

Name: Barakat Akinsiku

Subject: Bikeshare Analysis Project Report

BIKE SHARE ANALYSIS PROJECT REPORT

INTRODUCTION

Bike sharing is a bicycle-sharing system and transport business model frequently found in some developed cities around the world. Alien to Nigeria because of the bad road infrastructure, highway risks, and the 'not so much of biking culture' (learning to ride bicycles at a young age), bike sharing in developed cities like London, New York, and a host of others involves users (customers) borrowing for a temporary use one of several bikes electronically docked at different stops and using them to commute to other similar stations at another location. It is an innovative transport model that helps the 'go green' initiative by reducing carbon emissions from fossil fuels as we would find in cars and other transport means that rely on fossil fuels, while at the same time providing a leisurely and relaxing way of travel that frees one from the perils of a road traffic jam.

PROJECT OVERVIEW

This rideshare analysis project draws the curtain on my 3-month data analysis starter program at HertechTrail Academy with the aim to put into practice the exploratory and visualization analytics skills I learned in the program. With this project, I aim to decipher from the ride-sharing dataset, the type of customers the bike-sharing company has, the gender distribution of customers it has, what routes are available, and which routes (start and end stations) receive the most customer traffic, what the average trip duration is and which days of the week receive the most activity and customer traffic. I also aim to find out which age group tends to opt for bike riding as a means of commuting the most.

DATA SOURCING

The data for the ride-sharing analysis was provided by tutors at the Academy as an excel file to test our understanding of the taught skills.

DATA CLEANING PROCESS

As one of the first key steps in the data analysis process following data sourcing, data cleaning involves preparing data for analysis by removing and modifying incorrect, irrelevant, incomplete, and improperly formatted and duplicated data.

After importing the bike share dataset into Microsoft Excel, I transformed it by looking for inconsistencies and errors within the data as well as adding new columns and parameters that would aid my analysis.

CLEANING PROCESS

The ride share dataset had a total of 661 rows and 9 columns including columns for Customer ID, Start time, End time, Trip duration, End Station, Start Station, User Type, Gender, and Birth Year. I checked for blank rows and found a few in the Gender and Birth year columns. I proceeded to fill in the blank rows under birth year with 0 and the spaces under gender with Not Available (NA) since the customers might have declined to fill in those bits and there was no logical way to impute the correct gender or birth year. I also checked the dataset to ensure all the columns had the correct data type.

Next, I created a new column to calculate the age from the birth year given. I calculated the age by subtracting the current year 2022 from each year already given in the birth year column.

The Start time and End time both had dates attached to the time and I had to use a text-to-column function to separate the time from the date and created a new column for the date. Thereafter, I used the Number format in the Home tab to convert the DateTime to time only. This way I had a start time, end time, and a date column respectively.

While there was a trip duration given in the columns, the values were higher than usual, and because it wasn't specified if it was a duration in distance (miles covered) or duration by time, I created a new column I called 'Duration time' to find duration by time. I did this by subtracting the newly separated start time from the end time and with this new duration time, was able to calculate the average trip duration, maximum trip duration, and minimum trip duration.

Next, I sought to establish the trip routes by creating a column called trip data where I combined both the start station and end stations on corresponding rows using the CONCATENATE function and with a COUNTIF function, determined the male and female customer distribution from the range of gender column and also the customer type distribution from the User type column. I also created a new column to extract the day of the week from the start date column. For this, I used a text TEXT(COLUMN 'ddd') function.

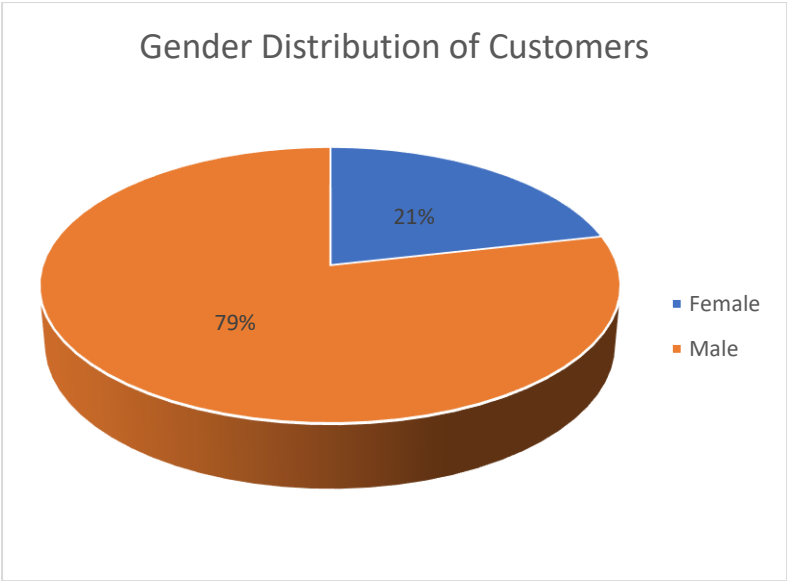
Finally, I created age bins by grouping the calculated customer ages into groups of 20-29, 30-39, 40-49... etc using the COUNTIF function to get the numbers for each age group. Following this, I moved on to creating pivot tables to gain a more in-depth analysis of my dataset.

