Data Visualization and Communication

ΒΟΥΛΓΑΡΑΚΗΣ - ΚΑΤΡΙΝΑΚΗΣ - ΜΑΡΓΩΝΗΣ

June 2024

Part 1

Daily Bike Usage

The total daily bike ride pattern shows a noticeable trend: usage is lowest on Mondays, gradually increases through the week, peaks on Saturdays, and then declines on Sundays. This pattern is consistent across all subcategories, with Saturdays experiencing the highest number of rides and Mondays the lowest.

The hourly usage of bikes varies depending on the day of the week. On weekdays, the distribution is bimodal, with two distinct peaks around 8 AM and 5 PM, corresponding to typical commuting hours. In contrast, weekends exhibit a nearly normal distribution, with peak usage occurring between 1 PM and 2 PM.

Bike Usage Line Plot

First of all, we observe an increasing usage of bikes over the years. Initially, there are two bike types: docked and electric. At the end of 2020, the classic bike is introduced. Following the introduction of the classic bike, the usage of docked bikes declines, suggesting that the classic bike might be a new version replacing the docked bike. The docked bike usage completely ceases by mid-September 2023.

Additionally, the monthly pattern indicates peak usage at the end of September and the start of October, followed by a decrease, likely due to the colder weather conditions in the winter months.

Stations and Bike lanes

From the bike lanes map, it is evident that bike stations are strategically built around existing bike lanes. The west side of the city, which has a more robust bike lane infrastructure, features a higher density of bike stations. In contrast, the east side of the city lacks sufficient bike lanes to support many bike stations, leading to fewer stations in that area.

Part 2

Deciphering bike types

A docked bicycle is a bike that is designed to be locked into a dock, rack, or sharing station. All Capital bikeshare bicycles are of this type, in contrast to "dockless" bicycles, that can be left or picked up anywhere, and that are offered by private companies, such as LimeBike, ofo, Mobike, Spin and JUMP. If so, then why are there 3 types of bicycle in the dataset:

classic, docked, and electric? Capital bikeshare's own website advocates that there are two types of bikes: classic and electric.
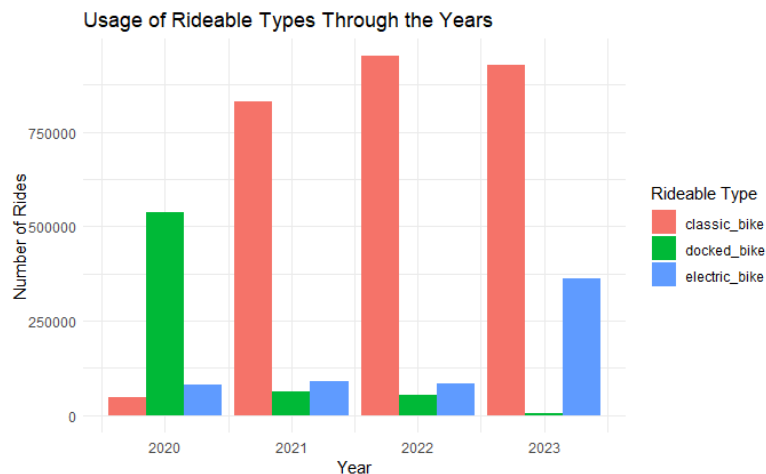
Looking at the usage of docked bikes we notice one very important thing: That docked bikes' use falls dramatically, while the classics' use takes its place.

This is further implied by the Daily bike rides, where we see the classic type spawn in the end of 2020, take the place of docked by 2021, and docked bikes vanish from use by September of 2023. Therefore, it is logical to conclude that docked bikes and classic bikes are a similar type of rideable, that one was introduced to replace the other gradually, and that they coexisted for about 2 years. Looking at the scatterplot of distance-duration, the different slopes tell us that docked and classic bikes have the same average speed- contrary to electric, which are faster.

Overall use grows over the years, perhaps as the company grows to more areas or it becomes more and more popular.
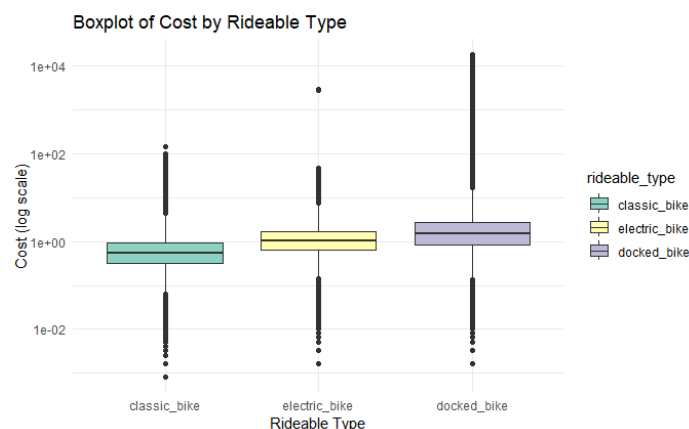
The stolen bikes?

