

# Riyadh Real Estate Market: Visualization Using AQAR Dataset

Noha Alawwad

17214080

Submitted as a requirement for the course CA682P Data Visualization. MSc in  
Data Analytics. Spring 2018

## 1. Introduction

This project aims to investigate and analyze Riyadh real estate market to draw conclusions related to the prices and the other features in the dataset. Many questions were asked during the analysis process, which is then answered by the resulted charts and plots.

- Did the district position (north, west, east, south) affect the price of the property?
- Did the property prices decreased or increased by years? If yes, what are the factors?
- Are there any districts who have property offers more than the others? If yes, what are the reasons?
- If any villa has some auxiliary services such as, driver room, or furniture, what is the effect on the villa price?
- Is there any relationship between the villa price and street width?

## 2. Dataset

### 2.1. AQAR dataset

The dataset was collected from AQAR website which is a well-known website for posting sales and rentals offers in Saudi Arabia.

The dataset includes 37 features and 99,969 instances collected in three years, 2014, 2015 and 2016.

Features in the dataset can be divided into two parts to make it understandable. Features related to the post like post ID, user ID, date of publishing the post, date of updating the post, closed (offer closed or yet opened for buyers), type of person offering the property (marketer or owner), paid or free post.

The other features are related to the offered property and can be divided further into a quantitative and qualitative feature. Quantities features such as the age of the property, number of apartments, number of rooms, number of bedrooms, number of living rooms, number of bathrooms, floor number, number of kitchens, price, area, street width.

Qualitative features include features like, residential or commercial, has driver room or not, property type (villas, apartments, farms, or stores), duplex (duplex or not). It also includes furnished or not, has swimming pool or not, rental period (annual or monthly), street direction, for sale or rent, district name in English and Arabic, region name in English and Arabic, province name in English and Arabic, nearest city name in English and Arabic and zip code number.

### 2.2. Longitude and Latitude for Riyadh Districts

This dataset was collected manually for 100 districts from Google Map to draw the districts circles in the right position in d3.js map.

## 3. Process

### 3.1. Preprocessing

Preprocessing part was done using Alteryx and Python. At first, Alteryx was used to filter and transform the main dataset in order to produce two smaller datasets that will be used in d3.js dashboards, one for villas that offered for sale and the other for rental apartments.

The main transformation process in Alteryx as shown in the following figure include:

- Exclude records with null values in district name.
- Exclude records with zero values in the area column.
- Filter the area of villas to include the range of 280 to 850 square meter, to exclude the outliers.
- Remove the outliers in the price column.
- Calculate (price per unit area) column through dividing the price of the villa by the area.
- Transform the date for the published post into a year of the form (yyyy) as a new column.
- Change the needed columns' names to from Arabic to English.