ANALYSIS

For my analysis and to answer pressing data questions, I made use of pivot tables. With the pivot tables I was able to gain the following interesting insights:

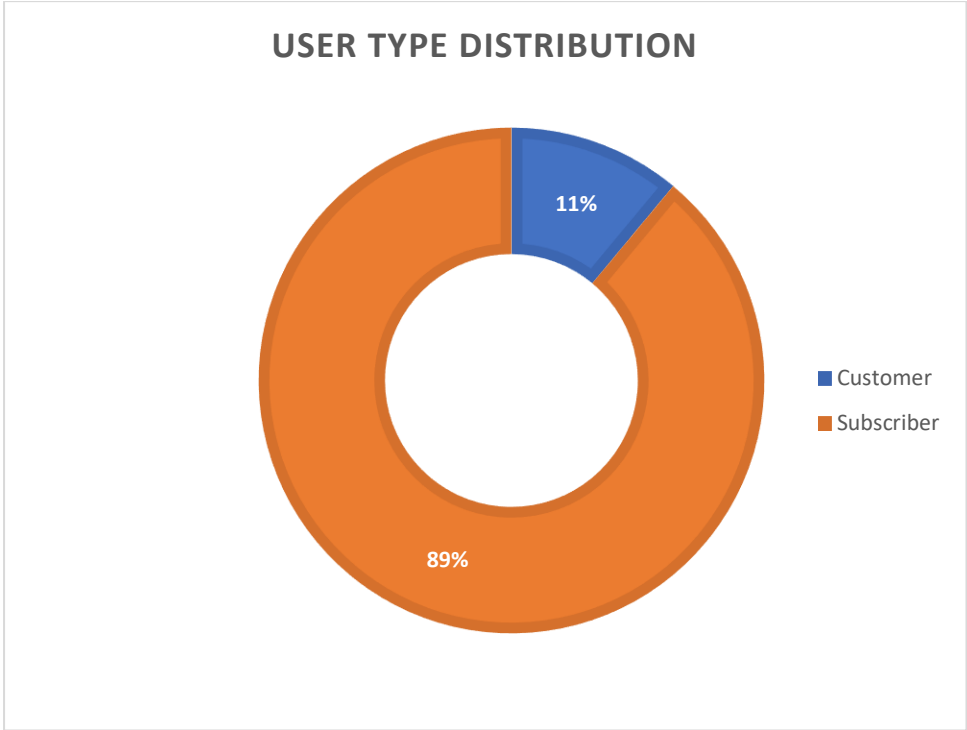
GENDER DISTRIBUTION

Male Customers accounted for more than half of the total customers making up 79% (457 male customers) to 21% (125 female customers)



