

# Visual Real Estate

## *Process Book*

COM-480

Élise  
**BOYER**  
Charles  
**GIRADOT**  
Gauthier  
**NIELLY**

# **CONTENT**

- 01** Introduction & Project Goals
- 02** Initial Research and Ideas
- 04** Challenges and Solutions
- 06** Final Project

# 01 INTRODUCTION & PROJECT GOALS

*Long story short:* We want to extract insights from real estate data across different cities and make them easy for anyone to understand.

*Longer story:* The real estate market is an unavoidable part of everyone's life. Whether you're traveling or making a long-term investment, informed decisions depend on understanding property locations, features, value, and evolving trends. Our goal is to turn complex real estate data into clear, visual stories that help people make smarter choices — wherever they are in life.

Our project explores and compares **real estate landscapes** across three of Europe's most dynamic cities: Berlin, London, and Madrid. Each city represents

a distinct slice of the European market:

**Berlin** – a city in transformation, shaped by rent control policies and rapid urban evolution.

**London** – a premium market with historic charm and sharp price contrasts.

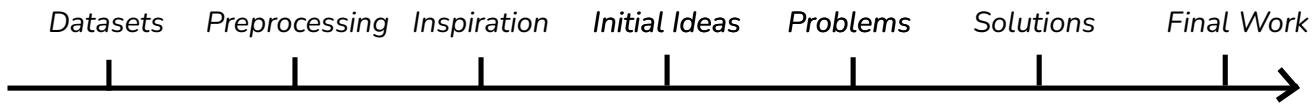
**Madrid** – vibrant, affordable, and full of neighborhood-driven diversity.

Using **data-driven visual analysis**, we turn complex real estate data into accessible insights — highlighting trends, price variations, and key drivers of property value. Whether you're looking for your next home, rental investment, or just the perfect neighborhood for your stay, our project helps you compare options with **clarity** and **confidence**.

**"Real estate cannot be lost or stolen, nor can it be carried away. Purchased with common sense, paid for in full, and managed with reasonable care, it is about the safest investment in the world."**

