

Bike Share Project Plan

Joseph Smith

OBJECTIVE

The purpose of the Bike Share Expansion Program is to evaluate the current routes of popular routes, population centers and destination attractions, for both amusement and employment bikesharing and add stations in X,Y, Z

HYPOTHESIS OR PROBLEM STATEMENT

- Define and focus the purpose for your analysis.

If COVID has impacted the number of rideshare users, then how has individual station popularity changed since the start of 2020?

- Clearly identify the benefit to the business or organization.

By doing this analysis we will be able to see how station popularity has changed and if we need to expand to new areas.

- What are you hoping to accomplish in the analysis? What are your desired outcomes?

Goal: eliminate any stations that have decreased popularity drastically since 2019 and relocate them to new areas

DATA PRIMARY AND SECONDARY SOURCES

- Which datasets have you selected?

Citi Bike NYC Data - [link](#), 2016-2019

- Based on the "Story of One Row" what is the potential value of each dataset?

Based on the first dataset we can see what type of user the rider is, where they started and where they ended, how long the ride was, how old and what gender the user is, if the user is a subscriber, and more.

- Which columns or calculations do you expect to provide key insights for the analysis? How do you expect they will contribute to the data story?

Columns to create: ride_id,

Calculations: qty of rides 2019,2020,2021,2022 by month, % of user_type by year, concat of popular routes

- Profile each dataset that you plan to include in the analysis listing its general attributes (size, range, descriptive statistics, as appropriate).

All data not yet known

Citibike2019 - total rows: 404,947

- Secure secondary data to expand the support for the analysis.

Will be securing additional rideshare data from citibike website

Bike counts from the city - [link](#)

- Create a data dictionary for your selected data. Note any columns to be excluded.

❖ Start_time

❖ Stop_time

❖ Start_station_id

❖ End_station_id

❖ Bike_id

❖ User_type

EXCLUDED

- | | |
|--------------|-------------|
| ❖ birth_year | ❖ Longitude |
| ❖ Name | ❖ Docks |
| ❖ Latitude | ❖ Gender |

ANALYSIS DATA DRIVEN INSIGHTS

- What "Key Performance Indicators" will provide you with insights toward your answer the goals outlined in your Purpose statement.

- ❖ Counts
- ❖ Percentages
- ❖ Change over time

- Create a focused list of questions you will be using as starting points for your analysis.

1. Look at num_rideshare_users increased/decreased?
2. Has subscriber v customer % changed?
3. What were the most popular stations then v now?
4. Are there any stations that are no longer being used?

- Describe the process you will implement to draw observations.

- Are there any models or “what if” scenarios that you plan to build and to test your data with?

Not yet - but will continue to keep it top of mind

- What method(s) will you use to confirm your findings?

- Make note of any limitations or assumptions in the data that will affect your analysis.

Limitations: Data can be incomplete for 2020.

Data collection methodology has changed slightly since 2019.

Assumptions: Riders are using their own account (not subscribers).