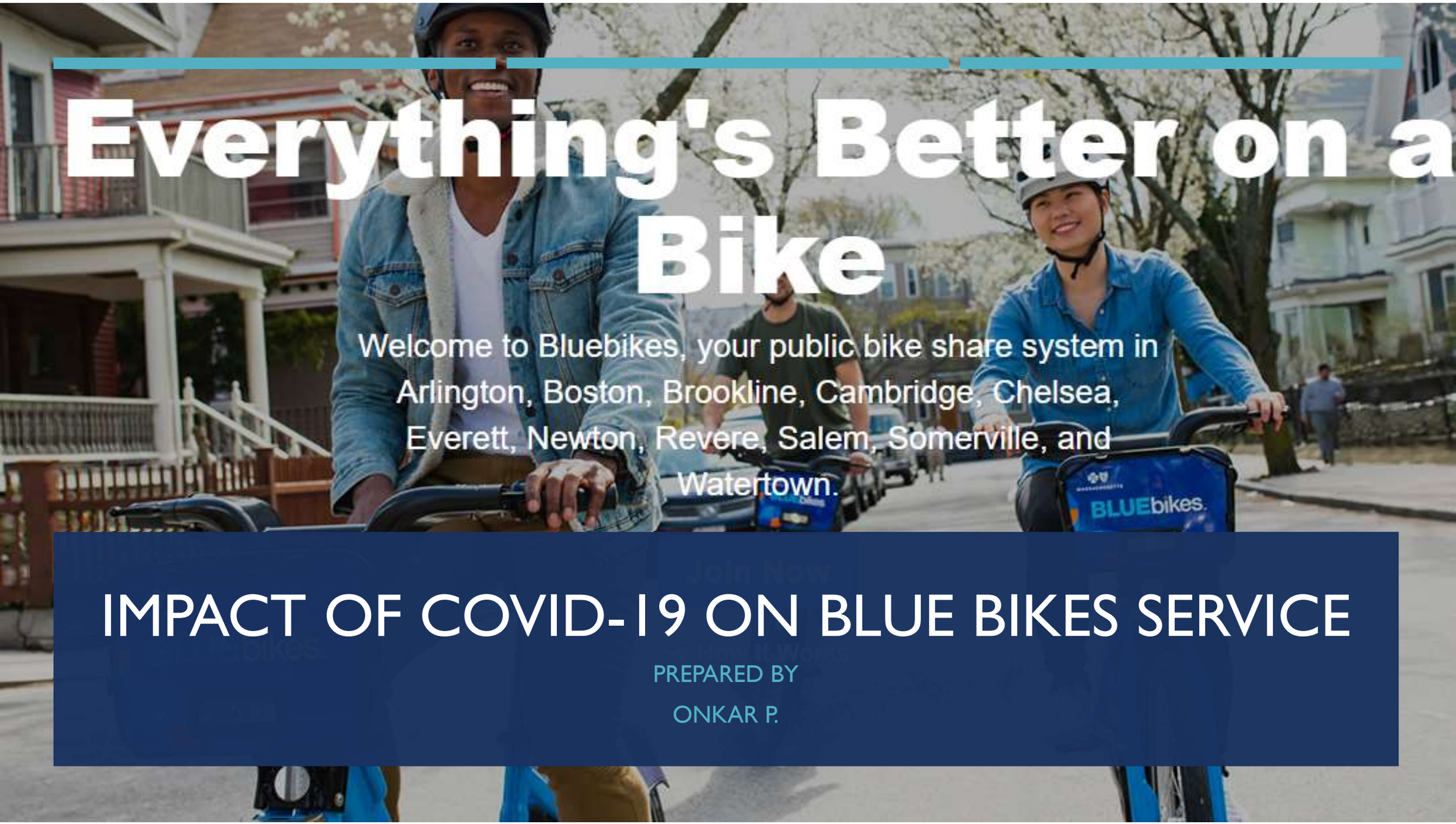


IMPACT OF COVID-19 ON BLUE BIKES SERVICE

PREPARED BY

ONKAR P.

INTRODUCTION TO BLUE BIKES

Bluebikes is a bike rental/sharing service in Boston.

Started with 3203 annual members in 2011 (610 bikes)

Grew to 21,261 members in 2019 (3500+ bikes)

Users can pick up a bike at any location, ride for specific time & return to any location for re-docking.

Vans of BlueBikes redistribute the bikes to ensure uniform availability at all locations.

BLUEBIKES DATASET

- The dataset is from Kaggle & can be downloaded by clicking on the logo on left.
- 2 Datasets are provided for 2019 & 2020.
- The objective is to perform EDA on both datasets & gain insights for effects of COVID on the service usage.
- 2 separate IPYNB are prepared for each year at GITHUB. All charts in further slides are “generated in Jupyter”
- At the end, the effects of COVID are summarized & other observations are noted.

BRIEF HISTORY OF LOCKDOWN & REOPENING IN BOSTON

FEB 2020
- 1ST CASE FOUND IN BOSTON

MAR 2020
- STAY – AT – HOME ADVISORY

MAY 2020
- MASKED MANDATE, PHASED REOPENING

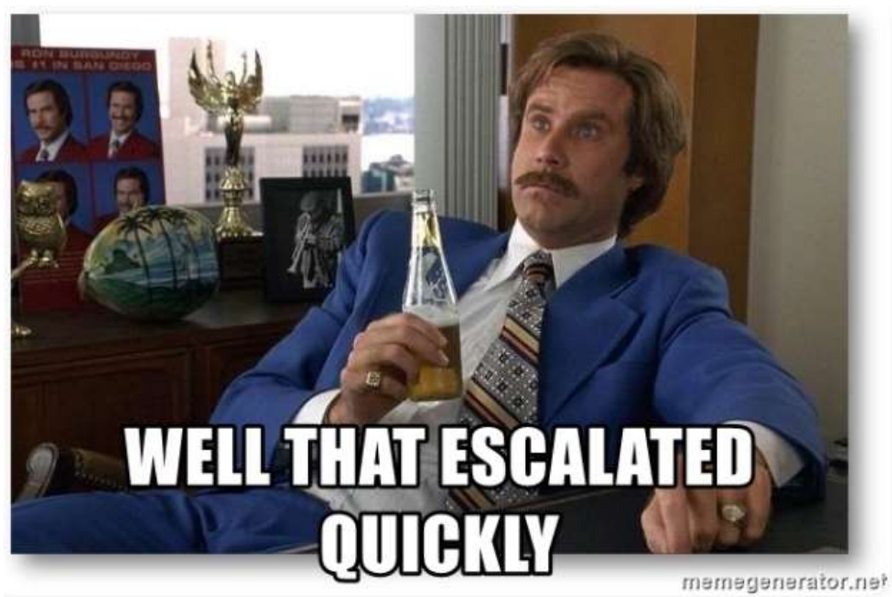
JUN, JUL 2020
- PHASED REOPENING

AUG 2020
- REOPENING PAUSE

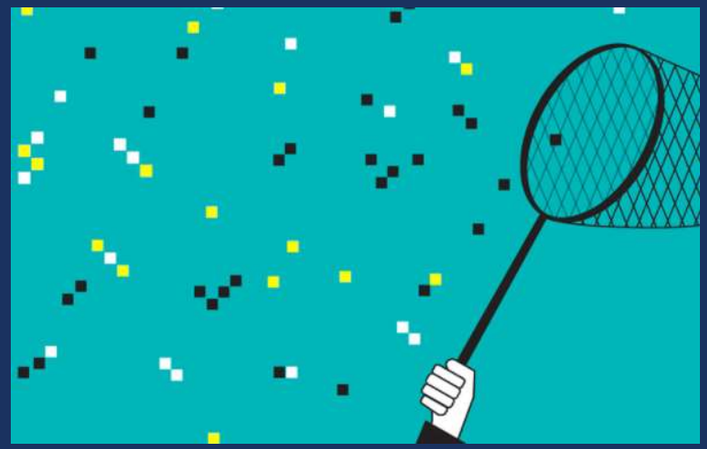
OCT 2020
- INEVITABLE 2ND WAVE!!

NOV 2020
- NEW MASK MANDATE + STAY-AT-HOME ORDERS

DEC 2020
- ROLLBACK OF REOPENING, VACCINE PLANNING



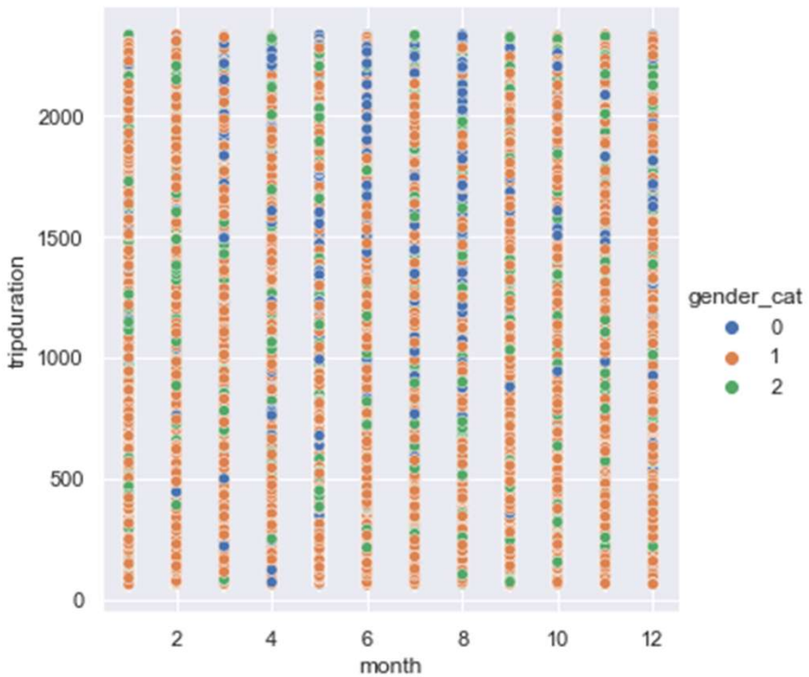
DATA WRANGLING



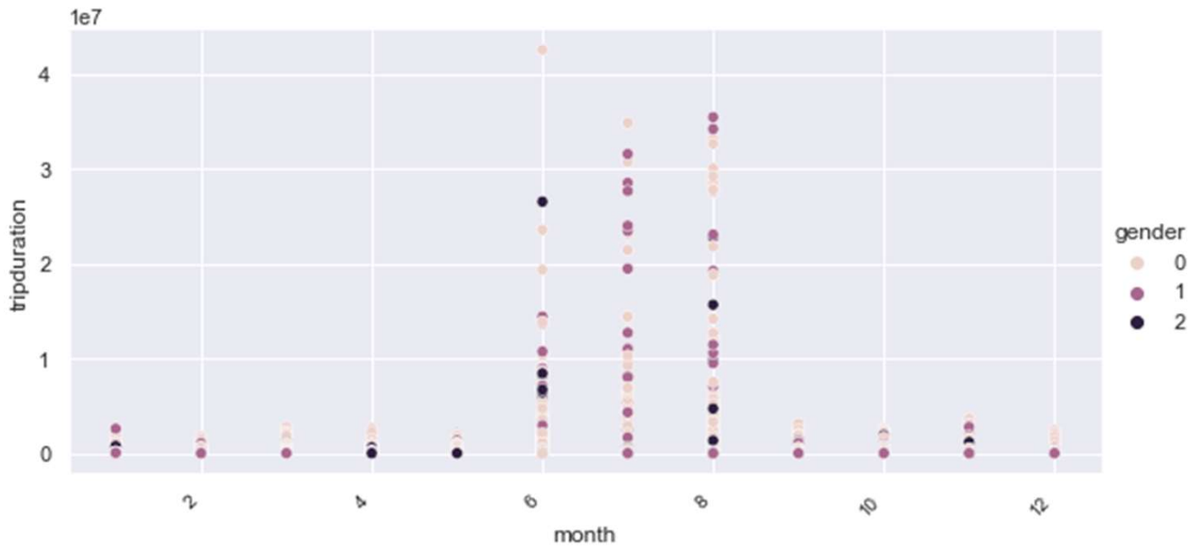
	2019 DATASET	2020 DATASET
MISSING VALUES	NONE ☺	POSTAL CODE, BIRTH YEAR, GENDER.
DATA TYPES OF COLUMNS	CHANGED FROM OBJECT TO DATETIME / INT /CATEGORICAL AS SUITABLE	CHANGED FROM OBJECT TO DATETIME / INT / CATEGORICAL AS SUITABLE
COLUMNS – DROPPED	-	POSTAL CODE
COLUMNS – ADDED	DISTANCE, DAY, MONTH ETC	DISTANCE, DAY, MONTH ETC
WHY COLUMNS WERE DROPPED?		TO BRING BOTH DATASETS TO THE SAME BAR OF COMPARISON WITHOUT ANY BIAS.