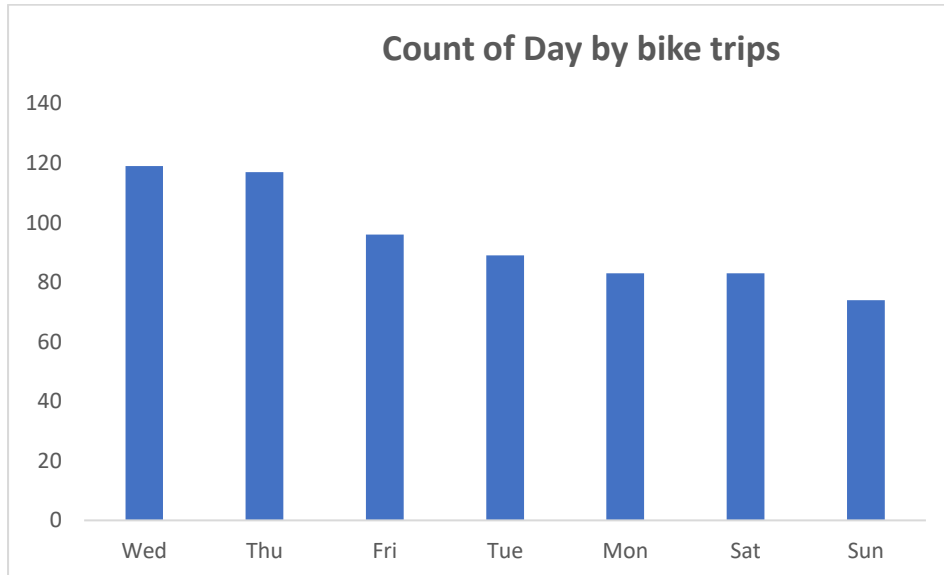
CUSTOMER SEGMENT

Subscribed users were the bulk of the bike-sharing company’s users at 588 Subscribers (89%) compared to 73 one-off users (11%)



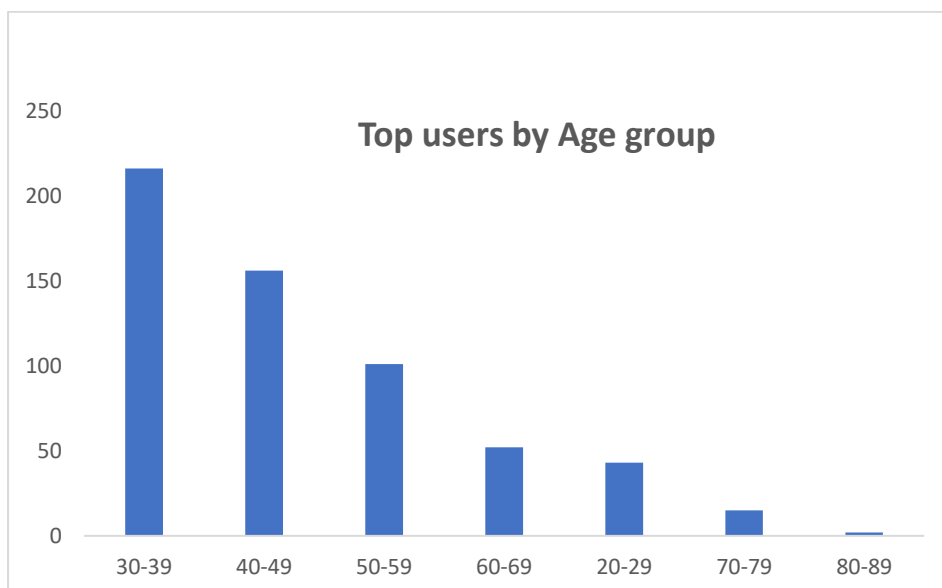
DAY WITH THE HIGHEST ACTIVITY

Thursday was the day with the highest bike rental activity with 119 trips (17.07%) of total recorded trips.



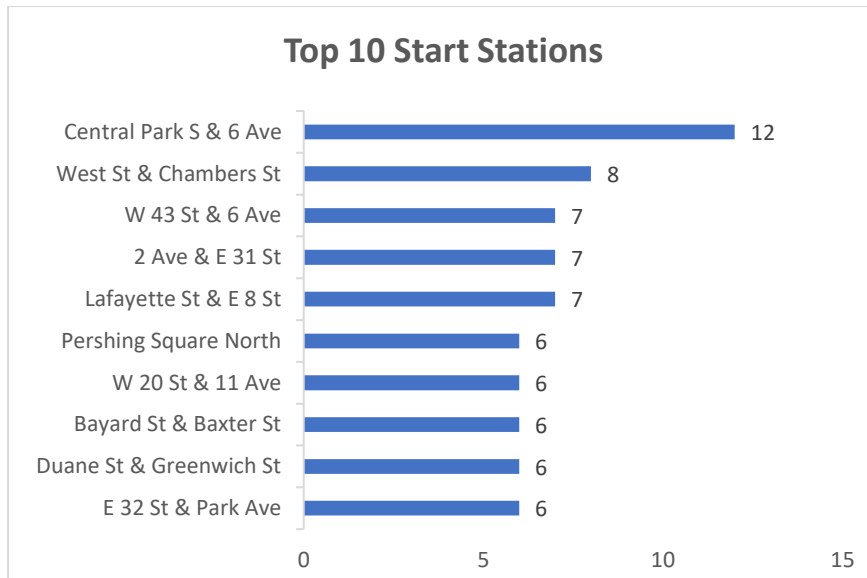
AGE-GROUP WITH THE HIGHEST BIKE USE

The age group with the highest bike use was the 30-39 age group with 216 trips made of the total while the age group with the least use was the 80-89 group.



