REPORT: CYCLISTIC BIKESHARE PROJECT

INTRODUCTION

The files for this Cyclistic bikeshare project can be viewed in **Github** and **Tableau**.

The project files are saved in **Github** can be accessed by clicking the file name links listed below:

- **Report-Cyclistic Bikeshare Project.pdf** which explains this data analytics project
- **Case study** guidelines for the project
- Changelog to document the data cleaning & verification done using SQL (Structured Query Language), and the query files containing SQL programming code:
    - **Changelog Clean Verify Data SQL.xlsx**
    - **1-SQL Data Clean & Verify**
- Log to document the data analysis done using SQL, and the query files containing SQL programming code:
    - **LOG- DATA ANALYSIS- SQL-File1.xlsx**
    - **LOG- DATA ANALYSIS- SQL-File2.xlsx**
    - **2-SQL Data Analysis**
- R programming code files**:**
    - **R Markdown Data Analysis.pdf** (a PDF report explaining the data analysis using R)
    - **rmarkdown_data_analysis.rmd** (the R Markdown Notebook)
    - **rmarkdown_data_analysis.html** (a Chrome HTML document)
- Original dataset consisting of 13 CSV files:
    - **Dataset Cyclistic – Folder 1**
    - **Dataset Cyclistic – Folder 2**
- Dataset cleaned in SQL exported and saved as CSV files to the folder **SQL Cleaned Dataset**
- Dataset used to create the dashboards in Tableau:
    - **tableau_import_data.xlsx**
    - **bike_routesv2.xlsx**

Click the following links to view the dashboards designed in **Tableau:**

- **Cyclistic Project Dashboard of Bike Rides by Season, Day & Time**
- **Cyclistic Project Map of Top 10 Bike Routes & Stations**

SQL, EXCEL, TABLEAU, GITHUB, R PROGRAMMING

The programming languages and software I used for this project are **Microsoft SQL Server, Microsoft Excel, RStudio, Tableau,** and **Github**. The data cleaning and verification were done using SQL and Excel. The data analysis was performed using SQL and R Programming. The data visualizations were designed using R and Tableau.

The **SQL** queries and functions I used included the following:

- IMPORT FLAT FILE (to import CSV files into SQL Server and to do data mapping)
- INSERT INTO (to merge 12 data tables into a single data table)
- SELF JOIN basd on an INNER JOIN (to join data in two tables)
- COALESCE (to identify and replace NULL values in various fields)
- TEMPORARY TABLES (to perform complicated calculations)
- CASE (to find and replace NULL values)
- SUBQUERY (to nest a query inside a second query)
- CALCULATED FIELD (to use various other fields to calculate and create a new field)
- CONCAT (to combine fields)
- DATEDIFF (to return the time difference between dates in days, minutes or seconds)
- DATENAME (to return a name for the month or day from a date & time)
- DATEPART (to return a number for the month or hour from a date & time)
- CONVERT (to change the data type of fields and include a style)
- CAST (to change the data type of fields)
- GROUP BY & HAVING (to group data and filter)
- WHERE (to filter records)
- Operators AND, OR, EQUALS, IN, NOT
- IS NULL and ' ' (to identify NULL values)
- DISTINCT (to identify unique records)
- ALTER TABLE (to delete fields)
- ALIAS (to label newly created fields and tables in a JOIN)
- TRIM (to remove leading & trailing spaces in fields with data type string)
- COUNT (to count records (rows))
- ORDER BY DESC or ASC (to sort records)
- SUBSTRING (to extract characters out of a string and to split data into various fields)
- SP_HELP procedure (to view the data type of each field in a table)
- TOP 10

I applied some **Excel** spreadsheet functions and tools. These included the following:

- PIVOT TABLES with CALCULATED FIELDS
- VLOOKUP function
- IF function
- CONDITIONAL FORMATTING

I used the following packages, notebooks, operators and functions in **R:**

- R MARKDOWN NOTEBOOK (to document R code, output and visualizations)

