

UNCOVERING CONSUMER PREFERENCES THROUGH BEER REVIEW ANALYTICS

Team #143: Ramon Benavides, Rory Covarrubia, Cesar Fabrega Tapia, Angie Camila Perez, Alejandra Sevilla



INTRODUCTION / MOTIVATION

What is the problem?

The craft beer market is experiencing rapid growth. While this offers opportunities, breweries face increasing competition. Consumer preferences evolve quickly, and traditional market research methods may struggle to keep pace.

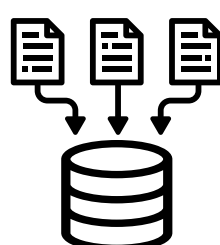
This project analyzes a large-scale beer review dataset to identify key factors influencing consumer preferences. We examine flavor, aroma, ABV, and other attributes to reveal what drives positive (or negative) consumer sentiment.

Why is it important?

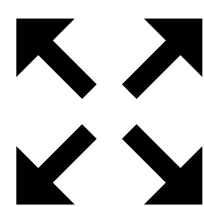
Our analysis provides breweries with data-driven insights to:

- Identify which styles and flavor profiles resonate with different consumer segments.
- Develop products that specifically target consumer preferences.
- Optimize marketing strategies for greater reach and impact.

DATASET



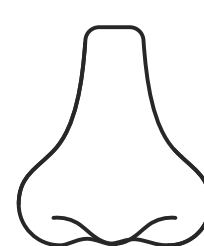
Publicly available dataset from Kaggle (Beer Reviews Data).



Over 1.5 million records from the BeerAdvocates website



Categorized beer styles, enabling analysis of trends within specific types (IPAs, stouts, etc.)



Beers rated on key sensory attributes, including aroma, taste, appearance, and overall palate experience.



Alcohol by volume (ABV) is included for each beer, a key factor influencing consumer preferences.



Geospatial data of breweries allows for analysis of regional trends and brewery popularity.

OUR APPROACH

Our Methods:

- Data-Driven Segmentation:** We applied the k-mean clustering algorithm to a dataset of 1.5 million beer reviews. This groups similar reviews, uncovering distinct consumer preference segments.
- Visualizing Trends:** We built interactive Tableau dashboards to display beer ratings, flavor preferences, geographical patterns, and more.
- Geolocation:** We integrated geographical data from the Google Places API to analyze how consumer preferences vary across different regions – a unique aspect of our analysis.

Why This Works:

It combines the power of data analysis with clear, interactive visualizations. This gives breweries deep insights into the types of beers consumers favor and the regional trends shaping the market.

EXPERIMENTS / RESULTS

1. Data Preprocessing:

- Enriched dataset with geographical coordinates using Google Places API
- Handled missing values, inconsistencies, and removed duplicate entries through data cleaning

2. Exploratory Data Analysis (EDA)

- Visualized overall ratings through histograms and ABV distributions using box plots
- Identified prevalent ABV range and common beer styles, revealing general positivity in ratings

3. Clustering Analysis

- Applied K-means clustering with Elbow Method for cluster selection
- Assessed cluster quality using Davies-Bouldin index and Calinski-Harabasz score
- Train-test split validation

4. Key Results

- 7 distinct beer review clusters identified
- Positive correlation between ABV and overall review scores ($r = 0.23$, $p < 0.001$)
- Top reviewed styles: American IPA, Russian Imperial Stout, American Double IPA
- High-performing breweries concentrated in established craft beer regions



