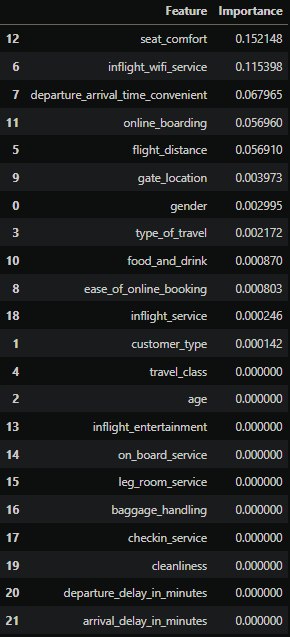
Model:

For 1 of the modeling notebooks, the preprocessing involved casting all of the ratings features to strings so they could be one hot encoded and dropping the ID column. The numeric features included: age, flight distance, departure delay (minutes), and arrival delay (minutes); the binary features included: gender, customer type, and type of travel; finally, the categorical features included all of the ratings features and travel class. 5 separate pipelines were then tested to determine if label encoding or ordinal encoding the categorical features was better and to determine how the ratings features impacted the models. Comparing Logistic Regression, ADA Boost Classifier, Gradient Boosting Classifier, XGBoost Classifier, and LightGBM on accuracy and ROC AUC, the best models looked to be the Gradient Boosting, XGBoost, and LightGBM for the datasets that included all of the ratings features. Off all of them, the LightGBM for the One Hot Encoded categorical features was the best because the feature importances placed a fairly even importance on all of the features. For the Gradient Boosting and XGBoost, 1-2 features had an importance over 10% while the next highest feature was about half as important, while the LightGBM had the highest feature having 8% importance and going from there.

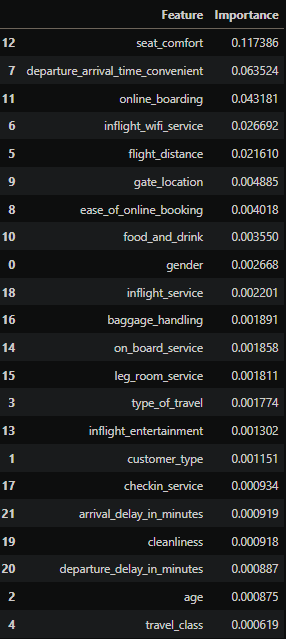
After determining the best models, Grid Searches were run on the LGBM for the One-Hot Encoded and Label Encoded data for all the features and the LGBM and XGB for the One-Hot Encoded data with some (but not all) features removed. The grid searches looked to find the best n\_estimators and learning\_rate by scoring on the f1 score and refitting on the f1 score. Based on f1 score the full feature list with One-Hot encoding and Label encoding were essentially identical. However, the LGBM for the One-Hot Encoding was slightly better, so it was used for cross validation.

The cross validation produced the following results for the One-Hot Encoded LGBM: mean accuracy: 0.9646; mean f1 score: 0.9584; mean precision: 0.9757; mean recall: 0.9417; mean ROC AUC: 0.9953. However, these score were worse than the models run with some of the numeric features binned and the rating features treated as numeric values, so those models were used instead.

GB Feature Importance:



XGB Feature Importance:



LGBM Feature Importances:

