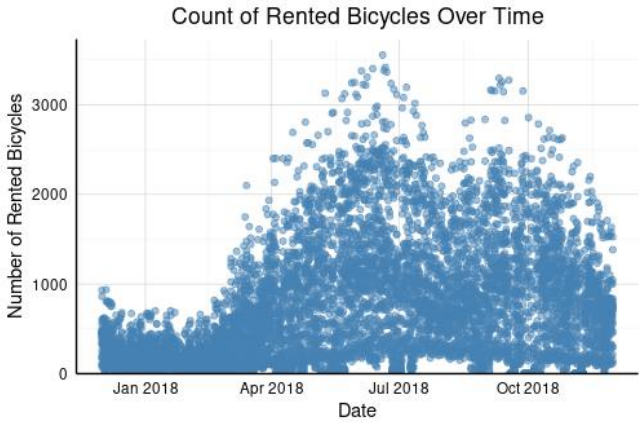

Fig. 2. Count of Rented Bicycles Over Time. This visualization reveals a clear seasonal pattern.

The hourly analysis (Fig. 3) reveals the most important usage pattern for daily fleet management, confirming the importance of including the hour of the day as an essential prediction variable.

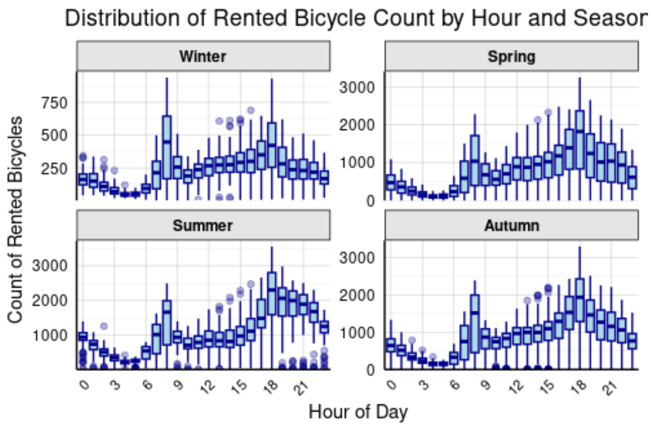

Fig. 3. Distribution of Rented Bicycle Count by Hour and Season.

Exploring relationships between multiple variables (Fig. 4) helps to formulate hypotheses about interactions. There is a general positive correlation between temperature and the number of rentals; in all seasons, higher temperatures tend to be associated with higher demand.

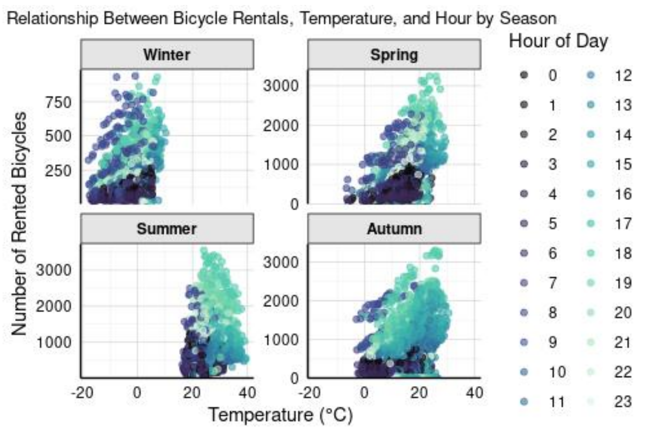

Fig. 4. Relationship between Bicycle Rentals, Temperature, and Hour by Season.

IV. CONCLUSION

This project demonstrates the application of a structured data mining process, CRISP-DM, to develop a complete and functional Decision Support System. With the R language and its libraries, it was possible to execute each phase of the methodology.

The results confirm that the demand for shared bicycles in Seoul is strongly influenced by temporal factors (seasonality, day of the week, hour of the day) and environmental conditions (temperature). The "Elastic Net" regression model was able to capture these relationships effectively, and its implementation in an interactive dashboard with "R Shiny" serves as an effective prototype of how data science insights can be democratized and made available to non-technical decision-makers. This work not only achieves its practical objectives but also serves as a didactic example of the integrated application of data quality, programming, analysis, and modeling concepts covered in the Decision Support Systems course.

REFERENCES

[1] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL LECTURE 07 Fundamentals of R Programming PART 01.pdf*. Universidade Autónoma de Lisboa.

[2] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL LECTURE 08 Fundamentals of R Programming PART 02.pdf*. Universidade Autónoma de Lisboa.

[3] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL Lecture 09 Fundamentals of R Programming PART 03.pdf*. Universidade Autónoma de Lisboa.

[4] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL Lecture 10 WORKING WITH DATA IN R - PART 04.pdf*. Universidade Autónoma de Lisboa.

[5] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL - Lecture 11 - EDA in R.pdf*. Universidade Autónoma de Lisboa.

[6] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL Lecture 12 Model Development in R.pdf*. Universidade Autónoma de Lisboa.

[7] Decision Support Systems UAL. (2025). *DECISION SUPPORT SYSTEMS UAL - Lecture 13 and 14 - CRISP-DM.pdf*. Universidade Autónoma de Lisboa.