— Franklin D. Roosevelt, 32<sup>nd</sup> U.S. President

# 02 INITIAL RESEARCH & IDEAS



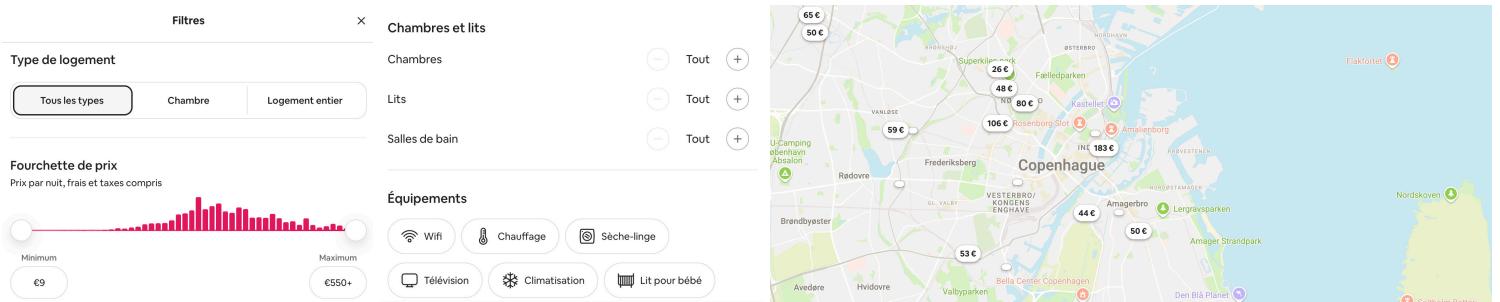
*Timeline of the project*

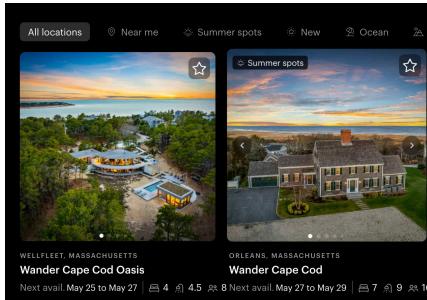
## • Dataset

Based on the idea of doing a site about real estate data, we looked for dataset on [Kaggle](#) and found 3 datasets for the following cities : Berlin, London and Madrid. The Berlin dataset includes around 5,000 entries with numerical and categorical features such as price, area, energy source, heating type, number of rooms, and zip code. London's dataset contains 1,000 well-labeled entries across 17 features, with no missing values. Madrid's dataset has over 15,000 listings with a wide range of features and limited missing data.

## • Preprocessing

The **Berlin dataset** contains around 5,000 diverse property listings across various neighborhoods, with features such as price, surface area, location, energy source, heating type, and construction year. Preprocessing included one-hot encoding of categorical variables and outlier filtering to support correlation analysis. The **London dataset** includes 1,000 well-labeled entries from 10 neighborhoods, featuring attributes like heating systems, interior styles, view classifications, and construction materials. Minimal cleaning was needed, with one-hot encoding applied for analysis. The **Madrid dataset** holds over 15,000 listings, with a wide range of property types and amenities; preprocessing involved handling limited missing values, filtering out high-end outliers, cleaning neighborhood names, and applying one-hot encoding for correlation analysis.





**Find your happy place.**

Book a Wander with inspiring views, hotel-grade amenities, dreamy beds, meticulous cleaning and 24/7 concierge service. It's a vacation home, but better.

Join 451,890 Wanderers and counting...

Location Wherever Dates Whenever Guests Whoever

**Wander Cape Cod Oasis**

4 bedrooms, 4.5 baths, 4 beds, 8 guests, 2,800 sqft, Pool outdoor, Guest house

Overview Amenities Sleep Work Location, transport, activities Pricing Testimonials

Discover the ultimate retreat at Wander Cape Cod Oasis, a reimagined mid-century modern masterpiece nestled on a private 1.6-acre wooded lot in Wellfleet's Indian Neck area. With a heated pool, 8-person hot tub, fitness center, and proximity to secluded beaches and tennis courts, this home offers the perfect blend of relaxation and adventure. Unwind in luxurious bedrooms with private terraces, cook in the fully-stocked chef's kitchen, or gather around the wood-burning fireplace under the stars.

C04710-43180

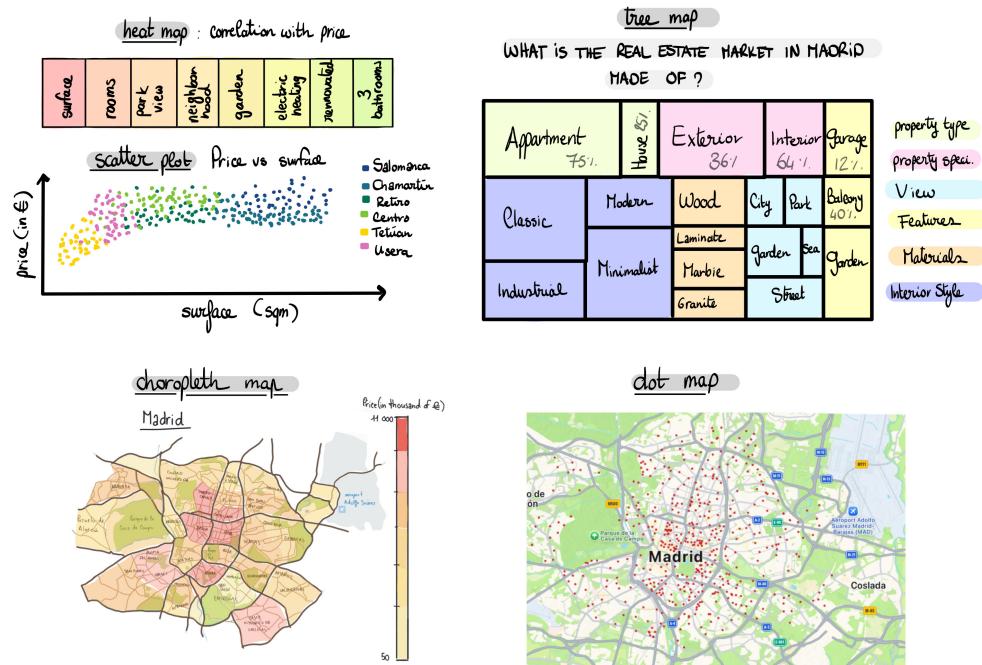
9.8/10 average guest rating Explore this location