- TIDYVERSE, GGPLOT2, DPLYR, TIDYR & READR (packages for data wrangling, analysis & visualization)
- LUBRIDATE (package to work with dates)
- GGPLOT + AES + GEOM_BAR (functions for bar chart visualizations)
- GGPLOT + AES+ GEOM_POINT (functions for scatter plot visualizations)
- GEOM_SMOOTH (function for regression lines to visualize trends in the data)
- FACET_WRAP (function to display separate plots from data subsets)
- ANNOTATE (function to highlight important data points in a visualization)
- LABS (function to label plots)
- PIPE (used for sequence of multiple operations)
- SELECT (function to select columns)
- ARRANGE (function to sort data)
- FILTER (function to filter data)
- COUNT (function to count data)
- UNITE (function to concatenate columns)
- READ_CSV (function to import CSV files into R)
- GLIMPSE (function to preview a data frame)
- STR (function to display the structure & data types in a data frame)

I used the following features in **Tableau** to create data visualizations and to derive meaningful insights:

- Dashboards
- Map
- Bar Chart
- Column Chart
- Table
- Interactive Filter (the dashboard included a filter to link the table of data and the map)
- Contextual Filter (the dashbaord included filters to link the bar chart and column charts)

STATEMENT OF THE BUSINESS TASK

The scenario of this case study is that Cyclistic, a bikeshare company offers various models of bikes including classic, docked and electric bikes, with over 600 docking stations in Chicago. The bikes are geotracked. Cyclistic wants to increase profit by increasing their annual memberships. The marketing strategy is to convert existing casual riders to member riders rather than attracting new customers. Therefore my capstone project is focused on trying to understand how annual members and casual riders use Cyclistic bikes differently.

The statement of the business task is as follows: **“Analyse how annual members and casual riders use Cyclistic bikes differently”.**

The problem I am trying to solve is identifying several different ways that member riders and casual riders use the bikes, so that Cyclistic can use these insights to encourage casual riders to convert to member riders, and ultimately increase business profit.

In order to solve this problem, I asked various questions and found answers. The details are explained in this report under the data analysis section. The business questions I asked are as follows:

- Did casual riders or member riders have the highest number of rides for the year 2021, during the weekend and during the weekdays?
- Which start stations are used most frequently by casual riders compared to member riders?
- What are the top 10 most popular bike routes and the top 10 least used bike routes by looking at pairings of start stations and end stations? Where are these bike routes located on a map of Chicago?
- During which seasons did the highest number of bike rides occurr? During which season did the most popular start stations have the highest number of rides?
- During which days and times were the bikes used most frequently? During which days and times did the most popular start stations have the highest number of rides?
- What types of bicycles (classic, docked or electric bikes) are in the highest demand by riders? At which start stations are these bikes located?
- Did any bike rides incurr overtime charges based on the Cyclistic price plan? What was the Length of Time in Minutes for rides done by casual riders and members?

DESCRIPTION OF DATA SOURCES USED

The data sources I used are as follows:

- I read a news article by NBC news regarding Chicago being ranked as the third most congested city in the USA in 2020. This provided background information about roads congested with vehicle traffic.
- The datasets for this bikeshare case study were provided by Google for the time period January 2021 to December 2021. I also obtained background information from the Divvy bikeshare website. These datasets consisted of 13 CSV files which I uploaded into the SQL Server database, to create 12 tables (for each month in 2021). I then merged the 12 tables into a single data table for the full year 2021.

Citations

- **NBC News (9 March 2021).** *Chicago Ranked Third-Most Congested City in US, Worst Traffic Stretch Nationwide,* https://www.nbcchicago.com/news/local/chicago-ranked-third-most-congested-city-in-us-worst-traffic-stretch-nationwide/2457591/
- **Divvy bikeshare**, https://divvybikes.com

DATA CLEANING & VERIFICATION USING SQL

During the data cleaning and verification process I checked the quality of the data to ensure the data was accurate, complete, consistent and secure.

I used the tools **Microsoft SQL Server** and **Microsft Excel Spreadsheets** to clean the data and to verify the data was cleaned properly. I used SQL as my table had a large volume of data consisting of over 5 million records (rows) for the full year 2021.

The details of the data cleaning and verification I performed are documented in the changelog Excel file titled **Changelog Clean Verify Data SQL.xlsx**. A summary of the data cleaning is in **tab "Clean Data"** and the summary of the verifications are in **tab "Verify Cleaned Data**". Also various tabs contain screenshots of the SQL queries executed and the results of these queries.

