# Divvy Bike Data Exploration

Presented by: Tina Carr

# Overview

## Purpose

Divvy is Chicago's bike sharing system which is owned by the Chicago Department of Transportation with the aim to "promote economic recovery, reduce traffic congestion and improve air quality".

With over 6,000 bikes and 580 docking stations in the Chicagoland / Evanston area that are available to the public 24/7 the higher the accuracy rate Divvy's forecasting models can achieve translates into more happy customers looking for a sustainable and fun way to get around the city.

Therefore, this capstone project will use forecasting models with the aim of creating a daily forecast for the the Lake Shore Dr & Monroe St Divvy station.

# Overview

## Data

The source of the data can be found here and was provided by the City of Chicago

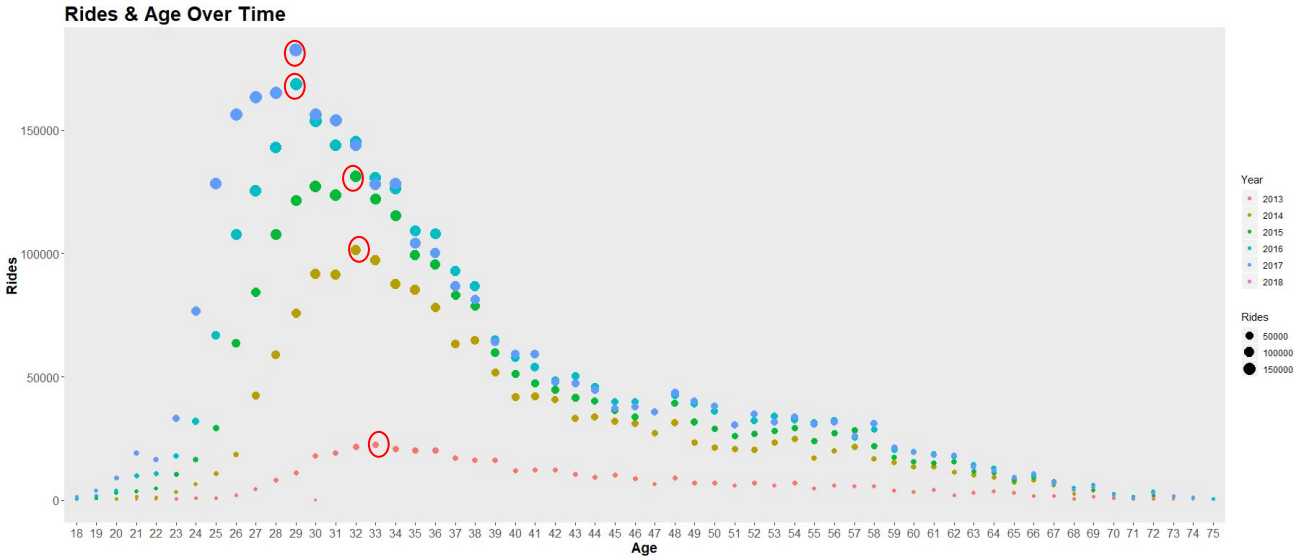
This data set has a total of 13,821,994 observations and 22 columns including:

```
> variable.names(Dataframe_Divvy)
 [1] "1..TRIP.ID"          "START.TIME"          "STOP.TIME"           "BIKE.ID"                "TRIP.DURATION"
 [6] "FROM.STATION.ID"     "FROM.STATION.NAME"   "TO.STATION.ID"       "TO.STATION.NAME"        "USER.TYPE"
[11] "GENDER"              "BIRTH.YEAR"          "FROM.LATITUDE"       "FROM.LONGITUDE"         "FROM.LOCATION"
[16] "TO.LATITUDE"         "TO.LONGITUDE"        "TO.LOCATION"         "Boundaries...ZIP.Codes" "Zip.Codes"
[21] "Community.Areas"     "Wards"
```

For forecasting purposes there are 2 key columns for every station: Bikes brought TO and FROM the station.

