CSP 571 Data Preparation and Analysis

Project Plan and Detail

Analyzing Divvy Bike Usage Patterns in Chicago

Date: 06/29/2024

Mohammad Hamza Piracha
mpiracha@hawk.iit.edu
A20554741

Poshan Pandey
Ppandey6@hawk.iit.edu
A20519852

Usman Matheen Hameed
uhameed@hawk.iit.edu
A20564338

Objective

The objective of this project is to analyze Divvy bike-sharing data in Chicago to identify usage patterns and optimize the distribution of bike docks and cycles. By examining the busiest and least busy areas and times, we aim to provide data-driven recommendations to improve the efficiency of the Divvy bike-sharing system. The analysis will also compare the usage patterns of electric bicycles, classic bicycles, and scooters, as well as differences between regular guest users and Chicago resident users.

The methodology includes comprehensive data collection, preparation, cleaning, and transformation. We will address outliers, perform distribution analysis, and employ clustering techniques to identify high and low activity areas. Dimensionality reduction techniques such as PCA will be used for visualization, and relevant features will be selected for predictive modeling. Regression models will be trained to forecast dock and cycle needs based on historical usage patterns.

Visualizations will be created to highlight key insights and findings, which will be documented in a detailed report. The report will interpret the results and provide actionable recommendations for optimizing the Divvy bike-sharing system. Attached are the teamwork schedules and tasks assigned in the chart.

Tasks:

Data Preparation

1. Data Collection: Gather datasets from the Divvy bike data repository.
2. Data Preparation: Combine individual datasets into a comprehensive dataset.

Data Cleaning and Transformation

1. Data Cleaning: Remove duplicate entries and handle missing values.
2. Data Transformation: Convert date and time fields to appropriate formats.
3. Outlier Detection and Handling: Identify and address outliers to ensure data quality.

Data Analysis and Modeling

1. Distribution Analysis: Analyze ride counts, durations, and distances.
2. Clustering: Identify clusters of high and low activity areas.
3. Dimensionality Reduction: Use PCA to reduce dataset dimensionality for visualization purposes.
4. Feature Selection: Identify the most relevant features for analysis.
5. Model Selection and Training: Use regression models to predict dock and cycle needs based on historical usage patterns.

Visualization and Reporting

1. Data Visualization: Create visualizations to highlight insights and findings from the data.
2. Output Visualization: Visualize analysis results and recommendations.

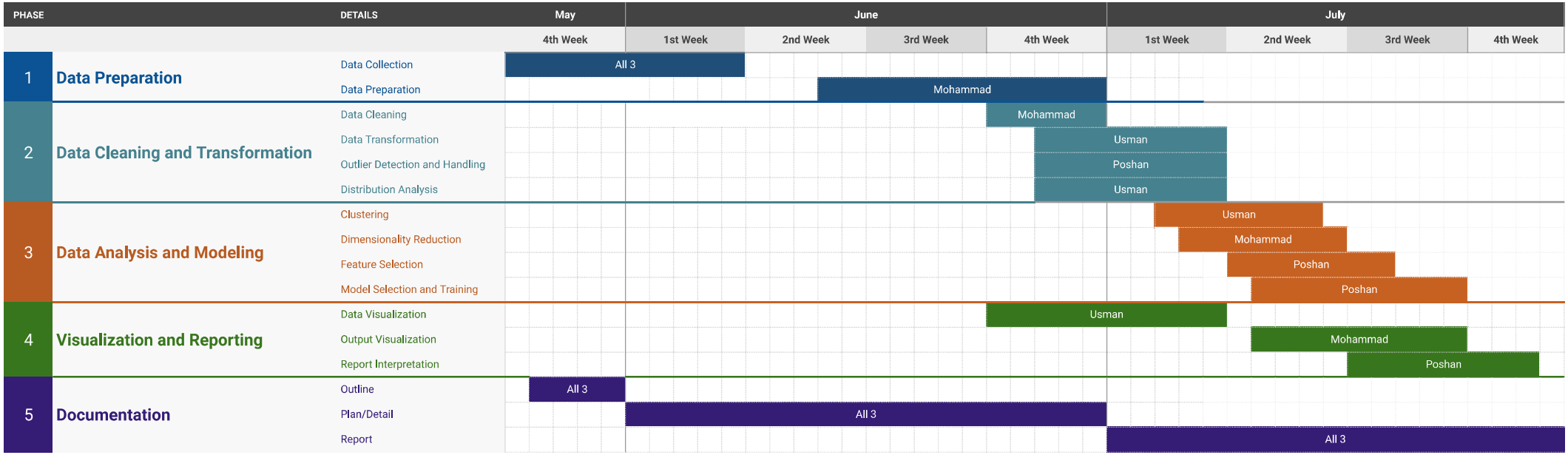
Documentation

1. Report Interpretation: Interpret findings and provide actionable recommendations.
2. Outline: Develop an outline of the report.
3. Plan/Detail: Detail the plan for the report.
4. Report: Write the final report documenting the methodology, analysis, and findings.

TITLE **Analyzing Divvy Bike Usage Patterns in Chicago**

Teammates
Mohammad Hamza Piracha (A20554741)
Poshan Pandey (A20519852)
Usman Matheen Hameed (A20564338)

Date: 06/30/2024



Note: All tasks are performed and completed through combined discussion and collaboration among the team members. The segregation here refers only to the primary person responsible for each task.