A close up of a logo

Description automatically generated

**Week 2 Assignment - Final Project.**

*Bike Sharing Data from Capital Bikeshare System in 2011 and 2012*

**Huskies Viz Team**

Siva Ramakrishna Sai Chand Chivukula

Thi Thai Nigan Le

Yang Wu

Bodian Li

80926 – Communication & Data Visualization - Spring 2019

Instructor: Jack Bergersen (Prof.)

**Introduction**

Bike sharing is one of the emerging ways of transportation in the present society, the central aim of these projects is proving affordable access to short distances in urban areas as an alternative means of public transport and private vehicles. These can reduce the sound pollution, air pollution. There are many health benefits mainly it improves the heart health. This paper would show insights about the bike sharing system – Capital Bikeshare - in Washington DC between 2011 and 2012. This data will be visualized by Tableau – one of the most common data visualization tools.

**Dataset overview**

There are two datasets which illustrates the same information about sharing bike in the system. One is separated by hours, while other is arranged by days. The data includes elements influencing the number of rented bikes such as time (date, season, year, month, holiday, weekday, working day) and weather (weather situation, temperature, real feel temperature, humidity, windspeed, number of casual users, number of registered users), and other related variables (record index, number of bike sharing). The only different column between two datasets is the “hour” which is extracted from the day.

In dataset, there are 17,379 records in “hour.csv” dataset , 731 records in “day.csv”, and no missing values in both dataset. It is assumed that the readers of this report are managers or marketing team of the Bikeshare System. Two factors (time and weather) can be used to draw the conclusions about the number of rented bikes, then the managers can know how the elements affect their business. Our team decided to choose variable “hour” and “working day” to answer the questions by comparing with weekday and weekends, related to time what time has the most demand in a day? and does the day off have more customers than in the working day? Between 2011 and 2012, we can compare the number of casual and registered users which impact the revenue of the company. Regarding weather elements, we think the temperature would mostly illustrate the customer behaviors compared to other weather-related factors.

**Visualize the data with Qlik Sense with research questions**

**Q1: Is number of sharing bikes on day off higher than that on working days?**

To answer this question, we require the information from columns “working day” and “cnt” (count number of rented bikes). The days off include holidays and weekends, otherwise working days. Because there are 5 working days and 2 days off (or slighty different because of the holiday), it is better to use average (mean) of number of bikes share to compare values in working days and days off.

Bar chart is the most suitable graph to compare the absolute values with zero base-line. Light green and light purple are chosen not to distract readers from the colors. Decluttering steps such as removing gridlines, axis titles, adding value labels are performed.

It can be seen that the average number of sharing is slightly higher on working days than that on Day-off. This can be understood because the main purpose of customer is commuting to workplace or schools.

A screenshot of a cell phone

Description automatically generated

***Figure 1:*** Average Number of Bikes Sharing between Day-off and Working days

**Q2: What time has the most demand in a day?**

The number of bikes sharing can be demonstrated by hours using data set “hour.csv” with line graph as we want to see the trend. The number of bikes rented is calculated by summing the total values by hours in column “cnt”. The same clutter steps are applied. The y-axis is hidden but the trend line includes the value labels. This can reduce the cognitive load.

According to the line graph below, we can conclude that 17:00 or 5pm has the highest number of rented bikes, while 4am has the least number of sharing bikes. 8 am and 5pm which are high demand hours as it is the time that people go to and from work or school. Hence, this conclusion would strengthen the one in the question 1 which states that the main purpose to rent a bike is commuting to work or school. To represent the trend from lower value to higher we used the gradient gray scale for making people differentiate the values easily.

A close up of a map

Description automatically generated

***Figure 2:*** Demand by Hours

**Q3: Is there any significant growth in the number of users?**

There are two variables (number of registered and casual) and years (2011 and 2012). The side-by-side bar chart is best fitted to compare overall values from 2011 to 2012, and between the count of casual and registered. The values shown in the bars are total in “cnt” column. Gridlines are removed, while axis tittles are changed to be suitable with the readers.

It can be seen from the graph that the number of registered users is always higher than that of casual. This means that the temporary customers take small part compared to the regular customers who use the bikes to commute. Besides, the demand for sharing bikes increased from 2011 to 2012. We used two contrast colors to differentiate the casual and registered users.

