Electrical Power Prediction For the Next 20 Years

using Time Series Modeling

By: Nick Catanglan

Agenda:

1. Problem Overview
2. Data Wrangling
3. Exploratory Data Analysis
4. Model Selection
5. Conclusion

PROBLEM OVERVIEW:

What is U.S. electricity generation by energy source?

- In 2018, about 4,178 billion kilowatthours (kWh) were generated at utility-scale electricity generation facilities in the United States
- About 63% of this electricity generation was from fossil fuels (coal, natural gas, petroleum, and other gases)
- 20% was from nuclear energy
- 17% was from renewable energy sources(Solar, Wind, Hydro).

Uncertainties that affects energy system of the future

- Growing electrification to the expansion of renewables.
- Upheavals in oil production and globalisation of natural gas markets.
- Policy choices made by governments about fuels and renewable energy.
- Planned policies that can meet long-term climate goals under the Paris Agreement, reduce air pollution, and ensure universal energy access.

How does this project help?

- Power Plants, Electrical Companies, Government Agencies and Small scale Power Producers can utilize this model to predict and monitor power demand and supply values in real time to be able to lay out plans in advance to meet the said Electrical Power supply and demand.
- The model will help companies to decide how much they can produce in daily, monthly and yearly basis.

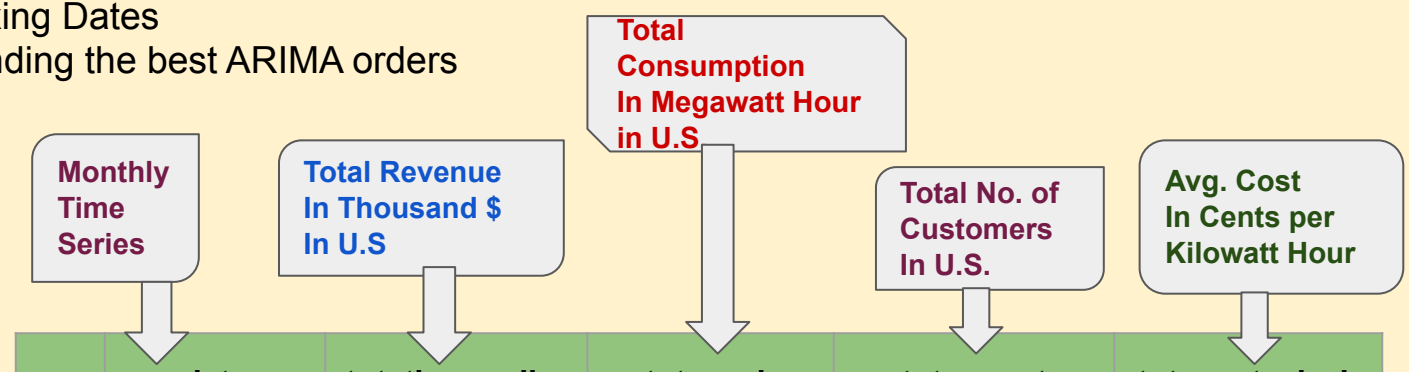
About the dataset :

- I got the data from the US Energy Information Administration(EIA). URL “https://www.eia.gov/electricity/data.php”
- The data is the monthly revenue, energy consumption in mwh, number of customers and price per kwh from 1990 to present.
- Residential, Commercial and Industrial is collected separately from all the US State and join them together in one data reference to time intervals.
- From Original Data shape(18108, 28), After joining datas, cleaning, grouping, fixing the date and be able reduce datashape (151, 18).
- I choose the final data, U.S. overall total of Revenue, Consumption in MWH, Number of customer , Cost per kwh in accordance to Date collected.

