Analytics Methodology

1. Data Collection

2. Data Transformation/Exploration
   - Time Trend Analysis
   - Normality Analysis
   - Variable Correlations

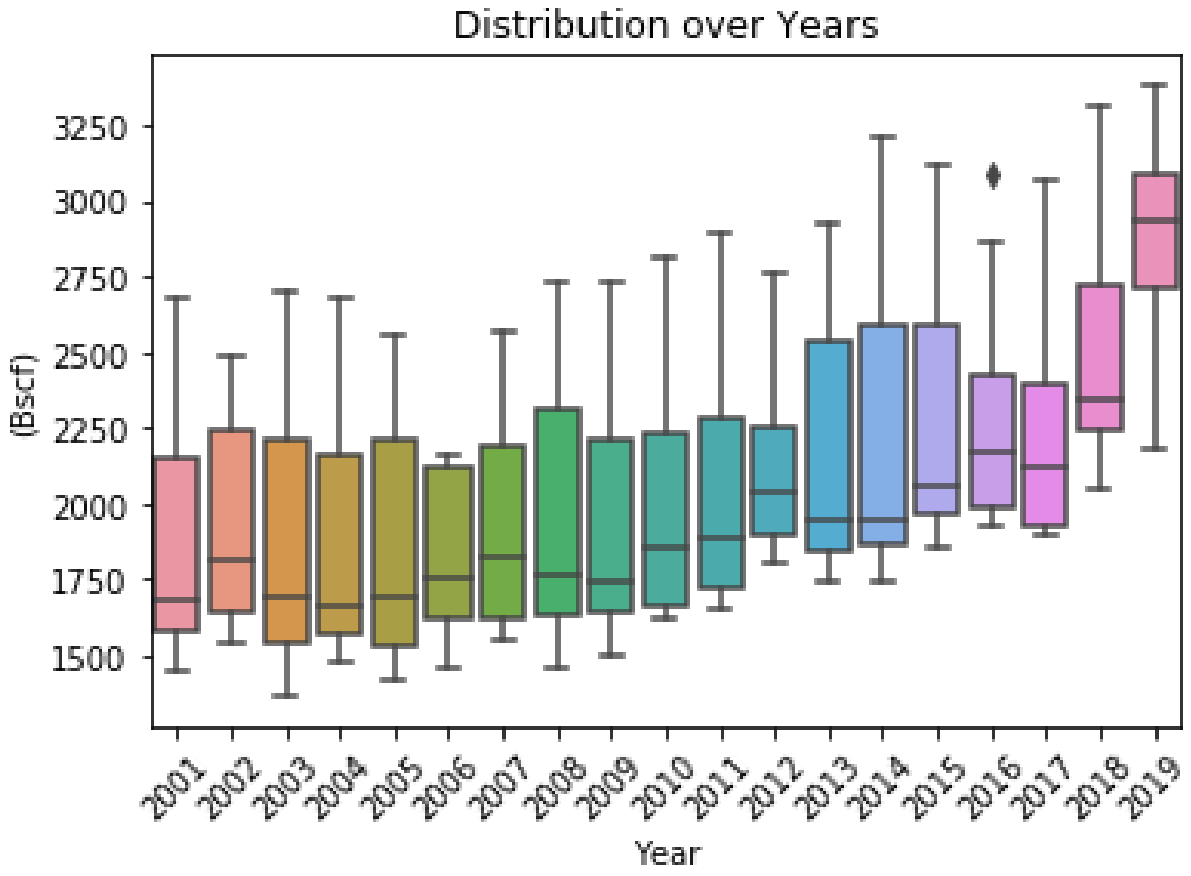
3. Model building – Econometric Model
   - ARIMA
   - SARIMAX

