Name: **Muley, Tushar**
Assignment: **DSC 680 - Week 5 Project Proposal Milestone 1**
Date: **January 16, 2021**

Metro Bike Share System

Contents:

Topic:

Bike sharing has come a longs way since it was first seen in 1965 in Amsterdam. Since urban city populations are growing and cars are trapped in traffic jams bike are becoming viable options to getting the city population around. Another reason is climate change. As automobiles are contributor to health and global warming many of the world's largest city are looking to limit their usage or band them from entering city centers. Enter a possible solution to getting around. New bike sharing companies are showing up in a majority of city. In the United States bike sharing program started in 2007 in Tulsa, Oklahoma. The growth has been stead since and is expected to increase from $1.5 billion in 2020 to $4.4 billion by 2027.[2]

Question to research:

The data I located is from 2021. I would say well enough into the pandemic to return to some kind of a normalcy. Below are the research questions I am proposing to answer.

1. Does ridership change over the course of a year? Does it go up or does it go down?
2. How is usage of the bikes? Do weekend or weekdays impact usage?
3. How is the ride duration change between membership riders and casual riders (one-time users)?
4. Identify the busiest bike stations
5. What features (station location, electric bites) are influencing trip count?
6. Building a model to predict the number of trips from a station?

Dataset:

I obtained the data set from Metro Bike Share. Metro partnered with the City of Los Angeles to provide bikes 24/7, 365 days a year in Downtown Los Angeles, Central Los Angeles and North Hollywood. I pulled three quarters of data for 2021, which is the most recent data available. The appendix section contains column names and description of the attributes for two tables. One table contains bike usage data (titled Bike Usages Table) and the second table contain station information (titled Station Information Table).

Data obtained from: https://bikeshare.metro.net/about/data/

Methods:

1. Perform analysis for the data sets
    a. Pull samples from the Bike Usage data
    b. Pull samples from the Stations information data
    c. Merge the four Bike Usage data files into one (files are quarterly a single file would be easier to analyze)
    d. Merge the Station information data into the now very large Bike Usage data
    e. Assess values types
        i. Remove NaN (Nulls) or updated them depending on which column data is missing
        ii. Split date and time data into their own columns (date and time is in single column)
        iii. Update any columns names if needed
        iv. Update any letter types to lowercase
        v. Change human readable to indicators for analysis and model consumption
        vi. Update numeric strings to int
2. Perform graphical analysis (histogram, scatter plots and correlation) to better understand what kind of data I am dealing with. Build any visuals that are need for the analysis
3. Collect/Connect additional data if needed
4. Based on the data select a few models to predict trips from a station
5. Run model
    a. Review data
    b. Determine changes
    c. Build visuals of the data story telling
6. Pull all my analysis together
    a. Polish visuals with correct titles, axis labels and colors
    b. Prep PowerPoint presentation
    c. Prep script for presentation
    d. Complete presentation

Ethical and Other Considerations:

This data is publicly available due to the partnership with City of Los Angeles. The data does not contain any personally identifiable information. The information can be use to determine bike usage and consideration for additional stations. The data might be used to determine better bus routes or other public transportation options. Metro Bike does do some scrubbing of the data prior to releasing it. If financial information was provided like type of credit card used or age or birth dates this information could be used for rider segmentation.

If age data was available it could be used to determine age range of riders. This would allow for information to promote on the Metro Bike app. This could be used to generate other add revenue similar to gas station adds as your filling up your tank. This information could be used to tie in other sponsors like local restaurants or bars or shopping venues. This would increase Metro Bikes revenue from alternative sources.

Outside of the promotion aspect of the data if more attributes were provided one could determine better bike stations locations based on address of users. The data could be used to increase pedestrian and bike lanes. Usage of the data could be used to possibly close street to cars to increasing foot traffic. This would reduce parking needs and may lead to increase business for nearby business.

Challenges/Issues:

The data seems straight forward. It was obtained from the source meaning no combining of data with other information that might cause unknown arrogation or interpretation problems. I might try and combined the data with day or week information. In that cause I would need to table the provides this information or find the formula to determine day of week. My biggest concern is how would I deal with the large volume of data? It could cause memory issue or other unknown issue I have yet to confront.

Reference:

1. LA metro bike share network https://bikeshare.metro.net/about/
2. McEvoy, J. (January 2022). “Bike Sharing Market Accelerating to $4.4 Billion by 2027”. From Startup Savant. https://startupsavant.com/news/bike-sharing-market
3. “Bicycle sharing System”. (December 2021). From Wikipedia.com. https://en.wikipedia.org/wiki/Bicycle-sharing_system
4. Walker, A. (December 2019). “The quiet triumph of bike share”. From curbed.com. https://archive.curbed.com/2019/12/16/20864145/bike-share-citi-bike-jump-uber
5. Malouff, D (January 2017). “All 119 US bikeshare systems, ranked by size”. From Greater, Greater Washington.org https://ggwash.org/view/62137/all-119-us-bikeshare-systems-ranked-by-size
6. Hosford, K & Winters, M & Sersli, S. (January 2020). “More people are using bike share programs, but the gender gap persists”. From Green Biz.com https://www.greenbiz.com/article/more-people-are-using-bike-share-programs-gender-gap-persists
7. Perry, S. (May 2020).” More people commute by bike in cities with bike-sharing programs, study finds”. From MINN POST.com https://www.minnpost.com/second-opinion/2020/05/more-people-commute-by-bike-in-cities-with-bike-sharing-programs-study-finds/

Appendix:

Bike Usage Table:

Column Name	Column Description
trip_id	Locally unique integer that identifies the trip
Duration	Length of trip in minutes
start_time	The date/time when the trip began, presented in ISO 8601 format in local time
end_time	The date/time when the trip ended, presented in ISO 8601 format in local time
start_station	The station ID where the trip originated (for station name and more information on each station see the Station Table)
start_lat	The latitude of the station where the trip originated
start_lon	The longitude of the station where the trip originated
end_station	The station ID where the trip terminated (for station name and more information on each station see the Station Table)
end_lat	The latitude of the station where the trip terminated
end_lon	The longitude of the station where the trip terminated
bike_id	Locally unique integer that identifies the bike
plan_duration	The number of days that the plan the passholder is using entitles them to ride; 0 is used for a single ride plan (Walk-up)
trip_route_category	"Round Trip" for trips starting and ending at the same station or "One Way" for all other trips
passholder_type	The name of the passholder's plan
bike_type	The kind of bike used on the trip, including standard pedal-powered bikes, electric assist bikes, or smart bikes.

Station Information Table:

Column Name	Column Description
Station ID	Unique integer that identifies the station (this is the same ID used in the Trips and Station Status data)
Station Name	The public name of the station. "Virtual Station" is used by staff to check in or check out a bike remotely for a special event or in a situation in which a bike could not otherwise be checked in or out to a station.
Go live date	The date that the station was first available
Region	The municipality or area where a station is located, includes DTLA (Downtown LA), Pasadena, Port of LA, Venice
Status	"Active" for stations available or "Inactive" for stations that are not available as of the latest update