- Our Final Time Series Data will 5 columns and 151 rows (4 numeric, 1 Datetime)

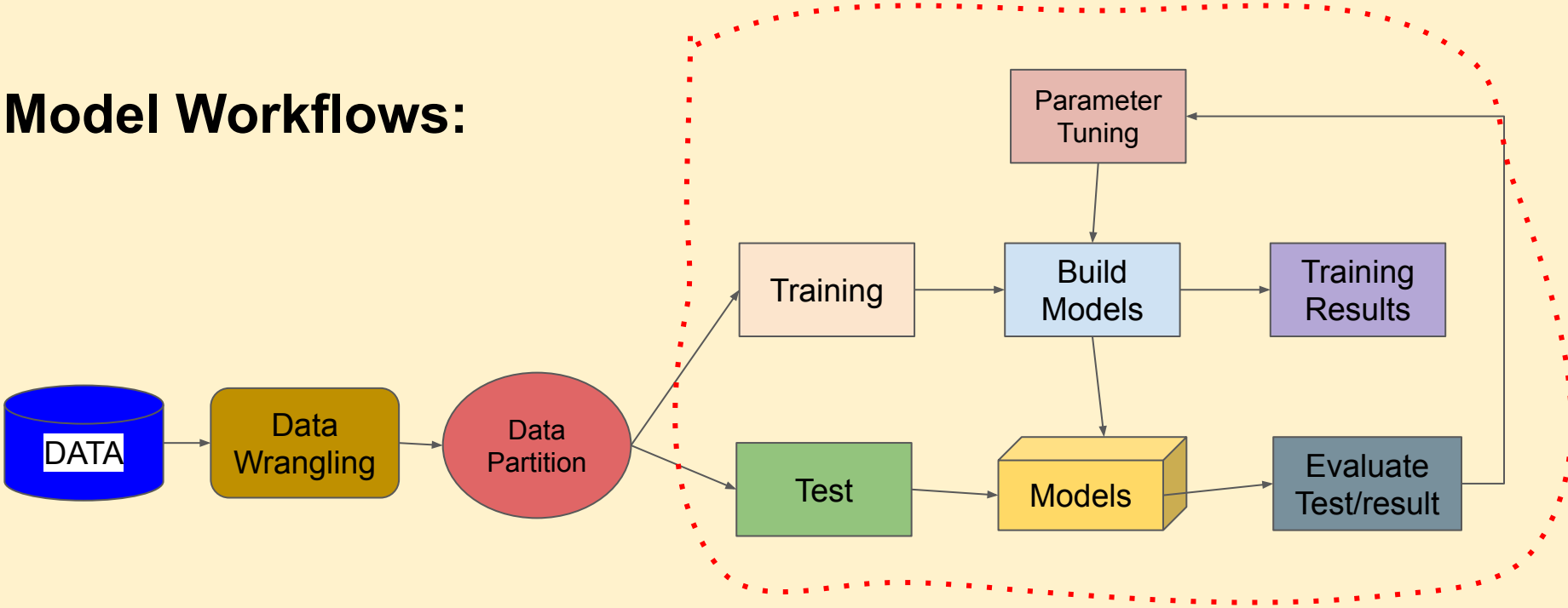
Challenges:

- Joining all the 3 dataset and put them together in accordance to the date it was collected.
- Cleaning the data
- Fixing Dates
- Finding the best ARIMA orders



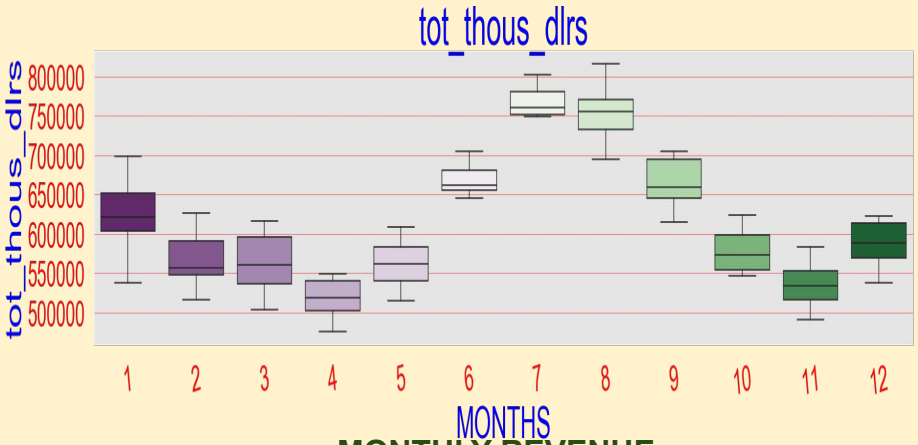
	date	tot_thous_dlrs	tot_mwh	tot_count	tot_cents_kwh
0	2007-01-01	538191.078431	6.178505e+06	2.301115e+06	8.800588
1	2007-02-01	516131.941176	5.908457e+06	2.285210e+06	8.866471
2	2007-03-01	503264.274510	5.718830e+06	2.307898e+06	8.893137
3	2007-04-01	476401.588235	5.398833e+06	2.297670e+06	8.883137
4	2007-05-01	514545.196078	5.741182e+06	2.309057e+06	8.996863

