

PREDICTING VOLUME OF LIQUOR SALES ACROSS IOWA SUPERMARKET CHAIN

MONIKA HELAK

THE PROBLEM



There is a belief that higher unemployment rates drive higher liquor consumption.

Hy-Vee Supermarkets was interested in a model to predict volume of liquor sales across the supermarket chain in Iowa.

DATA ACQUISITION

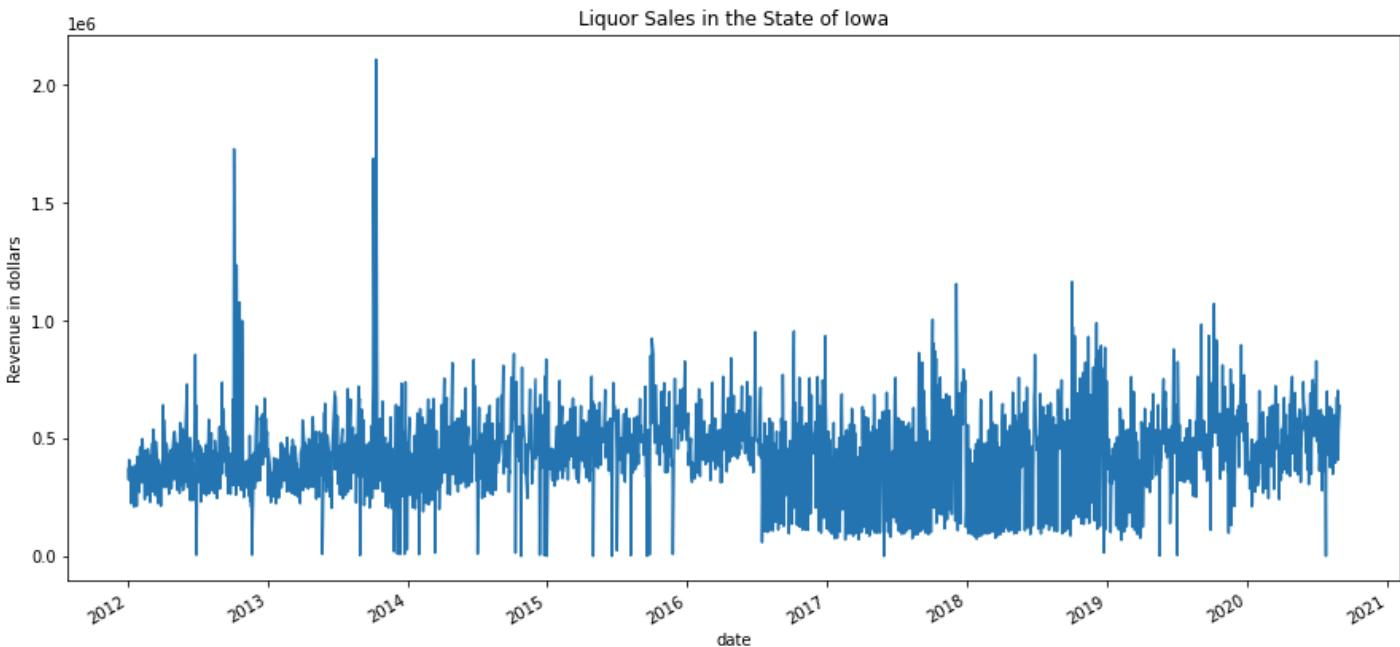
- Liquor Sales data was obtained from the Iowa data.gov website

