Time Series Forecasting and Hypothesis Testing on Electricity Demand in United States

Sayak Kumar Dutta
Praxis Tech School
Roll: A23036

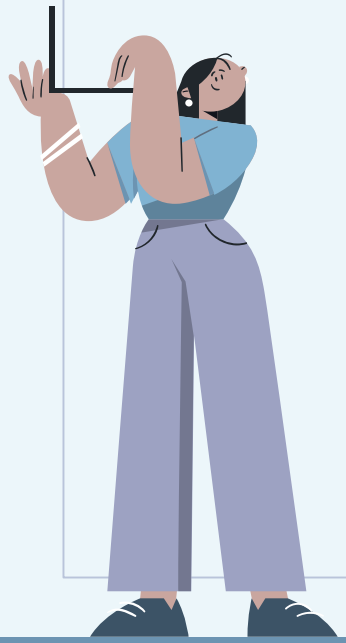
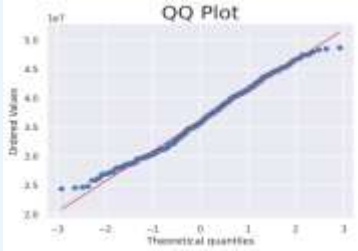


Table of contents

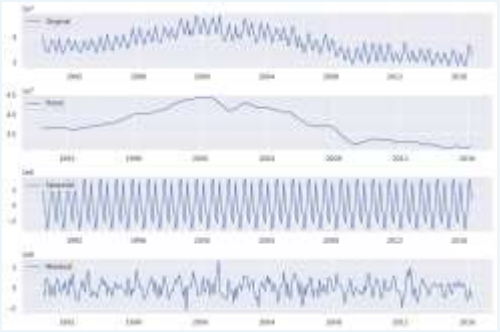
Objective of the Project

Our objective is to examine the data and hypotheses for precise future electricity demand predictions in the USA. Utilizing time series forecasting to optimize power distribution and strategize the implementation of new power sources.



- **AIM:** To explore and predict month-wise electricity demand in the USA.
- **EDA Insights**:

1. The dataset is normally distributed from QQ plot
2. The dataset has seasonality



- **Evidence Validation:** ADF Hypothesis Testing & ACF plots validated the results

Data Source:https://www.eia.gov/electricity/data/state/

Hypothesis Testing

➢ **T-Test:**

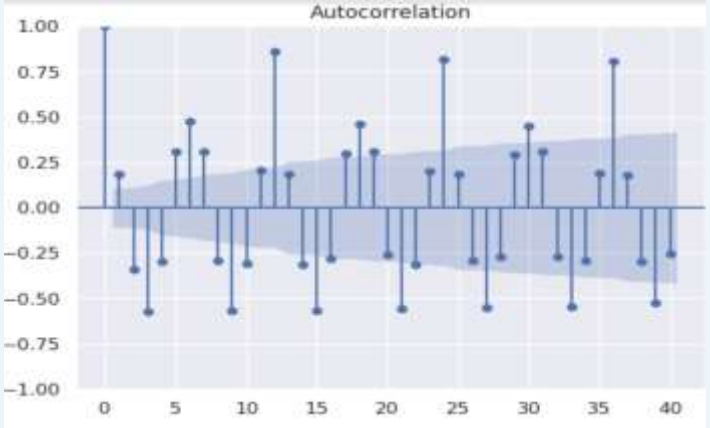
Null Hypothesis (H0): There is no significant difference in the mean 'Unit Generated' between two groups

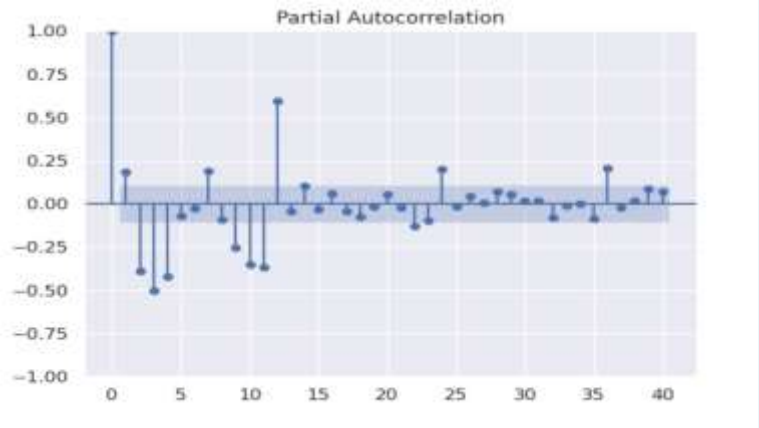
Alternative Hypothesis (H1): There is a significant difference in the mean 'Unit Generated' between the groups.