Model Workflows:

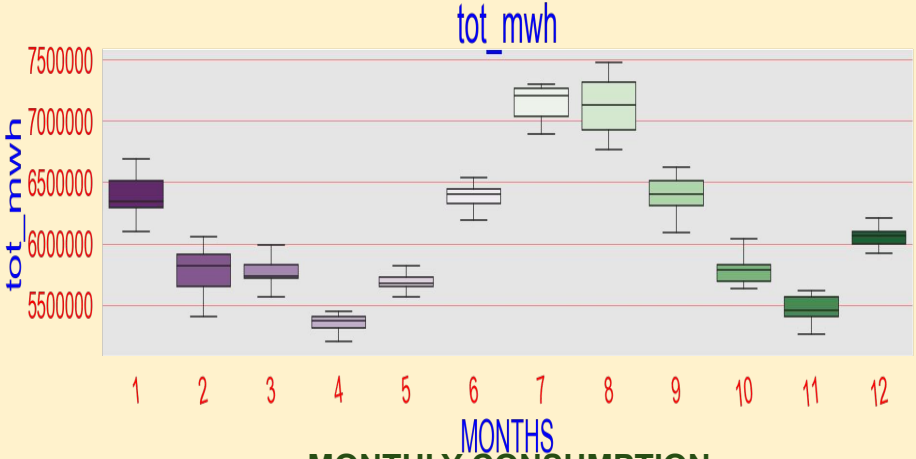


Model Training & Tuning

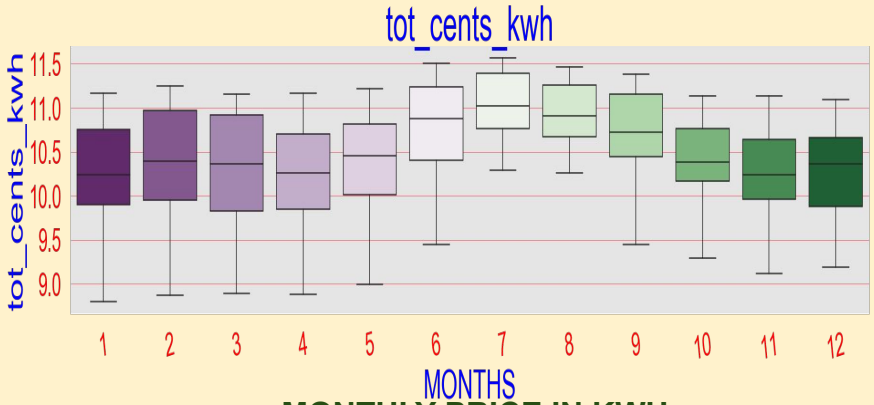
EXPLORATION AND ANALYSIS:



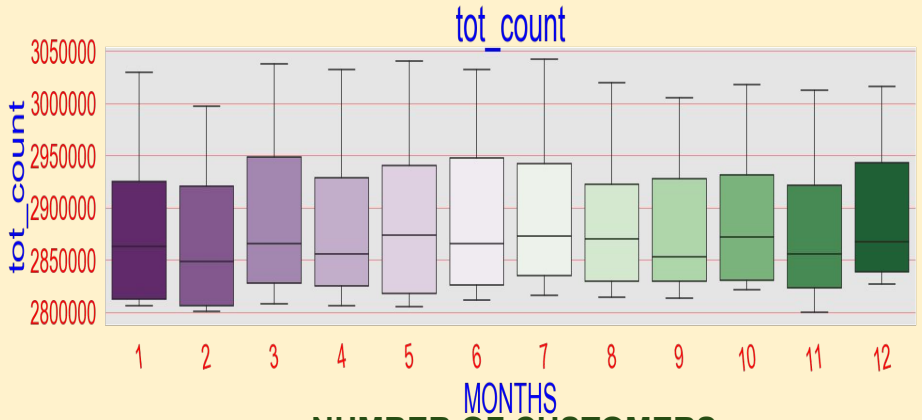
MONTHLY REVENUE



MONTHLY CONSUMPTION



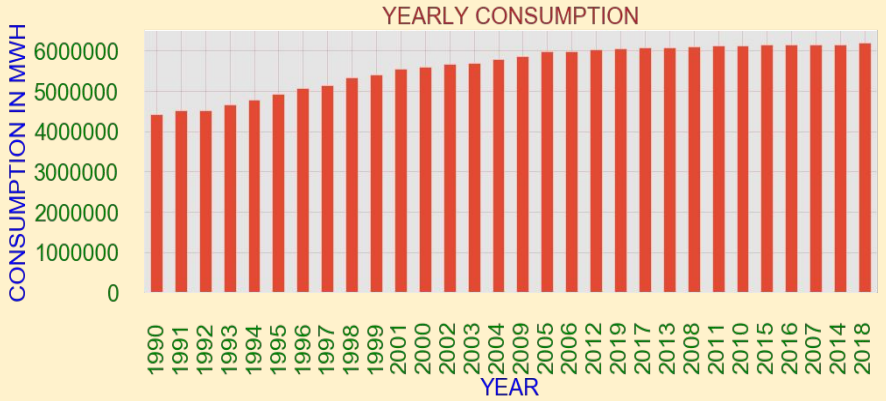
MONTHLY PRICE IN KWH



NUMBER OF CUSTOMERS

YEARLY BARPLOT

2018, 2014, 2015 has highest Revenue

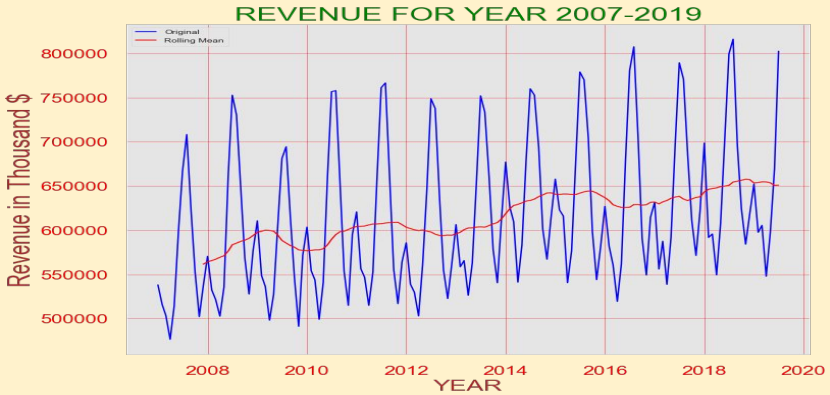


2018, 2014, 2007 has highest Consumption

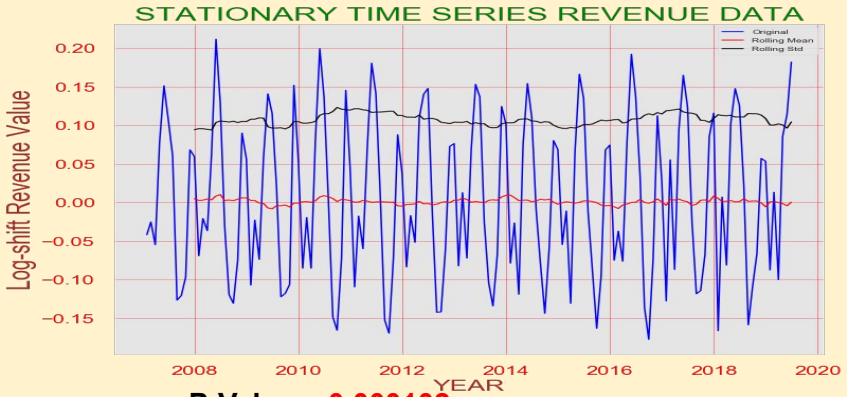
2019, 2018, 2017 has highest Customers

2019, 2018, 2017 has highest Cost

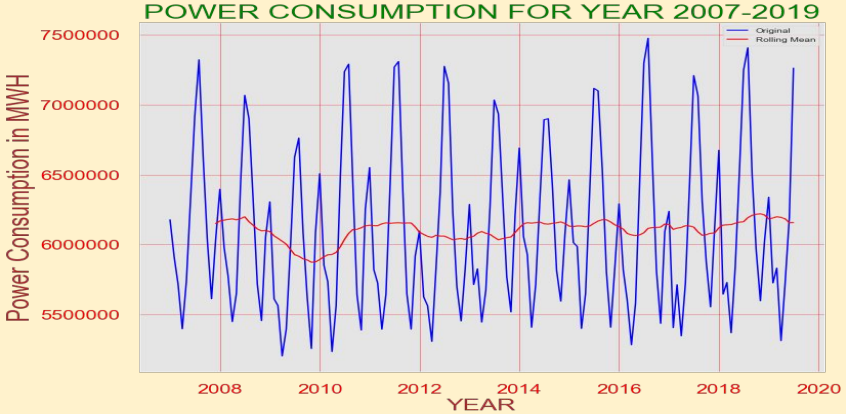
