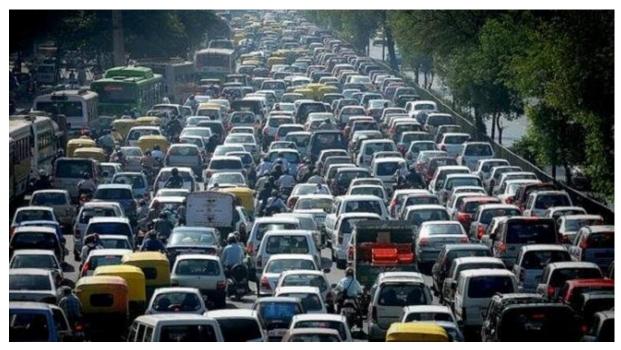# CORRELATION BETWEEN HEAVY TRAFFIC AND VENUES IN ISTANBUL

IBM DATA SCIENCE CAPSTONE

JULY 12, 2020

The biggest problem of Istanbul, Turkey's greatest and most populated city is heavy traffic. In this project, I'm going to try to find out the reason behind that. In my opinion, it's the way that city is serperated. In this city, the districts are built so different from each other. While some districts are consisting of farms, parks, greenplaces and venues like that, other districts contains only offices, malls etc. Because of that, everyone has to go to the same districts no matter which part of the city they live.

I'm going to use the counting data from hundreds of censors all around the city. This data is published by the government at:

https://data.ibb.gov.tr/dataset/gunluk-arac-sayimi/resource/15e1332d-e476-4899-b3b9-8574ac8f6eb8

With that dataset , I'm going to use the foursquare API to gather the data of venues around those censors. This is going to provide me enough information to find out if I'm right or wrong.

Plan of Attack

First, I'm going to use the dataset to serperate the amount of traffic into three different clusters: low , medium and heavy traffic.

Then, I'm going to gather the first 100 venues around every censor and get the most frequent type of venue for each censor.

After that, I will use KMeans to cluster censor locations to get different districts.

Then, I'll group the dataframe by those districts and add a column that contains the most frequent type of venue.

In the end, I'm going to plot those clusters of locations and heaviness of the traffic there by coloring them and write down the name of the most frequent type of venues.