Global Housing Prices Analysis

Tableau modeling

Profile - nicholas.catani | Tableau Public

INTRODUCTION

The global real estate market reflects a complex interplay of economic, social, and environmental factors that vary widely across different geographies. Utilizing a comprehensive dataset in Tableau, this analysis provides an overview of the housing market dynamics, highlighting how different variables such as location, property type, and economic conditions impact real estate values. London, Dubai, Seattle, Sydney, and Barcelona, serve as a case study to illustrate broader market trends and regional differences. These vibrant cosmopolitan cities known for their rich history and modernist architectures, present a unique case study in the global real estate market. The cities' real estate prices show significant diversity across their districts, influenced by factors such as proximity to cultural landmarks, the beach, and the central business district.

For instance, properties in Barcelona's upscale areas like Eixample and Gracia command higher prices due to their desirable location, historic value, and modern amenities. Contrastingly, peripherical suburbs often show more moderate pricing but offer potential growth due to recent urban development initiatives. Expanding our analysis globally, the business case compares Barcelona with properties in other cities within the dataset, such as those in emerging markets where real estate dynamics are influenced by rapid urbanization and economic development. This juxtaposition highlights how local factors such as economic growth, infrastructure development, and political stability influence property values differently across regions. The dataset allows for a detailed examination of how property prices have evolved, factoring in the age of the property and any remodeling efforts. Older properties that have been well-maintained or recently remodeled tend to retain or increase their values, reflecting a preference for historical buildings that are modernized to meet current living standards. Using Tableau, a temporal analysis reveals trends in property appreciation over decades, showing that properties built during certain periods, such as the pre-Olympic era, have seen varied price trajectories compared to newer constructions. The effect of remodeling is particularly pronounced, as updated features often make older properties more competitive in the fast-paced real estate market.

The insights gleaned from this analysis have significant implications for various stakeholders, including investors, real estate developers, and policymakers. For investors and developers, understanding the factors that affect real estate appreciation – such as location, property type, and the impact of remodeling – can guide investment and development decisions. For instance, investing in older properties in historically significant areas and undertaking thoughtful renovations can yield substantial returns. For policymakers, the data underscores the importance of supporting policies that preserve cultural heritage while encouraging modernization. Initiatives that facilitate property remodeling, with an emphasis on sustainability and efficiency, can enhance property values and attract further investment into the region. Overall, this detailed exploration of the global real estate market through the lens of the provided dataset not only sheds light on current market conditions but also serves as a tool for forecasting future trends. By leveraging the powerful visual analytics capabilities of Tableau, stakeholders can make informed decisions that capitalize on the dynamics of real estate markets worldwide.

DATA PREPARATION

This project aimed to consolidate and analyze housing price data from major global cities – Barcelona, Dubai, London, Seattle, and Sydney.

- Housing Prices in Barcelona (kaggle.com)
- Housing Prices in London (kaggle.com)
- Housing Price Dubai, UAE (kaggle.com)
- House price prediction (kaggle.com) Seattle
- Sydney House Prices (kaggle.com)

The datasets, which varied in structure and content, required meticulous cleaning and standardization to ensure reliable, unified insights. This section details the procedures undertaken to prepare the data for subsequent analysis and visualization in Tableau.

Each city's dataset was initially loaded to examine the structure, content, and integrity of the data. This preliminary inspection revealed variations in column naming conventions, data types, and the presence of missing values. Specifically, the datasets comprised the following entries and features:

- **Barcelona:** 3252 entries (columns: Price, Price_per_SQFT, Bedrooms, etc.)
- **Dubai:** 1905 entries (columns like Barcelona but with variations)
- **London:** 3480 entries (similar structure, varying column order)
- **Seattle:** 4600 entries (consistent columns, varying order)
- **Sydney:** 199504 entries (large dataset with similar columns; some missing values)

To facilitate analysis across these datasets, several steps were necessary:

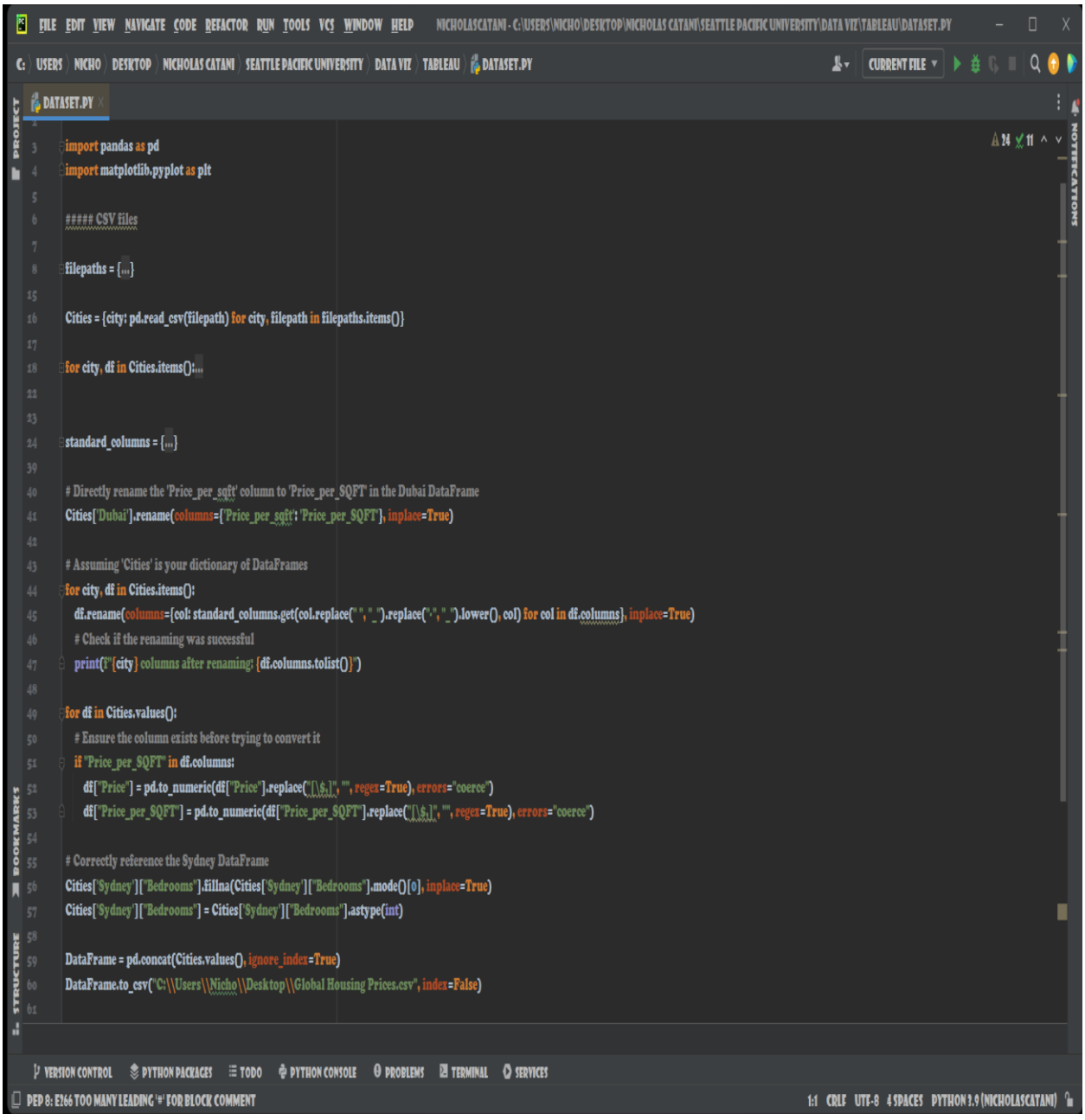
1) The datasets had inconsistencies in column names. A unified naming convention was established across all datasets to prevent potential errors during data manipulation.
2) Fields were formatted variably across datasets, often including non-numeric characters such as currency symbols. These were converted to numeric data types to enable mathematical operations.
3) The "Bedrooms" column in the Sydney dataset had missing values. These were filled using the mode (the most common value), ensuring no data points were excluded from analysis due to missing information.
4) To maintain consistency, each dataset was checked to ensure that key columns like "City" and "Zip Code" were correctly formatted and populated.

After cleaning, the datasets were merged into a single dataset. This combined dataset contained 212741 entries, with all data standardized and aligned. Key characteristics of the unified dataset included:

- Data types: Numeric conversions were successfully applied to price-related columns.
- Missing Data: Resolved issues with missing values to ensure comprehensive coverage across all variables.

- Consistency: Standardized column names and formats across datasets facilitated seamless integration.

The data preparation phase was critical in ensuring the reliability and usability of the global housing prices dataset for deeper analysis and visualization. The standardized dataset is a robust foundation for creating dynamic, informative visualizations in Tableau, designed to uncover patterns and insights in global housing market trends.



```
import pandas as pd
import matplotlib.pyplot as plt

##### CSV files

filepaths = {...}

Cities = {city: pd.read_csv(filepath) for city, filepath in filepaths.items()}

for city, df in Cities.items():...


standard_columns = {...}

# Directly rename the 'Price_per_sqft' column to 'Price_per_SQFT' in the Dubai DataFrame
Cities['Dubai'].rename(columns={'Price_per_sqft': 'Price_per_SQFT'}, inplace=True)

# Assuming 'Cities' is your dictionary of DataFrames
for city, df in Cities.items():
    df.rename(columns={col: standard_columns.get(col.replace(" ", "_").replace("-", "_").lower(), col) for col in df.columns}, inplace=True)
    # Check if the renaming was successful
    print(f"{city} columns after renaming: {df.columns.tolist()}")

for df in Cities.values():
    # Ensure the column exists before trying to convert it
    if "Price_per_SQFT" in df.columns:
        df["Price"] = pd.to_numeric(df["Price"].replace("[\$,]", "", regex=True), errors="coerce")
        df["Price_per_SQFT"] = pd.to_numeric(df["Price_per_SQFT"].replace("[\$,]", "", regex=True), errors="coerce")

# Correctly reference the Sydney DataFrame
Cities['Sydney']["Bedrooms"].fillna(Cities['Sydney']["Bedrooms"].mode()[0], inplace=True)
Cities['Sydney']["Bedrooms"] = Cities['Sydney']["Bedrooms"].astype(int)

DataFrame = pd.concat(Cities.values(), ignore_index=True)
DataFrame.to_csv("C:\\Users\\Nicho\\Desktop\\Global Housing Prices.csv", index=False)
```

CALCULATIONS

In the real estate market analysis, understanding how property prices have grown over time is crucial for assessing market trends and making informed investment decisions. The "Price Growth Rate" is a key metric that indicates the rate at which property values have appreciated since their construction or last major remodeling. This calculation can help identify potential investment opportunities and understand market dynamics in different regions.

