PROJECT : COMMUNICATE WITH YOUR DATA

Mady by :Yousef Mohamed Abd Elmonem

INTRODUCTION

Dataset overview

- **The dataset comes from the Ford GoBike System (February 2019).**
- **It contains bike-sharing trip data, including:**
  - **Trip Duration (in seconds)**
  - **Start & End Stations**
  - **User Type (Subscriber or Customer)**
  - **Rider Age & Gender**
  - **Bike ID & Rental Details**

PURPOSE OF ANALYSIS

1.

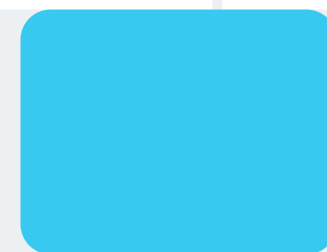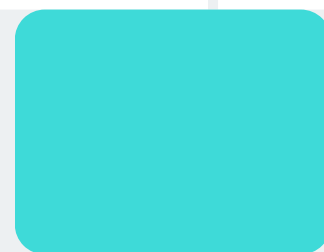- Understand user behavior patterns in bike-sharing.

2.

- Explore ride duration variations across age groups.

3.

- Identify correlations between user demographics and ride characteristics.

4.

- Present data-driven insights to optimize the bike-sharing experience.

KEY QUESTIONS & OBJECTIVES (WHAT YOU AIMED TO FIND)

- **Key Questions**
  - **How does ride duration distribution change across different age brackets?**
  - **What is the most rented month for bike usage?**
  - **How do ride durations vary by user type (Subscriber vs. Customer)?**
  - **What are the most popular hours for bike rentals?**
  - **How is gender distribution represented among riders?**

- **Purpose of Analysis**
  - **Understanding user behavior and ride patterns**
  - **Identifying key trends in bike rentals**

DATA CLEANING & PREPROCESSING

- **Steps taken:**
  - **Handling :**
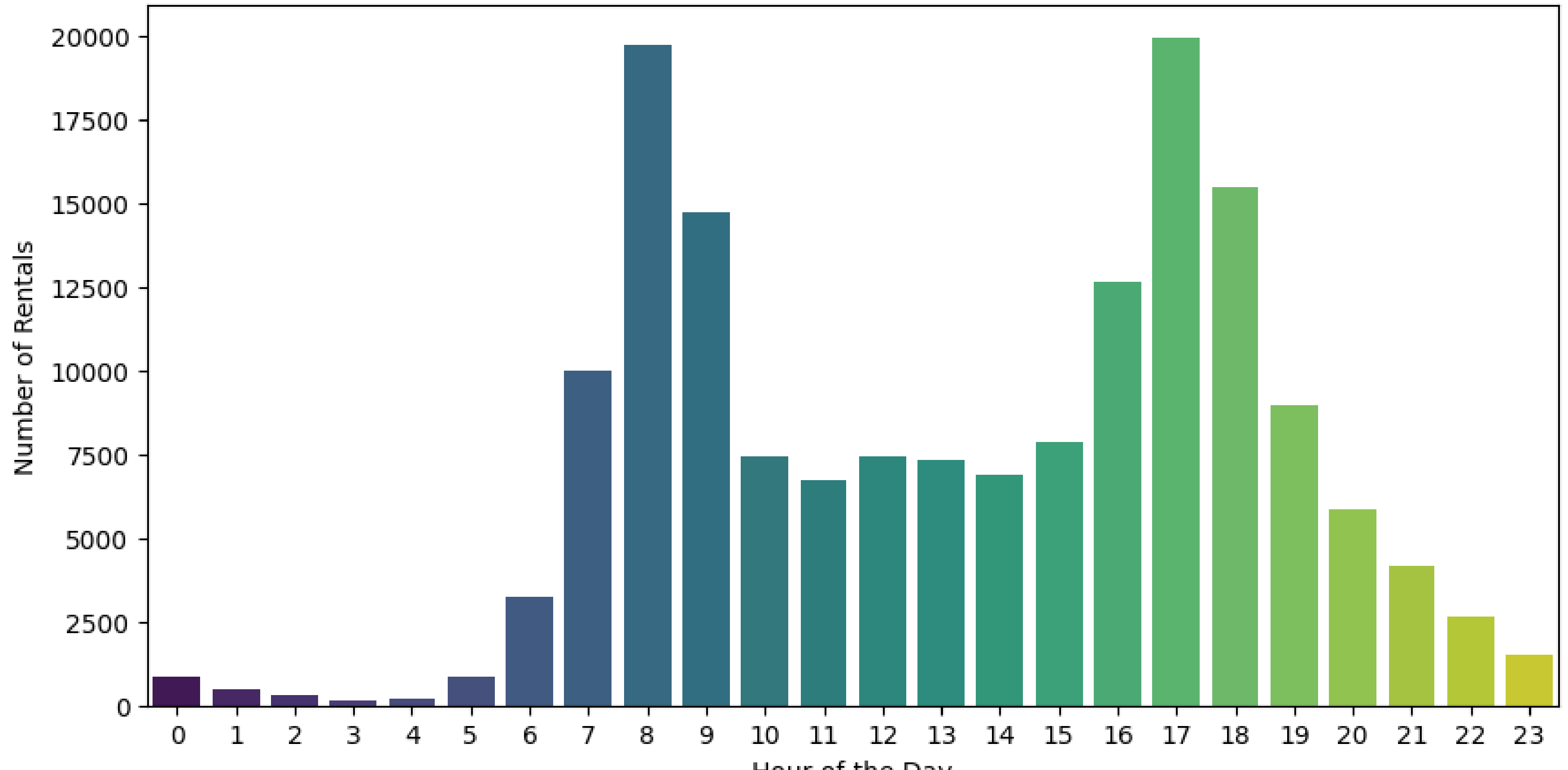   - **Missing values**
    - **Outliers,**
- **Data formatting**

