Data Processing and Preparation

In this document, each and every step of data processing and preparation is described.

1) The dataset for the report consists of 12 different worksheets and each worksheets is not very big. Hence, Excel was sufficient for data manipulation and processing.
2) Cleared the variables which were not important for the analysis. For example the columns which shows the location of each station.
3) Added a few other columns in the dataset.
   a) Ride Length: Subtracted the values of the "ended_at" column with the "started_at" column. The column would tell how long the user used the bike for the trip. This would help in distinguishing the purpose of using the bike. Changed the variable to numeric format.
   b) Weekday: Used WEEKDAY() function on the "started_at" column. This returns the on the day of the week on which the trip began. This would help in identifying on which day of the week the bikes are being used the most.
4) Followed these steps for other months' datasets.
5) Saved these edited datasets in another folder so that datasets do not get mixed with an original.
6) After editing datasets, transferred them to BigQuery.Combined them using FULL JOIN combined.
7) Now there is a dataset with combined data from all 12 months.
8) Filtered out the rows for which the ride length value was 0. This would help in reducing the size of dataset and help in getting more accurate result.
9) Transferred the filtered to Tableau for Data Visualisation.

Analysis

In this document, the analysis part of the projet is described.

1) After all the filtering of the columns and addition of other variables in each worksheet, Pivot table was used to find patterns in the dataset. Here 2017 worksheet is used to describe the steps.

2)

Row Labels	Count of ride_id
casual	442056
member	380354
Grand Total	822410

The first Pivot Table shows that the number of Casual customers is more than Member customers.

3)

Row Labels	Average of ride_length
casual	0.022770953
member	0.009888415
Grand Total	0.016812946

The pivot table shows Causal users tend to have longer rides than the Members users.

4)

Count of ride_id	Column Labels		
Row Labels	casual	member	Grand Total
1	69744	39746	109490
2	48316	46503	94819
3	45403	52383	97786
4	46552	53540	100092
5	56846	64429	121275
6	71968	64220	136188
7	103227	59533	162760
Grand Total	442056	380354	822410

The pivot table shows the number of rides completed by the users throughout the week. This shows that Causal users tend to have the more rides completed during the weekend whereas the Member have completed more rides during weekdays.

5)

	Average of ride_	Column Labels		
	Row Labels	casual	member	Grand Total
1		0.025554018	0.01115	0.02032478
2		0.025466178	0.00982	0.01779347
3		0.019501669	0.00935	0.01406134
4		0.019867517	0.00926	0.0141926
5		0.02122652	0.00934	0.01491113
6		0.021319761	0.00964	0.01581264
7		0.024238647	0.011	0.01939793
Grand Total		0.022770953	0.00989	0.01681295

The pivot table shows the average ride length completed by the users throughout the week. This shows that Causal users tend to have the longer rides during the weekend whereas the Member have completed have longer rides during weekdays.

6)

Count of ride_id	Column Labels		
Row Labels	casual	member	Grand Total
classic_bike	241489	265420	506909
docked_bike	57698		57698
electric_bike	142869	114934	257803
Grand Total	442056	380354	822410

The pivot table shows that both types of customers have similar behaviour with the type of bikes but the docked bikes are only used by casual bikes.

7) The findings indicates that the Casual users use the app for the leisure activities and the Member users use it for daily commute activity.

8) This analysis is repeated on other workbooks. This establish overview of the behaviour of different types of customers throughout the year.