Also the SQL query files for data cleaning and verification are uploaded to **Github** as listed below:

1A_1C_SQL_FullYear2021_Create

1D_SQL_Verify_FullYear2021

1E_SQL_Verify_Data_Types

2A_2B_SQL_Data_Type_SmallDateTime

2C_SQL_Verify_Data_Type_SmallDateTime

3A_3C_SQL_Duplicate_Rows

3D_3E_SQL_Verify_Duplicate_Rows

4A_SQL_Spaces_Remove

5A_SQL_NewTable_Create_Version3

5B_5C_SQL_Delete_Rides_Not_Started

5D_SQL_Verify_Rides_Not_Started

6A_SQL_NULL

6B_SQL_Verify_NULL_Included

6C_SQL_Verify_NULL_Excluded

7A_7B_SQL_Incorrect_DateTime

The data cleaning and verification completed are as follows:

1) Create a Single Merged Dataset in SQL Server

- The Cyclistic data was contained in 13 CSV files for the 12 months January 2021 upto December 2021. I imported these CSV files into SQL Server and merged the data into a single table to show the data for the full year 2021. I also reviewed the data mapping and changed the data types for the fields. For example applying the data type datetime for the columns started_at and ended_at
- *Verification was done by counting the number of rows in each CSV file imported. This was then compared to the count of total rows in the merged data table for the full year 2021 in SQL Server which is 5,595,063. Also verification of the data types for each fields was done by running the SQL Procedure SP_HELP.*

2) Ensure a Consistent Data Type & Format for Fields Containing Dates & Times

- The fields started_at and ended at had their data type changed from DateTime to SmallDateTime and to format yyyy-mm-dd-hh-mm-ss. This ensured a consistent date and time format. Also a new table version was created titled **full_year_2021_ver2**, and the fields were re-named as started_datetime and ended_datetime.
- *Verification was done by splitting the fields to show the date in a separate column from the time.*

3) Identify Duplicate Records

- 12 duplicate records were identified based on the criteria of searching for duplicates in a single field which was ride_id. But these 12 records had different data in their fields. Therefore these 12 records were not considered duplicates.
- 200,678 duplicate records were identified based on the criteria of searching for duplicates in a combination of multiple fields. But these records had different ride_id numbers. I made the assumption that these rides may have been taken by pairs of people or groups of people travelling together at the same date and time and the same start stations and end stations. Therefore these records are not considered duplicates.
- *Verification that duplicate records did not exist was done by using DISTINCT to find 5,595,051 unique records in the field ride_id. The remaining 12 records were previously identified as not being duplicates. Also verificaton was done to find 5,595,063 unique records based on multiple columns (all fields) in the table.*

4) Identify Leading or Trailing Spaces in Fields with Data Type String

- A count of distinct values in a field including any leading or trailing spaces was done, and then compared to a count of the same field excluding spaces. The result was that the count number was the same in both queries and therefore confirmed that the data in the field did not have leading or trailing spaces. This check was performed on all fields with data type string and confirmed there were no leading or trailing spaces.

5) Delete Records which had the Identical Data in the Fields Start_DateTime and End_DateTime

- 51,074 records were deleted because the date and time for the start of a ride was identical to the end of a ride. This revealed that the ride was never started by the rider. A filter using WHERE identified these records. Also a new version of the table was saved as **full_year_2021_version3.**
- *Verification was done to confirm that the rides previously identified as having the identical date & time for the start of a ride and end of a ride were identified correctly. Verification was done by using a Self Join based on an Inner Join.*

