

# Tableau Project

---

## A. Cleaning step: Tableau prep builder

- We imported independently each data file (From 2015 to 2019)
- We changed the type for price from string to number decimal
- Then we did a union to have all the data in the same file:
  - o Some variables had different names in the different data file so we merged them under the same name.

After this done, we performed a general cleaning:

- We kept only sale and sale before completion transactions
- We kept only apartment and house transactions
- We removed transactions without price, postal code, rooms or living area
- We replaced null values for land area by a 0 (considering that there is no land area for example for an apartment)
- We merged city and commune because it was the same
- And we removed the “arrondissements” of the French cities because the postal code is sufficient
- We fixed some data type (city for city and postal code to ZIP Code/Postcode) in order to allow Tableau Desktop to work with them.

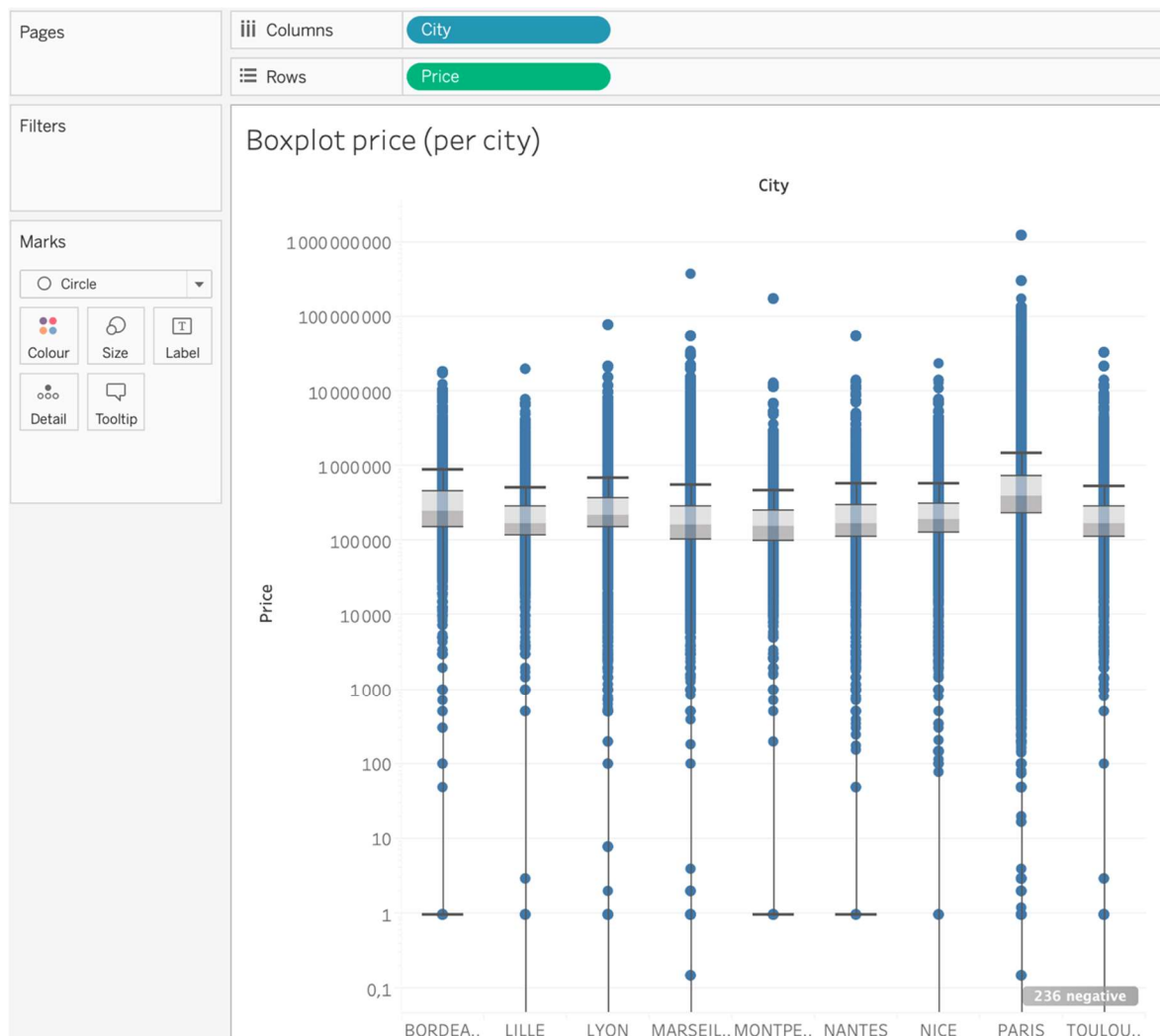
## B. Removing outliers

At this stage we have cleaned up a significant part of the data but there are still some outliers that we need to eliminate like when the price of houses that is lower than 1000 euros and apartments with a few square meters of surface area that cost thousands of euros. We therefore decided to study these atypical values through boxplot using Tableau Desktop.

At this point we decided to work on common real estate.

The idea of the project is to understand the French real estate market in order to predict fluctuations and allow for potential investment. Working on common real estate is fairer than taking into account typical values which will not represent the bottom of this market.

So after getting these boxplots we first reduced the study strictly to the values that they contained.



For each city:

#### Bordeaux

- Upper Whisker: 894 105 €
- Lower Whisker: 1 €

#### Lille

- Upper Whisker: 532 000 €
- Lower Whisker: 0 €

#### Lyon

- Upper Whisker: 711 250 €
- Lower Whisker: 0 €

#### Marseille

- Upper Whisker: 559 000 €
- Lower Whisker: 0 €

#### Montpellier

- Upper Whisker: 476 000 €
- Lower Whisker: 1 €

#### Nantes

- Upper Whisker: 590 400 €
- Lower Whisker: 1 €

Nice

- Upper Whisker: 592 200 €
- Lower Whisker: 0 €

Paris

- Upper Whisker: 1 497 200 €
- Lower Whisker: 0 €

Toulouse

- Upper Whisker: 544 930 €
- Lower Whisker: 0 €

But some variables were still weird to be considered as a common real estate, so we arbitrarily removed them.

Price under 20 000€

Living area under 5m<sup>2</sup>

More than 10 rooms.

Then we started making visualization with Tableau Desktop.

### C. Visualization step: Tableau Desktop

- How does your visualizations leverage at least one “pop-out effect” or “pre-attentive attribute?” Which one(s) was (were) chosen and why?