EXPLORATORY DATA ANALYSIS (EDA) OVERVIEW

UNIVARIATE, BIVARIATE & MULTIVARIATE ANALYSES

WHAT IS THE MOST POPULAR HOUR IN BIKE RENTAL?
ANSWER : 5 PM

Most Popular Hours for Bike Rentals

INCREASE THE NUMBER OF BIKES AT 5 PM

NUMBER OF TRIPS BETWEEN STATIONS?

Heatmap of Trips Between Top 20 Stations

ANSWER : THE HIGHEST TRIP COUNTS ARE IN RED AND LIGHT BLUE, INDICATING THE BUSIEST STATION-TO-STATION ROUTES.
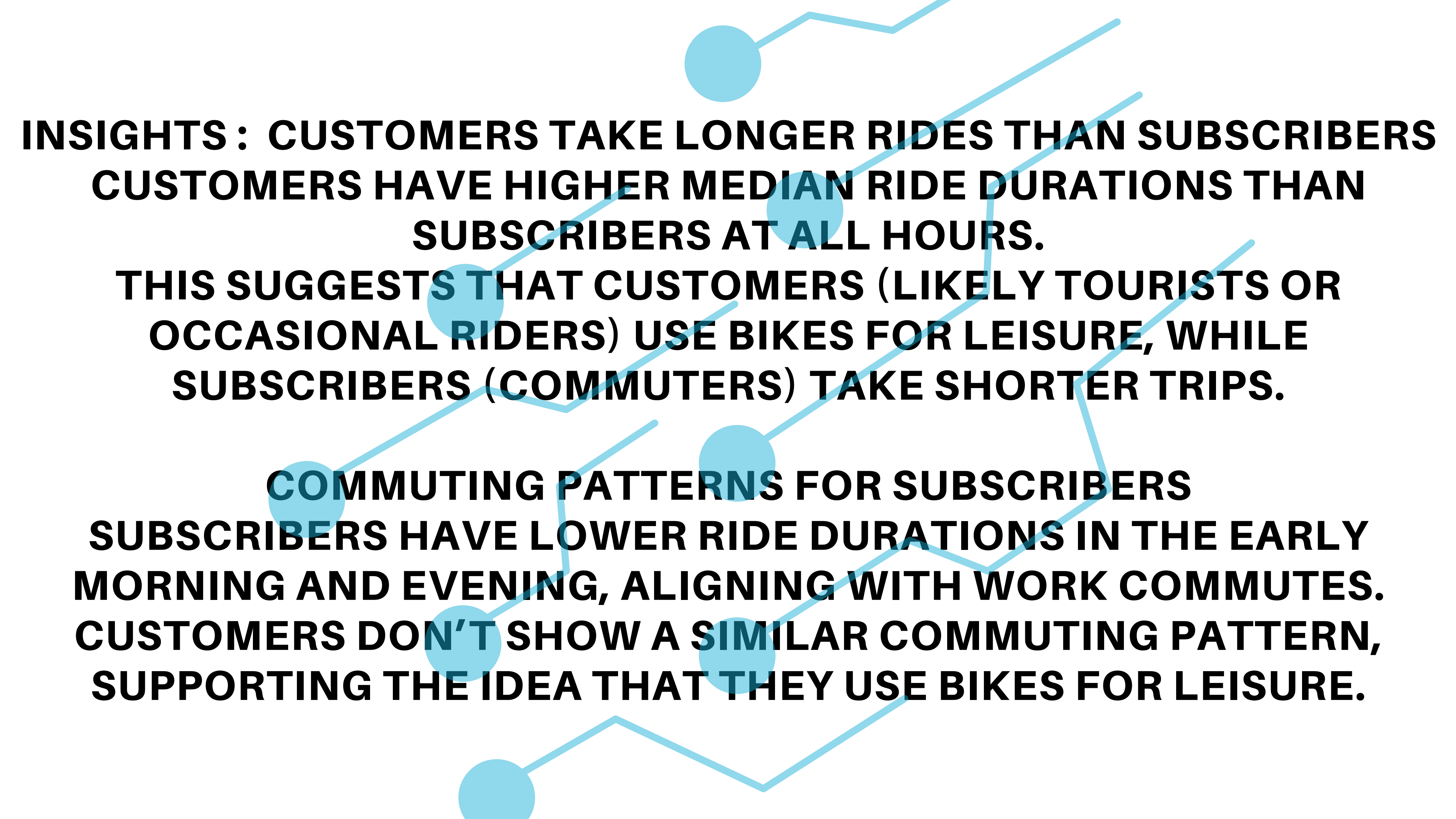
EXAMPLE: THE TRIPS BETWEEN STATIONS 6 & 15, 50 & 58, AND 81 & 58 APPEAR FREQUENTLY (TRIP COUNTS OVER 200+).
THIS SUGGESTS THAT THESE STATIONS ARE MAJOR HUBS WITH A LOT OF TRIPS BETWEEN THEM.