I have chosen the Lake Shore Dr & Monroe St Divvy station to forecast since it is the station the most bikes are taken from and the 2nd busiest station customers return bikes to.

# Overview
## Exploratory Data Analysis




*Data was cleaned to represent the 18-75 age range only

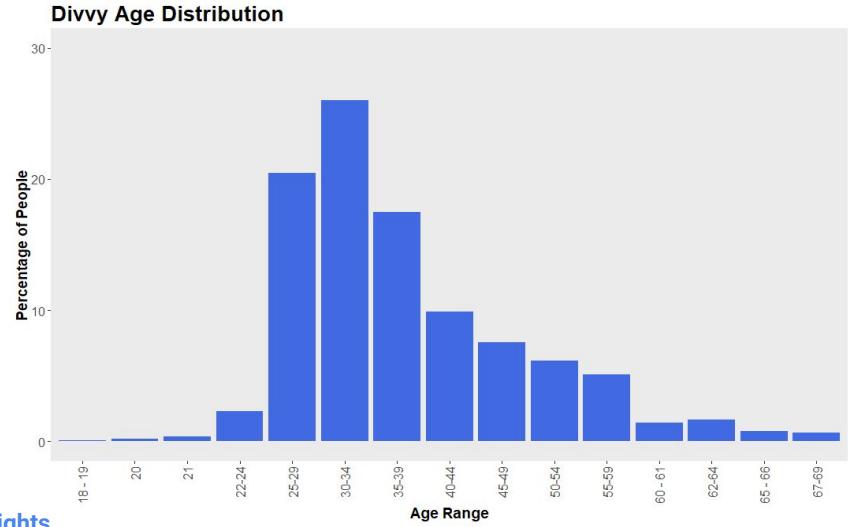
**Insights**

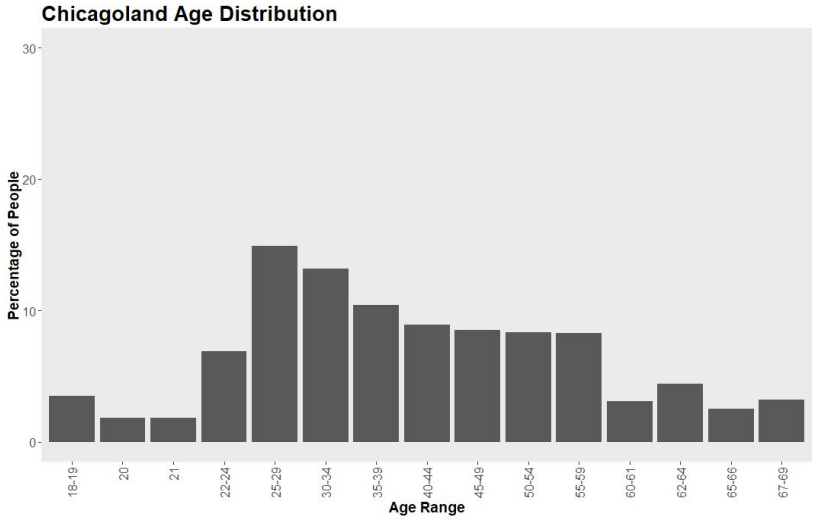
- We can see that over time the “peak” age for riders has shifted from 33 in 2013 to around 28 in 2018 and
- According to the US Census 18% of the Chicagoland population lies within the 20-29 range and 17% within the 30-39 range

**Recommendation**

- Divvy seems to capture the late 20s / early 30’s demographic quite well however based on US Census data there is a larger untapped segment in the early 20’s demographic
- Divvy could improve this early 20’s demographic by targeting marketing around universities and work with university orientation leaders to introduce their service

