COVID Movement Tracking

How has COVID-19 affected movement throughout Chicago?

An Analysis by: Noni Bell, John Carroll, and Dan Cusick

HYPOTHESES

General

There is a notable correlation between the onset of COVID-19 (and its accompanying restrictions) and transit in Chicago.

Null

There is no notable correlation between the onset of COVID-19 (and its accompanying restrictions) and transit in Chicago.

Questions

Question 1

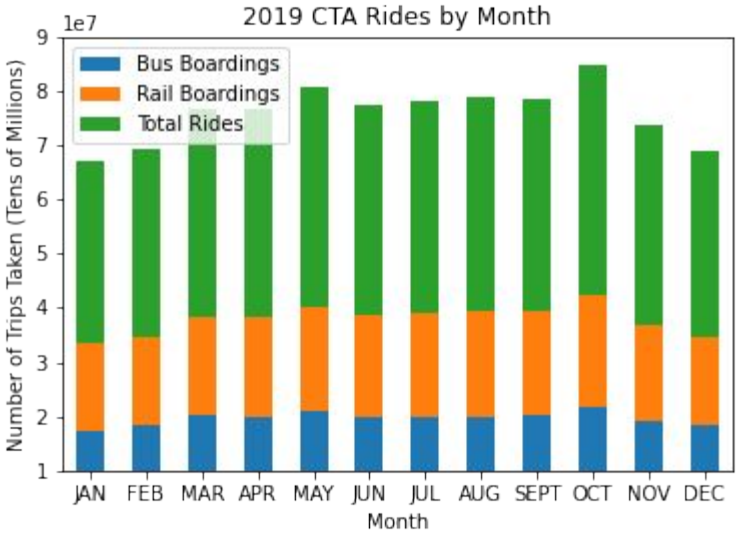
How has each main transport method changed between 2019 and 2020? Has any one increased? Decreased?

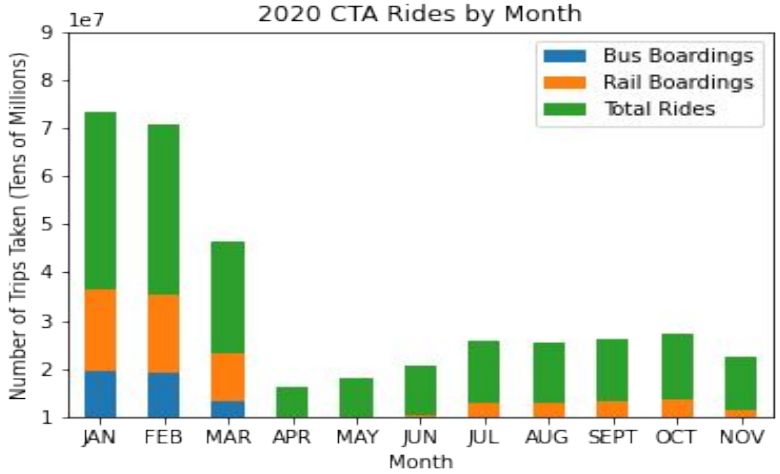
Question 2

How do the different transport methods compare?

Question 3

What other axes (trip times, trip distance) can be compared?