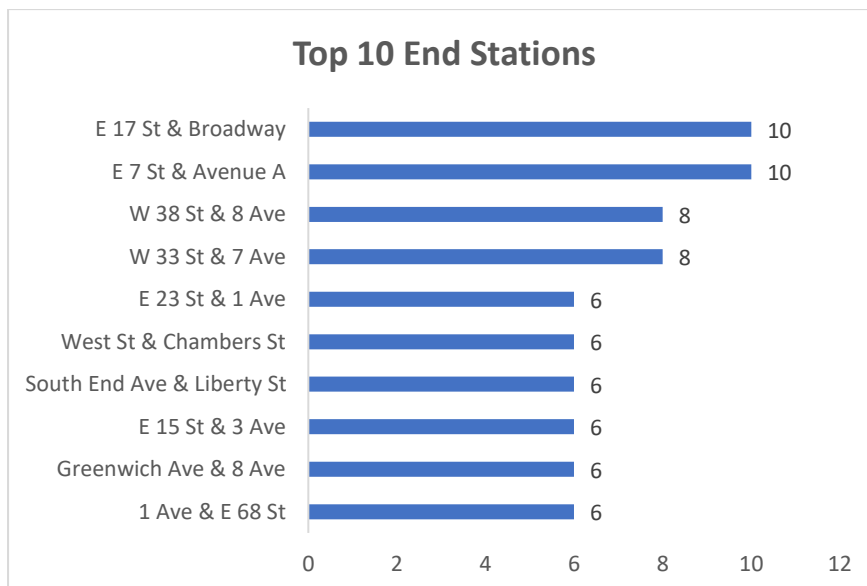
TOP START STATION

The top start station was Central Park S & 6 Avenue with 12 recorded trips



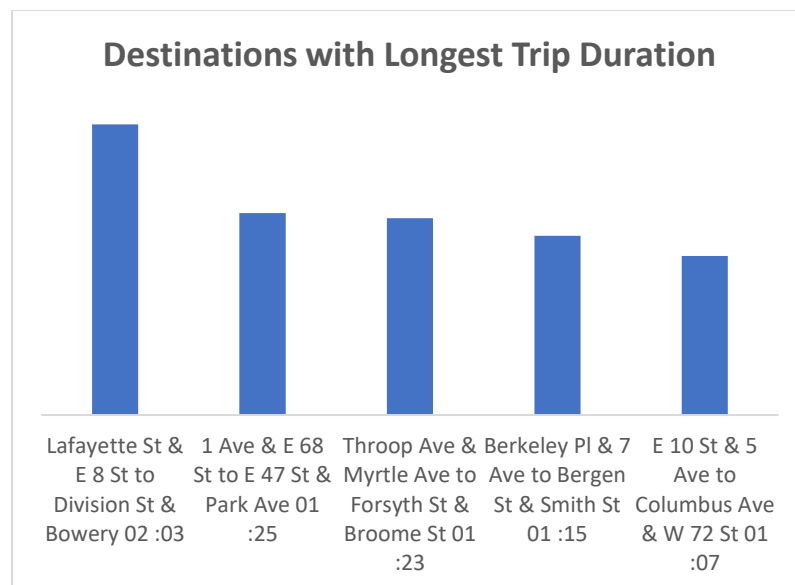
TOP END STATION

The top-end station was a tie between E 17 St & Broadway and E 7 St & Avenue with a total of 10 trips each



LONGEST TRIP DURATION

The longest trip duration was a 2hr 3min trip from Lafayette St & E 8 to Division St & Bowery.



AVERAGE, MINIMUM, AND MAXIMUM TRIP DURATION TIME

The average trip duration time was 13min, 36 secs while the shortest trip (MINIMUM) trip duration lasted 1min, 16secs. The longest trip duration was for 2hrs and 3min.

CONCLUSION AND RECOMMENDATION

A total of 661 trips were analyzed in the bike share dataset of which 79% were male and 21% were female. The 30-39 age group was the most active user while there were more subscribers than one-off customers. While this is a positive step, the 11% one-off users (customers) are a potential revenue source that can be targeted by the rideshare company to become repeat customers (subscribers). This will help improve the company's bottom line by bringing in more revenue and profit.

The majority of the users are also male and this could mean that men are more likely to use a bike share service or that a potential marketing opportunity exists where the bike share company can target and grow its female customer segment. To this end, more user research is recommended.

Next, Central Park S & Avenue was the top start station while there was a tie for the top-end station. An interesting observation made during my analysis was the trip from the start station Lafayette St to the end station Division St which took over 2 hours. It would be fitting to find out the distance in miles between both stations taking into consideration the various routes to determine why this trip took that long especially as the average trip duration for all trips was 13:36 secs.

