IBM DATA SCIENCE

PROFESSIONAL CERTIFICATE

APPLIED DATA SCIENCE CAPSTONE

Capstone Project - The Battle of Neighborhoods

Autor :

Pairs :

Janvion LOUKOU

...

12 juin 2020

Table des matières

1   Introduction

Since the 2000s, the price of housing in the French market, in terms of income and rents, is undeniably very high.

To this end, we can emphasize that from 2000 to 2010, on average and across France, the housing price index increased by 107% while the rent index and income per household did not only increased by 27% and 25% respectively.

In 2011, and this for almost 70 years, the price of housing in France was at its peak.

In 2015, 68% of French people lived in a house. Definitely higher than the European average of 58%.

During the years 2016 and 2017, the prices undoubtedly increase. We say more than 0.8% in 2016, and an increase of almost 1.7% in 2017. Not exempt from this increase, 2018 is the year of continuity !

The sales volume remains stable at its highest level. There were 960,000 transactions in 2017, compared to 948,000 in September 2018.

In conclusion, this is a very fluid market which represents more than 3 sales per year for 100 households.

Note that in the first half of 2018, 35% of the 50 largest cities in France saw their average prices decrease.

The Métropole Européenne de Lille (MEL) is not exempt from this increase. Real estate investors are therefore faced with the problem of the profitability of investments in the center of large cities. Thus, the average price per m2 in Lille center is 3,201 euros. Under these conditions, investments are no longer profitable and investors generate zero or even negative cashflows. To overcome this phenomenon, one idea is to make investments in neighboring cities while remaining close to all amenities. This study therefore aims to explore the neighborhoods of Lille to identify small towns in which the cost of real estate transactions would be lower and therefore generate more profit.

2   Data available

Data are from the France government site : Demandes de valeurs foncières.
We have two databases :

The first database concerns real estate transactions in France in 2019. This database contains 40 variables and 2,535,791 rows. Among the variables, we find : the nature of the transaction, the value of the property, the name of the town in which the transaction takes place, the surface area of the property, the longitude and latitude of the property, etc... A first choice of variables will therefore be performed for Exploratory Data Analysis (EDA).

The second database from the Wikipedia site concerning information on the municipalities of la Métropole Européenne de Lille (MEL) will be used to extract real estate transactions from Lille. The Jupyter notebook presents all the stages of this data selection.

3   Data cleaning