

Coursera Capstone Project



Real estate market in Barcelona



Background

- COVID-19 crisis has been a huge shock for millions of citizens and it has exposed some fundamental weaknesses of our economy
- People are afraid of losing their jobs, their social security and the quality of life.
- Real estate market is a good KPI to measure the real impact of the situation

Main Goal

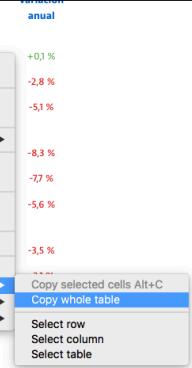
- Finding an accurate prediction about real state market situation for next months

Data acquisition

Histórico de precios de venta en Gràcia

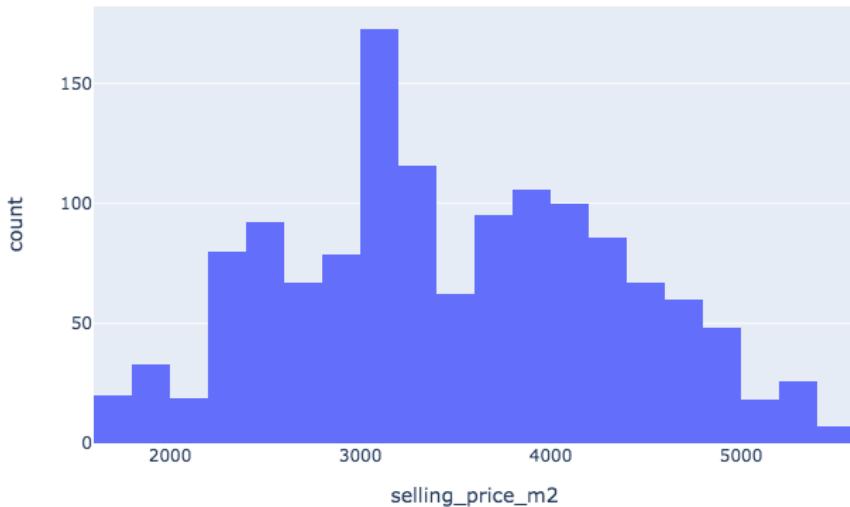
| Mes | Precio m ² | Variación mensual | Variación trimestral | Variación anual |
|----------------|------------------------|-------------------|----------------------|-----------------|
| Mayo 2020 | 4.214 €/m ² | +0,5 % | +0,7 % | +0,1 % |
| Abril 2020 | 4.193 €/m ² | +0,2 % | +0,4 % | -2,8 % |
| Marzo 2020 | 4.185 €/m ² | 0,0 % | -0,7 % | -5,1 % |
| Febrero 2020 | 4.185 €/m ² | +0,2 % | -2,4 % | -8,3 % |
| Enero 2020 | 4.176 €/m ² | -1,0 % | -2,9 % | -7,7 % |
| Diciembre 2019 | 4.217 €/m ² | -1,6 % | -1,5 % | -5,6 % |

| Mes | Precio m ² | Variación mensual | Variación trimestral | Variación anual |
|-----------------|------------------------|-------------------|----------------------|-----------------|
| Noviembre 2019 | 4.286 €/m ² | | | |
| Octubre 2019 | 4.300 €/m ² | | | |
| Septiembre 2019 | 4.283 €/m ² | | | |
| Agosto 2019 | 4.234 €/m ² | | | |
| Julio 2019 | 4.275 €/m ² | | | |
| Junio 2019 | 4.229 €/m ² | | | |
| Mayo 2019 | 4.211 €/m ² | | | |
| Abril 2019 | 4.312 €/m ² | | | |
| Marzo 2019 | 4.409 €/m ² | | | |



- The information was collected from one of the most popular real estate sites on line in Spain (<https://www.idealista.com>)
- I tried to develop some lines of code but web scrapping was not possible. This site is protected from it.
- A Firefox plugin called Table2Clipboard was installed. It allowed table capturing and export easily from HTML code to CSV file.
- I got several CSV files for each district and all the information was collected in one single CSV file

