

Exploring an international market to understand factors impacting price

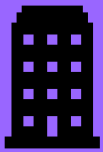
Case Study by Jackson Lanier

Project Overview

My interests in real estate and market trends on an international level connected with my role as a data analyst to produce a project which explores variables impacting price in Portuguese real estate. I completed this project over the course of a month with the guidance of my tutor, Ankur Bhargave, and my mentor, Anurag Srivastava.

Project Objectives

- Source, clean, and analyze data from the web
- Use statistical analyses, including linear regression and cluster analysis, to explore relationships within the data
- Further explore the data through geospatial and time series analysis
- Create a Tableau storyboard to present relevant findings in a digestible format



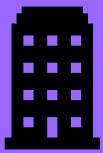
Data Overview

- Portuguese real estate data sourced from [Kaggle](#)
- Historical residential property prices source from the [Federal Bank of St. Louis](#)
- Geographic data on Portugals region's sourced from [simplemaps](#)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Price	District	City	Town	Type	EnergyCertificate	Floor	Lift	Parking	HasParking	ConstructionYear	TotalArea	GrossArea
2	250000	Faro	São Brás de Alportel	São Brás de Alportel	Apartment	A+	2nd Floor	TRUE	1	TRUE		114	
3	9500	Faro	Albufeira	Albufeira e Olhos de Água	Apartment	NC	1st Floor	TRUE	0	FALSE	1990	27	
4	580000	Faro	Vila do Bispo	Budens	Apartment	D	3rd Floor	FALSE	1	TRUE	2003	84	
5	350000	Faro	Portimão	Portimão	Apartment	C	4th Floor	TRUE	0	FALSE	1985	68	
6	175000	Faro	Faro	Faro (São e São Pedro)	House	NC		FALSE	0	FALSE	1950	78	
7	1485000	Faro	Loulão	Quarteira	House	D	Ground Floor	FALSE	2	TRUE	2004	472	
8	1500000	Faro	Faro	Faro (São e São Pedro)	Apartment	A	5th Floor	TRUE	3	TRUE	2018	421	313
9	25000	Faro	Portimão	Portimão	Garage	NC	Basement Level	TRUE	1	TRUE	1982	15	
10	260000	Faro	Lagos	São Gonçalo de Lagos	Apartment	C	2nd Floor	FALSE	0	FALSE	1999	105	114
11	235000	Faro	Faro	Faro (São e São Pedro)	House	NC		FALSE	0	FALSE		63	
12	320000	Leiria	Peniche	Peniche	Land	NC		FALSE	0	FALSE		3320	0
13	595000	Faro	Lagos	São Gonçalo de Lagos	Apartment	A	Ground Floor	TRUE	1	TRUE	2022	98	112
14	100000	Faro	Alcoutim	Alcoutim e Pereiro	Land	NC		FALSE	0	FALSE		24920	
15	420000	Faro	Portimão	Alvor	House	D		FALSE	1	TRUE	1998	140	185
16	1700000	Faro	Albufeira	Albufeira e Olhos de Água	House	B-		FALSE	3	TRUE	2001	800	
17	340000	Faro	Loulão	São Sebastião	Apartment	A	3rd Floor	TRUE	1	TRUE	2023	94	110
18	780000	Vila Real	Valpaços	Carracedo de Montenegro e Curros	Farm	NC		FALSE	0	FALSE		552450	200
19	250000	Faro	São Brás de Alportel	São Brás de Alportel	Apartment	A+	1st Floor	TRUE	1	TRUE		114	
20	2950000	Faro	Albufeira	Albufeira e Olhos de Água	Building	NC		FALSE	0	FALSE	1998	265	465
21	158000	Faro	Portimão	Portimão	Apartment	D	2nd Floor	TRUE	1	TRUE	1993	21953	48
22	250000	Faro	Loulão	São Clemente	Apartment	C	Ground Floor	TRUE	0	FALSE	2009	85	96
23	620000	Faro	Lagos	São Gonçalo de Lagos	Apartment	A	2nd Floor	TRUE	1	TRUE	2022	94	116
24	595000	Faro	Lagos	São Gonçalo de Lagos	Apartment	A	1st Floor	TRUE	1	TRUE	2022	80	106
25	435000	Faro	Loulão	São Clemente	House	D	Ground Floor	FALSE	2	TRUE	1958	226	
26	615000	Faro	Lagos	São Gonçalo de Lagos	Apartment	A	2nd Floor	TRUE	1	TRUE	2022	80	106
27	185000	Faro	Loulão	Quarteira	Apartment	D	Above 10th Floor	TRUE	0	FALSE	1978	38	48
28	185000	Faro	Portimão	Portimão	Apartment	D	Above 10th Floor	TRUE	0	FALSE	1982	38	
29	210000	Faro	Faro	Conceição e Estoi	Farm	NC		FALSE	3	TRUE	1951	4600	177
30	495000	Faro	Loulão	Quarteira	Apartment	C	1st Floor	FALSE	1	TRUE	2003	139	
31	469000	Faro	Loulão	São Clemente	Building	E		FALSE	2	TRUE	1978	203	
32	139000	Faro	Loulão	Boliqueime	Land	NC		FALSE	0	FALSE		32966	
33	19000	Faro	Lagoa (Algarve)	Estão mbar e Pargal	Land	NC		FALSE	0	FALSE		3710	3710
34	272000	Faro	Albufeira	Albufeira e Olhos de Água	Apartment	A	1st Floor	TRUE	1	TRUE		10560	67
35	291000	Faro	Albufeira	Albufeira e Olhos de Água	Apartment	B	1st Floor	TRUE	1	TRUE		10560	67

