

Real Estate Market Data Exploration in Toulouse and its suburbs

The objective is to perform real estate data exploration near the city of Toulouse. Your company (Immo-Inv) is a real estate agency who wants to understand very well the real estate market in Toulouse and its suburbs. You are the big data analyst of the company, and you have access to a 5-years real data history of houses and apartments sales transactions in Toulouse and its suburbs. The dataset contains details for each transaction: sale date, localization (city, postal code), type of residence, type of sale, land area, living area, number of rooms, price, etc.).

You should use Spark for this analysis because you should be able to apply your analysis to a dataset with the entire real estate market in France for instance (big data file). But for presenting your key results, you can of course use python visualization libraries such as seaborn or other libraries.

Challenges here are to explore all possible aspects of this real estate market (variables, relationships between variables, trends, patterns, outliers, etc.). But at the end you should focus on at least 5 keys findings (tables or visualizations) in your final notebook. You can also comment these findings. You can explore for instance:

- The evolution of the prices (e.g. price per square meter)
- Identification of sales outliers
- Differences of prices per cities or postal codes
- Differences between apartment and houses
- Differences between types of sales
- Differences between Toulouse and its suburbs
- High cost, low cost, or emerging cities