# Overview
## Exploratory Data Analysis

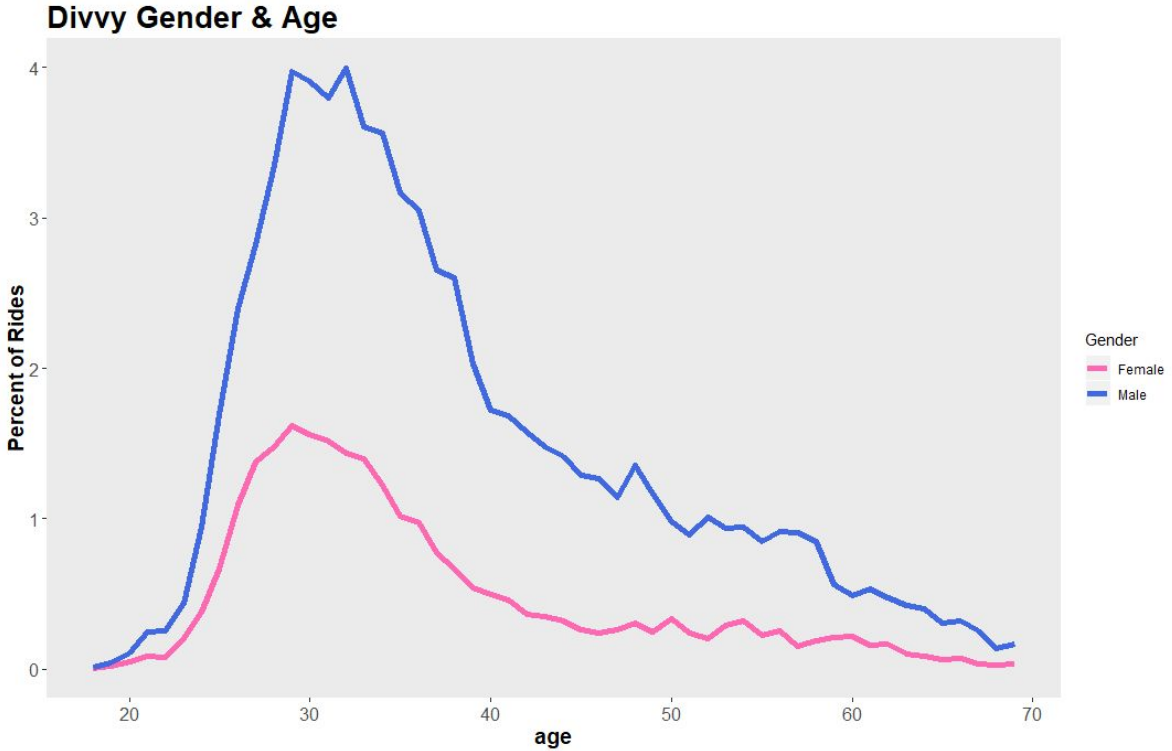



**Insights**

- The Divvy age distribution is skewed to the left compared to the overall population in Chicago which demonstrates there is untapped potential in the mid-age market segment
- As seen in the previous slide there is also potential in the 18-24 market
- Divvy age standard deviation = 10.4, Chicago age standard deviation = 14.2
- The wider spread of the Chicago standard deviation could in part be explained by the fact older people which are considered part of the Chicago population are not interested or physically unable to bike ride thus preventing Divvy bike from capturing this segment.
- In any case taking standard deviation into consideration if Divvy targets new age segments for which their customer base is currently much lower compared to the Chicago population the Divvy standard deviation will naturally increase towards the Chicago standard deviation