4. Model building – Machine Learning Model
   - Regression
   - Random Forest

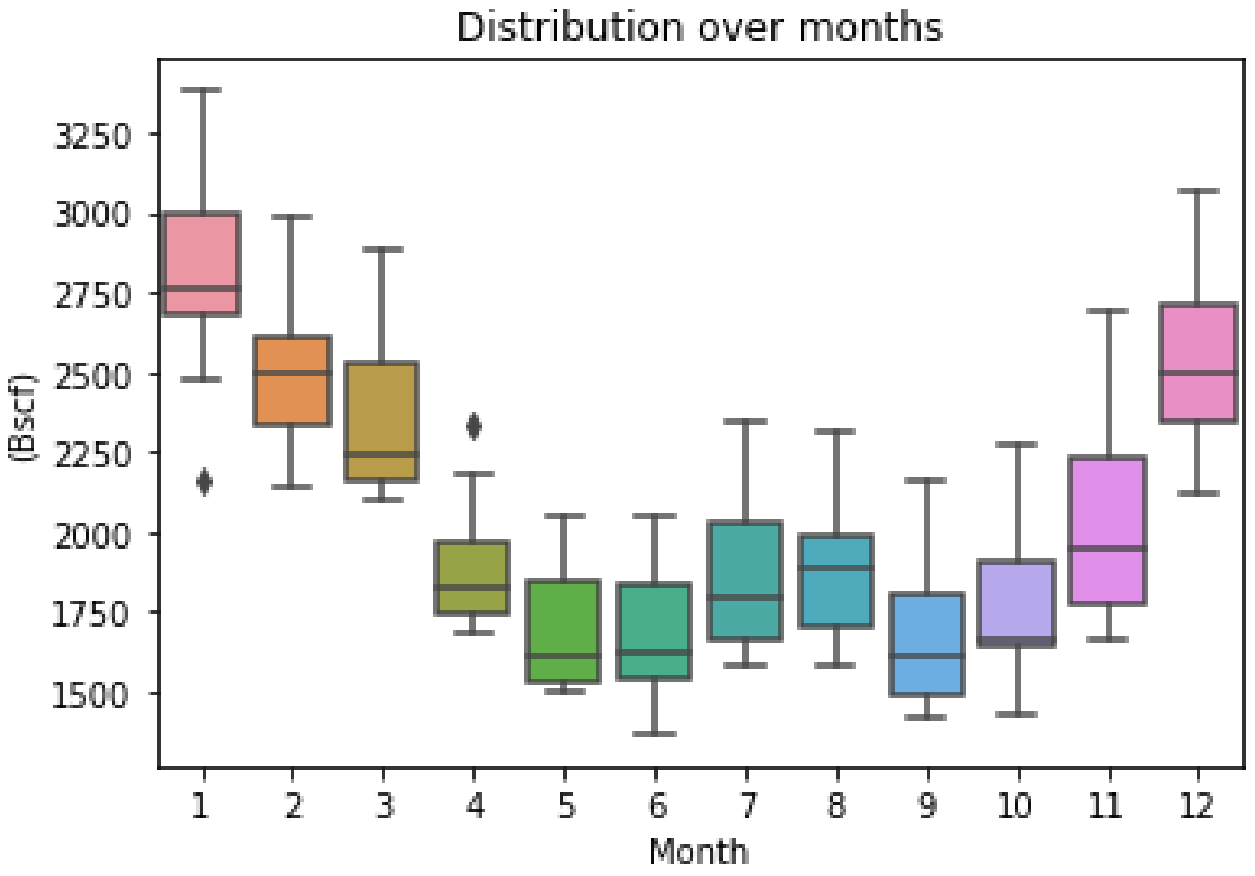
Data Collection

- Natural Gas Consumption and Prices
  - Source https://www.eia.gov/
  - Method: API
  - Range: daily value for 15 years from 01/01/2001 – 01/07/2019
- S&P Information:
  - Source: https://finance.yahoo.com
  - Method: API call using module fix_yahoo_finance
  - Range: daily value for 15 years from 01/01/2001 – 01/07/2019.
  - Data include: Ticker, Date, Close, and Volume.
- Economics Stats:
  - Source: Federal Reserve Economic Data https://fred.stlouisfed.org/
  - Method: direct download.
  - Range: quarterly value for 20 years from 01/03/2004 – 01/07/2019.
  - Data include: CPI, GDP, GDP_Change, and Household income