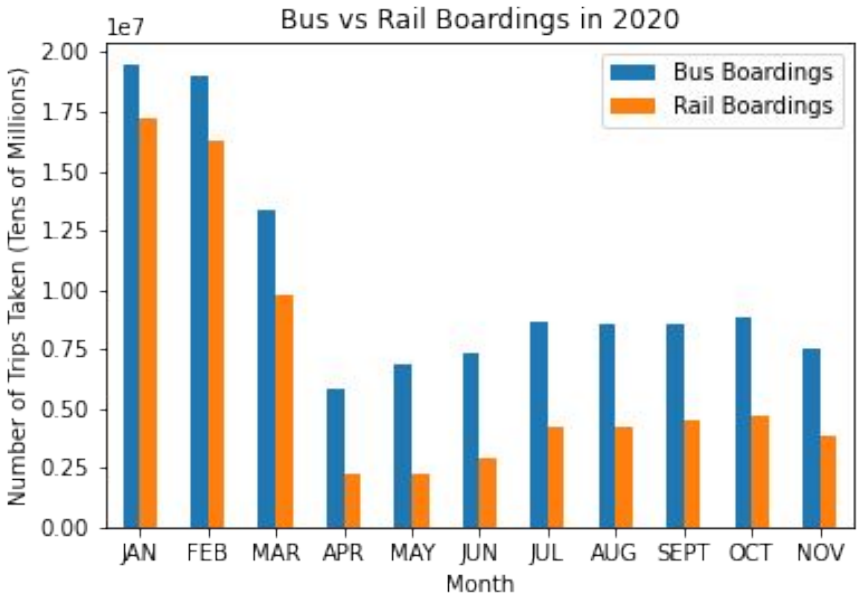


1. How do daily boarding totals between 2019 and 2020 compare?

Source: https://data.cityofchicago.org/Transportation/CTA-Ridership-Daily-Boarding-Totals/6iiy-9s97

- .csv file tracks January 2001 - November 2020
- simple cleanup slicing 2019 and 2020 rows by range

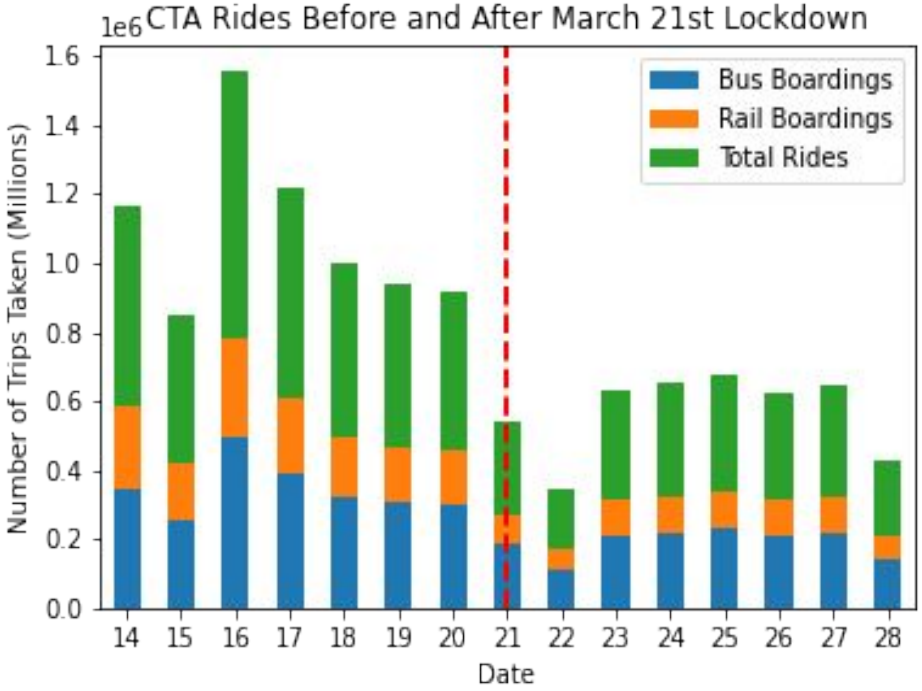
Bus vs. Rail



2. What is the comparison between bus and rail boardings in 2020?

Why was bus more popular than rail?

Although buses could be more "cramped", it's more likely that people were willing/had to travel shorter distances than over rail.

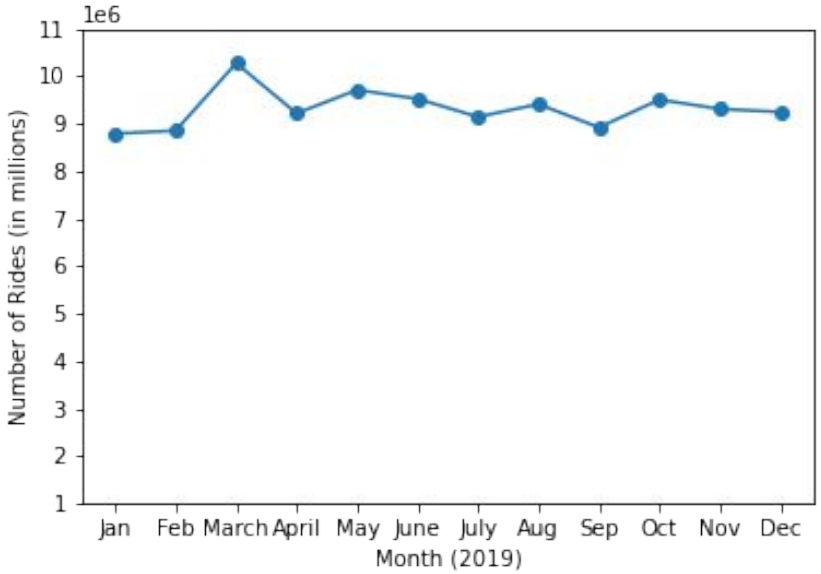


3. What was the effect of the March 21st lockdown announcement on boarding totals?

Between 2019 and 2020:

- The p-value for bus boardings was **4.971929233847977e-05**.
- The p-value for rail boardings was **2.2127920757801055e-05**.
- The p-value for total boardings was **3.2453159317145606e-05**.

Null Hypothesis? **Refuted**. There appears to be a correlation between COVID's onset and CTA ridership.

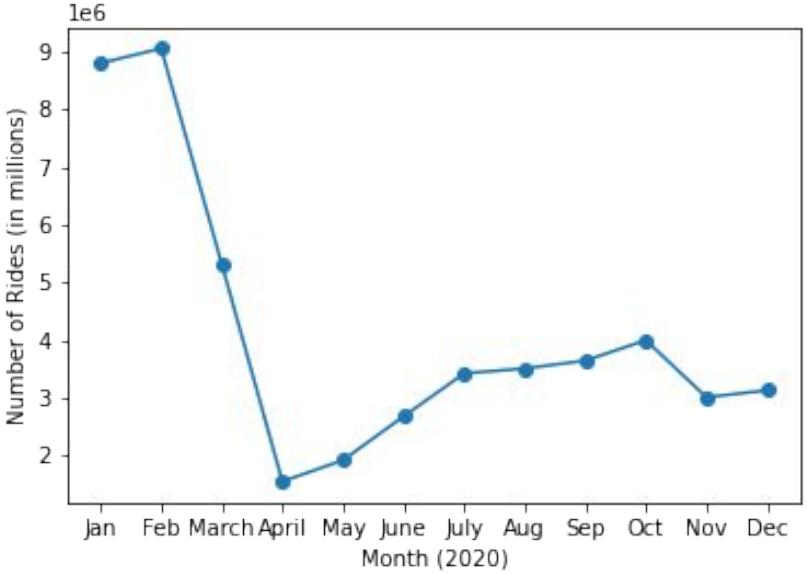


Rideshare

Question 1: Did COVID affect the number of rideshares in the Chicago area?

Question 2: Was the average distance of each rideshare impacted by COVID?

Question 3: Was the average ride length (measured in minutes) impacted by COVID?


1e6
9
8
7
6
5
4
3
2
Number of Rides (in millions)
Jan
Feb
March
April
May
June
July
Aug
Sep
Oct
Nov
Dec
Month (2020)

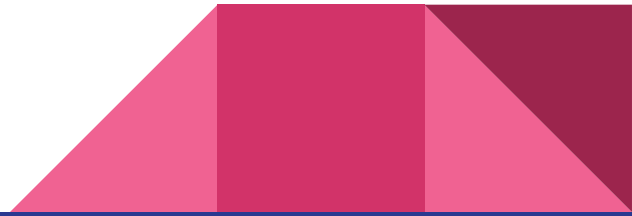
The Drop Off

Data Retrieval

Links:
https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2020/rmc8-eqv4

https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2019/iu3g-qa69

Above links provided all rideshare data that took place to the city of Chicago in 2020 and 2019 respectively



