Zuha Ahmad

DS-670: Capstone: Big Data & Business Analytics

Dr. Jaume

17 February 2017

**ASSIGNMENT 8**

**Midterm Proposal**

**OBJECTIVE**

The objective of the project is to execute a statistical modeling by performing a regression analysis and a time series forecasting based on crime and performance metrics. We will use programming tools and techniques, specifically the Zepellin notebook, to import the geo-spatial and regional data, inputting a superlative algorithm, and processing the data with visualization. The contributing variables in this dataset are street signs, potholes, wires, accidents. In addition, crimes such as battery, criminal damage, theft, etc. will be taken into consideration. The idea is to find correlation between crime and performance metrics and determine if there is a pattern of sorts with the provided dataset.

**DATA COLLECTION**

There are two datasets that will be derived from the United States in the city of Chicago. The first is from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) where it reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. The second dataset is from the Chicago Department of Transportation (CDOT). There are thirty-five different performance metrics tracked in this dataset, which reports the performance target, actual performance, and number of requests completed for a given metric.

The examples of performance metrics (variables) that will be looked upon are:

* Pothole in Street
* Traffic Signal Out
* Traffic Signal Timing
* CDOT Construction Complaints
* Landscape Median Maintenance
* Sign Repair - All Other Signs

These metrics are based on data calculated from the city's 311 system.

**BACKGROUND**

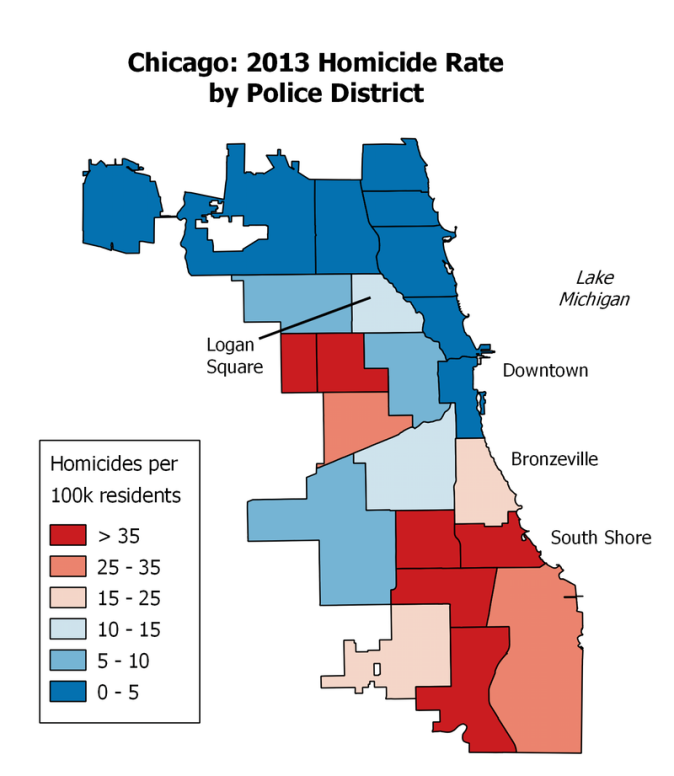
Chicago has a long history of open debasement that frequently draws the consideration of government law requirement and elected prosecutors. Chicago's political scene has been immovably under the control of the Democratic Party for more than 50 years and has been generally depicted as a political machine. In the 1980s, the FBI's Operation Greylord revealed monstrous and systemic debasement in Chicago's legal framework. Greylord was the longest and best covert operation ever, and brought about 92 government arraignments, including 17 judges, 48 legal counselors, eight policemen, 10 representative sheriffs, eight court authorities, and one state official. Almost all were indicted on an assortment of charges including renumeration, kickbacks, extortion, vote purchasing, racketeering, and medication trafficking.

