Digital Driven Business

Data Presentation and Visualization

Visualization Plan 2022

Group 2

(Joanna Antończak - 500884145,

Simone van Hooft - 500738996,

Henrike Dietrich - 500843432

Eriona Bytyci - 500862083)

1. Problematization

As a platform, Funda connects the brokers with potential house buyers. Even if the Netherlands is in a housing crisis with a high demand of buyers, private marketplaces are gaining popularity.
Funda needs to provide additional benefits in the big cities as in the countryside.
Every municipality of the Netherlands has its own characteristics and to give one advice that fits all of them, is not possible. Therefore, it needs to identify which factors of the advertisement have an influence on the price or the selling time of a house, to support the brokers as users of Funda but also to improve their performances.

2. Research Questions

Based on the story telling provided to the problematization mentioned above, certain research questions have been raised which focus on the main purpose of the Funda Case such as:

Main Research Question: Which features of the house influence its sale price and time to sell?
Research Question: How should brokers in the Netherlands assist their clients during times of housing crisis, considering regional differences?

Based on the Funda Dataset:
Sub-question 1: Does the selling price have an impact on the selling time?
Sub-question 2: Do the plot size and the living surface have an impact on the selling price and selling time?

Sub-question 3: Does the description count have an impact on the selling price and selling time?

Based on CBS Datasets:

Sub question 4: Does the Age distribution differ amongst different municipalities?
Sub question 5: Does the population density differ amongst different municipalities?
Sub-question 6: Which distances to essentials facilities apply in different municipalities?

3. Hypothesis

Funda Dataset:
H0: Description count has a significant impact on the selling time.
Ha: Description count has no impact on the selling time.

H0: Living area/Plotsize has a significant impact on the selling price.
Ha: Living area/Plotsize has no impact on the selling price.

CBS Datasets:
H0: Every municipality has population characteristics which differ and create different needs for clients.
Ha: Every municipality has population characteristics which don’t differ or create different needs for clients.

H0: The importance of the distances to different essential facilities varies between different age groups.
Ha: The importance of the distances to different essential facilities does not vary between different age groups.

4. Description of datasets

In this project two main datasets have been used, Funda dataset and CBS dataset. The Funda dataset was scraped from the Funda.nl Website and contains all relevant information about the published advertisements. The main features that were used from this dataset were:

- Postcode
- Addresses
- Selling Days
- Selling Price
- Living Area
- Plot Size
- Description Count

The CBS Dataset was downloaded from the Website cbs.nl with focus on population characteristics and distances as an additional asset to the project. During the previous models, both Datasets were merged in python. After being able to finalize the features that have an impact on the selling price and selling time based on the Funda dataset, additional data was put to use in order to provide more insights and better recommendations to the Funda brokers in improving their performance towards increasing their selling price and decreasing their selling time. The main features provided from the CBS dataset were:

- Municipalities
- Age Groups
- Distances to essential facilities (train station, highschool, primary school, doctor, daycare)
- Population Density
- Male and Female distribution

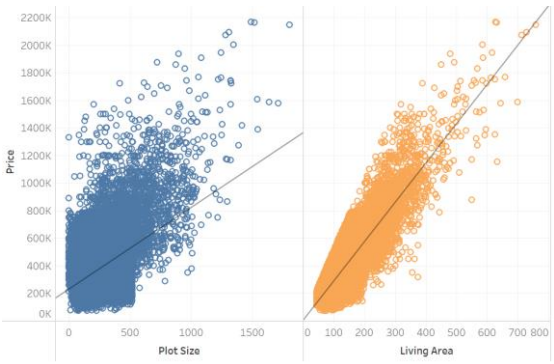
After creating the storyline, extra data from CBS were added to complete and visualize the problem statement and its narrative.

- New Houses built per Year
- Population development per Year
- Overbidding per municipality

Figure 1: Joined Datasets used in Tableau.