DATA WRANGLING
WHY OUTLIERS WERE NOT REMOVED?

WITHOUT OUTLIERS



WITH OUTLIERS



WITH OUTLIERS WE CAN DRAW CONCLUSIONS LIKE GENDER 0 & 1 WERE TAKING LONGER TRIPS DURING THE SUMMER MONTHS THEREFORE OUTLIERS WERE RETAINED .

IMPACT OF COVID ON BLUEBIKES
TOTAL DISTANCE BY USERS

2019

6087501
KMS

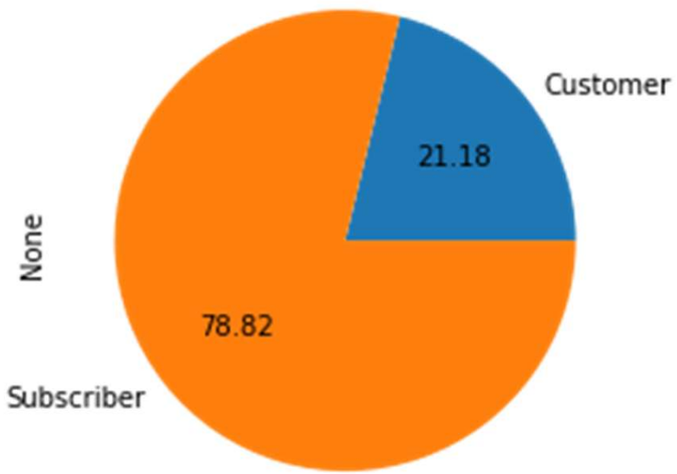
2020

4754295
KMS

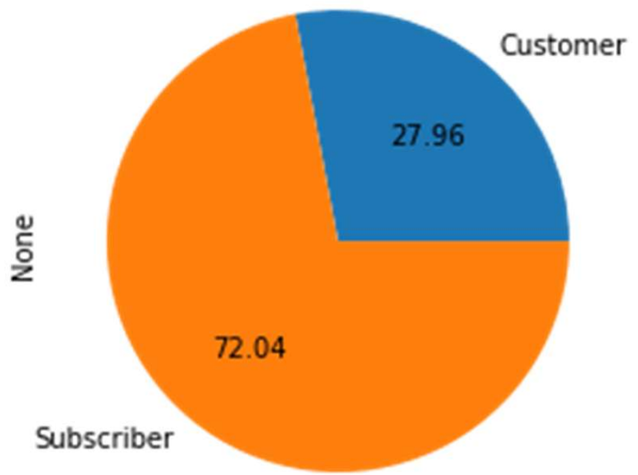
~22% DROP IN 2020

IMPACT OF COVID ON BLUEBIKES
PERCENTAGE - USER BASE

2019



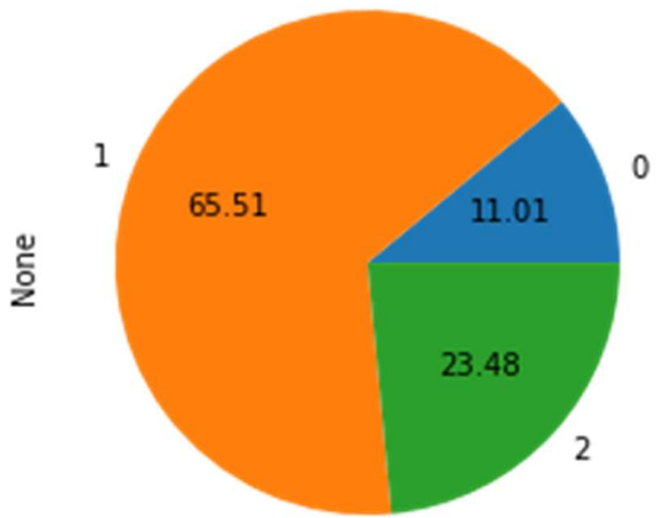
2020



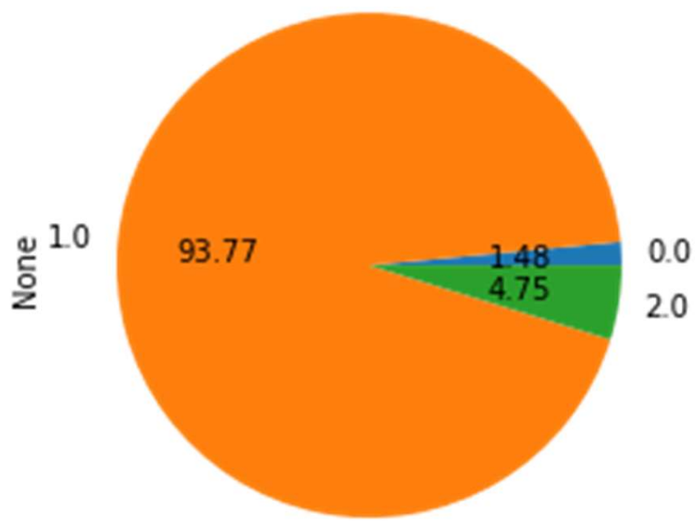
SUBSCRIBERS ↓ & RANDOM CUSTOMERS ↑ IN 2020

IMPACT OF COVID ON BLUEBIKES
PERCENTAGE - GENDER OF USER BASE

2019



2020



GENDER 2 & 0 HAVE REDUCED NUMBERS IN 2020.