Dataset Selection



Dataset Cleaning



EDA



Geocoding API



Conclusions



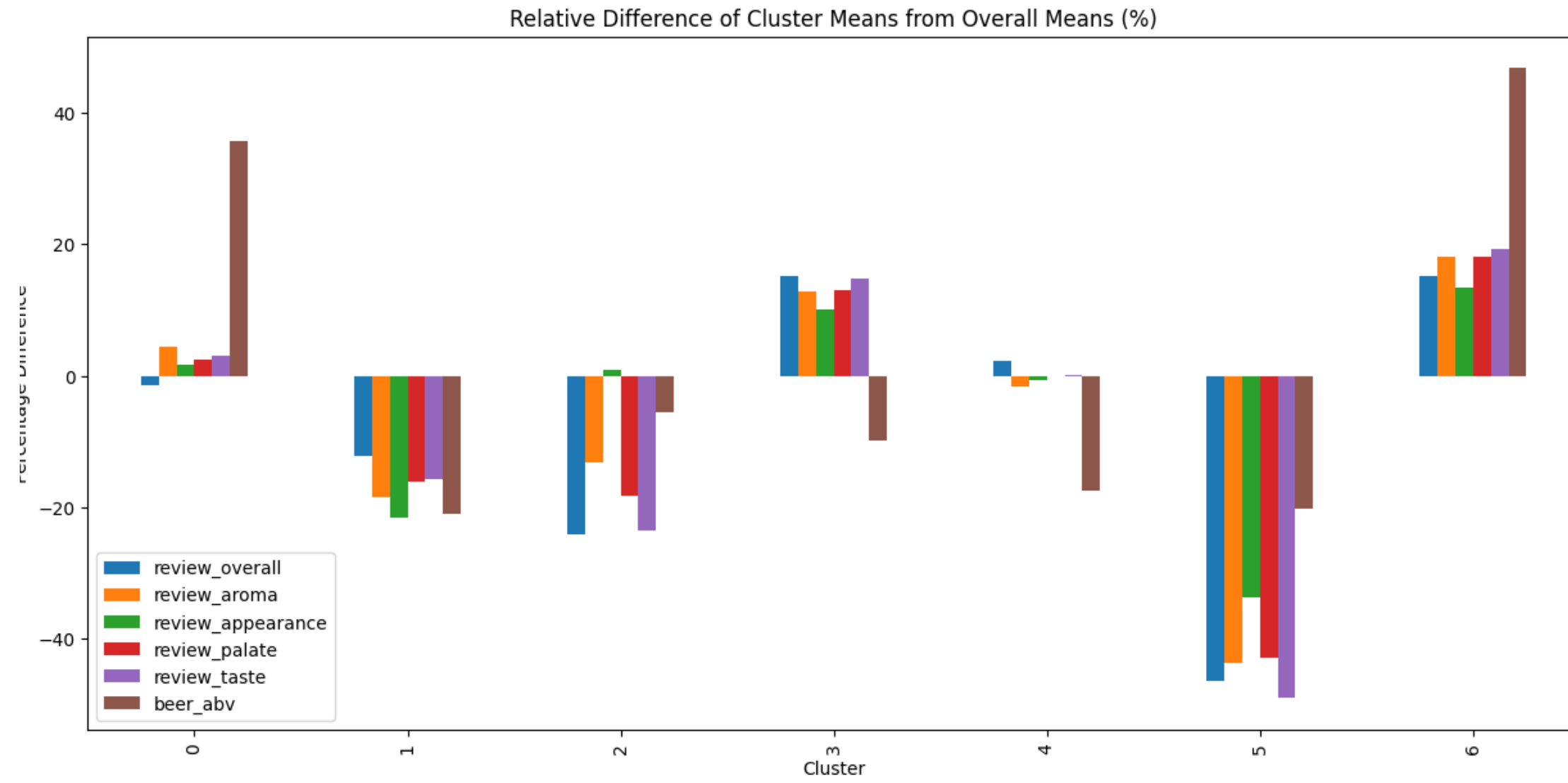
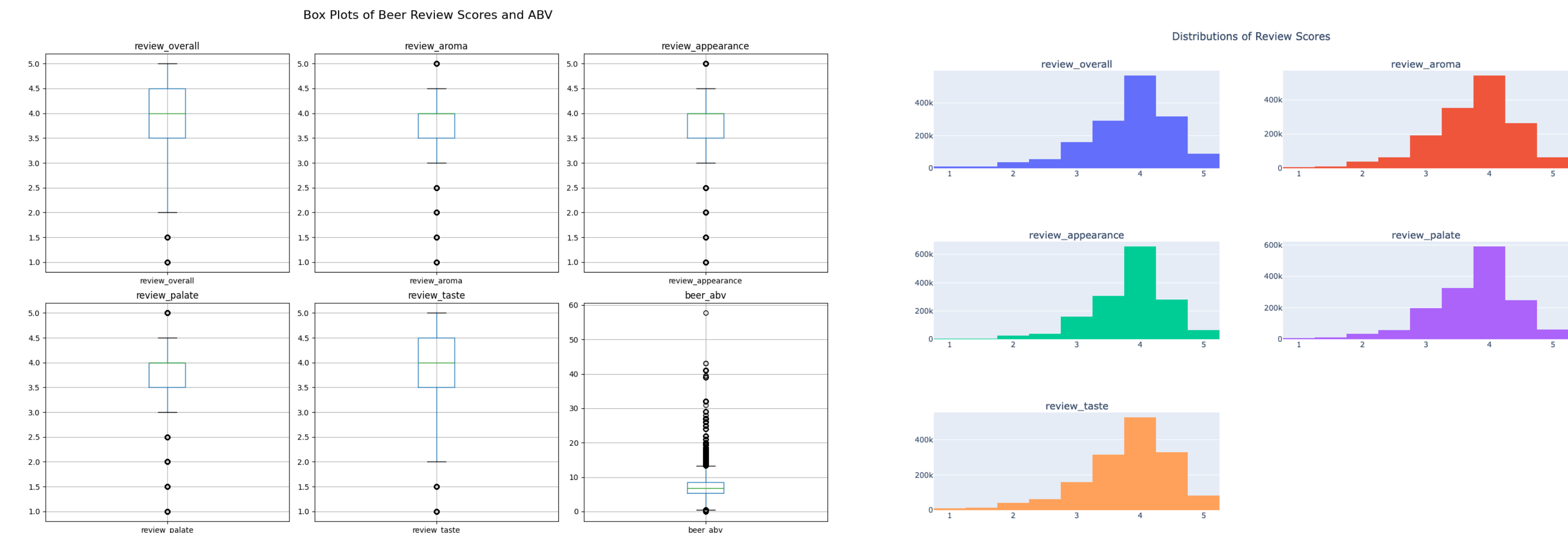
Visualization



