Forecasting real estate prices in Portugal

A data science approach

Sanam Samadani
ISEG - Lisbon School of Economics & Management,
University of Lisbon
Lisbon, Portugal
samadanisanam@aln.iseg.ulisboa.pt

Carlos J. Costa
Advance/CSG, ISEG - Lisbon School of Economics &
Management, University of Lisbon
Lisbon, Portugal
cjcosta@iseg.ulisboa.pt

Abstract — **The real estate market is vital from an economic and social perspective. It is why it is essential to understand its behaviour. In this context, our work aims to analyze the prices' evolution and understand what the main components are impacting the price. To understand the evolution and predict prices, we perform a time series analysis, including ARIMA. To understand the most relevant components, we perform a regression analysis. In this analysis, we used several approaches. We have collected our data from INE and PORDATA, the Databases of Contemporary Portugal, to construct the models and forecast house prices in Portugal.**

Keywords – real estate; prices; prediction; ARIMA; Regression.

I. Introduction

1.1.Statement of the problem

The real estate sector is of the utmost importance to the Portuguese economy. In fact, housing is the main asset of households: for instance, in 2017, real estate accounted for 48% of total family wealth.[1] Residential real estate provides shelter, preserves wealth. Therefore, access to accurate house price prediction is vital to central banks, financial supervision authorities, investors, and homeowners.[2] Due to covid, it is vital to analyze if the house prices fall or increase in the following years. Many foreign buyers may have put plans on hold due to uncertainties of Brexit and the coronavirus pandemic. However, despite such uncertainties, Portugal's property market has been relatively resilient to the crisis, with prices rising in 2020. [3] According to Eurostat data, the increase in house prices and rents is much higher in Portugal than the average for E.U. and Eurozone countries. In year-on-year terms, i.e. in the third quarter of 2020 compared with the same period in 2019, houses prices rose by 4.9% in the eurozone and 5.2% in the E.U., with Portugal rising above the average with property price increases of 7.1%.[4]

1.2. Objective

In this context, our purpose is to analyze the prices' evolution and understand what the main components are impacting the price. In other studies, the quality of the buildings has been studied and used to assess. Our paper evaluates variables such as crime, waste, purchasing power, and tax rate in various municipalities of Portugal. Although there are other possible variables, we find the mentioned factors essential to determining the properties' value. For example, changing levels of crime are likely to induce more immediate responses at the individual level. Increases in crime will directly impact an individual's perceptions regarding safety in a neighbourhood.[21]To figure out the evolution and predict prices, we perform a time series analysis, including ARIMA. To understand the most relevant components, we perform a regression analysis. In this analysis, we used several approaches and compared the results. Linear regression, Ridge Regression, Lasso Regression, Bayesian Regression, Polynomial Regression and Neural Network are the models we used in our analysis.

II. Literature Review

In Portugal, the rate of home ownership has risen from 57% in 1981 to 73% in 2011 and the number of secondary homes- from 7% in 1981 to 20% in 2011 of total family dwellings(INE,2012).[5] Portugal has one of the highest rates of homes per household in Europe- 1.70 in 2011(rising from 1.16 in 1980), coming third at the European level, after Spain and Ireland. Also, the growth of the financial markets directly impacted the housing sector, contributing to the predominance of policies focusing on demand that have stimulated homeownership through the use of credit, to whose expansion the European integration contributed by successive lowering of interest rates.[6]

John M. Quigley, in his studies, focused on the linkage between economic 'fundamentals' and property prices which is based upon a detailed body of data from the U.S. in particular, it includes a detailed comparison of the importance of "fundamentals" upon housing prices relative to the importance of "history" in affecting outcomes. Also, it focuses on answering the question- the potential for a causal role between outcomes in the property market and the subsequent health of the overall economy relevant to the economic conditions in many Asian economies from the '80s until the end of the century. Their study shows clear economic fundamentals do not explain most of the variation in the property prices in the short run, and bubbles in Asian property markets had real

2021 16th Iberian Conference on Information Systems and Technologies (CISTI)
23 – 26 June 2021, Chaves, Portugal
ISBN: 978-989-54659-1-0

consequences for the course of national and regional economic conditions during the late 1990s. [7]

