CITI BIKE 2022
CREATING A STRATEGIC DASHBOARD

2.2 RESEARCH QUESTIONS & ANALYTICAL APPROACH

2.2 RESEARCH QUESTIONS & ANALYTICAL APPROACH

CREATED BY OLA GAFFAROVA

BY OLA GAFFAROVA

RESEARCH QUESTIONS & ANALYTICAL APPROACH

Strategic Dashboard Planning

Questions

How can Citi Bike optimize its station network to balance demand and supply efficiently across New York City?

What will our team show on the dashboard?

We will create a strategic dashboard, which will aim to help management make informed decisions about the expansion of the brand's supply.

We need to look at the strategic placement of new stations across the city to boost the service supply. Installing stations and maintaining bikes is an expensive endeavor, and nailing the exact spots where the extra bikes are needed will help ensure the investment generates more revenue for the company.

There are a couple of research questions to consider when establishing how to approach the problem. The best way to do this is to try and think of different hypotheses you could explore:

Think about why there might be bike shortages in certain places. It's safe to assume that the stations where shortages happen are the most popular. This raises the question:

What are the most popular stations in the city?

Let's assume that bike trip numbers aren't the same throughout the year. Perhaps, you can hypothesize that the seasons and months play a role in resulting in more trips in warm weather and fewer trips in colder weather:

Which are the months with the most trips taken? Is there a weather component at play?

You can also assume that there are popular routes between stations, e.g., that not only start/end stations are popular but that people also tend to journey between certain stations over others:

What are the most popular trips between stations?

Given the main problem is a lack of demand in certain locations, you could also explore whether the bike stations are evenly distributed. The hypothesis here is that maybe there are locations in the city where some stations have larger gaps than others:

Are the existing stations evenly distributed?

RESEARCH QUESTIONS & ANALYTICAL APPROACH

Strategic Dashboard Planning

- Use descriptive analysis and apply aggregations of bike trips across New York to discover the most popular starting locations and summarize data yearly to find seasonal patterns.
- Apply geographic plotting to identify problem areas in station distribution and explore the most common routes.

Research Question	Hypothesis / Analytical Focus	Planned Analysis	Insight for Strategy
What are the most popular start and end stations in the city?	Certain stations (near transport hubs or residential/commercial centers) have much higher trip counts, indicating higher demand.	Count total trips per start and end station; rank top 10 stations.	Identify top-demand locations; potential need for more bikes nearby.
Which are the months with the most trips taken?	Warm months (spring/ summer) generate more trips due to favorable weather.	Aggregate trips by month; compare trip counts across months.	Understand seasonal demand; plan maintenance and redistribution schedules.
Is there a weather component at play?	Higher temperatures and less precipitation correlate with increased bike usage.	Merge weather data with trip data; analyze relationship between temperature/precipitation and trip count.	Adjust supply and marketing based on weather impact.
What are the most popular trips between stations?	Certain station pairs form frequently used routes.	Group by start–end station pairs; rank top 10 routes.	Highlight corridors for capacity expansion.
Are the existing stations evenly distributed?	Some neighborhoods may have fewer stations, creating access inequality.	Map all stations and compute average nearest-neighbor distances or density.	Identify underserved areas and prioritize new station placement.

RESEARCH QUESTIONS & ANALYTICAL APPROACH

Strategic Dashboard Planning

Once the plan is established, I will begin working on data preparation. Since the Citi Bike data is divided into multiple monthly CSV files, I will use list comprehension in Python to read and combine them efficiently. This method will make the code more concise and easier to maintain. To handle the large data volume, I will also use a generator expression to concatenate all datasets into one large DataFrame without overloading memory.

After consolidating the data, I will explore its structure and perform initial descriptive analysis. This will include checking column names, data types, and missing values.

I expect to find that the most popular start and end stations will be located near transportation and commercial hubs, such as Hoboken Terminal or Grove Street PATH, which are typically high-traffic areas for commuters. Identifying these high-demand locations will help guide later recommendations on where new stations or additional bikes might be most needed.

To better understand external factors influencing usage, I will also integrate weather data into the analysis. Using NOAA's public API, I will source daily weather observations for 2022, focusing on temperature and precipitation. After retrieving and cleaning the data, I will merge it with the Citi Bike trip dataset based on date, creating a unified dataset that includes both operational and environmental variables. This will allow me to explore whether warmer temperatures or lower rainfall correspond to higher trip counts, helping to confirm or reject the hypothesis that weather significantly affects bike demand.

Throughout this process, I will pay attention to clarity and reproducibility in the notebook. I will include comments explaining each step, particularly how list comprehension and generators work, to make the code easier to follow and understand. Each analytical step will directly prepare the data for visualization and interpretation in later parts of the project.

By the end of this stage, I will have a fully prepared dataset and a clear analytical framework for the dashboard. The next step will be to perform descriptive and comparative analyses to identify seasonal trends, spatial demand patterns, and correlations between weather conditions and trip frequency. These results will then be visualized in the dashboard, helping management pinpoint areas with unmet demand and prioritize where to expand or rebalance Citi Bike stations.

RESEARCH QUESTIONS & ANALYTICAL APPROACH

Strategic Dashboard Planning

The final deliverable will be a strategic dashboard that brings together several complementary visualizations, each answering one of the core research questions. The dashboard will be designed to provide both a quick overview for management and interactive elements for deeper exploration.

To understand which stations are the most popular, I will include an interactive horizontal bar chart showing the top ten start and end stations by number of trips. Users will be able to hover over each bar to view the exact trip count and switch between start and end stations using a simple filter. This visual will help quickly identify which areas experience the highest demand and may need capacity expansion.

To explore seasonal usage patterns, the dashboard will feature a line chart of monthly trip volumes across the year. The chart will show the fluctuation of ridership over time, helping confirm whether warm weather leads to higher demand. Next to it, a dual-axis chart will display the same monthly trip counts alongside average monthly temperature data, allowing users to see how weather conditions correlate with ridership trends. Together, these visuals will help identify seasonal peaks and dips, supporting planning decisions for maintenance and staffing.

To reveal the most frequent travel connections, I will include a network map or flow visualization highlighting the most popular routes between stations. Each line between two points will represent a trip connection, with line thickness indicating the number of trips. This component will give a clear picture of how riders move through the system and whether certain routes could benefit from additional bikes or docks.

To analyze how stations are distributed across the city, the dashboard will include an interactive geospatial map. Each station will appear as a point on the city map, with color intensity or clustering indicating station density. Users will be able to zoom in on specific neighborhoods to identify underserved areas or regions with few nearby stations. Overlaying trip frequency or population density (if available) will provide further insight into potential expansion zones.

All these elements will be connected through interactive filters—such as time period, station name, or neighborhood—allowing the user to focus on particular parts of the city or specific time frames. The combination of bar, line, and map visualizations will make the dashboard both strategic and exploratory, supporting management in making informed, data-driven decisions about where to allocate new stations and how to optimize existing ones.