Inès KOUYATE, Inès AGOSSOU, Nandy MAVUA

**The exploration of our data about the real estate market in France
(from 2015 to 2019)**

Question 1: Explain the main choices you made when preparing your data.
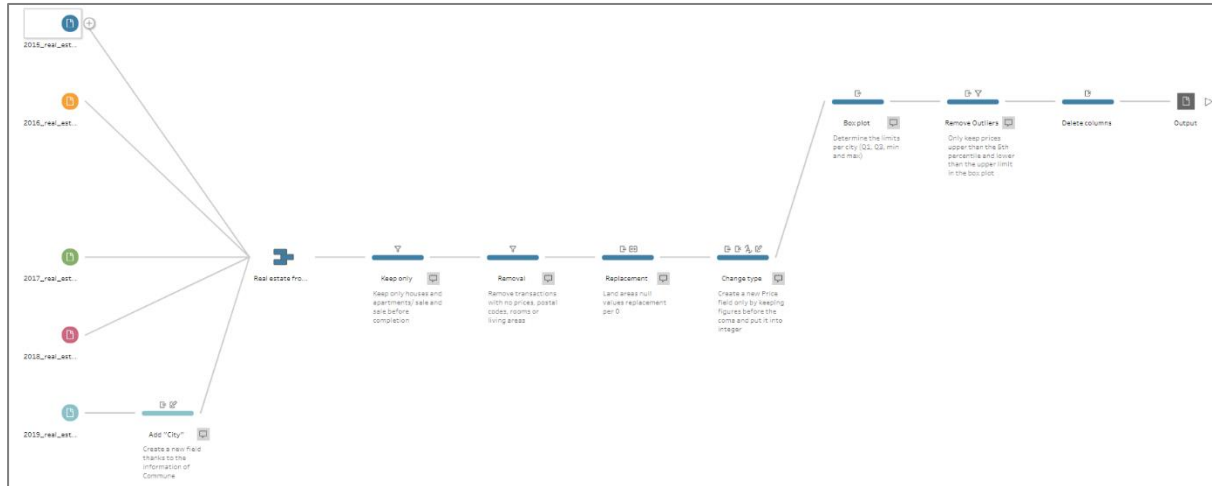

Figure 1

Our first step was to prepare and clean the data with Tableau Prep Builder. We had to merge first the data since it was in different datasets (one dataset per year), so that we could explore them all together. However, we found that in the 2019 database, the "City" column did not exist unlike in previous years' databases. This created an imbalance between them.

We have therefore created a calculated field to remedy this problem using the following formula: TRIM(SPLIT([Commune], " ", 1)), to split the "Commune" column into two parts and only keep the name of the cities in it (Figure 1).
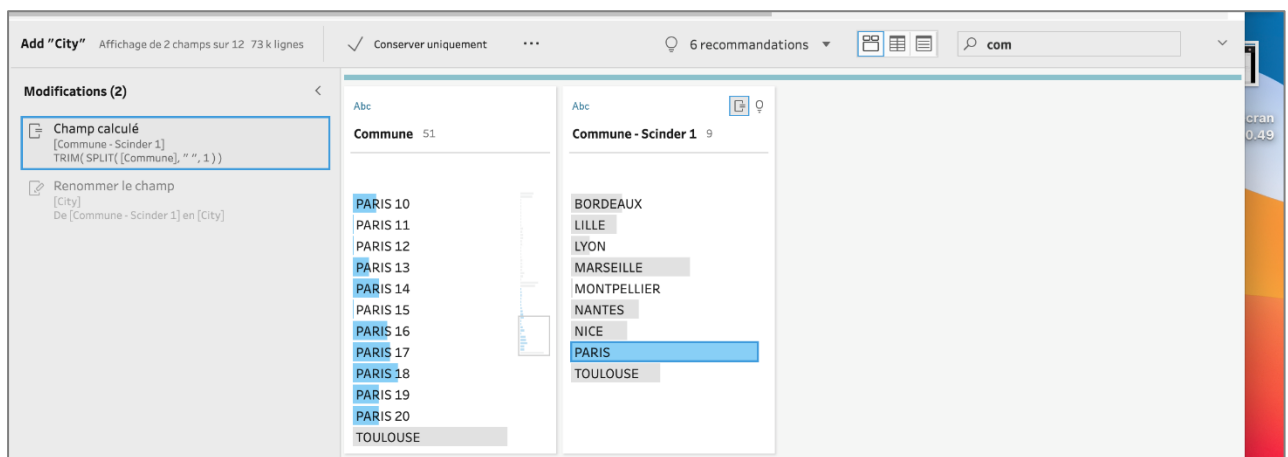