IMPACT OF COVID ON BLUEBIKES
TOTAL DISTANCE BY USERS

2019

6087501
KMS

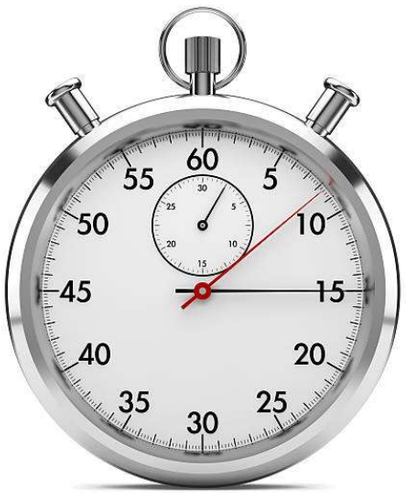
2020

4754295
KMS

~22% DROP IN 2020

IMPACT OF COVID ON BLUEBIKES
AVERAGE TRIP DURATION IN MINUTES

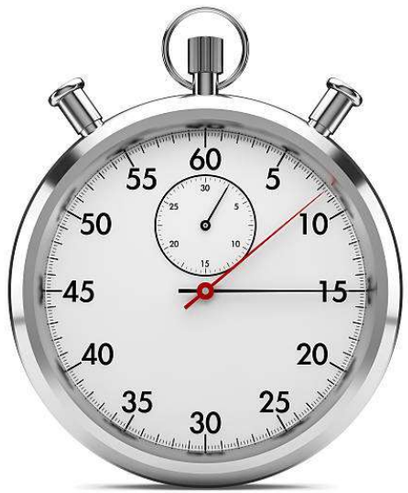
2019



~36

2020

~31

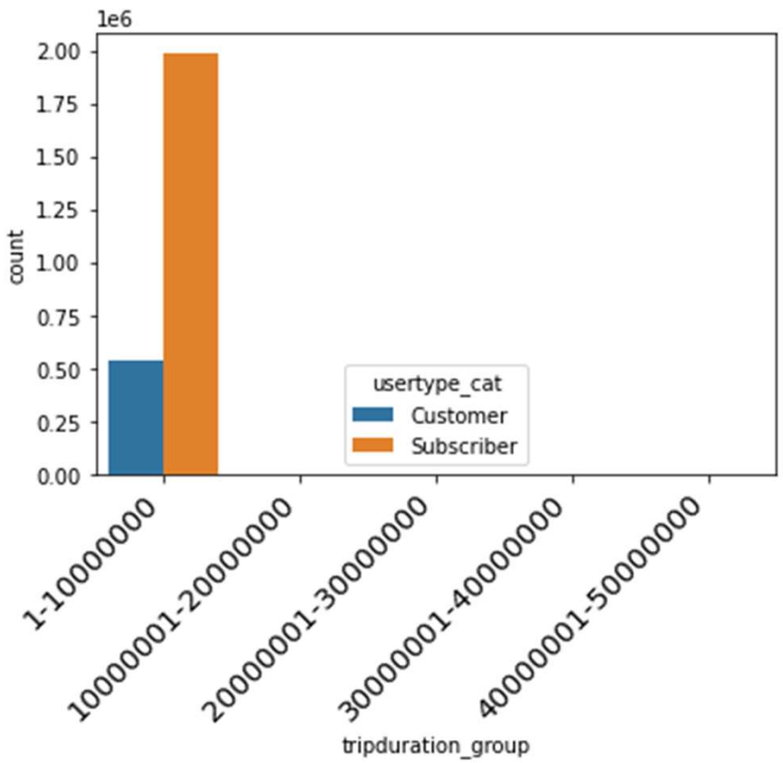


AVERAGE TRIP DURATION REDUCED SLIGHTLY IN 2020

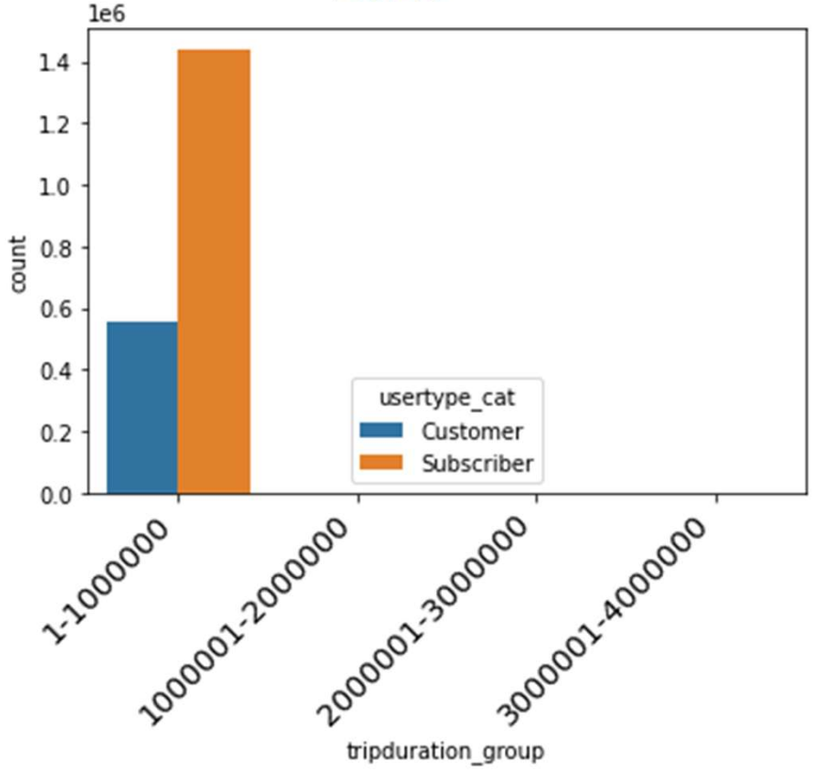
UNIVARIATE ANALYSIS

IMPACT OF COVID ON BLUEBIKES TRIP DURATION

2019



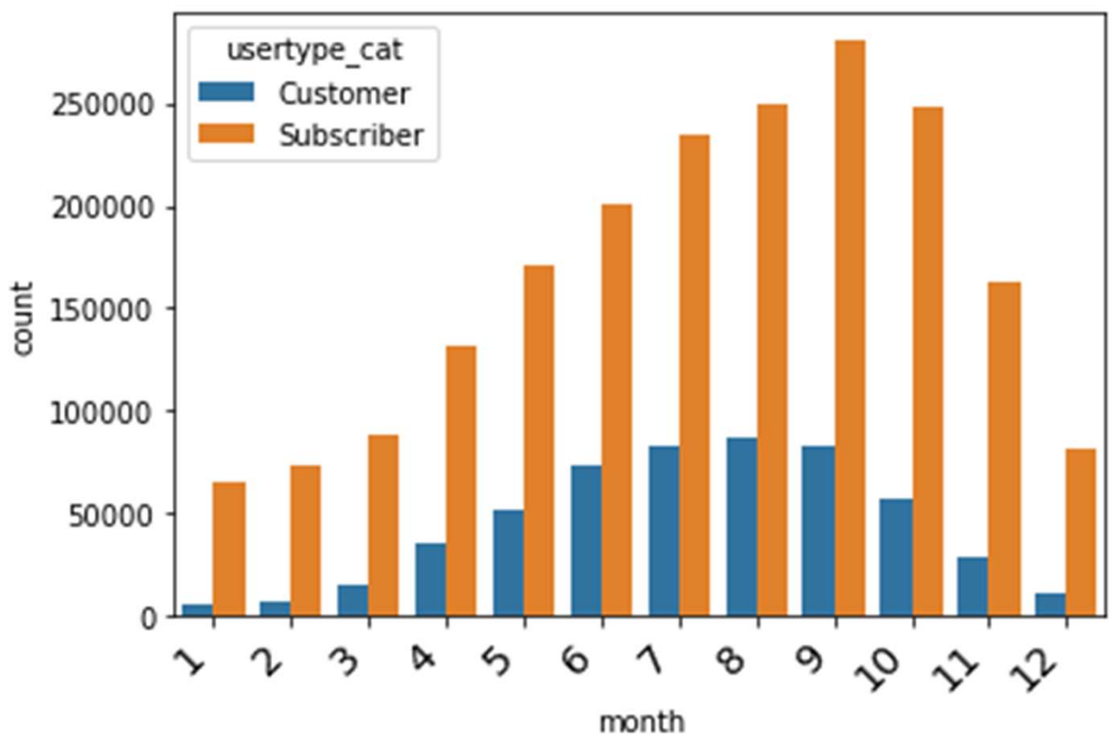
2020



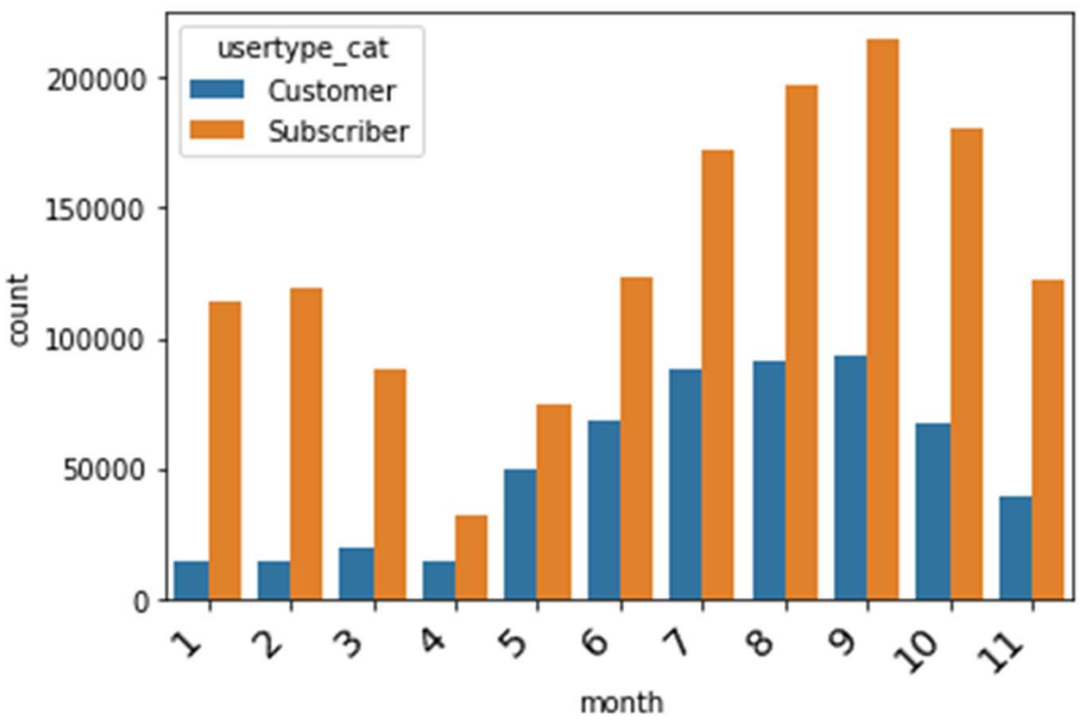
SHORTER TRIPS BY SUBSCRIBERS & LONGER TRIPS BY CUSTOMERS IN 2020