Data Exploration – Time Trend Analysis

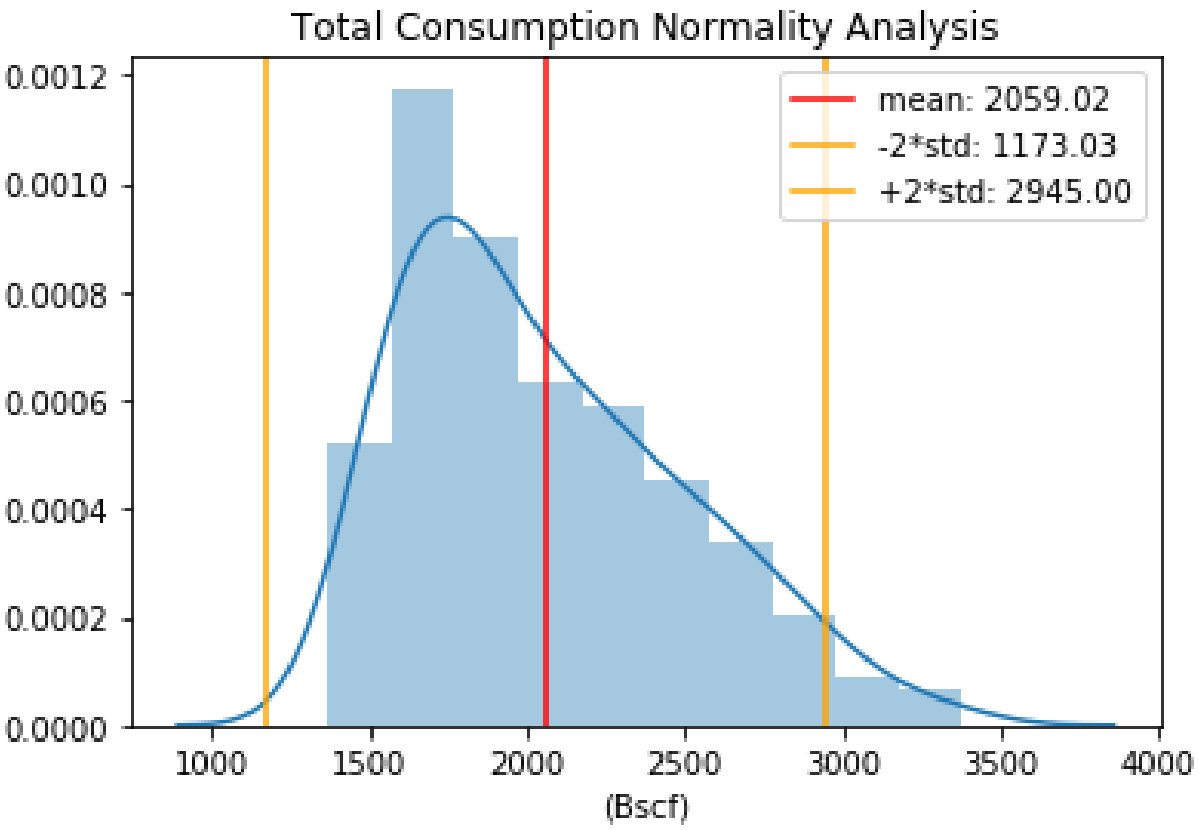


Consumption has been increasing over the years



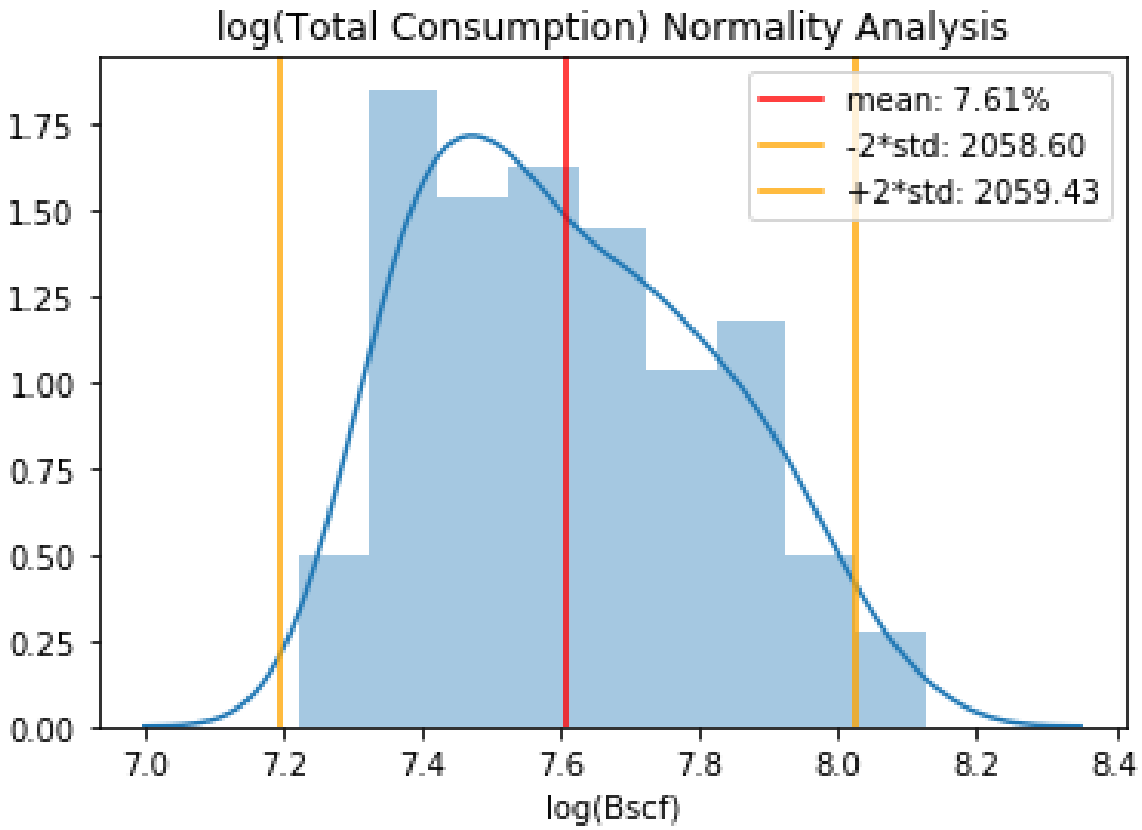
Drastic intra-year consumption variations

