Final Project: The MTA's Capital Programs and Customer Experience.

Principal Investigator: [Mateo Vargas Castillo]
Email: mvc319@stern.nyu.edu (mvc319@stern.nyu.edu)

This project will study how the feedback received by MTA customers has influenced or driven the MTA to allocate their Capital (via Capital Programs) strategically to tackle the current issues. In other words, the research question that will be addressed throughout the project is : "To what extent does the MTA's data recollection influence their Capital Investment Allocation, as a means to improve costumer experience?". To restore, improve, and expand this irreplaceable public asset. The MTA has implemented capital programs between 1982 and 2018. As the MTA continues to improve and expand under its 2010-2014, as well as its 2015-2019 Capital Program. It is accelerating critical improvements to its century-old subway system under the Subway Action Plan, announced in July 2017, which aims at both immediate performance gains and a thoroughly modernized subway system to serve New York in the 21st century.

The project will include three equally important dataframes.

- Capital Allocation Dataframe: The MTA Capital Dashboard data was downloaded from Kaggle.com. This data contains descriptions of each project ran by the MTA from 2009-2019. In order to use this data I will remove duplicates of projects whose budget had been recorded monthly. I can divide the data into 2 different time periods, 2010-2014 and 2015-2019, as a means of analyzing it against Costumer Feedback and Performance Indicators. Will sort the data based on the sum of the original budget associated to each project. I could also sort the capital projects based on their Category and Element descriptions. It is important to note that the data includes projects in different phases (completed, design, construction). I believe in including all of these as it is an indicator of the fields that the MTA is planning to improve in the future.

- Customer feedback Dataframe: The MTA Customer Feedback data was downloaded from Kaggle.com. This data lists the both complaints and commendations from MTA costumers from 2014 to 2018. Each field specifies the Agency, Subject Matter, Subject Detail, Issue Detail, Year, Quarter and Branch/Line/Route of each complaint or commendation. I will sort the data based on the number of complaints, filtered to NYC Subways. The count of complaints are sorted throughout time by Line or Route, and Subject Matter. I will sort both of these fields in descending order based on the count of complaints. The data can be used as a means to understanding whether the MTA is addressing costumer satisfaction as a main factor to their Capital Budgeting. A number of complaints where related to Lines with "No Value", I will eliminate these data points in order to simplify the data.

- Performance indicators Dataframe: MTA Performance Indicator data was downloaded from Kaggle.com. This data lists the monthly values of the MTA Performance Indicators. The data displays the Period, YTD Target, YTD Actual, Monthly Target, Monthly Actual, Desired Change for each indicator. In order to use analyze this data, I will filter the Agency Name to NYC transit and within this agency, and will only take into account those indicators that are related to subways. The values for each indicator are recorded in a monthly basis, and compared to the target value. As means to analyze the MTA performance based on Period Year, I found the average of YTD Actual and YTD Target monthly values. It is important to know that all indicators are percentages (%) except for Total Ridership and Mean Distance Between Failure. The rest of the values have been sorted in descending order, from best performing (e.g. highest percentage of OTP) to worst performing.

I anticipate that the project will have four main sections.

- Basic statistics regarding all three dataframes. An introduction to familiarize ourselves with the data.
- Visualizations of the relationship between Customer Feedback and Performance indicators at a given time period.
- Visualizations of the relationship between Customer Feedback and Capital Allocation at a given time period.
- Visualizations of the relationship between Capital Allocation and Performance indicators at a given time period.

Data Report

Overview: The data behind my project comes from Kaggle.com. As I explained above, I will use three different dataframes in order to find a correlation between Customer Feedback and Capital Allocation.

Important Variables: There are several important variables. These include On-Time Performance per lane, customer feedback, as well as the most important areas of capital investment.

Based on the data that I reviewed for this analysis, I ballparked that the most significant investments were allocated to **Signals & Communications, Passenger Stations, and Subway Cars.**

Access I have downloaded the data and it is currently saved on my laptop. Below I demonstrate that I have the ability to access the data.

Industry Overview

The Metropolitan Transportation Authority (MTA) is North America's largest transportation network, serving a population of 15.3 million people in the 5,000-square-mile area fanning our from New York City throughout Long Island, southeastern New York State, Connecticut.

The MTA network comprises the nation's largest bus fleet and more subway and commuter rail cars than all other U.S. transit systems combined. It provides over 2.6 billion trips each year, accounting for about one-third of the nation's mass transit users and two-thirds its commuter rail passengers. MTA Bridges and Tunnels, which recorded a record 310 million crossings in 2017, carries more vehicles than any other bridge and tunnel authority in the nation.

The MTA's provision of safe, clean, efficient public transportation is the lifeblood of the New York City area, one of the world's major economic hubs. It opens up employment opportunities for millions of area residents, linking them to jobs miles from their homes. It revives old neighborhoods and gives rise to new business corridors. It links millions of residents and visitors to cultural, educational, retail, and civic centers across the region.

The Project

Requisite Packages Below I bring in the packages I need...

In [185]:
```
import pandas as pd # We know this one...
import numpy as np # For performing numerical analysis
import matplotlib.pyplot as plt # Plotting
```

Importing Dataframes Below I will establish the Dataframes in the following order: Capital Allocation, Customer Feedback, and Performance Indicators. These Dataframes will be stored into their respective variables.