Although the recession during 2003 to 2013, In China's top cities, actual prices grew 13.1 per cent annually, increasing the number of employed people in construction, which is 16 per cent of urban employment. However, the United States and Spain accounted for 8 per cent and 13 per cent of that, respectively. Also, it has been mentioned a housing crash after a boom is not inevitable in China and only if new construction is sufficiently restricted can prices remain high.[8]

According to Temur and colleagues, the surplus in the housing supply may cause undesired price reductions, resulting in firms operating in this sector facing various problems, such as not being able to sell what they produce or selling at a lower cost. The failure of the housing supply to meet demand affects the welfare of individuals who want to buy housing as a shelter of investment. Therefore, an accurate estimation of real estate sales is of great importance for balancing supply and demand in the housing market. Also, as primary housing markets have become more integrated with secondary markets, the computation of housing prices has become of great practical importance to investors who confront choices among portfolios composed of housing securities and other investment assets. [9] On the other hand, several researchers published empirical works showing that decreasing crime does, in fact, benefit lower-valued properties disproportionately, reducing the inequality among properties.[10] They mention the price inequality can potentially decrease as low-priced properties react more to crime reductions.[11]

Giudice and colleagues analyzed the effect of COVID in the Italian real estate markets. According to them, the housing prices will reduce 4.16% in the short-run and 6.49% in the mid-run, "predatory" housing may occur in the short-run, leading to change in the national and local economic geography. The greatest danger for the national and local economy is the income impoverishment that will arise as an effect induced by forced inaction.[12]

In 2020, Nicola and colleagues summarized the socioeconomic effects of COVID-19 on the world economy's aspects. Other authors compared this pandemic situation with previous similar situations. Economic research's main findings on the 2003 Hong Kong SARS epidemic estimated a 1.75% loss in annualized Gross Domestic Product (GDP), or 5.1% monthly loss at a peak, 1.3% increase in unemployment rate, residential real estate prices, and real estate transactions. [13]

P. Angelov and colleagues assess the effect of crime on the sale price of properties and make predictions with three machine learning algorithms. They used two data sets containing the property values and tax information for a County in Washington state and also the crime information for the same County; 13 property attributes, 5 crime attributes. The results show the Sale Price has negative correlation coefficients with all crime types. They contribute the buyers of a property to be aware of the neighbourhood's crime rate where the property is located since most real estate companies, or websites do not necessarily release such data. The customers mainly consider the physical attributes such as the size of the house or apartment. [14] While homeownership is often viewed as a way to enable households to build wealth, threats to that investment's value may limit its appeal. One such threat is crime, which may reduce the desirability of ownership in affected neighbourhoods.[21] Regarding the levels of crime, in his research, Dugan found that if victimization near one's home increases the probability that people move to a new residence.[22]

P. Ofori, in his study, demonstrates the importance of waste dumping on real estate rentals. According to his study, most people are not ready to pay higher rents if buildings are close to a filthy area like a dumpsite, irrespective of building type.[15] Moreover, in their studies, A. Hamid and Christopher assess the locational effect of noise and water pollution on nearby property values. Their analysis revealed that noise and water pollution were seen as a discomfort and therefore all properties located close to noise pollutants were sold at lower prices.[16]

Cecilia Rocha et al. analyses the noise influence on real estate values as the Portuguese Noise Code enforces building restrictions on municipal urban areas exceeding established noise limits. On these areas, and until mitigation measures enable noise reduction, private contractors will have their building permits refused on an excessive noise reduction. In 2003, the Portuguese government issued a law, concerning Real Estate Taxation. As municipalities will have to communicate the inadequacy of vacant land for building construction, the following steps will reduce nominal real estate value and the taxes income.[23]

III. Methodology

In what concerns the methodology, we followed a data science approach [17]. We started by analyzing the building market. Then, we identified the most critical variables available to work-study. We processed the data and developed the main models to answer the main questions. Supported in the literature review, we selected purchasing power (purchacingPower), crime rate (CrimeRate), taxes (IMTpercapita and IMIpercapita) and selected usage (waste).

This study's data are drawn from two sources: PORDATA, the database of contemporary Portugal, and INE (Instituto Nacional de Estatistica). The data includes values and the numbers of houses in Portugal, quarterly, from 2009Q2 to 2020Q3, containing 47 observations.

