**Sai Mounika Tadaka**

**Z1784462**

**Project 2**

**Divvy Bikes – Trips – Analysis – Visualization**

**Group B:**

**Sai Mounika Tadaka**

**Ahil Nishanthan Gunasekaran**

**Deepak Kotrike**

**Initial Steps:**

After having a look at the given data from divvy, we had a lot of discussion of what we need to do or what can we do to analyze and visualize the data given so that it will be handy to get a look into how the Divvy bikes are performing.

So we decided analyze the data for every year, 2013 and 2014 separately and create a dashboard for the respective years. We thought this will be an effective way to keep track of how the business progress and in turn businesses tend to view review their yearly reports and prepare themselves for the upcoming years.

**What should be present in the dashboard:**

After a lot of discussion, we came up with a few ideas of what can we do to best represent these data visually providing a complete performance year round and below are the ideas,

* Total fleet, trips and duration
* Subscriber Age group and Gender
* Customer Subscriber Ratio
* Top Destinations of the year
* Most frequent time - Month, Day and Hour
* Most used bikes
* Busiest Routes
* Where are the stations located, their fleet count and the day the station was opened?

2013:

The data for 2013 was pretty straight forward. But it had the data only for the final two quarters of the year because of which we will not be able to get a complete picture of the divvy bikes performances in 2013. We didn’t have much of data manipulation work to be done for this since the given data file consists of all the info we needed.

2014:

The data for 2014 was also pretty straight forward apart from the fact that the data was separated into different files based on Quarters and even months. We decided to merge all these files together which we felt would make it easier to make an analysis for the year 2014. The only problem we had here is the verification of the data once the files are merged. Since the merged file was too huge, and excel doesn’t allow to open a file with many millions of rows we had a hard time verifying the data. But finally we decide to move on with the visualization since if there were any discrepancies in the data it would reflect in the visualization.

We used “R” to merge all these files.

**Data Visualization:**

Once we are ready with the data, each member in our team made themselves responsible for generating the visualization of the ideas we generated.

**Sai Mounika Tadaka:**

I took the task of generating visualization for the below three,

* Customer Subscriber Ratio
* Top Destinations of the year
* Most used bikes

***Customer Subscriber Ratio:***

For this task I need to identify the total number of customers and subscribers for each year, creating two graphs for 2013 and 2014.

As given in the description of the data file, Customers are the ones having the Gender and Birth year available in the data. Such details were not provided for the 24-hour customers. The attribute “Usertype” was helpful in creating this viz.

For representing the ratio, I felt a pie chart would be much clean since we had only two categories, Namely Customers and Subscribers.

This viz was created with the usertype and the number of records for each user type.

***Top Destination of the Year:***

This viz helps to understand and know what are all the most frequent destinations for the users riding the divvy bikes. Also depending on this result, one can easily judge whether a particular station needs more bikes or less.

To Station Name and the Number of Records were used to create this viz and the it was filtered to show the top destinations of the riders.

***Most used Bikes:***

There will be some point where the company plans to invest on more bikes or replace existing bikes. This viz will act as a guide for them to replace the old bikes. Bikes that have travelled a lot or used a lot based on the trip duration is shown in this graph. It is filtered to show the most used bikes.

**Ahil Nishanthan Gunasekaran**

I was responsible for the below tasks,

* Total fleet, trips and duration
* Subscriber Age group and Gender
* Most frequent time - Month, Day and Hour

***Total fleet, trips and duration:***

These are not any kind of graphical viz rather they are simple numbers which provides the summary of the bikes, trips and duration for a given year. This information helps the company to know, what serves as the base for other visualizations. Numbers are all important at the end.

Total fleet – Calculated using Bike id’s - Unique number of bike id’s – dragged and dropped to text label.

Total Trips – Calculated using Trip id’s - Unique number of trip id’s – dragged and dropped to text label.

Total Duration – Converted trip duration to hours (n/3600) – Sum of hours - dragged and dropped to text label.

***Subscriber Age group and Gender:***

This viz provides information on the Age group of people and the gender who are most likely to take a divvy bike.

This graph can represent only the information about the subscribers and not the customers, sine we do not have the required data for the customers.

Calculated Age using DATEDIFF function in calculated field.

Created a stacked bar chart representing the age and the gender of people in that age group.

***Most frequent time – Month, Day and Hour:***

This viz provides information on the time at which the users use the divvy bikes the most. Be it which time of the year, which day of the week or what hour in a day.

This viz was created with the help of trip start time and the number of records.

I used the star time and stop time to find what is the most frequent start and stop time.

I used the Month from the start time to find out the months in which the business reach its peak.

I used the Weekly option in tableau to find out which is the day of the week when the trips are more.

**Deepak Kotrike**

I took the responsibility of the below tasks,

* Busiest Routes
* Where are the stations located, their fleet count and the day the station was opened?

***Busiest Routes:***

This viz is a grouped bar chart providing information on the busiest routes for the given year. The graph was filtered to show the top results.

This graph also helps us to know where do people start their ride and where do they end. On looking into it closely, people generally start their trip at one station and end at the other.

***Where are the stations located, their fleet count and the day the station was opened?***

This is a map visualization providing information on the location of the stations. This viz is created for both 2013 and 2014. There were 301 stations around Chicago by the end of 2014.

The size of each stations (Number of docks) is encoded and represented by the size of the stations (circle) plotted in the graph

Also the colors represent month, the station went public or online. I also used the existing map layer from tableau representing the median Age group of the people living around Chicago which also gives us information on which Age group of people are most likely to use the divvy bikes.

Finally, with all the visualizations created, we put up these visualizations together to create a dashboard for the year 2013 and 2014.