

Aqar Historic Data

CASE STUDY

Abdulaziz Salem Almalki
aalmalky88@gmail.com

Data

Historical data for sales deal indicators, including the number of sale deals, their total values, and the average price per meter according to location and property classification

Objectives:

1. EDA which exploring the data examining and understanding the characteristics of your dataset
2. Cleaning the data and make it more ready for the analysis and visualization
3. visualizing the data for gaining insights, identifying patterns, and communicating findings in a clear and effective manner
4. applying a regression technique for to more understanding of the relationship between prices and time

Tools:

1. Python.
2. SQL.
3. Excel.
4. Tableau

Step 1:

EDA :

- A. Basic Information about the Dataset:
 - the first few rows of the dataset to get an overview.
 - the data types of each column.
 - missing values.
 - Summary of the dataset

jupyter EDA + Regression Last Checkpoint: 9 minutes ago (autosaved)

```

File Edit View Insert Cell Kernel Widgets Help


```

In [2]:

```

import warnings
warnings.filterwarnings("ignore")
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
pio.templates
from plotly.offline import init_notebook_mode, iplot # plotly offline mode
init_notebook_mode(connected=True)

```

In [3]:

```

AqarData = pd.read_csv('AqarRep.csv')

```

In [3]:

```

# Display the first few
print(AqarData.head())

```

Display data types and missing values

```

print(AqarData.info())
print(AqarData.isnull().sum())

```

	yearnumber	quarternumber	quartername_ar	quarterid	region_ar	city_ar	\
0	2023	1	منطقة الرياض	20231	الربع الأول		
1	2018	3	منطقة الرياض	20183	الربع الثالث		
2	2022	2	منطقة الرياض	20222	الربع الثاني		
3	2021	3	منطقة الرياض	20213	الربع الثالث		
4	2020	4	منطقة الرياض	20204	الربع الرابع		

	district_ar	typecategory_ar	deed_counts	RealEstatePrice_SUM	\
0	120000.0	1	آخر		الأندلس
1	1000000.0	1	آخر		الأندلس
2	167177.0	1	آخر		الأندلس
3	475000.0	1	آخر		الأندلس
4	1000000.0	1	آخر		الأندلس

	Meter_Price_W_Avg_IQR
0	Nan
1	Nan
2	Nan
3	Nan

In [4]:

```

# Summary statistics
print(AqarData.describe())

```

	yearnumber	quarternumber	quarterid	deed_counts	\
count	32730.000000	32730.000000	32730.000000	32730.000000	
mean	2020.808280	2.448029	20210.530828	58.224809	
std	1.687726	1.087204	16.799506	1108.821263	
min	2018.000000	1.000000	20181.000000	1.000000	
25%	2019.000000	1.000000	20194.000000	2.000000	
50%	2021.000000	2.000000	20213.000000	6.000000	
75%	2022.000000	3.000000	20224.000000	18.000000	
max	2023.000000	4.000000	20233.000000	70678.000000	

	RealEstatePrice_SUM	Meter_Price_W_Avg_IQR
count	3.273000e+04	1.632000e+04
mean	5.247090e+07	3.501168e+03
std	6.657379e+08	1.126499e+05
min	1.000000e+00	9.500000e-04
25%	1.520000e+06	9.667075e+02
50%	6.153582e+06	2.163789e+03
75%	2.325657e+07	3.499858e+03
max	7.393738e+10	1.437036e+07

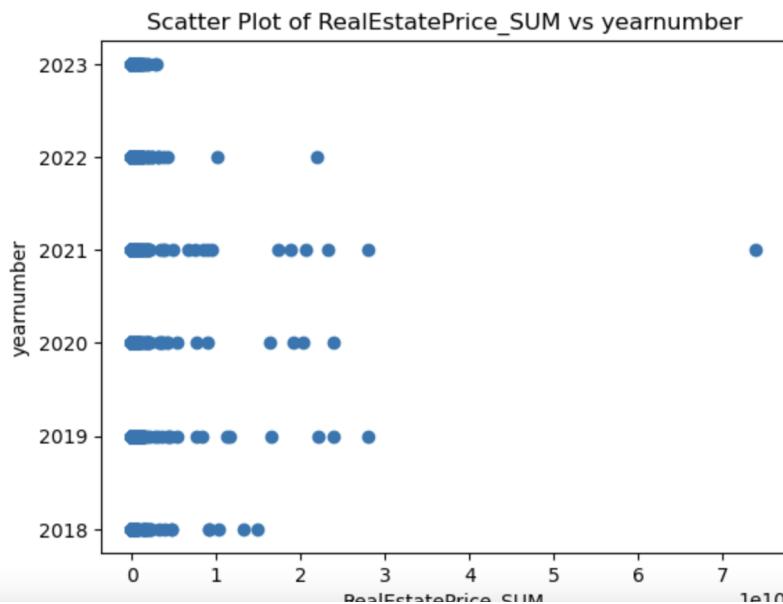
B. visualization for more clear understanding