6) Identify NULL Values

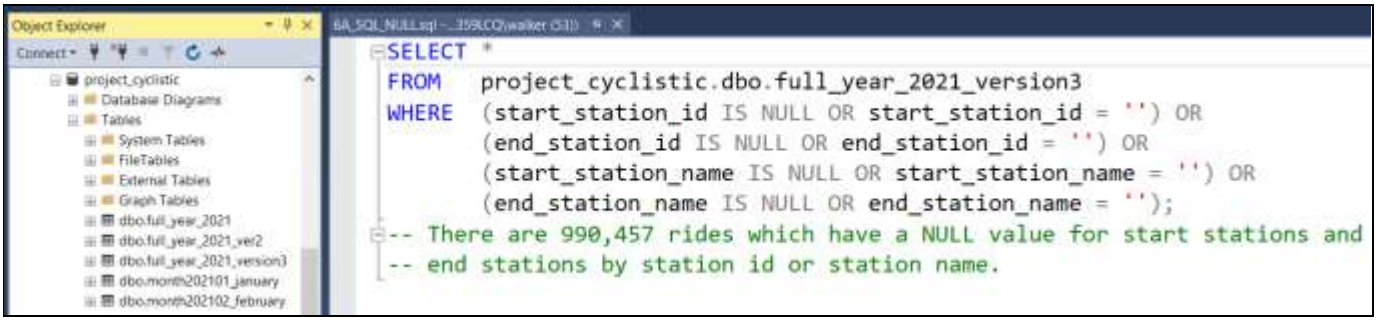
- There are 990,457 rides with a NULL value for their start stations and end stations by station id or station name. A filter using WHERE was used to identify these records. There is no additional information available to enable me to update these NULL values. Also the records with NULL values were not deleted as the other fields for each record contained relevant information. Therefore the NULL values will be excluded by filter for any future data analysis.
- *Verification was done to confirm the 990,457 records including NULL values were identified correctly during the data cleaning process. COALESCE and a Temporary Table were used for this verification. Also the remaining 4,553,532 rides exlcuding NULL values were confirmed to be correct.*

7) Identify Length of Ride Times which were Negative Minutes or Greater than One Day

- There are 86 rides identified as having a negative time for their length of ride time in minutes. This was because the time at the start of a ride was later than the time at the end of a ride. Also 4,015 rides were identified as having a length of ride time greater than 1 day. I made the assumption that this occurred due to a data error or because the bike was not docked properly at the end of a ride. These records will

not be deleted as there is insufficient information. Also the data in the other fields for each record is needed for the data analysis of the rides. Instead the inaccurate data will be excluded by filter for any data analysis.

- *Verification was done to confirm the 86 rides and the 4,015 rides were identified correctly. This verification was done on Excel using pivot tables with calculated fields, conditional formatting, and the functions IF and VLookup.*



DATA ANALYSIS USING SQL

I chose to use the tools **Microsoft SQL Server** and **R Studio** to perform the data analysis

The details of the data analysis completed is documented in the log files titled **LOG- DATA ANALYSIS-SQL-File1.xlsx** and **LOG- DATA ANALYSIS-SQL-File2.xlsx** A summary of the data analysis is in **tab "Data Analysis"**. Also various tabs contain screenshots of the SQL queries executed and the query results. The SQL query files for data analysis have also been uploaded to **Github** as listed below:

A1_A3_SQL_CasualRider_Member

A4_SQL_Member

B1_SQL_StartStation_Highest_Rides

B2_B3_SQL_StartStations_UsedBy_Casuals_Members

C1_C5_SQL_Bike_Routes_Popular

D1_SQL_Month

D2_D3_SQL_Month_StartStation

E1_E5_SQL_Day_Time_StartStation

F1_SQL_Bicycles_InDemand

F2_SQL_Bicycles_InDemand_StartStationId

G1_G3_SQL_Length_of_Ride_Time_Casuals

G4_G5_SQL_Length_of_Ride_Time_Members

Various SQL queries were run for the data analysis. These insights helped to answer my business questions listed below:

A) Did casual riders or member riders have the highest number of rides for the year 2021, during the weekend and during the weekdays?

- The SQL queries reveal that in 2021 member riders did over 3 million rides which was more than the casual riders who did 2.5 million rides.
- However, casual riders did 53% of the rides during the weekend and Friday in comparison to member riders who did only 47%
- Member riders did 62% of the rides during the weekdays (excluding Friday) in comparison to casual riders who did only 38%

B) Which start stations are used most frequently by casual riders compared to member riders?

- The top 10 most frequently used start stations by casual riders included the stations Streeter Dr & Grand Ave, Michigan Ave & Oak St and Millemium Park. Google Maps displays these stations are located close to scenic areas and tourist sites like Lake Michigan, Museum of Contemporary Art, Millenium Park, Shedd Aquarium, Theater on the Lake & the Chicago Theatre.
- The top 10 most frequently used start stations by member riders included the stations Clinton St & Madison St, and Clark St & Elm St. Google Maps displays these stations are located further inland and closer to the Chicago city centre. However, these station locations may not be convenient for riders.