Figure 2

We could also do it manually by duplicating the column "Commune", grouping the districts by city (e.g. Paris 01, Paris 02, …, Paris 20) and rename this duplicated column by "City".
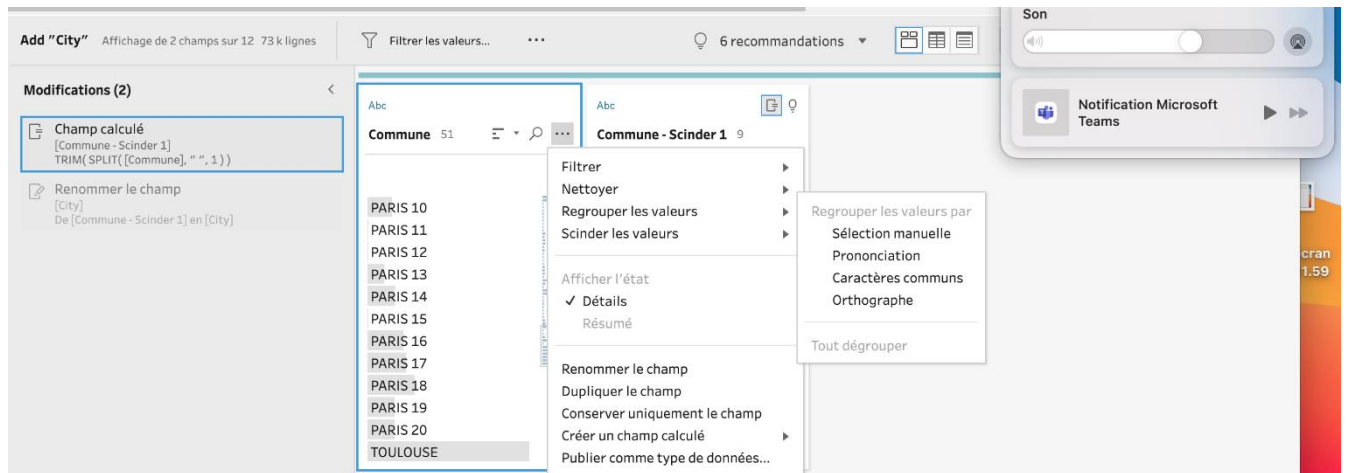
Figure 3

Once we have done that, we decided to make the cleaning and adjustments required, one by one (Figure 1). We also added descriptions to each adjustment to make it more understandable.

Finally, concerning the outliers, we have first created calculated fields allowing to delimit, by city, a boxplot (min, max, Q1, Q3) using simple calculations of percentiles. This allowed us to remove the upper outliers from each city (i.e., all lines whose price was higher than the maximum bar of the respective boxplots) but not the lower outliers. To overcome this, we decided then to remove, by city once again, the lowest 5% of prices remaining, in order to keep 95% of their relevant values for the study.

Question 2: How does your visualizations leverage at least one "pop-out effect" or "pre-attentive attribute?" Which one(s) was (were) chosen and why?

To enable the personas to easily perceive and understand our visualizations without them having to think about it, the best way was to directly address their iconic and short-term memory, that is why we used these leverages:

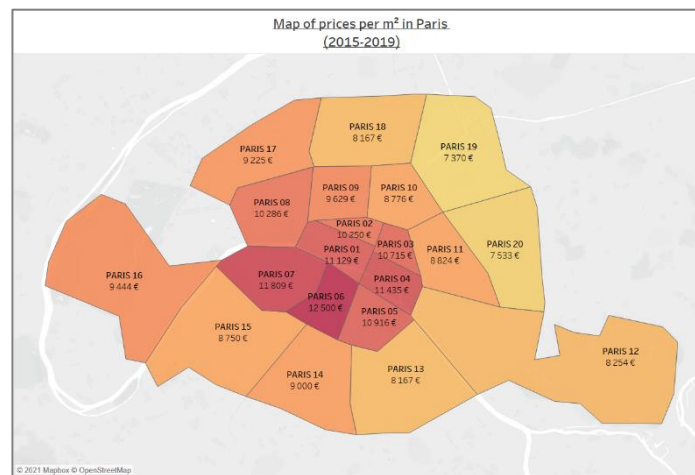|	
<u>Spatial position and Size/Area:</u> These were particularly used in the dashboard focused on giving a global overview of the French Real Estate Market. Both a map and a bar chart have been relevant, allowing to show the relationship between total sales volume, prices, and all the other features for each city. The larger the bars, the higher the metric (bar chart), and the bigger the circle, the higher the sales (map). This makes it easier to have a quick overview of the dataset.	 *Figure 4*
<u>Hue/Color:</u> Different colors were used to classify the cities and to make clusters. We also used shades of colors as well as color brightness to make the information clear at a glance (e.g. we created a colored map of Paris with the darker meaning the more expensive. Finally, to highlight the top 3 selling cities, color saturation has been particularly adequate.	 *Figure 5*

Pop-ups:
We made sure that each value was well described in textual and sober pop-up windows depending on both the relevance and the space available in the dashboard. This provides important information and precision which do not appear directly on the dashboard, and which illustrates our points.

