THE FLATIRON SCHOOL: DATA SCIENCE FELLOWSHIP PROJECT

TENNESSEE FUEL QUALITY ANALYSIS

PREPARED BY / KATE HAYES & MISHA BERRIEN

JULY 31, 2019

OVERVIEW

- I. Our Question
- II. The Dataset
- III. Key Volatility Metrics
- IV. Time Series Analysis
- V. Logistic Regression Analysis
- VI. Conclusion & Next Steps

OUR QUESTION

Can we use data collected by fuel inspectors to better predict gas station fuel test failures in the state of Tennessee?





THE DATASET

Five year fuel quality inspection records for the state of Tennessee

- Source: State of Tennessee
 Department of Agriculture
- Time Period: Mid 2014 to early2019
- Number of Fuel Products: 11
- Number of Test Types: 72

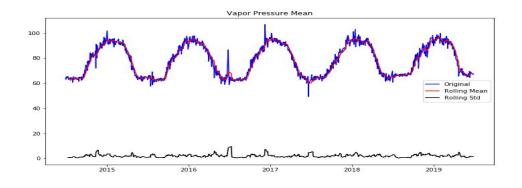
KEY VOLATILITY METRICS

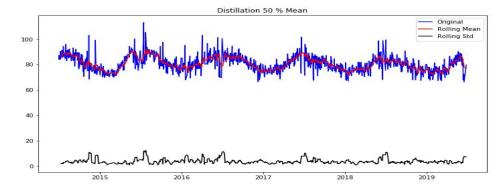
Distillation 50% - The temperature at which 50% of a sample is evaporated. Relates to the driveability and idling characteristics for the fuel.

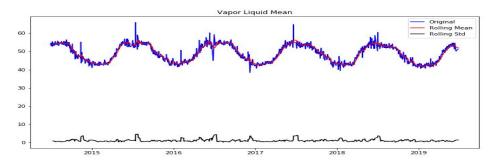
Vapor Pressure - is a measure of the amount of vapor that is produced by a gasoline sample at 37.8°C (100°F). Vapor pressure most affects an engine's ease of starting.

Vapor-liquid ratio is the ratio of the volume of vapor to the volume of liquid at atmospheric pressure. Used to evaluate a gasoline sample's tolerance to changes in temperature.

Dicky Fuller Test of The Mean



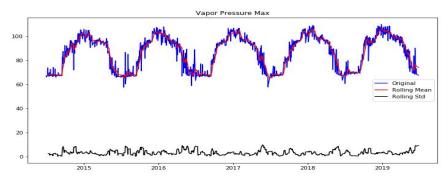


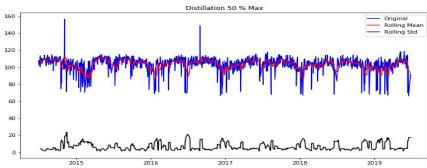


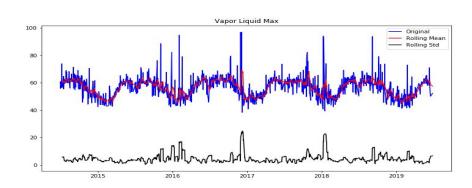
Dicky Fuller Test

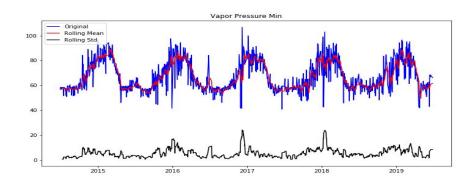
- Distillation 50% average is stationary above the 1% critical value
- Vapor Pressure and vapor liquid are stationary at all critical values