ACTION: INCREASE NUMBER OF BOKES AND THE AREA OF THE BIKE STATION

Ride Duration by Hour and User Type
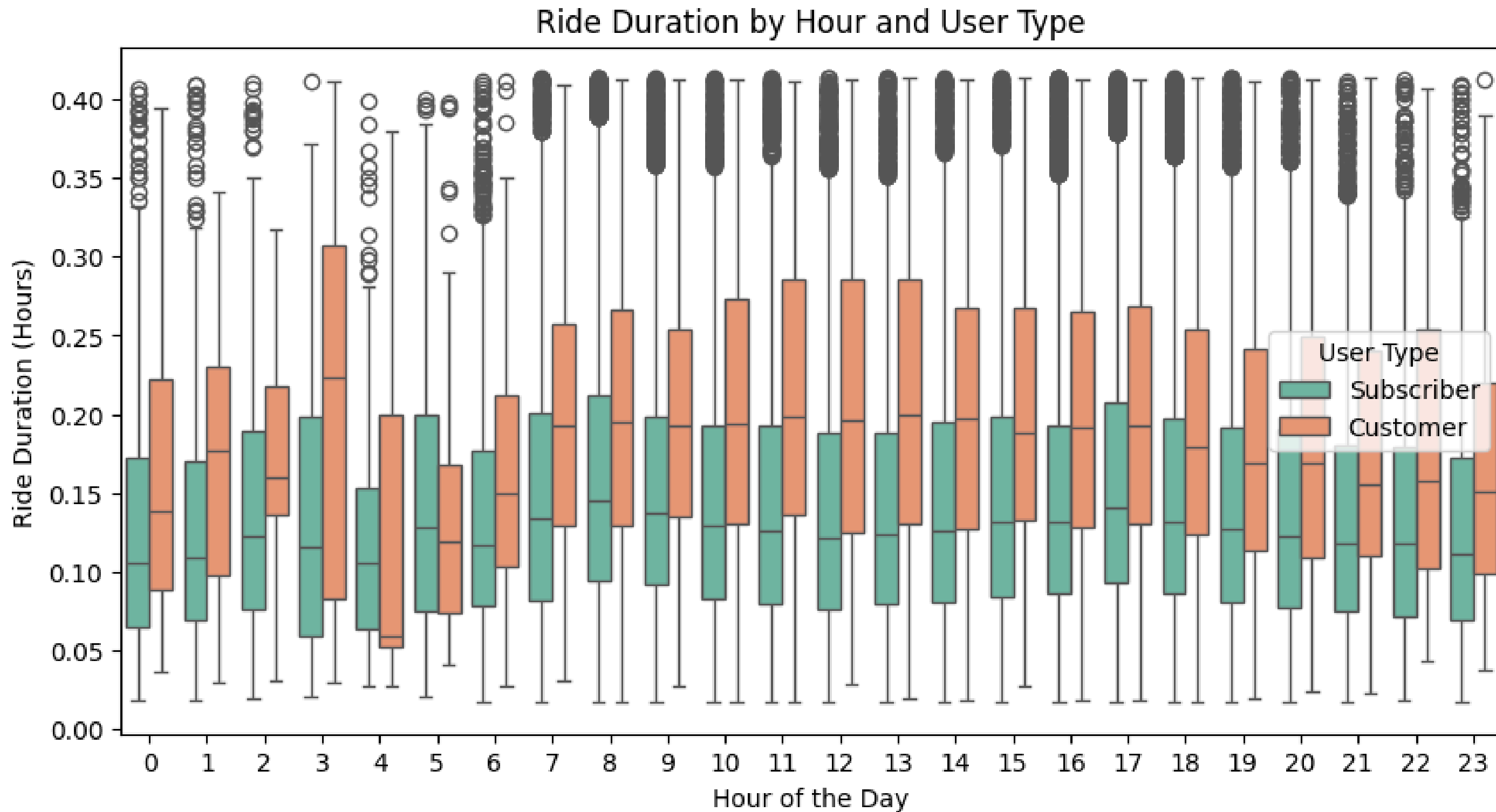
Multivariate Analysis

How the type of the user affects the ride duration?