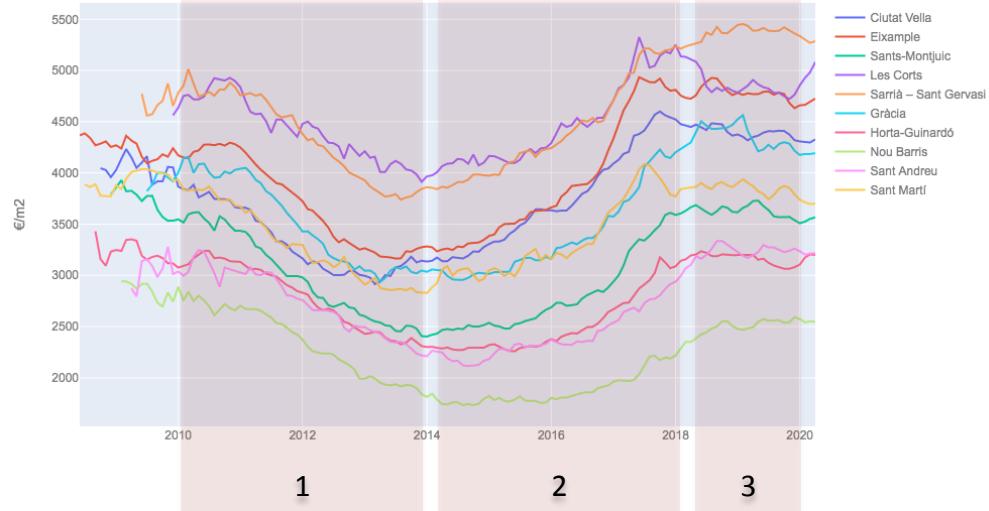
Global analysis



Global pricing distribution

- Just as expected I found a Gauss Curve
- There is more values in the high band between 3500€/m² and 4500€/m² than in the low band between 2500€/m² and 3500€/m².
- Clear trend towards high values (More expensive houses)

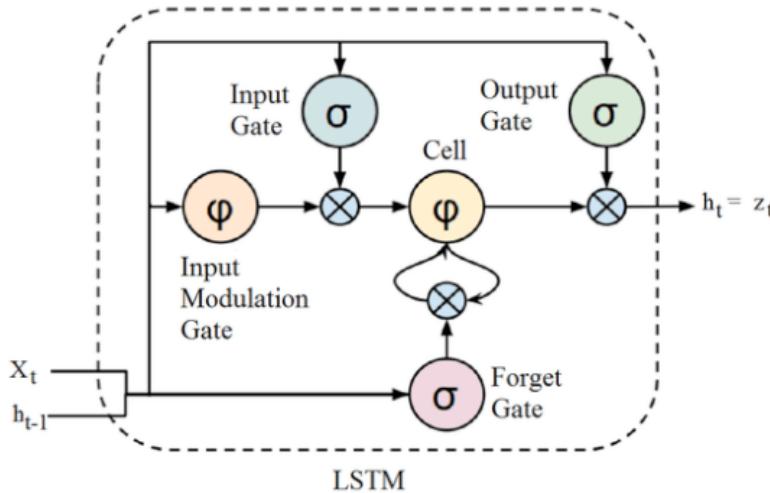
Real estate market pricing in Barcelona (by District)



Pricing distribution time line by district

- Alteration in trend from 2010 -2014 with decreasing values. Real estate bubble consequences (1)
- Increasing pricing trend from 2014-2018, employment rate increasing. Economic regeneration (2)
- Stabilization in highest price from 2018 to 2020 (3)
- COVID-19 crisis detected in 2020 February in Spain

Model definition and deployment



RNN

A **recurrent neural network (RNN)** is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence.

LSTM

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture

Normalizing

I used Scaled values (between 0 and 1) because the algorithm gives better performance than using original values.

Parameters

The best option for choosing the right value for the parameters is basically understanding the behaviour of results based on them.

Units = 200, Epochs = 100, activation method ='relu', loss ='mse'

Little trick!

I discovered the EarlyStopping function that is very useful for Stop Training When Generalization Error Increases. More information here:

[Prevent overtraining with EarlyStopping](#)

Pricing analysis steps

Predicted vs. current values comparison for Ciutat Vella



Future predicted values for Ciutat Vella



Step 1. Market price evolution & test over existent data

- Dataset is divided in two time frames (train and test).
- With training information a prediction is done over know data (testing data)
- We can compare the precision of the model using root mean squared error (rmse).
- We can see line blue (real values) and red line (predicted value) sharing the same time period

Step 2. Predicted future values

- New time frame of 6 periods were created
- The process returned new records with new information based on previous calculations.
- We can see line blue (real values) and red line (predicted future values) as a continuation of the full timeline

Step 3. Root mean squared error & certainty

- The performance indicator were RSME. It is very important to remember that the RMSE value is evaluated in the same units of the values, in this case, €/m²

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Ciutat Vella | train-test | 26.45 | 0.61 | 99.39 |
| Ciutat Vella | future | 56.30 | 1.29 | 98.71 |

Pricing analysis by district (I): Ciutat Vella

Predicted vs. current values comparison for Ciutat Vella



Future predicted values for Ciutat Vella



Analysis

In the predicted values I observed a good overlapping between selling price and prediction. In the future prediction the algorithm indicates a little pricing increase at the start and it keeps flat after that. Good performance near 99% in future predictions and a deviation of 56,30€ of total price.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Ciutat Vella | train-test | 26.45 | 0.61 | 99.39 |
| Ciutat Vella | future | 56.30 | 1.29 | 98.71 |

Pricing analysis by district (II): Eixample

Predicted vs. current values comparison for Eixample



Future predicted values for Eixample



Analysis

In the predicted values I observed a poor overlapping between selling price and prediction. In the future prediction the algorithm indicates that the prices goes down slowly. Good performance in both values.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Eixample | train-test | 61.28 | 1.31 | 98.69 |
| Eixample | future | 81.71 | 1.72 | 98.28 |