Data Exploration – Normality Analysis



Normality Test: Shapiro Test

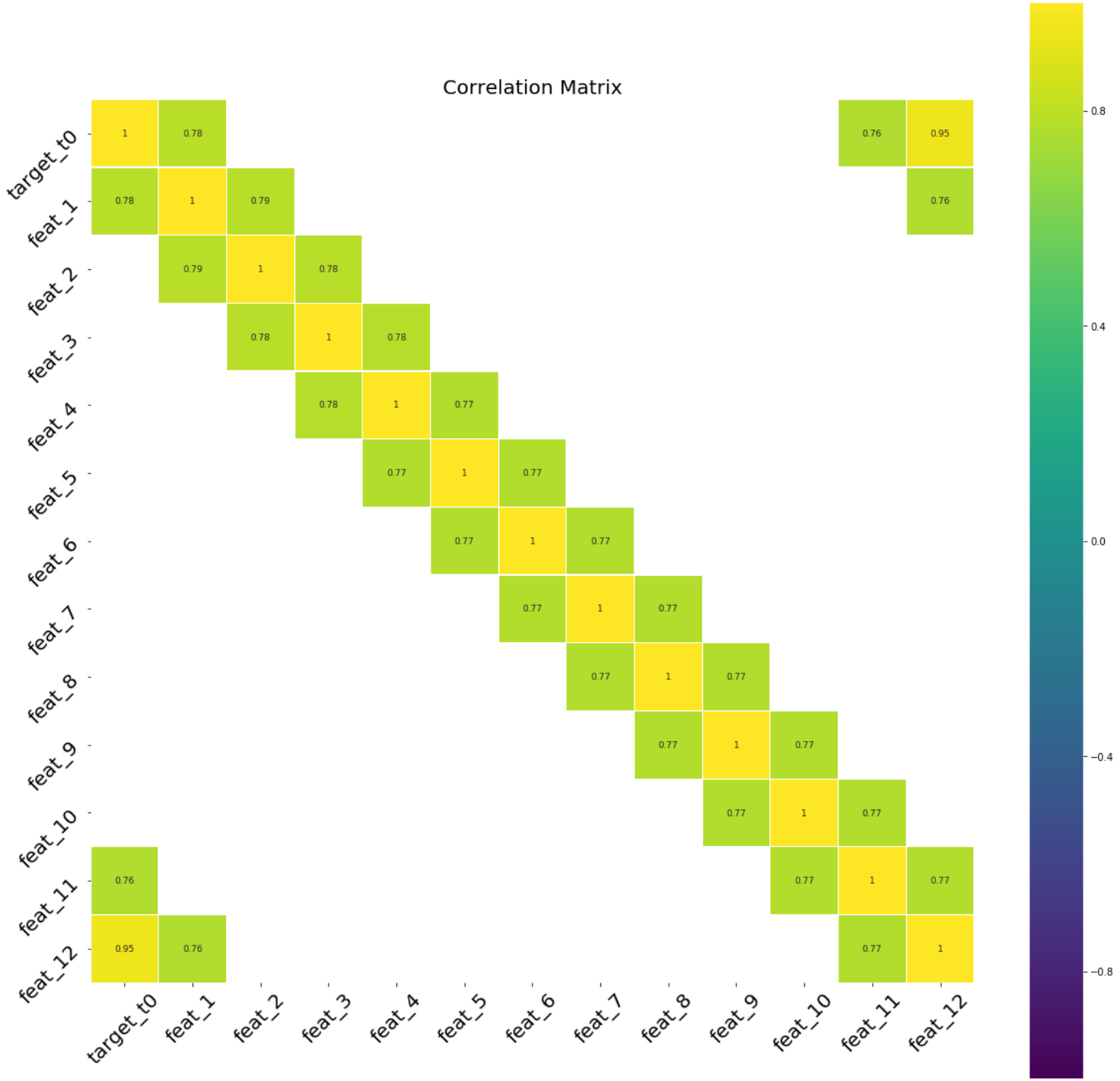
$p = 1.2 \times 10^{-7}$