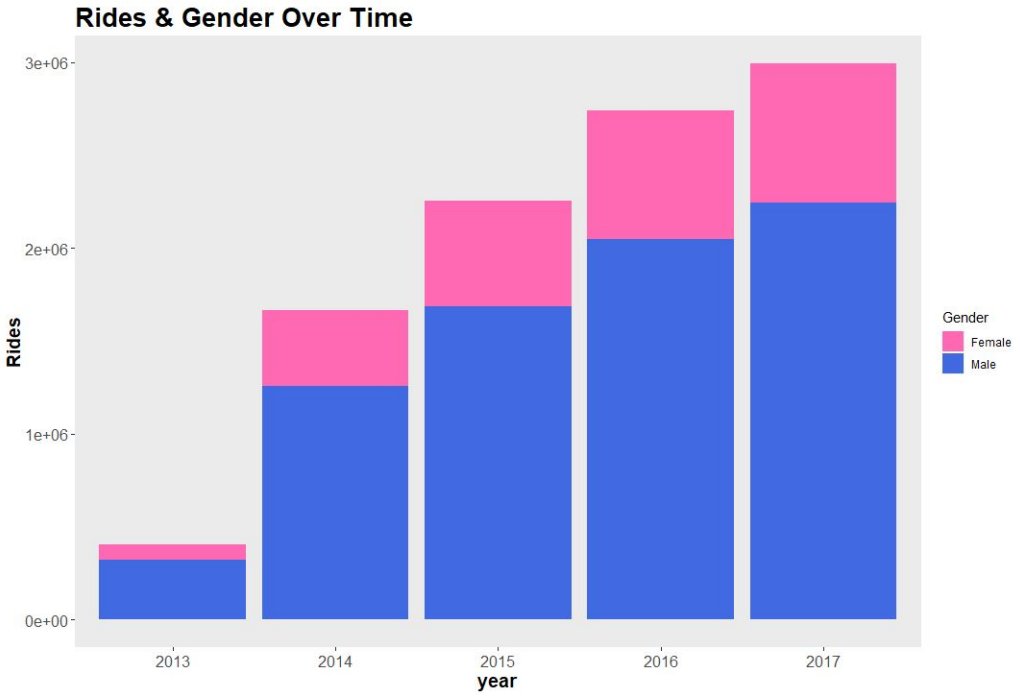
# Overview
## Exploratory Data Analysis




**Insights**

- The male and female data experiences similar peaks and dips with some exceptions
- A peak in male an female riders happens around the late 20's - however males experiences another peak in the early 30's
- There is a peak in male riders in their late 40's while female riders experience a dip in that age - a similar instance happens in late 50's

# Overview
## Exploratory Data Analysis




*not all customers revealed their gender

**Male / Female Trip Ratio:** 7,552,990 / 2,495,777*

### Insights

- Divvy was able to increase its female riders trips from only 21% in 2013 to 25% in 2017
- However, the US Census reports Chicago's female population at about 51%

### Recommendations

- Divvy should conduct customer research on marketing their product to women since this segment is massively underrepresented in their customer base compared to the overall population of Chicago