13m²	55m²	1	2
270m²	103m²	0	4
38m²	56m²	1	3
250m²			
18m²			
221m²			
20m²			
305m²			

Cluster: "High sales with Low prices (<3500€/m²)"

There are 13 districts in this cluster,
representing a total sales volume of 89 287 (APARTMENTS).

APARTMENTS to invest in:
Nb Rooms moy.:   3
Nb Lots moy.:   1
Living Area moy.:   56m²
Land Area moy.:   38m²
Mediane Price m²: 3 068 €/m²
Median Sale Price: 159 500 €

Figure 6

Question 3: How does your design reflect an understanding of cognitive load and clutter?

Cognitive load

Simplicity, ease, and relevance; these are three words that helped us in the construction of all our visualizations. When a design requires a huge effort to be understood, it means that it is not optimal, and therefore, can be improved. That is why we tried our best to avoid that and make sure the information was always easy to interpret, especially since it is particularly important for Aude P. who wants quick and clear credible updates. To do this, we intended to:

- Adapt the presentation to the expertise of the personas: Gerard, marketing director, who is able to identify by himself some findings, and Aude, general manager, who is data and finance oriented.

- Reduce the "Problem Space": through the dashboards and the story, we broke the problem down into parts by giving first a quick overview of the real estate in France, then showing the evolution of sales and prices in each city, and finally suggesting practices to adopt be it in Paris or not, to let notably Gerard T. anticipate next strategies (Figure 7).

- Reduce "Split-Attention Effect": the structure of the story allows time to the personas to become familiar with the data first, before introducing the technical parts. We also strategically used relevant labels into the visualizations, and designed differently headings, titles, content, to make sure they always get the right information at the right time; no more nor less (Figure 7).

| The French real estate market analysis (2015-2019) |
| Context | 1. Overview of the market | 2. Key figures and forecasts | 3. Strategy 1 | 4. Strategy 2 | 5. Focus on Paris | Conclusion |

Figure 7

Clutter

To avoid clutter, we tried to demonstrate organization in each sheet, dashboard and story, to prevent distraction; we put for instance all the filters on the side and gave spaces for the main visualizations. We also tried to keep the colors sober and meaningful by using a limited color palette (blue for graphs, maps mostly in orange-red, etc), and we made apparent sorting of the data for direct comparison when necessary, and favored 2D to 3D (Figure 8).
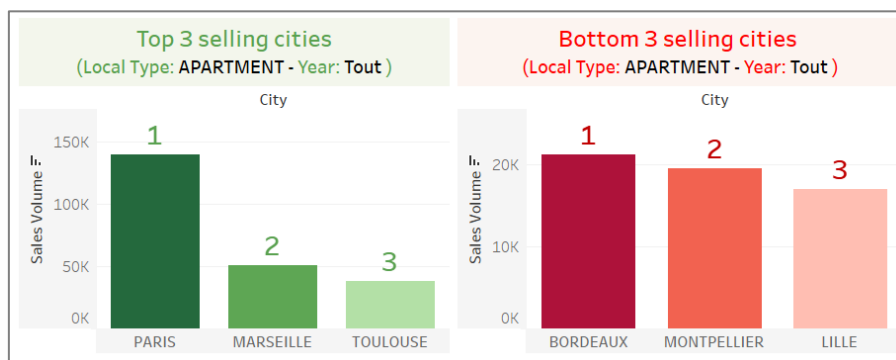

Figure 8

Question 4: Which metric(s) have you used? and why?

Prices: We tried to exploit all the metrics at our disposal. However, we mainly used "Prices" and "Prices per m²" to be able to analyze the most expensive cities or neighborhoods according to our visualizations. The formula of Prices per m² is: [Price]/[Living Area].

Total Sales Volume: We also used the "Sales Volume" to be able to determine the number of houses or the number of apartments in different cities with the following function COUNT ([Sale Type]). But also, to be able to filter certain visualizations, if we ever wanted to display only houses or only apartments.

Forecasts: We have also added the "Forecast" tool to be able to make forecasts for the coming years. This tool is very useful and indispensable as it helps managers in their decision making

Question 5: What is (are) the most important(s) dashboard design principle(s) you have used?

Clear and intuitive visuals
To create our dashboards, our strategy was to design and assemble very precise visualizations, very clear and understandable by all. We put ourselves in the shoes of someone who has never used Tableau Desktop software and who knows nothing about data. Our dashboards had to be understandable immediately.

No Scrolling
In addition, we also wanted to avoid having to scroll through each display and have everything readable.

