**CitationPy – Team 3**

**(03/30/2019)**

**Introduction:**

Team 3 picked a project to uncover patterns in parking activity around the city of Los Angeles. The first purpose was to examine a relationship between the makes of vehicles and number of parking citations. The second purpose was to examine a relationship between the colors of vehicles and parking citations.

**Description of data and technology used:**

The following data sets were used throughout the project.

* Sample dataset was obtained from the City of Los Angeles for the months of May through July 2018 in csv format.
* A dataset from the Los Angeles DMV was separately analyzed to check for normalization.
* Folium Mapping was used for plotting latitude and longitude.
* Jupiter notebook was used for coding of analysis and cleaning of data.
* Git Hub was used as a repository to store analysis and reports.

**List of dependencies used**

* import matplotlib.pyplot as plt
* import pandas as pd
* import numpy as np
* import requests
* import time
* import scipy.stats as stats
* import folium
* import folium.plugins as plugins
* from folium.plugins import MarkerCluster
* from folium.plugins import FastMarkerCluster
* import pyproj
* import sys

**API Call**

* [City of Los Angeles](https://data.lacity.org/resource/8yfh-4gug.json)
* [Folium Mapping](https://python-visualization.github.io/folium/)

**Methodology:**

* A horizontal bar chart was plotted to show the top 25 makes of vehicles that had the most citations.
* A vertical bar chart was plotted to show the top 10 colors of vehicles that had the most citations.
* location and heat maps were plotted to show the exact locations where citations were made.
* A separate graph was created to show the possible correlation between parking citations and the surrounding area.
  + Restaurants
* A comparison between the means of two sample vehicle makes were carried out in order to determine whether they are different. Chi-Square analysis was used for this purpose.

**Hypothesis Testing:**

* Q1. What is the relationship between the make of vehicle and the number of citations received in the City of Los Angeles between May – July 2018?
  + Ho: there is no significant relationship between the make of vehicles and the number of citations.
  + Ha: there is a significant relationship between the make of vehicles and the number of citations
* Q2. What is the relationship between the make of vehicles and the number of citations received in the e City of Los Angeles between May – July 2018?
  + Ho: there is no significant relationship between color of vehicles and the number of parking citations.
  + Ha: there is a significant relationship between color of vehicles and the number of parking citations.
* Q3. Are the means of the citations received of two groups of vehicles different (independent)?
  + Ho: there is no significant relationship between the means of the two groups.
  + Ha: Alternate Hypotheses: there is no significant relationship between the means of the two groups.

**Findings:**

* Toyota tops the list of the make of vehicles that received the most citations at 74K followed by Honda at 47K and Ford at 35K. Nissan was fourth in the list at 31K and Chevy ranked no.5 at 27K.
* White colored vehicles received the most citations followed very closely by black and gray colored vehicles. Silver ranked at no.4 and blue ranked no. 5.
* While comparing the top 25 list of vehicle makes that received citations to the top 25 list of vehicles registered with the DMV, there were 23 vehicles that were common in both. This shows that if we take count only, it could probably be because there were more vehicles of that make.
* Since the chi-square value of 25599 at a confidence level of 95% exceeds the CV of 7.8114. We, therefore, conclude that the difference between the two groups is statistically significant. Also, the p-value was zero, which indicates that the null hypothesis is false. We rejected the null hypothesis.

**Limitations of data and analysis:**

The data set was very big and the most recent data available was up to July 2018. For this reason, we only looked at three-month data from May to July 2018. This limited us in looking for seasonality for months during which the citations peaked. Also, a separate analysis had to be done to normalize data to check if the citations were occurring more for a certain make and color of a vehicle just because they were the most registered vehicles in the city during that particular period. The dataset obtained from DMV of the city of LA didn’t have the color of a vehicle which limited the analysis to make of vehicles only.

**Conclusions/Recommendations:**

From our evaluation of the data used we recommend that you park in designated areas and do not violate the parking laws.

The Analysis is limited to the sample data set of May – July of 2018.

Use a more current dataset to further evaluate the findings.

**Appendix of All Vizzes:**

1. Color chart
2. Make chart
3. Citation Map
4. Citation Heat Map
5. Restaurant Map
6. Restaurant Heat Map