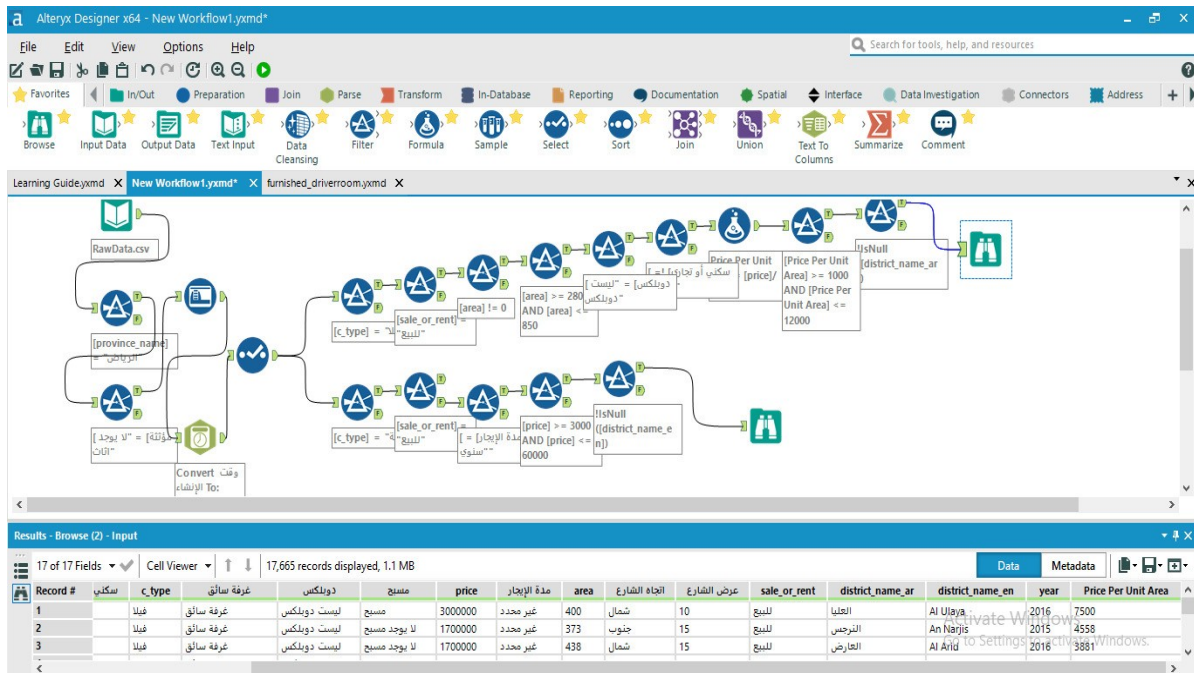


Figure 3.1: Preprocessing in Alteryx

This process of transformation was done again for a smaller range of area 300 to 650 square meter in order to use it for python charts since the majority of villas falls in this range. Also, to not deviate the conclusions.

After that, Python was used through Jupyter notebook to do some further transformation:

- Create pivot table to calculate the average of prices per district for the three years 2014, 2015, and 2016.
- Drop districts that have null values in two years.
- If any district has a null value in one year, substitute the null with the average of the other two years.
- Calculate the total offers in each district by using the *groupby* function.
- Use *query* function to create subsets of the dataset, for instance, query the villas that have driver room and it is furnished (Figure 3.2).

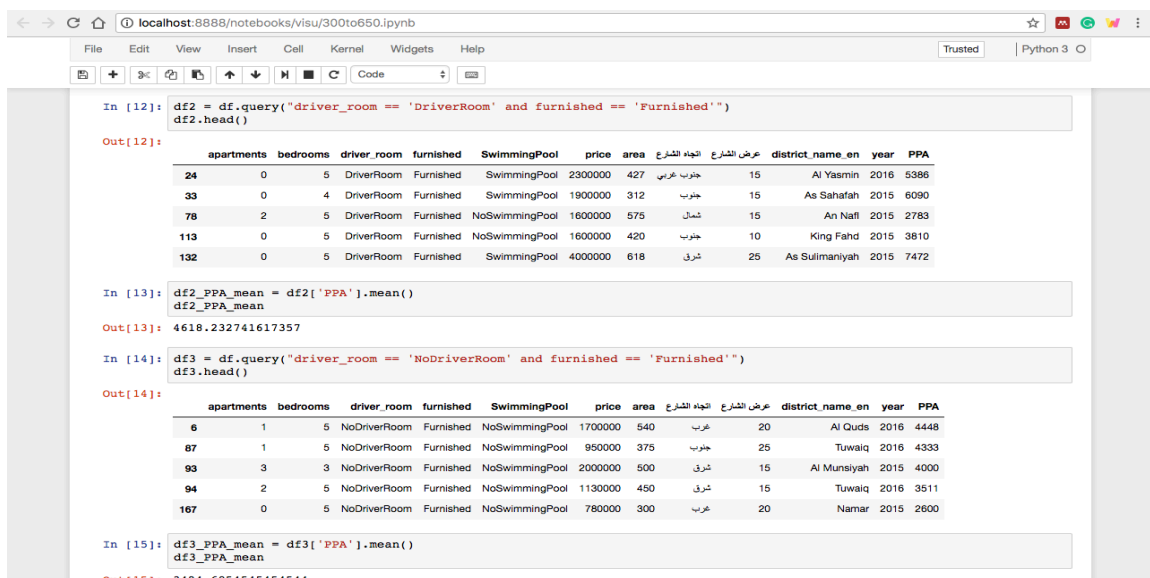


Figure 3.2: Preprocessing in Python

## 3.2. Visualization tools

### 3.2.1. D3.js library

D3.js was used to create two interactive dashboards, one for villas sales prices and the other for rental apartments prices in a sample of 100 district. Three main visualization components were used in the dashboards, map as JSON file, line chart, and bar chart. Map used to show districts information, line chart to show price change over the years, while bar chart was used to indicate offers number for each district.