- Identified significant differences in electricity demand levels between two groups Total Power Unit generated in USA (from 1999 to 2009) and (from 2010 to 2020)
  Output: T-statistic: 13.211182757799744, p-value: 1.9541992558764684e-29
  We reject Null Hypothesis

- Identified significant differences in mean electricity demand levels between two groups i.e Total Power Unit generated in USA (from 1999 to 2009) and (from 2010 to 2020) in two states and doing it for all states
  Output: All p value <0.05
  So, We don't reject Null Hypothesis

Time Series Forecasting





After Differencing ACF & PACF Plot

ADF TEST

•Null Hypothesis (H0): The time series has a unit root, indicating it is non-stationary.

•Alternative Hypothesis (H1): The time series does not have a unit root, indicating it is stationary.

Since significant so we do not find evidence of stationarity in dataset. p=0.69

Forecasting Graph

Interpretation

1. Model Specification:
    1. Model: SARIMAX(3, 1, 3) - This indicates the SARIMA model order. SARIMA(3,1,3) means you have a Seasonal Autoregressive Integrated Moving Average model with order (p=3, d=1, q=3).

2. Coefficients and Significance:
    1. ar.L1, ar.L2, ar.L3: Autoregressive coefficients at lags 1, 2, and 3. Interpretation depends on context. Here, ar.L1 is positive but not statistically significant, ar.L2 is negative and not significant, and ar.L3 is negative and highly significant.
    2. ma.L1, ma.L2, ma.L3: Moving average coefficients at lags 1, 2, and 3. Negative ma.L1 and ma.L3 are statistically significant.
    3. sigma2: Estimated variance of the residuals. It seems quite large, possibly indicating volatility in the residuals.

3. Statistical Tests:
    1. Ljung-Box (Q) Statistic: Tests whether there is serial correlation in the residuals. A high p-value (0.74) indicates that there is no significant autocorrelation up to lag 1.
    2. Jarque-Bera (JB) Statistic: Tests whether the residuals are normally distributed. A high p-value (0.60) suggests that the residuals are approximately normally distributed.
    3. Heteroskedasticity (H) Statistic: Tests whether the residuals exhibit changing variance over time. A p-value of 0.78 suggests no evidence of heteroskedasticity.

SARIMAX Results

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	321
Model:	SARIMAX(3, 1, 3)	**Log Likelihood**	-5054.790
Date:	Sun, 28 Jan 2024	**AIC**	10123.580
Time:	21:52:53	**BIC**	10149.958
Sample:	01-31-1990 - 09-30-2016	**HQIC**	10134.113
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.1008	0.091	1.109	0.268	-0.077	0.279
ar.L2	-0.0378	0.103	-0.367	0.714	-0.240	0.164
ar.L3	-0.8233	0.093	-8.869	0.000	-1.005	-0.641
ma.L1	-0.2154	0.100	-2.145	0.032	-0.412	-0.019
ma.L2	-0.0461	0.120	-0.384	0.701	-0.282	0.189
ma.L3	0.7475	0.105	7.116	0.000	0.542	0.953
sigma2	2.858e+12	7.02e-14	4.07e+25	0.000	2.86e+12	2.86e+12

Ljung-Box (L1) (Q):	0.11	**Jarque-Bera (JB):**	1.02
Prob(Q):	0.74	**Prob(JB):**	0.60
Heteroskedasticity (H):	0.95	**Skew:**	-0.13
Prob(H) (two-sided):	0.78	**Kurtosis:**	2.93

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.58e+41. Standard errors may be unstable.

Conclusion

The project's findings offer crucial insights into forecasting the energy demand of USA. We find the data fits the SARIMA model best because of seasonality. With these we can predict future values of energy demand thus optimizing total energy production.

Thanks!