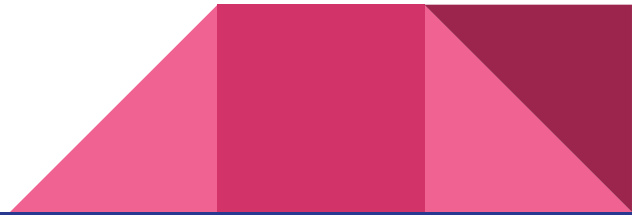
Data Cleanup

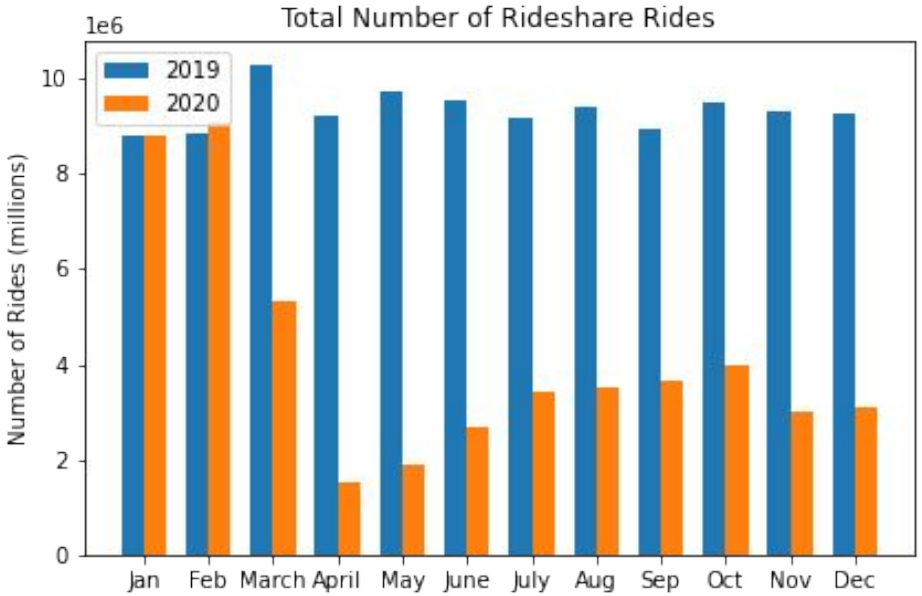
The csv file for 2019 was too large!

Used sed -n ‘1p;0~2p’ Trips_2019.csv>new_trips_2019.csv to print the first line (the header) as well as every other line to a new csv file

Did the same for 2020

Imported the csv files to pandas and used dropna() to get rid of rows with null values





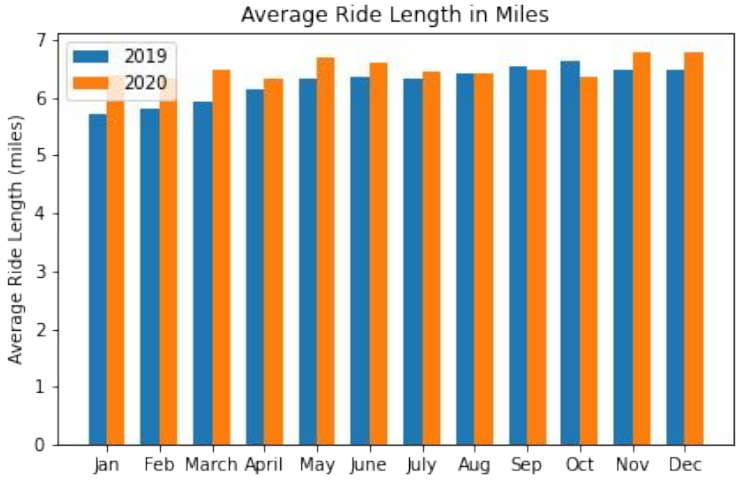
Total Number of Rides

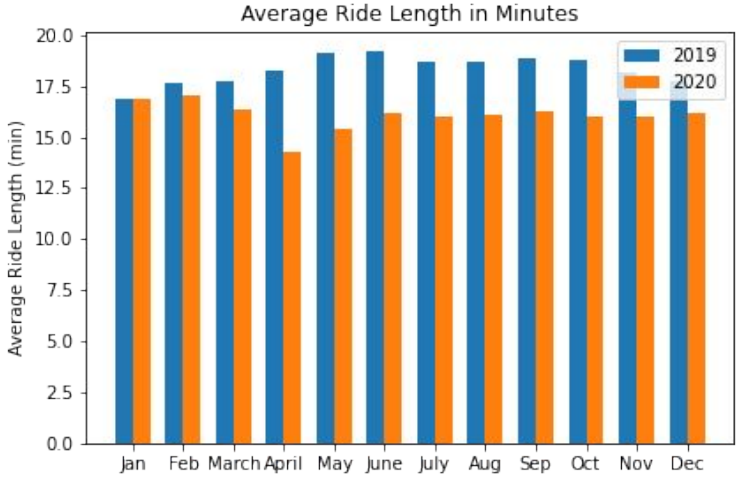
-83.24%

Largest Decrease In Number of Rides was Between April 2019 and April 2020

March	-48.20%
April	-83.24%
May	-80.30%
June	-71.92%
July	-62.58%
August	-62.74%
September	-59.16%
October	-57.94%
November	-67.71%
December	-66.17%