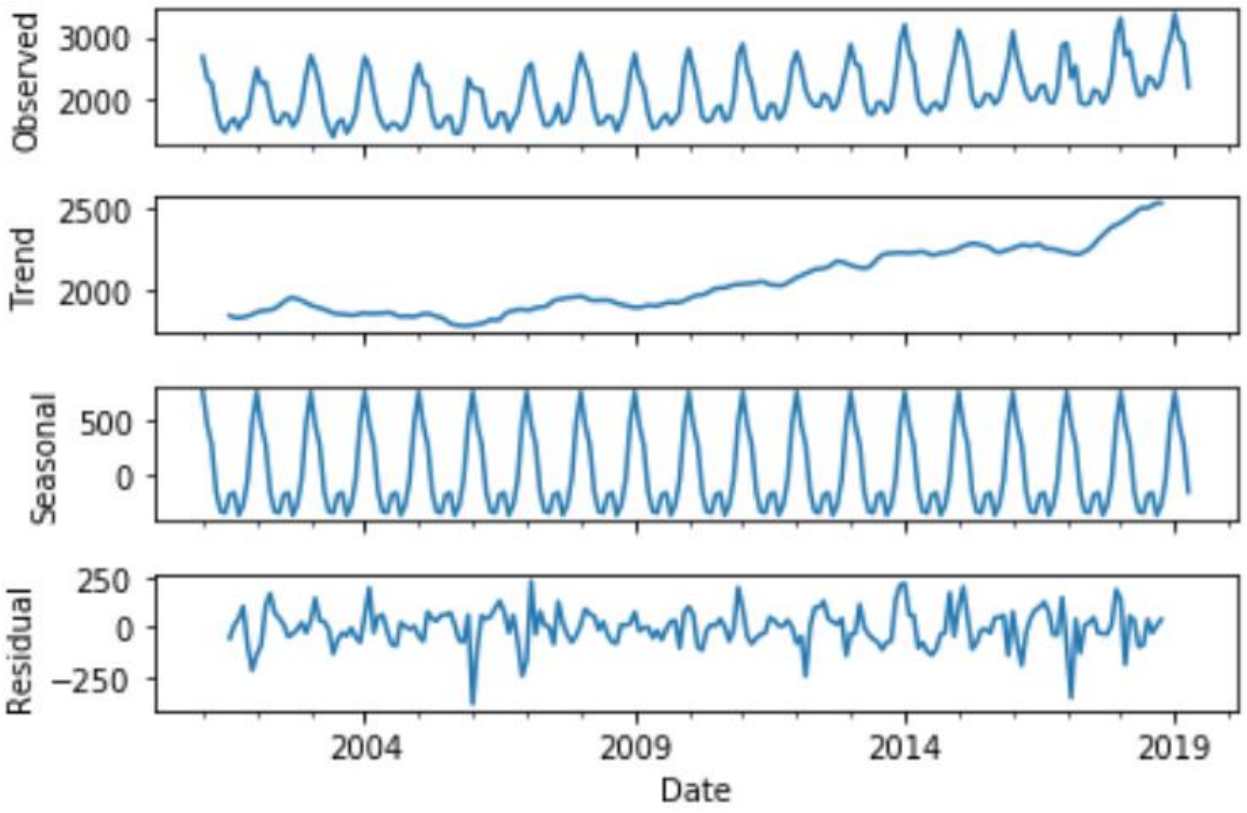
Normality Test: Shapiro Test

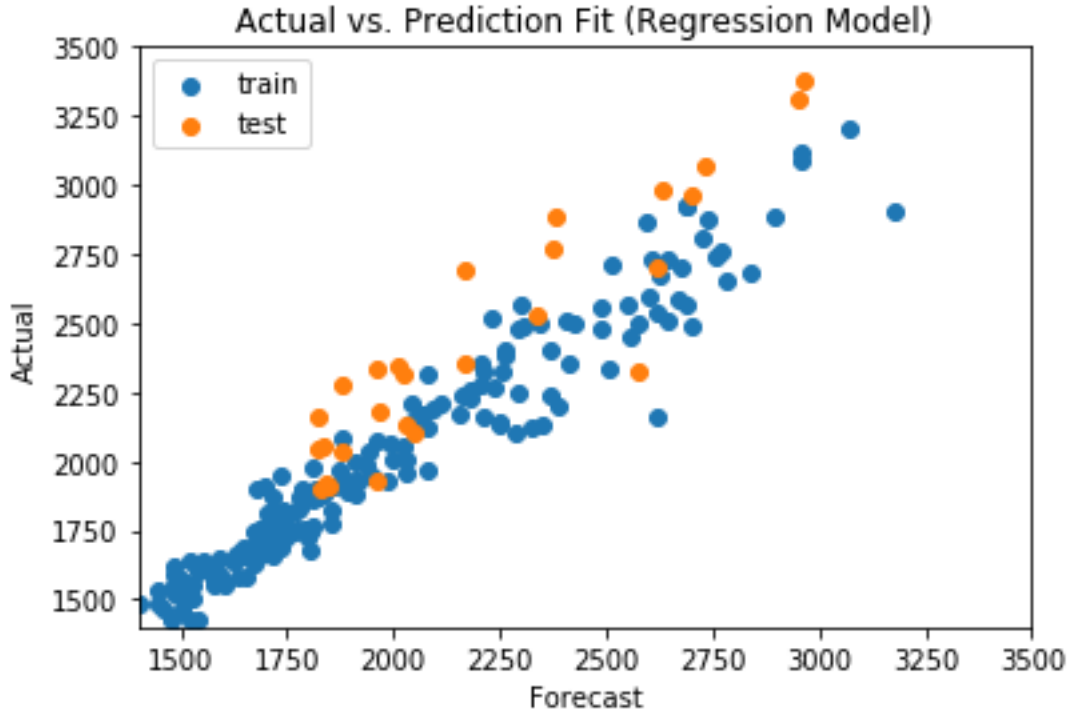
