

Project Documentation: Airbnb Market Analysis

Purpose

The project is created to show real state investors, property owners, analysts, and other professionals how the revenue of Airbnbs in California is correlated to the commodities inside the properties. The app provides interactive visualizations to perform a market analysis in four Californian cities: Big Bear Lake, Big Bear City, Joshua Tree, and Yucca Valley. Users can explore key metrics like nightly rate and occupancy based on different property features and compare revenue across cities.

Data Description

The data was extracted from the Kaggle dataset “*Airbnb Market Analysis & Real Estate Sales Data*.” Source: [<https://www.kaggle.com/datasets/computingvictor/zillow-market-analysis-and-real-estate-sales-data>]

Two files are utilized “market_analysis.csv” and “geolocation.csv.” The files are merged to get the location of all Airbnb points.

The dataset “market_analysis.csv” displays a collection of Airbnb properties listed in 2019. It encompasses metrics like generated revenue, occupancy, nightly rate, and availability. It also shows the commodities inside the properties, such as the number of bathrooms and bedrooms. Each row represents a single property, and all rows have a unique ID associated with them.

Questions to answer

1. How do different property features (bedrooms, bathrooms, guests) relate to average revenue, nightly rate, and occupancy?
2. What are the top-earning Airbnbs in a selected city?
3. How does revenue correlate with nightly rate across different cities?

Insights from Data

After being able to filter out Airbnbs their revenue based on input these are the most valuable insights from the data:

1. Airbnbs with three bedrooms and three bathrooms have the highest revenue average.
2. Even though there exists a relationship between nightly rate and average revenue, there exist some outliers that have a high nightly rate but not as much revenue compared to other Airbnbs in the same range.
3. By filtering all highest earning Airbnbs in different cities most of them are dispersed, meaning that there is one popular Airbnb within each city section that has a high revenue.

Reproducibility

The data was merged between two datasets based on the unique `id` field that was constant in both files. Then the data was cleaned by setting all necessary field to numeric. Because the floats used “,” instead of “.” a substitution was needed in all decimal fields. Additionally, all points that did not have a revenue field were filtered out. Finally, because the data was gathered monthly there were various rows for a single observation, thus I decided to unite them based on their id and get their average revenue, nightly rate, occupancy, and length of stay to create a single row per property.

Design Decisions

What: Visualizations include maps, bar charts, and scatter plots to represent various commodities and numerical features of Airbnb market data.

Why: The visualizations aim to provide insights into revenue, nightly rate, and occupancy based on different property features and city selections.

How: The Shiny App uses leaflet for maps and ggplot2 for other visualizations. Different encodings through a blue color palette are used to represent relevant metrics and property features, as well as the time log visualization.

In order to make the app consistent I decided to use a blue palette of colors in all of my visualizations. I decided the main visualization to be an interactive map so the users can play with the data before looking for factor relationships. For the comparison by bedrooms and the relationship of nightly rate and revenue I decided to create scatter plots so the user could identify outliers within the data. Finally, the comparison by bathrooms and the time log are in bar plot format.

What needs improvement?

Creating more visualizations that use the fields “length_of_stay,” and comparing this metric based of bedrooms and bathrooms. Moreover, if I wanted to publish this app I would need a personal section where people could contact me about the market analysis and if they have any questions.

References

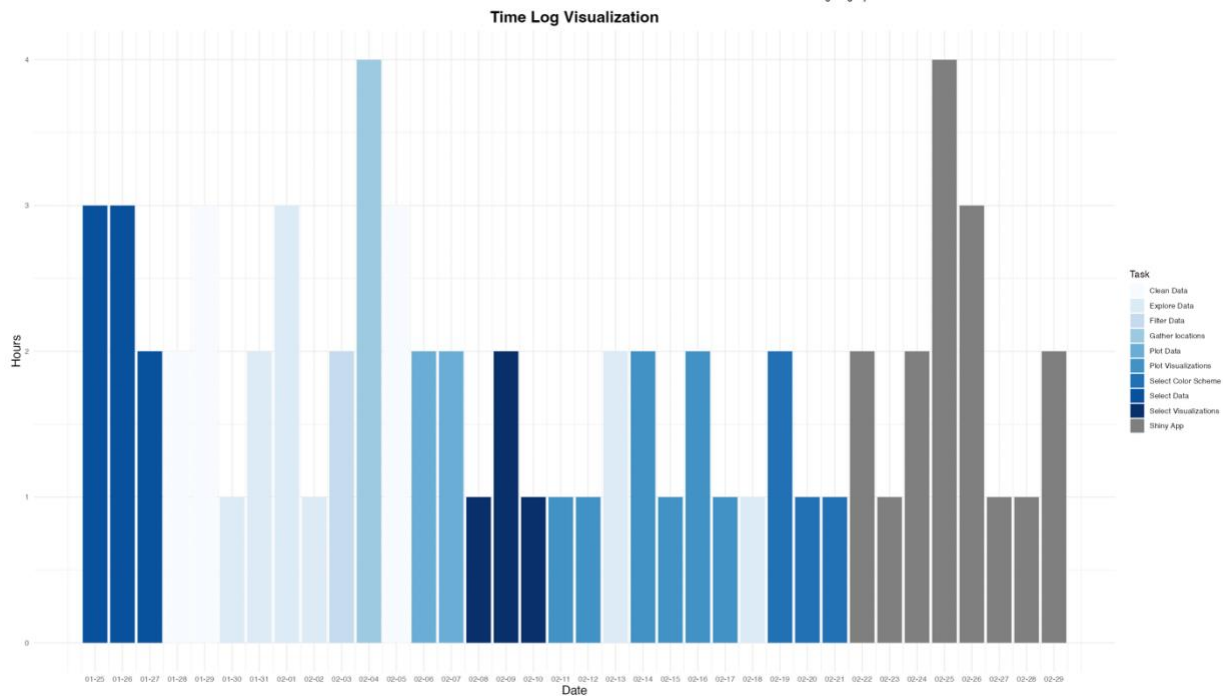
 *Airbnb Market Analysis & Real Estate Sales Data*. 2024. Kaggle. January 26, 2024.
<https://www.kaggle.com/datasets/computingvictor/zillow-market-analysis-and-real-estate-sales-data>.

Beeley, Chris. *Web Application Development with R Using Shiny: Integrate the Power of R with the Simplicity of Shiny to Deliver Cutting-Edge Analytics over the Web, Second Edition*. 2nd ed. Birmingham: PACKT Publishing, 2016.

Munzner, Tamara, and Eamonn Maguire. *Visualization Analysis & Design*. Boca Raton, FL: CRC Press/Taylor & Francis Group, 2015. <https://www.cs.ubc.ca/%7Etm/vadbook/>.

Appendix

1. Log chart



2. Requirements description:

- A particularly nice piece of code to display the main map visualization. I utilized the library leaflet with latitude and longitude information to display pins wherever there is a registered Airbnb.
- Combined two datasets, one for geolocation of the Airbnbs and the other one with the characteristics of each Airbnb.
- Good recording exploring the visualizations and app, under the name "App_Demo" in the Github repository.
- Well-designed time log and chart displayed above.