- **Estimated Original Price**

To estimate what the property was likely worth when it was built or last remodeled, considering how property values have appreciated over time.

- Open Tableau and connect to the dataset (CSV file by clicking to other in this case).
- Go to the "Data" pane, right-click, and choose "Create Calculated Field".
- Create the "Age of Property" calculated field:
  - Name: Age of Property
  - Formula: 2024 – MIN([Built], [Remodeled])
- Create the Parameter for the Annual Appreciation Rate:
  - Right-click in the "Data" pane and select "Create Parameter".
  - Name the parameter: Annual Appreciation Rate
  - Set the data type to Float.
  - Configure the range and step size as needed (for example, set a range from 0.01 to 0.10 with a step of 0.01 to represent 1% to 10% appreciation).
- Create the "Estimate Original Price" calculated field:
  - Name: Estimated Original Price
  - Formula: [Price] / (1 + [Annual Appreciation Rate]) ^ [Age of Property]

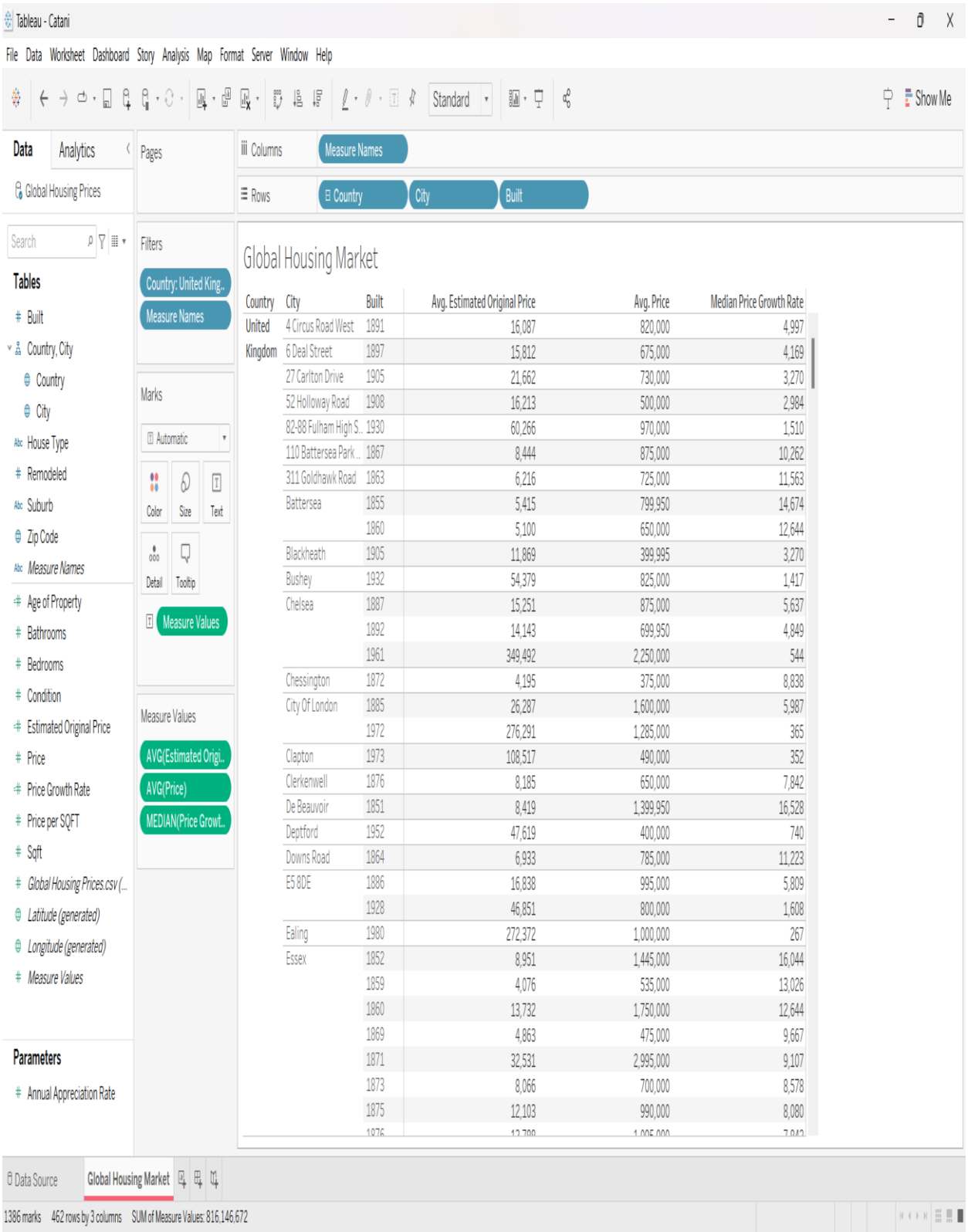
This formula adjusts the current price backward by the compounded annual appreciation rate over the number of years since the property was built or remodeled. The exponential factor accounts for the compounding effect of property value appreciation, providing a more accurate estimation of the property's initial price.

- **Price Growth Rate**

To calculate the percentage growth in the property's price from its estimated original price to the current listed price.