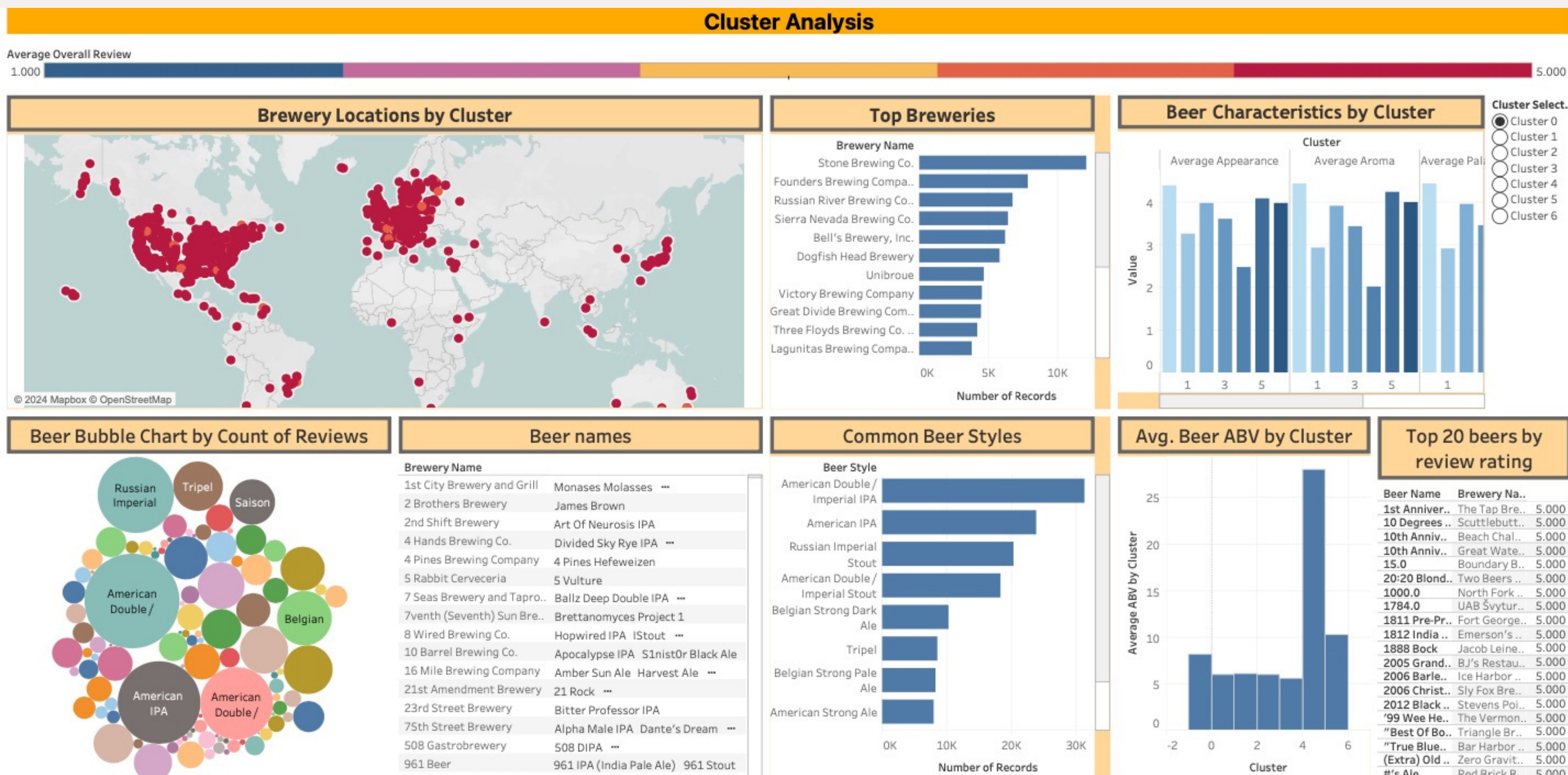
Results



Clustering



VISUAL DASHBOARD



CONCLUSIONS AND FUTURE WORK

Conclusions

- Main Findings:** Our analysis segmented beer reviews into clusters that delineate distinct consumer preferences and sensory attributes of beers. Key findings include a preference for robust, high-quality craft beers and specific sensory profiles that resonate with diverse consumer groups.

- Implications:** The clustering information can assist breweries in targeting specific consumer segments more effectively. By aligning product characteristics with the preferences highlighted in different clusters, breweries can tailor their marketing strategies and product development.

Future Work

- Limitations:** The study predominantly focused on data from reviews, which may not fully represent the broader consumer base. The geographical analysis, although enriched with coordinates from the Google Places API, was limited to the data available in the beer reviews dataset, which may not cover all regions comprehensively.
- Future Directions:** Further research could incorporate additional data sources to capture a wider array of consumer preferences. Implementing time series analysis could also uncover trends over time, providing information into the temporal dynamics of beer popularity and consumer preferences.