### 3.2.2. Python Altair library

Altair library mainly used to build charts as layers. This library was used to create bar chart for villas average prices for the additional villas services (furniture and driver room), and on top of the bar chart a rule and text were added.

### 3.2.3. Python Plotly library

This library allows you to make interactive graphs in python. Interactive box plot was created using this library to show street width and villa price relationship, where information related to the minimum, maximum, and median values appear in the graph by hovering.

## 4. Results

The answers to the previous asked questions during the analysis process shown in the following part.

### 4.1. D3.js dashboards

Did the district position (north, west, east, south) affect the price of the property?

Sale prices have been categorized into five categories ranging from less than 2,000 up to 10,000 as indicated in the legend. In the past three years starting from 2014 up to 2016, the categories indicate that the highest villas prices were located in the north of Riyadh with five districts dominated the others namely, AlWahah, An Nakhil, Hittin, Ar Rabi, and AlGhadir.

Urban expansion is always toward the north, and most of the services such as universities, government agencies, and the airport are provided in the northern districts making it more appealing for people to move to the north area. Besides, the south area is overpopulated. That's why people want to move to the north. Figure 4.1 shows Al Wahah district in the darkest color.

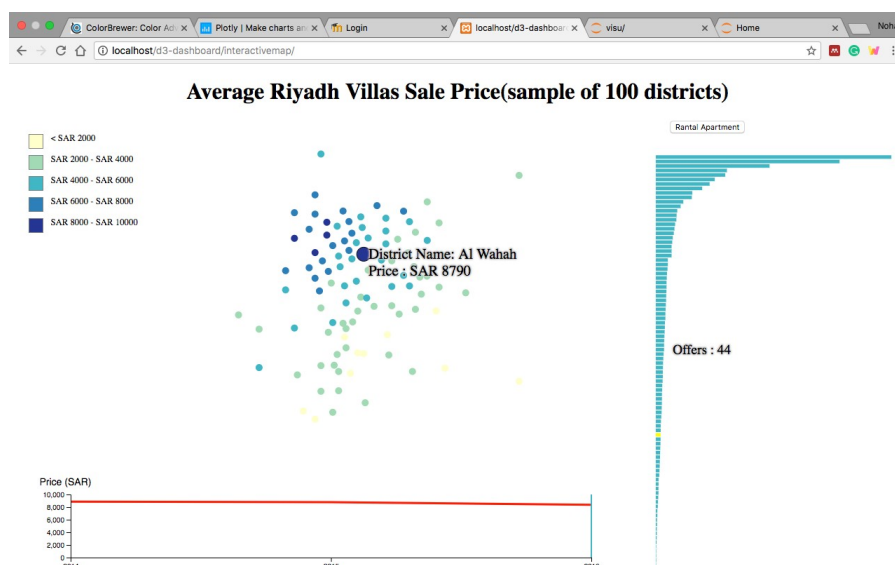


Figure 4.1: Northern districts prices in villas sale price dashboard

Did the property prices decreased or increased by years? If yes, what are the factors?

We have seen a drop in property prices starting from 2015 to 2016 compared to 2014 prices. However, we have noticed that some of the districts' price remain the same throughout this sample period. The drop in prices are not surprising since it was highly correlated to the drop in government spending and the decrease in household income and welfare as the government had increased the energy and some of the services prices in 2016. Moreover, this was coupled with an increase in the inflation rate in Saudi Arabia, which ultimately reduced the individual purchasing power (Figure 4.2).