- Create the "Price Growth Rate" calculated field:
  - Name: Price Growth Rate
  - Formula: (([Price] – [Estimated Original Price]) / [Estimated Original Price])

This measure helps investors and analysts understand the return on investment for properties, reflecting the market's inflation or deflation of property values. It is crucial to compare the performance of properties across different markets or within the same market over different periods.



*Dollar signs and rectifications have been implemented in the workbook.

PARAMETERS

- **House Type Parameter**

This parameter will allow users to select the type of property they want to view, such as apartments, villas, etc.

- Right-click in the Parameters area in the lower part of the Data pane and select Create Parameter.
- Name: House Type Parameter
- Data Type: String
- Allowed Values: List
- Add from Field: choose column “House Type” to automatically populate the list.
- Drag the House Type parameter into the view or dashboard.
- Right-click on the parameter and select Show Parameter Control for users to interact with it.

- **Condition Rating Parameter**

This parameter enables users to filter properties based on their condition rating.

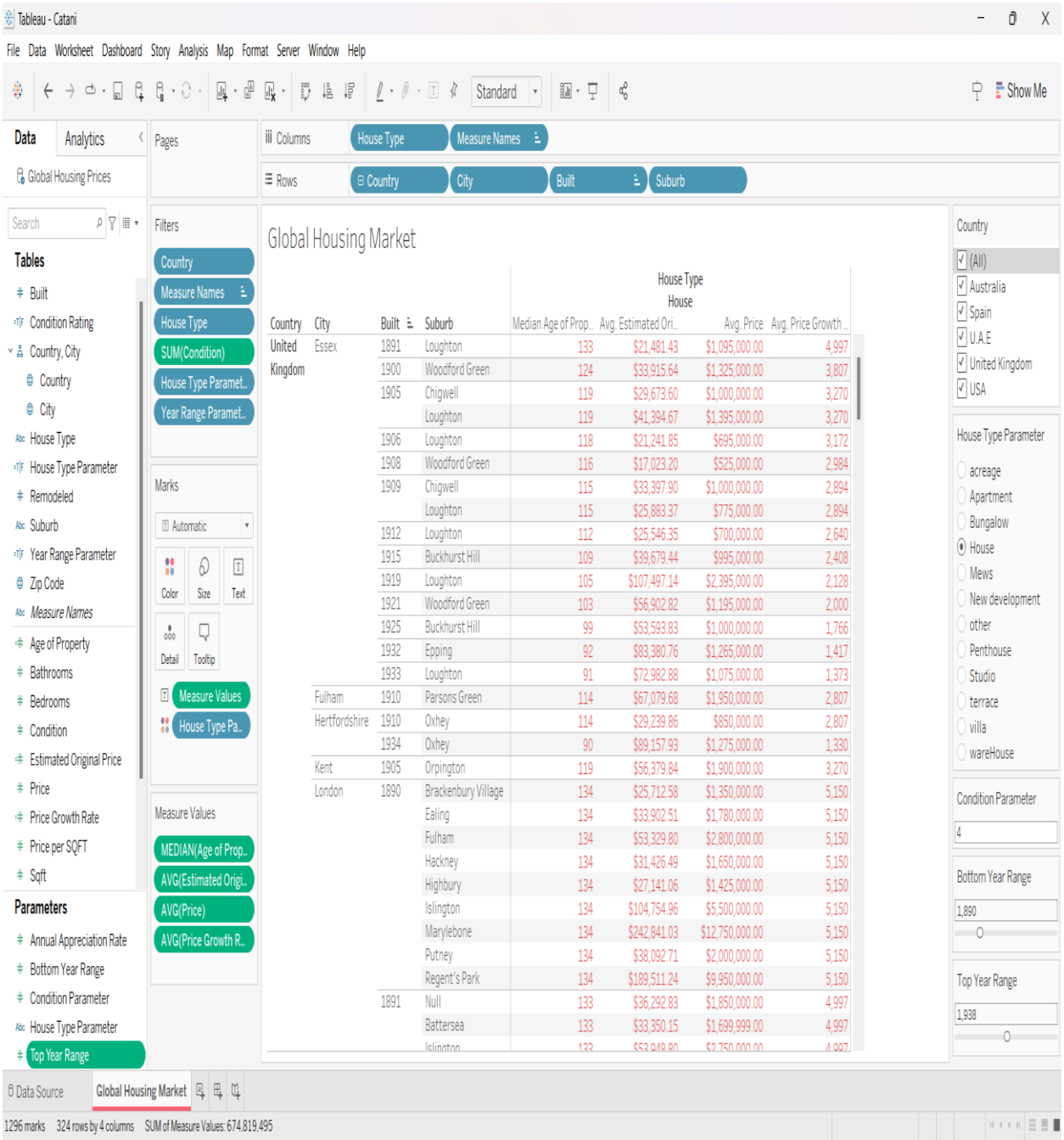
- Right-click in the Parameters section and select Create Parameter.
- Name: Condition Rating
- Data Type: Integer
- Allowable Values: Range
- Set the Minimum and Maximum based on the scale. In this example, 1 to 5.
- Step Size: 1 (allow whole numbers only).
- Like with the Property type parameter, drag the Condition Rating parameter to your view or dashboard and enable the parameter control.

- **Year Range Parameter**

This slider will allow users to select a range of years for properties built or remodeled.