5. Choices for Visualization Models and Argumentation

a. Correlation graphs - Scatterplots



Correlation graphs were used to analyze the Funda dataset features on the selling price and the selling days. The choice of visualization model was made to check the impact that plot size, living area and description count have on the selling time and selling price of a house and its relationships. The model was visualized using scatter plots and a modeling line that helps in identifying similar or opposing trends. Furthermore, the outliers were cleaned from the final version of the graphs.

Figure 2: Correlation between selling price and the picked Funda variables.

In the sketching phase, it was experimented with different versions of a line graph with dots. This sketch was based on municipality level since this is the geographical aspect of the story. However, the scatterplot per advertisement was more powerful and highlighted the correlation stronger. Therefore, we introduced the municipality level after the scatterplots.

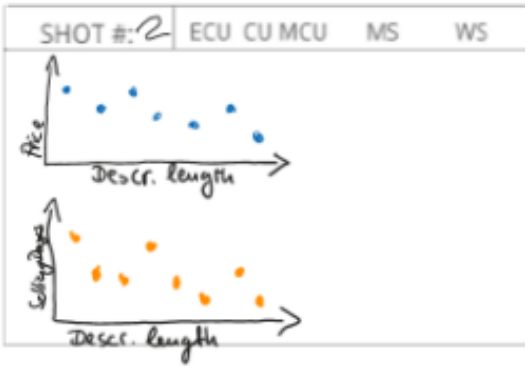


Figure 3: Initial sketches of the graphs.

b. Line graphs

The Line graph was applied to show the population development and the new building of houses in the Netherlands over the Years. With both lines in one graph, a contrary trend was outlined. Which supported the point of the Housing crises of the Netherlands.

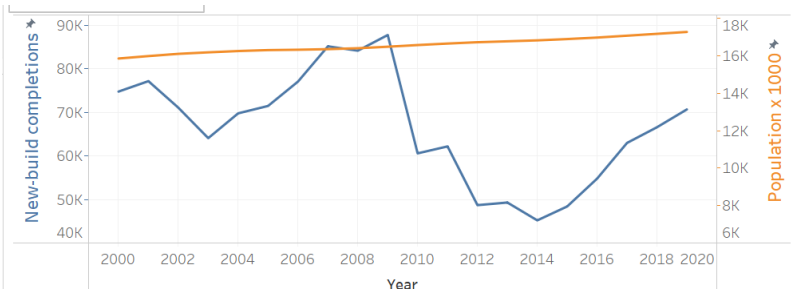


Figure 4: Housing Crisis graph.

c. Horizontal Bar charts

The horizontal Bar Chart was chosen to visualize the distances to essential facilities. The Bar embodies the Distance and makes it easy to understand the differences. Since the Distances have long names, a horizontal graph makes the graphs names comfortable to read.

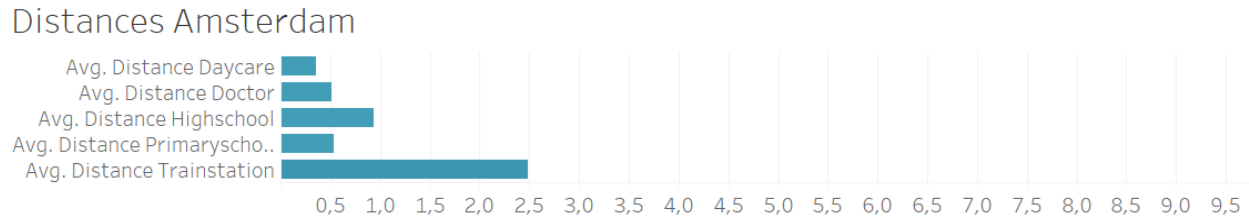


Figure 5: Distances to essential facilities in Amsterdam.

d. Treemap

To display the distribution of age groups in different municipalities, the bar chart was the first sketch.

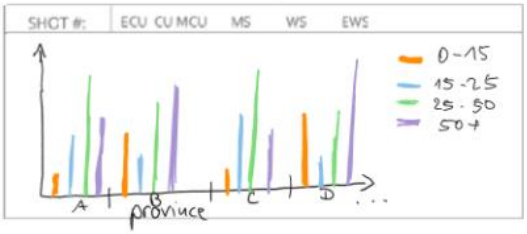


Figure 6: Initial sketch of the graph.