## • Inspirations

We sought two types of inspiration for this project: one for the website's overall design and another for the visualization of our data. For the design, we aimed for an elegant, modern, and inspiring aesthetic. This visual inspiration comes from the **Wander** website. For our data visualization, we focused primarily on map-based displays, as they offer a clear and intuitive way to present analytical insights based on location. We also aimed to implement user-interaction tools with filtering options to provide a personalized and engaging experience. These approaches were inspired by the interactive maps and filter option used on the **Airbnb** platform.

## • Initial Idea & Sketches

Below are some diagrams illustrating the key ideas we aimed to explore in this project. Scatter plots and choropleth maps were successfully implemented for each city, while the dot map, heat map, and tree map encountered challenges, which are detailed in the next section.



# 03 CHALLENGES & SOLUTIONS

- **Data availability**

Despite our initial analysis of the dataset, we encountered missing information when trying to implement concrete visualization ideas. From our early sketches, two key limitations stood out: the *lack of a sufficient number of “extra features”* (beyond basic ones) to build a meaningful *treemap*, and the *absence of precise property locations*. The London dataset highlighted both issues, it had a relatively rich set of features that showed the potential for deeper analysis, but the geolocation data was limited. Many properties were clustered along main avenues, likely due to privacy concerns, making it impossible to visualize them at the street level. In contrast, only the Madrid dataset offered low-scale geolocation data, which allowed for more accurate neighborhood price insights. However, even Madrid lacked the exact property-level precision required to fully execute the “*Airbnb-style*” map we envisioned due to lack of precise property location.

**Takeaway:** Even when data *appears* detailed, a deeper analysis often reveals hidden limitations. Verifying the precision and completeness of the dataset is essential before committing to specific visualization ideas.

- **Overcoming Data Gaps Through Diverse Visualizations**

Due to the lack of certain data, we had to explore alternative ways to extract meaningful insights from the datasets. Instead of relying solely on precise geolocation or detailed feature sets, we focused on a variety of *other map types* such as density maps, dot maps, heat maps, and choropleth maps. We also incorporated *more analytical visualizations* like scatter plots, box plots, and bar charts.

- **Enhancing the Accessibility of Analytical Plots**

This shift introduced a new challenge: some of these plots were not immediately user-friendly or easy to interpret. To address this, we consistently included an “Insight” section beneath each complex plot to summarize the key takeaways.