Change in Number of Rides





Average Distance Traveled Vs Average Trip Time

Analysis

Less traffic making it easier to travel longer distances in shorter amounts of time

People were probably trying to use rideshares only when they needed to (for longer distances)

People are creatures of habit, so the distance they’re going to want to be traveling will stay about the same

DIVY

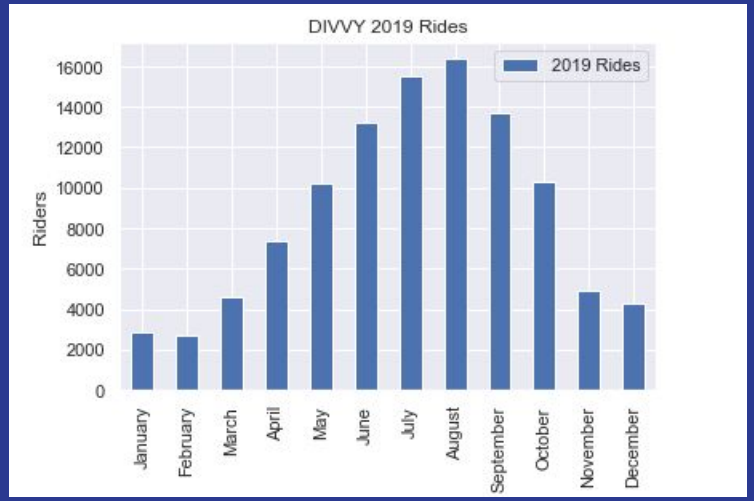
2020 -7.24 %

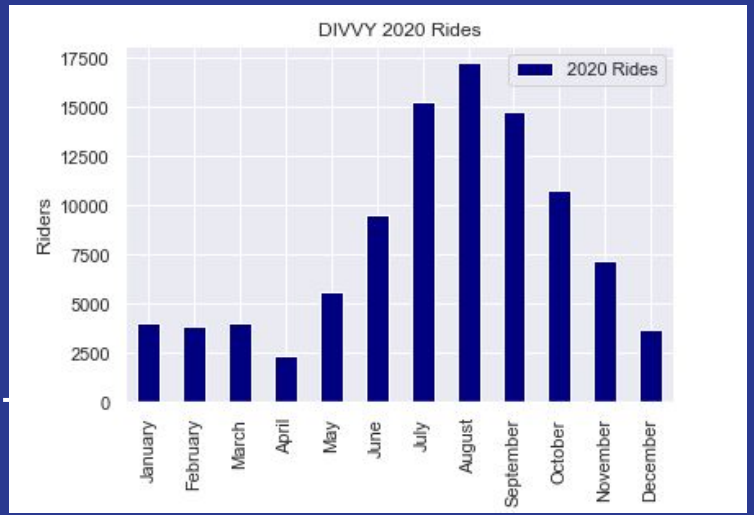
https://www.divvybikes.com/system-data

DIVVY

Ride Volume

2019	3,818,008
2020	3,541,692





divvy.ipynb

Code

Python 3

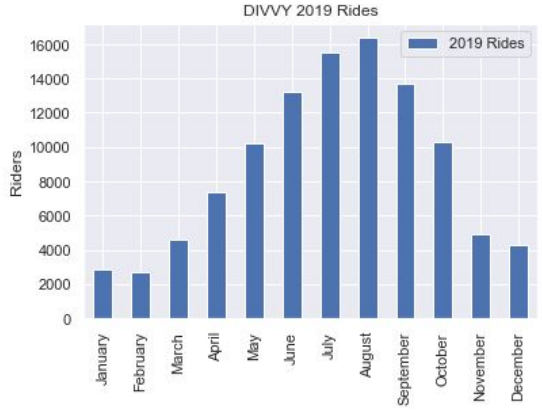
[347]:
```
# Lose some unused colmns
divvy19_df.drop(['end_time','bikeid','usertype','gender','birthyear'],axis=1, inplace=True)

# Organize the trips by month
divvy19_df['start_time'] = pd.to_datetime(divvy19_df['start_time'],errors = 'coerce',format = '%Y-%m-%d %H:%M:%S').dt.strftime("%B")

# Sum the Monthly data, reorg base on month order and ## Find a more perfect solution to this.
div19cnt = pd.value_counts(divvy19_df['start_time'])
div19cnt = div19cnt.to_frame().reset_index()
div19cnt = pd.DataFrame(div19cnt, index=[10, 11, 8, 6, 5, 3,1,0,2,4,7,9])

# Fix Column Names to reflect the data
div19cnt.rename(columns = {'index':'Month','start_time':'2019 Rides'},inplace=True)
div19cnt # .index

# Plot 2019
div19cnt.plot(x='Month',y='2019 Rides',kind='bar',title='DIVVY 2019 Rides',xlabel='Month',ylabel='Riders');
```

