**Story:**

As a CitiBike business, there are two most important elements of my business:

**Optimizing Bike Station Placement**

In a city like New York CitiBike would like to ensure that all the bikes are optimally placed in areas that encourage a high demand from bikers. To tackle this problem, the BI project utilized historical data from CitiBike. By analyzing this dataset, our team identified popular areas and routes, peak demand times, and underutilized stations.

**Customer Segmentation**

What type of users most use the bikes. One of the main things for us is to make sure we attain high user satisfaction, engagement, and loyalty. We operate in a region of high competition, hence this element for our business is severely important. By gathering information such as user demographics, user types, and age groups, our project incorporated user profiling into the analysis of the Citibike dataset. Moreover, based on this User Profiling we would like to target specific customers to further increase the use of bikes, and achieve a higher stream of customers.

We will be solving our Customer Segmentation using our **User Profiling** Dashboard:

A screenshot of a computer

Description automatically generated with medium confidence

We have firstly segmented our Users using their dimensions:

* A picture containing text, screenshot, circle, diagram

  Description automatically generatedGender

We can see a clear Majority of Male, however Female also participate in high trip durations. Unknown gender is a problem for us, we need to refine our collection methods such that percentages of Unknown decreases.

* A picture containing text, screenshot, circle, diagram

  Description automatically generatedAge

We can see that trip duration is very evenly divided between 20–50-Year-olds. These are Customers that are independent, and are financially held, hence CitiBike is a viable option. <20 and 60+ is in minority, showing that our labourous product is not suitable for these ages.

* User Type, total trip duration

A picture containing text, screenshot, circle, graphic design

Description automatically generated

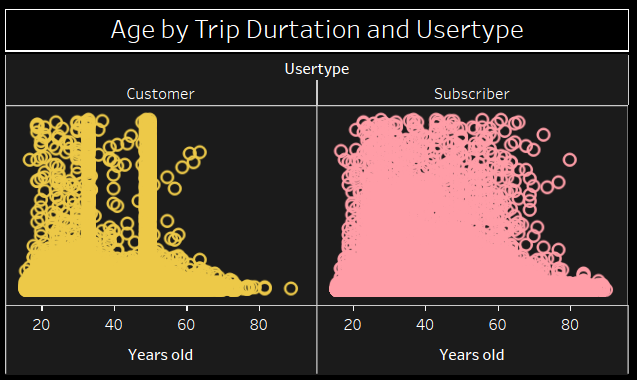
Who are these types? Subscriber are users that have bought a subscription over a certain period of time. These are constants, that will be charged a fixed amount irrespective of their trips. Customers are those that on their way for one time use buy our services.

* User Type, average trip duration

A picture containing text, screenshot, circle, font

Description automatically generated

Interesting find, when we look at the average trip duration, we see that Customer type has a much higher average TD. These are one time users, hence we should devise a package, that sets the price by trip duration. Higher TD, higher price leading to maximizing revenue.

Further analysis of User types led us to this chart:

We can clearly see that our customers are less spread compared to Subscriber. An interesting find shows that for some ages, like 45-50, and 30-35, our customers peak at trip durations. This suggests that our customers that generate higher TD’s, are in this age bracket. On the other hand, customers of ages, 70+ are very scarce. This suggests that most of our customers are tourists, and high age people do not travel much, and due to obvious health reasons avoid biking.

A picture containing text, screenshot, font

Description automatically generated

On the other hand, with Subscribers we generate a steady stream of usage. The chart is well spread and says that for all ages it is generating high TD’s. The frequency of 70+ users is much higher, showing that old people that have a structured schedule, along which they operate use our bikes. This could be for recreational purposes. Since old people are not bounded by any schedules, a subscription helps them to keep their activity levels high.

A screen shot of a graph

Description automatically generated with medium confidence

Next we look at hour our genders are spread by User types.

A screenshot of a computer

Description automatically generated with low confidence

For subscribers we see that male are in majority, however female produce over 800k records, which clearly means that despite the relative small percentage, we should not ignore them as a valuable customer.

A screenshot of a computer

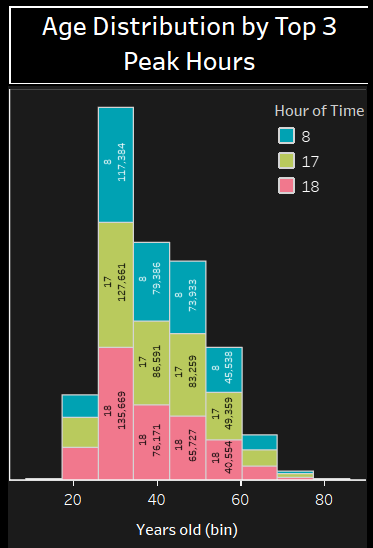
Description automatically generated with low confidence

Next for customers, we have a vert interesting find. All of Gender Unknown, lie in the Customer type of our dataset. Zero instances of Unknown exist in subscriber. Which shows, that for the registration process of our customers, we should generate a form that ensures a reply. We can do this by compulsory fields or adding options to cater for all. This would enrich our data, and allow to grow with better quality of data.

A screenshot of a computer

Description automatically generated with medium confidence

Next we look at our peak hours:



Our top 3 peak hours are:

8 am

5 pm

6 pm

These are very well distributed by age, and we can see majority lies in the age bracket of 25-50, showing most of our users are working people and use our product as a mean of commute.

Lastly a year wise overview of our records:

In 2017 February was our highest performing month. This is an interesting insight, since February has a lower number of days compared to other months yet peaked in 2017. We can further investigate the count of working days to see whether February was a month of higher working days or not.

A picture containing text, screenshot, font, number

Description automatically generatedA picture containing text, screenshot, font, rectangle

Description automatically generated

In 2018 we see March peaking with a high margin. There is a rise in Avg trip duration.

A picture containing text, screenshot, font

Description automatically generatedA picture containing text, screenshot, font, rectangle

Description automatically generated

Lastly in 2019, March retains its lead, while January and February maintain equal levels. There is a drop in the average trip duration from last year.

A picture containing text, screenshot, font, graphics

Description automatically generatedA picture containing text, screenshot, font, rectangle

Description automatically generated

January is consistently to the lower end, this can be due to the rough weather conditions in New York in January.

In conclusion these are the factors employed that allow us to design User Profiles.

We will be solving our Optimizing Bike Station Placement using our **Trip Profiling** Dashboard:

A screenshot of a computer

Description automatically generated