PREDICTING RESTAURANT CHURN

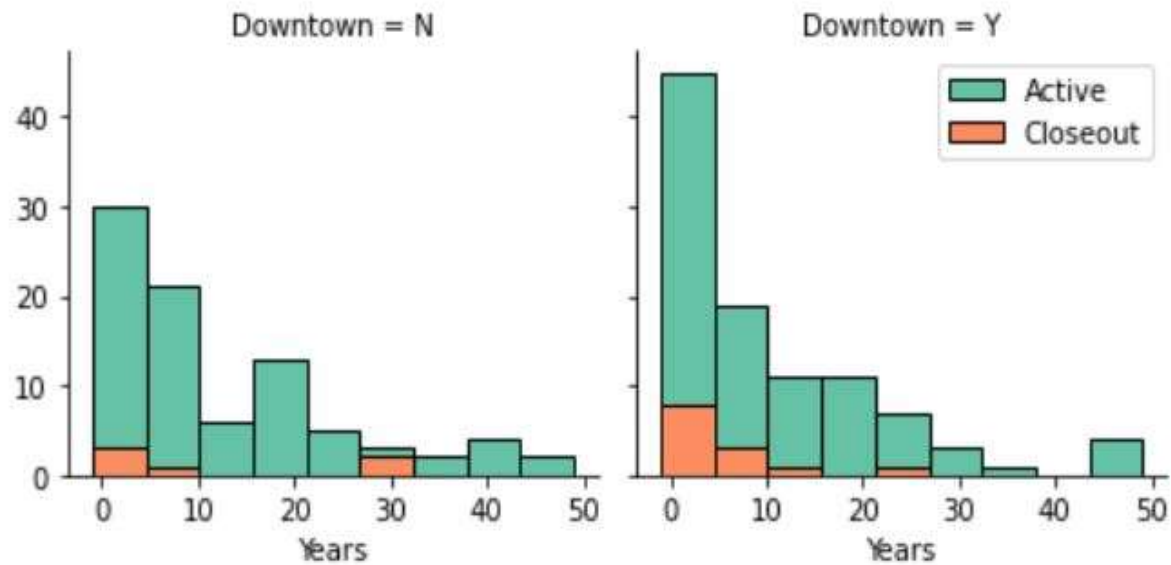ARE FOURSQUARE RATINGS A GOOD INDICATOR OF IMPENDING BUSINESS FAILURE?

PREDICTING BUSINESS FAILURE IS IMPORTANT TO LOCAL GOVERNMENT

Early recognition of failing businesses could help city hall develop policies to aid local restaurants and prop up its tax base.

DATASET

- Restaurant data from San Rafael business license records

- Key restaurant features extracted for the analysis: district name associated with the address, form of ownership (sole proprietor, corporation, etc.), number of years in business, and whether a restaurant has been "red tagged" (temporarily shut down by county officials for health code violations)

- Foursquare API used to locate restaurant rating

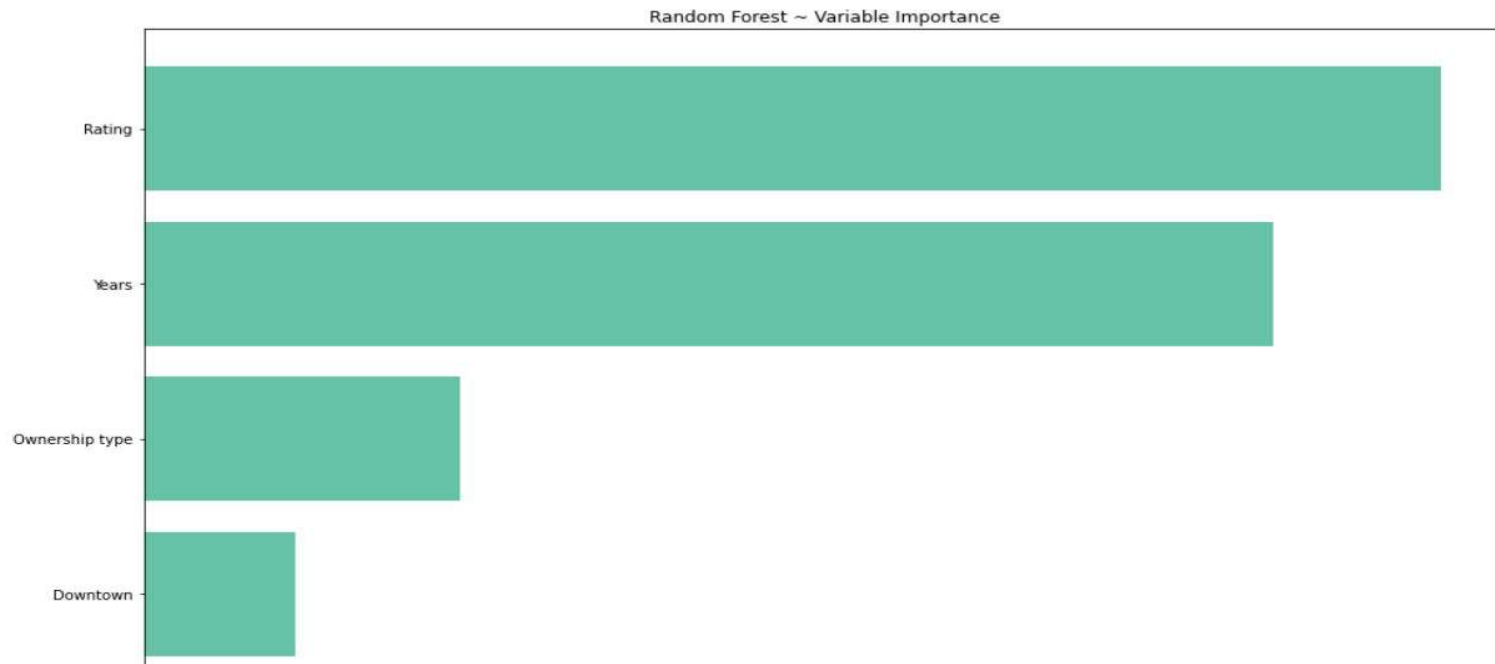- Restaurants without ratings were dropped from the analysis

EXPLORATORY DATA ANALYSIS

Most restaurant failures occur in first 10 years of operation, with a second wave around the 25-year mark.

.

CLASSIFICATION

- Numerous classifiers employed (KNN, SVM, etc.)

- No classifier outperformed the accuracy of simply predicting no business failure

- K Nearest Neighbors did, however, have the highest weighted F1 score

Random Forest ~ Variable Importance

FEATURE IMPORTANCE

Foursquare rating was top 1 or 2 feature for the 3 ensemble classifiers.

ROOM FOR IMPROVEMENT?

- Supplement missing Foursquare ratings with Yelp ratings?

- Fill in missing Foursquare ratings with an arbitrary low score, e.g., 0?

- Take steps to compensate for imbalanced dataset?