CHECK STATIONARITY:



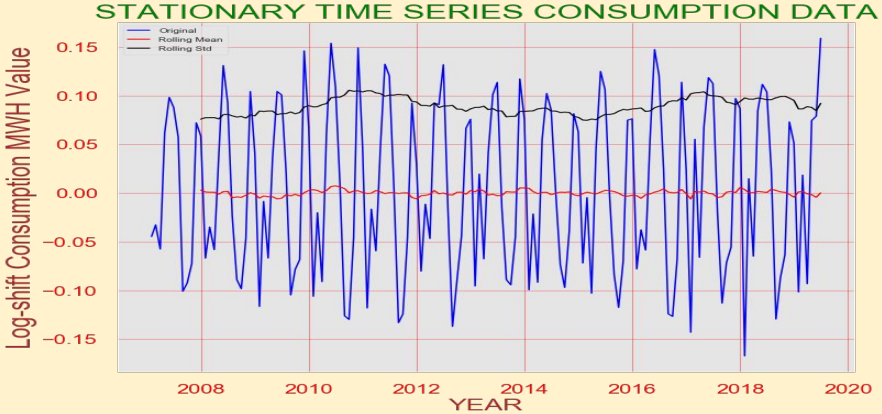
P-Value = 0.55
Test Statistic (-1.5)> Critical Val.(-3.5)



P-Value = 0.000102
Test Statistic (-4.7) < Critical Val.(-3.5)



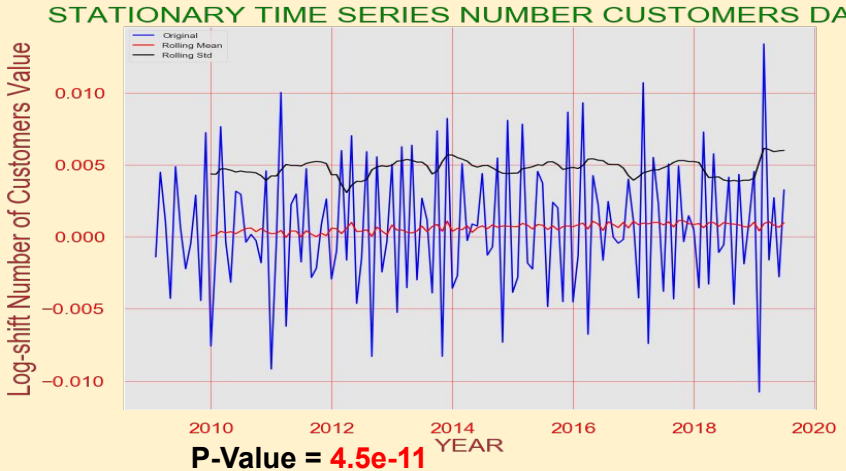
P-Value = 0.071
Test Statistic (-2.719974) > Critical Val (-3.478648)



P-Value = 0.000002
Test Statistic (-5.5) < Critical Val (-3.5)

P-Value = 1.0

Test Statistic (5.8) > Critical Val (-3.5)