IMPACT OF COVID ON BLUEBIKES
COUNT OF TRIPS BY USER TYPE

2019



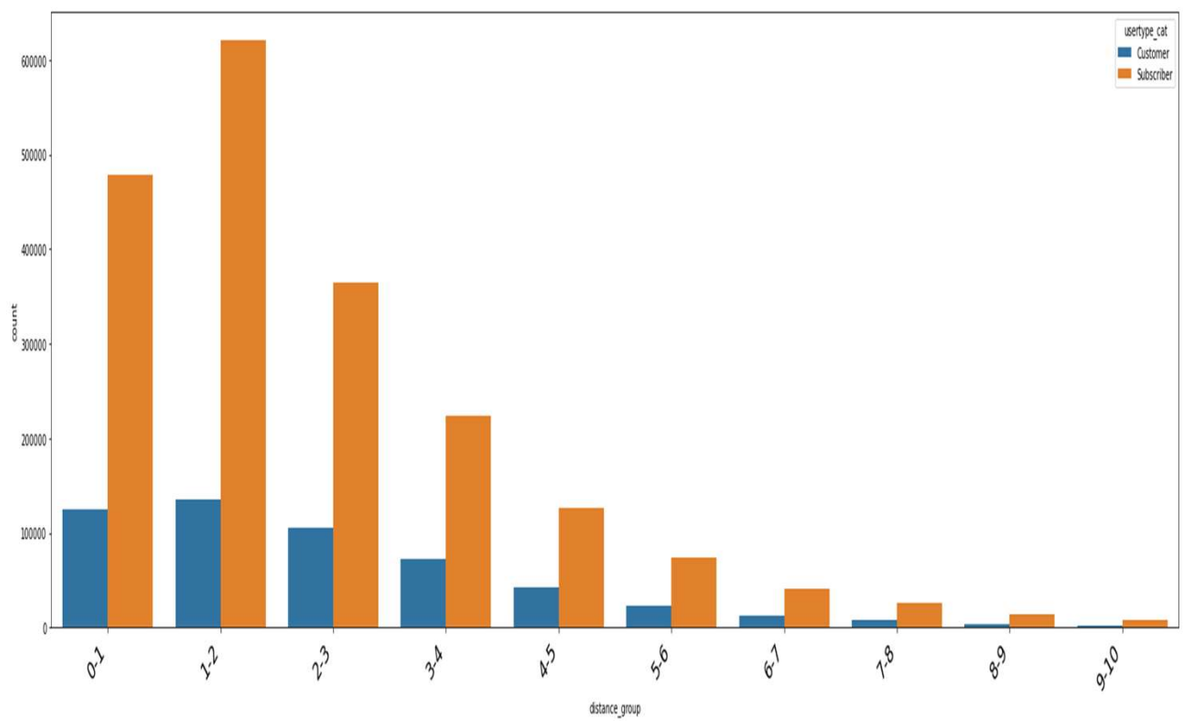
2020



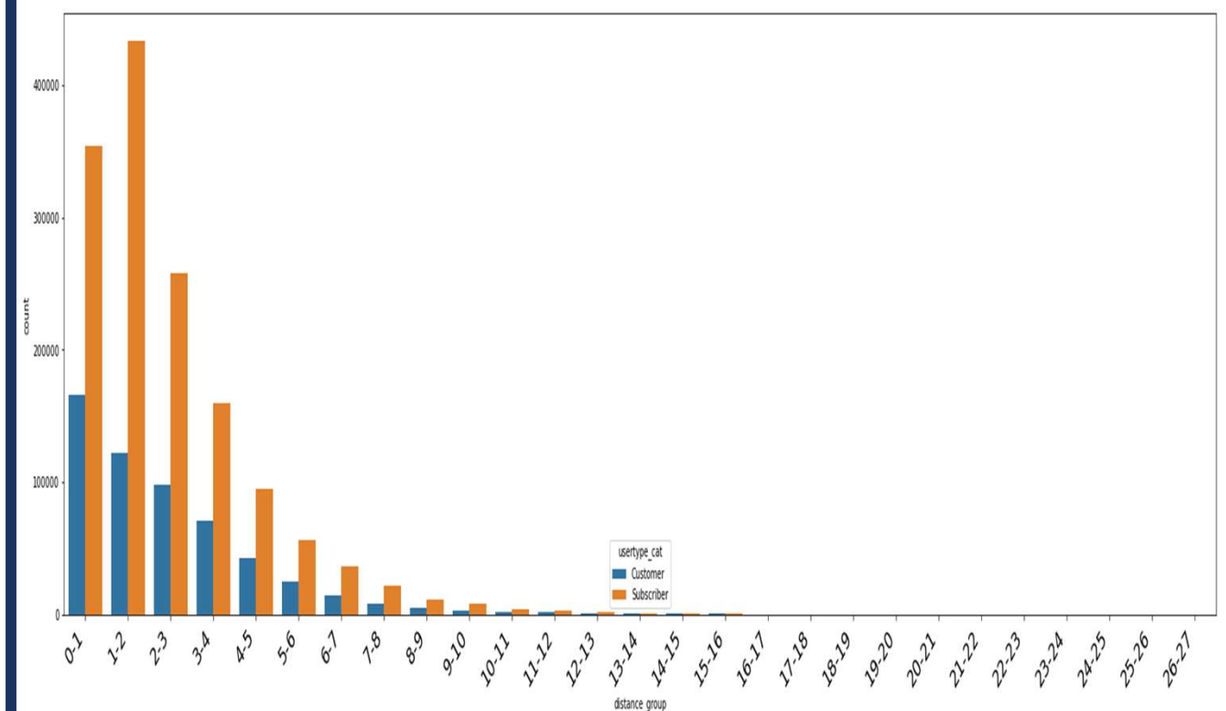
IN 2020 TRIPS DROP DURING LOCKDOWN & RISE LATER THOUGH LESS THAN 2019 LEVELS

IMPACT OF COVID ON BLUEBIKES TRIP DISTANCE

2019



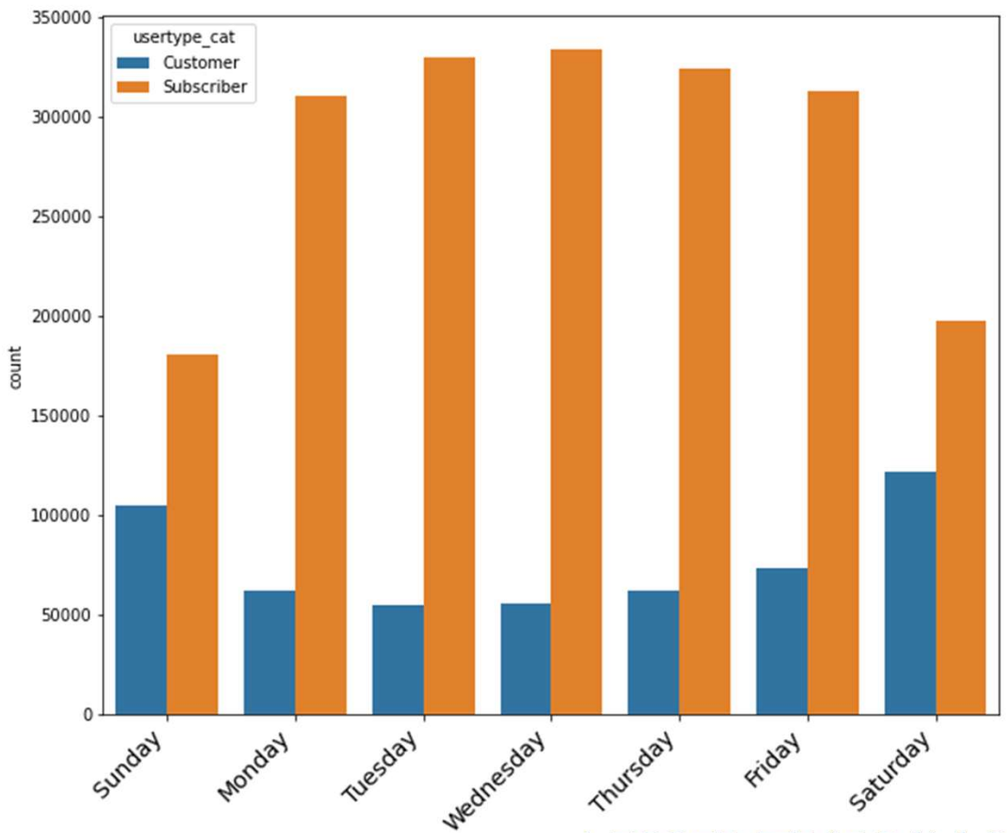
2020



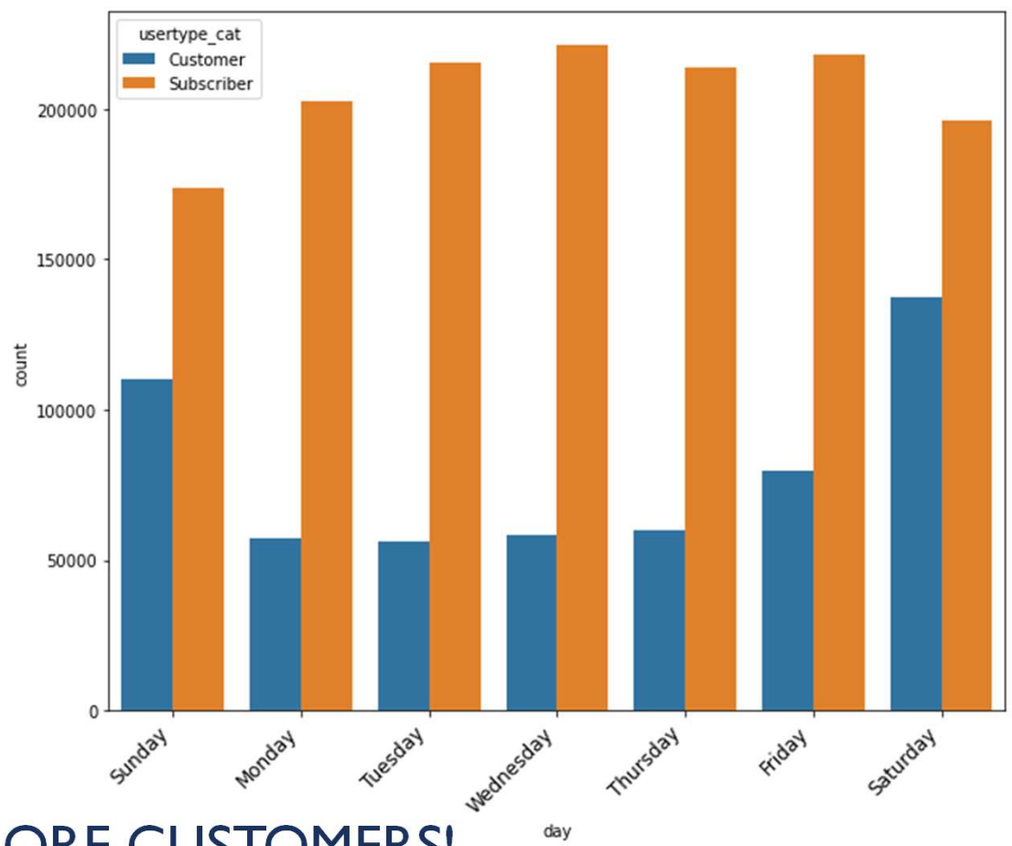
FEWER TRIPS IN 2020.

IMPACT OF COVID ON BLUEBIKES TRIPS ON VARIOUS DAYS

2019



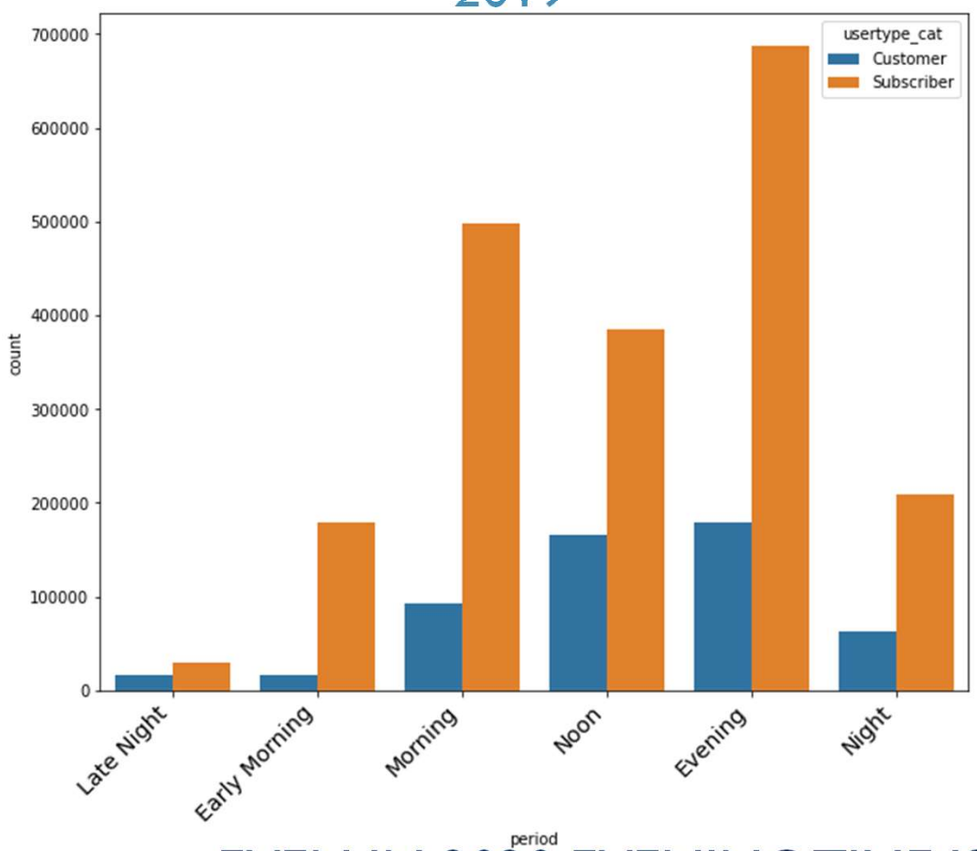
2020



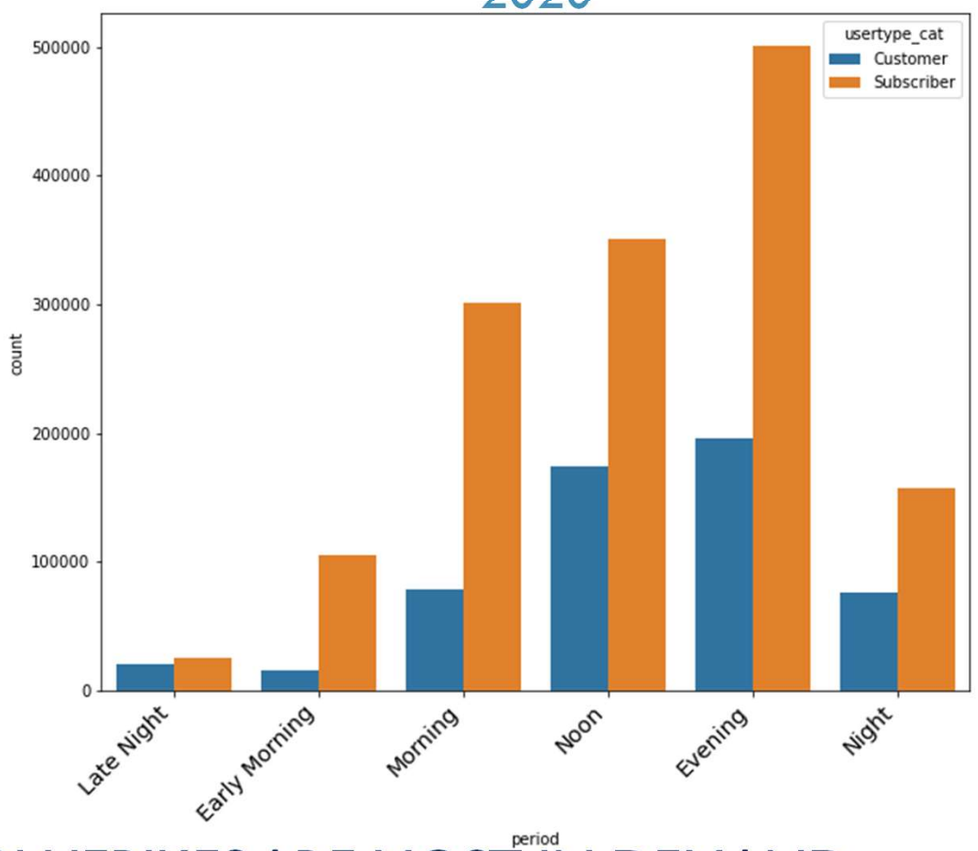
LESSER TRIPS IN 2020 BUT MORE CUSTOMERS!