- Right-click in the Parameters section and select Create Parameter.
- Name: Bottom Year Range
- Data Type: Integer
- Allowable Values: Range
- Determine the Minimum and Maximum values based on the dataset’s “Built” field.
- Step Size: You might choose a side step of 1 year or more, depending on the granularity desired.
- Display Format: Choose “Slider” to allow users to easily adjust the year range.
- Iterare the process for the Top Year Range.

- Create a Calculated field “Conditional”
- Name it “Year Range”
- Type the formula “[Built] >= [Bottom Year Range] And [Built] <= [Top Year Range]
- Drag the Year Range calculated field into the filter box and enable True.
- Enable the parameter control as a slider for easy interaction.



MAP VISUALIZATION

The first step is to ensure that the dataset has the necessary geographical information (City, Zip Code, Country) and that these fields are correctly recognized by Tableau as Geographical fields.

- Check Geographical Role: Right-click on the City, Zip Code, and Country fields in the data pane and make sure they are set to their appropriate geographical roles.
- In a new worksheet, drag the “Country” field to the Rows shelf. Tableau will automatically recognize this field as geographical data and attempt to create a map view.
- If you want to drill down to a more localized view, you can also add “City” or “Zip Code” to the Detail mark card to refine the geographical plotting.
- Drag the “Price” and “Price_per_SQFT” fields to the Color mark card. This action will color-code the map based on property prices, giving a visual representation of the distribution of both.
- Adjust the color setting: Click on the Color mark card and choose a color palette that provides good visual contrast. You might want to use a graduated color scheme where higher values are represented by more intense colors.
- Customize the tooltips to display detailed information about each property when hovered over. Include data like Price, House type, Bedrooms, Bathrooms, and any other relevant information.
- To edit tooltips, go to the Tooltip shelf and modify the HTML to include the fields you want.