We also obtained data from PORDATA corresponding to the information about each municipality. The number of observations according to the municipalities are 341. The independent variables which we want to study their importance on real estate prices in each municipality are : Purchasing power, crime rate, waste rate, IMI and IMT.

To understand the evolution and predict prices, we perform a time series analysis, including ARIMA. [20] To understand the most relevant components, we perform a regression analysis. In this analysis, we used several approaches.

SARIMA (Seasonal Autoregressive Integrated Moving Average). or Seasonal ARIMA is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (A.R.), differencing (I) and moving average

(M.A.) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. Four seasonal elements are not part of ARIMA that must be configured: P: Seasonal Autoregressive order, D: Seasonal difference order, Q: Seasonal moving average order and m: The number of time steps for a single seasonal period.

IV. Results

We started by analyzing the evolution of volume, quantity and prices. This analysis was performed considering new buildings and used buildings.

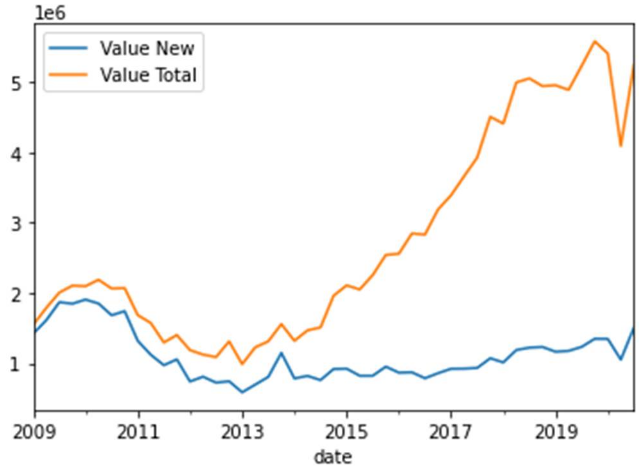


Figure 1. Evolution of the total value of buildings sold. Values of new buildings and the total real estate market.

The evolution of the quality of building sold is similar to the evolution of the total value.

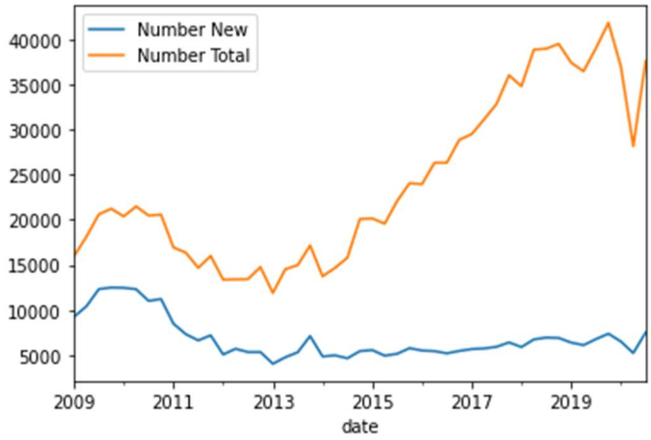


Figure 2. Evolution of the number of buildings sold. The number of new buildings and the total real estate market.

Analysis prices it is possible to identify an increasing trend with a slight price reduction in 2013.

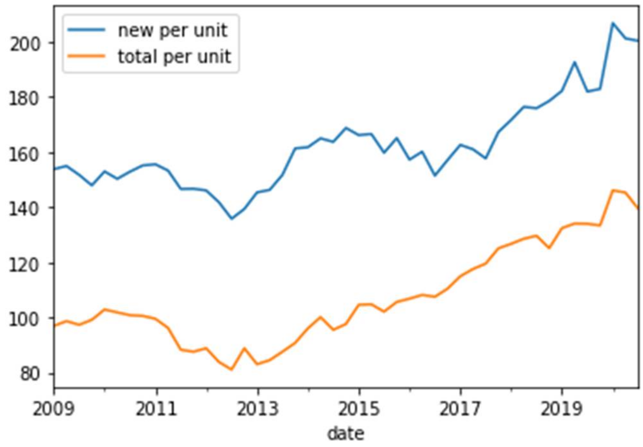


Figure 3. Evolution of prices of buildings sold. Prices of new buildings and total real estate market.

Considering the prices, in the following analysis, we show the main results of an ARIMA analysis.

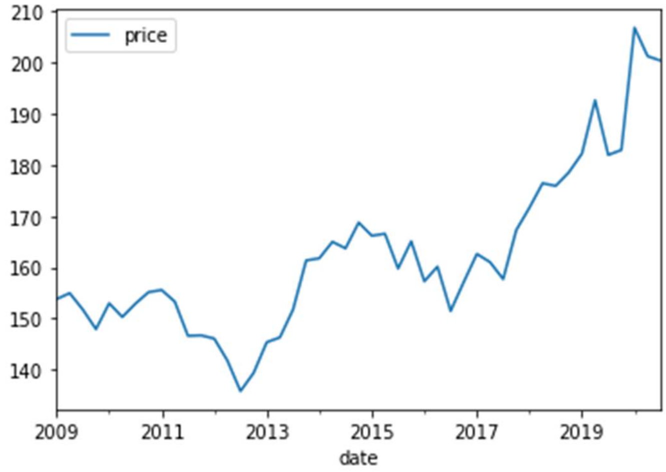
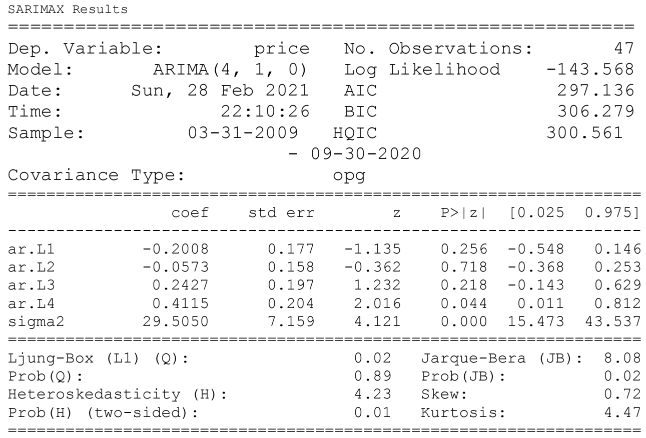


Figure 4. Evolution of prices of building sold.

The ARIMA model was created and fitted. The best result was for AR(4) with one first difference. A low P-value, less than 0,05, means we can reject the null hypothesis, here only ar.L4 seems to be less than 0,05 and statistically significant. Also, sigma2 is 0.000 and statistically significant. It calculated the t-statistic: t = estimated coeff. /std.error of coeff to 1.96. ACF of the residuals, if it is a good model, all autocorrelations for the residual series should be non-significant. Box-Pierce (Ljung) tests for possible residual autocorrelation at various lags.



```
SARIMAX Results
=========================================================
Dep. Variable:           price   No. Observations:      47
Model:          ARIMA(4, 1, 0)   Log Likelihood    -143.568
Date:         Sun, 28 Feb 2021   AIC                297.136
Time:                 22:10:26   BIC                306.279
Sample:             03-31-2009  HQIC               300.561
                     - 09-30-2020
Covariance Type:                opg
==========================================================
                 coef    std err          z    P>|z|  [0.025  0.975]
----------------------------------------------------------
ar.L1         -0.2008      0.177     -1.135    0.256  -0.548   0.146
ar.L2         -0.0573      0.158     -0.362    0.718  -0.368   0.253
ar.L3          0.2427      0.197      1.232    0.218  -0.143   0.629
ar.L4          0.4115      0.204      2.016    0.044   0.011   0.812
sigma2        29.5050      7.159      4.121    0.000  15.473  43.537
==========================================================
Ljung-Box (L1) (Q):                 0.02  Jarque-Bera (JB):  8.08
Prob(Q):                            0.89  Prob(JB):          0.02
Heteroskedasticity (H):             4.23  Skew:              0.72
Prob(H) (two-sided):                0.01  Kurtosis:          4.47
==========================================================
```

Figure 5. Evolution of prices of buildings sold. Prices of new buildings and total sales of buildings.

This ARIMA model was used to forecast housing prices. One way to measure the accuracy of the forecast is with the RMSE. For this specific model, we obtained 8.463. In the graph, the red line represents the forecast, and the blue is the price series.