C) What are the top 10 most popular bike routes and the top 10 least used bike routes by looking at pairings of start stations and end stations? Where are these bike routes located on a map of Chicago?

- The top 10 most popular bike routes included from start station Streeter Dr & Grand Ave to end station Millennium Park, and from Shedd Aquarium to Streeter Dr & Grand Ave. Another popular bike route was from University Ave & 57th Street to Ellis Ave & 60th Street. Google Maps shows these routes are located close to parks, Lake Michigan and the University of Chicago.
- The top 10 least used bike routes are located inland and close to expressways. For example the route from start station Racine Ave & Belmont Ave to end station Damen Ave & Division St. Another example is Ada St & Washington Blvd to Honore St & Division St.
- The insights revealed by my data analysis are supported by a news article by NBC news (*website link is listed under Citations*) which explained that the worst vehicle traffic congestion in Chicago occurred on Edens Expressway and Eisenhower Expressway.
- My cycle routes analysis excludes stations with NULL values and station pairings with the identical start station and end station.

D) During which seasons did the highest number of bike rides occurr? During which season did the most popular start stations have the highest number of rides?

- The Summer season in Chicago (June- Sep) had the highest number of rides while the Winter season (Dec- Feb) had the lowest number of rides.
- The most frequently used start stations by casual riders (such as Millenium Park, Streeter Drive & Grand Avenue, and Michigan Avenue & Oak Street) and also by member riders (for example Clark St & Elm St, and Dearborn St & Erie St) had their highest number of rides during the summer season.

E) During which days and times were the bikes used most frequently? During which days and times did the most popular start stations have the highest number of rides?

- The highest number of rides occurred during the weekend and Friday rather than during weekdays (Monday to Thursday). This implies the bikes were used mainly for leisure purposes during the weekend rather than for travel to work purposes on weekdays.
- The time at which the highest number of rides occurred during the weekend and Friday was between 12pm and 6pm
- The time at which the highest number of rides occurred during the weekdays (excluding Friday) was between 7am-9am and 4pm-6pm. This is the peak hour traffic times for commuters to work.
- In addition the most frequently used start stations had the highest number of rides during the weekends & Friday, between the times 12pm & 6pm.

F) What types of bicycles (classic, docked or electric bikes) are in the highest demand by riders? At which start stations are these bikes located?

- The classic bikes and electric bikes are in the highest demand by both casual riders and member riders. The docked bikes are in the least demand.
- The most popular start stations (for example Streeter Dr & Grand Ave, and Wells St & Concord Ln) provide a variety of bike types (classic bikes, electric bikes and docked bikes) but classic bikes are in the highest demand while docked bikes are in the lowest demand.

G) Did any bike rides incurr overtime charges based on the Cyclistic price plan? What was the Length of Time in Minutes for rides done by casual riders and members?

The price plan for Cyclistic is summarised in the table shown below. This pricing information was obtained from the Divvy bikeshare website (*website link is listed under Citations*). The table shows that overtime charges may be incurred depending on the length of time in minutes for a ride, the type of bike used (classic, docked or electric bike), if the rider is a member or casual rider, and the type of price plan chosen by the rider.

	Single Ride	Day Pass	Annual Membership	Membership Ebikes Zone1	Membership Ebikes Zone2	Day Pass Electric Bikes Zone 1	Day Pass Electric Bikes Zone 2
Cost per Ride	$3.30	$0.00		$0 to unlock +$0.15 per minute	$0 to unlock +$0.15 per minute	$3.30 to unlock +$0.20 per minute	$3.30 to unlock +$0.20 per minute
Cost per Day		$15.00					
Cost per Year			$9 per month at $108				
Ebike Zone 1							
Ebike Zone 2							
Overtime charges	**0.15 per minute**	**0.15 per minute**	**0.15 per minute**	**0.15 per minute**	**0.15 per minute**		**0.20 per minute**
Included minutes per ride	30 minutes	180 minutes	45 minutes	None	45 minutes	None	30 minutes
Unlimited Rides	No	Unlimited Rides per Day	Unlimited Rides per Year				
Type of bike included	**Classic bike**	**Classic bike**	**Classic bike**	**Electric Bike**	**Electric Bike**	**Electric Bike**	**Electric Bike**
Dockless bike which does not require a Divvy docking station and bike can be locked anywhere in Chicago within the bike zone	No	No	No	Yes	Yes	Yes	Yes