## • Work Distribution

### Elise

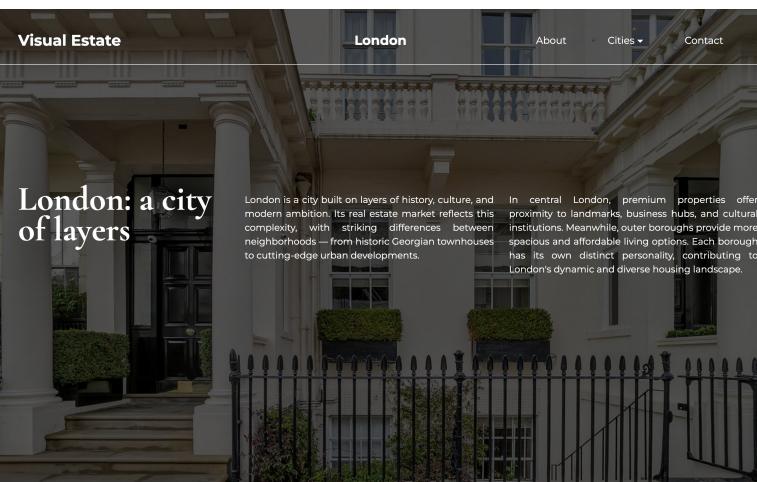
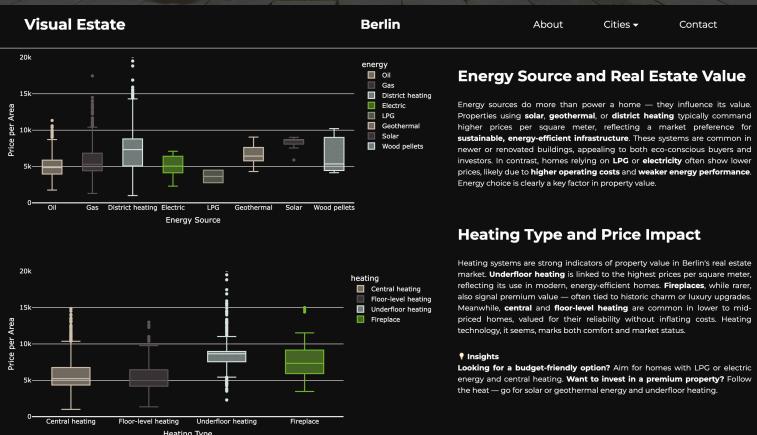
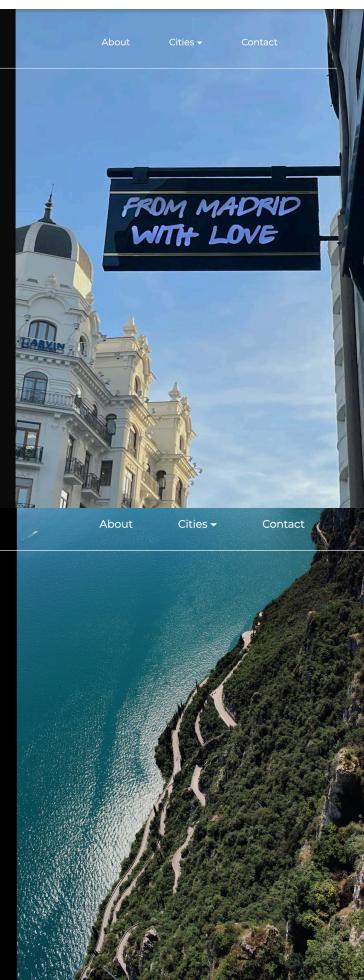
EDA of London  
 Design and structure of the full web site  
 All plots & analysis of the London page  
 Plots & analysis of the Berlin page  
 Density & Dot map of Madrid  
 Screencast  
 Process Book

### Charles

Dataset research on Madrid  
 Heatmap of average property in Madrid neighborhoods  
 Creation price vs. surface area graph for Madrid  
 Plot & analysis of Madrid

### Gauthier

EDA for Madrid & Berlin  
 Implementation of all D3.js visualizations (Berlin map, scatter plots for Madrid, feature-based plot for all cities)  
 Madrid and Berlin Analysis

# 05 FINAL WORK

And here we are! After all this work, the final project is complete. We've built a full website with an original and elegant design, accompanied by analytical visualizations to better understand the real estate landscape across different countries.

This project allowed us to grow on multiple fronts:

- On the technical side, we gained hands-on experience with basic web development using HTML, CSS, and JavaScript, as well as libraries that enabled us to create our best visualizations.
- On the design and data side, we learned how visualizations can sometimes mask messy or misleading data, making it appear clean and trustworthy a reminder to always interpret visuals with care.

We also recognized the critical importance of working with clean, well-processed datasets, and the value of designing from multiple perspectives to reach a wider audience. With more refined data and additional cities, this project could evolve into a powerful global tool for making real estate more accessible and transparent for everyone.