$p = 9.4 \times 10^{-5}$

Data Exploration – Correlation Matrix

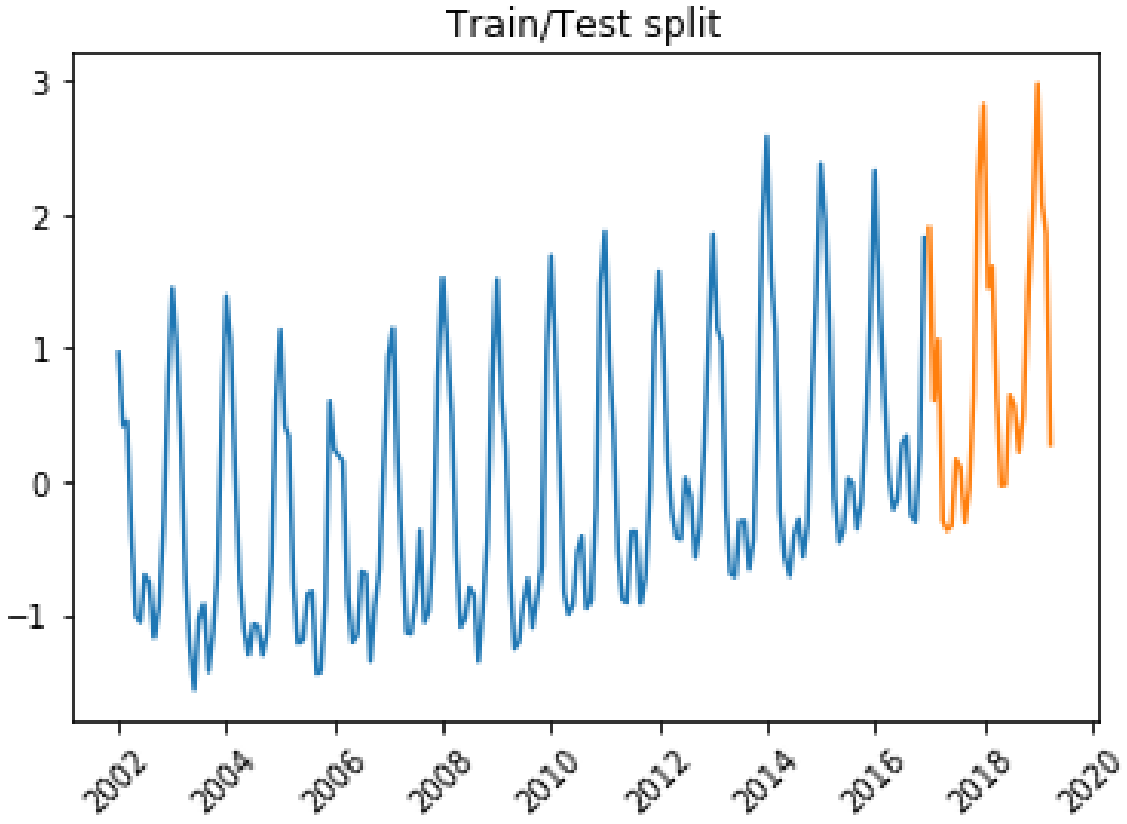


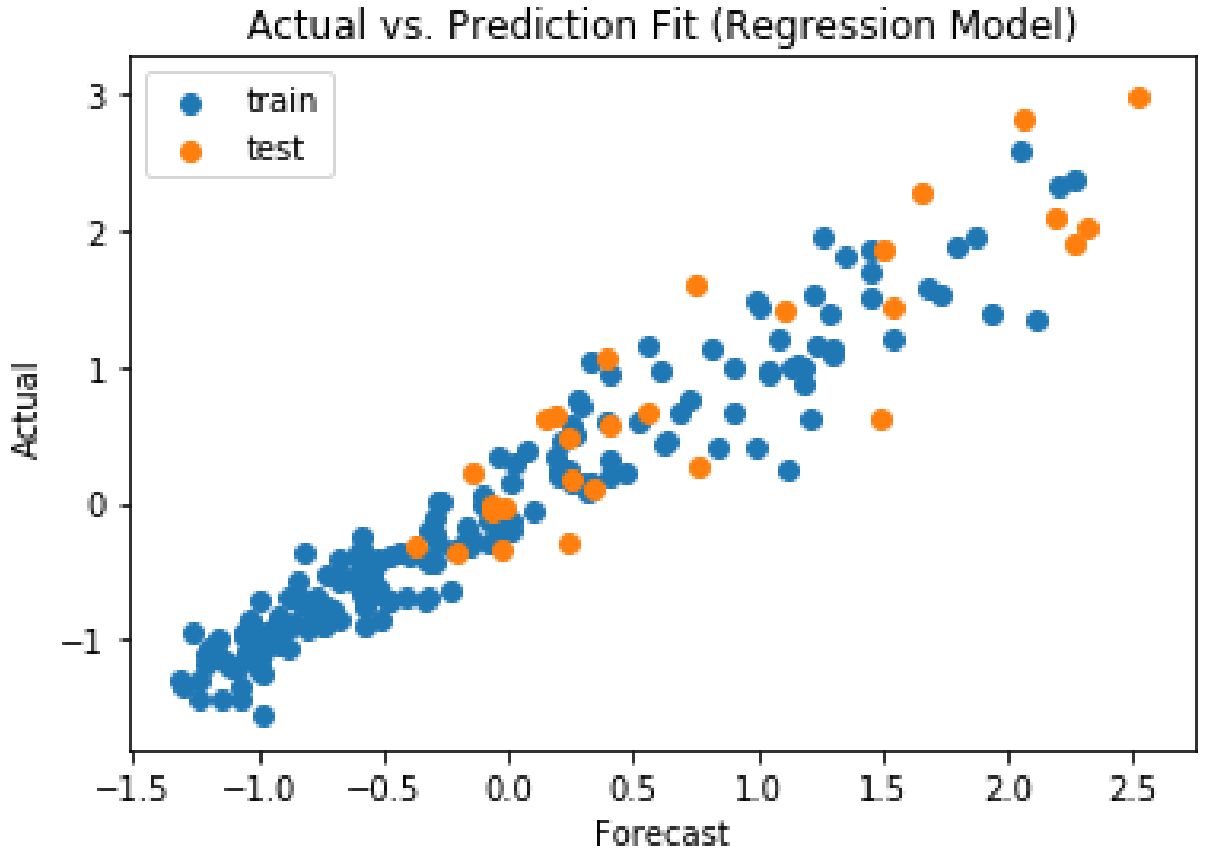
