Exploratory Data Analysis on MTA Turnstile Data

By:Tarfah Alabbad

Outlines

- BACKSTORY
- AIMS AND OBJECTIVES
- TOOLS
- WORK FLOW
- RESULTS
- CONCLUSION

BACKSTORY

A client reach out to me to do exploratory data analysis because she want to open a booth that sells breakfast, snacks and coffee to go, she wants to set her booth to have a strategic place since COVID 19 pandemic closed her shop and now she wants to know where is the stations that have the most foot traffic after COVID where people mostly will need to take a coffee and breakfast before getting on to the train or after getting off it in the morning time specifically

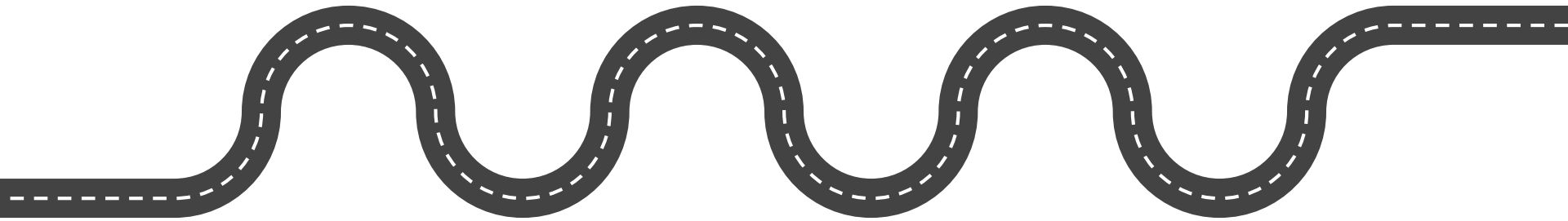
AIMS AND OBJECTIVES

- To compare the entries and exits before and after COVID
- To find the stations that have the most traffic
- To visually present the results

TOOLS

WORK FLOW

1. Uploding MTA dataset from November and December 2019 to November and December 2020
2. Data cleaning
3. Compairing the daily entries and exits before and after COVID
4. Extracting tbe total entries for busiest stations
5. Using sqlalchemey to make queries on data
6. Knowing the top 10 busiest stations in the morning period



