# The Battle of the Neighborhoods

## Coursera Capstone Project Report

### By Muhammad Musaddiq Sajjad

# Introduction

### Background

I live in the charming city of Al-Khobar, situated near the sea in the Eastern Province in the beautiful country of Saudi Arabia. It is one of the twin cities of Al-Khobar and Dammam, that reside side-by-side and have a lot in common.

Both cities are hustling and bustling with life, being the centers of the major activity in the Eastern Region. Everything from industrial zones, shopping malls, resorts, luxury apartments, lakes to restaurant chains and traditional markets can be found here.

Needless to say, both the cities, while being quite close and similar, are still a substantially varied mix of businesses and venues.

### Problem

For this report, I will be focusing on developing a system to recommend locations for businesses of different categories to open in one of these two cities. Any business that is entering the market can make use of this data to determine likely opening locations.

We aim to provide an idea of which district, location and cluster (from K-means clustering) the business should consider opening in.

### Interest

This report will be of interest to any business owner looking to open a new business in the Al-Khobar and Dammam area, as it will deliver, given an input business category, locations where the business is likely to succeed.

It can also be of use for anyone looking for recommendations of businesses, such as where to find a good Italian Restaurant.

# Data

### Data Sources

The venue data will be sourced from the Foursquare API, which will return location data and nearby venues, along with the venue category. This will help us in categorizing the different kinds of businesses to build our models.

Furthermore, we will be using publicly available district boundaries data and parsing that into a GeoJSON to establish our district boundaries.

Source: https://github.com/homaily/Saudi-Arabia-Regions-Cities-and-Districts

### Data Cleaning

Since the Foursquare API returns data that is already formatted, there is no data cleaning involved in that part.

However, the district boundary data sourced from the above linked Github repository does need to be prepared.

Specifically, it is JSON data, in the below format:

{'district\_id': 10100003002,

'city\_id': 3,

'region\_id': 1,

'name\_ar': 'حي النموذجية',

'name\_en': 'Al Namudhajiyah Dist.',

'boundaries': [[[24.65018372, 46.70227584],

[24.64939455, 46.7014039],

[24.64915715, 46.70115918],

[24.64892224, 46.70091159],

[24.64868987, 46.70066116],

[24.64857349, 46.70053129],

[24.64846099, 46.70039739],

[24.64835249, 46.70025959],

[24.6482481, 46.70011803],……

}

Note that it is using latitude,longitude coordinates system, so it will have to be reversed to be used as a GeoJSON. More information here, see "Position" section: <https://macwright.com/2015/03/23/geojson-second-bite.html>

This will have to be parsed into the standard GeoJSON format as below:

{

"type": "FeatureCollection",

"features":

[{

"type": "Feature",

"geometry":

{"type": "Polygon",

"coordinates": [[[50.13177952, 26.42026955], [50.13205218, 26.42399558]

…

}

}

}