IMPACT OF COVID ON BLUEBIKES TRIPS AT VARIOUS TIMES OF THE DAY

2019



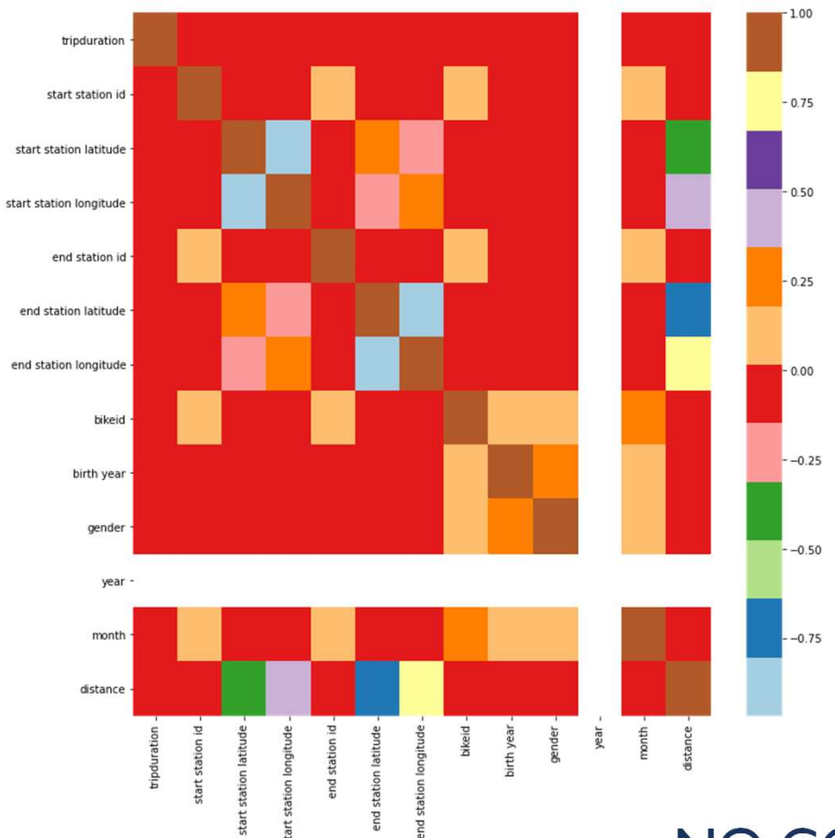
2020



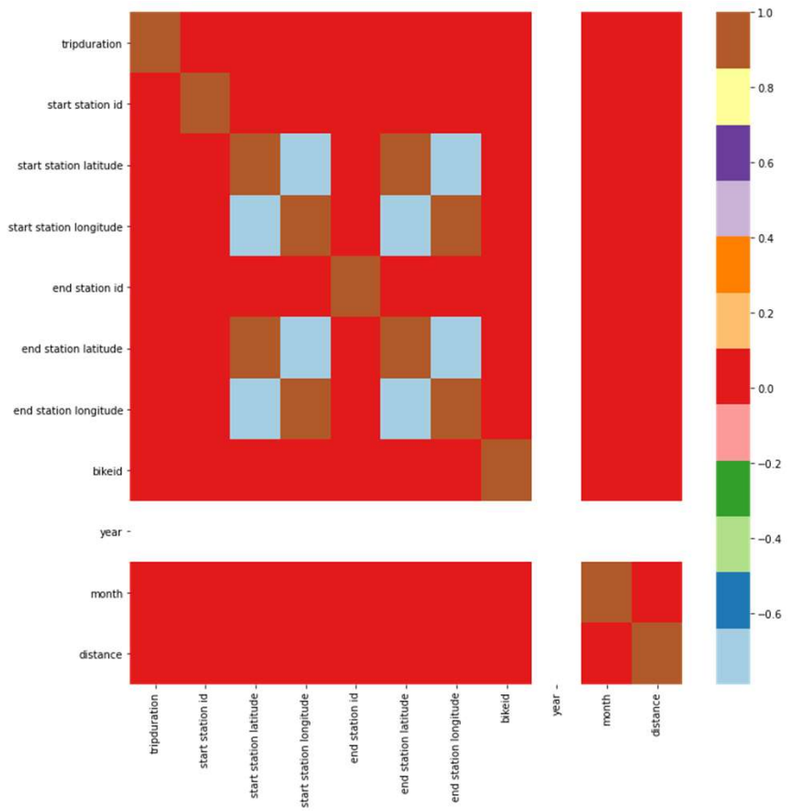
EVEN IN 2020 EVENING TIME IS WHEN BLUEBIKES ARE MOST IN DEMAND.