Figure 1: Chicago Police Department

The late 1990s saw advance endeavors by the FBI to indict Chicago's open wrongdoing syndicates. Operation Incubator acquired around twelve feelings or liable supplications, including those from five individuals from the City Council and an associate to previous Mayor Harold Washington. Later Operation Gambat brought an extensive variety of charges against a Chicago judge, a state congressperson, a representative, and two others identifying with defilement in the Cook County Circuit Court, the Illinois Senate, and the Chicago City Council. 4 were sentenced and a fifth kicked the bucket amid trial. The most broad operation by the FBI of the 1990s, Operation Silver Shovel, tried to reveal debasement inside Chicago worker's parties, sorted out wrongdoing, and other city government authorities. Operation Silver Shovel brought about the conviction of 6 Chicago Alderman and twelve other nearby authorities on an extensive variety of debasement related charges.

From 1972 to 2012, 33 Chicago Alderman were sentenced on defilement accusations, a conviction rate of generally ⅓ of those chose in the day and age. A report from the Office of the Legislative Inspector General noticed that over portion of Chicago's chosen council member took unlawful battle commitments in 2013. In 2015, chairman delegated Barbara Byrd-Bennett, the CEO of Chicago Public Schools, was indicted in a 23 million dollar kickback plot and was sentenced to 7 and a half years in jail. Notwithstanding the Bennett conviction, a joint investigative report issued by the Office of the Inspector General and government specialists recorded across the board debasement inside Chicago Public Schools in 2015. The review noticed the criminal squeeze of a CPS seller, a records adulteration conspire by a key, various cases of workers manhandling CPS's duty absolved status to buy individual things at enormous box retailers, wrongfully utilizing citizen financed assets to battle for political causes and taking from citizen subsidized records proposed for buying understudy materials.

Executive of the Cook County Democratic Party and province assessor, Joseph Berrios, has been consistently blamed for nepotism, cronyism and damaging political support. Days into his residency, Berrios contracted his child and sister to his staff, holding them even after 58 others were laid off because of spending cuts. Berrios likewise contracted his hit or miss sweetheart, Vicki LaCalamita, to be his new executive of Human Resources, with a compensation of $107,841 and later enlisted LaCalamita's child and cousin.

A 2015 report discharged by the University of Illinois at Chicago's political science office pronounced Chicago the "defilement capital of America", refering to that the Chicago-based Federal Judicial District for Northern Illinois detailed 45 open debasement feelings for 2013 and an aggregate of 1,642 feelings for the a long time since 1976 when the U.S. Division of Justice started assembling the measurements. UIC Professor and previous Chicago Alderman Dick Simpson noted in the report that "To end debasement, society needs to accomplish more than convict the folks that get got. An extensive hostile to defilement procedure must be fashioned and completed over no less than 10 years. Another political culture in which open defilement is no longer endured must be made".

Cases of other prominent Chicago political figures indicted on debasement related charges incorporate Rod Blagojevich, Carol Moseley Braun, Jesse Jackson, Jr., Isaac Carothers, Arenda Troutman, Edward Vrdolyak, Otto Kerner, Jr., Constance Howard, Fred Roti and Dan Rostenkowski.

In October 2015, the FBI declared that Michael Anderson would assume control for a resigning Robert Holley as Special Agent in Charge of the Chicago Bureau. Anderson, a debasement veteran who composed the FBI Public Corruption Field Guide, called Chicago "target rich" for cases in a meeting with the Chicago Tribune. Anderson orders a group of 850 specialists in Chicago alongside investigators and care staff.

Most defilement cases in Chicago are arraigned by the US Attorney's office, as legitimate purview makes most offenses culpable as a government wrongdoing. The present US Attorney for the Northern area of Illinois is Zachary T. Fardon. In a question and answer session in January 2016, in the wake of the conviction of previous Chicago City Hall official, John Bills, for taking 2 million dollars in rewards, Fardon remarked "Open defilement [in Chicago] is an ailment and where open authorities disregard the general population believe, we need to consider them responsible. Furthermore, I do trust that thusly, it sends an impediment message."