- Histogram.
- distribution of categorical variables.
- Boxplot to identify outliers.

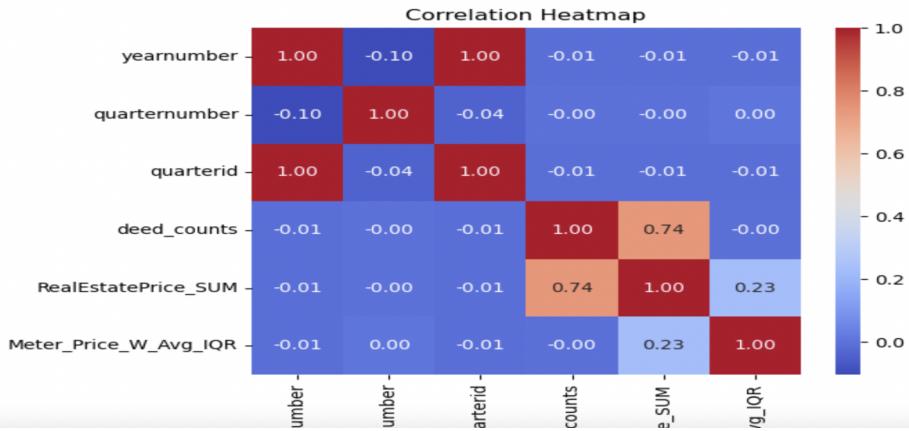
```
In [7]: plt.hist(AqarData['RealEstatePrice_SUM'], bins=20, color='blue', edgecolor='black')
plt.title('Distribution of prices')
plt.xlabel('RealEstatePrice_SUM')
plt.ylabel('Frequency')
plt.show()
```



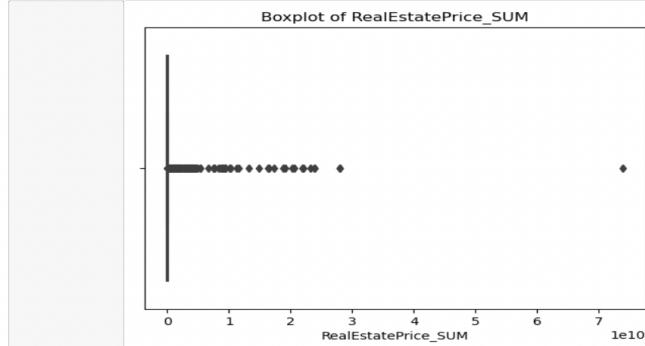
```
In [8]: # Scatter plot of two variables
plt.scatter(AqarData['RealEstatePrice_SUM'], AqarData['yearnumber'])
plt.title('Scatter Plot of RealEstatePrice_SUM vs yearnumber')
plt.xlabel('RealEstatePrice_SUM')
plt.ylabel('yearnumber')
plt.show()
```