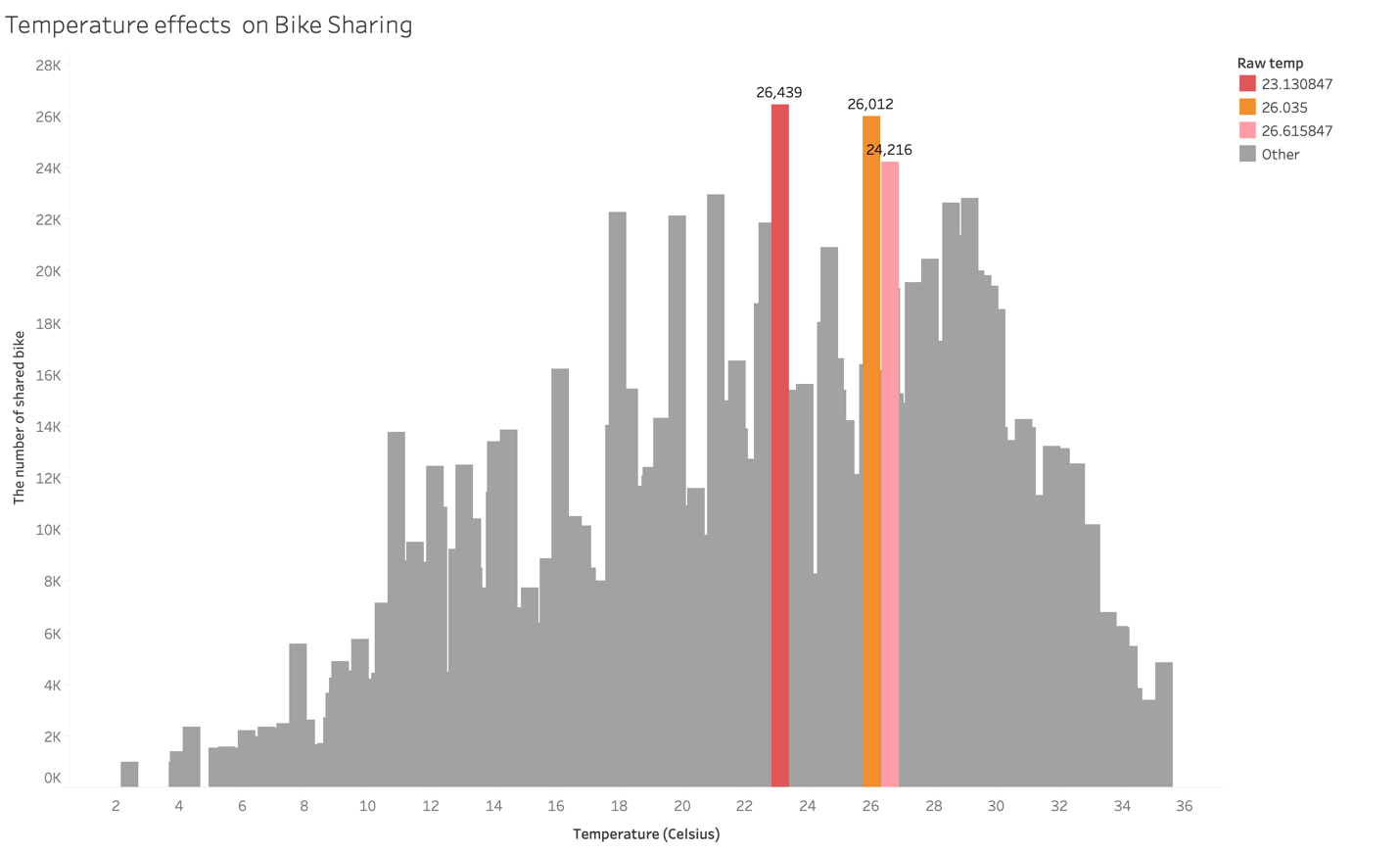
***A screenshot of a cell phone

Description automatically generated***

***Figure 3:*** The Number of Casual and Registered Users.

**Q4: What temperature does affect the number of sharing bikes?**

The temperature variable in the dataset show normalized values which are divided to 41. The relationship between number of sharing bikes and temperature can be demonstrated by the line bar chart. Gridlines are also removed, and similar formats are applied as above. We kept the y-axis labels because there are a greater number of values in the chart if we provide the labels for each bar it includes more clutter and make the chart messed. The bars’ color is changed to the grey so that we can clearly present the information behind the dataset and at same time we try to represent the top three values with labels to focus the temperature effect on human beings.



***Figure 4:*** Number of Sharing Bikes by Types of Users in 2011 and 2012.

Main reason behind analyzing this check the temperature effect, we generally have the perception when the temperature is a little higher than normal people does not prefer to walk, so obviously they go for the transport and it is clear evident that people are preferring the bike share when the temperature is little higher than normal. If it is hot people general prefer public transport and private vehicles as they need air conditioners to compensate.

**Conclusion.**

The whole story can be summarized in the dashboard below. Overall, readers can understand the type of customers with their specific purposes and their behaviors. Then, managers can determine the plans for the next coming time. This helps a lot to the business because we tried to provide the insights in the form of research questions based on the two systems which are very crucial for getting into conclusions.

A screenshot of a cell phone

Description automatically generated

***Figure 5:*** Dashboard of Bike Sharing from Capital Bikeshare System in 2011 and 2012