In [186]:
```
capital_allocation_pd = pd.read_excel('/Users/Mateo/Desktop/DATA_PROJECT/MTA_Capital_Allocation.xlsx')

customer_feedback_pd = pd.read_excel('/Users/Mateo/Desktop/DATA_PROJECT/MTA_Customer_Feedback1.xlsx')

performance_indicators_pd = pd.read_excel('/Users/Mateo/Desktop/DATA_PROJECT/MTA_Performance_Indicator.xlsx')
```

Must clean Datasets: If you visualize each data set separetly you will be able to tell that there are many fields within the datasets that are of little use for my analysis. Therefore, in order to make this project clearer and more concise I must clean up the datasets.

Capital Allocation

Capital Allocation Dataset: This Dataset contains the Capital Expenditures that the MTA has implemented throughout 2005 to 2019. We will ignore the data the entails capital investments of 2005 throughout 2010 given the nature of the time at which the Customer Feedback Dataset was created. In other words, the Customer Feedback ranges from 2014 to 2018. Given that we are going to analyse two main Capital Allocation Programs, we will merge the two to view the focus areas of the investments. **However**, before accomplishing this, I will clean the General Dataset by deleting elements (columns & rows) that do not serve any purpose to this investigation for the reasons stated above.

In [187]:
```
# Dropping unnecessary columns

capital_allocation_pd = capital_allocation_pd.drop(["Project Number", "L
oad Date", "Plan Revision Keys",
                                                    "Plan Revision Displa
y Keys", "Agency Code", "Category",
                                                    "Element", "Project D
escription", "Scope Objective",
                                                    "Phase", "Original Bu
dget", "Original Budget vs Current Budget",
                                                    "Overall Scope Chang
e", "Original Start Month",
                                                     "Original Start Yea
r", "Current Start Month", "Original Completion Month",
                                                    "Original Completion
 Year", "Current Completion Month",
                                                    "Design Start Month",
"Design Start Year", "Design Completion Month",
                                                    "Design Completion Ye
ar", "Construction Start Month",
                                                    "Construction Start Y
ear", "Construction Completion Month",
                                                    "Construction Complet
ion Year", "Percentage Complete",
                                                    "Current Quarter Fla
g", "Location Indicator"], axis=1)
```

In [188]:
```
# Dropping unnecessary rows
# First, set target rows as index by making the column the index

capital_allocation_pd = capital_allocation_pd.set_index("Capital Plan")

# Dropping rows

capital_allocation_pd = capital_allocation_pd.drop("Capital Plan 2005 -
 2009", axis = 0)
```

It is **important** to note that for simplicity purposes, this project will be focused solely on the **MANHATTAN SUBWAY**.

What does this entail?

Well, this means that we must drop the redundant data that focuses on the MTA's subway services outside of Manhattan as well as the Buses and Depots because these do not affect the subway data

In [189]:
```
# Repeat previous step to drop more unnecessary rows

capital_allocation_pd = capital_allocation_pd.loc[capital_allocation_pd[
"Agency Name"] == "New York City Transit"]

capital_allocation_pd = capital_allocation_pd.loc[capital_allocation_pd[
"Category Description"] != "Buses"]

capital_allocation_pd = capital_allocation_pd.loc[capital_allocation_pd[
"Category Description"] != "Depots"]
```

Final Capital Allocation Dataframe: The Dataframe below depicts the important information that we will use from this specific Dataframe.

We are therefore left with the most important variables in the dataset.

- Capital Plans of 2010 to 2018
- Agency Name (Manhattan focused)
- Category Description (Operational areas for investment)
- Element Description (Specific subcategory of "Category Description")
- Current Budget (allocated in each area)
- Current Start Year
- Current Completion Year

We can view the cleaned dataset below

In [190]:
```
capital_allocation_pd
```

Out[190]:

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2010 - 2014	1	New York City Transit	Subway Cars	Subway Cars ...	2.910900e+08	2010	
Capital Plan 2010 - 2014	1	New York City Transit	Subway Cars	Subway Cars ...	2.723857e+08	2011	
Capital Plan 2010 - 2014	2	New York City Transit	Subway Cars	Subway Cars ...	6.380000e+08	2010	
Capital Plan 2010 - 2014	2	New York City Transit	Subway Cars	Subway Cars ...	7.355614e+08	2012	
Capital Plan 2010 - 2014	3	New York City Transit	Subway Cars	Subway Cars ...	1.100000e+08	2012	
Capital Plan 2010 - 2014	3	New York City Transit	Subway Cars	Subway Cars ...	1.243858e+07	2012	
Capital Plan 2010 - 2014	3	New York City Transit	Subway Cars	Subway Cars ...	0.000000e+00		
Capital Plan 2010 - 2014	4	New York City Transit	Subway Cars	Subway Cars ...	1.000000e+06		
Capital Plan 2010 - 2014	1	New York City Transit	Passenger Stations	Fare Collection ...	7.271000e+07	2010	
Capital Plan 2010 - 2014	1	New York City Transit	Passenger Stations	Fare Collection ...	7.271000e+07	2011	
Capital Plan 2010 - 2014	2	New York City Transit	Passenger Stations	Fare Collection ...	2.310000e+06	2010	