IMPACT OF COVID ON BLUEBIKES CORRELATION

2019



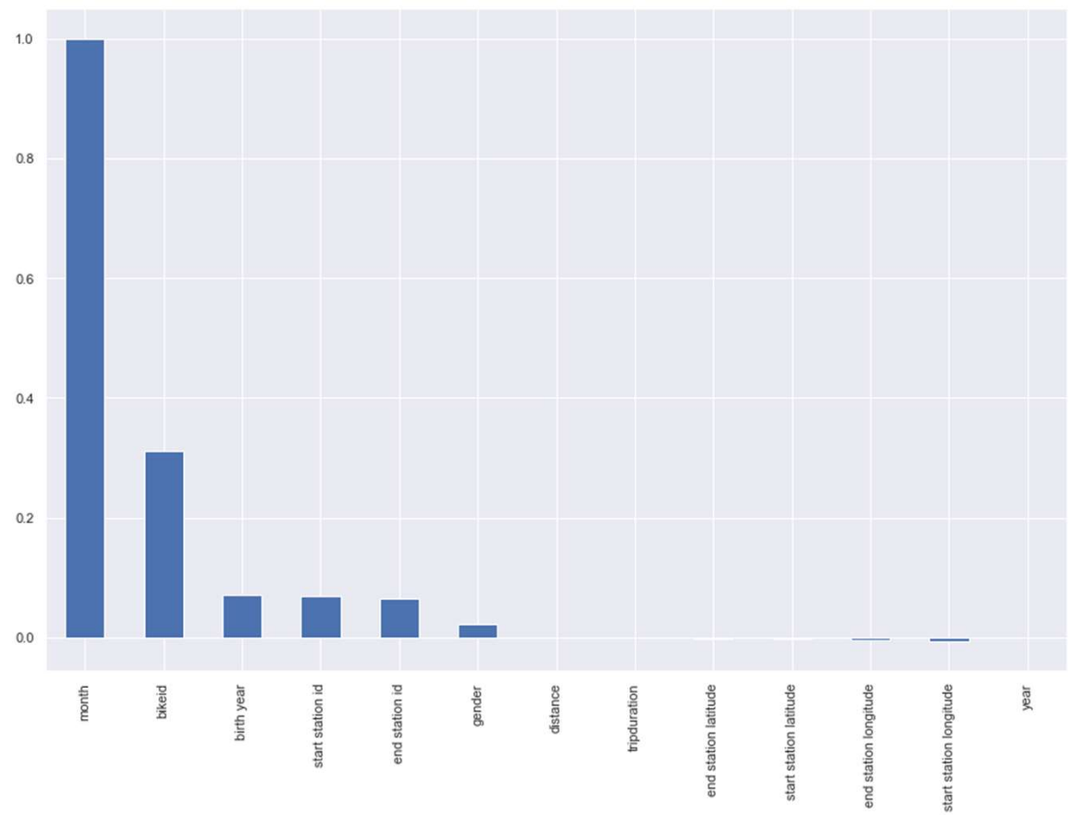
2020



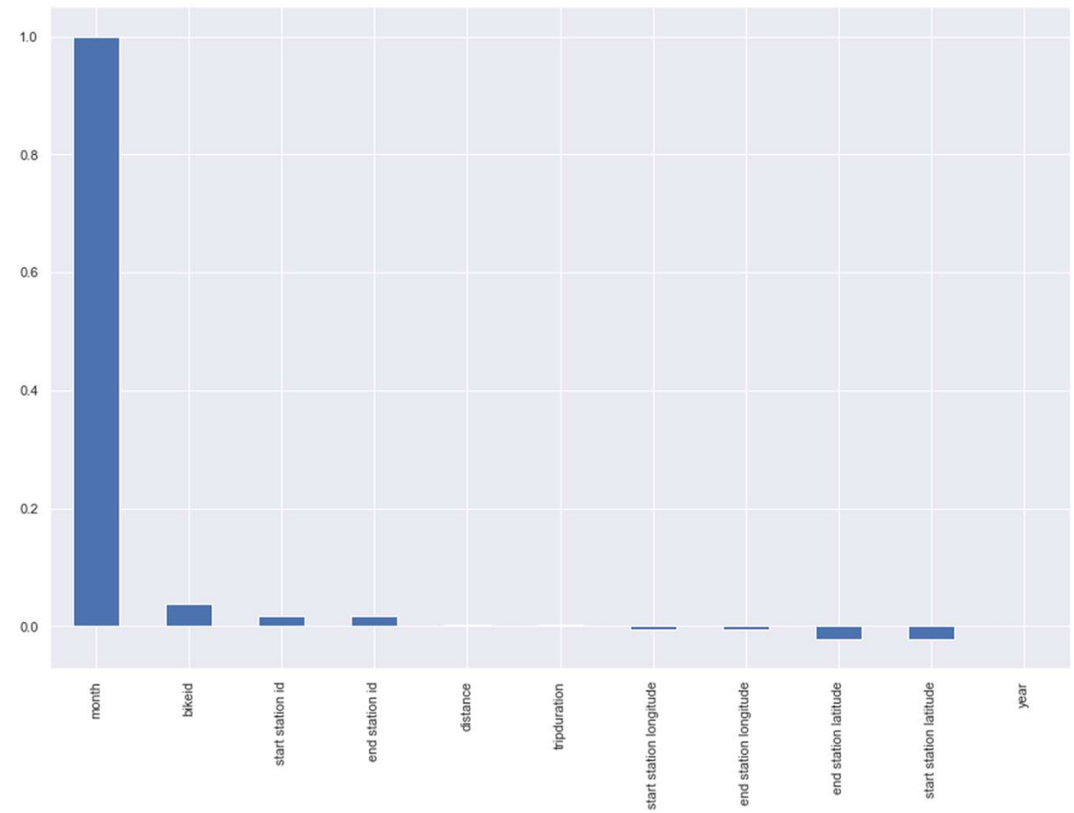
NO CORRELATION WAS VISIBLE

IMPACT OF COVID ON BLUEBIKES CORRELATION

2019



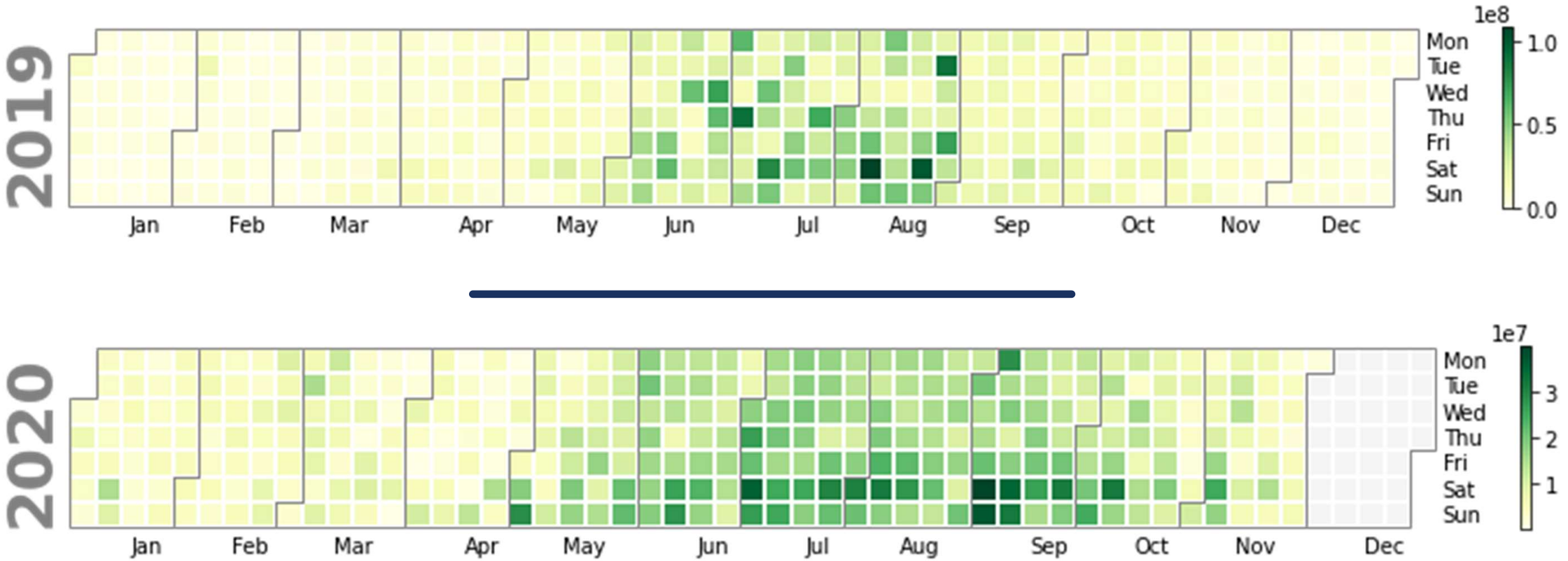
2020



NO CORRELATION WAS VISIBLE

BIVARIATE ANALYSIS

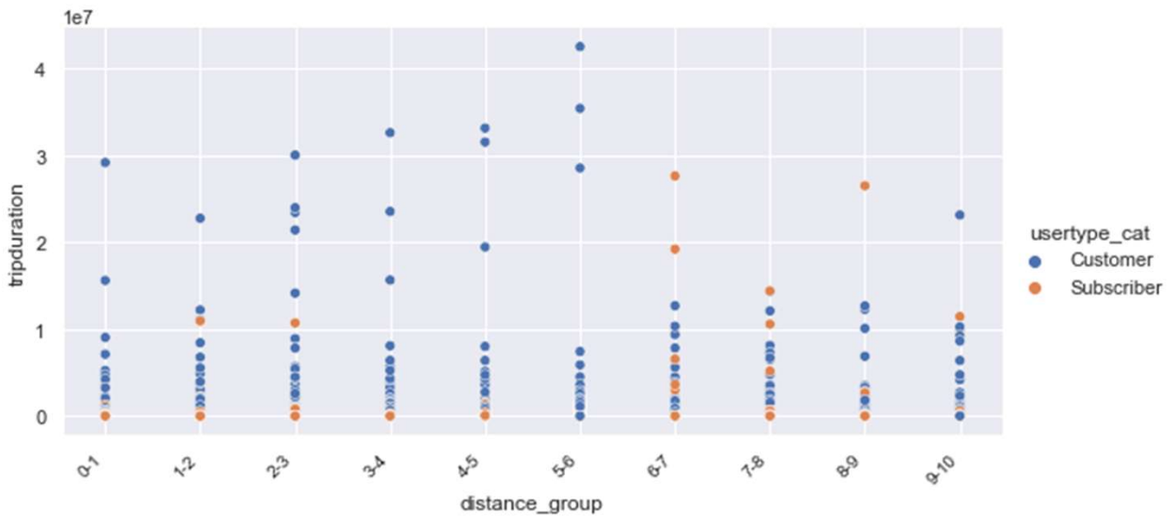
IMPACT OF COVID ON BLUEBIKES
VIEW OF TRIPS FOR A FULL YEAR SHOWING DAILY, WEEKLY, MONTHLY