The data analysis done by SQL queries for **casual riders shows that 311,408 rides incurred overtime charges**. This included the following:

- **8,534 rides by casual riders on classic bikes with a ride length of time greater than 180 minutes.** A Day Pass incurrs overtime charges for rides greater than 180 minutes whereas for a Single Ride Pass it must be greater than 30 minutes. The overtime charge is $0.15 per minute. An assumption was made in this SQL query that these casual rides were done using only Day Passes (as the dataset does not identify Single Ride Passes). Therefore the query filtered for rides greater than 180 minutes.
- **157,113 rides by casual riders on electric bikes with a ride length of time greater than 30 minutes.** Casual riders of electric bikes in Zone 2 are allowed up to 30 minutes of riding time without incurring overtime charges. The overtime charge is $0.20 per minute. An assumption was made in this query that these rides were done only by Zone 2 riders (and not Zone 1 as the dataset does not identify Zone1 riders)
- **145,761 rides by casual riders on docked bikes with a ride length of time greater than 30 minutes.** The dataset does not provide information about the docked-bikes nor their price plan on the Divvy Bikeshare website so I made the assumption that overtime charges were incurred after 30 minutes.

The data analysis for **member riders shows that 54,535 rides incurred overtime charges**. This included the following:

- **33,364 rides by member riders on classic bikes with a ride length of time greater than 45 minutes.** Riders with an annual membership are allowed up to 45 minutes of a riding time without overtime charges. The overtime charge is $0.15 per minute. This SQL query filter was based on 45 minutes.
- **21,171 rides by member riders on electric bikes with a ride length of time greater than 45 minutes.** Riders with an annual membership in Zone 2 are allowed up to 45 minutes of riding time without overtime charges. The overtime charge is $0.15 per minute. An assumption was made in this query that these rides were done only by Zone 2 riders (and not Zone 1 as the dataset does not identify Zone1 riders). The SQL query filter was based on 45 minutes.

DATA ANALYSIS & VISUALIZATION USING R PROGRAMMING

Some of the SQL query code used for the data analysis was long and complicated with nested queries. Therefore R Programming was also used for data analysis and data visualization. This demonstrated that the results obtained in SQL can also be achieved in R but by using shorter programming code.

The data wrangling and data cleaning was completed using SQL and then exported as a CSV file titled sqlexport.csv to my desktop computer. This large CSV file contained the dataset from the SQL Server table titled **"full_year_2021_V3"** and therefore was saved as several smaller size files in the Github folder **SQL Cleaned Dataset** The CSV file was imported into RStudio to perform data analysis and data visualization (see example plot below) by using R Markdown.

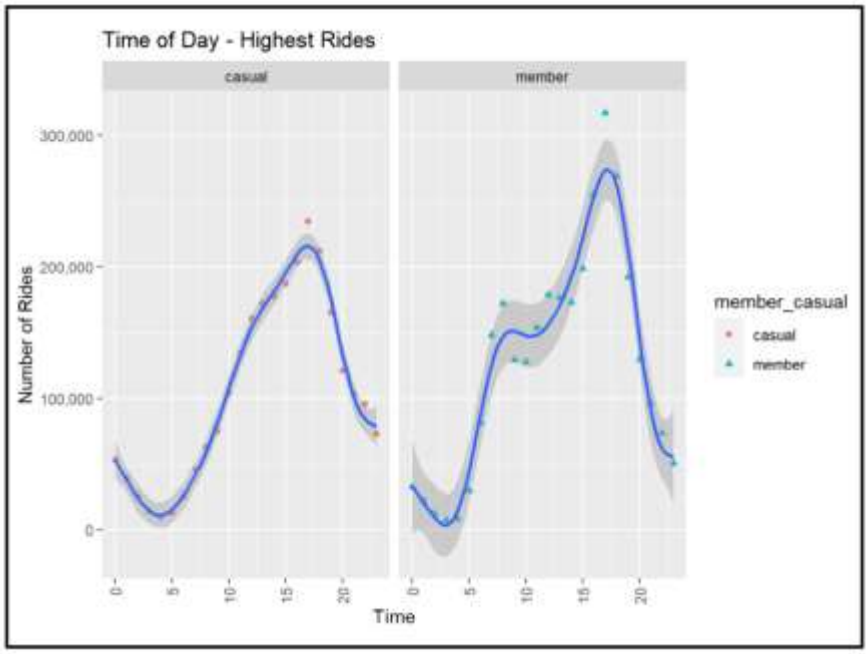
The following R files can be viewed in **Github:**