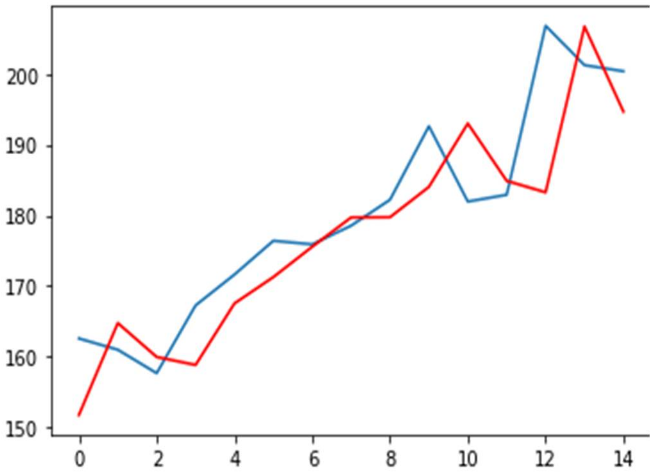


Figure 6. Evolution of prices of building sold. Prices of new buildings and total sales of buildings.

To understand the most important variables that explain the price, we created and estimated a regression model. We called it Model 1.

Purchasing power is a measure of the wealth of a region. In this case, we collected data corresponding to 2017

The crime rate is the number of crimes reported per 1000 inhabitants.

IMI is a tax over the transaction. Our analysis used the IMT paid in the municipality per capita (IMT 2018 per capita). The other tax analysis is the land value tax (IMI 2018, per capita). A land value tax or location value tax (LVT), also called a site valuation tax, split rate tax, or site-value rating is an ad valorem levy on the land's unimproved value.

The form of processing waste may be a possible way to evaluate the comfort of a building sold and also the sophistication level of the market. The variable used is the percentage of waste separated and recycled.

TABLE I. FEATURES DESCRIPTIONS

Feature Name	Description
purchacingPower	Purchase power per capita of the population form the municipality
CrimeRate	Crime rate per capita of the population form the municipality
IMTpercapita	Building purchasing tax
IMIpercapita	Real estate annual tax
waste	Percentage of selected waste

The model has as observations average data from municipalities. So the granularity does not give enough information to analyze this reality with precision.

We created a regression model where the price of 2018 houses is explained by purchasing power, crime, taxes, and processing. We used the OLS (Ordinary least square) and fitted the model. The main statistics may be seen in the following table. The main statistics may be seen in the following table. As R-squared is 80% and F-statistics is higher than 4, we conclude that the model is a good fit.

TABLE II. MODEL 1 MAIN STATISTICS

Dep. Variable:	price2018	R-squared:	0.804
Model:	OLS	Adj. R-squared:	0.801
Method:	Least Squares	F-statistic:	274.0
Date:	Tue, 26 January 2021	Prob (F-statistic):	5e-116
Time:	20:44:26	Log-Likelihood:	-3907.1
No. Observations:	341	AIC:	7826.
Df Residuals:	335	BIC:	7849.
Df Model:	5		
Covariance Type:	nonrobust		

After fitting the model, the relative importance and significance of each variable is presented in Table III.

TABLE III. MODEL 1 FEATURES

	coef	std err	t	P>|t|	[0.025	0.975]
const	-6.22e+04	6645.580	-9.360	0.000	-7.5e+04	-4.9e+04
purchacingPower	1393.6988	82.410	16.912	0.000	1231.592	1555.805
crimeRate	-773.0046	176.526	-4.379	0.000	-1120.24	-425.766
IMTpercapita	376.1759	23.760	15.832	0.000	329.437	422.914
IMIpercapita	-21.2276	30.317	-0.700	0.484	-80.864	38.409
waste	5.927e+04	1.75e+04	3.378	0.001	2.48e+04	9.38e+04

A brief analysis allows verifying that all the features are significant for 0.001, except IMIpercapita.

ISBN: 978-989-54659-1-0

TABLE IV. MODEL 1 ADDITIONAL STATISTICS

Omnibus:	40.391	**Durbin-Watson:**	1.536
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	214.887
Skew:	0.256	**Prob(JB):**	2.18e-47
Kurtosis:	6.855	**Cond. No.**	2.57e+03

Figure 7. Evolution of prices of building sold. Prices of new buildings and entire building market.

The regression allows understanding that prices are related to the purchasing power and that crime and Taxes may impact the price. We performed a new linear regression to understand the evolution of the prices. In this case, Target is to understand the evolution of prices from 2000 to 2018 (Y=Price2018-Price2000).