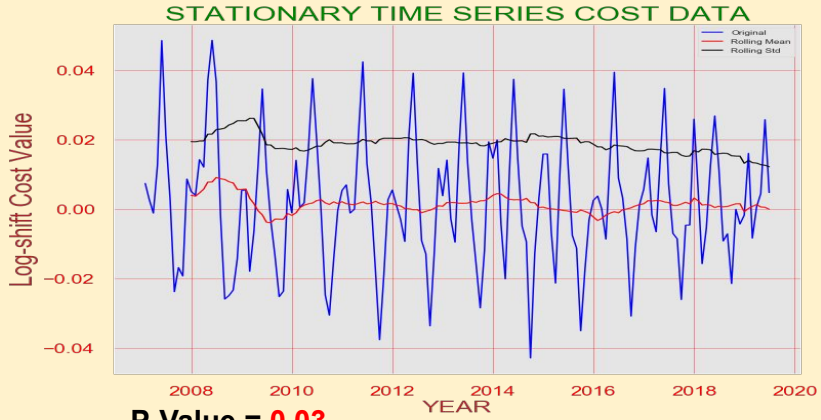


P-Value = 4.5e-11

Test Statistic (-7.5) < Critical Val (-3.5)

P-Value = 0.39

Test Statistic (-1.7) > Critical Val (-3.5)



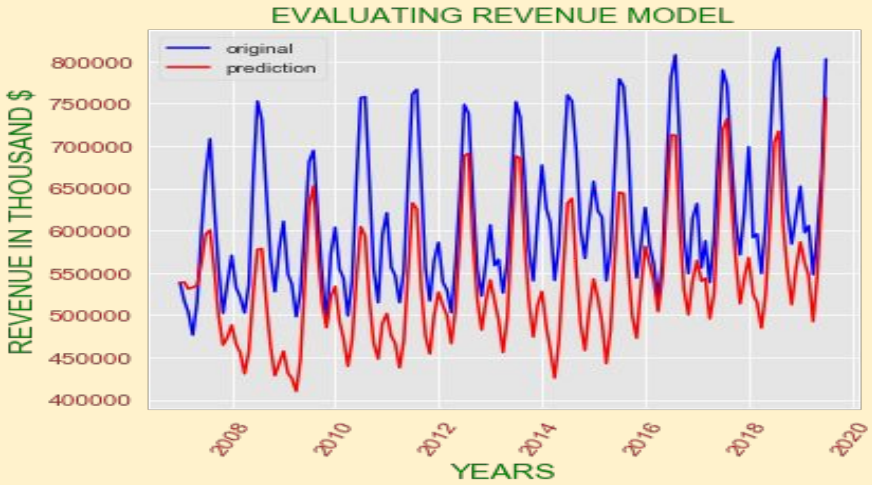
P-Value = 0.03

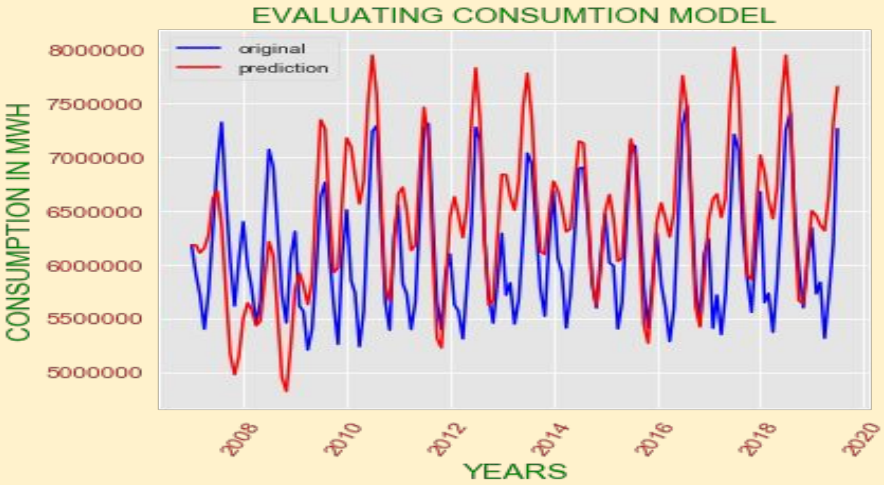
Test Statistic (-3.9) < Critical Val (-3.5)

