Sales Forecast

Pushkar Chougule

Content

Business Problem and Background

- Global Superstore
- Product offerings : Furniture, Office Supplies and Technology
- Customer Base : Consumers, Corporates and Home Offices
- Generate Insights from Sales data
- Sales Forecast to help maintain adequate staffing through year, popular products' inventories, supply chains for high sales regions / states and make delivery arrangements

Data Explanation and Preparation

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
1	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600
2	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400
3	CA-2017-138688	12/6/2017	16/06/2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036.0	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200

Features with NULL Values

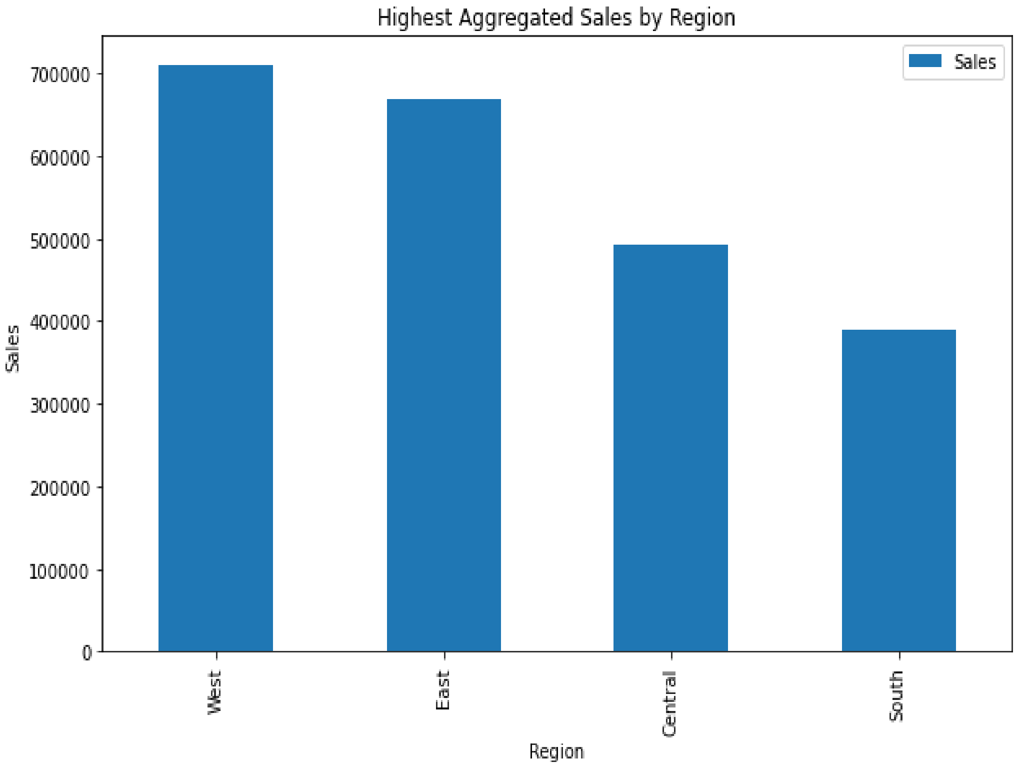
Row ID	False
Order ID	False
Order Date	False
Ship Date	False
Ship Mode	False
Customer ID	False
Customer Name	False
Segment	False
Country	False
City	False
State	False
Postal Code	True
Region	False
Product ID	False
Category	False
Sub-Category	False
Product Name	False
Sales	False

