**Student: Lester Mcpherson**

**Customer Service Requests Analysis.**

Course-end Project 1

DESCRIPTION

You've been asked to perform data analysis of service request (311) calls from New York City. You've also been asked to utilize data wrangling techniques to understand the pattern in the data and visualize the major types of complaints.

**Note:**Download **311-service-requests-nyc.zip** file using the link given in the **Customer Service Requests Analysis** project problem statement and extract the **311\_Service\_Requests\_from\_2010\_to\_Present.csv**file

**Perform the following steps:**

1. Understand the dataset:
2. Identify the shape of the dataset
3. Identify variables with null values

2. Perform basic data exploratory analysis:

1. Utilize missing value treatment
2. Analyze the date column and remove the entries if it has an incorrect timeline
   1. Draw a frequency plot for city-wise complaints
   2. Draw scatter and hexbin plots for complaint concentration across Brooklyn

3. Find major types of complaints:

1. Plot a bar graph of count vs. complaint types
2. Find the top 10 types of complaints
3. Display the types of complaints in each city in a separate dataset

4. Visualize the major types of complaints in each city

5. Check if the average response time across various types of complaints

**To download the detailed probelm statement click here -**[https://cfls5.simplicdn.net/frontend/images/Download.png](https://lms.simplilearn.com/user/project/download-attachment?file=1669035134_project_02_customer_service_requests_analysis_caltech.pdf)

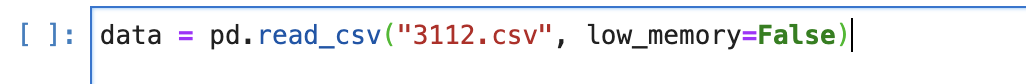
**The dataset is available under course resource section in the Self-Learning Module.**

Solution Approach:

* Understand the dataset:
  1. To identify variables with null values, we would need to use a function, such as .isnull() or .isna() in pandas library, to check for missing values in the dataset.
* Perform basic data exploratory analysis:
  1. To utilize missing value treatment, we would need to decide on an appropriate method for handling missing values, such as dropping the rows with missing values, imputing the missing values with a specific value or using a regression model to predict the missing values.
  2. Analyze the date column and remove the entries if it has an incorrect timeline, you would need to look for any invalid date entries, such as invalid format or a date that falls outside of the range of the dataset and remove them.
  3. To draw a frequency plot for city-wise complaints, we would need to group the dataset by city and then plot the number of complaints for each city.
  4. To draw scatter and hexbin plots for complaint concentration across Brooklyn, we would need to filter the dataset to only include complaints from Brooklyn and then plot the data using scatter and hexbin plots to see the concentration of complaints across the area.
* Find major types of complaints:
  1. To plot a bar graph of count vs. complaint types, you would need to group the dataset by complaint type and then plot the number of complaints for each type using a bar graph.
  2. To find the top 10 types of complaints, you would need to group the dataset by complaint type and then sort the complaints in descending order and select the top 10.
  3. To display the types of complaints in each city in a separate dataset, you would need to group the dataset by both city and complaint type, and then create a new dataset or a pivot table displaying the types of complaints for each city.
  4. Visualize the major types of complaints in each city, you would need to create different plots and graphs, such as bar plots, stacked bar plots, map, etc to visualize the major types of complaints in each city.

