1ASBS

Hello guys my product name is A S B S which stands for Analyze smart buy smart. You will get why at the end of the presentation

2 Content

This is a quick look at the content that we will cover

3 Inro

I'm thinking to buy with my wife an apartment in a couple of years, and the question is where and based on what to choose a neighborhood? Could I do an analysis to choose it?

4Objective

The objective is to find a statistical methodology to help you choose a location base on data.

5 Problem

I choose Madrid city to do the neighborhood search and I identified a couple of problems, there are 131 different neighborhoods, there are a lot of apartments in sale with daily updates, I have specific requirements for the apartment, and I am on a tight budget.

6 Solution

As a solution, I created a rating system to define two neighborhoods by sampling 2 potential locations, in this case, Puerta del Angel and Comillas.

7 Sample

Regarding the sample: my source was Idealista a very well-known real estate App in Spain. Four hundred and fifty euros was the cost for the sample. The cost was calculated based on a junior data analyst's salary in Dublin for the hours needed. I sample apartments with specific characteristics like maximum budget, and other features like no ground floor, and minimum 2 rooms. I got 16 features per apartment and I assigned 3 quality qualifications based on the pictures for the kitchen, bath, and an overall score for the apartment.

8 Preliminary

I had preliminary questions like Which neighborhood is cheaper, the best price slash quality, which location has more spacious flats, and how to choose between them?

9 Now let's take a quick look at the data. (15 sec)

Tableau 1 General Overview
Here is a quick overview so we can understand the data better.
We have the overall rating ratio assigned, the distribution of the apartments by the number of rooms, and the average price per floor. We can zoom in on each location (click on each one).

Tableau 2 Analysis overview
Over here we have a small sample of the analysis that was performed in the “E” “D” “A” with Python. Where we analyzed the correlation of price versus square meters, also the price distribution in order to understand how the price fluctuates between the 2 locations. One of the features that had to be discarded from the analysis was the orientation because it had too many options and a lack of data. And we could analyze them one by one.
So now the question is how this could help me to choose.

10 Hypotesys productive model 03
And the answer is with a rating system based on hypothesis testing

11 Hyp 1
Here we have the first preliminary question about the cheapest location. I wanted to prove if Comillas was more expensive than Puerta del Angel.
 Based on the hypothesis result, I can not reject the null hypothesis with the data that I have, so I assigned a tie to the rating system.

12 Hypo 2
I found out that Comillas apartments trend to be more spacious, so now we have a score 2-1, comillas is leading.

13 Hypo 3
For the last one, and with a mindset ‘’Best quality with the best price’’, Puerta del Angel got the last point, overall we got a tie for those locations. I tried to find a tiebreaker but I couldn't, it was always a tie. And this is because of how I sample the apartments, with the previous restrictions that I mention, my samples were very similar in those specific neighborhoods.

14 Predictive model

We can also have a price reference

15 Multilinear reggression model

I mention a price reference because with this sample I didn't get a good score for the model. But we got some prices so they can be used as reference.

16 Conclusions

To grab up, This product could be used for 2 other locations, or even for more than two, then we will get different outcomes. So guys if you have four hundred and fifty euros I can sample 2 more locations for you.
Also, we can define different approaches, for example, we could analyze and score the location in a more general way not so restrictive.

17 Q&A

So do you have any questions?