By sheer size perhaps of the dataset, there are many outliers. So many in fact, and so extreme that they dominate the range of the distribution, leading the boxplot to be unintelligible. Using a log scale, to get a better view of the distribution, the message was still

clear; docked bikes, much more than classic or electric, are ridden to… oblivion. While it could be that the data themselves contain mistakes, the public owned company has to either... fix their data collection system or find ways to better secure their vehicles.

And perhaps it did just that, with the introduction of "classic" bikes to replace the "docked", though no official data regarding that have been published.

Fall 2020

But if we look at the total duration for each year, it tells a different story. In 2020, from September to November, a total of almost 2 billion hours were ridden only on docked bikes. There could be a handful of reasons for that:

1. The 2020 pandemic meant the city was impacted with ongoing restrictions and public health measures. Metro services were reduced, with reduced frequencies and operating hours to accommodate decreased ridership.
2. BLM movement. Following the killing of George Floyd, there were widespread protests for racial justice and against police brutality, including significant demonstrations in Washington DC in September and October. Metro stations near protest sites were occasionally closed or had limited access.
3. The 2020 presidential elections took place in November leading to a period of rallies and demonstrations in the capital, prior to and after the elections.

Barplot Duration Years
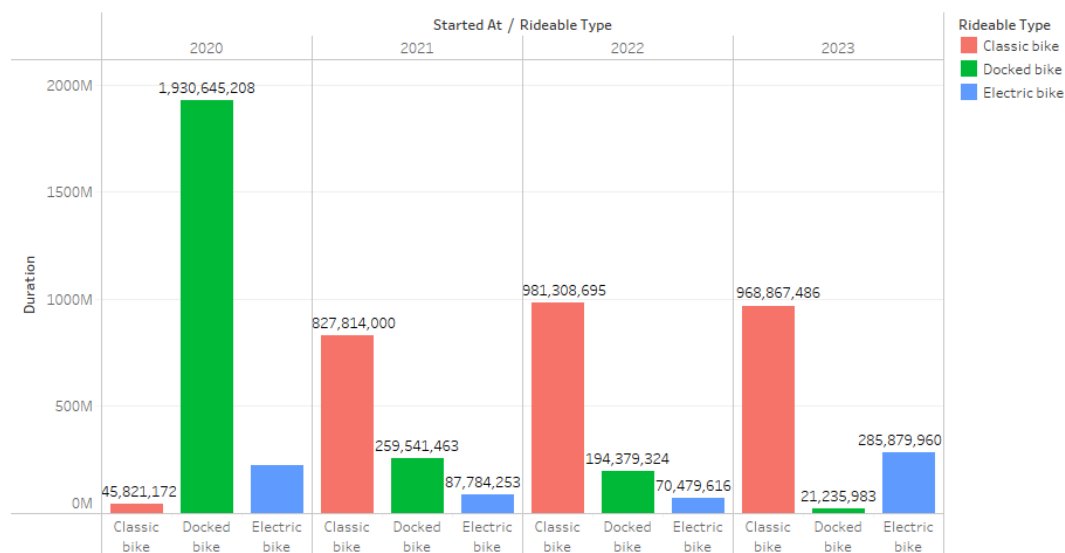
All the above reasons may be perhaps why the fall of 2020 was so busy for the city and hence for bike sharing. Reduced public transportation meant that people may have relied more on bikes to get around due to both the pandemic and protests, and conversely, political rallies and marches may have attracted more people to the city of DC.