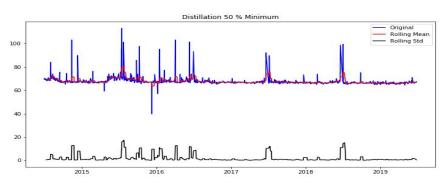
Dicky Fuller

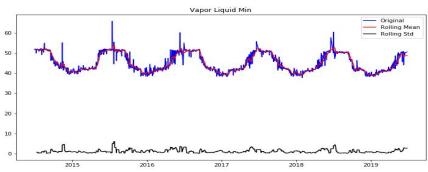




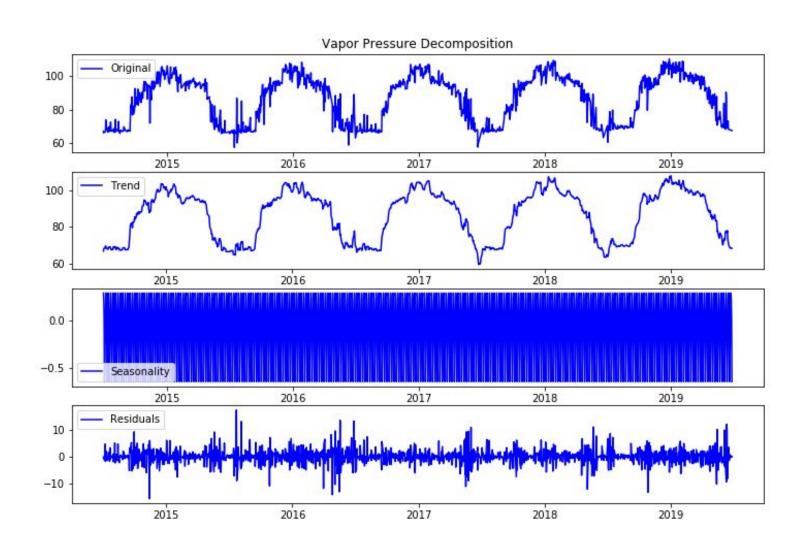








Seasonal Decomposition



out of order











OUT OF ORDER

MODEL SELECTION

Chosen Model:

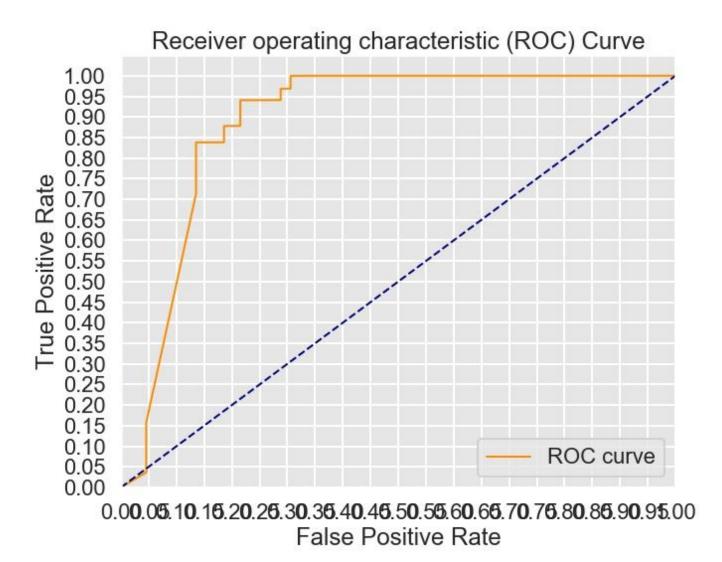
- Target Variable: Vapor Liquid-Ratio Test Outcome
- Predictor Variables: Tennessee Retailers Season & Grade

Model Accuracy:

	PRECISION	RECALL	ACCURACY	F1-SCORE
TRAIN	0.88755	0.74867	0.79480	0.81221
TEST	0.94850	0.76702	0.83020	0.84816

ROC CURVE

Vapor Liquid-Ratio Test Outcome ~ Tennessee Retailers Season & Grade



CONCLUSION & NEXT STEPS

- There is obvious seasonal trends due to the nature of the regulations.
- Further analysis is required in order to interpret the logistic regression.
- Do tests that pass but barely pass lead to more complaints?
- Do certain regions of Tennessee have more failures in certain seasons
- Do certain brands of gas stations or distributors deliver more failures or have more complaints against them?

THANK YOU!

QUESTIONS?