A sample image of the Portuguese real estate data downloaded as a CSV file from Kaggle







Techniques Applied

- Sourcing data suitable for analysis
- Cleaning data, including the use of machine learning to impute missing values
- Analyzing the relationships between variables within the data to identify the impact variables upon one another
- Forming and answering research questions related to the data
- Presenting relevant findings, including limitations and future possibilities

Tools Used

-  Kaggle
 - Data Sourcing
-  Anaconda's Jupyter Notebook
 - Data analysis using Python
 - Python libraries including NumPy, Pandas, Seaborn, Matplotlib, and scikit-learn
-  Tableau
 - Visualization of analyses
 - Presentation of findings

03. Exploring & Cleaning Data

```
[1980]: # Check dataframe's shape
df.shape

[1980]: (114623, 25)

[1981]: # Print the first 5 rows of the dataframe
df.head()

[1981]:
```

	Price	District	City	Town	Type	EnergyCertificate	Floor	Lift	Parking	HasParking	...	Elevator	ElectricCarsCharging	TotalRooms	NumberOfBed
0	250000.0	Faro	São Brás de Alportel	São Brás de Alportel	Apartment	A+	2nd Floor	True	1.0	True	...	NaN	NaN	2.0	
1	9500.0	Faro	Albufeira	Albufeira e Oihos de Água	Apartment	NC	1st Floor	True	0.0	False	...	NaN	NaN	0.0	
2	580000.0	Faro	Vila do Bispo	Budens	Apartment	D	3rd Floor	False	1.0	True	...	NaN	NaN	2.0	
3	350000.0	Faro	Portimão	Portimão	Apartment	C	4th Floor	True	0.0	False	...	NaN	NaN	2.0	
4	175000.0	Faro	Faro (São e São Pedro)	Faro (São e São Pedro)	House	NC	NaN	False	0.0	False	...	NaN	NaN	4.0	

5 rows x 25 columns

```
[1982]: # View all column names
df.columns.values

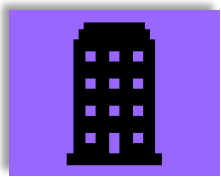
[1982]: array(['Price', 'District', 'City', 'Town', 'Type', 'EnergyCertificate', 'Floor', 'Lift', 'Parking', 'HasParking', 'ConstructionYear', 'TotalArea', 'GrossArea', 'PublishDate', 'Garage', 'Elevator', 'ElectricCarsCharging', 'TotalRooms', 'NumberOfBedrooms', 'NumberOfWC', 'ConservationStatus', 'LivingArea', 'LotSize', 'BuiltArea', 'NumberOfBathrooms'], dtype=object)

[1983]: # Check for null values
print(df.isnull().sum())

Price          244
District        0
City            0
Town            2
Type            16
EnergyCertificate 14
Floor          86694
Lift           46488
Parking         154
HasParking     46488
ConstructionYear 41873
TotalArea      6457
```

This screenshot shows my Jupyter notebook used for the initial data exploration and cleaning



