

# SMART INTERNZ PROJECT REPORT

Submitted by

1.Arun C

(20BCE0294)

arun.c2020@vitstudent.ac.in

2.Syed Sha Suheb

(20BDS0394)

Syedsha.suheb2020@vitstudent.ac.in

3.Keertheeswar

(20BDS0304)

Keertheeswar.2020@vitstudent.ac.in

4.Samay Sharma

(20BCG10002)

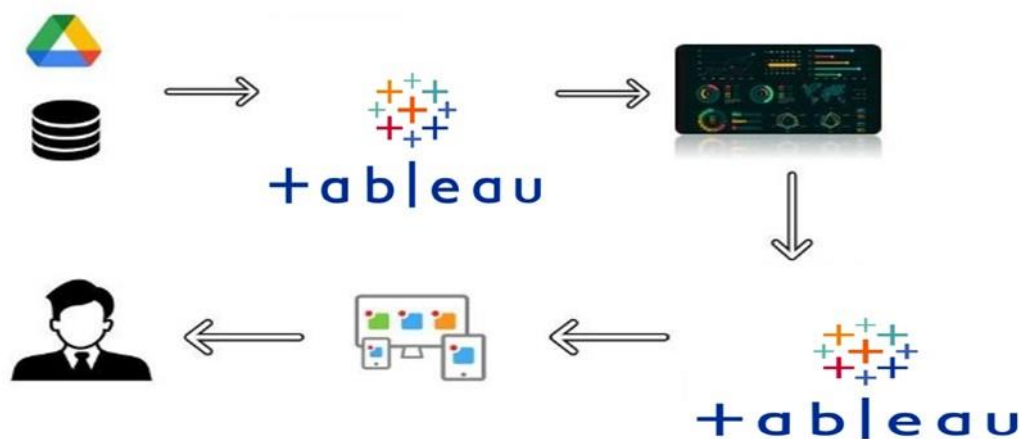
Samay.sharma2020@vitbhopal.ac.in

# *Ford Gobike System Data Analysis using Tableau*

The purpose of this report is to provide data-driven insights on the Ford GoBike system using Tableau, a powerful data visualization tool. The Ford GoBike system is a bike-sharing program that operates in various cities, providing users with a convenient and eco-friendly mode of transportation. By analyzing the data collected from the system, we can gain valuable insights into user behavior, popular routes, and other key metrics.

Technical Architecture:

## **Purpose**



The purpose of this project report is to present the findings and insights derived from a data analysis of the Ford GoBike system using Tableau. The report aims to provide a comprehensive overview of the project's objectives, methodologies, and outcomes, highlighting the value of data-driven decision-making in optimizing the bike-sharing program.

## **Problem Understanding**

The analysis of the Ford GoBike system using Tableau seeks to address several key problem areas:

- Usage Patterns
- User Demographics
- System Efficiency
- Seasonal Variations

To effectively address the business problem, we need to deeper into the following aspects:

a. User Behavior Analysis: Analyzing user behavior patterns, such as trip frequency, duration, and preferred routes, can provide insights into user preferences and interests. Understanding these patterns allows for the development of personalized recommendations and incentives to increase user engagement.

b. Pricing and Membership Optimization: Analyzing user subscriptions, pricing structures, and membership benefits can help identify opportunities to optimize pricing plans. By offering competitive pricing options and tailored membership benefits, the system can attract and retain a larger user base.

c. Targeted Marketing Strategies: Exploring user demographics, such as age, gender, and location, enables the development of targeted marketing campaigns. By understanding the preferences and characteristics of different user segments, personalized marketing strategies can be implemented to increase user engagement and attract new users.

d. Operational Efficiency: Assessing the availability and distribution of bikes across different stations can help optimize operational efficiency. By ensuring an adequate supply of bikes in high-demand areas and improving maintenance schedules, the system can provide a seamless user experience, ultimately leading to higher user retention.

## **Business Requirements**

Business Requirements for Data-driven Insights on Ford GoBike System using Tableau:

Data Collection and Integration: Obtain comprehensive and up-to-date data from the Ford GoBike system, including trip details, user information, and system performance metrics.

Ensure seamless integration of data from multiple sources to create a unified dataset for analysis in Tableau.