Econometric Model – ARIMA





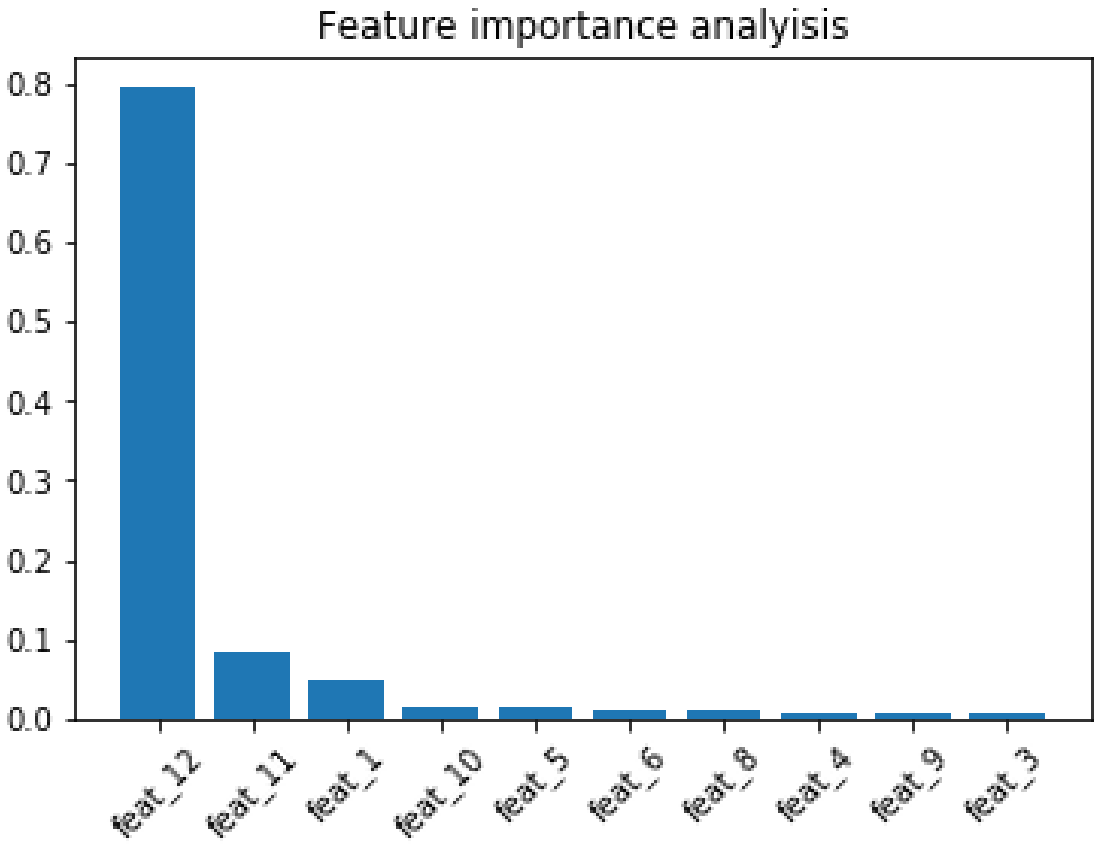
Machine Learning Model – Linear Regression