TABLE V. MODEL 2 MAIN STATISTICS

Dep. Variable:	y	R-squared:	0.742
Model:	OLS	Adj. R-squared:	0.739
Method:	Least Squares	F-statistic:	193.1
Date:	Tue, 26 January 2021	Prob. (F-statistic):	2.2e-96
Time:	20:44:26	Log-Likelihood:	-3830.4
No. Observations:	341	AIC:	7673.
Df Residuals:	335	BIC:	7696.
Df Model:	5		
Covariance Type:	nonrobust		

The most significant variables are essentially the same as the previous model.

TABLE VI. MODEL 2 FEATURES

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.887e+04	5179.46	-5.574	0.00	-3.9e+04	-1.87e+04
purchacingPower	498.0330	65.141	7.645	0.00	369.89	626.170
varCrime	0.3513	3.453	0.102	0.91	-6.441	7.144
IMTpercapita	328.3973	19.437	16.895	0.00	290.16	366.632
IMIpercapita	-77.1252	24.047	-3.207	0.00	-124.42	-29.822
waste	2.89e+04	1.3e+04	2.08	0.04	1608.6	5.6e+04

A brief analysis allows verifying that all the features are significant for 0.001, except crime. It is interesting verifying that land value tax or location value tax has a negative impact on the price variation.

TABLE VII. MODEL 2 ADDITIONAL STATISTICS

Omnibus:	77.468	**Durbin-Watson:**	1.499
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1550.606
Skew:	-0.257	**Prob(JB):**	0.00
Kurtosis:	13.434	**Cond. No.**	5.47e+03

Figure 8. Evolution of prices of building sold. Prices of new buildings and total building market.

To improve the analysis, we used other models. We spit the sample into 70% to train and 30% to test. We compared linear regression with ridge regression, lasso regression, Bayesian regression, polynomial regression and neural networks. Our network's architecture has two layers with 9 and 7 neurons, ReLU activation, max_iter of 5000, and as solver the 'lbfgs' (Limited-memory Broyden–Fletcher–Goldfarb–Shanno).

TABLE VIII. COMPARISON OF MODELS

Model	Accuracy on the training subset	Accuracy on the test subset
Linear Regression	0.742	0.751
Ridge Regression	0.742	0.749
Lasso Regression	0.742	0.751
Bayesian Regression	0.740	0.741
Polynomial Regression	0.815	0.552
Neural Network	0.834	0.734

III. DISCUSSION

Considering the total values of building sales, there was an increase until 2010. Then, the total value reduced but increased since 2013. Before the mid of 2012, house prices fell around 3% on average per year from the beginning of the financial crisis.[12] Since 2013, Portugal's housing sector has experienced significant changes, with a strong upturn in real estate transactions and a rise in housing prices. This rise in housing prices has resulted from the conjunction of a low supply of real estate property and significant growth in demand.[13]

As we expected, linear regression allows an acceptable accuracy on the training subset. Neural networks and polynomial regression improve the training subset's accuracy but with lower accuracy on the test subset. This is a consequence of the possible overfit limitation of those approaches. But more sophisticated approaches may be used, like in other studies. [19]

IV. CONCLUSIONS

Our work aims to analyze the prices' evolution and understand the primary components impacting the price. To understand the evolution and predict prices, we perform a time series analysis, including ARIMA. We concluded an essential variability of quantity sold from this analysis, but prices have a crescent trend. It is possible to predict prices using ARIMA with some error even with the last year's covid impact. To understand the most relevant components, we perform a regression analysis. We conclude that purchase power, crime and taxes may have a substantial effect on the prices. In this analysis, we used several approaches, like neural networks. Results were not much better than with linear regression. It is also essential to consider that our approach has some limitations, especially regarding the analysis's granularity.

Acknowledgement

We gratefully acknowledge financial support from FCT Fundação para a Ciência e a Tecnologia (Portugal), national funding through research grant UIDB/04521/2020.

References

[1] D. Belo, T. Gil Pinheiro, Portugal and the future of housing, Caixa Bank Research, 2019

[2] G. Milunovich, "Forecasting Australian Real House Price Index: A Comparison study of Machine Learning and Time Series Methods", https://www.researchgate.net/publication/334388950: 2019.

