- Question/Problem statement:

Did the public transportation differ before and after the lockdown that happened in New York city from march 2020 until April 2020?

- Data Description:

The source of the data is the dataset provided by Metis of the metropolitan transportation authority in New York. The duration that I choose is six months before march and six months after April which is the lockdown period.

- Columns:

There are 11 columns, however I will be using five of the existing columns and they are: station, date, time, entries, and exits, and I will add a new one named the people number for the calculation of the number of people (not cumulative number). Down is the table with the column's names, description, and data type.

C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESc	ENTRIES	EXIT
control area	remote station unit	the address for the device of calculating exits and entries	station name	train lines of this station	line originally the station belonged to BMT, IRT, or IND	Date	Time	scheduled audit event (every 4 hours)	cumulative entry	cumulative exit
Object	Object	Object	Object	Object	Object	Object	Object	Object	Integer	Integer

- Added columns: total number of people using the public transportation in a specified date.
- Data size (between 204795 and 217832 rows ,11 columns).
- Tools:
  - Programs: spyder.
  - Libraries: pandas, seaborn, NumPy, SciPy, matplotlib.
  - Functions: calculating the total number of people in a specific data (not cumulative) using pandas.dataframe.diff () function.
  - Plots: scatter plot, distribution plot, bar chart.
- MVP Goal:

The goal is to know whether the public transportation differ before and after the lockdown that happened in New York city from march 2020 until April 2020.