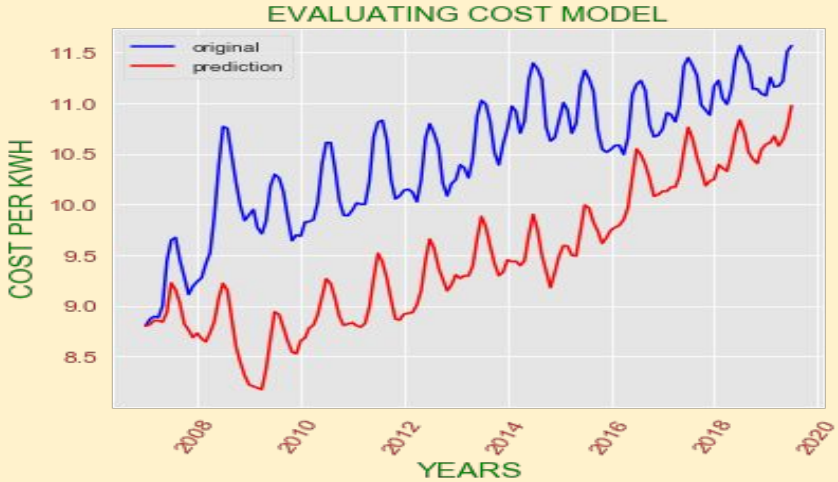
MODEL SELECTION:

Model Order Rev Time series	RSS Error	AIC	Model Order Mwh Time Series	RSS Error	AIC	Model Orders Count Times Series	RSS Error	AIC	Model Order Cost TIme Series	RSS Error	AIC
AR (8,1,0)	0.38	-450	AR (2,1,0)	0.63	-343	AR (2,1,0)	0.001	-1107	AR (8,1,0)	0.0194	-848
MA (0,1,6)	0.58	-388	MA (0,1,4)	0.58	-395	MA (0,1,2)	er	er	MA (0,1,3)	0.0316	-833
ARIMA (8,1,6)	0.18	-559	ARIMA (2,1,4)	0.30	-456	ARIMA (2,1,2)	0.001	-1109	ARIMA (8,1,3)	0.0146	-910

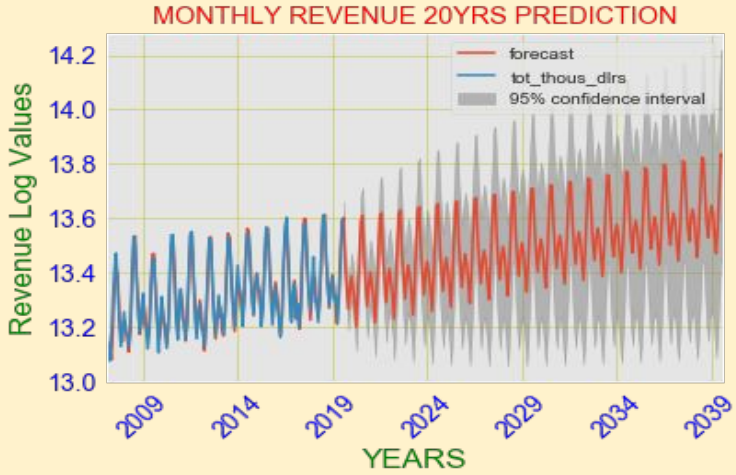
MODEL EVALUATION:

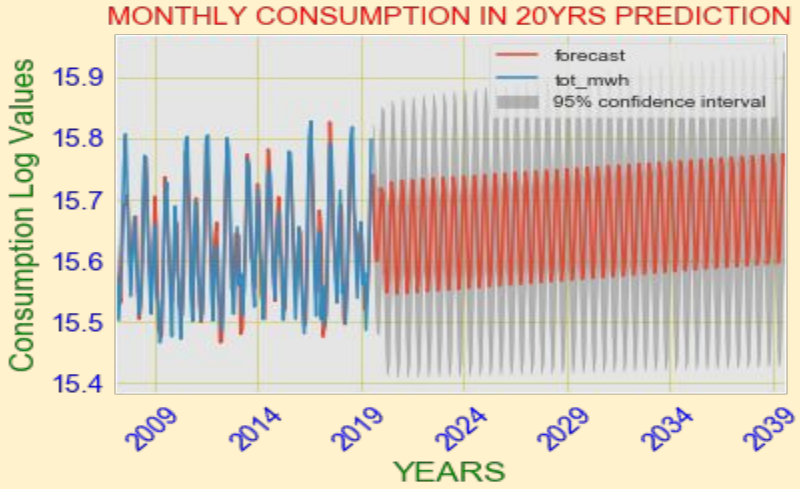






6243956.77 5815382.09 5928201.45





20 yrs Monthly Prediction

Date yr/mnt/day	Revenue Prediction Values
2019-09-01	$690,514,000
2019-10-01	$641,091,000
2019-11-01	$711,162,000
TO	TO
2039-06-01	$974,148,000
2039-07-01	$101,1897,000
2039-08-01	$983,044,000