[3] E. Donaldson, The effects of COVID-19 on house prices in Portugal in 2021 Idealista.pt, 2021. https://www.idealista.pt/en/news/property-for-sale-in-portugal/2021/02/22/877-the-effects-of-covid-19-on-house-prices-in-portugal-in-2021 (accessed on 20 January 2021)

[4] Redaction, Property prices in Portugal have soared in the last decade

Idealista.pt, 2021.

[5] A.C. Santos, N. Serra, N.Teles, Finance and housing provision in Portugal, Journal of Economic Literature classification codes, ISSN 2052-8035, January,2015.

https://www.idealista.pt/en/news/property-for-sale-in-portugal/2021/01/21/845-property-prices-in-portugal-have-soared-in-the-last-decade (accessed on 20 January 2021)

[6] S. Gudel Information from Past Pandemics, and What We Can Learn: A Literature Review, Zillow Economic Research. 2020. Available online: https://www.zillow.com/research/pandemic-literature-review- 26643/ (accessed on 20 January 2021).

[7] J. M. Quigley, the real estate prices and economic cycles, powered by the California Digital Library, University of California, https://escholarship.org/uc/item/58c6v2kx : 2002

[8] E.Glaeser, W. Huang, Y. Ma, A. Shleifer, "A real estate boom with Chinese Characteristics", The Journal of Economic Perspectives, Vol. 31, No. 1, winter 2017, pp.93-116.

[9] A.Soy Temur, M. Akgun, G. Temur, "Predicting housing sales in Turkey using arima, LSTM and Hybrid models", Journal of Business Economics and Management, Vol.20 Issues 5: 920-938, 2019.

[10] R. J. Shiller, "Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures." Journal of Finance, vol. 48(3), 1993, pp. 911-931

[11] C. Frischtak, B.R. Mandel," Crime, House Prices, and Inequality: The Effect of UPP in Rio". Federal Reserve Bank of New York Staff Reports,no. 542, 2012.

[12] V. D Giudice, P. D. Paola, F.D. Giudice. "COVID-19" infects real estate markets: Short and mid-run effects on housing prices in Campania region(Italy), Soc. Sci. 2020, 9,114.

[13] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, and R. Agha.. The socioeconomic implications of the coronavirus pandemic (COVID-19): A review. International Journal of Surgery 78: 2020 185–93.

[14] P. Angelov, H. Le, E. Tolentino, B. Kim, "Using Machine Learning Algorithms to Analyze Impact of Crime on Property Values", Issues in Information Systems, Volume 21, Issue 1, pp. 55-61, https://doi.org/10.48009/1_iis_2020_55-61: 2020.

[15] P. Ofori, "Waste Disposal Sites and Residential Rental Values Nexus: An Appraisal of Agogo Asante Akyem Dumps", African Journal of Science, Technology, Innovation and Development, https://doi.org/10.1080/20421338.2020.1830542: 2021.

[16] A. Hamid, M. I., and G.Christopher. 2016. Community Loss of Residential Value from Water and Noise Pollution.

[17] C. J. Costa, and J. T. Aparicio. POST-DS: A Methodology to Boost Data Science. In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE, 2020

[18] R. F. Lourenço, P. M. M. Rodrigues, "House prices in Portugal- what happened since the crisis?" Exchange seminar at Banco de Portugal, 2017 .

[19] J. P. G. Custódio, C. J. Costa and J. P. Carvalho, "Success Prediction of Leads – A Machine Learning Approach," *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, Seville, Spain, 2020, pp. 1-6, doi: 10.23919/CISTI49556.2020.9141002.

[20] D. Tomás, C. Costa, J. P. Gaivao and J. P. Carvalho "Time series for incidences, orders and invoicing forecast". CAPSI 2018 Proceedings. 36. 2018 https://aisel.aisnet.org/capsi2018/36

[21] G. E. Tita, T. L. Petras, R. T. Greenbaum, "Crime and Residential Choice: A Neighbourhood Level Analysis of the Impact of Crime on Housing Prices", J Quant Criminol (2006) 22:299–317

[22] D. L, "The effect of victimization on a household's moving decision. Criminology (1991), 37:903-931

[23] C. Rocha, A. Carvalho, "Portuguese Real estate Taxation, Land Use and Noise", 37th International Congress and Exposition of Noise Control Engineering, 2008.