Two problems appear during the work with tableau. Firstly, it is not possible to display five age groups for 12 provinces at once. Secondly, the bars didn't add up to one unity.

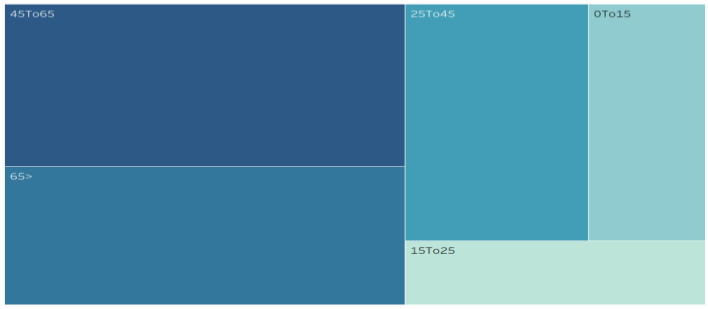


Figure 7: Distribution of age groups.

It was decided to pick 3 representative municipalities and create 3 similar sheets to characterize them. To illustrate the age, the Treemap was the chosen graph. It visualizes the distribution of the age groups in one municipality and gives a clear idea what the dominant group is. Several tree maps after each other illustrate the changes in the age groups per municipality and set the focus to the relevant bar chart about the distances to essential facilities on the same dashboard.

e. The Map of the Netherlands

Symbol maps were used to present the map of the Netherlands on the municipality level, in order to highlight variables which are interesting for brokers.
During the sketching phase, the idea was created to display various variables at once in one map. In addition, a detailed box with detailed information will appear.

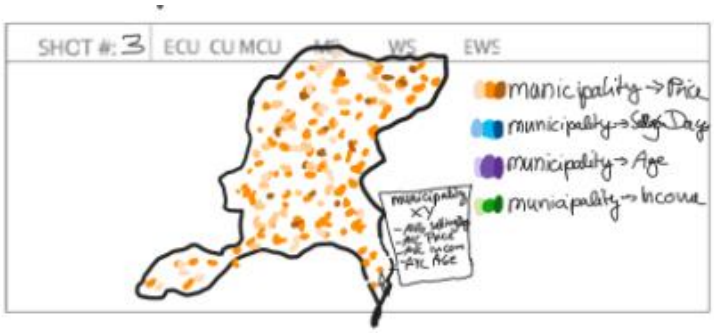


Figure 8: Initial sketch of the map.

In the final version of the map, a heat map was added to define the color of the different dots. In addition, only one color was chosen to not overload the map. The size of the bullet gives an indication about the total number of inhabitants in gemeente.

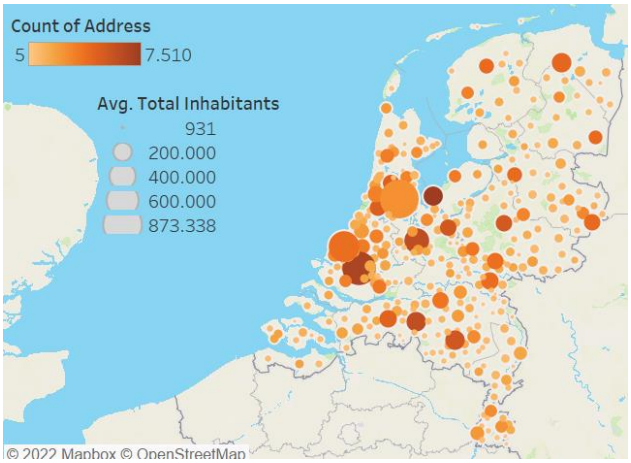


Figure 9: Map created in Tableau.

f. Highlight tables

Highlight tables were used multiple times in this dashboard. The reason why it was used so often is that it helps in reducing the time of the analysis, the highest and lowest values are visible right away. Moreover, it improves the accuracy of the analysis and gives insights that in normal tables cannot be discovered so easily.