It consisted of 24 features and ~20 million observations

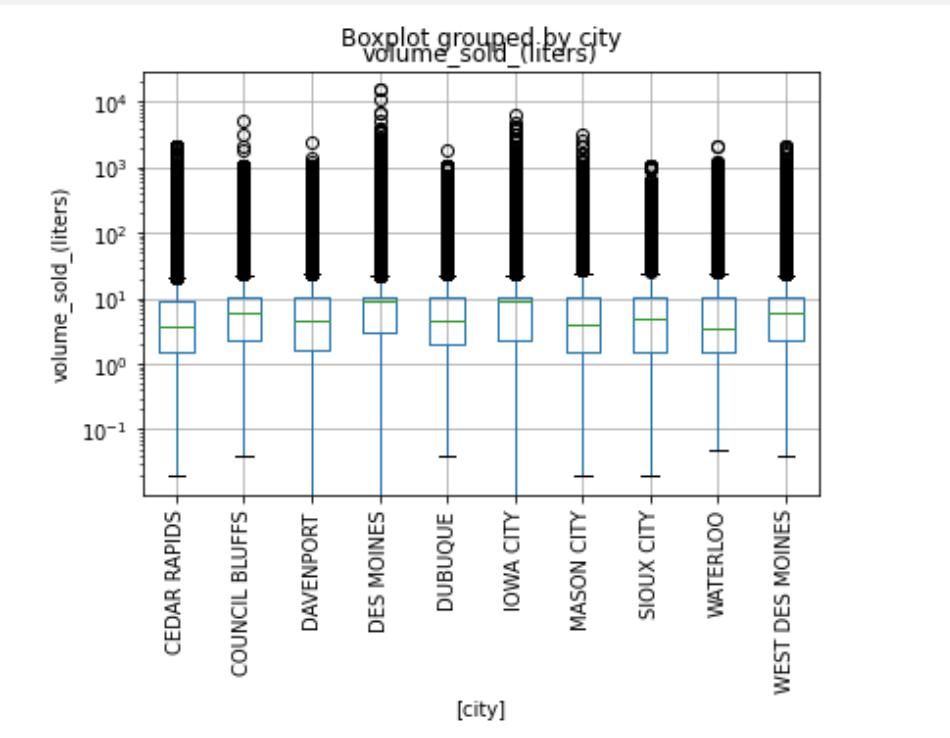
- Weekly unemployment data for the State of Iowa by county was obtained from the same source.

