Process Overview

06 April 2024 05:29

1) Objective:

Forecasting number of antidiabetic drug prescriptions

2) checking for through visualizations

'Trend' ✓ & 'Seasonality' ✓

3) checking for stationarity

ADF statistic → should be less
& p-value < 0.05

ours is not stationary

4) So we apply transformations to make it stationary

fixed { i) 1st diff → again not stationary
ii) Seasonal diff → m=12

Now it is stationary ✓

& d=1, D=1, m=12

& $d=1$, $D=1$, $m=12$

5) Model selection & parameters till now

SARIMA $\longrightarrow$ $(p,1,q)$ $\overbrace{(P,1,Q)}^{\text{Seasonal}}$ $m=12$

(No. exogenous variables)

6) Optimizing the model

to get the best possible combination of parameters through 'AIC' criterion

Minimum AIC $\longrightarrow$ $p=2, q=3, P=1, Q=3$

$\therefore$ Final Model

SARIMA $(2,1,3)(1,1,3)12$

7) Residuals Analysis

✓ 1) Variance : Constant Variance

(No Heteroskedasticity problem)

(No Heteroskedasticity problem)

✓ ii) Distribution: Normal (Histogram + Density plot)

✓ iii) Normal Q-Q plot: Straight line

✓ iv) Correlogram:

No significant co-efficients after lag 0
just like white-noise

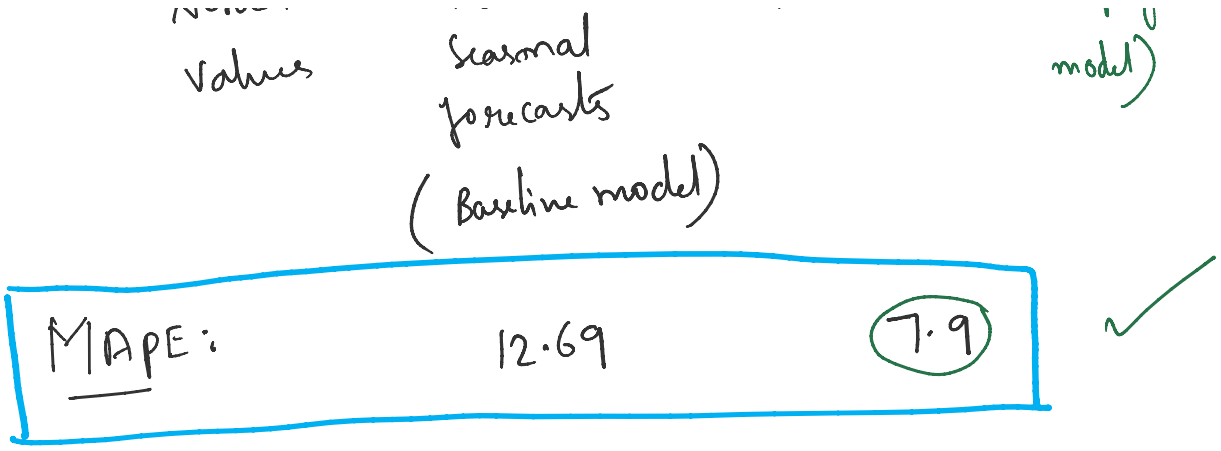
So, with this analysis, residuals seem to closely resemble white noise

8) Ljung-Box test:

to check whether the residuals → Independent (or) correlated ??

9) Forecasting & Comparision

- Actual values
- naive seasonal
- SARIMA ✓ (Best performing model)



Since the SARIMA model achieves the lowest MAPE, we can conclude that the SARIMA(2,1,3)(1,1,3)12 model should be used to forecast the monthly number of antidiabetic drug prescriptions.