**Background**

The Pronto Cycle Share system consists of 500 bikes and 54 stations located across Seattle.

**Problem statement**

The company has noticed that there is a drop of earning of the bicycle-sharing business and increasing number of complaints that there are not enough bicycles for usage.

**Hypothesis**

* **Does users have preferred location for end and starting locations**
* **Are any timings that affects the usage rate?**

**Data**

The dataset adopted provides an overview of the bicycle-sharing service in Seattle over the past three months. With this, we will be able to identify the stations and timings which are in higher demand and enable an ease in locating the movements of bicycles. We will therefore, be able to quickly redeploy and cater sufficient bicycles to appropriately leverage on the demanding needs.

Studying the usage of trends of ‘Usertype’ will also aid the business in coming up with attractive subscription plans to incentivise short-term pass holders and therefore translate them into earnings through a membership-tailored package.

**Data Type**

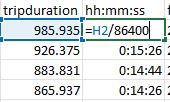
In the data preparation, there were multiple instances of blank data, which does not contain useful data in multiple files. In order not to affect the overall results, the affected rows were deleted.

There were erroneous duplicative value and these were also omitted.

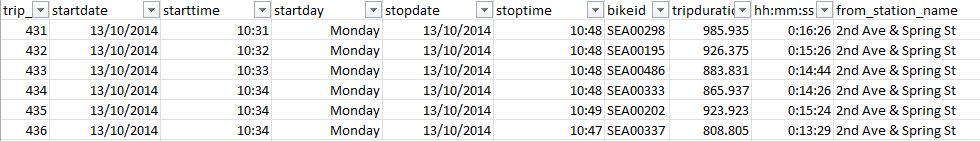
The available dataset had the values of starttime and stoptime condensed into “dd/mm/yyyy hh:mm” format. Therefore, using “text to columns” function, we subdivided them into “startdate”, “starttime”, “stopdate” and “stoptime” fields.



Additionally, by assigning a weekday function (=TEXT(B2,"dddd")) value to the “startdate”, and formatting it to display the weekday, we can further see weekday usage patterns later on.



The “tripduration” field was captured in a per second format. As such, for a simpler view, we can divide the cell value by /86400 and display the timing in hh:mm:ss.



**Data Dictionary**

**Worksheets**

|  |  |  |
| --- | --- | --- |
| **Worksheet** | **Description** | **Remarks** |
| **Original** | **Original Data from Kaggle – Cycle Share Dataset -**https://www.kaggle.com/datasets/pronto/cycle-share-dataset |  |
| **Trip** | **Cleanup Data from ‘Original’ File** | 1. Cleanup blank data in multiple columns as data is not necessary and won’t affect the outcome 2. Did conditional formatting to ensure there is no duplicative value and omitted those values. 3. Did ‘Text to Column’ function of ‘starttime’ and ‘stoptime’ to condense from ‘dd/mm/yyyy hh:mm’ format into dd/mm/yyyy and ‘hh:mm’. Stored those data into 4 columns – ‘startdate’, ‘starttime’, ‘stopdate’ and ‘stoptime’. 4. Convert time duration of ‘tripduration’ of seconds into ‘hh:mm:ss’ (Created a new column – ‘hh:mm:ss’) |
| **Popular Places by Day** | **Pivot and bar graph chart of popular places by day** | Create a pivot and bar graph chart to see the most popular location by day of the week from ‘startday’, ‘from\_station\_name’ and count of ‘ from\_station\_name’ |
| **Peak Usage Timings** | **Pivot and bar graph chart of Peak Usage Timing** | Create a pivot and bar graph chart to see the peak usage timings from ‘Hours’, ‘starttime’, ‘from\_station\_name’ and count of ‘from\_station\_name’. |
| **8am (Top 10)** | **Pivot and bar graph chart of Top 10 locations during 8am timing** | Create a pivot and bar graph chart to see the ‘Top 10 locations during 8am timing’ from ‘Hours’, ‘from\_station\_name’ and count of ‘from\_station\_name’. |
| **8am (top 10 – Popular Routes)** | **Pivot and bar graph chart of Top 10 Most Popular Routes during 8am timing** | Create a pivot and bar graph chart to see the ‘Top 10 locations during 8am timing’ from ‘Hours’, ‘from\_station\_name’ and ‘to\_station\_name’ and count ‘from\_station\_name’ |
| **5pm (top 10)** | **Pivot and bar graph chart of Top 10 Most Popular Routes during 5pm timing** | Create a pivot and bar graph chart to see the ‘Top 10 locations during 5pm timing’ from ‘Hours’, ‘from\_station\_name’ and count of ‘from\_station\_name’ |
| **5pm (top 10 – Popular Routes)** | **Pivot and bar graph chart of Top 10 Most Popular Routes during 5pm timing** | Create a pivot and bar graph chart to see the ‘Top 10 locations during 5pm timing’ from ‘Hours’, ‘from\_station\_name’ and ‘to\_station\_name’ and count ‘from\_station\_name’ |
| **MembershipTrend** | **Pivot and bar graph chart of Membership Trend** | Create a pivot and bar graph chart to see the trend of usage of different user types based of ‘Dayofweek’, ‘Hours’ and count of ‘from\_station\_name’ |
| **MemberType** | **Pivot and bar graph chart of Member Type** | Create a pivot and bar graph chart to see the trend of usage of different user types based on Count of ‘usertype’ |
| **Dashboard** | **Dashboard of all the various charts** |  |