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2010 - 2014	2	New York City Transit	Passenger Stations	Fare Collection ...	2.310000e+06	2011	
Capital Plan 2010 - 2014	5	New York City Transit	Passenger Stations	Fare Collection ...	2.000000e+08	2010	
Capital Plan 2010 - 2014	6	New York City Transit	Passenger Stations	Fare Collection ...	3.000000e+07	2013	
Capital Plan 2010 - 2014	6	New York City Transit	Passenger Stations	Fare Collection ...	3.000000e+07	2013	2014
Capital Plan 2010 - 2014	6	New York City Transit	Passenger Stations	Fare Collection ...	5.245253e+06	2013	2014
Capital Plan 2010 - 2014	7	New York City Transit	Passenger Stations	Fare Collection ...	2.000000e+06	2013	
Capital Plan 2010 - 2014	7	New York City Transit	Passenger Stations	Fare Collection ...	2.000000e+06	2013	
Capital Plan 2010 - 2014	7	New York City Transit	Passenger Stations	Fare Collection ...	3.560891e+06	2015	2016
Capital Plan 2010 - 2014	8	New York City Transit	Passenger Stations	Fare Collection ...	1.997398e+07	2014	2016
Capital Plan 2010 - 2014	9	New York City Transit	Passenger Stations	Fare Collection ...	2.391626e+06	2013	2014
Capital Plan 2010 - 2014	10	New York City Transit	Passenger Stations	Fare Collection ...	7.000000e+05	2017	
Capital Plan 2010 - 2014	11	New York City Transit	Passenger Stations	Fare Collection ...	5.000000e+07	2015	2017

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2010 - 2014	1	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	2.472380e+07	2011	
Capital Plan 2010 - 2014	3	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	1.427000e+07	2012	
Capital Plan 2010 - 2014	4	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	4.837000e+07	2012	
Capital Plan 2010 - 2014	5	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	6.113000e+07	2013	
Capital Plan 2010 - 2014	5	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	2.413774e+07	2014	
Capital Plan 2010 - 2014	6	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	2.076405e+07		
Capital Plan 2010 - 2014	6	New York City Transit	Passenger Stations	Station Escalators/Elevators ...	1.384270e+07	2019	
...	...	...	...	...	...	...	...
Capital Plan 2015 - 2019	26	New York City Transit	Misc./Emergency	Engineering Services ...	7.500000e+04		
Capital Plan 2015 - 2019	1	New York City Transit	Misc./Emergency	Environmental And Safety ...	2.628065e+07	2017	2021
Capital Plan 2015 - 2019	3	New York City Transit	Misc./Emergency	Environmental And Safety ...	1.201697e+07	2017	2020
Capital Plan 2015 - 2019	5	New York City Transit	Misc./Emergency	Environmental And Safety ...	6.500000e+06	2016	2019

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2015 - 2019	6	New York City Transit	Misc./Emergency	Environmental And Safety ...	1.149473e+07	2018	2023
Capital Plan 2015 - 2019	1	New York City Transit	Misc./Emergency	Employee Facilities ...	2.786907e+07	2019	
Capital Plan 2015 - 2019	3	New York City Transit	Misc./Emergency	Employee Facilities ...	1.531964e+07		
Capital Plan 2015 - 2019	4	New York City Transit	Misc./Emergency	Employee Facilities ...	9.739980e+06	2015	2017
Capital Plan 2015 - 2019	5	New York City Transit	Misc./Emergency	Employee Facilities ...	4.869990e+06	2016	2017
Capital Plan 2015 - 2019	5	New York City Transit	Misc./Emergency	Employee Facilities ...	4.869990e+06	2016	2017
Capital Plan 2015 - 2019	6	New York City Transit	Misc./Emergency	Employee Facilities ...	4.085238e+06		
Capital Plan 2015 - 2019	6	New York City Transit	Misc./Emergency	Employee Facilities ...	4.085238e+06		
Capital Plan 2015 - 2019	7	New York City Transit	Misc./Emergency	Employee Facilities ...	5.106547e+06		
Capital Plan 2015 - 2019	8	New York City Transit	Misc./Emergency	Employee Facilities ...	2.680236e+06		
Capital Plan 2015 - 2019	15	New York City Transit	Misc./Emergency	Employee Facilities ...	2.500000e+07		
Capital Plan 2015 - 2019	17	New York City Transit	Misc./Emergency	Employee Facilities ...	5.230000e+06	2017	

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2015 - 2019	17	New York City Transit	Misc./Emergency	Employee Facilities ...	5.230000e+06	2017	
Capital Plan 2015 - 2019	18	New York City Transit	Misc./Emergency	Employee Facilities ...	8.300000e+06	2019	
Capital Plan 2015 - 2019	21	New York City Transit	Misc./Emergency	Employee Facilities ...	9.948947e+05	2018	
Capital Plan 2015 - 2019	22	New York City Transit	Misc./Emergency	Employee Facilities ...	1.500000e+06		
Capital Plan 2015 - 2019	23	New York City Transit	Misc./Emergency	Employee Facilities ...	5.000000e+06	2018	2020
Capital Plan 2015 - 2019	24	New York City Transit	Misc./Emergency	Employee Facilities ...	5.000000e+06		
Capital Plan 2015 - 2019	25	New York City Transit	Misc./Emergency	Employee Facilities ...	5.000000e+06		
Capital Plan 2015 - 2019	27	New York City Transit	Misc./Emergency	Employee Facilities ...	4.080980e+05	2018	2020
Capital Plan 2015 - 2019	28	New York City Transit	Misc./Emergency	Employee Facilities ...	5.000000e+06		
Capital Plan 2015 - 2019	29	New York City Transit	Misc./Emergency	Employee Facilities ...	5.000000e+06		
Capital Plan 2015 - 2019	30	New York City Transit	Misc./Emergency	Employee Facilities ...	1.500000e+07		
Capital Plan 2015 - 2019	31	New York City Transit	Misc./Emergency	Employee Facilities ...	1.000000e+07		