Data Exploration and Visualization: Utilize Tableau to explore and visualize the Ford GoBike data, enabling stakeholders to gain a holistic understanding of the system's performance and user behavior.

Develop intuitive and interactive dashboards that allow users to easily explore and analyze the data, ensuring a user-friendly experience.

User Behavior Analysis: Identify and analyze user behavior patterns, including trip duration, frequency, and preferred routes.

Determine the factors influencing user engagement, retention, and satisfaction within the Ford GoBike system.

Demographic Analysis: Investigate user demographics, such as age, gender, and location, to understand the system's appeal to different demographic segments.

Analyze user preferences and behavior based on demographic factors, enabling targeted marketing strategies and personalized recommendations.

Performance Optimization: Identify areas for operational improvements, such as bike station placement, bike distribution, and maintenance schedules.

Optimize the availability and accessibility of bikes to meet user demand and enhance the overall system efficiency.

Seasonal and Geographic Analysis: Analyze usage patterns and demand fluctuations across different seasons and geographic regions.

Identify seasonal trends and regional variations in bike usage to optimize resource allocation and operational planning.

Insights Communication: Present insights and findings in a clear, concise, and visually appealing manner through Tableau visualizations and reports.

Enable stakeholders to easily understand and interpret the data-driven insights, facilitating informed decision-making processes.

Continuous Improvement: Establish a feedback loop to gather user feedback, ratings, and complaints, and incorporate them into ongoing analysis and improvement efforts.

Regularly update and refine the data analysis process to adapt to changing business needs and emerging trends in the bike-sharing industry.

## **Literature Review**

The literature review section of this report aims to provide a comprehensive overview of existing studies, research papers, and relevant literature on the topics related to data-driven insights and the use of Tableau in analyzing bike-sharing systems. This review serves to establish a foundation of knowledge and highlight the current understanding of data analysis in the context of bike-sharing systems.

### Bike-Sharing Systems:

Literature on the evolution and impact of bike-sharing systems in urban areas.

Studies on the benefits of bike-sharing systems in terms of reducing traffic congestion, improving air quality, and promoting sustainable transportation options.

### Data Analysis and Visualization:

Research papers and articles on the role of data analysis and visualization techniques in extracting meaningful insights from large datasets.

Studies highlighting the advantages of using Tableau as a powerful tool for data visualization and exploratory data analysis.

### User Behavior Analysis:

Literature on user behavior analysis in bike-sharing systems, including factors influencing usage patterns, trip duration, and route preferences.

Research on the impact of user characteristics, such as age, gender, and socioeconomic status, on bike-sharing system utilization.

### Factors Affecting User Engagement and Retention:

Studies investigating the factors affecting user engagement and retention in bike-sharing systems, including pricing structures, convenience of station locations, and service quality.

Research on user satisfaction, loyalty, and the role of user feedback in improving bike-sharing systems.

### Operational Efficiency and Optimization:

Literature on optimizing operational aspects of bike-sharing systems, including bike station placement, redistribution strategies, and maintenance schedules.

Studies on demand forecasting, resource allocation, and system performance improvement.

### Case Studies and Best Practices:

Case studies and real-world examples of successful data-driven analyses in bike-sharing systems.

Research on best practices and recommendations for improving the effectiveness of bike-sharing programs through data-driven insights.

### Challenges and Future Directions:

Literature addressing the challenges and limitations of data-driven analysis in bike-sharing systems.

Research on emerging trends, technologies, and future directions for data analysis and visualization in the context of bike-sharing programs.

By reviewing the relevant literature, this report establishes the theoretical foundation and contextual understanding necessary for conducting a data-driven analysis of the Ford GoBike system using Tableau. It also provides insights into the current state of knowledge, identifies gaps in the literature, and informs the methodology and approach used in the project.

## **Social & Bussiness Impact**

Socially, the analysis can lead to improved transportation planning, reduced traffic congestion, and increased adoption of eco-friendly modes of travel. It can also enhance user experiences, satisfaction, and safety. From a business perspective, the insights can optimize operational efficiency, enhance marketing strategies, and increase user retention and engagement. This can result in improved profitability, growth, and competitiveness for the Ford GoBike system. Ultimately, the data-driven insights generated through Tableau can transform the bike-sharing program, benefiting both the community and the business.