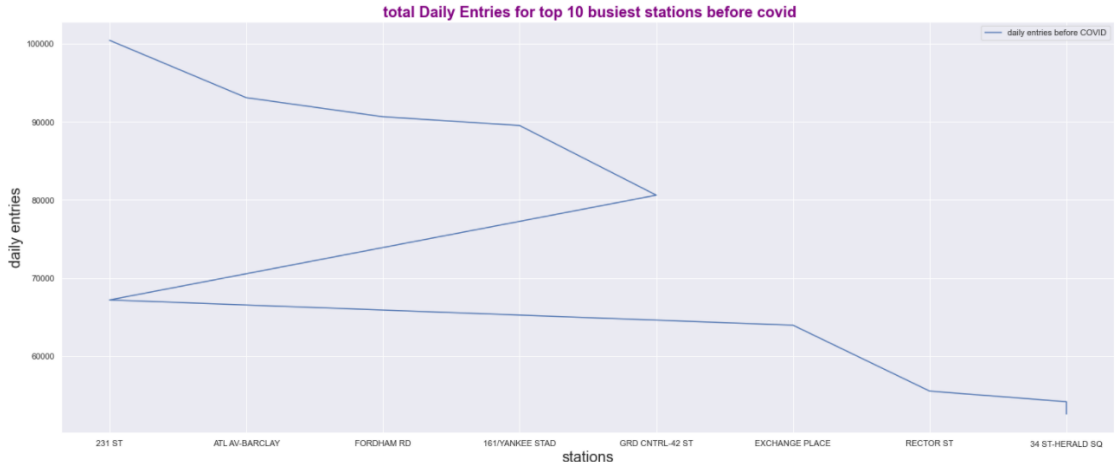
RESULTS

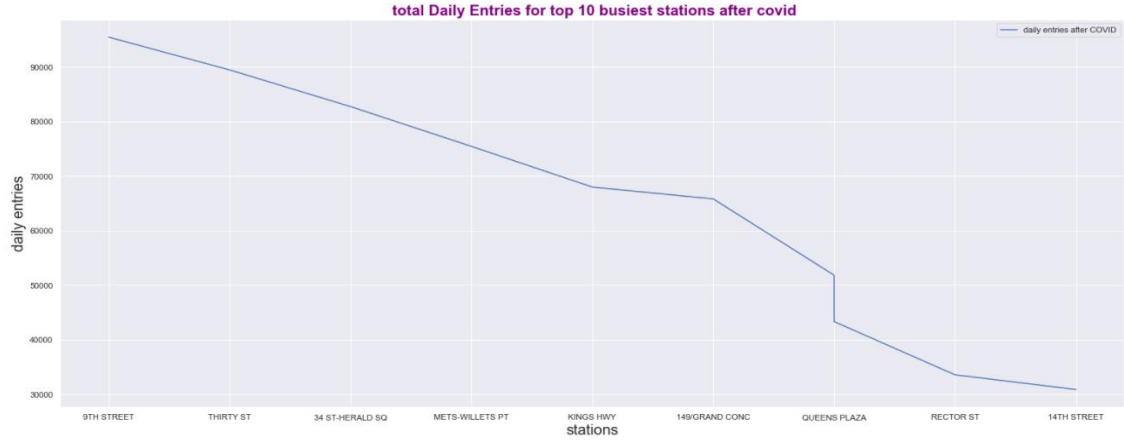
Top 10 busiest stations before COVID:

- 231 ST 10 million
- ATL AV-BARCLAY
- FORDHAM RD
- 161/YANKEE STAD
- GRD CNTRL-42 ST
- 231 ST
- EXCHANGE PLACE
- RECTOR ST
- ST-HERALD SQ
- 34 ST-HERALD SQ

Top 10 busiest stations after COVID:

- 9TH STREET 90 thousand
- THIRTY ST
- 34 ST-HERALD SQ
- METS-WILLETS PT
- KINGS HWY
- 149/GRAND CONC
- QUEENS PLAZA
- QUEENS PLAZA
- RECTOR ST
- 14TH STREET





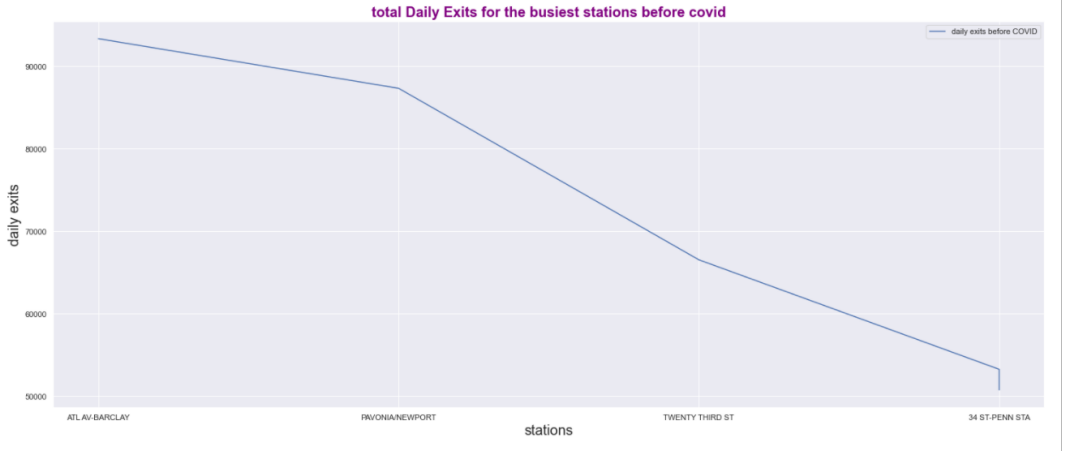
RESULTS

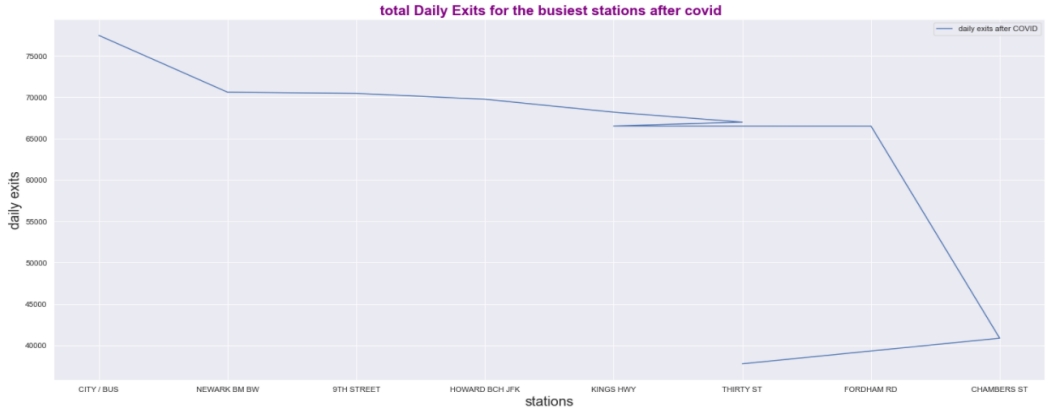
Top 10 busiest stations before COVID:

- ATL AV-BARCLAY 93 thousand
- PAVONIA/NEWPORT
- TWENTY THIRD ST
- 34 ST-PENN STA
- 34 ST-PENN STA
- 34 ST-PENN STA
- 34 ST-PENN STA
- 34 ST-PENN STA
- 34 ST-PENN STA
- 34 ST-PENN STA

Top 10 busiest stations after COVID:

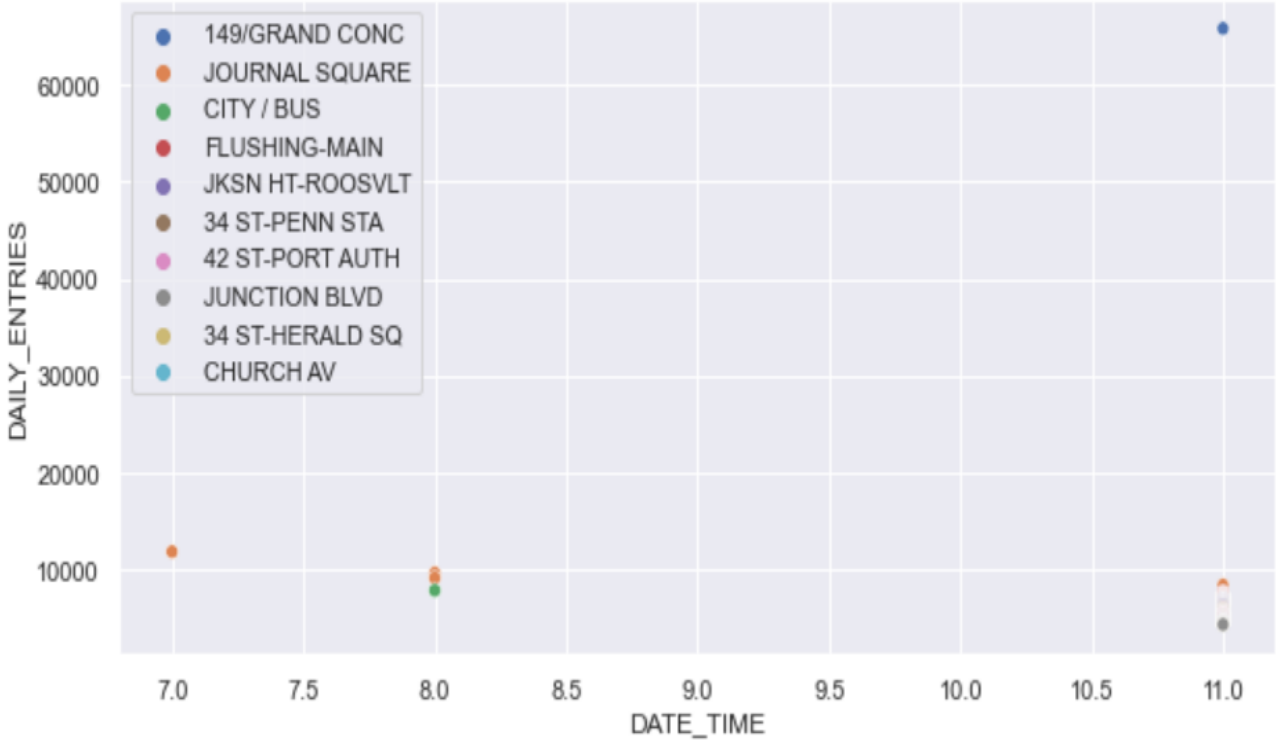
- CITY / BUS 77 thousand
- NEWARK BM BW
- 9TH STREET
- HOWARD BCH JFK
- KINGS HWY
- THIRTY ST
- KINGS HWY
- FORDHAM RD
- CHAMBERS ST
- THIRTY ST





RESULTS

The final result of my EDA is the busiest stations in the morning period



CONCLUSION

In conclusion my final result will help my client to choose the location she want from the busiest stations to open her booth based on my exploratory data analysis on the MTA dataset.

THANK YOU FOR LISTINING!

Any questions?