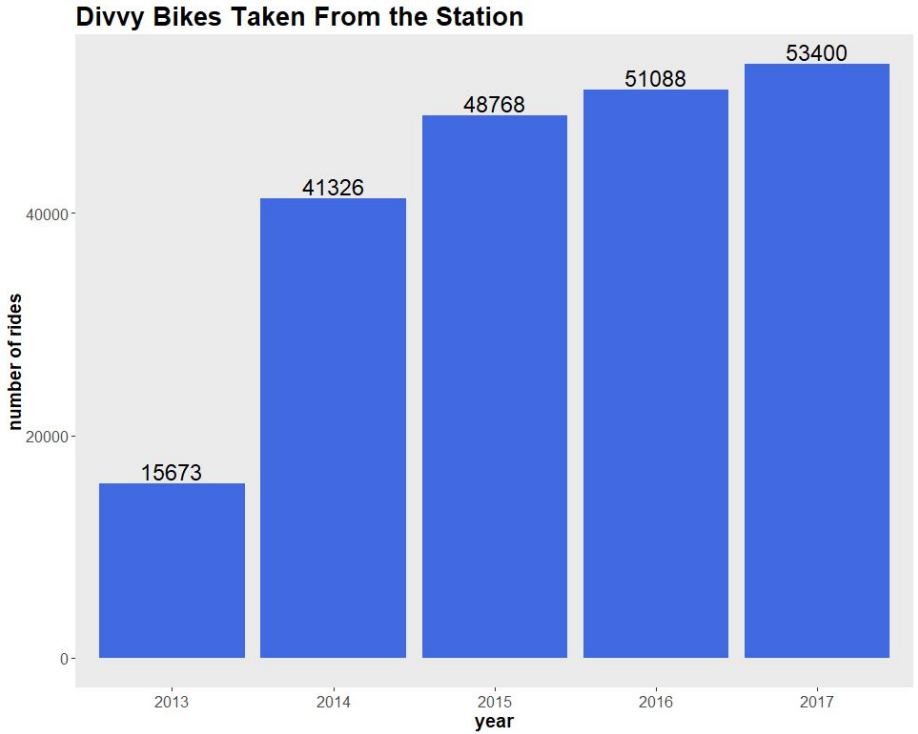
# Overview

## Data: Lake Shore Dr & Monroe Station

The graph to the right is an overview of the number of bikes taken from the Lake Shore Dr. & Monroe Divvy Station

**Observation Dates:** Jun 27 2013 - December 28 2017
**Total Observations:** 210,255

- Divvy Bike began operations in June 2013
- The significant increase from 2013 - 2014 can in part be explained by the fact the 2013 only covered 6 months of the year while 2014 - 2017 data was recorded for a full 12 month calendar year
- We can see a positive trend of rides as each year passes which is natural for a brand just starting out as more and more people become aware of the brand



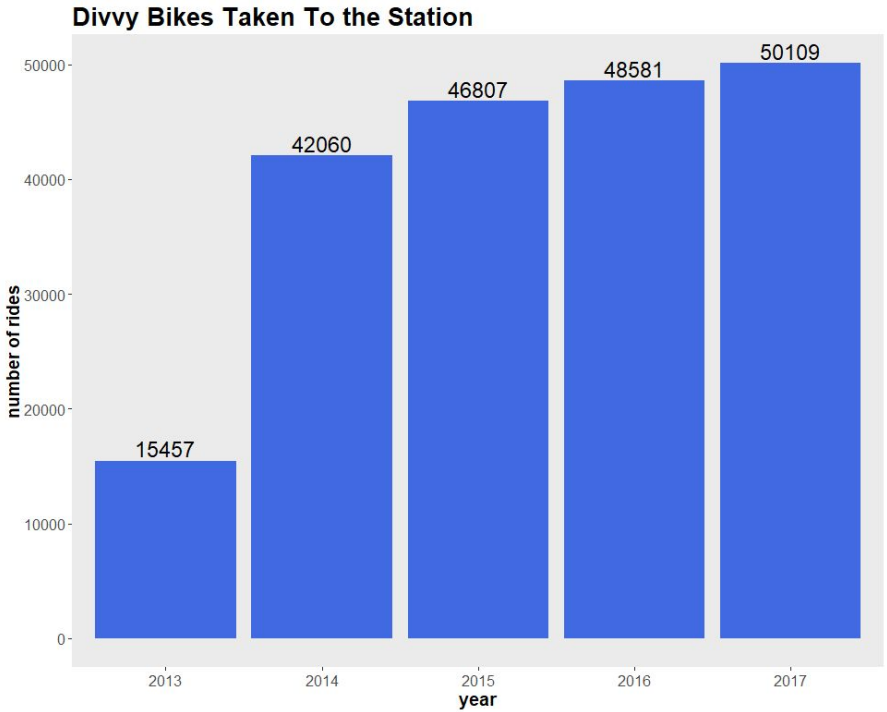
# Overview

## Data: Lake Shore Dr & Monroe Station

The graph to the right is an overview of the number of bikes taken to the Lake Shore Dr. & Monroe Divvy Station

