FinalProjectPaper

Ashani Jayasekera

2024-11-20

1. Introduction

1.1 Topic

This project analyzes the usage of the Austin Metro BikeShare, the bike-sharing program of Austin, TX. There are 3 main inquiries of focus:

1. The distribution of bike stations across the city of Austin and their capacity.

2. The usage of Austin BikeShare and whether the usage is dependent on the month in the year and whether it changes across years.

3. The usage of the bikesharing system across user types.

Austin, TX is currently expanding their bikeshare program with a $21 million investment. Using these inquires, city developers can determine where area where new docking stations are needed and where there is no existing stations. Furthermore, city officials can understand their user demographic (subscribers or single user) better and when during the year bikes are used more than others.

1.2 Source of Data

The data used for this project is "Austin Bikeshare", a public database from Austin, TX hosted on Google BigQuery. There are two 2 tables in this dataset: "bikeshare_stations," which includes the locations and the number of bikes that can be docked at a given station of the 82 active bike stations, and "bikeshare_trips," which include data from 2.27 million bike trips taken from 2013 to 2024. The trip date including the start and end times, duration, start and end stations, and the user type are the variables that are used in this study.

1.3 Github Repository, Packages, maps

The link to the GitHub repository is: https://github.com/ashanijay/SURV727.

The main R packages used in this study are `DBI`, `ggplot`, `ggmap`, and `dbplyr.` Stadia Maps is used for map visualizations.

2. Analysis

2.1 Bike Station analysis

To begin,the data is accessed via a connection to BigQuery using `DBI`'s `dbConnect` function.

The number of bike stations that are active in the city of Austin are identified. This is defined as I first checked how many working bike stations there are, defined as stations that have a non-zero number of bikes and are available for renting and returning bikes.

<div align="center">

Table 1: Number of active, working bike stations

f0__
82

</div>

There are 82 working bike stations currently. The proposed exapansion in the city of Austin will add an additional 1000 bike stations.

These bike stations are visualized below using `ggmap`.

From Figure 1, we can see that as of 2019, bike stations are mainly in downtown Austin. They are also some stations in Zilker Park and the UT Austin campus. There is one in the far West part of the city and a couple in the far East part of the city.

Most bike stations have a capacity of 13 or fewer bikes. These are the purple stations in the map. The orange stations are those that have 15 or bikes at the station. There are more of these in the north part of the city in "The Drag."

The mean number of bikes per station is 13.46, and the median is 13, suggesting again that most bike stations are small with exception of a few larger stations where the number of bikes is closer to 20.

2.2 Analyzing the number of trips taken across years and months

Next, the number of bike trips taken by month and by year are analyzed. Figure 2 below shows the result.
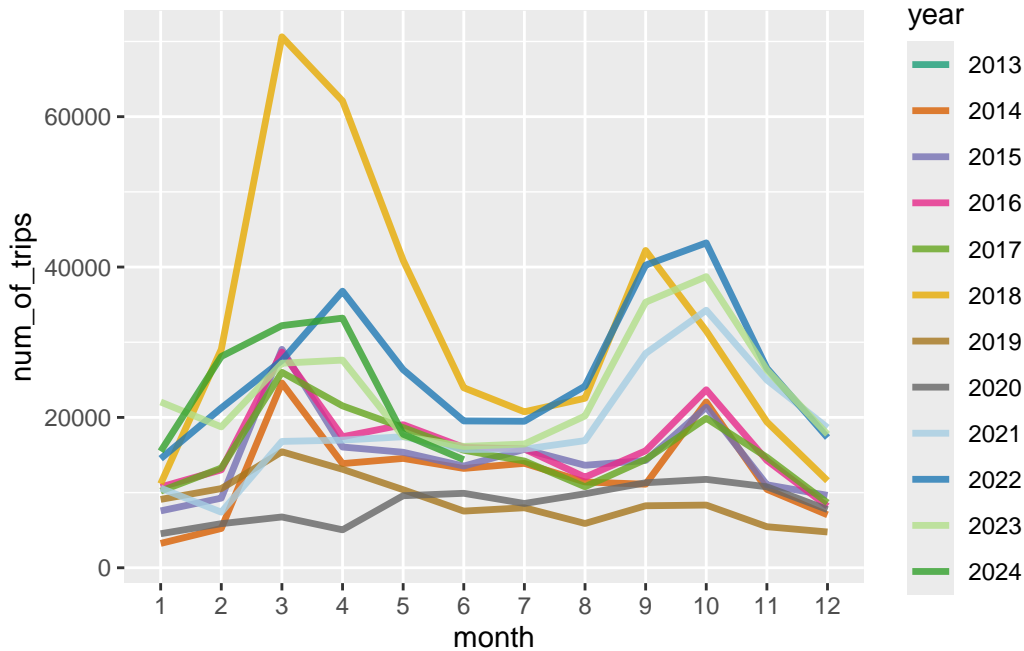