## **Data Collection & Extraction From Database**

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

### **Collect The Dataset**

Check out the below link to understand the dataset in detail:

<https://www.kaggle.com/code/nancyalaswad90/data-visualization-trip-data-ford-go-bike>

### **Activity 1.1: Understand the data**

Explanation video link:

[https://drive.google.com/file/d/1EMOS6l-e1cYVRmoyBXwzadvH3\\_I6CSG9/view?usp=sharing](https://drive.google.com/file/d/1EMOS6l-e1cYVRmoyBXwzadvH3_I6CSG9/view?usp=sharing)

### **Activity 2: Connect MySQL and Tableau with the dataset**

Explanation video link:

<https://drive.google.com/file/d/1a24aFTycmHTirozwNEI0Z29jjhJJpcC8/view?usp=sharing>

## **Data Preparation**

Data preparation for Tableau involves the process of organizing, cleaning, and transforming raw data into a format that can be effectively visualized and analyzed within the Tableau software. This includes tasks such as data cleaning, data integration, data formatting, and data aggregation. The goal is to ensure that the data is accurate, consistent, and structured in a way that enables meaningful insights and visualizations in Tableau.

### **Prepare The Data for Visualization**

Explanation video link:

<https://drive.google.com/file/d/1f5Wd8CZsqq5pK5HciFkmWGQAEBWJpwxx/view?usp=sharing>



## **Data Visualization**

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex datasets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

### **No Of Unique Visualizations (Filters Applied)**

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the data include bar charts, line charts, heat maps, scatter plots, pie charts, maps, etc. These visualizations can be used to compare performance, track changes over time, show distribution, and show relationships between variables.

**Activity 1.1:** Statistical information about the dataset

Explanation video link :

<https://drive.google.com/file/d/1EavmeihRDAj9GQs3YrqAcAK6IMqQAQs0/view?usp=sharing>

**Activity 1.2:** User distribution information.

Explanation video link :

[https://drive.google.com/file/d/1\\_msYkYdgdC5cTs7wDRGhezXUFD8P0BBJ/view?usp=sharing](https://drive.google.com/file/d/1_msYkYdgdC5cTs7wDRGhezXUFD8P0BBJ/view?usp=sharing)

**Activity 1.3:** Find the total number of user type.

Explanation video link :

<https://drive.google.com/file/d/1LcwU15PmR0PGzy-gHTzHXxUSIC0qmdWq/view?usp=sharing>

**Activity 1.4:** Find which day in the week has the highest number of trips.

Explanation video link :

<https://drive.google.com/file/d/1XTzOIwWLReXYY2B0RkjQhxZj7MalQTKl/view?usp=sharing>

**Activity 1.5:** Find which hour in the day has the highest number of trips.

Explanation video link :

<https://drive.google.com/file/d/1Yx17XVof1fK4MsJdA-pgQg8RFy9KTZUm/view?usp=sharing>

**Activity 1.6:** Find the distribution between age range and bike share interest.

Explanation video link :

<https://drive.google.com/file/d/1kS06LhVR4zq2s5EUa9I0dah0hPfCHK3l/view?usp=sharing>

**Activity 1.7:** Find the distribution of the age of the user.

Explanation video link :

<https://drive.google.com/file/d/1dqZl-nNtDE2QwiQqyN4tehsiDGP3-yQj/view?usp=sharing>

**Activity 1.8:** Find the destination location with travel more than 8 hours.

Explanation video link :

<https://drive.google.com/file/d/1XyzRCgO4X41ELbofE8PX3kmfLW-a7t5c/view?usp=sharing>

**Activity 1.9:** Find the top 10 end station with age range.

Explanation video link :

[https://drive.google.com/file/d/1G39jESdJohoaGv4xco1xz9a\\_pZvbPVb0/view?usp=sharing](https://drive.google.com/file/d/1G39jESdJohoaGv4xco1xz9a_pZvbPVb0/view?usp=sharing)

**Activity 1.10:** Find the top 10 date with estimate prices.

Explanation video link :

<https://drive.google.com/file/d/1KMwKCx9wRPNZiZfXmQqAiBctmh36kn3G/view?usp=sharing>