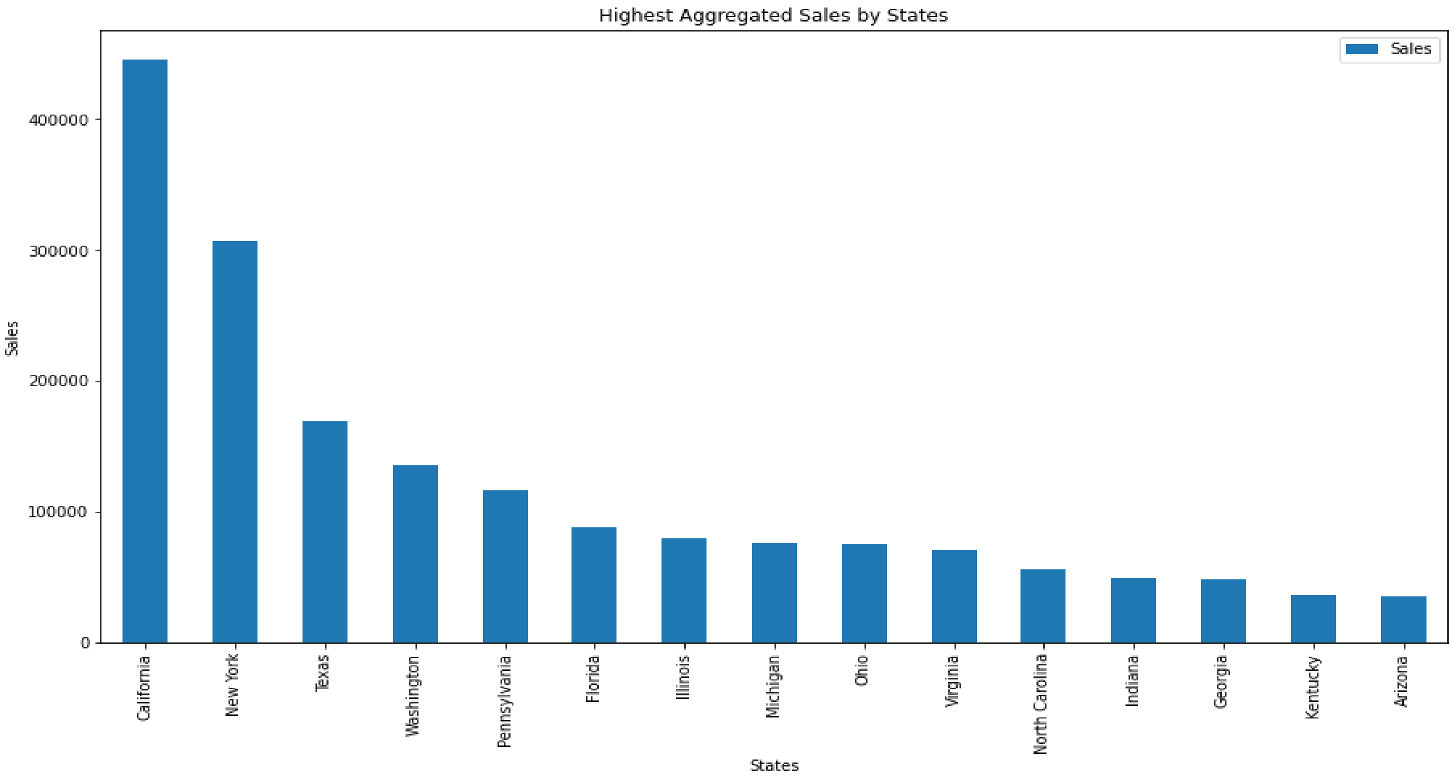
Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region
5/12/2018	10/12/2018	Standard Class	QJ-19255	Quincy Jones	Corporate	United States	Burlington	Vermont	NaN	East
7/11/2016	9/11/2016	Second Class	SV-20785	Stewart Visinsky	Consumer	United States	Burlington	Vermont	NaN	East
6/4/2017	10/4/2017	Standard Class	VM-21685	Valerie Mitchum	Home Office	United States	Burlington	Vermont	NaN	East