On the maps we used colors as a pre-attentive effect, to show in green when the prices were low and in red when they were high.

We also worked with size, for example for the pies or more specially for the living area. The bigger the circle, the bigger the average living area.

- How does your design reflect an understanding of cognitive load and clutter?

We tried to keep our design as simple as possible trying to minimize the cognitive load and clutter. We never used 3D graphics because it was not necessary: when we wanted to represent 3 different variables in a single graph, we used colors, which were easier to understand immediately. Moreover, when we used the color, a legend is associated with the graph to be able to understand it

We also grouped the sheets with the same topic in the same dashboard so that at one point in the story, our interlocutor can fully understand the topic without having to remember previous slides.

- Which metric(s) you have used? and why?

We used the price per meter squared because it's an important parameter in the real estate market. This field provides important information about how profitable the market is in a

certain area or in a certain year: the higher it is, the more houses are sold at high prices for the same surface area.

- What is (are) the most important(s) dashboard design principle(s) you have used?

We have created dashboards to be aesthetic and interactive. We have also avoided scrolling for ease of use. The use of filter allows to have differentiated analyzes on the same subject. Finally, we have chosen to keep only the essentials and to make our dashboards as uncluttered as possible.

- What is (are) the most important(s) data stories design principle(s) you have used?

We really wanted to create a story that could be understood by anyone. For this we introduced the context and then we first showed the facts by studying the situation on the French real estate market. Then we focused our studies on the conclusions that could be drawn for investment or decision-making in the field of real estate, what in our conclusion was put forward by proposals.

- Why do you think your data story meets the audience's needs?

First of all, we believe that the story was designed for real estate professionals. Then that it was built in order to understand initially the situation and the evolution of the market for 5 years. In a second step, to highlight the essential elements for decision-making such as seasonality and the fall in prices. The audience after that would have all the keys in hand to understand and take actions in consequence. Also, the data history is interactive and, thanks to the pre-attentive attributes, it would keep the audience captive. The presentation is uncluttered, and the principles of gestalt make it easy to understand general information without looking in detail.