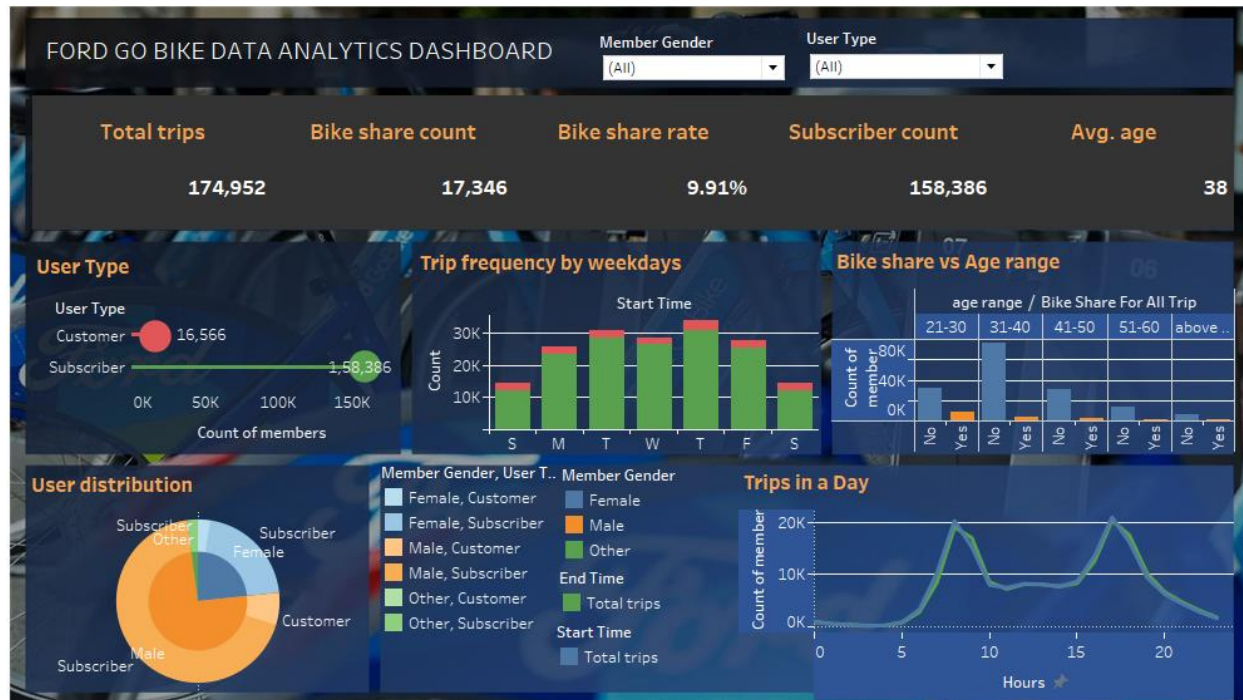
## **Dashboard**

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case.

Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

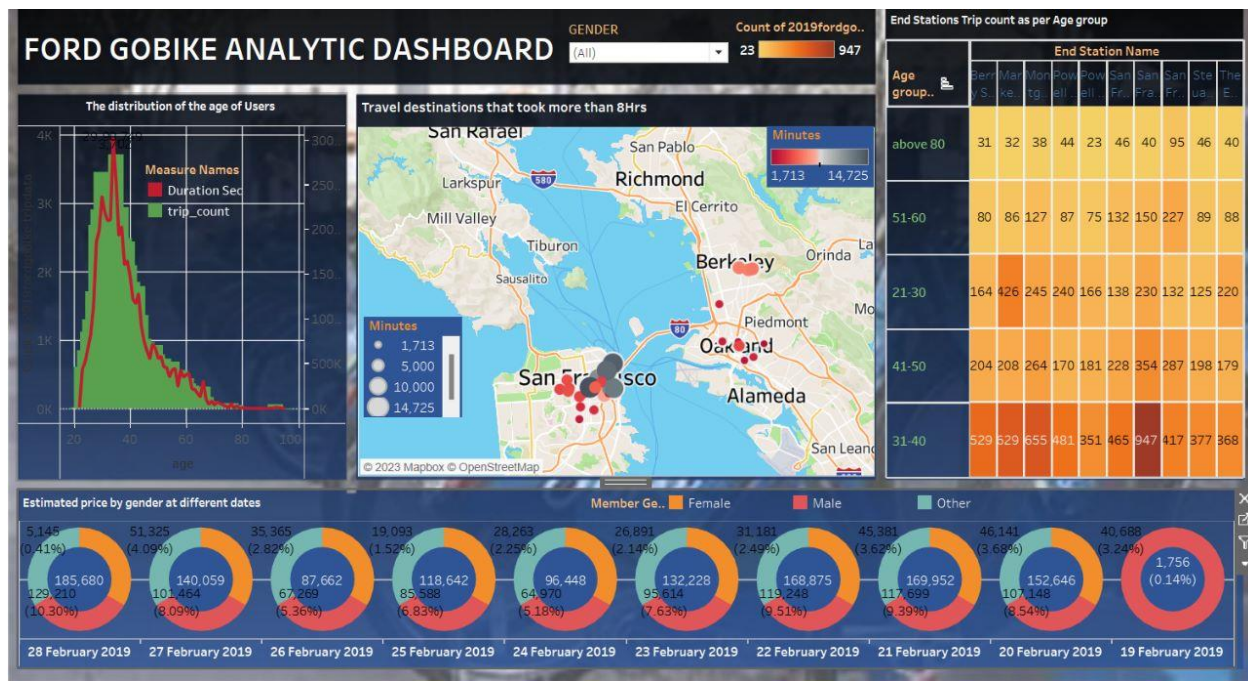
## Responsiveness And Design Of Dashboard