	Project	Agency Name	Category Description	Element Description	Current Budget	Current Start Year	Current Completion Year
Capital Plan							
Capital Plan 2015 - 2019	32	New York City Transit	Misc./Emergency	Employee Facilities ...	2.500000e+07		
Capital Plan 2015 - 2019	33	New York City Transit	Misc./Emergency	Employee Facilities ...	1.500000e+07		

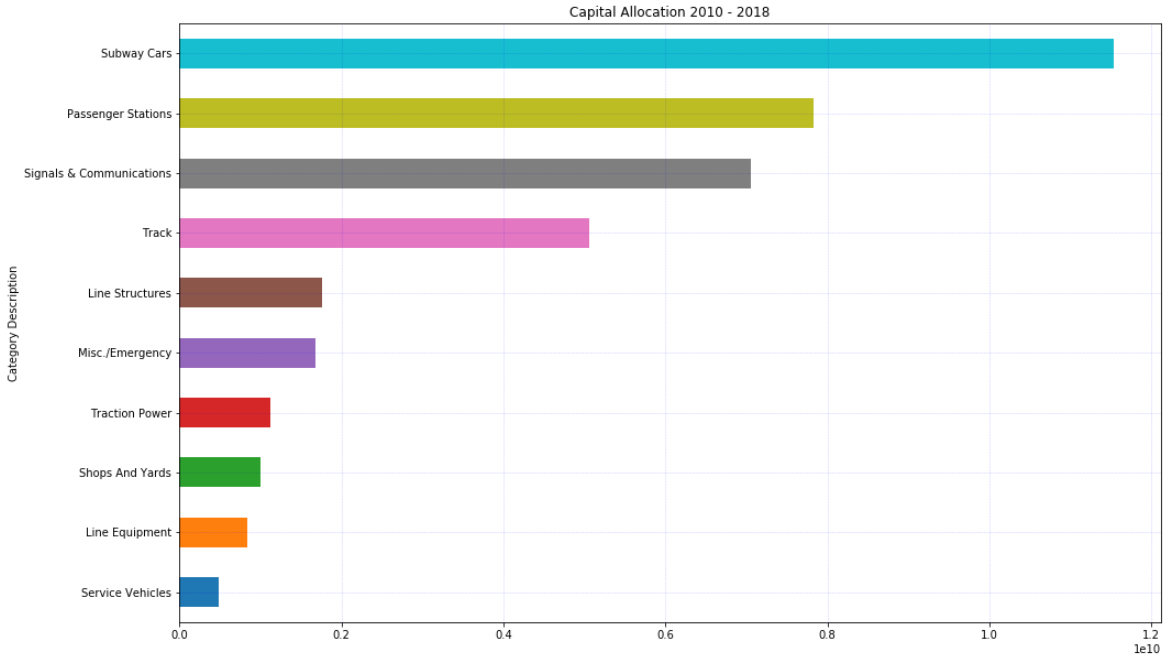
1423 rows × 7 columns

Now we want to be able to visualize our data in a simple and comprehensive manner.

Given that we want to understand how much capital is being allocated where, I believe the most useful way to illustrate this is through a bar chart separated by category.

Below is the code to come up with the bar chart.

In [191]:
```
capital_allocation_pd.groupby("Category Description")["Current Budget"].
sum().sort_values().plot(kind='barh', title='Capital Allocation 2010 - 2
018', figsize=(16,10)).grid(linestyle='--', linewidth='0.5', color='blu
e', alpha = 0.2)

# Start by grouping the data by Category AND Budget.
# This will identify the relevant data we need.
# Use sum() as we want to know the total capital allocated per category
# Use sort_values() to specify where the values will be going
# Use plot() to illustrate the bar chart horizontally (barh)
# Use grid() to introduce a grid on the chart for an easier viewing of d
ata.
```



The bar chart above illustrate the Capital Expenditure allocation of the MTA's two most recent projects. Based on the data that we have used for this analysis, I confirm that the most significant investments were allocated to Signals & Communications, Passenger Stations, and Subway Cars.

Signals & Communications - This area tackles the MTA staff 's, such as drivers and logistics analysts, effectiveness to communicate while they are underground. If a certain problem arises at any given moment in time, all scheduled trains must adapt accordingly. Failure to communicate such issues may result in more significant delays.

Passenger Stations - This area is concerned with the quality, cleanliness, and the consumer-friendly standards of the Subway Stations. It is a popular belief that the NYC Subway can be, at times, unappealing to everyday customers and even dangerous due to a variety of reasons.

Subway Cars - This area regards the improvement & renovation of the Subway Wagons. In comparison to other major metropolitan hubs around the world such as London or Madrid, the NYC Subway Cars are significantly older. Investments in newer Wagons may increase both efficiencies and customer comfort.

Next we will analyze the Customer Feedback Dataset

Research Question

To what extent does the MTA's data recollection influence their Capital Investment Allocation, as a means to improve costumer experience?

Customer Feedback

Customer Feedback Dataframe: This Dataset is simpler than the one before. The Customer feedback is classified in two different ways. The subject matter (the area that is being complained about), and the metro line. We will concern ourselves more with the latter. But first, we must clean up the data so that it is more concise.

In [192]:
```
# Dropping columns

customer_feedback_pd = customer_feedback_pd.drop(["Subject Detail", "Issue Detail", "Quarter"], axis=1)
```

In [193]:
```
# Dropping rows that do not concern with ONLY SUBWAY matters

customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Agency"] == "Subways"]
```

In [194]:
```
# Set the year as the index for viewing purposes

customer_feedback_pd = customer_feedback_pd.set_index("Year")
```

Final Customer Feedback Dataframe: The Dataframe below depicts the important information that we will use from this specific Dataframe.

