**ETL PROJECT**

2016 NYC Crime & Weather Analysis

Authors – Aziz Farouk, Walid Ahmed & Emil Kwarteng



# Executive Summary

As a newly appointed team in the FBI crime investigation department, we have been assigned to the NYC metropolitan area to investigate the relationship between the weather patterns and crime in the NYC boroughs, and to determine if there are any correlations between the changes in the weather and patterns in crime. This is a document describing the ETL process that we carried out in conjunction with the NYC State Police dept and the NY weather services department.

## Sources of Data

## Data Dictionary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field\_Name** | **Table** | **Data\_Type** | **Short description** | **Long description** |
| crime\_id,pk | CRIME\_NYC\_2015 | int |  |  |
| weather\_key | CRIME\_NYC\_2015 | timestamp |  |  |
| police\_dept\_code | CRIME\_NYC\_2015 | int |  |  |
| offense\_code | CRIME\_NYC\_2015 | int |  |  |
| occurence\_date | CRIME\_NYC\_2015 | date, |  |  |
| report\_date | CRIME\_NYC\_2015 | date, |  |  |
| crime\_completed | CRIME\_NYC\_2015 | varchar, |  |  |
| offense\_level | CRIME\_NYC\_2015 | varchar, |  |  |
| resp\_jurisdiction | CRIME\_NYC\_2015 | varchar |  |  |
| borough | CRIME\_NYC\_2015 | varchar, |  |  |
| police\_precinct | CRIME\_NYC\_2015 | varchar, |  |  |
| weather\_location | CRIME\_NYC\_2015 | varchar, |  |  |
| location\_description | CRIME\_NYC\_2015 | varchar |  |  |
| weather\_key,pk | WEATHER\_NYC\_2015 | timestamp |  |  |
| temperature | WEATHER\_NYC\_2015 | int, |  |  |
| dew\_point | WEATHER\_NYC\_2015 | int, |  |  |
| humidity | WEATHER\_NYC\_2015 | int, |  |  |
| wind\_speed | WEATHER\_NYC\_2015 | int, |  |  |
| wind\_direction\_degrees | WEATHER\_NYC\_2015 | int, |  |  |
| wind\_direction | WEATHER\_NYC\_2015 | varchar, |  |  |
| visibility | WEATHER\_NYC\_2015 | int, |  |  |
| pressure | WEATHER\_NYC\_2015 | int, |  |  |
| weather\_condition | WEATHER\_NYC\_2015 | varchar, |  |  |
| fog | WEATHER\_NYC\_2015 | varchar, |  |  |
| rain | WEATHER\_NYC\_2015 | varchar, |  |  |
| snow | WEATHER\_NYC\_2015 | varchar, |  |  |
| hail | WEATHER\_NYC\_2015 | varchar, |  |  |
| thunder | WEATHER\_NYC\_2015 | varchar, |  |  |
| tornado | WEATHER\_NYC\_2015 | varchar, |  |  |
| weather\_date | WEATHER\_NYC\_2015 | date, |  |  |
| weather\_time | WEATHER\_NYC\_2015 | time |  |  |
| police\_department\_code,pk | POLICE\_DEPT\_CLASSIFICATION | int |  |  |
| police\_department\_desc | POLICE\_DEPT\_CLASSIFICATION | varchar |  |  |
| offense\_code,pk | OFFENSE | int |  |  |
| offense\_level | OFFENSE | varchar |  |  |

## Two graphs that lead to cleaning

## ERD Diagram

## 

## Create table schema

-- Create a new table

DROP TABLE CRIME\_NYC\_2015;

CREATE TABLE CRIME\_NYC\_2015 (

crime\_id int NOT NULL,

weather\_key timestamp NOT NULL,

police\_dept\_code int NOT NULL,

offense\_code int NOT NULL,

occurence\_date date,

report\_date date,

crime\_completed varchar,

offense\_level varchar,

resp\_jurisdiction varchar ,

borough varchar,

police\_precinct varchar,

location varchar,

location\_description varchar

);

DROP TABLE POLICE\_DEPT\_CLASSIFICATION;

CREATE TABLE POLICE\_DEPT\_CLASSIFICATION (

Police\_Department\_Code int NOT NULL,

Police\_Department\_Desc varchar

);

DROP TABLE OFFENSE;

CREATE TABLE OFFENSE (

offense\_code varchar,

offense\_desc varchar

);

DROP TABLE WEATHER\_NYC\_2015;

CREATE TABLE WEATHER\_NYC\_2015 (

weather\_key timestamp NOT NULL,

temperature int,

dew\_point int,

humidity int,

wind\_speed int,

wind\_direction\_degrees int,

wind\_direction varchar,

visibility int,

pressure int,

weather\_condition varchar,

fog varchar,

rain varchar,

snow varchar,

hail varchar,

thunder varchar,

tornado varchar,

weather\_date date,

weather\_time time

); -- Create a new table

DROP TABLE CRIME\_NYC\_2015;

CREATE TABLE CRIME\_NYC\_2015 (

crime\_id int NOT NULL,

weather\_key timestamp NOT NULL,

police\_dept\_code int NOT NULL,

offense\_code int NOT NULL,

occurence\_date date,

report\_date date,

crime\_completed varchar,

offense\_level varchar,

resp\_jurisdiction varchar ,

borough varchar,

police\_precinct varchar,

location varchar,

location\_description varchar

);

DROP TABLE POLICE\_DEPT\_CLASSIFICATION;

CREATE TABLE POLICE\_DEPT\_CLASSIFICATION (

Police\_Department\_Code int NOT NULL,

Police\_Department\_Desc varchar

);

DROP TABLE OFFENSE;

CREATE TABLE OFFENSE (

offense\_code int NOT NULL,

offense\_desc varchar

);

DROP TABLE WEATHER\_NYC\_2015;

CREATE TABLE WEATHER\_NYC\_2015 (

weather\_key timestamp NOT NULL,

temperature int,

dew\_point int,

humidity int,

wind\_speed int,

wind\_direction\_degrees int,

wind\_direction varchar,

visibility int,

pressure int,

weather\_condition varchar,

fog varchar,

rain varchar,

snow varchar,

hail varchar,

thunder varchar,

tornado varchar,

weather\_date date,

weather\_time time

);

## Five queries

/\* standard weather query \*/

select w.weather\_condition, w.visibility,w.temperature,w.fog,w.rain,w.snow,w.hail,w.thunder,w.tornado

from WEATHER\_NYC\_2015 w

where w.weather\_date='2016-01-07' and w.weather\_time='18:14:00'

/\* Counting the total number of reported misdemeanors and felonies in NYC in 2015 \*/

SELECT COUNT(offese\_level)

FROM CRIME\_NYC\_2015;

/\* What were the most common crimes when there was heavy rain and fog in Queens \*/

select c.offense\_code,c.offense\_level, w.weather\_condition,w.fog,c.borough

from w.weather join c.crime\_nyc\_2015 on w.weather\_key = c.weather\_key

where w.fog='yes' and w.weather\_condition='Heavy Rain' and c.borough='Queens'

group by c.offense\_code,c.offense\_level w.weather\_condition,w.fog,c.borough

select c.offense\_code,o.offense\_description

from c.crime\_nyc\_2015 join o.offense on c.offense\_code = o.offense\_code

/\* Query 4 \*/

/\* Query 5 \*/

## Deliverables

## Python notebook with the data cleaning, filtering, transformation and loading

## SQL file showing table creation along with executable queries