Hours ridden on bikes
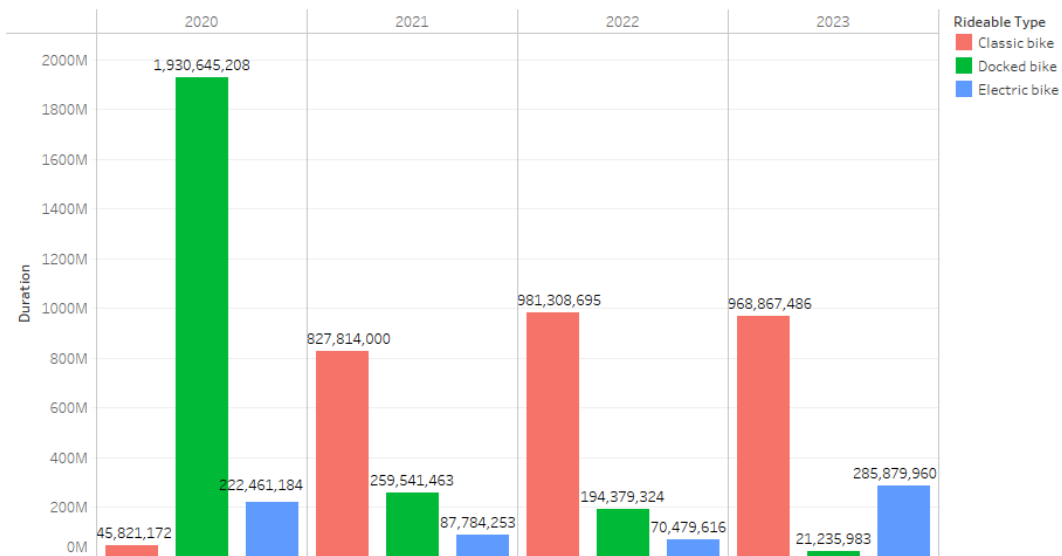
Part 3

The company's expansion over the years

The analysis commenced in September 2020, identifying a strong company presence within the Washington D.C. area, with smaller clusters in the surrounding suburbs. In the following year, the company increased its presence primarily in already established markets, specifically around Washington D.C. and the western area, while also attempting a minor expansion into the Greenbelt suburbs.

During 2022, the company's expansion efforts were concentrated within the D.C., Arlington, and Alexandria areas, which were deemed to be the most profitable. Finally, in 2023, the company supplemented established areas with new stations and expanded westward into the Fairfax area.

Examination of hourly station demand

In a city of almost 700,000 citizens, it is imperative to understand station usage patterns throughout the day. To this end, the analysis examines the number of outbound and inbound rides at each station, henceforth referred to as flow. If outbound rides predominate, the station is primarily used as a starting point during that time window, marked in green. Conversely, if inbound rides are prevalent, the station is used as an ending point, marked in red.

Several patterns emerge from this analysis. At 04:00, the city slowly wakes up with people moving from the surrounding areas toward downtown. This movement peaks at 10:00, where the map predominantly shows green stations, except for a few areas. Conversely, by

19:00, the pattern reverses, with most stations showing red, indicating people returning to their homes.

Profit by area

The above results directed the analysis towards examining total profits earned, with a primary focus on the year 2023, to extract insights that could inform the company's strategic plan. As expected, the Washington D.C. area dramatically outpaced the suburbs in profits, accumulating approximately 2.2 million dollars in 2023. Next, this dominant variable was excluded, and profits were examined on a per-county basis.

Initially, a strong correlation between a county's profitability and its proximity to the city center was observed. However, this assumption did not hold when comparing counties across different areas. For example, counties such as Fairland, Greenbelt, and Largo, which have been established markets for at least two years, generated less revenue than the newly established Fairfax. This finding led the analysis to examine the rides at each station, focusing on their profitability and hourly demand.

Route examination

To facilitate the comparison between counties, an interactive map was created, displaying the revenue generated by each station. Green indicates unprofitable stations, while red indicates profitable ones, with the color scale following a logarithmic progression to prevent outliers from dominating. Additionally, selecting a station reveals all rides to and from that station (traffic), and a bar chart shows the total traffic per hour for each station.

Using this interactive tool, the study examines edge cases within each area. Specifically, the most profitable stations from the counties of Arlington, Bethesda, National Harbor, Silver Spring, Reston, Fairfax, Largo, Rockville, and Fairland are analyzed. This comparison suggests a correlation between a station's distance from the city center, average traffic, and the distance of long-range rides that influence profit. Notably, stations further east tend to favor shorter trips, while those further west favor longer trips. When these stations' positions are overlaid on a street map, it is evident that the west side is connected to the city center by larger and wider roads.

Conclusions

At this point, two major assumptions can be made regarding the conditions for increasing profit. Firstly, the company must facilitate easy access from surrounding areas to the most popular ones, creating more opportunities for citizens to move within those areas and potentially generating positive feedback loops. This assumption is supported by the company's past expansion strategy detailed in previous sections.

Secondly, by examining citizens' hourly habits, the company must optimize the number and type of bikes allocated to each station. For example, in the morning, stations in the most popular parts of the D.C. area can have fewer bikes, reallocating most of them to neighboring areas where citizens reside. Alternatively, suburban stations that engage in long-range rides can be stocked with more electric bikes.

Moving forward, the demand and traffic patterns on the far east side indicate a stagnated area, suggesting that further investment there is discouraged. Instead, the rapid growth of Fairfax and the general demand exhibited by the western area, combined with the more advantageous road infrastructure, could present new opportunities worth exploring. Therefore, future analysis could investigate the company's potential expansion into the county of Centreville.

Contributions

The following table details the tasks completed by each participant:

Name	Report	Presentation
Dionisis Voulgarakis	Part 1	Pages 1-8
Christos Katrinakis	Part 2	Pages 9-20
Phevos Margonis	Part 3	Pages 21-95