We are therefore left with the most important variables in the dataset.

- Year
- Agency (Subway focused)
- Commendation or complaint
- Subject Matter (Area being complained about)
- Branch/Line/Route (Manhattan specific lines)

We can view the cleaned dataset below

In [195]:
```
customer_feedback_pd
```

Out[195]:

Year	Agency	Commendation or Complaint	Subject Matter	Branch/Line/Route
2015.0	Subways	Complaint	Customer	Q
2015.0	Subways	Complaint	Customer	3
2015.0	Subways	Complaint	Customer	M
2015.0	Subways	Complaint	Customer	F
2014.0	Subways	Complaint	Employees	C
2014.0	Subways	Complaint	Employees	7
2014.0	Subways	Complaint	Employees	Q
2014.0	Subways	Complaint	Employees	N
2014.0	Subways	Complaint	Employees	J
2014.0	Subways	Complaint	Employees	A
2014.0	Subways	Complaint	Employees	N
2014.0	Subways	Complaint	Employees	5
2014.0	Subways	Complaint	Employees	A
2015.0	Subways	Complaint	Customer	G
2015.0	Subways	Complaint	Customer	5
2015.0	Subways	Complaint	Customer	A
2015.0	Subways	Complaint	Customer	J
2015.0	Subways	Complaint	Customer	3
2015.0	Subways	Complaint	Customer	E
2015.0	Subways	Complaint	Customer	B
2015.0	Subways	Complaint	Employees	A
2015.0	Subways	Complaint	Employees	7
2015.0	Subways	Complaint	Employees	6
2015.0	Subways	Complaint	Employees	Q
2015.0	Subways	Complaint	Employees	D
2015.0	Subways	Complaint	Employees	4
2014.0	Subways	Commendation	Employees	A
2015.0	Subways	Complaint	Customer	C
2015.0	Subways	Complaint	Customer	E
2015.0	Subways	Complaint	Customer	A
...	...	...	...	...
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	R
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	7

Year	Agency	Commendation or Complaint	Subject Matter	Branch/Line/Route
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	D
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	Q
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	F
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	N
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	7
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	A
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	B
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	B
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	B
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	E
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	Q
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	7
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	7
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	E
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	No Value
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	Q
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	A
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	R
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	N
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	F
2018.0	Subways	Complaint	Travel Disruption / Trip Problem	7

80897 rows × 4 columns

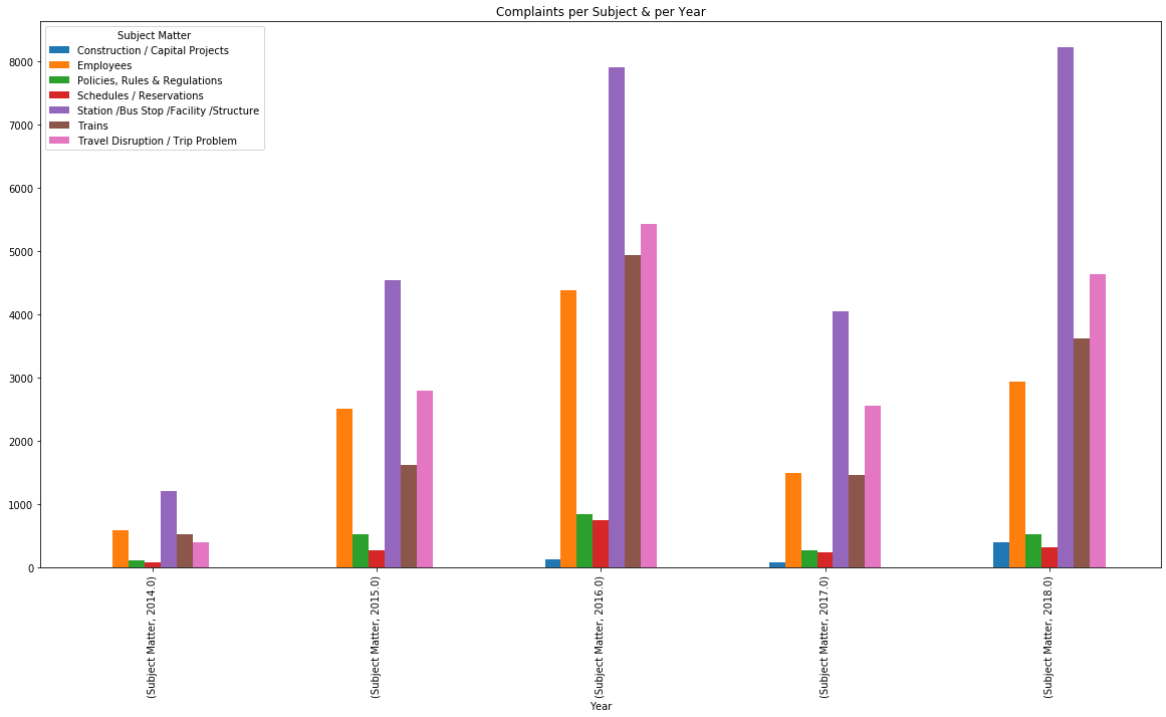
Now we want to visualize the data in the simplest manner.