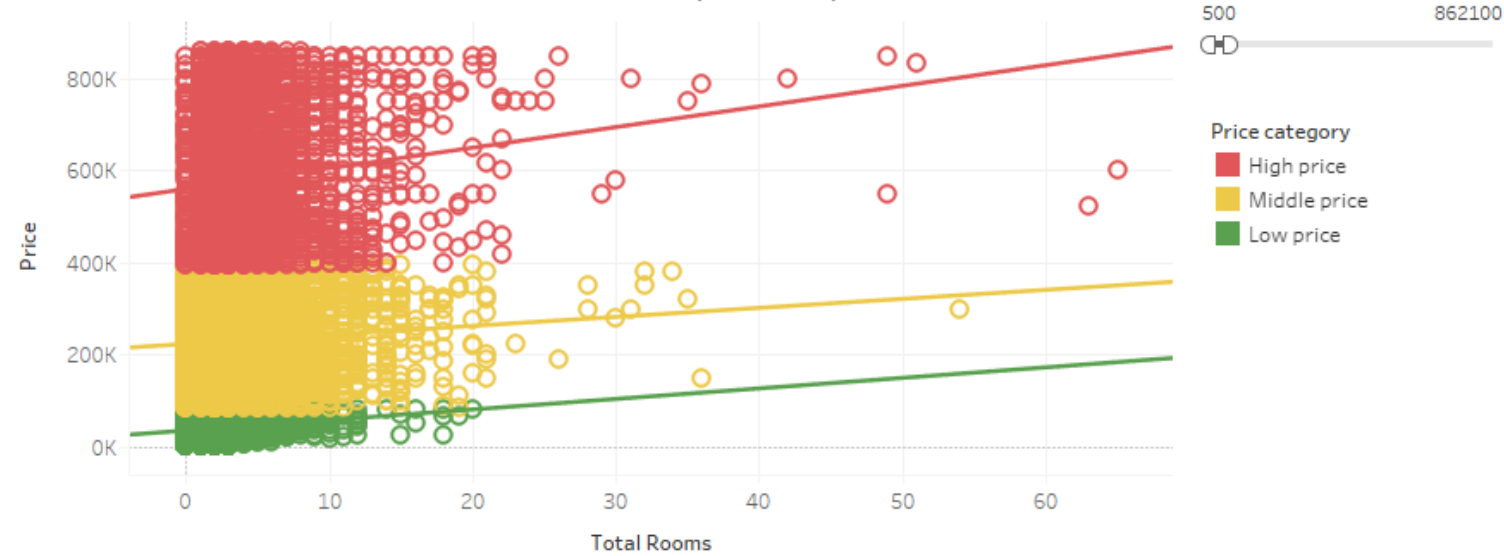


Analysis Highlights

- Portuguese real estate contains properties with selling prices as high as 36 million euros but 94% of properties are listed for less than 1 million euros.
- Even when removing high-priced outliers from the data, no strong relationship is found between price and one specific variable
- Property location and the state of the global economy appear to impact real estate prices but their impact needs further exploration

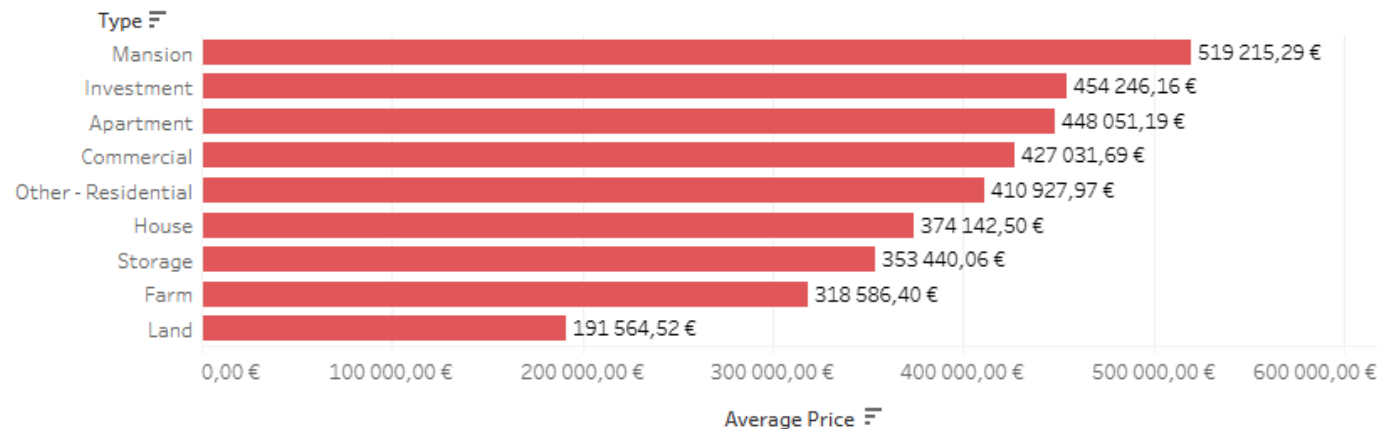
Note that some property types were grouped together during the cleaning process, such as “Mansion” and “Manor”