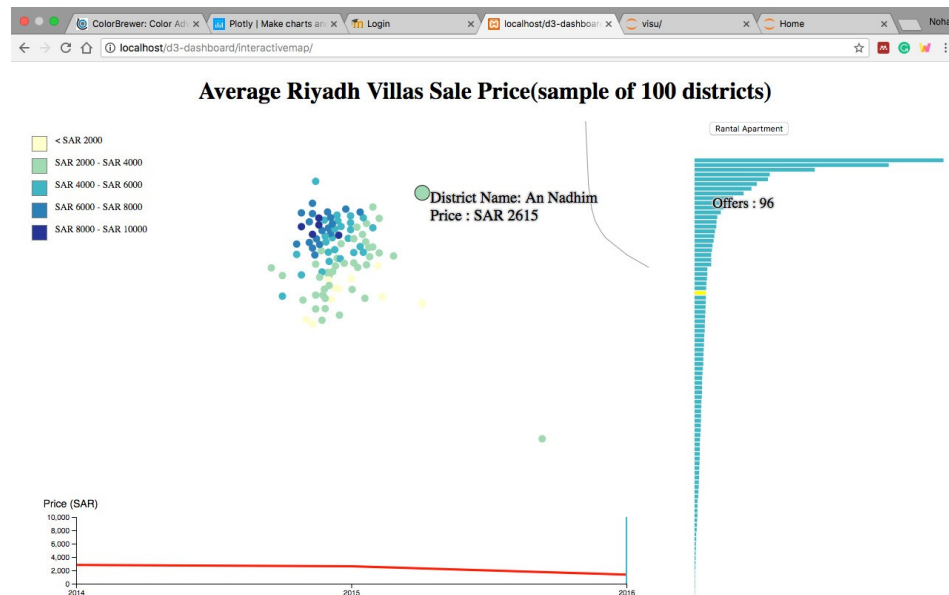


Figure 4.2: Price decreasing in 2016 in villas sale price dashboard

As people income is low and they cannot afford villas, they stay in rental apartments which keep the demand high for the apartments. That's why apartments did not react negatively with the decrease in villas prices as shown in Figure 4.3.

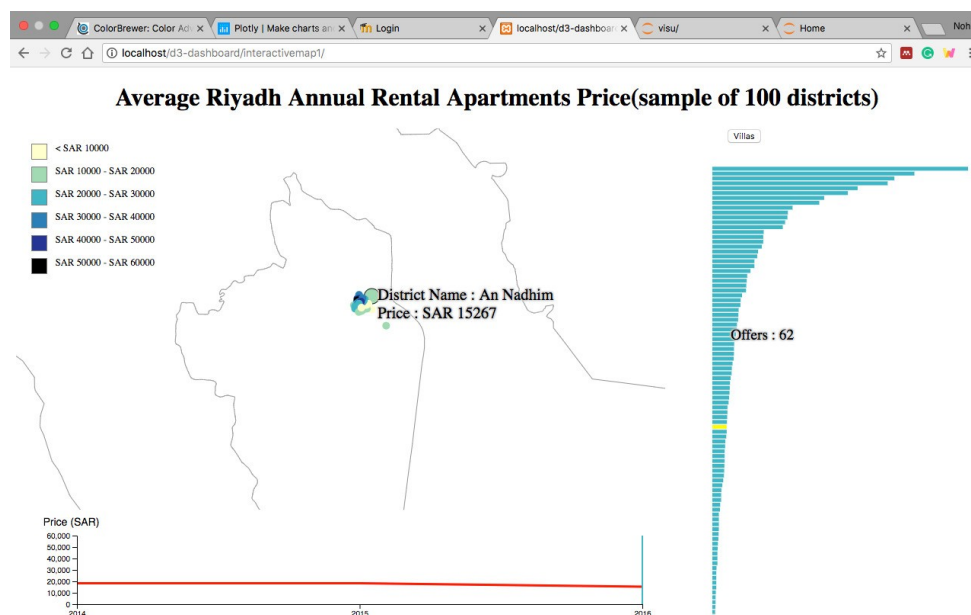


Figure 4.3: Annual rental apartments

Are there any districts who have property offers more than the others? If yes, what are the reasons?

More offers have been witnessed in Tuwaiq, and Dhahrat Laban since these two districts are considered new districts and has been in high demand right after the government constructed Western Ring Road which directly connected these two districts to the north and south of Riyadh. Figure 4.4 shows Tuwaiq district offers.