A vertical bar chart whereby the number of complaints per category per year will allow us to see the complaints trend over time

Below is the code to illustrate this

In [196]:
```
# Focused on only complaints
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Commendation or Complaint"] == "Complaint"]

# Drop reduntant complaint areas given that we already know which areas
# the MTA is spending most of its capital in.
# Therefore we want to see if these areas are also the most complained about.
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "Public Hearing"]
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "Customer"]
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "MTA Agency Cars / Trucks"]
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "MetroCard/Tickets/E-Zpass & Tolls"]
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "Telephone / Website / Mobile Apps"]
customer_feedback_pd = customer_feedback_pd.loc[customer_feedback_pd["Subject Matter"] != "Reasonable Modification"]

# Group the data by year and by subject matter because it is want we want to look at.
ready = customer_feedback_pd.groupby(["Year"])["Subject Matter"].value_counts()

# This bar chart was a little complicated to plot
# Given that we want to view data by year and by category, framing it and unstacking it was necesary.
ready.to_frame().unstack(level=0).T.plot(kind='bar', title='Complaints per Subject & per Year', figsize=(20,10)).set_xlabel("Year")
```

Out[196]: Text(0.5, 0, 'Year')



Our aim is to find a correlation between the data that is publically available on the MTA's official website and their trend of Capital Allocation over time. The bar chart above illustrates, by category, the level of complaints given by customers to the MTA over a period of 5 years. The data embodies an interesting facet of the analysis. The first facet that the chart allows us to see a direct link to the Capital Programs that were highlighted above where the three categories with the highest levels of complaints are the ones where most Capital is being allocated through the Expenditure Programs.

Station / Facilities / Structure ----> Passenger Stations

Travel Disruption ----> Signals & Communications

Trains ----> Subway Cars

Nevertheless, while this may appear to be a positive outcome of the MTA's consideration of their data collected, the chart reveals a counter-intuitive facet to the investigation. It seems essential to note that the levels of complaints have increased over time despite their heavy allocation of Capital Investment in these areas. Why?

The issue - Capital Programs fail to improve the significant issues that affect both Efficiency & Customer Experience

The belief is simple; the more capital that one allocates to a specific area within the industry, the larger the improvement on deficiencies of that specific field. However, the data proves that this assumption does not hold in practice. If it did, the dataset would display an improving trend regarding the efficiency of such areas. For now, our only point of reference has been the MTA's customer feedback. It has helped us deduce that the investments have not achieved their primary objective. We continue our analysis by digging deeper into a more representative dataset of efficiency: performance indicators.

Performance Indicator - On-Time Performance

Performance Indicators Dataframe: This dataframe is a little more complicated than the other ones. The issue is that we want to look at the data based on three different indicators: mean distance between failure, on time performance (OTP), and total ridership, for each line. The problem is that OTP is linked to a specific metro line. As we lack a column that specifies this metro line, it is difficult to organize the data based on OTP per metro line over time. Looking into how we will further analyse the data in the project, we will be looking at each OTP separately and graphing it. But before that, let's take a look at the data and clean it up.

In [197]:
```
# set new index
performance_indicators_pd = performance_indicators_pd.set_index("Indicat
or Name")

# drop reduntant variables
performance_indicators_pd = performance_indicators_pd.drop(["Indicator S
equence", "Parent Sequence", "Description",
                                                            "Category",
"Frequency", "Desired Change",
                                                            "Indicator Un
it", "Decimal Places", "Period Month",
                                                            "Monthly Targ
et", "Monthly Actual", "Period"], axis=1)
```

In [198]:
```
# Focus on Manhattan only
performance_indicators_pd = performance_indicators_pd.loc[performance_in
dicators_pd["Agency Name"] == "NYC Transit"]
```

We are therefore left with the most important variables in the dataset.

- Indicator Name (Which performance indicator)
- Agency (Manhattan focused)
- Period Year
- YTD Target
- YTD Actual
- Difference Target and Actual

We can view the cleaned dataset below

In [199]:
```
performance_indicators_pd
```

Out[199]:

Indicator Name	Agency Name	Period Year	YTD Target	YTD Actual	Difference Target and Actual
Mean Distance Between Failures - Subways	NYC Transit	2008	148244.0	148244.00	0.00
Mean Distance Between Failures - Subways	NYC Transit	2008	148476.0	148476.00	0.00
Mean Distance Between Failures - Subways	NYC Transit	2008	146941.0	146941.00	0.00
Mean Distance Between Failures - Subways	NYC Transit	2008	146277.0	146253.00	24.00
Mean Distance Between Failures - Subways	NYC Transit	2008	144999.0	142961.00	2038.00
Mean Distance Between Failures - Subways	NYC Transit	2008	144987.0	141011.00	3976.00
Mean Distance Between Failures - Subways	NYC Transit	2008	146516.0	139628.00	6888.00
Mean Distance Between Failures - Subways	NYC Transit	2008	149567.0	138836.00	10731.00
Mean Distance Between Failures - Subways	NYC Transit	2008	150155.0	136946.00	13209.00
Mean Distance Between Failures - Subways	NYC Transit	2008	152392.0	137332.00	15060.00
Mean Distance Between Failures - Subways	NYC Transit	2008	152365.0	136064.00	16301.00
Mean Distance Between Failures - Subways	NYC Transit	2008	155000.0	134795.00	20205.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	133546.00	11454.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	133325.00	11675.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	133711.00	11289.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	133021.00	11979.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	136152.00	8848.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	138972.00	6028.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	139897.00	5103.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	139889.00	5111.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	142538.00	2462.00