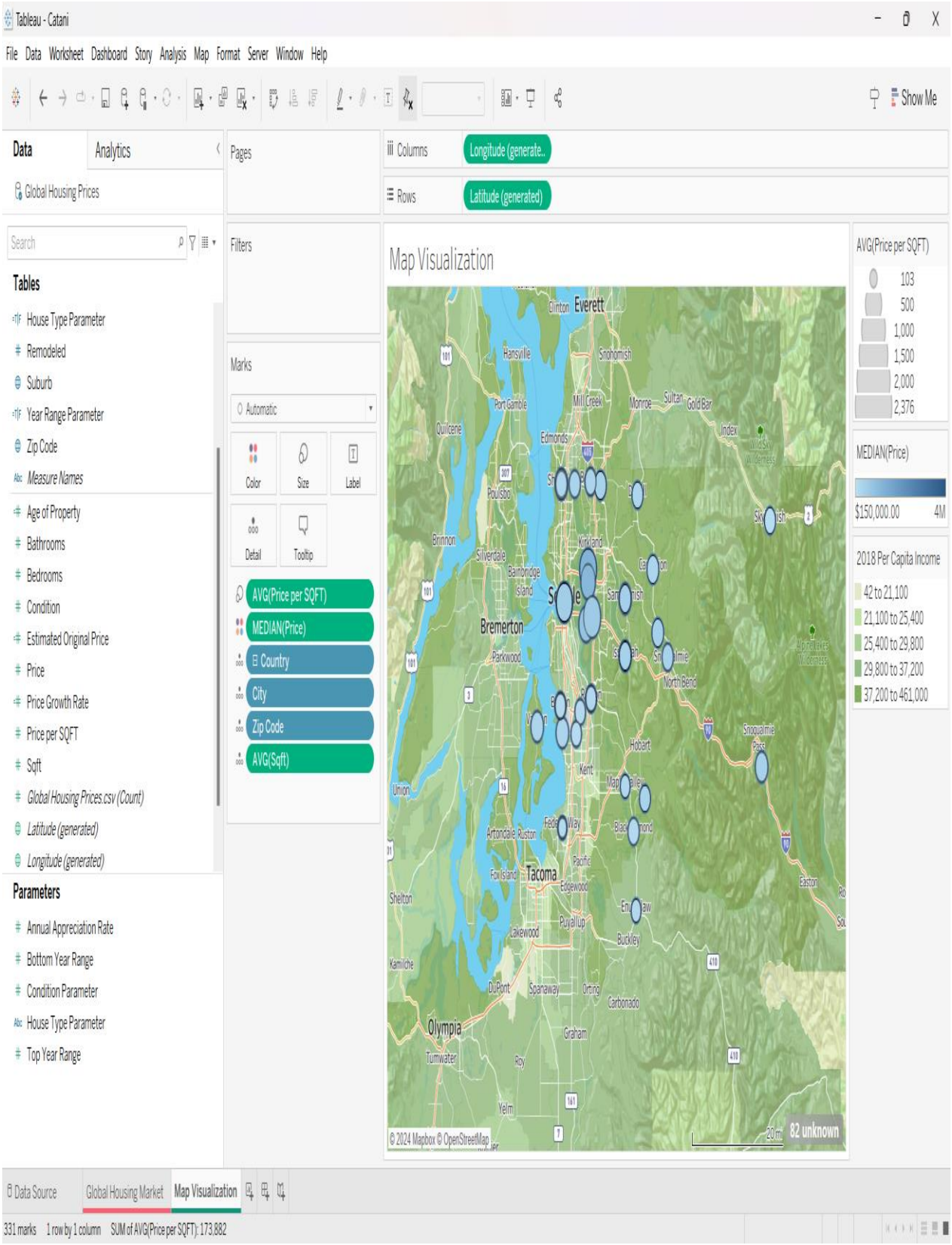
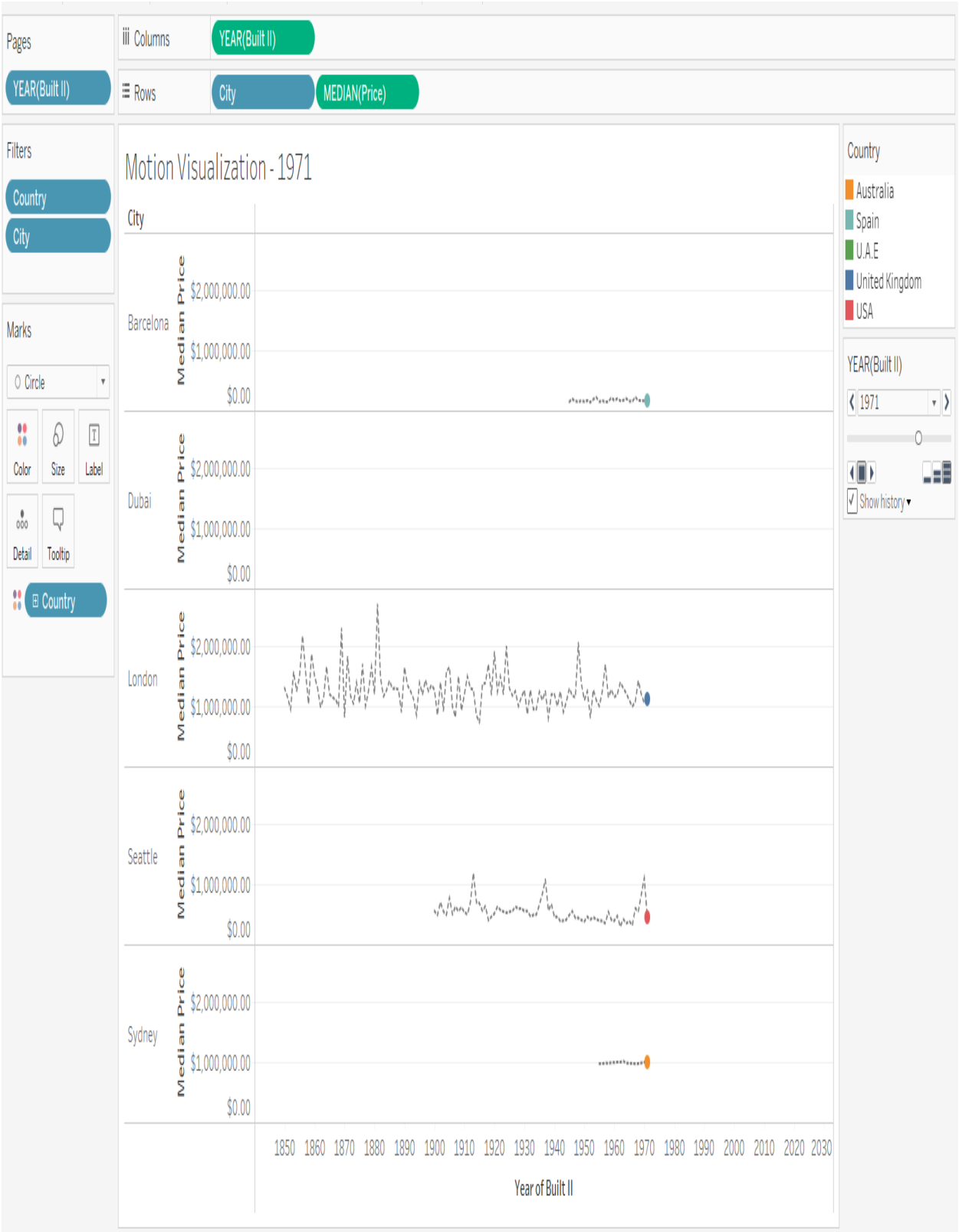


Tableau - Catani
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Global Housing Prices
Search
Tables
House Type Parameter
Remodeled
Suburb
Year Range Parameter
Zip Code
Measure Names
Age of Property
Bathrooms
Bedrooms
Condition
Estimated Original Price
Price
Price Growth Rate
Price per SQFT
Sqft
Global Housing Prices.csv (Count)
Latitude (generated)
Longitude (generated)
Parameters
Annual Appreciation Rate
Bottom Year Range
Condition Parameter
House Type Parameter
Top Year Range
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
AVG(Price per SQFT)
MEDIAN(Price)
Country
City
Zip Code
AVG(Sqft)
Columns
Longitude (generate..
Rows
Latitude (generated)
Map Visualization
AVG(Price per SQFT)
103
500
1,000
1,500
2,000
2,376
MEDIAN(Price)
$150,000.00
4M
2018 Per Capita Income
42 to 21,100
21,100 to 25,400
25,400 to 29,800
29,800 to 37,200
37,200 to 461,000
Everett
Bremerton
Tacoma
Olympia
© 2024 Mapbox © OpenStreetMap
82 unknown
Data Source
Global Housing Market
Map Visualization
331 marks
1 row by 1 column
SUM of AVG(Price per SQFT): 173,882

