Q&A

##### ***How effective is the predictive model in estimating the number of breweries in a city based on its population?***

The effectiveness of the predictive model in estimating the number of breweries in a city based on its population can be evaluated using the Mean Squared Error (MSE) and R-squared (R²) values obtained from the model's performance.

Mean Squared Error (MSE): The MSE is a measure of the average squared difference between the actual and the predicted values. A lower MSE value indicates that the model's predictions are closer to the actual number of breweries, hence a more effective model. An MSE of 5636.61, as obtained in the analysis, suggests that there is a significant average error in the predictions made by the model.

R-squared (R²): The R² value represents the proportion of the variance for the dependent variable (number of breweries) that's explained by the independent variable(s) in the model. An R² value closer to 1 indicates a model that explains more of the variance. With an R² of 0.0213, the model explains only about 2.13% of the variance in the number of breweries, which is quite low. This indicates that the population size alone does not effectively predict the number of breweries in a city and that other factors likely play a significant role

#### **The type of brewery most likely to be found in a city based on demographic data?**

To determine the type of brewery most likely to be found in a city based on demographic data, we would typically use a classification model. The performance of this model can be assessed by looking at various metrics such as accuracy, precision, recall, and F1-score for each brewery type category.

From the provided classification report, the model achieved an accuracy of approximately 54.04%. This means that for about 54% of the cities, the model could correctly predict the most likely type of brewery based on the city's population.

However, the detailed classification report shows varying precision and recall across different brewery types. The model performs well in identifying 'Brewpubs' with a high recall (0.98), indicating it can recognize this category effectively, but with a precision of 0.54, it also has a fair number of false positives. For 'Microbreweries', the precision is 0.37, indicating a lower rate of correct positive predictions, and the recall is 0.03, suggesting that the model rarely identifies this category correctly.

The classification model's reliance on a single feature, the city's population, limits its ability to predict the brewery type accurately. The feature importance score indicates that population is the sole feature used, which implies a need for a more complex model that includes additional demographic and possibly economic data to improve predictive accuracy.

In conclusion, the current model based on population alone can predict some brewery types to a certain extent but is not highly effective across all categories. Including more demographic factors could potentially increase the model's performance in predicting the type of brewery most likely to be found in a city.

#### **Are certain types of breweries (e.g., Brewpub, Microbrewery) more prevalent in specific regions or states?**

The heatmap would visually represent the number of each type of brewery in each state. Each cell in the heatmap would correspond to a specific brewery type in a state, with the color intensity indicating the frequency of that brewery type. This visualization would allow us to quickly identify which brewery types are most common in each state ( in this case its CA) and observe any regional trends or patterns.

A screenshot of a data visualization

Description automatically generated

#### **Which demographic factors most strongly influence the presence of a particular type of brewery in a city?**

The array [1.] indicates that the feature 'Population' has a feature importance score of 1.0. This suggests that in the Random Forest model, the population is the only feature used and is thus considered fully influential in predicting the type of brewery.

The classification report shows the precision, recall, and F1-score for each brewery type.

For most brewery types, both precision and recall are 0.00, indicating that the model is not able to predict these categories effectively based on population alone.

For 'Brewpub', the model has a precision of 0.54 and a recall of 0.98, suggesting that while it can identify most Brewpubs (high recall), it is less precise (many false positives).

'Microbrewery' has a slightly better precision of 0.37 but a low recall of 0.03.

The overall accuracy of the model is 0.54, meaning it correctly predicts the brewery type 54% of the time across all predictions.

The results suggest that while population size does have some influence on predicting brewery types, particularly for Brewpubs and Microbreweries, it is not sufficient on its own to effectively predict all types of breweries. The model struggles with less common brewery types, likely due to the limited diversity of features and possible imbalances in the dataset.

These insights indicate that additional demographic or economic factors might be needed to improve the model's predictive power for a wider range of brewery types.

#### **Find common characteristics among cities with a high number of a specific type of brewery (e.g., microbreweries)?**

## This result suggests that, according to the model, the population size of a city is highly influential in determining the likelihood of the city having a high number of microbreweries. However, it's important to note that this analysis is based on a single feature, and the addition of other relevant demographic or economic factors could provide a more comprehensive understanding of what drives the presence of microbreweries in cities.