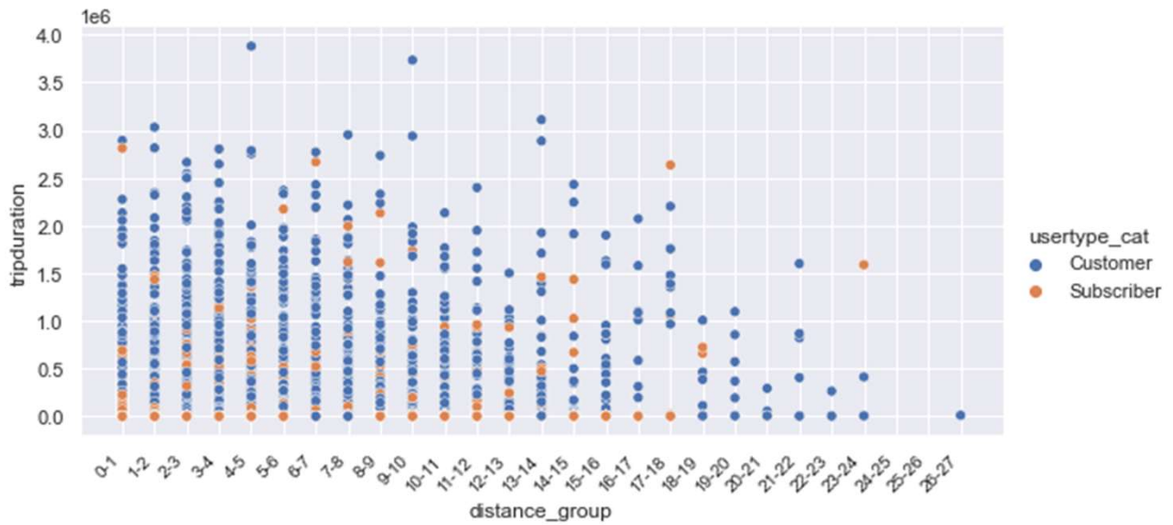
CALPLOT CLEARLY SHOWS LESS USAGE DURING MARCH APRIL 2020 & MORE DURING SUMMER

IMPACT OF COVID ON BLUEBIKES
TRIP DURATION & DISTANCE BY CUSTOMERS/SUBSCRIBERS

2019



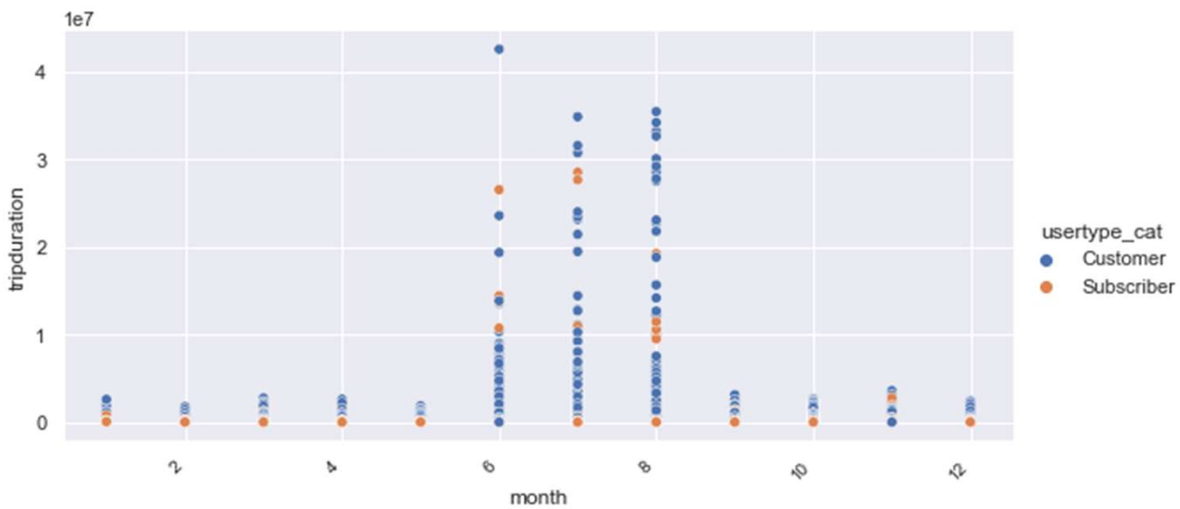
2020



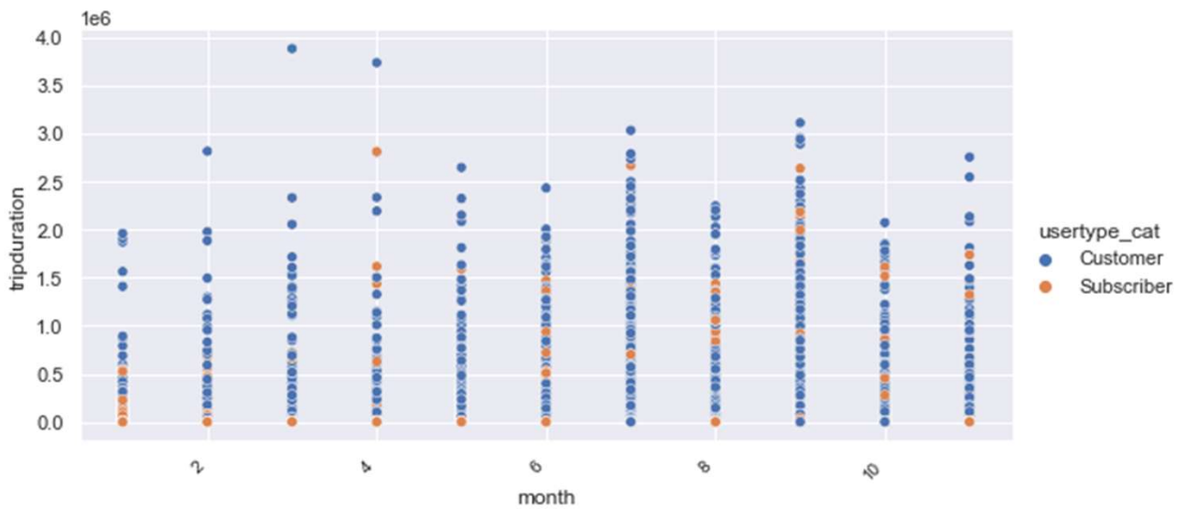
MORE CUSTOMERS USED THE SERVICE FOR LONGER TRIPS AS WELL AS LONGER DISTANCES

IMPACT OF COVID ON BLUEBIKES
TRIP DURATION & MONTHS BY CUSTOMERS/SUBSCRIBERS

2019



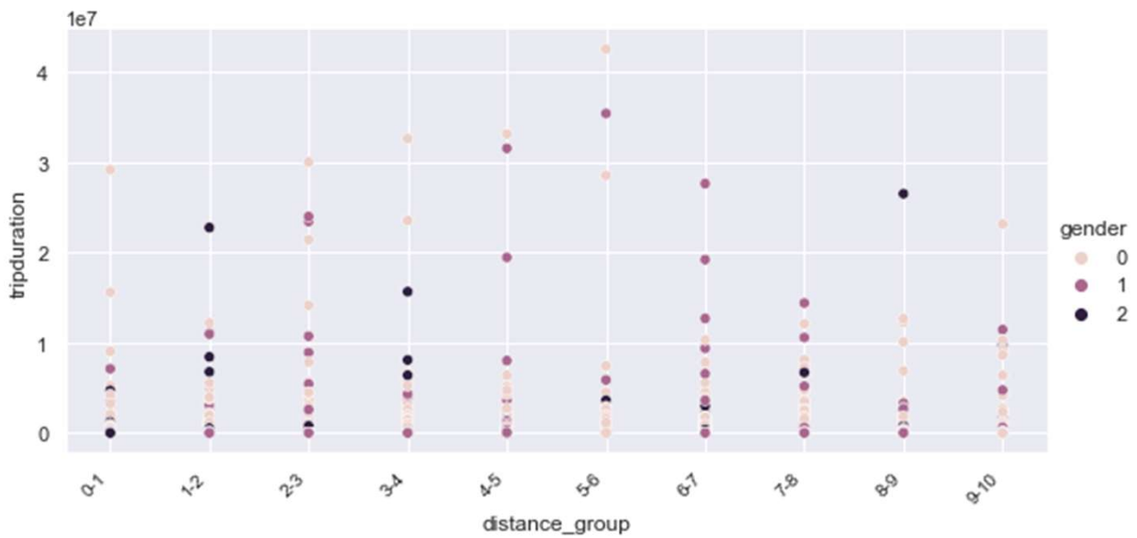
2020



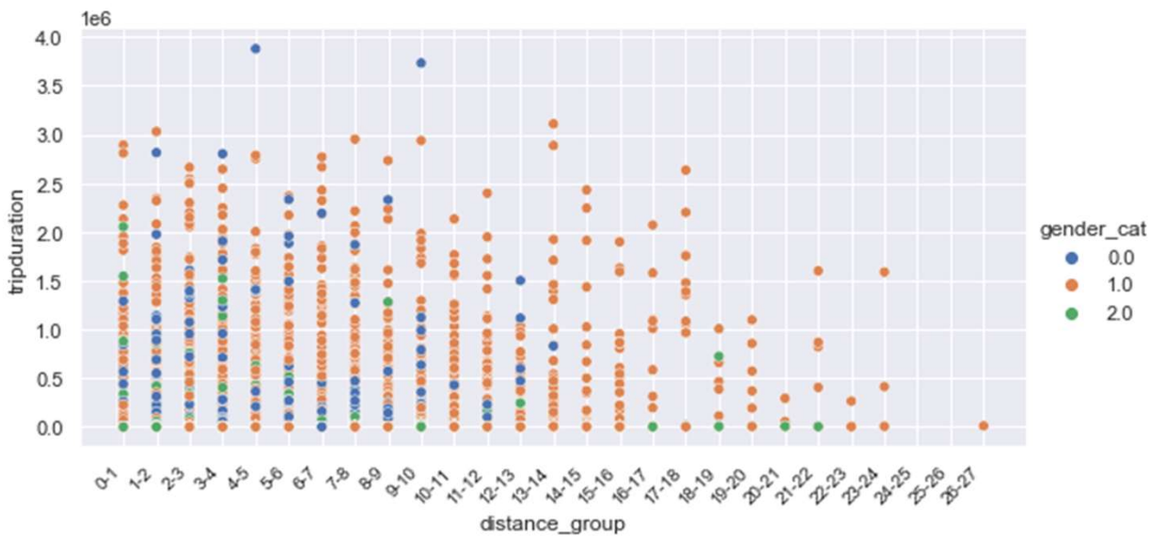
MORE CUSTOMERS USED THE SERVICE FOR LONGER TRIPS AS WELL AS LONGER DISTANCES

IMPACT OF COVID ON BLUEBIKES
TRIP DURATION & MONTHS BY GENDER

2019



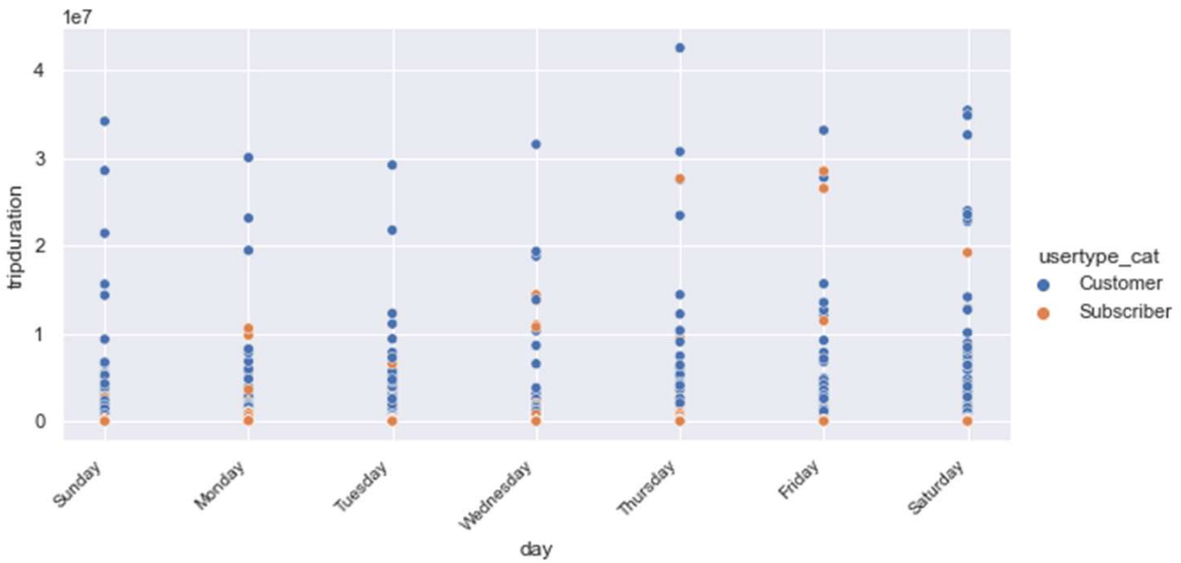
2020



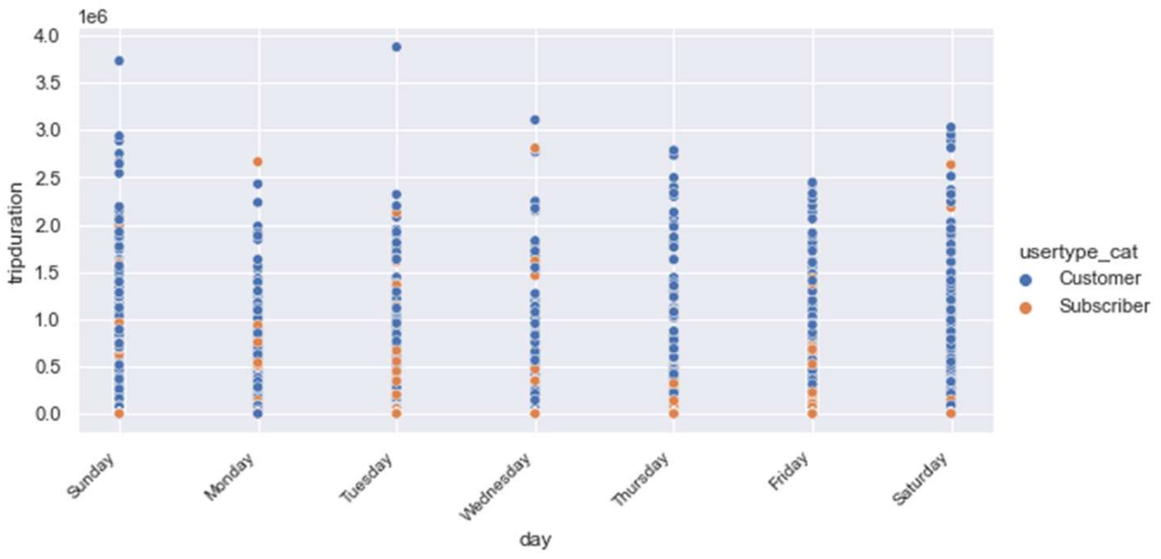
GENDER 1 TOOK LONGER & FARTHER TRIPS IN 2020

IMPACT OF COVID ON BLUEBIKES
TRIP DURATION & DAY BY CUSTOMERS/SUBSCRIBERS

2019



2020



MORE CUSTOMERS USED THE SERVICE FOR LONGER TRIPS AS WELL AS LONGER DISTANCES

FINDINGS, INSIGHTS & MORE!

IMPACT OF COVID ON BLUEBIKES
FINDINGS FROM EDA

- Average trip duration in 2020 has reduced to 31 mins from 36 mins in 2019.
- Total trip duration in 2020 has fallen to ~43K days from ~63K days in 2019
- **Bluebikes have reduced the no. of bikes from 4652 in 2019 to 3942 in 2020.**
- **41 New stations are added in 2020.**
- More subscribers use the service for short trips. Customers are only 25% of subscribers.
- Ridership by both customers & subscribers increases from spring to autumn & drops during the harsh winter seasons.
- Customers undertake longer & farther trips during the summer months.
- Due to lockdown, ridership dropped during March & April 2020.
- Ridership increased during the re-opening in 2nd half of 2020.
- Subscribers dropped during 2020 due to lockdown & customers increased.
- More subscribers use the service during the weekdays even in 2020.
- No. of customers is highest on Saturdays probably due to weekend outings. It was highest in 2020.
- Use of the service is the most during evenings. This is probably to beat the evening rush hour traffic.
- 2019 Ridership was more in morning, but it has decreased in 2020.

OVERALL – EVEN WITH REDUCED TRIP TIMES IN 2020 BLUEBIKES IS CLEARLY AIMING TO EMERGE STRONGER AT THE END OF PANDEMIC BY ADDING NEW STATIONS.

IMPACT OF COVID ON BLUEBIKES
FINDINGS FROM EDA

- Top 5 most used & bottom 5 least used start stations of 2019 are listed below.

```
MIT at Mass Ave / Amherst St                61056
Central Square at Mass Ave / Essex St       50997
MIT Stata Center at Vassar St / Main St     47197
South Station - 700 Atlantic Ave            44427
Ames St at Main St                          39128
                                            ...
Washington St at Walsh Playground              33
American Legion Hwy at Canterbury St           21
Mobile Temporary Station 2                      3
MTL-ECO5.1-01                                   1
8D QC Station 01                                1
Name: start station name_cat, Length: 363, dtype: int64
```

- Top 5 most used & bottom 5 least used start stations of 2020 are listed below.

```
Central Square at Mass Ave / Essex St                          32668
Charles Circle - Charles St at Cambridge St                    31712
MIT at Mass Ave / Amherst St                                   28286
Christian Science Plaza - Massachusetts Ave at Westland Ave    26696
Ames St at Main St                                             22410
                                                               ...
Revere City Hall                                                  13
Blue Hill Ave at Havelock St                                      11
Revere Public Library                                              7
BCBS Hingham                                                       4
MTL-ECO4-01                                                        1
Name: start station name_cat, Length: 386, dtype: int64
```

MIT IN TOP 3 INDICATES SERVICE IS MOSTLY USED BY STUDENTS/COLLEGE STAFF.