Interactivity
We also played the interactivity card. Indeed, we have put filters to allow users to select the data that interests them (year, type of housing or price range for example). It's necessary that our dashboards can be used by any member of the team, in the same way. Also, if you click on the name of a city in some dashboards, titles and values change automatically.
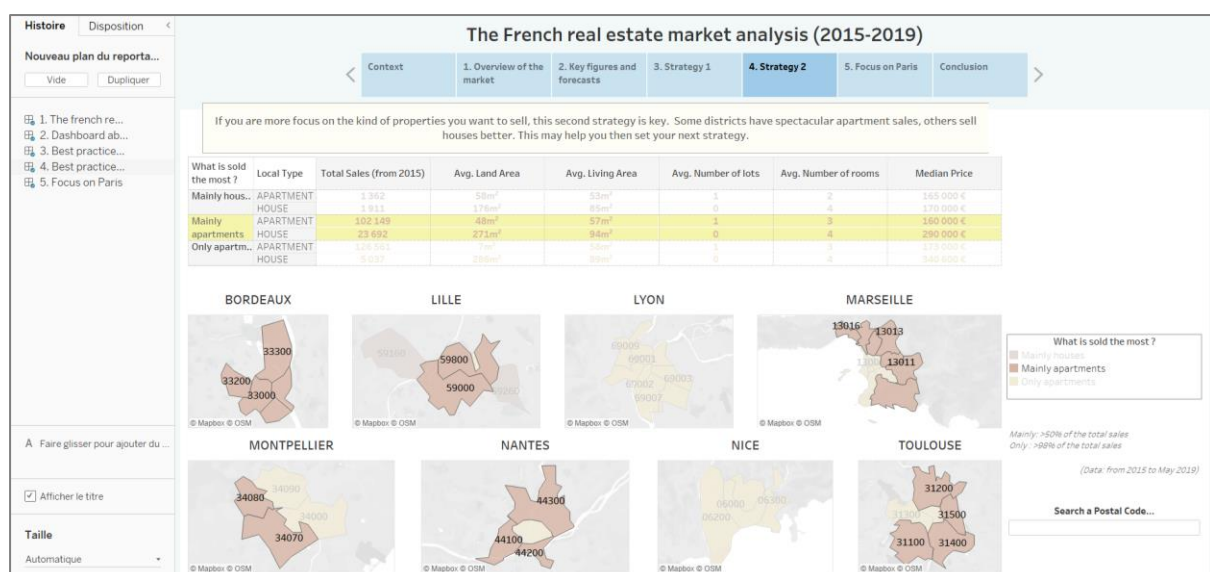

Figure 9

Few details
We have avoided flooding our visualizations with details. We feel that if the visualizations are relevant, few details are enough to understand what we wanted to express

Question 6: What is (are) the most important(s) data stories design principle(s) you have used?

Our objective while realizing our data stories was to be the foundation of a great decision-making process. In that way, we have considered our personas obviously, but we also tried to highlight determinant points useful for our business, while keeping our stories simple and understandable.

- Distinct types of data analysis presentation

We used distinct types of data analysis presentation such as maps, graphs, or charts, according to what we wanted to show.
- Enlighting tools

We used tools that we think were the most enlightening for people that have not been working on the project.

- Specific color code

We used a specific color code so that it will not be confusing for the audience, but more essentially meaningful (for instance, using red for the more expensive value).

- Removal of uninsighful extras

We were also tempted to add decorations, but we resigned because we thought it would be distractive for our audience. The major data (the data we think were the most insightful) were highlighted in. In addition to all this, we tried to make this data story interactive, so that the personas will really be actors and be able to interact when viewing it.

Question 7: Why do you think your data story meets the audience's needs?

First, we tried to make the data as visible as possible because we know that both our personas need to understand our analysis, while not necessarily going through all the data.

Gerard has more time and will be more likely to go through the filters to see the changes in real time, as well as the forecasts. That way, he could also find some aspects that he thinks are useful for him. Aude has less time and wants to see the essential points that we highlighted. She would be more interested in seeing numbers, that is why we combined graphs and charts with comments. Our dashboards need no scrolling, it will easier for her to see the analysis on her tablet.

Conclusion
In conclusion, we have created dashboards that can easily be used by our personas. We were able to analyse the 9 largest cities in France in terms of real estate. This data was used to establish a real strategy, starting with an analysis of the French property market. We then analysed the market in the main cities except for Paris and then ended with the latter.

Paris is the most expensive city in France but also the one with the highest sales, while Marseille is the least expensive city but comes just behind Paris in terms of sales. We have also identified the areas

where there are mainly houses or mainly flats so that we can position ourselves intelligently in the market.

Finally, we chose to focus on Paris because it is a very different city from the others. The districts in the center of Paris are very expensive and much less so on the outskirts. In addition, we have observed a seasonality in sales. Indeed, there is a peak in sales in July and then a very sharp drop in August. We therefore advise against trying to sell during the month of August.