The responsiveness and design of a dashboard for analyzing the factors important for A comprehensive analysis of the IT sector's salaries and roles analyzes various engagement metrics such as Company name, Job Title, Salary, Salaries reported, Location, Employment Status, Job roles, and Rating.



**Explanation video link:**

[https://drive.google.com/file/d/1fMrgX-iW0FXGxm5LXUD9c12gF0qnrYD/view?usp=drive\\_link](https://drive.google.com/file/d/1fMrgX-iW0FXGxm5LXUD9c12gF0qnrYD/view?usp=drive_link)



**Explanation video link:**

[https://drive.google.com/file/d/1SKs66NyBb3qID\\_-XsrzVA9sutfAKyFB4/view?usp=sharing](https://drive.google.com/file/d/1SKs66NyBb3qID_-XsrzVA9sutfAKyFB4/view?usp=sharing)

## Story

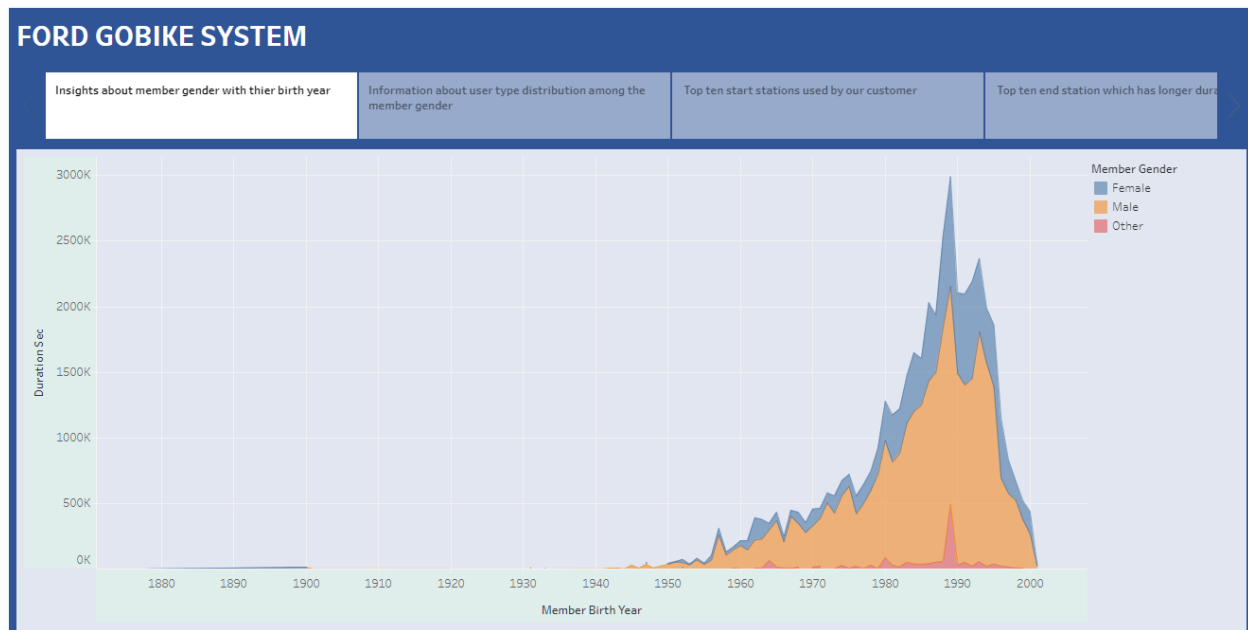
A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