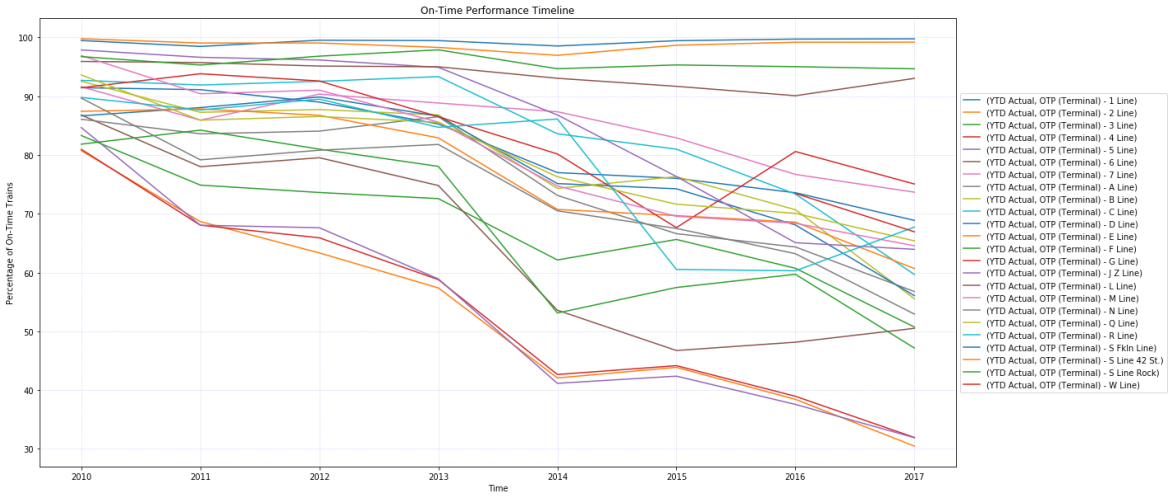
Indicator Name	Agency Name	Period Year	YTD Target	YTD Actual	Difference Target and Actual
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	145170.00	-170.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	148417.00	-3417.00
Mean Distance Between Failures - Subways	NYC Transit	2009	145000.0	153201.00	-8201.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	157233.00	-2233.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	160839.00	-5839.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	165396.00	-10396.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	170182.00	-15182.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	170314.00	-15314.00
Mean Distance Between Failures - Subways	NYC Transit	2010	155000.0	166263.00	-11263.00
...	...	...	...	...	...
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	82.10	-82.10
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	82.70	-82.70
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.10	-83.10
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.30	-83.30
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.40	-83.40
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.60	-83.60
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.70	-83.70
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.60	-83.60
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	83.90	-83.90
Subway Wait Assessment - W Line	NYC Transit	2009	0.0	84.00	-84.00
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	84.30	-84.30
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	84.40	-84.40

Indicator Name	Agency Name	Period Year	YTD Target	YTD Actual	Difference Target and Actual
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	83.90	-83.90
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	83.80	-83.80
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	83.60	-83.60
Subway Wait Assessment - W Line	NYC Transit	2010	0.0	83.70	-83.70
Subway Wait Assessment - W Line	NYC Transit	2016	80.7	72.00	8.70
Subway Wait Assessment - W Line	NYC Transit	2016	80.7	71.65	9.05
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	68.50	12.20
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	69.75	10.95
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	68.87	11.83
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	69.13	11.57
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	69.36	11.34
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	69.95	10.75
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	70.79	9.91
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	71.08	9.62
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	71.41	9.29
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	71.29	9.41
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	71.13	9.57
Subway Wait Assessment - W Line	NYC Transit	2017	80.7	71.00	9.70

9231 rows × 5 columns

In [200]:
```
# Now must filter for appropiate time frame and focus on ON TIME PERFORMANCE
otpOnly = performance_indicators_pd.loc[performance_indicators_pd["Period Year"] > 2009].filter(like='OTP', axis=0).drop(["Agency Name", "YTD Target", "Difference Target and Actual "] , axis=1)
```

In [201]:
```
# Must group by indicator name and period year
# Find average on time performance per line and unstack as we are playing with many variables
otpOnly = otpOnly.groupby(['Indicator Name', 'Period Year']).mean().unstack(level=0)

# Plot line graph
otpGraph = otpOnly.plot(kind='line', figsize=(20,10), title='On-Time Performance Timeline')

# Move legend to the side
otpGraph.legend(loc='center left', bbox_to_anchor=(1, 0.5))

# Set x label
otpGraph.set_xlabel("Time")

# Set y label
otpGraph.set_ylabel("Percentage of On-Time Trains")

# Set grid for viewing purposes
otpGraph.grid(linestyle='--', linewidth='0.5', color='blue', alpha = 0.2)
```



The graph above illustrates a timeline on which we can observe the On-Time Performance (OTP) for every subway line in the NYC transit system. On-Time Performance is a measure of the percentage of trains making all the scheduled station stops arriving at the destination terminal on time, early, or no more than five minutes late. Just as the segments of investments apply to the main areas of customer feedback, the same investment plans are necessary in order to make the subway system more efficient and improve OTP figures. Yet again, it can be observed through the graph, over its timeline and the time of the MTA's capital programs, that OTP figures keep falling. Why?

Analysis

As it can be seen, the MTA seems to focus its investment plans relative to its data collection on performance indicators and customer feedback. Indeed, It seems that the MTA was incentivized to invest more prominently on passenger stations, subway cars, and signals and communications, because their data outlines more heavily the mass-delays within the subway system. But were the MTA truly is successful in resolving the subway system's underperformance by reducing delays? To help the MTA foresee a better future performance, we chose to firstly determine wether or not there is an underlining correlation, between customer complaints and the On-Time Performance of subway lines. In doing so, we hope to show how relatable the MTA's investment choices are to customer feedback and performance indicators. Yet, while there may be an underlining relationship between all, capital programs, customer feedback, and performance indicators, have the MTA's Capital Programs actually brought improvement? How effective were the capital program's efforts in addressing the subway system's delays?

