## 

## Project Report : Extract Transform Load (ETL)

**Date : March 29, 2019**

**Data Cleanup & Analysis**

**-----------------------------**

This is the data from city of Oakland “Call Center , the reason why I chose to perform an ETL on this data is it because I was interested to know more about which way is easy for citzen to report or reach out to the city of Oakland administration in case they have anything to report or if they have any complain , I extracted two different data set one was with json format and another one was csv file . I downloaded them from data\_world and they loaded them into pandas . for the kind of dataset that I have I am required to make sure that clean data by droping NAN and replacing it with an value that cannot affect the outcome. After droping nan values also filtered only the columns that I needed and aggregated them later I joined two data set into one single dataset . in joining the data set I had an issue with with x and y in this case I used suffixes=("\_First" for replacement of x value and "Last" for replacement of y value .

* After performing all necessary cleaning that I mentioned above I loaded the final chosen data into pgadmin postgress sql ,
* Steps required to reproduce my ETL process , are the following .

\* download json and csv file from either cargo or data\_world. Then save them into single folder , in loading data , make sure that the data are loaded separately because they contain different format , after loading each file depending in it`s own format , we can now transform json data into csv before joining them together .