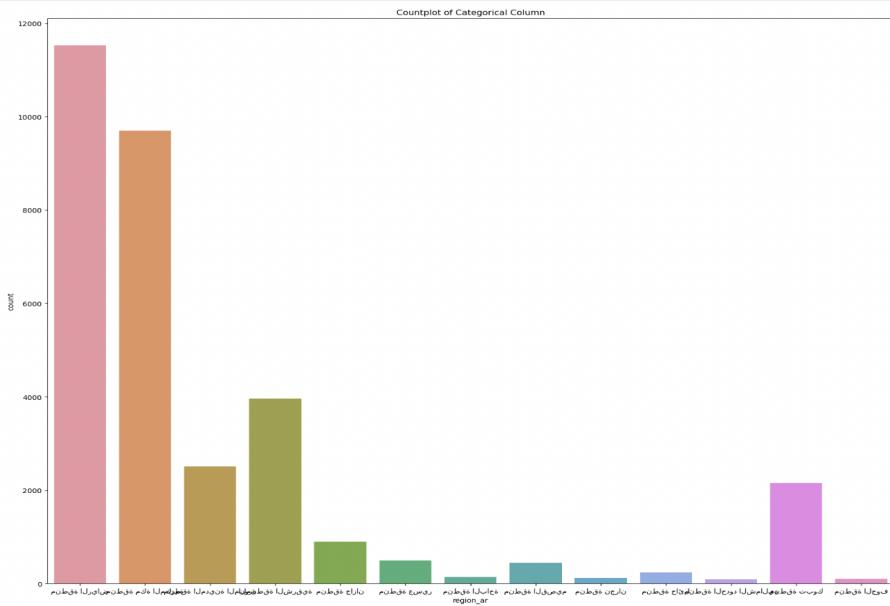
```
In [9]: #Correlation heatmap
correlation_matrix = AqarData.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



```
In [17]: # Boxplot to identify outliers in a numerical column
sns.boxplot(x=AqarData['RealEstatePrice_SUM'])
plt.title('Boxplot of Numeric Column')
plt.show()
```

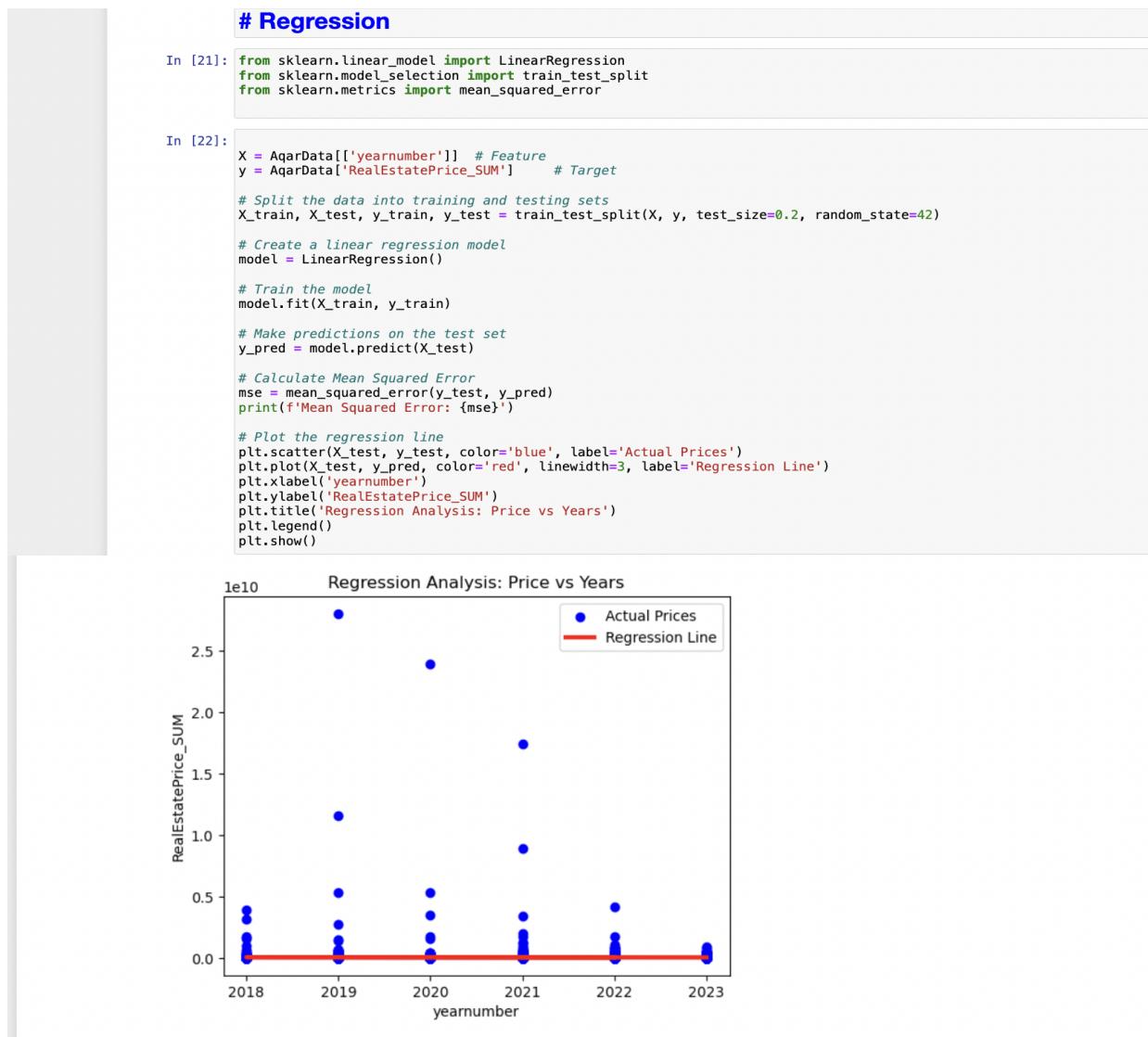