Pricing analysis by district (III): Sants-Montjuïc

Predicted vs. current values comparison for Sants-Montjuic



Future predicted values for Sants-Montjuic



Analysis

In the predicted values I observed a reasonable overlapping between selling price and prediction, at least in the firsts time series. In the future prediction the algorithm indicates that the prices goes down fast. Very good performance.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|----------------|-----------------|---------|----------|--------------|
| Sants-Montjuic | train-test | 34.35 | 0.97 | 99.03 |
| Sants-Montjuic | future | 29.75 | 0.84 | 99.16 |

Pricing analysis by district (IV): Les Corts

Predicted vs. current values comparison for Les Corts



Future predicted values for Les Corts



Analysis

In the predicted values I observed a very bad values between selling price and prediction, This behaviour is the same in the future prediction. The algorithm indicates that the prices goes down fast. Very bad performance comparing to other districts. Maybe related with the high pricing/m2 of this district.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Les Corts | train-test | 295.81 | 6.05 | 93.95 |
| Les Corts | future | 176.68 | 3.70 | 96.30 |

Pricing analysis by district (V): Sarrià – Sant Gervasi

Predicted vs. current values comparison for Sarrià – Sant Gervasi



Future predicted values for Sarrià – Sant Gervasi



Analysis

In the predicted values I observed a correct overlapping between selling price and prediction. This behaviour is totally different in the future prediction where the algorithm indicates that the prices keeps the value. Good performance near to 99% in both cases.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|-----------------------|-----------------|---------|----------|--------------|
| Sarrià – Sant Gervasi | train-test | 31.76 | 0.60 | 99.40 |
| Sarrià – Sant Gervasi | future | 77.82 | 1.44 | 98.56 |

Pricing analysis by district (VI): Gràcia

Predicted vs. current values comparison for Gràcia



Future predicted values for Gràcia



Analysis

In the predicted values I observed a good overlapping between selling price and prediction. As well as the future prediction offers a stable value. Very good performance of indicators

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Gràcia | train-test | 37.66 | 0.9 | 99.1 |
| Gràcia | future | 55.01 | 1.3 | 98.7 |

Pricing analysis by district (VII): Horta

Predicted vs. current values comparison for Horta-Guinardó



Future predicted values for Horta-Guinardó



Analysis

In the predicted values I observed a full deviation between selling price and prediction, This behaviour is the same in the future prediction where the algorithm indicates that the prices draws a curve. Upgradable performance.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|----------------|-----------------|---------|----------|--------------|
| Horta-Guinardó | train-test | 107.34 | 3.42 | 96.58 |
| Horta-Guinardó | future | 67.59 | 2.17 | 97.83 |

Pricing analysis by district (VIII): Nou Barris

Predicted vs. current values comparison for Nou Barris



Future predicted values for Nou Barris



Analysis

In the predicted values I observed an important deviation between both values. But in bot graphics I observed the same behaviour. The trend indicates decreasing values. Correct performance of the indicators.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Nou Barris | train-test | 67.44 | 2.64 | 97.36 |
| Nou Barris | future | 48.01 | 1.91 | 98.09 |

Pricing analysis by district (IX): Sant Andreu

Predicted vs. current values comparison for Sant Andreu



Future predicted values for Sant Andreu



Analysis

In the predicted values I observed a good approach in trend between selling price and prediction. As well as the future prediction offers a stable decreasing value. Very good performance of indicators.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Sant Andreu | train-test | 43.21 | 1.34 | 98.66 |
| Sant Andreu | future | 13.37 | 0.41 | 99.59 |

Pricing analysis by district (and X): Sant Martí

Predicted vs. current values comparison for Sant Martí



Future predicted values for Sant Martí



Analysis

I appreciated that both values (predicted & current) have different trajectories. With the interactive graphic I observed that the major distance is over 80€/m². Regarding to futures values the graphic show a clear descendent behaviour. Good performance of the indicators.

| District_name | Prediction_type | RMSE(€) | Error(%) | Certainty(%) |
|---------------|-----------------|---------|----------|--------------|
| Sant Martí | train-test | 59.88 | 1.60 | 98.40 |
| Sant Martí | future | 58.19 | 1.54 | 98.46 |

Conclusion



- The main goal were to predict the future values for the real state market in Barcelona. **The requirements have been covered.**
- Time series prediction is one of the most typical business case. The role of CNN in this type of projects is capital. If we can enrich the inbound information we can get better combinations and awesome outcomes.
- Specifically with this projects I detected a global trend for the majority of districts that indicates a stabilization of prices for the next months and in some cases going down prices. Will see in 6 months.
- Important improvements:
 - Designing an APP or web site for real state agents could be very useful
 - The correlation with other indicators as unemployment rate, purchasing power, supply and demand may be very useful specially because all the prediction are based in the same trend value (pricing/m²) has a poor reliability.