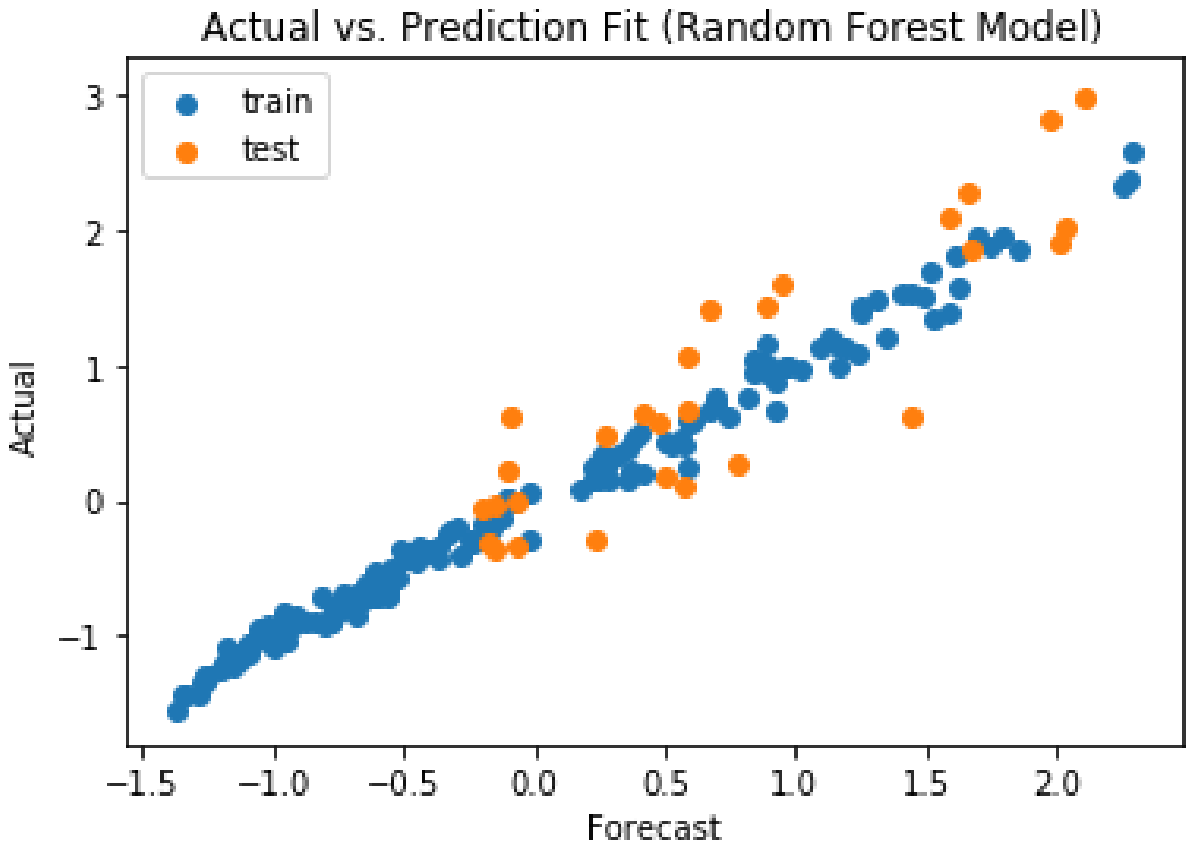




Train RMSE: 0.25
Test RMSE: 0.42
Train R2: 0.93
Test R2: 0.81

Machine Learning Model – Random Forest





Train RMSE: 0.10
Test RMSE: 0.47
Train R2: 0.99
Test R2: 0.77