- PDF report titled **R Markdown Data Analysis.pdf**

- R Markdown Notebook titled **rmarkdown_data_analysis.rmd**

- Chrome HTML document **rmarkdown_data_analysis.html**

- Datasets cleaned in SQL and exported as CSV files to the folder **SQL Cleaned Dataset**



VISUALIZATIONS USING TABLEAU

Tableau is a data visualization tool used for data analytics and business intelligence. Tableau was used for this Cyclistic project to create **two dashboards with interactive filters and contextual filters.** These dashoards can be **viewed in the Tableau website by clicking the file name links below:**

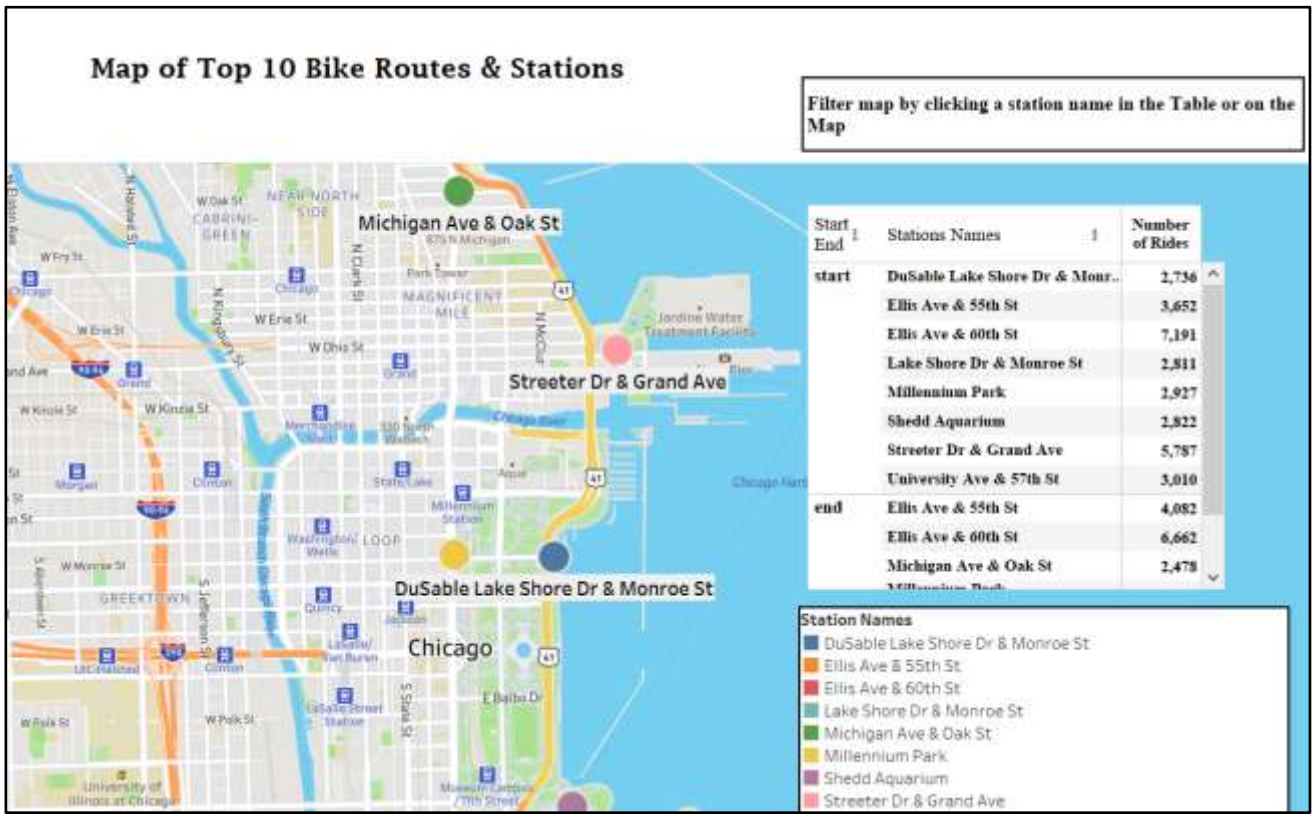
- **Cyclistic Project Dashboard of Bike Rides by Season, Day & Time**
- **Cyclistic Project Map of Top 10 Bike Routes & Stations**

