**NYC Fire Cases**

1. **Introduction**

**1.1 Background**

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the [most populous city](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) in the [United States](https://en.wikipedia.org/wiki/United_States). With an estimated 2018 population of 8,398,748[]](https://en.wikipedia.org/wiki/New_York_City#cite_note-2018Estimate-7) distributed over a land area of about 302.6 square miles (784 km2), New York is also the [most densely populated](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population_density) major city in the United States. Located at the southern tip of the state of [New York](https://en.wikipedia.org/wiki/New_York_(state)), the city is the centre of the [New York metropolitan area](https://en.wikipedia.org/wiki/New_York_metropolitan_area), the largest [metropolitan area](https://en.wikipedia.org/wiki/Metropolitan_area) in the world by [urban landmass](https://en.wikipedia.org/wiki/Urban_area) and one of the world's most populous [megacities](https://en.wikipedia.org/wiki/Megacity). Due to a huge population accident increases, fire being one of the most dangerous

**1.2 Problem**

With a tremendous increase in city food joints our aim is to find, are these joints have a contribution to it or not. Weather the areas with more joints have more fire breakouts or not?

1. **Data acquisition and cleaning**

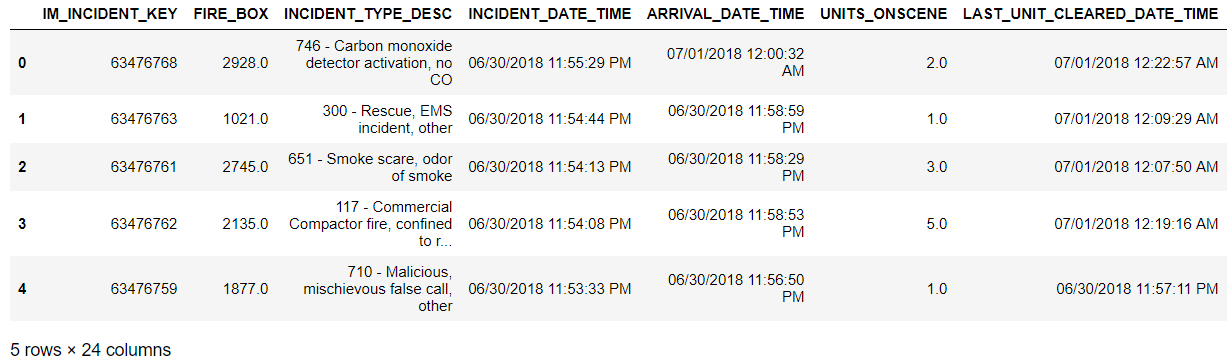
**2.1 Data Sources**

Several datasets will be used in the Project:

1. Fire Department Dataset (To analyse past fire incidents)
2. Location Database (to extract Coordinates)
3. FOURSQUARE api (to determine food joints in the proximity of fire effected place)

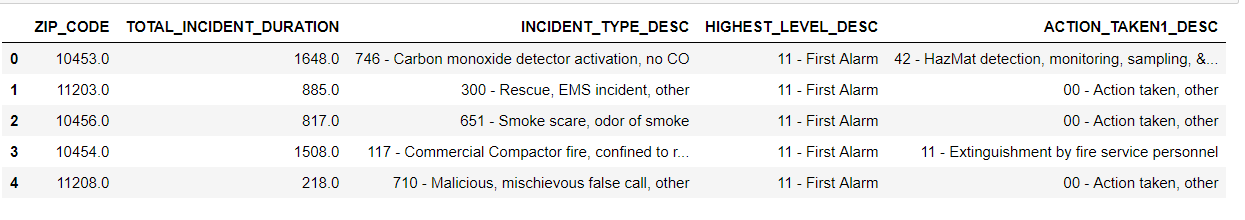
**2.2 Data Cleaning**

Fire department dataset have a lot of information including fire box number, type of fire, action taken, units deployed, date and time of fire break etc all of which is necessary.