MOTION VISUALIZATION

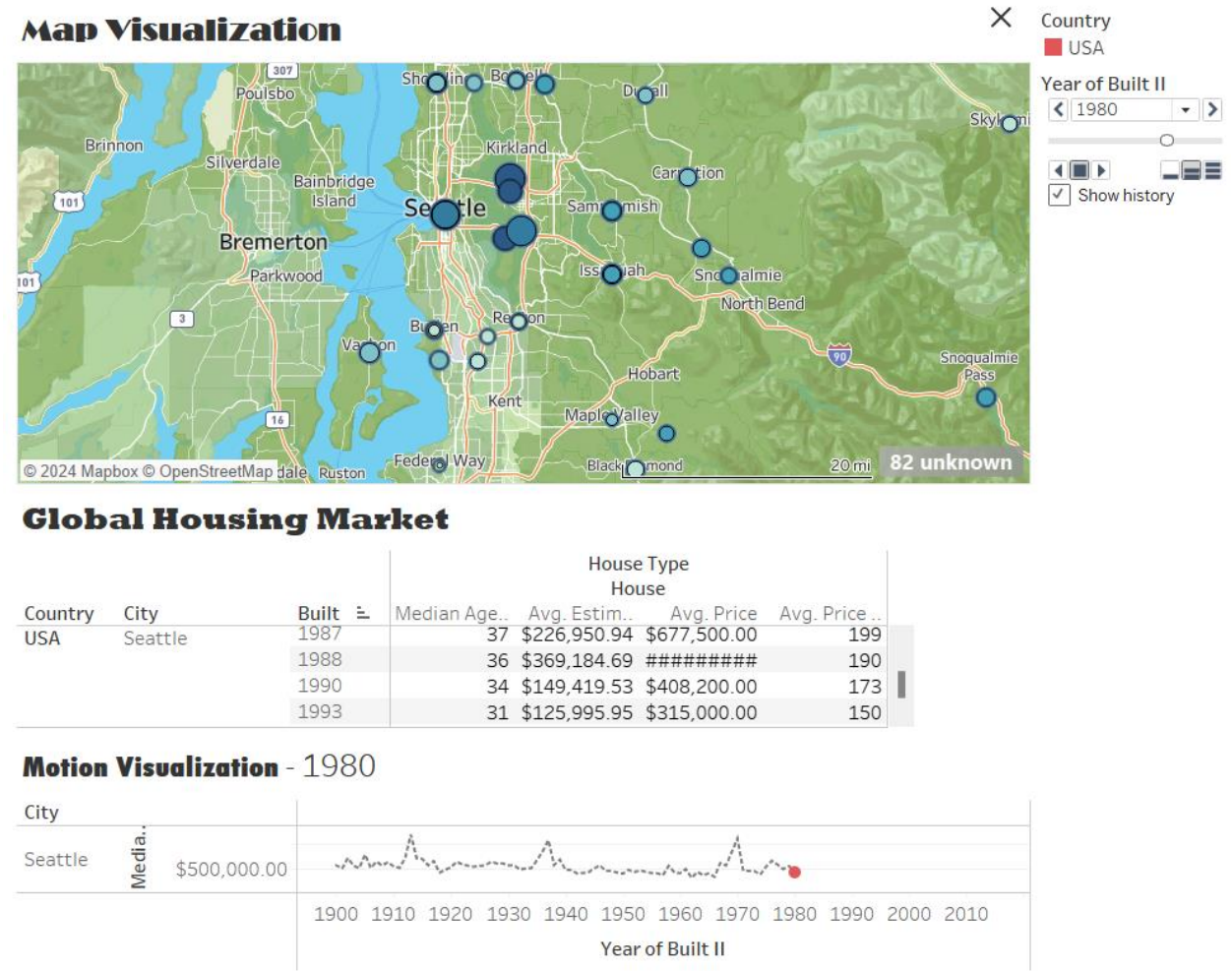
Creating a motion chart in Tableau to visualize price trends over time based on the year properties were built or remodeled can provide valuable insights into how real estate values have changed. This type of visualization uses animation to depict changes in data points through time, making it easier to identify trends and patterns.

- **Price Trends over time**:
- Go to the Data pane.
- Right-click and choose Create Calculated Field.
- Name the field "Year Considered" or a similar descriptive name.
- Enter the formula: MAX([Year Built], [Year Remodeled]). This formula will pick the latest year between the year built and the year remodeled for each property.
- Drag the "Year Considered" field to the columns shelf.
- Drag the price field to the rows shelf.
- If you want to analyze by location or property type, drag the city or house type field to the detailed mark card.
- Drag the year considered field to the Pages shelf. Tableau uses the Pages shelf to create motion in the visualization by animating the data points across the specified dimension – in this case, time.
- Adjust the size and color of the marks to make trends clearer. For instance, use different colors for different property types or sizes to reflect price ranges.
- Go to the Marks card and select the type of mark that best represents your data.
- Click on the play button located at the bottom of the Pages shelf to see how the data animates over time.
- Adjust the speed of the animation with the controls next to the play button to slow down or speed up the animation as needed.
- Utilize the history trails feature to show the path of each data point over time. Right-click on the Time field on the Pages shelf, select "Show History", and then adjust the settings to show trails.