## No Of Scenes Of Story

The number of scenes in a storyboard for a data-driven analysis of the Ford GoBike system using Tableau will depend on the complexity of the analysis and the specific insights that are being conveyed. A storyboard serves as a visual representation of the data analysis process, breaking it down into a series of steps or scenes. These scenes highlight key findings and data visualizations that help uncover insights into the performance and user behavior within the Ford GoBike system..

## Explanation video link:

<https://drive.google.com/file/d/1iHgdYHl5Okw5IHdVrqYEEYgvo57B-EwPT/view?usp=sharing>



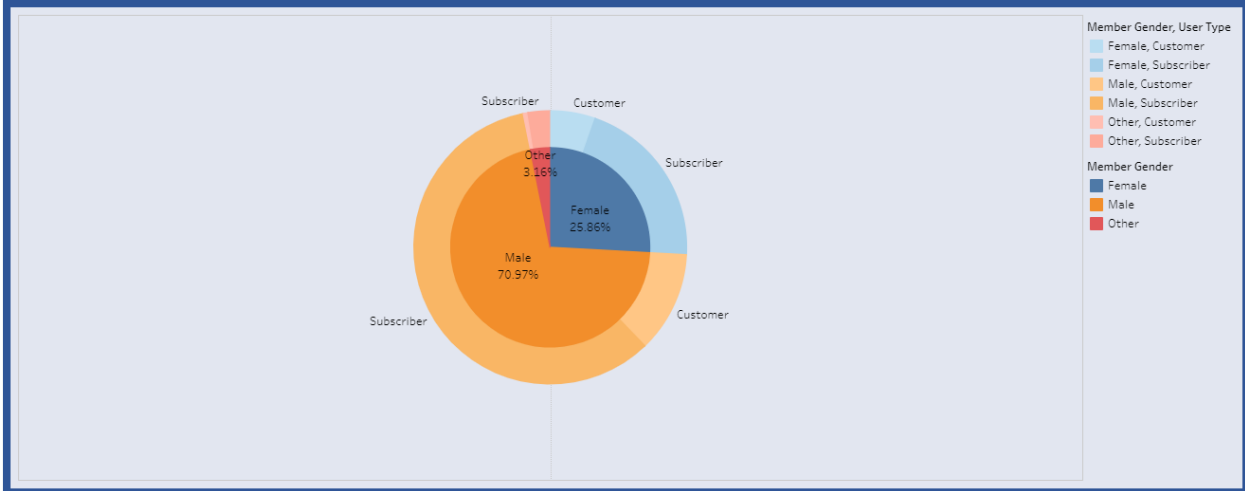
## FORD GOBIKE SYSTEM

Insights about member gender with thier birth year

Information about user type distribution among the member gender

Top ten start stations used by our customer

Top ten end station which has longer dura



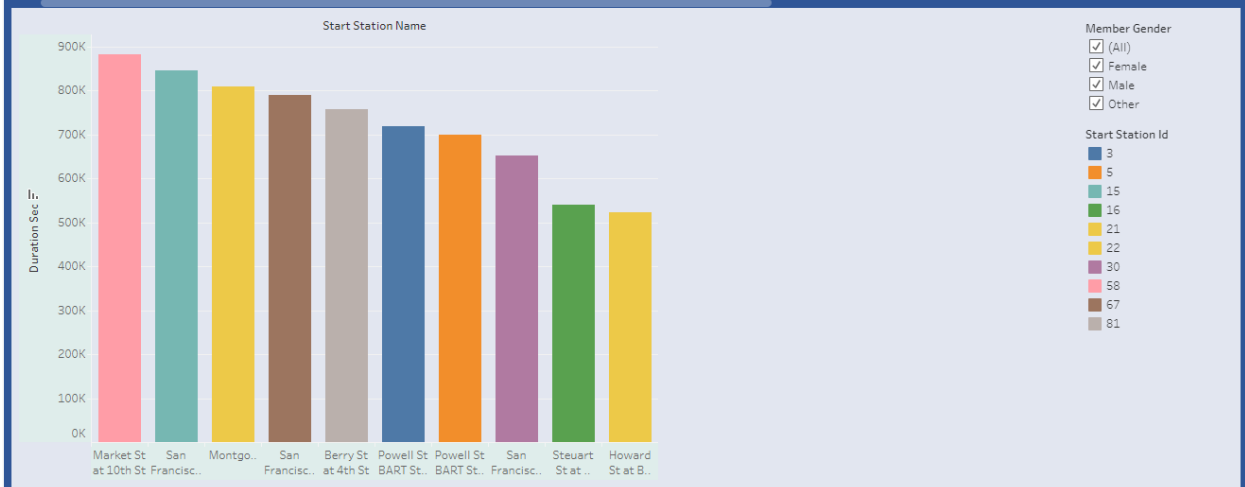
## FORD GOBIKE SYSTEM

Insights about member gender with thier birth year

Information about user type distribution among the member gender

Top ten start stations used by our customer

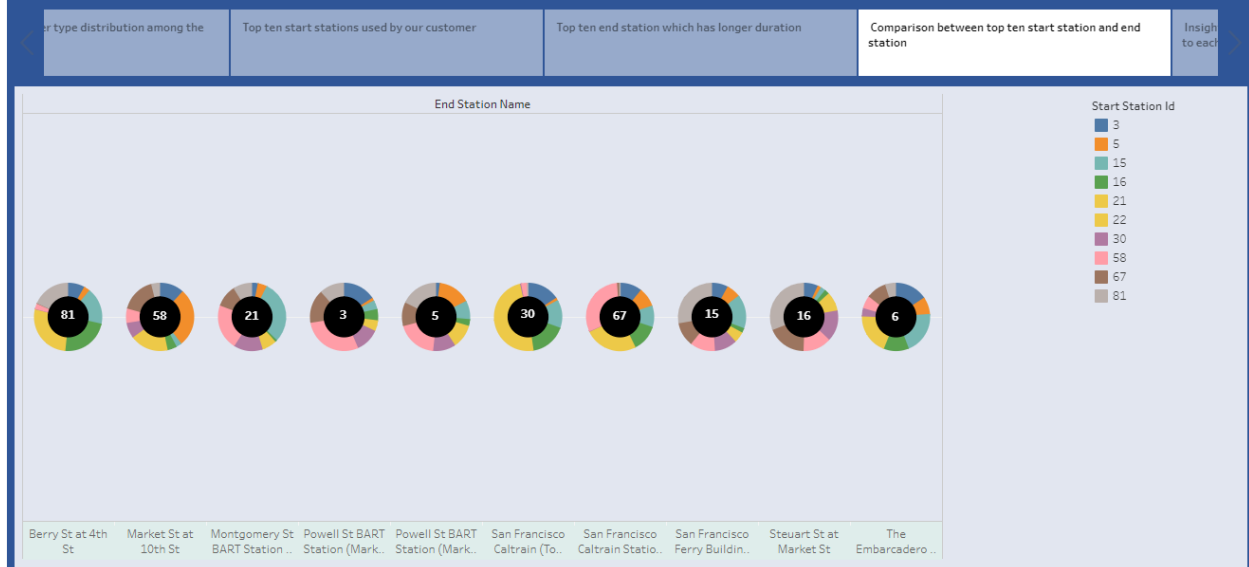
Top ten end station which has longer dura