Figure 1: Number of bike trips by year and month

There are three things that are of note from Figure 2.

1. 2019 and 2020 had the lowest number of trips. A possible reason for this could be the COVID-19 pandemic where individuals were not going to work and not traveling.

2. The number of trips increased year by year from 2013 to 2018. 2021-2023 there is another increase in the number of trips post the COVID-19 pandemic. The reason is likely twofold: over time more people started to know and use the shared bikes, and the bike program continued to expand their coverage and capacity, attracting more new users.

3. There is an apparent seasonal difference in terms of number of trips. There are two peaks. Once between February and May and another between September and November. This is when there are more moderate temperatures in Austin. It is very warm in the summer and in the winter, it may be too cold to bike.

2.3 Analyzing the time taken number of trips taken across years and months

Another factor to look at is the duration of trips. Figure 3 and Figure 4 below show the average trip duration (in minute) by year and by month.
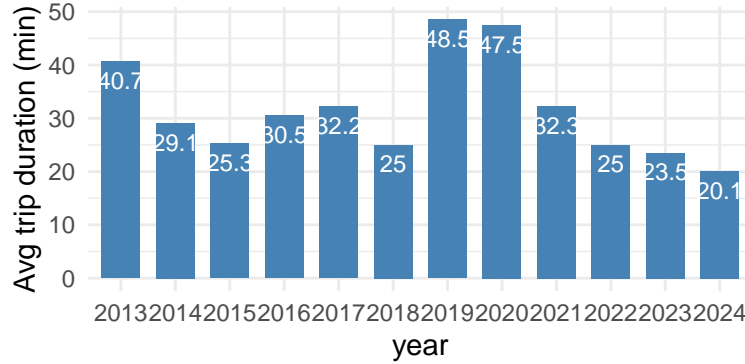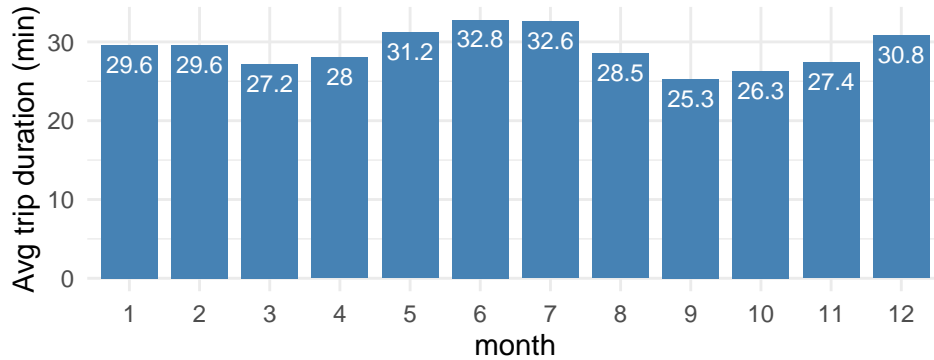
Figure 2: Trip duration by year

Figure 3: Trip duration by month

From Figure 3 and Figure 4, we can two things:

1. The average trip duration decreased over the yearsfrom 2013 to 2015 and then held fairly steady until 2019. When we were in the COVID-19 pandemic, the duration of the trips increased dramatically, nearly double that of 2018. (Note: 2024 is missing the last month and a half of the year as we are still in 2024.)