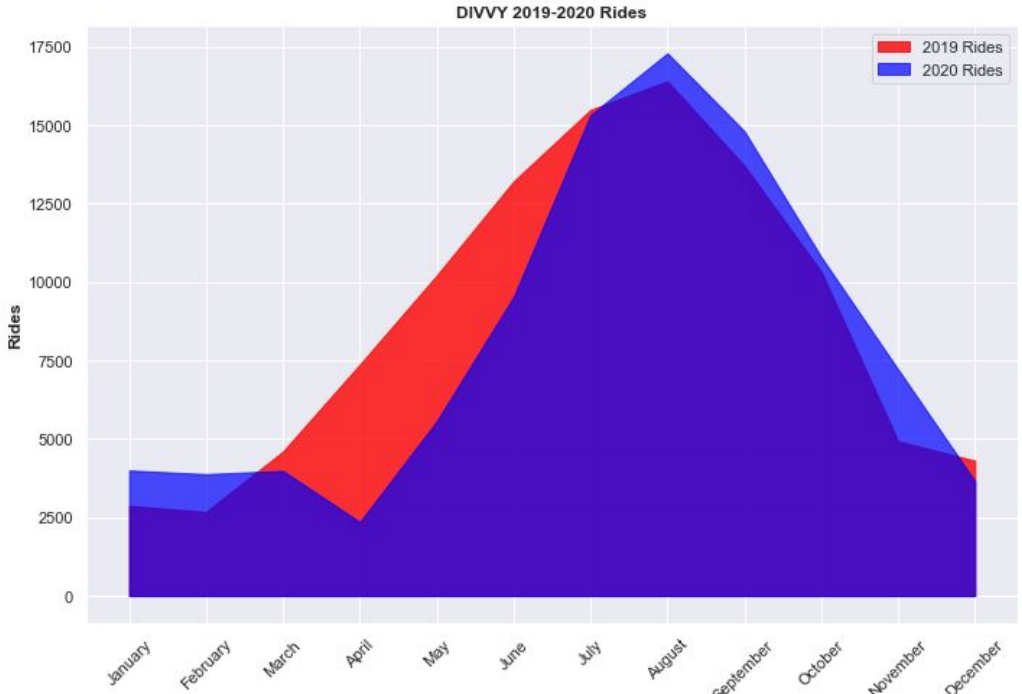


Jupyter Lab

I dropped columns not using at this time. Organized the trips by Month. Summed and plotted the Monthly Trip Rides.

DIVVY

```
[360]: plt.subplots(figsize=(10, 7))
plt.fill_between(div19cnt['Month'], div19cnt['2019 Rides'], color='red', alpha=0.8, label='2019 Rides')
plt.fill_between(div20cnt['Month'], div20cnt['2020 Rides'], color='blue', alpha=0.7, label='2020 Rides')
plt.title("DIVVY 2019-2020 Rides",fontweight='bold')
plt.ylabel('Rides', fontweight='bold')
#plt.xlabel("Month")
plt.ylabel("Rides")
plt.xticks(rotation = 45)
plt.legend()
plt.tight_layout()
plt.savefig('mpltComboArea.png')
plt.show()
```



Jupyter Lab

Combo Area graph created with Matplotlib
Showing the depth of the April Crash and the July+ ridership expansion.

divvy.ipynb

Code

Python

Jupyter Lab

[362]:
```
div19cnt.rename(columns = {'2019 Rides':'Rides'},inplace=True)
div20cnt.rename(columns = {'2020 Rides':'Rides'},inplace=True)
div19cnt.insert(0, 'Year', '2019')
div20cnt.insert(0, 'Year', '2020')
frames = [div19cnt,div20cnt]
divvydata = pd.concat(frames)
#divvydata.head

# divvydata.info()
```