## FORD GOBIKE SYSTEM



## FORD GOBIKE SYSTEM





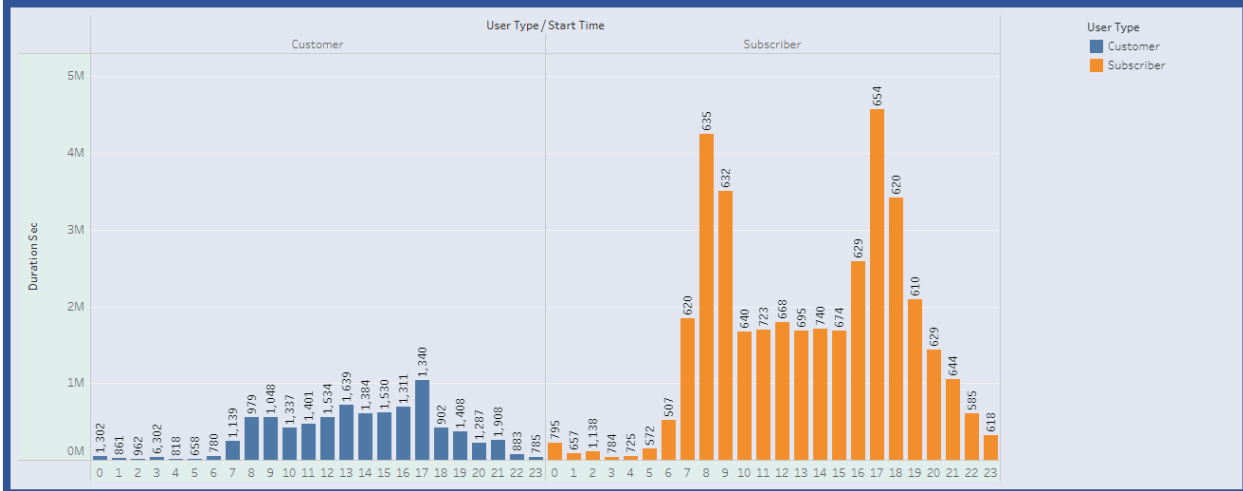
## FORD GOBIKE SYSTEM

stations used by our customer

Top ten end station which has longer duration

Comparison between top ten start station and end station

Insights about trip duration of our user with respect to each hour



## Performance Testing

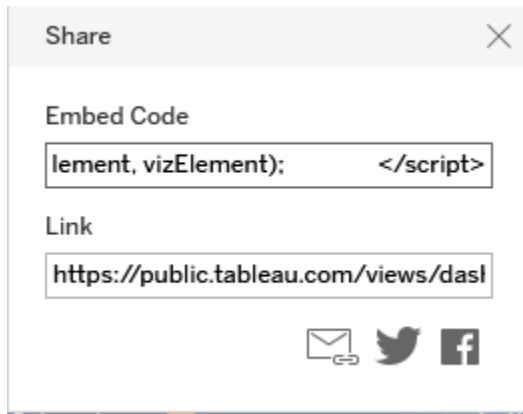
Performance testing for Tableau focuses on evaluating the software's speed, responsiveness, and scalability under various conditions and workloads. It involves measuring and analyzing key performance indicators such as query response time, data loading speed, dashboard rendering time, and concurrent user handling capacity. The testing process helps identify any performance bottlenecks, optimize system configurations, and ensure that Tableau can handle the expected workload efficiently, providing users with a smooth and responsive experience while working with large datasets and complex visualizations.

## Web Integration

Publishing helps us track and monitor key performance metrics, to communicate results and progress. Help a publisher stay informed, make better decisions, and communicate their performance to others.

Integrating dashboard/reports/stories to web

Step 1: Go to Dashboard/story/report, click on share button on the top right.



### **Activity 1: Integrating with Tableau Public**

**Explanatory video:**

<https://drive.google.com/file/d/1TuumjrEKEwqqFk00GX4xlIRULAnI76Os/view?usp=sharing>

### **Activity 2: Integrating with bootstrap website**

**Explanatory video:**

[https://drive.google.com/file/d/1DOMzo0\\_p9AWmvE7OR4ozwPARV6PvzOYI/view?usp=sharing](https://drive.google.com/file/d/1DOMzo0_p9AWmvE7OR4ozwPARV6PvzOYI/view?usp=sharing)

### **Activity 3: Implementing Flask**

**Explanatory video:**

[https://drive.google.com/file/d/1cKIhJ1F0\\_Vb3zKn5It9nS\\_05c1hIv0mC/view?usp=sharing](https://drive.google.com/file/d/1cKIhJ1F0_Vb3zKn5It9nS_05c1hIv0mC/view?usp=sharing)