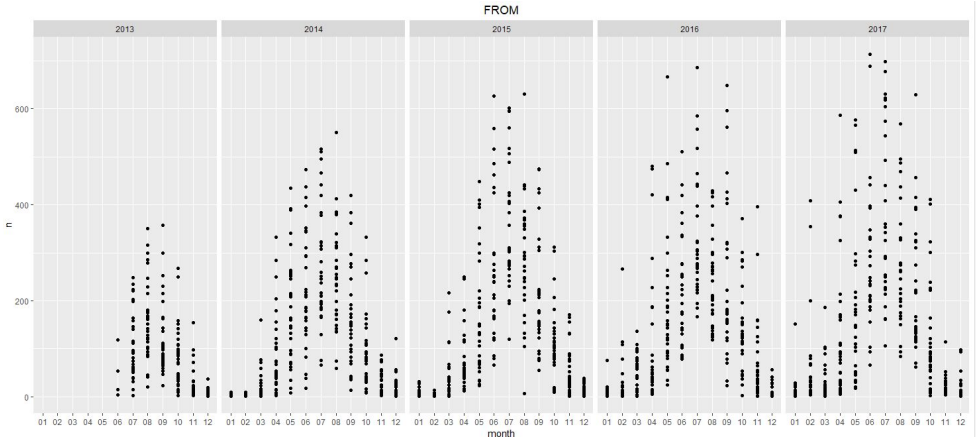
**Observation Dates:** Jun 27 2013 - December 31 2017
**Total Observations:** 203,014

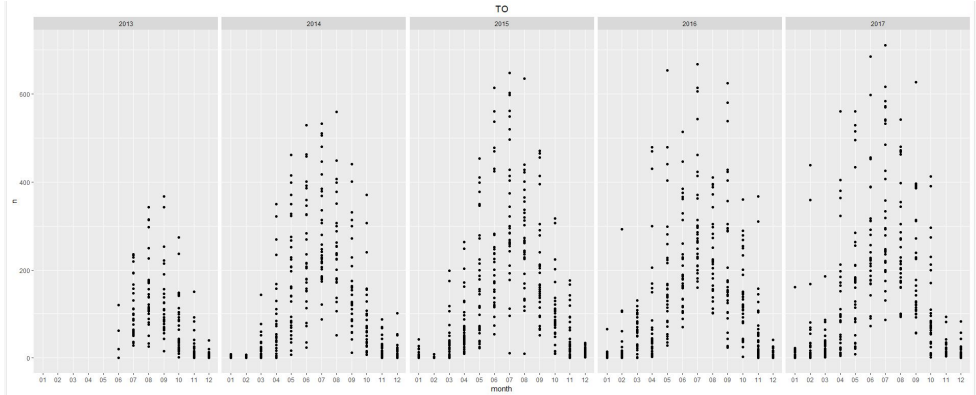
- 7,241 fewer bikes were taken to the station compared taken from the station from 2013-2017
- The significant increase from 2013 - 2014 can in part be explained by the fact the 2013 only covered 6 months of the year while 2014 - 2017 data was recorded for a full 12 month calendar year
- We can see a positive trend of rides as each year passes which is natural for a brand just starting out as more and more people become aware



# Overview

## Data: Lake Shore Dr & Monroe Station




### Insights

- There is a clear pattern of seasonality which develops which should be accounted for when modelling in both the to and from data sets

# Executive Summary

- Female trips in 2017 accounted for 25% of Divvy rides while the Chicagoland population is 51% female. When examining the overall dataset there is untapped potential in Divvy's female customer segment.

- Divvy has done a great job capturing the late 20's / early 30's age segment. However, there are other untapped age segments when compared to the population of Chicago which Divvy could also market their service to in order to have the Divvy customer population which appeals to a wider age range (and therefore has a wider spread or standard deviation). There is especially potential in the 18-24 and late 40's/early 50's age range.

- There is a clear pattern of seasonality in the Divvy data