Pages
YEAR(Built II)
Columns
YEAR(Built II)
Rows
City
MEDIAN(Price)
Filters
Country
City
Marks
Circle
Color
Size
Label
Detail
Tooltip
Country
Motion Visualization - 1971
City
Barcelona
Dubai
London
Seattle
Sydney
Median Price
$2,000,000.00
$1,000,000.00
$0.00
1850 1860 1870 1880 1890 1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030
Year of Built II
Country
Australia
Spain
U.A.E
United Kingdom
USA
YEAR(Built II)
1971
Show history

DASHBOARD

- In Tableau, go to the bottom of the screen where the sheet tabs are located and click on the “New Dashboard” tab.
- Drag and drop each visualization from the sheets list onto the dashboard area.
- Arrange logically. For example, place the map prominently at the top or side, with the table and motion chart underneath or beside it, depending on your layout preference.
- Drag relevant filters. Ensure you set them to affect all visualizations on the dashboard where applicable.
- Right-click on the filter or parameter control on the dashboard, select “Apply to worksheets”, and then choose “All Using this data source”.
- Add a clear title and necessary annotations or text boxes to guide users through the dashboard. Explain what each visualization represents and how to interact with them.
- Preview the dashboard in presentation mode to see how it looks to end-users.
- Interact with the filters and parameters to ensure they modify all relevant visualizations correctly.
- Check performance; if the dashboard is low to load or interact with, consider optimizing by reducing the complexity of calculations or the number of data points displayed.


Map Visualization
Country
USA
Year of Built II
1980
Show history
Poulsbo
Brinnon
Silverdale
Bainbridge Island
Bremerton
Parkwood
Kirkland
Seattle
Issaquah
Snoqualmie
North Bend
Snoqualmie Pass
Hobart
Kent
Maple Valley
Federal Way
Ruston
© 2024 Mapbox © OpenStreetMap
20 mi
82 unknown
Global Housing Market
House Type
House
Country
City
Built
Median Age..
Avg. Estim..
Avg. Price
Avg. Price ..
USA
Seattle
1987
37
$226,950.94
$677,500.00
199
1988
36
$369,184.69
#########
190
1990
34
$149,419.53
$408,200.00
173
1993
31
$125,995.95
$315,000.00
150
Motion Visualization - 1980
City
Seattle
Media..
$500,000.00
1900 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 2010
Year of Built II
g Market
Map Visualization
Motion Visualization
Story
Dashboard

WHAT WENT WELL

The analysis of global housing prices using Tableau leveraged my pre-existing knowledge and skills with the software, which greatly contributed to the success of the project. Having a strong foundation enabled me to efficiently manipulate and visualize the data, transforming raw datasets into insightful, actionable information. My ability to apply the CRISP-DM approach was instrumental in structuring the analysis effectively. This methodology guided me through understanding the business problem, data understanding, data preparation, modeling, evaluation, and deployment, ensuring a comprehensive analysis framework.

One of the major successes was the seamless integration of datasets from multiple cities into a single, coherent database. The meticulous data preparation, which included cleaning, standardizing, and merging data, was conducted with precision, thanks to the robust functionalities of Tableau and my prior experience with similar tasks. This groundwork facilitated accurate and dynamic visualizations that highlighted regional market trends and differences effectively.

Furthermore, the creation and use of parameters to filter and drill down into specific data points exemplified the application of my Tableau expertise. These elements not only enhanced the user experience but also provided stakeholders with a tool to explore the data according to their specific interests and hypotheses about the real estate market.

The application of the CRISP-DM approach ensured that each phase of the project was executed with attention to detail and adherence to the project's objectives. This structured approach helped to maintain focus on the end goal and facilitated a systematic exploration of the data, which was crucial given the complexity and scope of the analysis

WHAT DID NOT GO WELL

Despite the successes, there were several challenges encountered during the project. One significant issue was the initial underestimation of the time and resources required for the data preparation phase. The datasets from different cities varied greatly in structure and completeness, which necessitated extensive cleaning and transformation efforts. This phase took considerably longer than anticipated, which delayed subsequent analysis steps.

Another challenge was related to the limitations of Tableau in handling very large datasets efficiently. The Sydney dataset, for example, was particularly large and caused performance issues that affected the speed of data manipulation and visualization rendering. This sometimes led to frustration and hindered my ability to iterate quickly on the visualizations.

There were also difficulties in accurately modeling some of the more complex real estate market dynamics, such as the impact of economic indicators on property prices. While Tableau is excellent for visual analysis, its capabilities in statistical modeling are limited, which meant that some of the deeper analytical insights had to be simplified or foregone.

WHAT WOULD YOU DO DIFFERENTLY NEXT TIME

In future projects, there are several adjustments I would make based on the experiences and lessons learned from the analysis. Firstly, I would allocate more time and resources to the data preparation phase. Anticipating the complexities involved in cleaning and merging data from various sources would allow for a more realistic project timeline and reduce pressure during the initial stages.

I would also consider integrating Tableau with other analytics tools that are better suited for handling large datasets or performing complex statistical analyses. Tools such as Python or R could complement Tableau's visual capabilities and provide more robust options for data manipulation and modeling, particularly for large datasets like Sydney's

Additionally, expanding the team to include specialists in areas such as database management and statistical analysis could enhance the project's capability to deal with complex datasets and sophisticatd modeling requirements. This would not only improve the efficiency of the data preparion phase but also enrich the analytical depth of the project.

Lastly, engaging more frequently with stakeholders throughout the project could provide continual feedback and adjustment opportunities. This iterative approach would help ensure that the final deliverables more closely align with the stakeholders' needs and expectatioins, and could also help identify potential issues or additionatl requirements early in the process