IMPACT OF COVID ON BLUEBIKES
FINDINGS FROM EDA

- Top 5 most used & bottom 5 least used end stations of 2019 are listed below.

```
MIT at Mass Ave / Amherst St                 56986
Central Square at Mass Ave / Essex St        51442
Ames St at Main St                           46295
Nashua Street at Red Auerbach Way            46245
MIT Stata Center at Vassar St / Main St      41828
                                               ...
American Legion Hwy at Canterbury St            22
Mobile Temporary Station 2                       6
Warehouse Lab PBSC                               1
MTL-ECO4-01                                      1
8D QC Station 02                                 1
Name: end station name_cat, Length: 364, dtype: int64
```

- Top 5 most used & bottom 5 least used end stations of 2020 are listed below.

```
Central Square at Mass Ave / Essex St                          33493
Charles Circle - Charles St at Cambridge St                    32045
Christian Science Plaza - Massachusetts Ave at Westland Ave    26844
MIT at Mass Ave / Amherst St                                   26550
Ames St at Main St                                             26113
                                                                 ...
Revere City Hall                                                   9
Revere Public Library                                              8
BCBS Hingham                                                       4
Mobile Temporary Station 1                                         1
MTL-ECO4-01                                                        1
Name: end station name_cat, Length: 387, dtype: int64
```

MIT HAS SLIPPED FEW PLACES ON THE LIST IN 2020.

IMPACT OF COVID ON BLUEBIKES
QUESTIONS FROM KAGGLE

Q] Are there any null values or outliers? How will you wrangle/handle them?

A] There were no null values in 2019 data. In 2020 data there were ~86% null values in gender & birth year column. Ideally these columns with null should have been deleted but for the academic purposes of this EDA mode was imputed in gender & mean was imputed for birth year.

Q] Are there any useful variables that you can engineer with the given data?

A] The customer/subscriber variable is an important one for EDA which showed customers using the service more on weekends. From the trip start & end time stamps, day & period of day columns were added which showed the service being used more in evenings.

Q] Do you notice any patterns or anomalies in the data? Can you plot them?

A] In 2019 data, gender 0 was noted for major entries & birth year for this gender was 1969. This anomaly appears to be wrongly entered by large no. of users. In 2020 data, gender data was missing from the month of April onwards. Plots for these are in attached python notebooks.

Q] Which stations were most popular and vice versa?

A] This information has been provided in previous slides.

IMPACT OF COVID ON BLUEBIKES
QUESTIONS FROM KAGGLE

Q] Do you notice any differences between 2019 & 2020? Can you plot them?

A] We can note the trend of trips going down during the month of March & April of 2020 due to lockdown & again going up during reopening of June 2020. This has been plotted in the Univariate section.

Q] Do you think COVID-19 had any influence? Explain the rationale.

A] Covid-19 definitely had an impact on number of trips in 2020. The average trips per month in 2019 were ~210K & reduced to ~181K in 2020. The number of bikes was reduced by 710. The number of stations increased by 41.

Q] Which time of day was the busiest and vice versa?

A] Evening time was the most busiest & late night was least busiest.

Q] Can we segment/cluster our users into categories? What is their frequent destinations?

A] Since MIT appears in the top 5 end stations, we can segment the top users as students/college staff using the service on a daily basis.

Q] Do you have a suggestion for redistribution of bikes in the short run? (for e.g. during a specific time of day, and specific neighborhood, which possible stations would you choose for bike redistribution?)

A] The top 5 start & end stations in the evening should be targeted for redistribution of bikes.

WHAT NEW DID I LEARN IN THIS EDA?

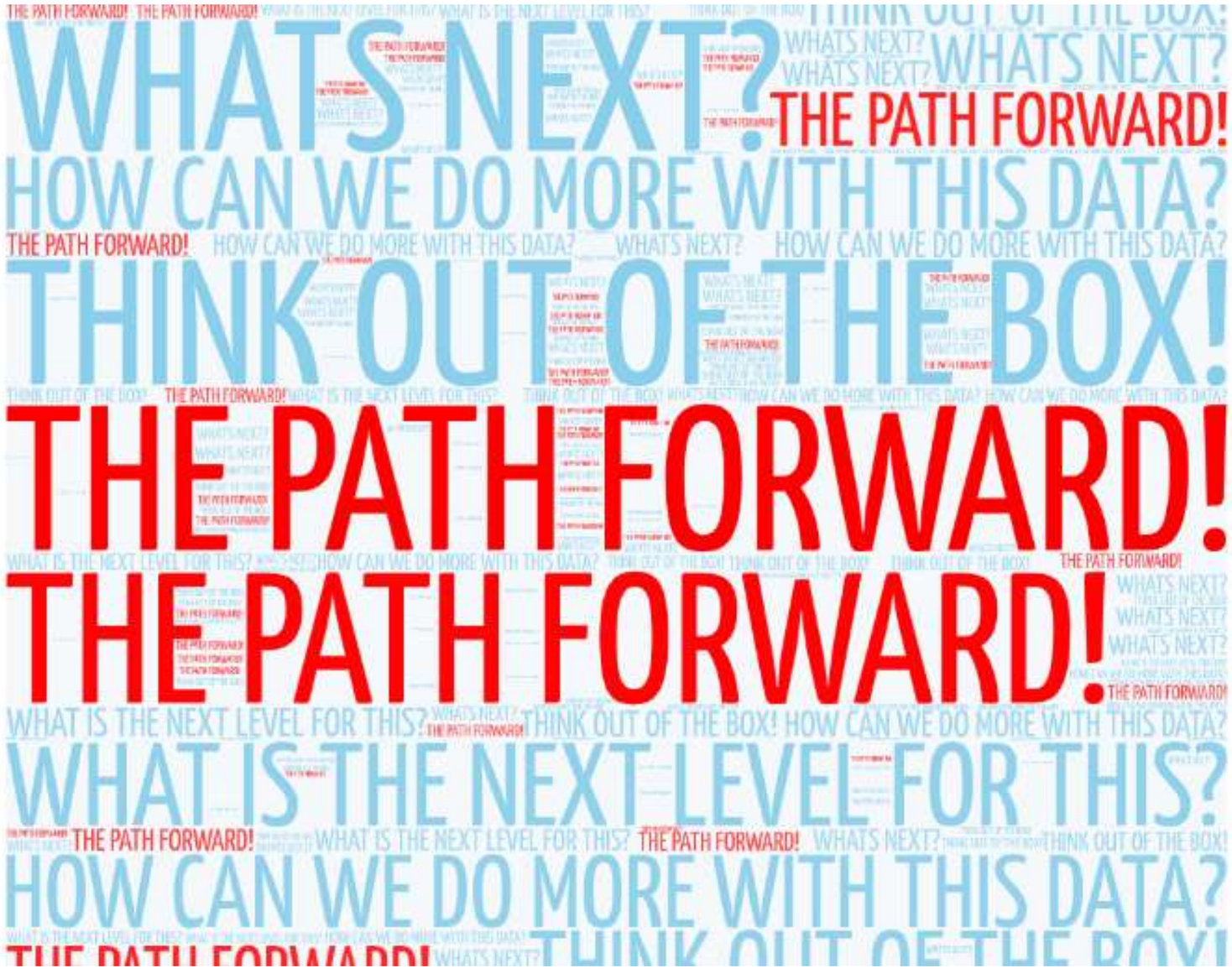
MY LEARNINGS FROM THIS EDA.

Calculating distance in kms from given trip data of start/stop latitude/longitude. Considering the curvature of the Earth, the distance is not calculated on a flat plane. (**Haversine Formula**)

Calculating the Weekday, time of day (morning/night/noon) from the trip start/stop time stamp.

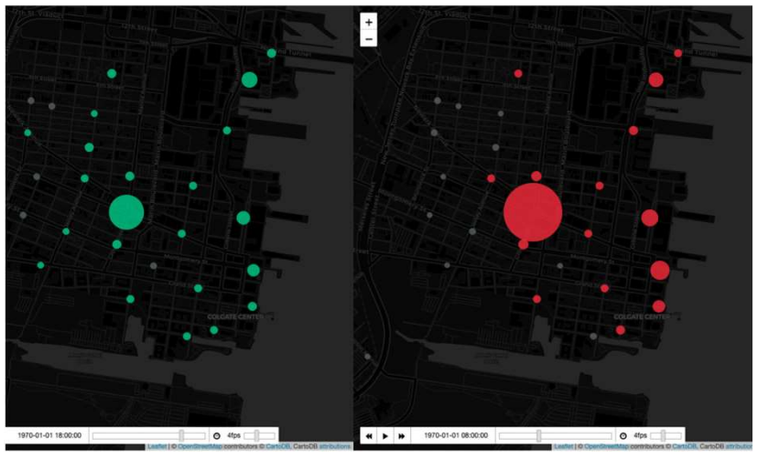
Discovered a library called **Calplot** on the web. It is useful to present a time series data for a full year arranged in neat weekly layout & brings the benefit of heatmaps.

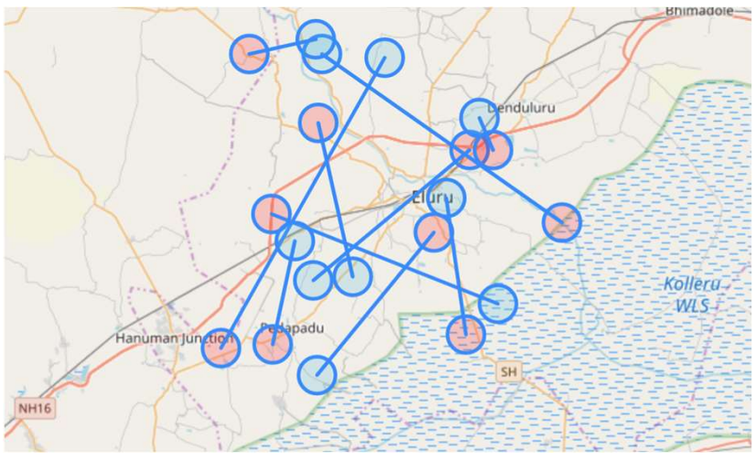
Discovered a mapping library called **Folium** which I shall discuss in depth in next section.



THE PATH FORWARD!

THE NEXT LEVEL FOR THIS EDA – PLOTTING ON MAP!





- Using the Folium library, I am planning to plot the start & end points of the bike trips on the map of Boston.
- This will give a clear picture of which areas are more popular.
- I am also planning to plot side by side the start & end stations for both years on map to understand which were discontinued in 2020 or started new.
- I have tried this unsuccessfully few times on my laptop but the code doesn't run due to large RAM requirements & the maps don't get plotted. However, I will keep searching for ideas on the web & succeed one day!
- Some examples are shown.

SIMILAR CHALLENGES!

SIMILAR CHALLENGES ON KAGGLE

- Google Data Analytics Certification Capstone Project. (Cyclistic)
- San Francisco Bike Sharing.
- London Bike Sharing.
- New York Bike Sharing.
- Washington Bike Sharing.
- Austin Bike Sharing.


THANK YOU