**Original Worksheet**

|  |  |  |
| --- | --- | --- |
| **Data Field** | **Description** | **Remarks** |
| **trip\_id** | Transaction number of the bicycle trip |  |
| **starttime** | Start time of trip | Format is in dd/mm/yyyy hh:mm |
| **stoptime** | End time of trip | Format is in dd/mm/yyyy hh:mm |
| **tripduration** | Duration of trip | Format is in Seconds |
| **bikeid** | ID of bike |  |
| **from\_station\_name** | Start location of trip |  |
| **to\_station\_name** | End location of trip |  |
| **from\_station\_id** | ID of Start location |  |
| **to\_station\_id** | ID of End location |  |
| **usertype** | Member type |  |
| **gender** | Gender of user |  |
| **birthyear** | Birth year of user |  |

**Trip Worksheet**

|  |  |  |
| --- | --- | --- |
| **Data Field** | **Description** | **Remarks** |
| **trip\_id** | Transaction number of the bicycle trip |  |
| **startdate** | Start date of trip |  |
| **starttime** | Day and time of trip started |  |
| **startday** | Day of trip | =weekday function to convert startdate into day of week. Number format |
| **dayofweek** | Day of week | =Vlookup function of ‘startday’ against ‘Day of Week’ worksheet to churn out the day of week |
| **stopdate** | End date of trip |  |
| **stoptime** | End time of trip |  |
| **bikeid** | ID number of bicycle |  |
| **tripduration** | Duration of trip | In per seconds format |
| **hh:mm:ss** | Duration of trip | Convert ‘tripduration’ from per seconds into Hours/Mins/Secs |
| **from\_station\_name** | Start destination of trip |  |
| **to\_station\_name** | End destination of trip |  |
| **from\_station\_id** | ID number of Start station |  |
| **to\_station\_id** | ID number of End station |  |
| **usertype** | Membership type of user |  |
| **gender** | Gender of user |  |
| **birthyear** | Birth year of user |  |

**Day of Week Worksheet**

|  |  |  |
| --- | --- | --- |
| **Data Field** | **Description** | **Remarks** |
| **Code** | Code for day of week | Field is created to match the ‘dayofweek’ field in ‘Trip’ worksheet |
| **Day of week** | Day of week |  |

**Analysis**

Clustering analysis was adopted in solving the problem statement. As categorical data is involved, segmenting customers based on similarity can aid in the optimization of bicycle availability and the development of subscription plans for targeted consumers.

We were able to identify the higher demand for bicycles at the following times of day:

|  |  |  |
| --- | --- | --- |
| **Weekdays** | Monday | 8am and 5pm |
| Tuesday |
| Wednesday |
| Thursday |
| Friday |
| **Weekends** | Saturday | 2pm and 3pm |
| Sunday | 1pm and 3pm |

In order to focus on monetizing the subscription plan, we need to also do a bicycles redeployment by shifting the bicycles at identified off-peak locations to meet the demands during peak periods.

Summit Ave. E. is the most popular starting point and yet the least popular destination. This implies that there will be a shortfall of bicycles after the morning peak which requires redeployment before the next peak. Similarly, both stations along 12th Ave followed the same pattern and will be redeployed.

Pier 69 station would not necessitate any actions because consumers begin and end their journey at the same location. Similarly, with Dexter and eBlaine stations, customers travel bidirectionally. Therefore, neither station requires any further action.

At King Street station, however, fewer customers are returning to meet the demand at 8 a.m. Hence, the redeployment of bicycles is required.