```
In [16]: #Countplot of a categorical variable
plt.figure(figsize=(20, 16))
sns.countplot(x="region_ar", data=AqarData)
plt.title('Countplot of Categorical Column')
plt.show()
```



C. Regression analysis

- LinearRegression.



Step 2:

Data Cleaning :

- Rename some columns.
 - Replacing blanks and nulls.

```
connect + DESKTOP-JASP270\SOLEXPRESS (SQL Server 16.0.1000 -)
| Databases
|   System Databases
|   Database Snapshots
|   Aqar
|     Database Diagrams
|   Tables
|     System Tables
|     FileTables
|     External Tables
|     Graph Tables
|     dbo.Data
|   Views
|   External Resources
|     Synonyms
|     Programmability
|     Query Store
|     Service Broker
|     Storage
|     Security
|   Jadar Database
|   PortfolioProjects
|     Security
|     Server Objects
|     Replication
|     PolyBase
|     Management
|     XEvent Profiler

Select *
From dbo.Data

--Rename a Column
sp_rename 'Data.RealEstatePrice_SUM', 'Price'

--Replacing Blanks or Nulls with the Value غير ممتد
UPDATE dbo.Data
SET region_ar = COALESCE(region_ar, 'غير ممتد')
WHERE region_ar IS NULL;

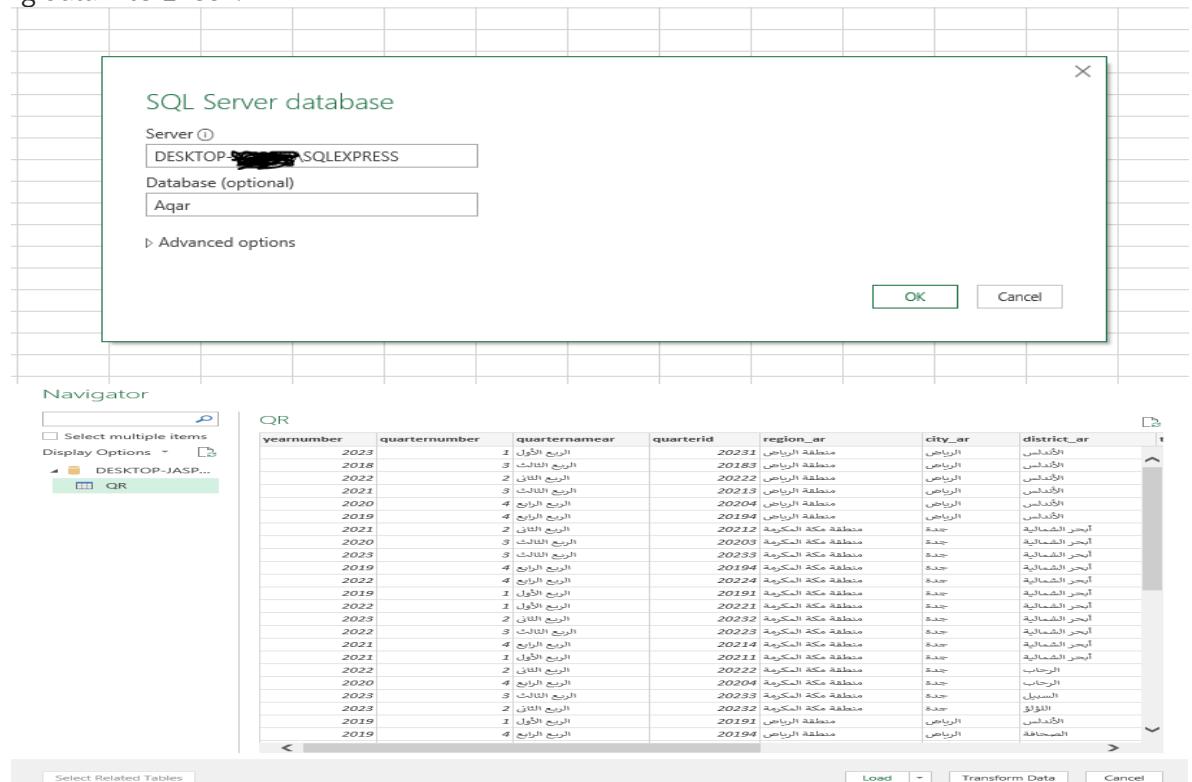
UPDATE dbo.Data
SET city_ar = COALESCE(city_ar, 'غير ممتد')
WHERE city_ar IS NULL;