INSIGHTS : CUSTOMERS TAKE LONGER RIDES THAN SUBSCRIBERS

CUSTOMERS HAVE HIGHER MEDIAN RIDE DURATIONS THAN SUBSCRIBERS AT ALL HOURS.

THIS SUGGESTS THAT CUSTOMERS (LIKELY TOURISTS OR OCCASIONAL RIDERS) USE BIKES FOR LEISURE, WHILE SUBSCRIBERS (COMMUTERS) TAKE SHORTER TRIPS.

COMMUTING PATTERNS FOR SUBSCRIBERS

SUBSCRIBERS HAVE LOWER RIDE DURATIONS IN THE EARLY MORNING AND EVENING, ALIGNING WITH WORK COMMUTES. CUSTOMERS DON'T SHOW A SIMILAR COMMUTING PATTERN, SUPPORTING THE IDEA THAT THEY USE BIKES FOR LEISURE.

Key Findings & Insights

- Peak usage times and age group trends
  - "Peak hour is 5 PM with 17500+ rentals").
- Subscriber vs. Customer ride behavior
  - That is clear in the day parts
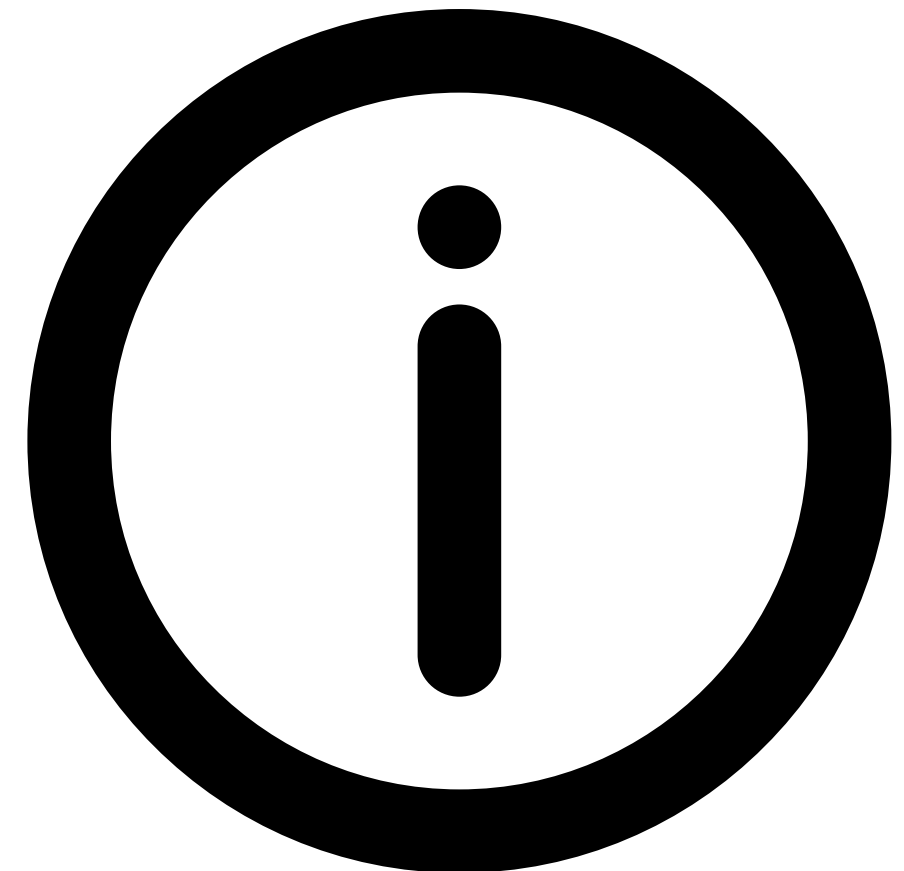- Gender-based differences in ride patterns
  - Males ride bikes by 75%!

Challenges & Limitations

- Outliers in ride duration
- Data gaps & missing values

CONCLUSION & RECOMMENDATIONS

- **Key takeaways from the data analysis**
  - **Suggestions for improving service**
    - **(adjusting bike availability in high-demand areas,**
    - **promoting usage during off-peak hours)**

Thank You! 🎉