Price per Total Rooms (Filtered)



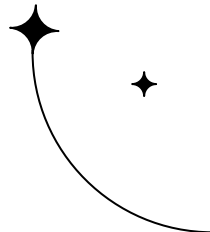
The properties with the most total rooms tend to be in the high price range. However, there is no statistically relevant strength to the relationship between price and total rooms.

Average Price of Property Types



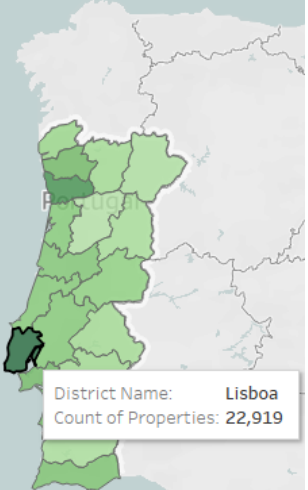


Additional Highlights



Mountains, river valleys, dry plains, marshes, and beaches make up the country of Portugal's diverse landscape. It is no surprise that a country with such varying landscapes is home to a complex real estate market.

Portuguese real estate data sourced from Kaggle provides current information on nearly 100,000 real properties for sale across the mainland and islands. Analyzing this data will provide a better understanding of the Portuguese real estate market.

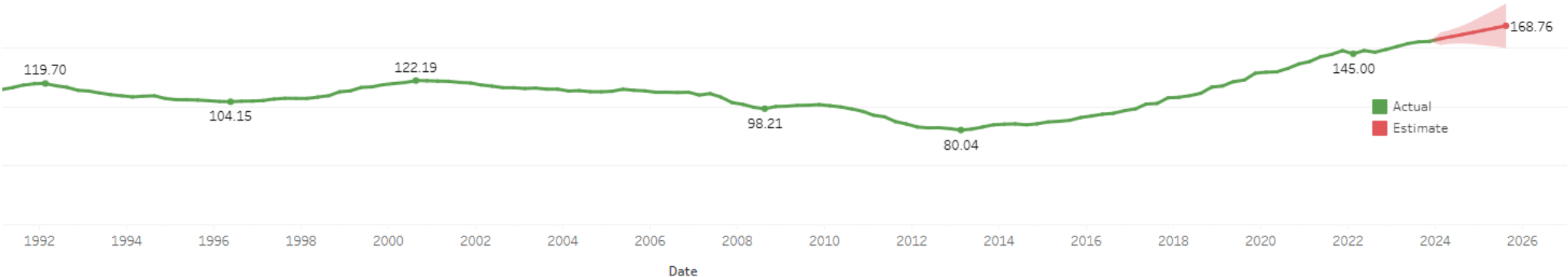


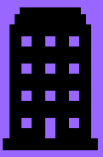
Lisboa, the coastal district containing Portugal's capitol, contains more available properties than any other district

Historical data on the Portuguese market allows for an understanding of prices over time and a projection of future prices



Indexed Real Estate Prices per Year





Data Limitations and Future Analysis

- Limitations included the presence of null and irregular values in the original data as well as the massive variety of property types and range of variables present in Portuguese real estate.
- Exploring real estate comes with the general limitation that the market is constantly changing and is affected by a near endless number of factors
- Future analyses could dive deeper into the relationship between price and location or price and current economic conditions. Additionally, future analyses could explore the differences between asking prices and selling prices in the market.

Personal Takeaways

- Data may not contain obvious relationships between variables. The lack of a strong relationship can be just as meaningful as a relationship with strength.
- The data cleaning process, such as removing outliers or imputing values, is a sensitive process. Comparing an analysis with and without the presence of outliers can be useful.
- Sourcing useable, relevant, available data can be a time-consuming process. The value of data can't be understated.

Thank you!

Please check out the links below for the complete project

[Link to GitHub Repository](#)

[Link to Tableau Storyboard](#)