Correlation justifies investment choices

In order to take a deeper look to the correlation. I ran a regression of complaints vs on time performance on Excel. The results illustrated a medium-strong negative correlation of -0.652991 between OTP and customer complaints. For every increase in customer complaints, we observe a proportional fall in NYC subway's OTP. The deviation to a perfect correlation may be caused by the relatively vague indicator names in customer complaints – indicator names apply to a range of complaints that may not consistently relate to OTP.

To my belief, the two correlated variables are at the source of the incentive to proceed with both MTA Capital Programs. So, the choice to focus investments on passenger stations, subway cars, and signals & communications, is an attempt at addressing the main areas of customer complaints and OTP rates. For instance, many signals are poorly maintained or misconfigured, triggering emergency breaks up to 10 to 15 miles below the listed limit and causing train delays. Similarly, it is only by renovating and updating subway cars and stations that the subway system can become more efficient, and that both customer feedback and performance indicators can improve. Yet, while this graph successfully demonstrates the relation between all, capital investment plans, customer feedbacks, and OTP rates, the efforts of the MTA Capital Plans seem to fail. For every year, both variables worsen with proportions, as OTP (in blue) falls, customer complaints (in orange) increase. Certainly, the MTA correctly determined the need for heavier investments in the 3 previously discussed categories, yet the MTA's efforts seem to only worsen the situation. Clearly, this means that capital is inefficiently being allocated.

A call for specifity

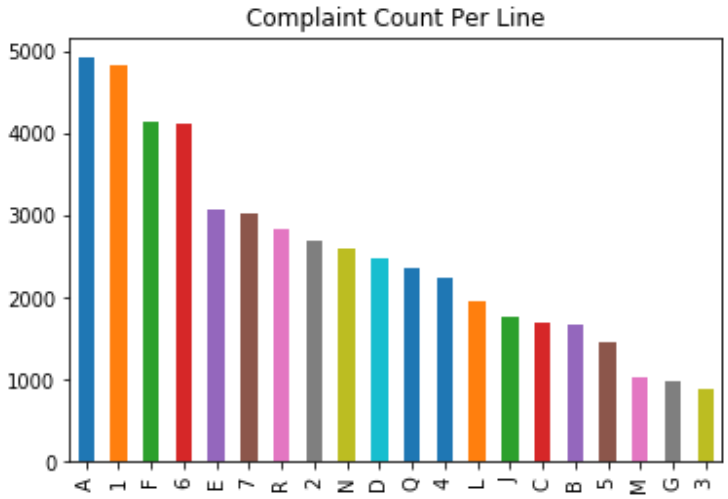
We want to look further into the complaints per metro line in NYC. This will give us a larger insight regarding the correlation.

I will plot a bar chart whereby I count the number of complaints per metro line.

The code is below

In [202]:
```
# Drop no value rows as they are redundant
customer_feedback_pd1 = customer_feedback_pd.loc[customer_feedback_pd["Branch/Line/Route"] != "No Value"]

# Drop
customer_feedback_pd1 = customer_feedback_pd1.drop(['Agency', 'Commendation or Complaint', 'Subject Matter'], axis=1)

# Plot the top 20 lines with most complaints
customer_feedback_pd1['Branch/Line/Route'].value_counts().head(20).plot(kind='bar', title='Complaint Count Per Line')
```

Out[202]: <matplotlib.axes._subplots.AxesSubplot at 0x12f36b2e8>



Of course, given the above described correlation between customer complaints and OTP, we can observe that the same subway lanes with worst OTP figures have also the most costumer complaints. In this bar chart, subway lines are graphed against one an other with their number of customer complaints as a mean of comparison. The 10 lines with worst OTP figures are also the ones that are granted most complaints by customers. For example, across both graphs, we observe that both subway lines, 6 and F, are heavily underperforming relative to their surrounding lanes. Again, while these graph support the previously described correlation, they both suggest the ineffectiveness in capital allocation, because figures are only worsening

Summary

After looking at the data for both Count of Complaints and On-Time Performance of each line over time, I discovered that worse performing lines match high complaint counts. So, why do both complaints and performance show a negative trend over time if capital is being directed towards the right subject matters?

My hypothesis was based on the idea that complaints are directly correlated to the performance of each line. Under this theory I wanted to explore the capital investments in passenger stations, the main factor to reduce complaints, and signal and communications, the main factor affecting on-time performance, by observing the geographical investment trends. I believe that while the MTA is correctly investing in the renewal/rehabilitation of Passenger Stations with highest complaint counts, it is not investing in improving the on-time performance of lines as a means to improve customer experience.

Data Sources

MTA Performance Indicator Dataset

https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#metropolitan-transportation-authority-mta-performance-indicators-per-agency-beginning-2008.csv (https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#metropolitan-transportation-authority-mta-performance-indicators-per-agency-beginning-2008.csv)

MTA Customer Feedback Dataset

https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#mta-customer-feedback-data-beginning-2014.csv (https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#mta-customer-feedback-data-beginning-2014.csv)

MTA Capital Dashboard Dataset

https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#metropolitan-transportation-authority-mta-capital-dashboard-agencies-summary-all-load-dates.csv (https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data#metropolitan-transportation-authority-mta-capital-dashboard-agencies-summary-all-load-dates.csv)

MTA General Information

http://web.mta.info/mta/network.htm (http://web.mta.info/mta/network.htm)