The first dashbboard displayed **three charts**: the number of bike rides done each month in 2021, the number of bike rides by day of the week during 2021 and the number of bike rides done by time of day for 2021. This dashboard included a **contextual filter**. If we click on a value inside any of the 3 charts then this will filter & update all charts on this dashboard. For example click on 814,841 in the chart "Rides by Season" so that we can visualize the efffect on the other charts.

This dashboard enabled us to visualize the data analysis and to display the following **key findings:**

- the summer season in Chicago had the highest number of rides
- the highest number of rides done during the weekend were by casual riders whereas for weekdays it was by member riders.
- the time of day during which the highest number of rides occurred was between 12pm and 6pm

The second dashboard included a **map** to display the streets of Chicago including scenic areas (parks, museums, lakes) and universities. A **table** also displayed the number of rides for these stations. The dashboard included an **interactive filter** so that clicking on a station name in the data table would highlight the station on the map. This map displayed the **key finding** that the locations of start stations and end stations for the top ten most popular bike routes were close to scenic routes.



The **datasets** used to create the visualizations in Tableau were obtained from SQL. The original datasets were mereged, cleaned and verified in SQL, exported from SQL, and then imported into Tableau as the following Excel files:

tableau_import_data.xlsx

bike_routesv2.xlsx

TOP THREE RECOMMENDATIONS BASED ON DATA ANALYSIS

My top three recommendations based on my data analysis are as follows:

A) I recommend that Cyclistic should create new cycle routes which are less traffic congested routes and where bike lanes are not near heavy vehicle traffic (such as buses, trucks & cars). Cycle routes closer to scenic areas like parks, lakes, beaches and cafes may take riders a longer time to travel to work office locations. But the scenic bike rides would be more enjoyable and would also have the health benefit of being away from traffic vehicle pollution. This may encourage casual riders to change to membership riders and to use their bikes to travel to work along scenic routes. My recommendation is supported by my data analysis which revealed that the top 10 most popular bike routes were located close to scenic areas whereas the least popular routes were located inland and close to expressways.

B) I believe that Cyclistic has the potential to increase revenue by encouraging riders to use bikes on weekdays for travel to work purposes (instead of for leisure on weekends). I recommend this be done by providing discounts on annual membership prices, such as increasing the included minutes per ride from 45 minutes to 1 hour, if bikes are used during the peak hour times of 7am -9am and 4pm-6pm on weekdays. This may encourage casual riders to convert to member riders and to use their bikes to travel to work during weekdays. My recommendation is supported by my data analysis which revealed that the highest number of rides occurred during the weekend and Friday rather than during weekdays (Monday to Thursday). This implies the bikes were used mainly for leisure purposes during the weekend rather than for travel to work purposes on weekdays.

C) I recommend that Cyclistic should invest in purchasing more electric bikes. The classic bikes have a disadvantage because they must be docked at a station to end a ride. This is not convenient for riders if their workplace location is not near a docking station. An electric bike has the advantage that it can be docked at a station or the rider can use a cable lock to lock the bike to any public bike rack or signpost. This cable lock option for electric bikes was mentioned on the Divvy bikeshare website (see). Therefore providing more electric bikes that can be locked closer to workplace locations may encourage casual riders to convert to member riders and to use their bikes for commuting to work. My recommendation is supported by my data analysis. This revealed that classic bikes and electric bikes are in the highest demand by both casual riders and member riders, whereas the docked bikes are in least demand.