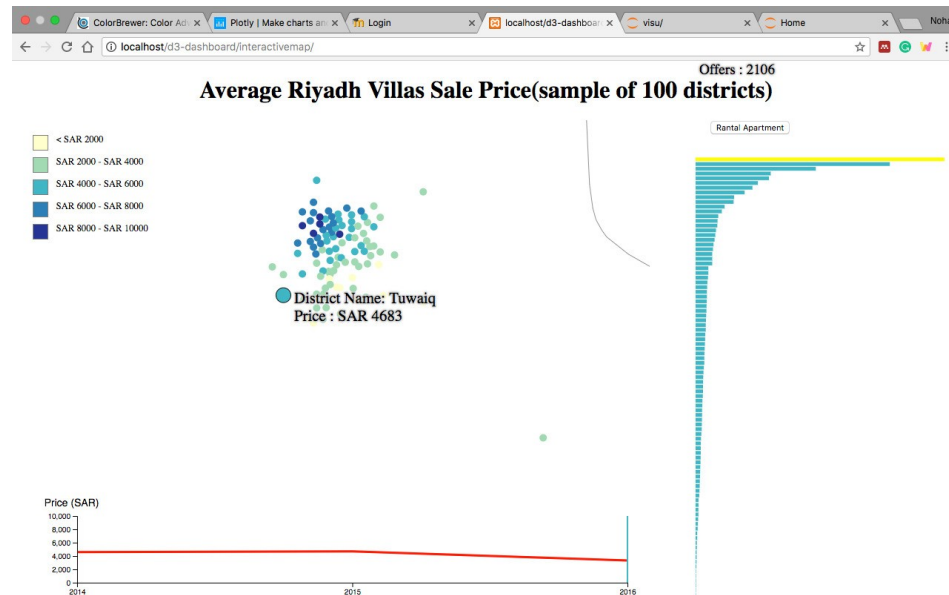


Figure 4.4: High number of offers in Tuwaiq district in villas sale price dashboard

## 4.2. Python plots

If any villa has some auxiliary services such as, driver room, or furniture, what is the effect on the price?

Compared to the base (not furnished villa with no driver room), results indicate that adding driver room will significantly increase the selling price. However, furniture will not add much to villas price.

Moreover, this could indicate that people depend heavily on their preference and selection of their homes where they attribute significant attention to the driver room. However, with the recent change in Saudi Arabia that allow women to drive starting from June 2018, we could see some alterations in people preference for their homes, and we might witness a slowing demand for driver room (Figure 4.6).