2. There is a slight seasonal difference in terms of trip duration. The summer months have longer average trip duration (31 -33 minutes per trip) compared to the rest of the year save for the month of December (25-30 minutes per trip).

2.4 Bike trip analysis – user types

Besides year and month, another interesting variable is whether the bike trip is made by a "subscriber", who holds an annual, monthly, or a student ridership plan, or a "single user", a person who either uses a day pass, a weekend pass, or a single trip ticket. There are 74 classifications of riders in this data. Using descriptions of the different riders the 74 classifications were collapsed into these two groups.

From the queried result above, we can see differences among subscribers and single users (and weekender) customers in terms of their bike usage. First, subscribers took fewer bike trips than customers – about 1.5 times more. Second, the trips taken by single users on average were around 34.8 minutes, while the trips taken by subscribers were much shorter on average 20 minutes; the difference in the average trip duration is around 15 mintues.

These differences are likely associated with the different purposes of trips: subscribers likely use the trip for commuting or short-distance connecting, while the customers are more likely to be tourists or occasional users who bought a day pass or single ride pass and used the bike to go around the city, and hence the longer trip duration.

2.4 Popular Bike Routes

Next, I analyzed the popular shared bike "routes", defined as different combinations of a start station (where the bike was rented) and an end station (where the bike was returned). Below is the result of the top 20 most popular routes, their number of trips and average trip duration.

Table 2: Most popular bike routes

route	num_trips	duration
Dean Keeton/Whitis to 21st/Speedway @ PCL	17672	8.4
21st/Speedway @ PCL to Dean Keeton/Speedway	16301	7.4
21st/Speedway @ PCL to Dean Keeton/Whitis	16116	12.7
Zilker Park to Zilker Park	14753	58.7
Dean Keeton/Speedway to 21st/Speedway @ PCL	14504	7.9
21st & Speedway @PCL to Dean Keeton & Speedway	14235	10.3
Dean Keeton & Speedway to 21st & Speedway @PCL	13543	10.4
26th/Nueces to 21st/Speedway @ PCL	12179	11.5
Riverside/South Lamar to Riverside/South Lamar	9764	52.5
21st/Speedway @ PCL to 21st/Guadalupe	9762	8.4

From the table, we can see that:

- There are some "routes" that share the same start station and end station. The fourth most popular routes is this type of route.

- For routes that have different start and end stations, the most popular route is "Dean Keeton/Whitis to 21st/Speedway @ PCL", which was taken 17672 times and for on average 8.4 minutes.

- The next two popular routes with different start and end stations are quite similar. Both start from 21st/Speedway @ PCL, and end at Dean Keeton/Speedway or Dean Keeton/Whitis. These 2 routes have a volume of 16301 times and 16116 times, and the average duration is 7.4 minutes and 12.7 minutes respectivlly.

However, when looking at popular routes taken by different types of users, we see very different patterns among subscribers and customers.
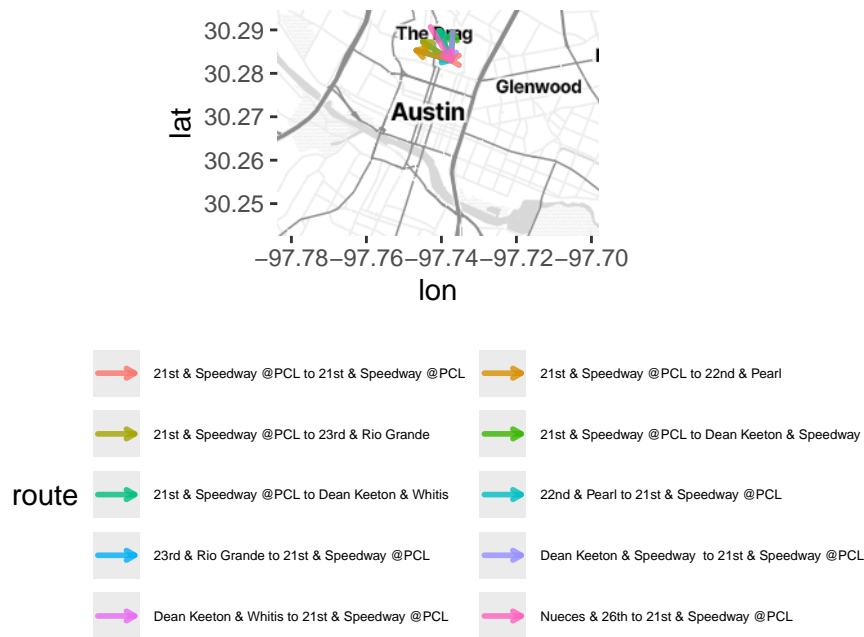