Finally, the day with the most bike trips was Thursday. It would be equally interesting to find out why this is so. Was it due to more or less vehicular traffic on the road? Finding out these and the other aforementioned parameters would help in determining what promotional efforts can be made by the bike-share company to increase the subscriber base and subsequently gain or maintain market share.

Thank you.

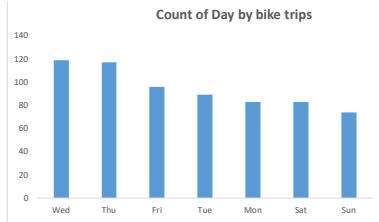
BIKE SHARE ANALYSIS DASHBOARD

AVERAGE TRIP DURATION
00:13:36

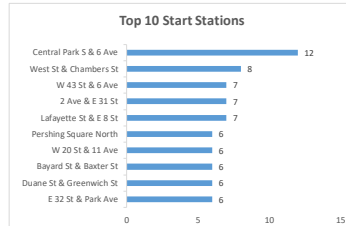
MAXIMUM TRIP DURATION
02:03:07

MINIMUM TRIP DURATION
00:01:16

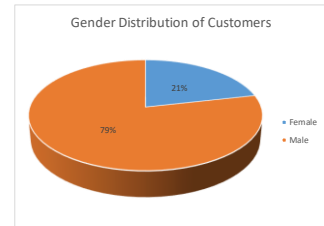
Thursday had the most bike trips with 119 trips (17.70%) of the total recorded trips



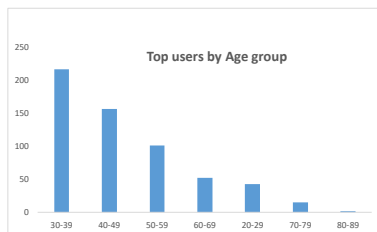
Central Park S & 6 Ave was the top Start Station with ten recorded entries out of the total



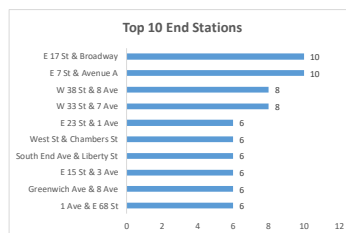
Male customers accounted for 79% of bike rentals compared to 21% of female customers



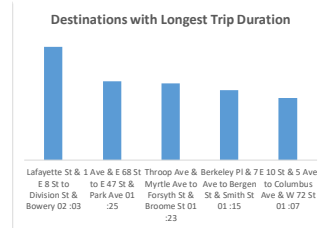
The age-group with most bike trips was the 30-39 age bracket while the least was the 80-89 age bracket



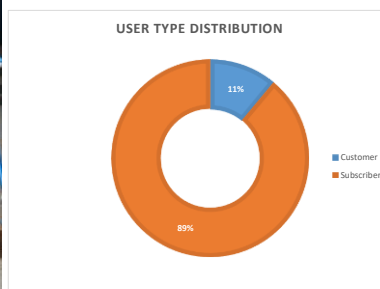
The top end station was E17 St & Broadway and E7 St & Avenue A



The longest trip duration of 2 hours, 3 minutes was a trip from Lafayette St & E to Division St & Bowery



The Customer segment distribution showed Subscribers are the bulk of users making up 89% of the total users



Summary of Insights and Recommendations

A total of 661 trips were analysed in the bike share dataset of which 79% were male and 21% were female. The 30-39 age group were the most active users while there were more subscribers than one-off customers. While this is a positive step, the 11% one-off users (customers) are a potential revenue source that can be targeted by the company to become repeating customers (subscribers). This will help improve the company's bottomline by bringing in more revenue and profit.

Majority of the users are also male and this could mean that men are more likely to use a bike share service or that a potential marketing opportunity exists where the bike share company can target and grow their female customer segment. To this end, more user research is recommended.

Next, Central Park S & Avenue was the top start station while there was a tie for the top end station. An interesting observation made was the trip from start station Lafayette St to end station Division St which took 2 hours. It would be fitting to find out the distance between both stations taking into consideration the various routes to determine why it took so long especially as the average trip duration for all trips was 13:36 secs.

Finally, the day with the most bike trips was Thursday. It would be equally interesting to find out why this is so. Was it due to more or less traffic on the road? Finding out these and the other afore-mentioned parameters would help in determining what promotional efforts can be made by the bike-share company to increase subscriber base and subsequently market share.