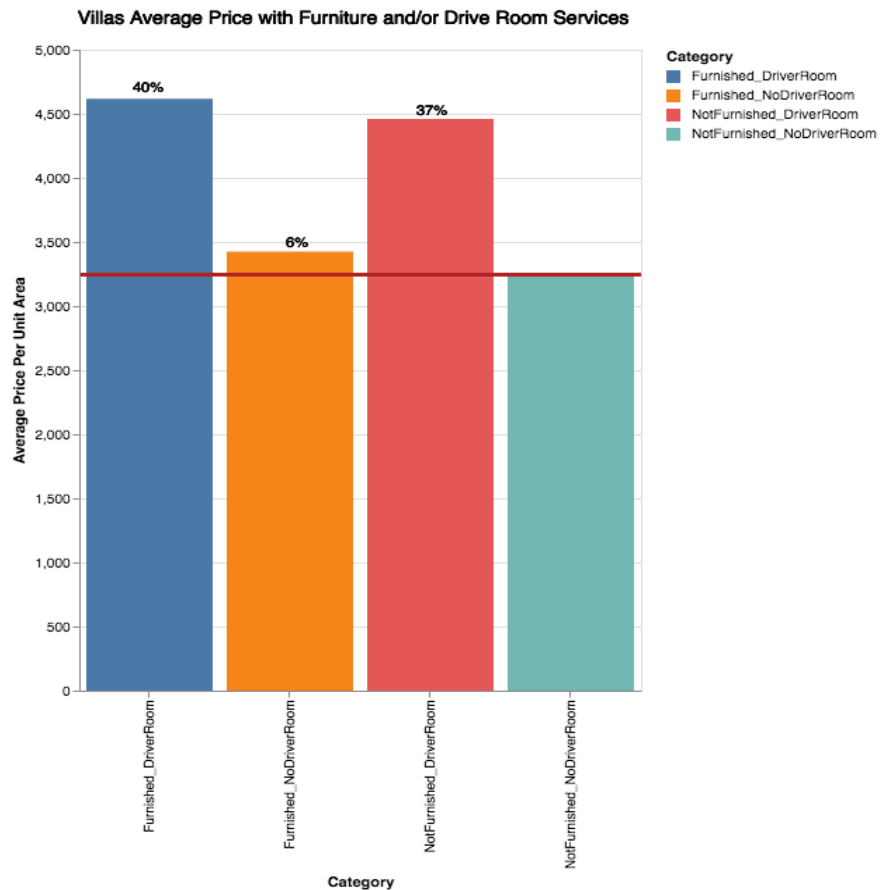


Figure 4.6: villas sale price with additional services

#### Is there any relationship between the villa price and street width?

The streets width of 15 and 20 meters are the most common streets width in Riyadh districts as its proven from our sample that most of the offers are for villas in these two types of streets. Price range starts from 1,911 SAR per unit area in villas that have a street width of 25, which is almost double the starting point of the villas that have a street width of 10, 15, or 20 (1,000 SAR). This indicates that the higher the street width, the higher unit rate. Moreover, it could indicate that high width streets are mostly located in modern high demanded areas. Figure 4.7 shows villa sale price categorized in four street types.

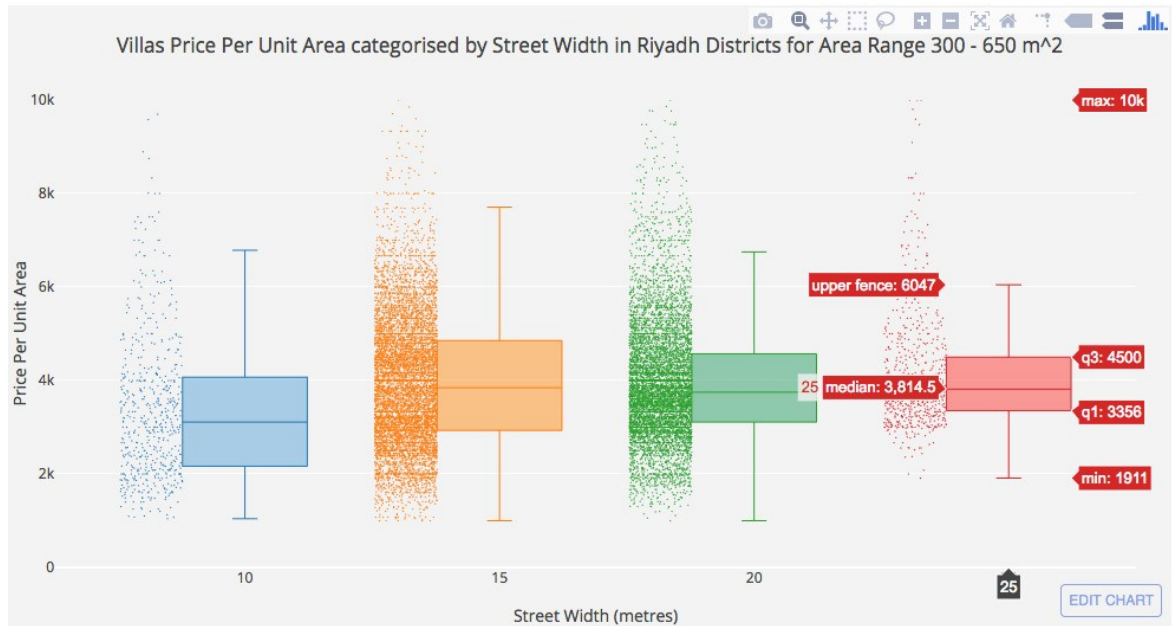


Figure 4.7: villas sale price in four street types