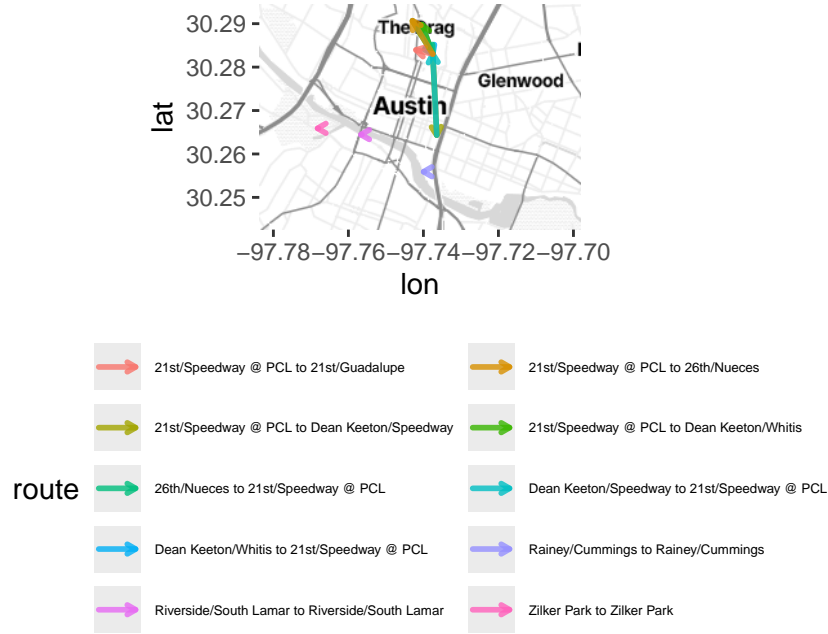Figure 4: Most popular bike routes taken by subscribers

Figure 5: Most popular bike routes taken by customers

We can see that:

1. Individuals that had annual, monthly, or student memberships did shorter trips from one location to another. Occasional users, the Single Users, did longer trips and tended to return the bike at the same station.

2. The most popular routes taken by subscription members are all in The Drag and are all short routes. These routes tend to be around 10 minutes long.

3. On the other hand, customers liked to rent and return bikes at stations in downtown Austin and the average duration was around 20-30 minutes.

3. Conclusion

3.1 Summary

This project analyzes the Austin bike share program from 2013 to 2024. The focus was on the distribution of the bike stations, their capacity, the users of the bike share program and the routes that were taken by the people. This information could be used to inform where additional stations should be built. Visualizations are presented with this information as well.

Need to update these sections

In conclusion, the number of bike trips increased year over year until 2019-2020 which can be explained by the COVID-19 pandemic. After 2020, the increase in rider trips again and has continue to grow year over year. However, in 2019 and 2020, the aerage duration of the rides was at its peak. There are seasonal differences in terms of number of bike trips taken. There are more trips taken in the spring and autumn. However, the trips that are taken are longer in the summer than during the rest of the year. Different types of users show very different shared bike riding behavior. The subscribers (who bought the annual ridership plan) took shorter trips than others. The most popular routes for subscriber riders where in the The Drag area of Austin. Customers with limited passes or single trip users took longer rides that more in different areas of the city. Customers were also more likely to rent and return the bike at the same station.

3.2 Limitation

There are a few limitations for this study. The data does not include time information. Therefore, it was not possible to look directly at when during the day were the trips being taken. The data also does not include the specific user that took the trip, just the type of user they are. If this information was avaialble we would be able to learn if there many repeat customers or just sinlge use customers. It would be interesting to repeat this study using the new stations that are to be built to determine if the trends in this study hold.