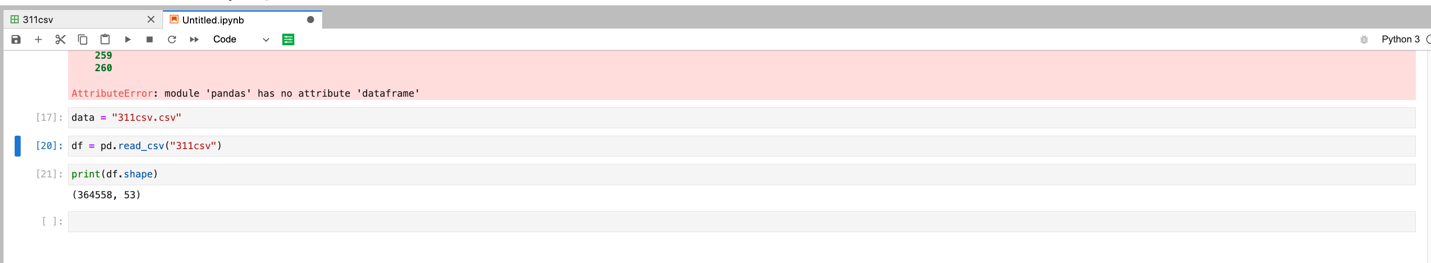
First, we get the environment setup:

Next, we’ll create a dataframe so we can identify the shape of the dataset in 311:



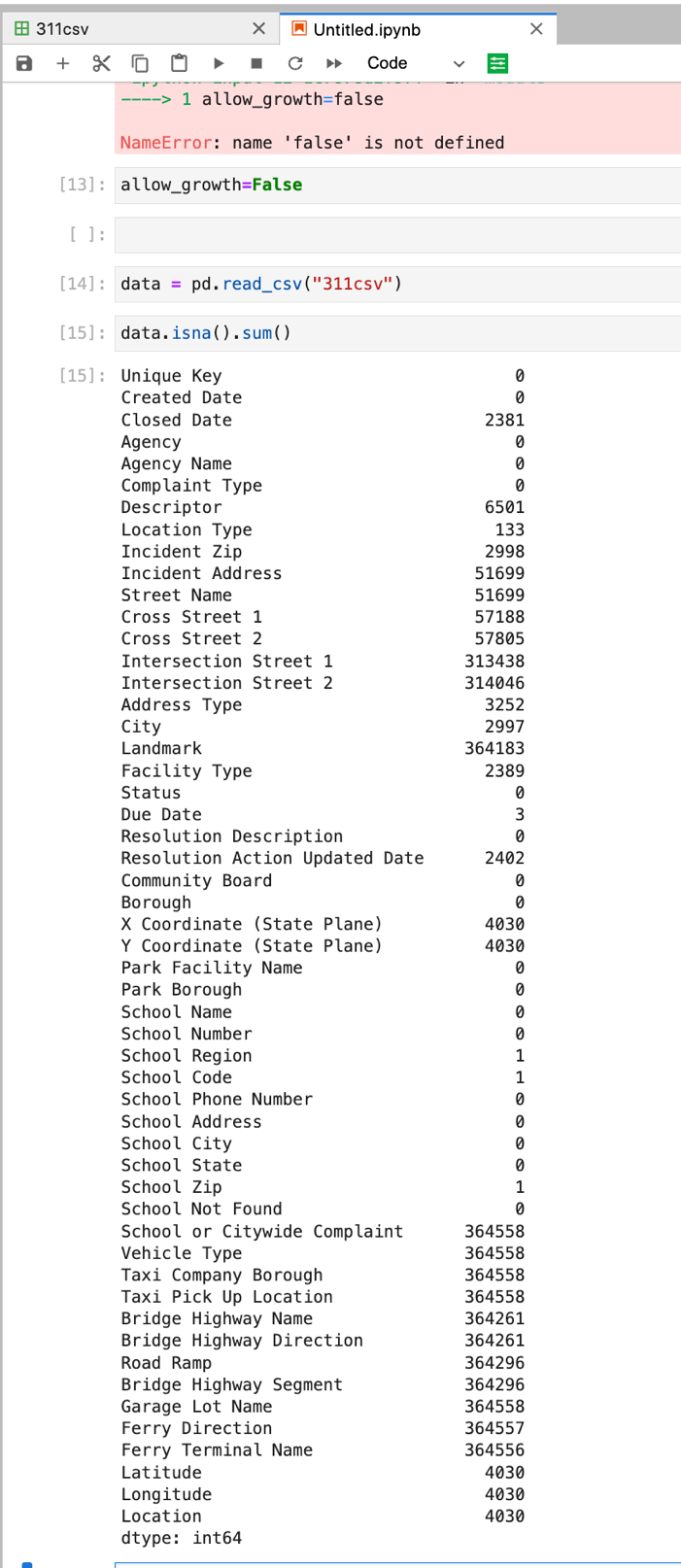
Next, we’ll print the dataframe to check out dataframe:

Our results gives us this (364558 Rows,53 Columns):

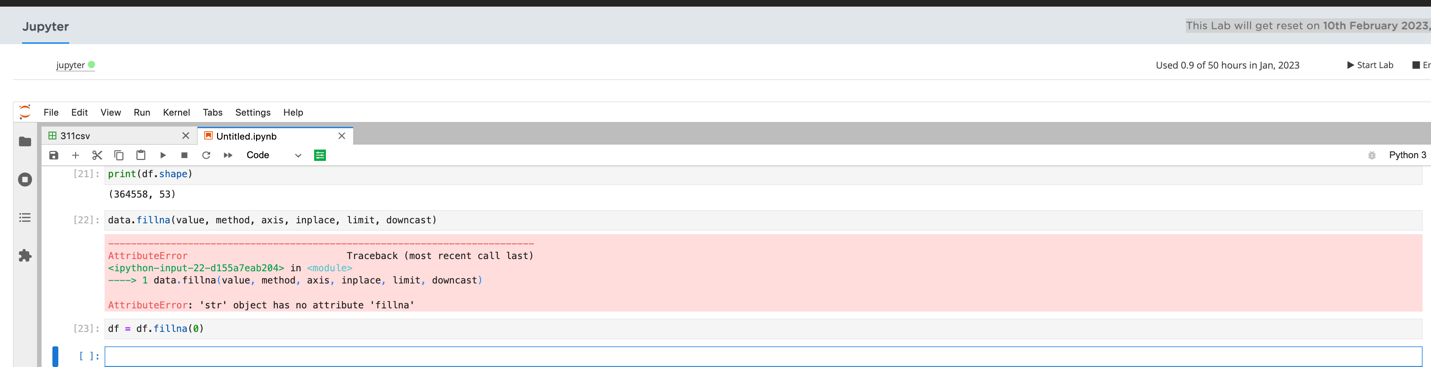




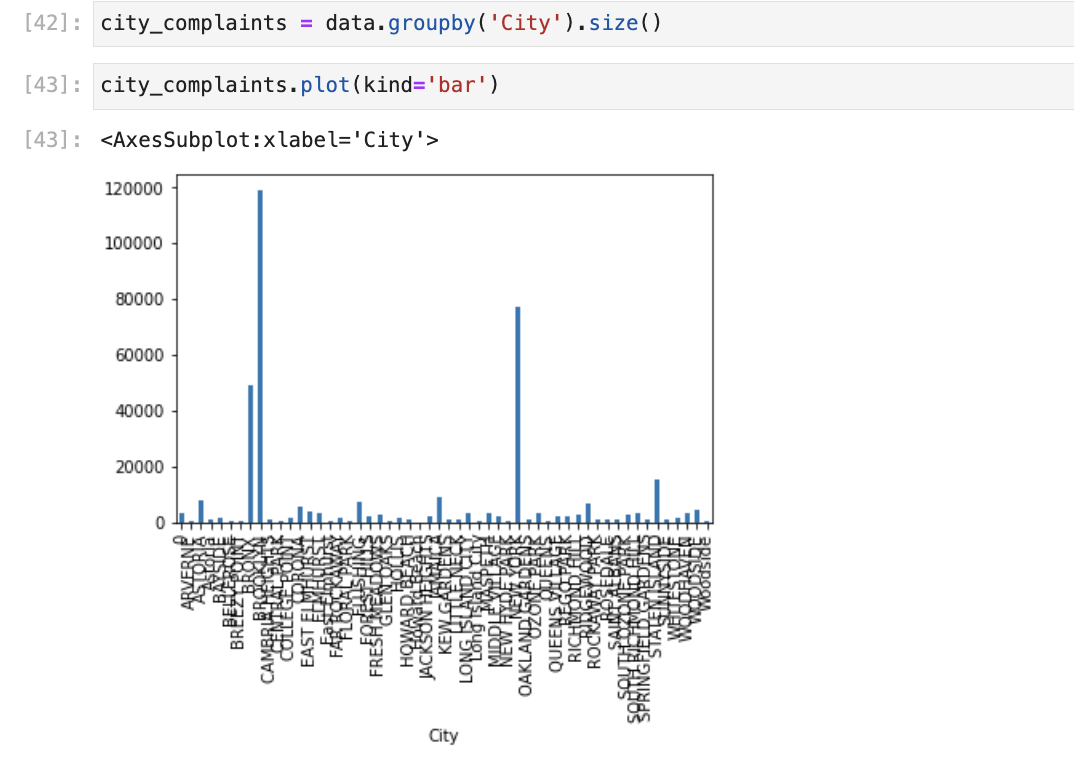
Next, we will import the 311 csv into our notebook, and then identify the number of missing values:

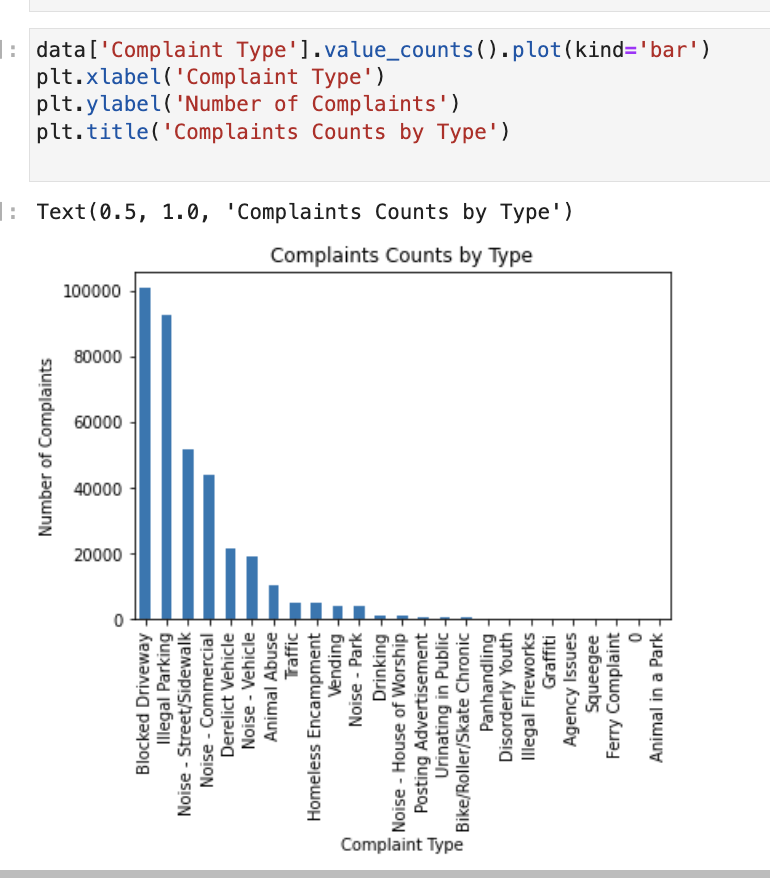


So next we want to deal with the missing values, so we’ll just add a “zero” anywhere we have a blank:

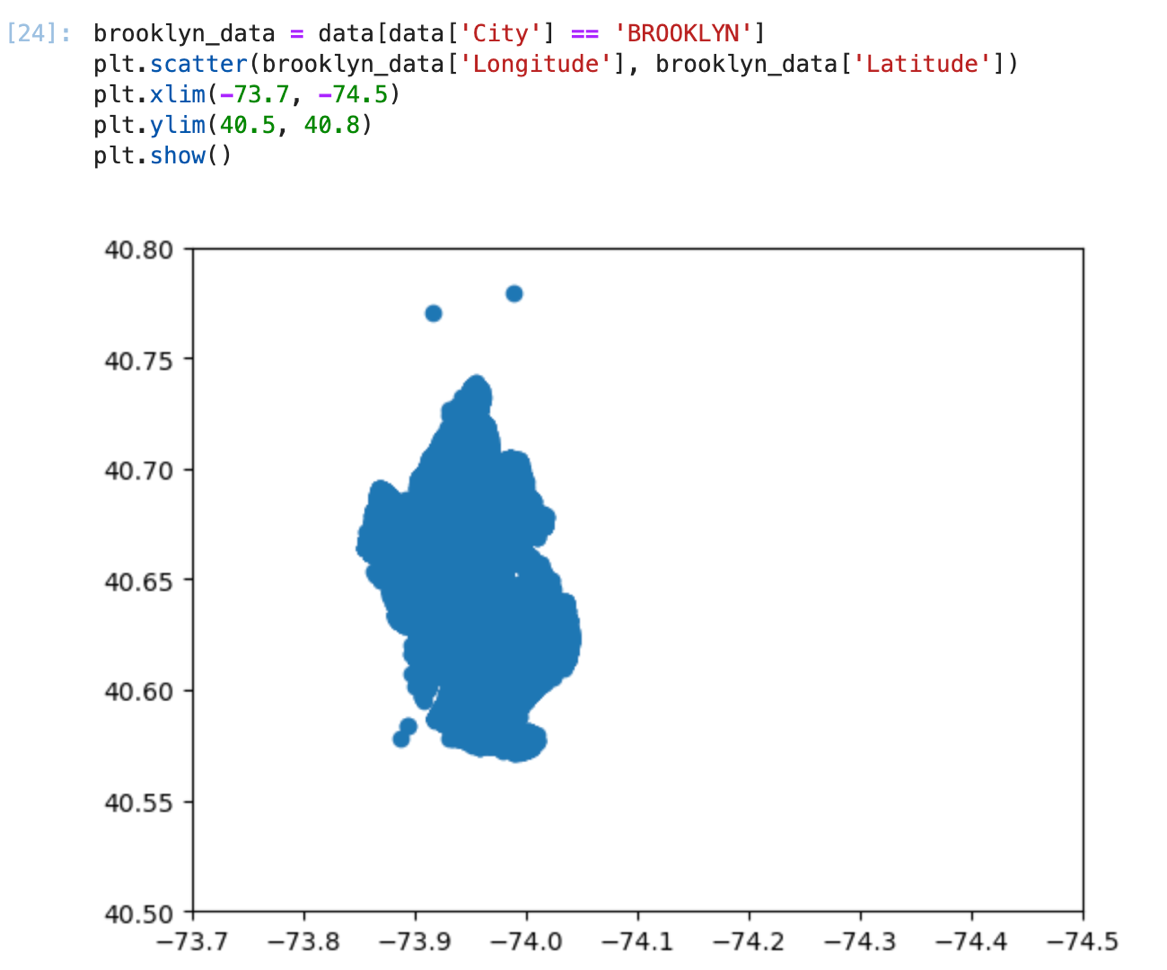


Next, we’re going to do a frequency plot for City Wise Complaints:

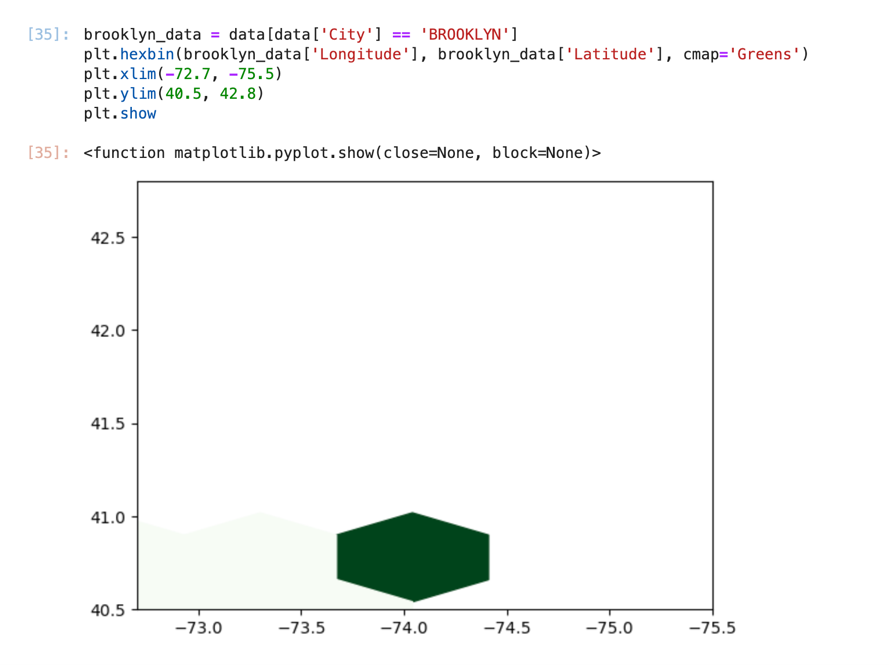




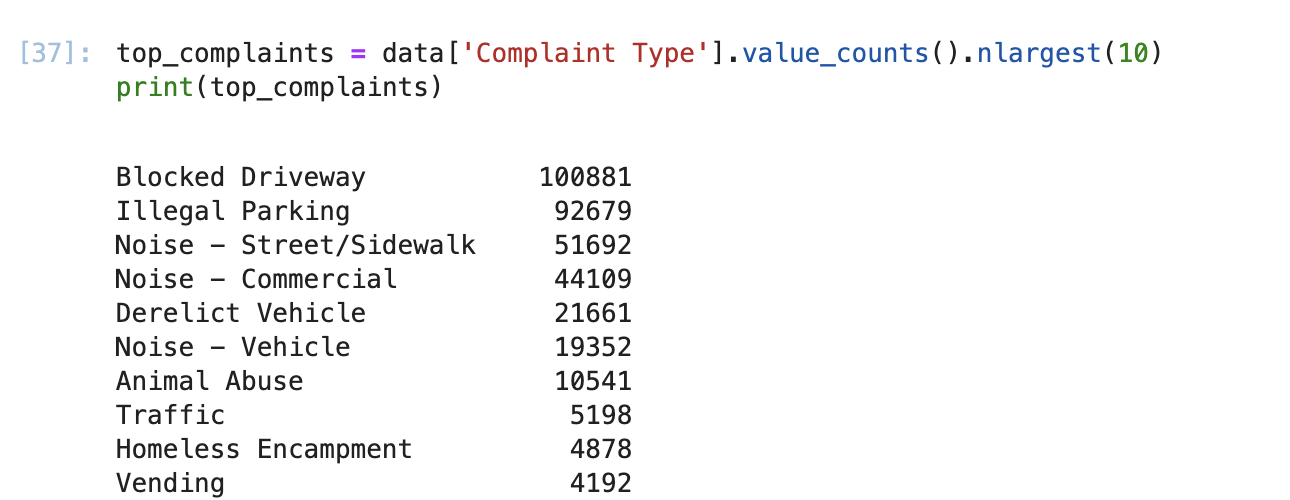
This is the scatter for complaint centration:



This is the hexabin for complaint concentration



Top 10 Complaints:



Average Response time needs the “Average closing Time” as per the instructions in the power point. This column is not listed in the .csv.