1 Sample T-Test using Paired data 2019-2020
Year over Year Does Not refute that Covid19
Had no effect on Divvy usage

[333]:
```
divvydata.describe()
```

[333]:

	Rides
count	24.000000
mean	8518.166667
std	5031.841911
min	2354.000000
25%	3993.750000
50%	7292.000000
75%	13329.500000
max	17287.000000

[348]:
```
#divvydata['Rides'].value_counts()
```

[335]:
```
# https://www.marsja.se/how-to-use-python-to-perform-a-paired-sample-t-test/
d19=divvydata.query('Year =="2019"')['Rides']
d20=divvydata.query('Year =="2020"')['Rides']

# Python paired sample t-test
ttest_rel(d19, d20)
```

[335]:
```
Ttest_relResult(statistic=0.9019753470116122, pvalue=0.3863939662782875)
```

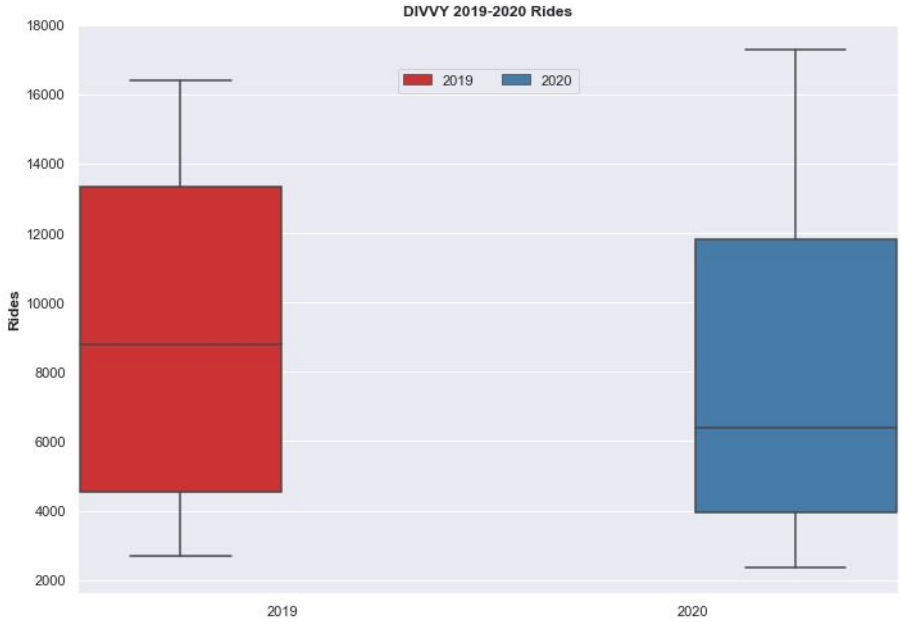
divvy.ipynb

Code

Python

Year

```
[366]: #divvydata.boxplot("Rides", by=["year","Month"], figsize=(20, 10))
sns.set(style="darkgrid")
df = sns.load_dataset('tips')
#sns.boxplot(x="Month", y="Rides", hue="Year", data=divvydata, palette="Set1", width=.66) #, showfliers=False)
plt.subplots(figsize=(10, 7))
sns.boxplot(x="Year", y="Rides", hue="Year", data=divvydata, palette="Set1", width=1) #, showfliers=False)
plt.legend(loc="center", bbox_to_anchor=(0.5, 0.9), ncol=2)
plt.title("DIVVY 2019-2020 Rides",fontweight='bold')
plt.ylabel('Rides', fontweight='bold')
plt.xlabel("")
plt.tight_layout()
plt.savefig('ComboBox.png')
plt.show()
```



Jupyter Lab

Matplotlib and Seaborn used on boxplots of Ridership data. I could not find any way to Get an ‘air-gap’ on the edges of these boxes.

Conclusions

1. Rides involving strangers had a marked decrease versus those involving only individuals, which had an increase.

2. Based on the massive decrease in the number of rides for rideshares and the CTA, many people chose to go nowhere or stay close to home due to COVID.

General Hypothesis **True**:

There is a notable correlation between the onset of COVID-19, its restrictions and transit in Chicago.