Average Selling Days

Municipalityna..	Average Selling Days
Sint Anthonis	30,714
Vaals	28,308
Tubbergen	26,862
Maassluis	26,425
Schiedam	26,347
Schouwen-Duiveland	25,994
Oostzaan	25,959
Oude IJsselstreek	25,569
Brummen	25,563
Tholen	25,460

Figure 10: Highlighted table on an average selling days.

6. Argumentation for storytelling / structure of the narrative arc

The Netherlands is one of the main countries in Europe which is constantly facing a housing crisis due to unmatching figures between the low number of new and affordable houses present in the market and the high number of citizens (especially internationals) looking for a house. Studies have provided stats which prove that a factor that has affected this situation is the increased average price by 10% which has also resulted in a decrease of houses sold by 12%. However, the numbers also highlight the short selling time of a house due the current situation in the market. *(Signorazzi, Aurora)*

Another factor that seems to reflect its impact on this situation is also the limited availability of the smaller houses. If every client would be offered a house that is compatible with the family size of the client in terms of living area surface, that would leave room for improvement to offer bigger houses to

bigger families. *("Older People Staying in Large Homes Is a Key Factor to Housing Crisis: Municipalities")*

This case closely looks into several features that may or may not have an impact on the selling time and the asking price of a house. However, it also highlights possible factors that can affect the selling time and the asking price by visualizing the datasets based on different variables provided by CBS database and Funda database such as municipalities, age, distance to essential facilities, living area surface, plot size, description count. As the most significant results provided from the Funda data analysis are visualized in the dashboards, group 2 aims in providing a deeper understanding to the figures visualized by highlighting what CBS database is able to provide as well. The purpose of using the CBS dataset is to provide more client specific recommendations to the brokers.

Based on the dashboards presented, decent recommendations are provided to the brokers in terms of leading their clients to the right properties which at same time can fulfill their needs taking into consideration their age, preferences to certain essential facilities *(Panel, Forbes Biz Council Expert)*. This could also lead to a decrease in the selling time and an increase in the number of houses being sold as well. Moreover, selling suitable properties to suitable clients would give its contribution to the housing crisis currently going on in the Netherlands (*"5 Factors That Determine How Long a Home Sits on the Market")*.

The narrative structure followed in this story line is drill down data story type, whereas the audience of this case are the Funda brokers.

7. Literature (APA style)

Signorazzi, Aurora. "The Housing Crisis in the Netherlands: What Is It and Why Should We Worry about It? – DutchReview." *DutchReview*, 22 Oct. 2018, dutchreview.com/expat/housing/housing-crisis-in-the-netherlands/.

"Older People Staying in Large Homes Is a Key Factor to Housing Crisis: Municipalities." *NL Times*, nltimes.nl/2021/12/30/older-people-staying-large-homes-key-factor-housing-crisis-municipalities. Accessed 1 Feb. 2022.

"Seven Things to Look for When Buying a House in the Netherlands." *Expat Guide to the NetherlandsExpatica*, www.expatica.com/nl/housing/buying/seven-things-to-look-for-when-buying-a-house-in-the-netherlands-1638218/. Accessed 1 Feb. 2022.

"Renting in the Netherlands: The Ultimate Guide." *DutchReview*, 6 Sept. 2021, dutchreview.com/expat/renting-in-the-netherlands-the-ultimate-guide/. Accessed 1 Feb. 2022.

"5 Factors That Determine How Long a Home Sits on the Market." *Inman*, www.inman.com/2017/11/10/5-key-factors-average-time-sell-house-every-homeowner-needs-know/. Accessed 1 Feb. 2022.

Panel, Forbes Biz Council Expert. “Council Post: Selling a Home? 12 Factors That May Impact SaleTime.”*Forbes*, www.forbes.com/sites/forbesrealestatecouncil/2020/04/13/selling-a-home-12-factors-that-may-impact-sale-time/?sh=7f64ed983208. Accessed 1 Feb. 2022.

8. Appendices

Appendix 1: Initial sketch of the storyline