1st 3 mnts prediction

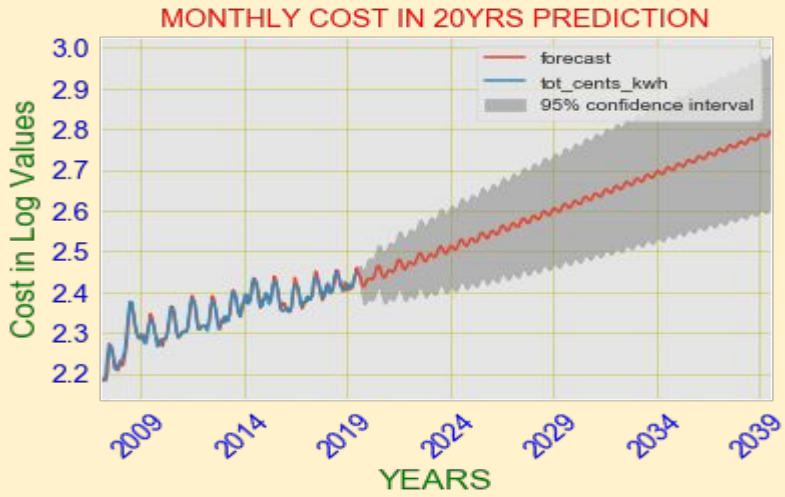
Last 3 mnts prediction

20 yrs Monthly Prediction

Date yr/mnt/day	Consumption Prediction Values(mwh)
2019-09-01	6,243,956.77
2019-10-01	5,815,382.09
2019-11-01	5,928,201.45
TO	TO
2039-06-01	6,748,108.42
2039-07-01	7,295,790.32
2039-08-01	7,238,184.50

1st 3 mnts prediction

Last 3 mnts prediction



20 yrs Monthly Prediction

Date yr/mnt/day	No. of Customers Prediction Values
2019-09-01	3,350,392
2019-10-01	3,352,535
2019-11-01	3,354,680
TO	TO
2039-06-01	3,839,318
2039-07-01	3,841,774
2039-08-01	3,844,232

1st 3 mnts prediction

Last 3 mnts prediction

20 yrs Monthly Prediction

Date yr/mnt/day	Cost Prediction Values(Cents)
2019-09-01	17.62
2019-10-01	17.64
2019-11-01	17.66
TO	TO
2039-06-01	20.47
2039-07-01	20.5
2039-08-01	20.53

1st 3 mnts prediction

Last 3 mnts prediction

CONCLUSION:

- We were able to make a model that learn from the past data and predict the monthly values for the next 20 years for Revenue, Demand/Consumption in MWH, Number of Customers and Cost in KWH with a small range of 95% confidence intervals.
- All except Consumption in MWH increase with time while the consumption has a little increase for the demand for the next 20 years.
- Knowing the trend, seasonality & values for the next 20 yrs will help the power producers companies to make decisions and plan in advance to meet the Electrical power supply and demand while in accordance with the planned policies of government and the upheavals in oil production and globalisation of natural gas market.

Before Implementing:

- Review with the engineering department for further study before deployment.
- Our inputs will be the dates intervals, monthly interval for monthly prediction, weekly intervals for weekly prediction and so on.
- We can also base in this model to be able to predict per State Revenue, Consumption in MWH, Number of Customers and Cost in KWH and per sector(Residential, Commercial, Industrial)

How can we maintain our models after implementing:

- Monitor - constant monitoring of the model is needed to determine the performance accuracy of the model.
- Evaluate - evaluating the metrics of the current model is calculated to determine if new algorithm is needed.
- Compare - the new model is compared against each other to determine which model perform the best.
- Rebuild - the best performing model is rebuilt on current state of the data.

THANK YOU!!!!!!

Any Question?