Merged with former dataset on county

EXPLORATORY DATA ANALYSIS

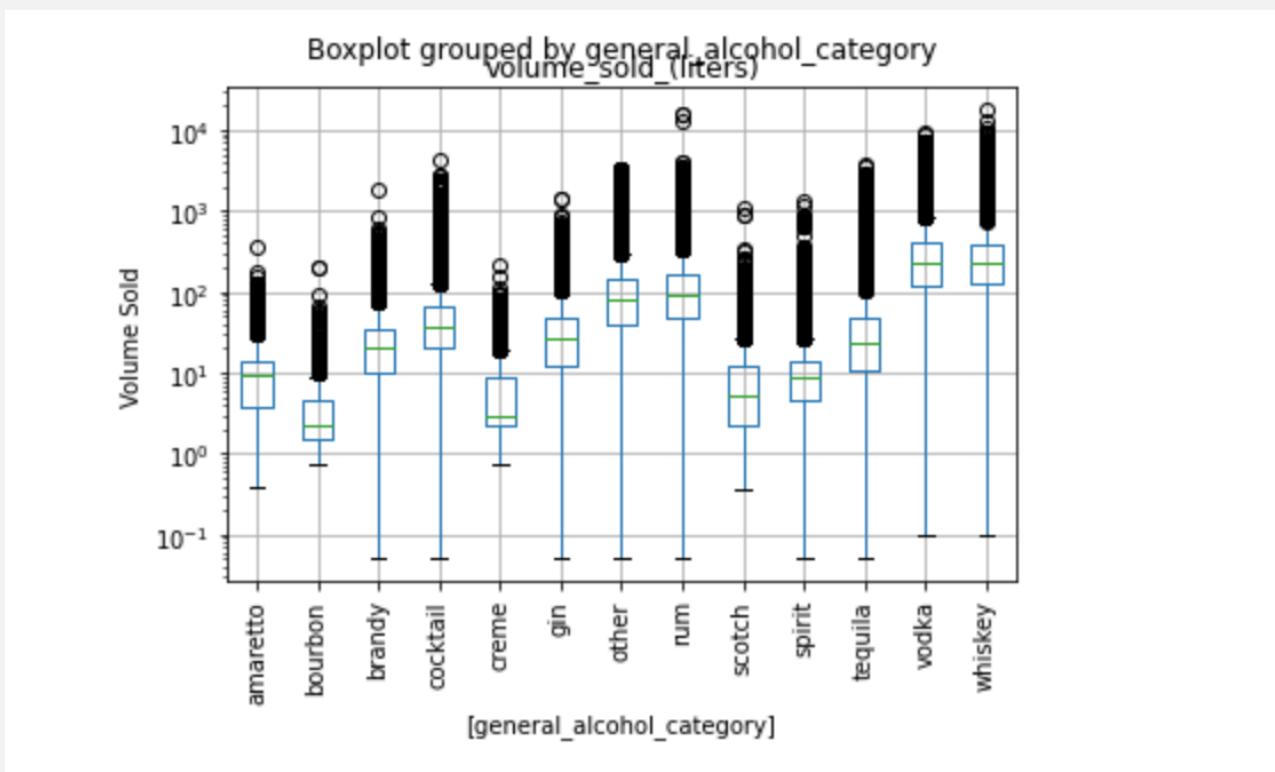


STATISTICAL TESTS



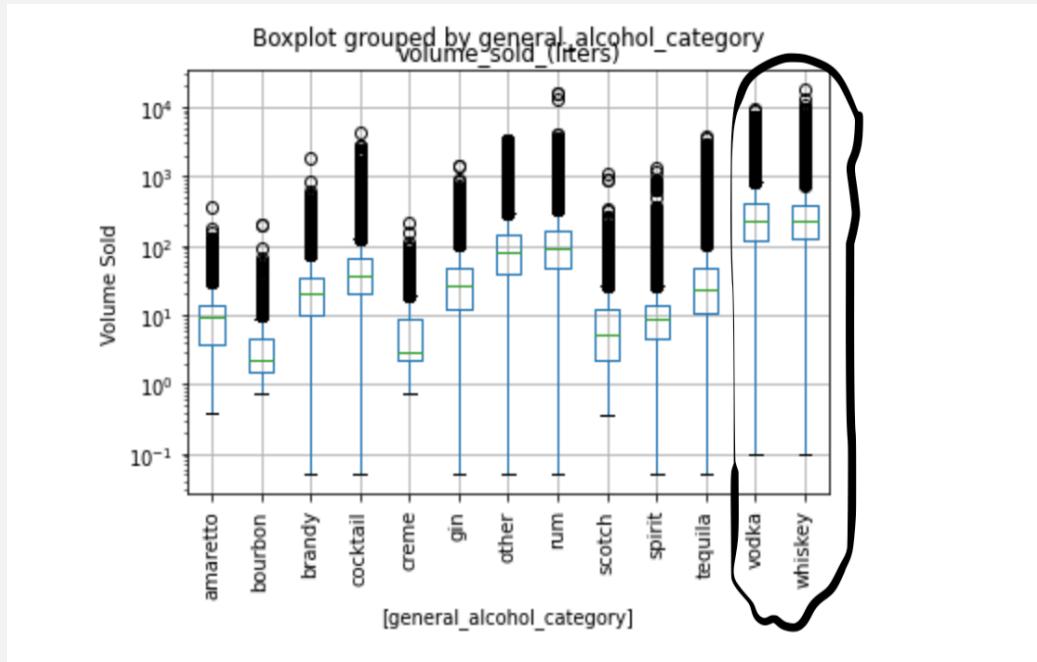
Distribution of liquor sales varied between county and city

ANOVA



Significant difference in volume of sales between liquor categories

UNPAIRED TWO SAMPLE T-TEST



Vodka vs.Whiskey
comparison

p-value < 0.05

Rejected the null
hypothesis that there
is no difference in
mean volumes sales
between the two
groups

FEATURE SELECTION



REMOVED CORRELATED
FEATURES WITH NO NEW
INFORMATION



FEATURE ENGINEERED FEWER
CATEGORIES FOR LIQUOR TYPES



LEFT WITH MANY CATEGORICAL
VARIABLES – CONVERTED TO
DUMMY VARIABLES

MODEL SELECTION

	Linear Regression	Decision Tree	Random Forest
Advantages	Simple implementation and interpretation	<ul style="list-style-type: none">• Very sensitive to noise• Poor out of sample prediction	<ul style="list-style-type: none">• High accuracy• Can handle many predictor variables
Disadvantages	<ul style="list-style-type: none">• Prone to underfitting• Sensitive to outliers	<ul style="list-style-type: none">• Prone to overfitting	<ul style="list-style-type: none">• Prone to overfitting
Performance (MAE)	102.79	124.98	69.48
Training Time	6.09 s	18.72 s	121.49 s

FUTURE RESEARCH



- Include more variables for better measure of socioeconomic condition of the state (measures beyond unemployment claims)
- Analyze which other products in the supermarket are sold at higher rates along with liquor



THANK YOU