**REGRESSION ANALYSIS**

In this measurable displaying, regression analysis is a factual procedure for assessing the connections among factors. It incorporates numerous systems for demonstrating and investigating a few factors, when the attention is on the relationship between a needy variable and at least one autonomous factors. All the more particularly, relapse investigation helps one see how the run of the mill estimation of the needy variable changes when any of the autonomous factors is shifted, while the other free factors are held settled. Most regularly, relapse investigation assesses the contingent desire of the needy variable given the autonomous factors, that is, the normal estimation of the reliant variable when the free factors are settled. Less normally, the emphasis is on a quantile, or other area parameter of the restrictive dispersion of the needy variable given the free factors. In all cases, the estimation target is an element of the free factors called the relapse work. In relapse examination, it is likewise important to portray the variety of the reliant variable around the relapse work which can be depicted by likelihood dissemination. A related however unmistakable approach is important condition investigation, which assesses the greatest (as opposed to normal) estimation of the needy variable for a given estimation of the free factor (roof line as opposed to focal line) with a specific end goal to recognize what estimation of the autonomous variable is essential yet not adequate for a given estimation of the reliant variable. The execution of relapse examination techniques by and by relies on upon the type of the information producing procedure, and how it identifies with the relapse approach being utilized. Since the genuine type of the information creating procedure is for the most part not known, relapse examination regularly depends to some degree on making presumptions about this procedure. These suppositions are now and again testable if an adequate amount of information is accessible.

Regression analysis is broadly utilized for expectation and determining, where its utilization has generous cover with the field of machine learning. Relapse investigation is additionally used to comprehend which among the autonomous factors are identified with the reliant variable, and to investigate the types of these connections. In limited conditions, relapse examination can be utilized to surmise causal connections between the free and ward factors.

In measurements, forecast is a piece of factual derivation. One specific way to deal with such deduction is known as prescient derivation, yet the expectation can be attempted inside any of the few ways to deal with factual induction. In fact, one depiction of insights is that it gives a method for exchanging information about a specimen of a populace to the entire populace, and to other related populaces, which is not really the same as forecast after some time. At the point when data is exchanged crosswise over time, regularly to particular focuses in time, the procedure is known as gauging.

Standard straight relapse models with standard estimation strategies make various suspicions about the indicator factors, the reaction factors and their relationship. Various expansions have been created that permit each of these suspicions to be casual (i.e. diminished to a weaker frame), and at times disposed of completely. A few strategies are sufficiently general that they can unwind various presumptions on the double, and in different cases this can be accomplished by consolidating diverse expansions. By and large these augmentations make the estimation method more intricate and tedious, and may likewise require more information so as to create a similarly exact model.

**TIME SERIES**

Time series involves strategies for breaking down time arrangement information with a specific end goal to separate important measurements and different attributes of the information. Time series is the utilization of a model to foresee future qualities in view of already watched values. While relapse investigation is frequently utilized so as to test speculations that the present estimations of at least one free time arrangement influence the present estimation of some other time arrangement, this kind of examination of time arrangement is not called "time arrangement examination", which concentrates on looking at estimations of a solitary time arrangement or numerous reliant time arrangement at various focuses in time.

Time series have a characteristic transient requesting. This sets aside a few minutes arrangement investigation unmistakable from cross-sectional reviews, in which there is no normal requesting of the Time arrangement examination is likewise particular from spatial information examination where the perceptions ordinarily identify with land areas. A stochastic model for a period arrangement will for the most part mirror the way that perceptions near one another in time will be more firmly related than perceptions facilitate separated. Furthermore, time arrangement models will regularly make utilization of the normal one-path requesting of time so that qualities for a given period will be communicated as getting somehow from past qualities, instead of from future qualities.

**PSEUDO CODE**

%python

import pandas as pd

import numpy as np

import matplotlib.pylab as plt

%matplotlib inline from matplotlib.pylab

import rcParams rcParams['figure.figsize'] = 15, 6

data = pd.read\_csv('Users/ZAhnad/Downloads/performance\_metrics/\*')

print data.head()

%sql

select carbon\_monoxide, timestamp from data1

**REFERENCES**

Kamisar, Y. (1972). How to Use, Abuse--and Fight Back with--Crime Statistics. Okla. L. Rev., 25, 239.

Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Data mining for discrimination discovery. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *4*(2), 9.