UPDATE dbo.Data
SET district_ar = COALESCE(district_ar, 'غير ممتد')
WHERE district_ar IS NULL;

Select *
From dbo.Data
```

Step 3:

Exporting data into Excel :

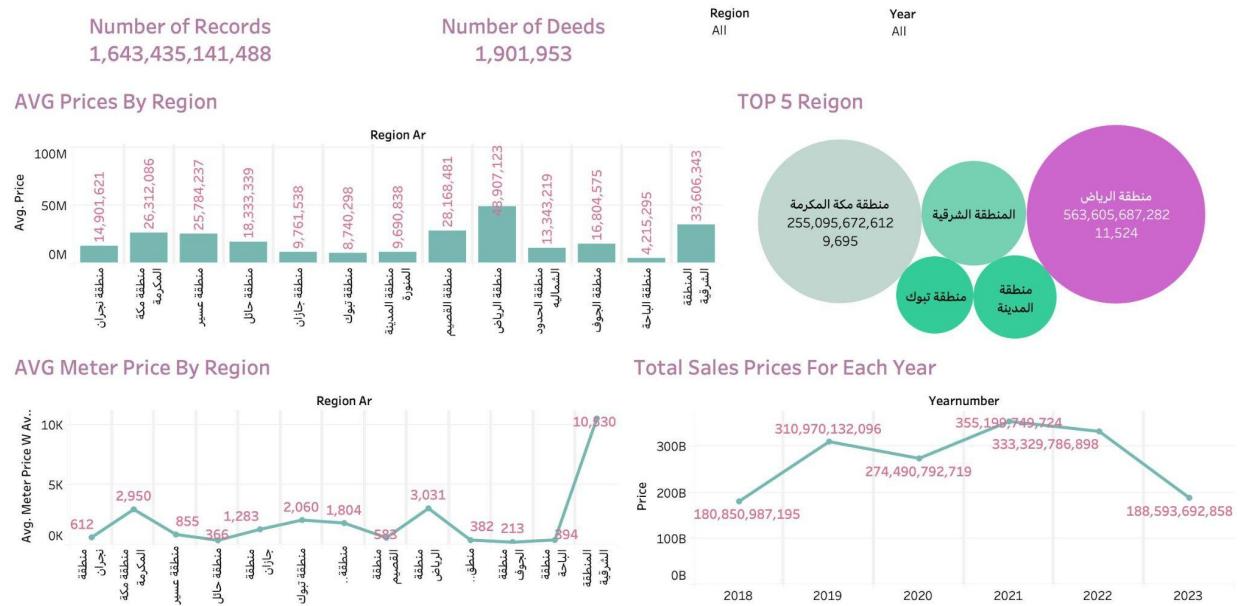


Step 4:

Creating the Dashboard Using Tableau:

- Number of total Records and Deeds.
- AVG for Prices by Region and Meter Price by Region.
- Total sales Prices by Regions and top 5 Regions.

Aqr Dashboard | Overview



Access the live Dashboard By Clicking [Here](#)

Insights

In this study, we try to describe and analyze historical Realestate “Aqr” data through the use of a type of analysis called descriptive analysis, and what we have achieved is describing the current data and clarifying the relationship between the columns and various factors such as “price changes in each year”.

The final result is a comprehensive analysis and understanding of the historical data.