We only require

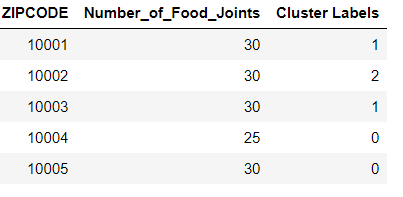
* ZIP\_CODE *(Zip code of the place)*
* TOTAL\_INCIDENT\_DURATION *(Time period for which the fire was not extinguished)*
* INCIDENT\_TYPE\_DESC *(Type of incident)*
* HIGHEST\_LEVEL\_DESC
* ACTION\_TAKEN1\_DESC *(Action taken by the authority)*



This Zip code is matched with the coordinated for plotting from the location database. Also, it is used to find the Food joints from FOURSQUARE api and cluster them on the bases of how prone they are to fire depending on previous data stored by fire department.

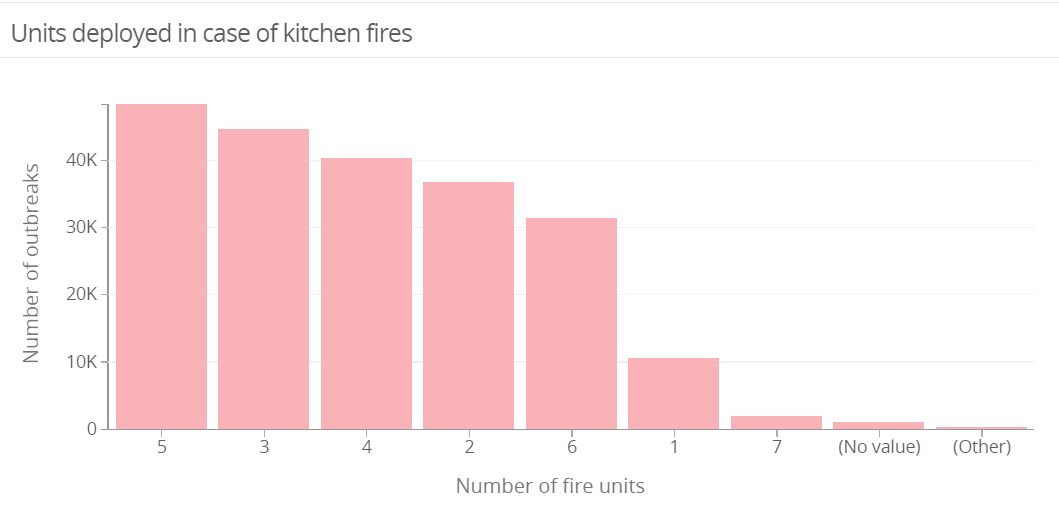
Feature scaling of incident duration is also done for improving performance.

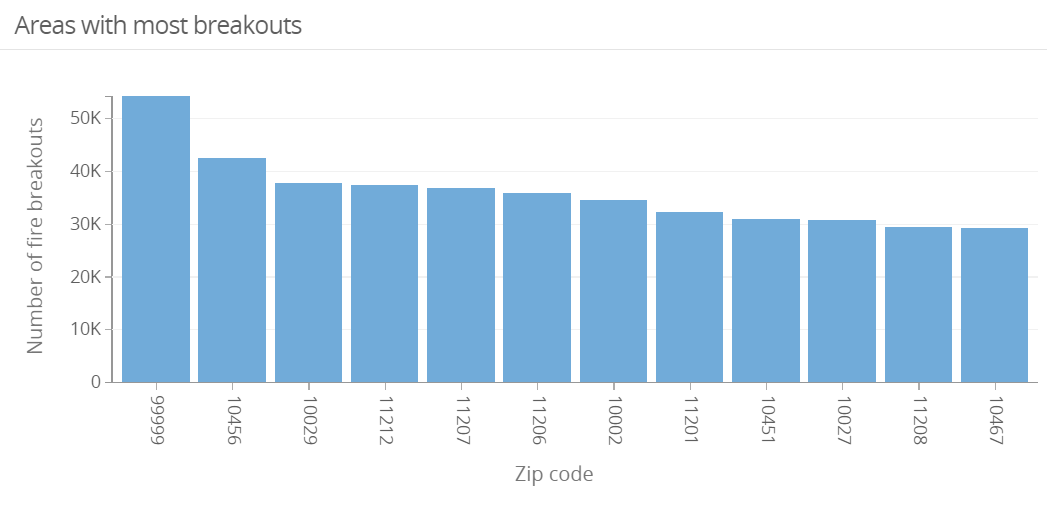
The following data was extracted using FOURSQUARE api.

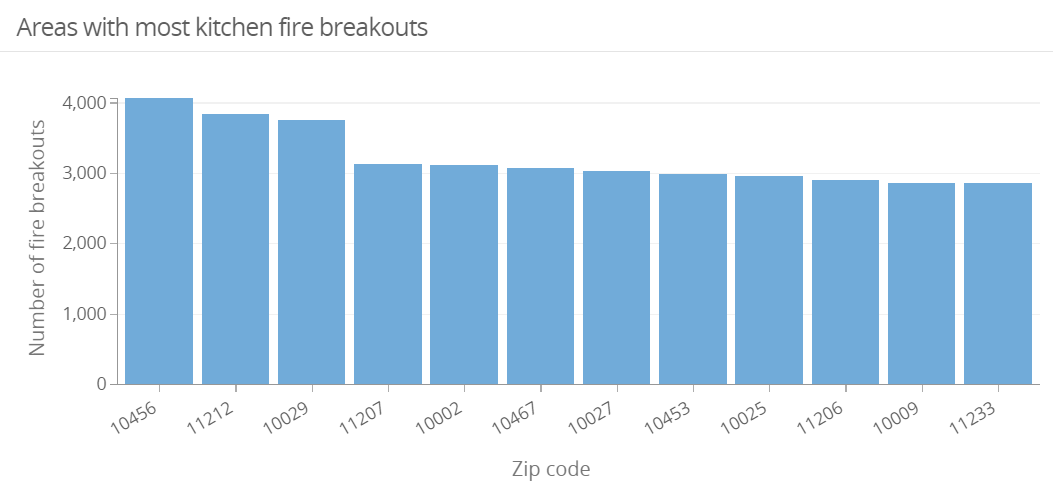


1. **Exploratory Data Analysis**

The fire outbreak is high if the time duration and fire units deployed to extinguished fire of the incident was large.

**Let’s observe its dataset**

**Let’s observe zip-code with the greatest number of fire break outs in NYC**

**Let’s observe number of kitchen fire-breakout with zip-code**

As we can see most fire breakouts occur near ZIP-CODE which had **23 food-joints** *(according to foursquare API)*

Data Analysis goes on like this:

* Extracting fire department dataset
* Extracting relevant data and applying feature scaling
* Group by the dataset on Zip Code
* Clustering the data
* Extracting coordinate data and merging (plotting)
* Calculating number of food joints using coordinates
* Classifying data and preparing final data

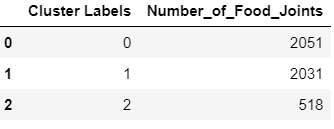
1. **Machine Learning Modelling**

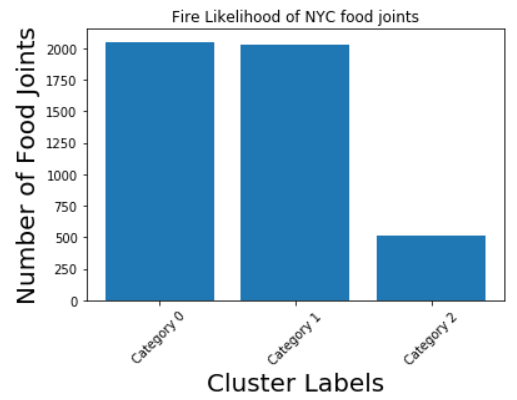
For separating the zip-codes on the bases of fire-breakouts K-nearest neighbours’ algorithms was used.

**KNN** depends on two things: A metric used to compute the distance between two points and the value of "k" the number of neighbours to consider. It will classify just based on the closest neighbours. For our dataset it groups data with most similarities. Using this model 3 classes were created

* Cluster 0: Zip-code with **most chances** of fire breakouts
* Cluster 1: Zip-code with **moderate chances** of fire breakouts
* Cluster 2: Zip-code with **least chances** of fire breakouts

1. **Results**





As from the dataset most fire breakouts were in the region with high number of food joints. Fire breakouts were less in the regions with less food joints.

1. **Discussion**

we conclude that most fire breakouts in NYC are due to food joints. This include kitchen fires or others due to high density of customers. Hence fire department must express its concern for the issue and work for safety of people in the region. People must be aware about first aids and emergency equipment must be present in the food joint premises. This will help tourist and residents for a better lifestyle.

1. **Conclusion**

Hence, we have concluded that most fire breakouts in NYC are due to food joints. This include kitchen fires or others due to high density of customers.