**Coursera**

**IBM Data Science Professional Course**

Capstone project – The Battle of Neighborhoods

**Predicting Residential Units Rental Rates in**

**Riyadh City**

Alaa J. Jaradat

April 2019

Table of contents:

[**I.** **Introduction:** 2](#_Toc5290194)

[**II.** **Data description:** 3](#_Toc5290195)

1. **Introduction:**

Riyadh is the capital and largest city in the Kingdom of Saudi Arabia with a population of more than 5 million people with a total surface area of 1,798 kilometers squared and population density comes to approximately 4,300 people living per square kilometer.

Riyadh city is divided into districts and has approximately 1 million residential units.

The residential units rental prices started to drop starting from a couple of years ago due to number of reasons such as expatriates leaving the country, and the delivery of a number of residential construction projects done by the ministry of housing for Saudi nationals, and those reasons caused a low demand in residential units rental market.

This report is part of the final project in the IBM Data Science Professional course provided by Coursera with a requirement to leverage the Foursquare location data to execute this project.

The goal of this project is to predict residential units rental rates for the current year half (first 6 months 2019) to help the target audience of this report to understand the real estate market in the city of Riyadh and study the effect of surrounding venues of each district on the rental rates. Another goal is to share knowledge learned from this course with other data scientists who are looking for enrolling in this course.

Let us find out if conducting a data science approach for this problem allows us to predict the rental rates and eventually help the audience of this report in making decisions related to this market or at least extract some insights by doing this study.

The audience of this report are:

* Real estate investors.
* Riyadh residents or people planning to relocate to Riyadh.
* Individuals interested in the real estate trends in Riyadh.
* Data analysts and Data scientists that can contribute or get benefit from this project.

1. **Data Description:**

The data used in this project is collected from below data sources:

* Foursquare.com:

Foursquare is a web application that provides its users with features including search and discovery of current location, and a social networking tool enabling users share their location with friends via the check-in feature.

Foursqaure API is used here to collect information about the surrounding venues to Riyadh districts such as venue name, venue category, venue coordinates.

* Address.gov.sa:

National address is a web application created by the Saudi post institute to unify and organize the locations in Saudi Arabia and approved by the Saudi government as the main addressing reference in Saudi Arabia.

National address website offers a web API that enables developers accessing its database to get geospatial information about regions, cities, districts, and other addressing components. This API is used in this project to get the location of Riyadh districts and its location boundaries.

* Aqar.fm:

Aqar is a popular web application in Saudi Arabia allowing its users that are looking for buying or selling real estates to search the app for the available real estates that are advertised by other users.

Aqar.fm has page that shows residential units rent prices statistics for each district in a year half time frames.

* Another data source is a web page I found while I was trying to collect data from the internet, and it has Riyadh districts locations, the webpage is part of Google MyMaps service.

Web scraping is used to extract the results from the Google MyMaps page but I did not include it in the analysis after I found a more trusted data source, which is the national address service of Saudi Arabia.