Total of 11 records : For the City of Burlington in the State of Vermont

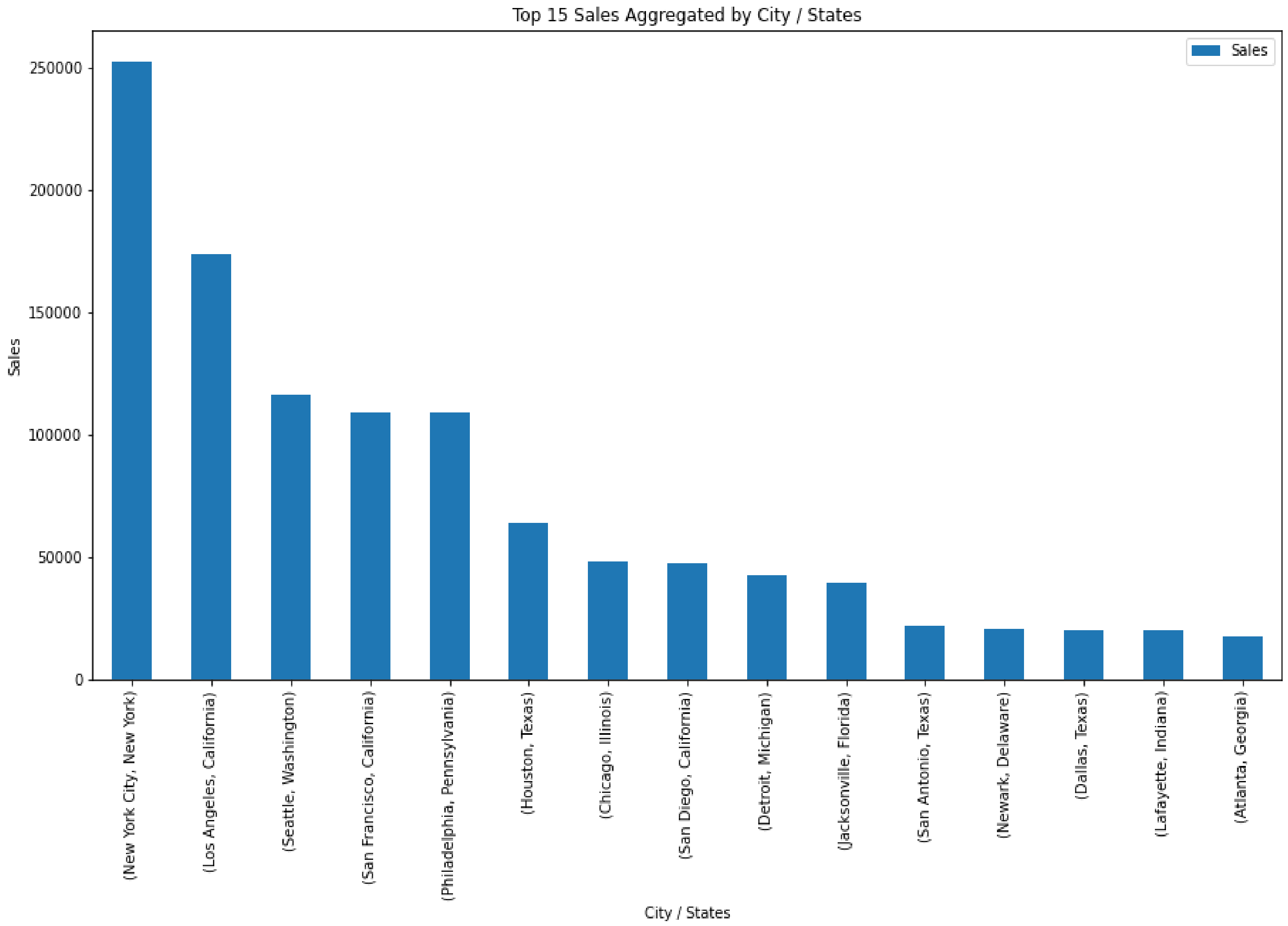
Replaced with ‘05401’ as default value

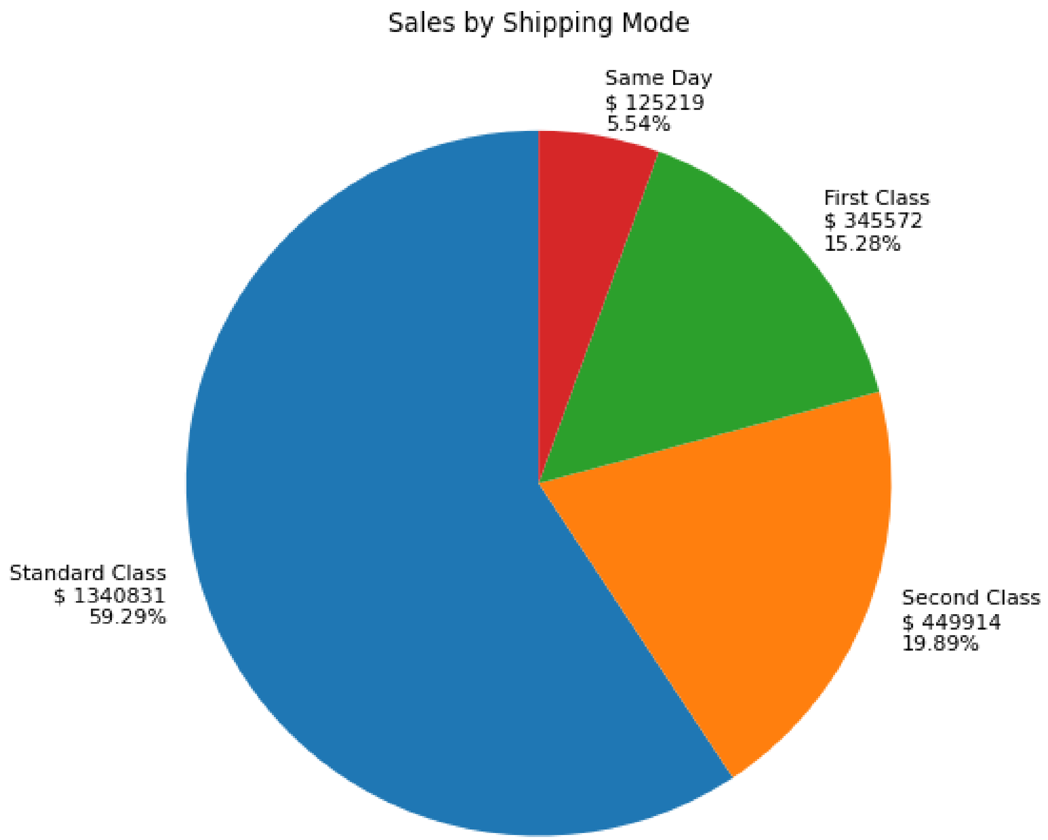
Features Analysis and Methods



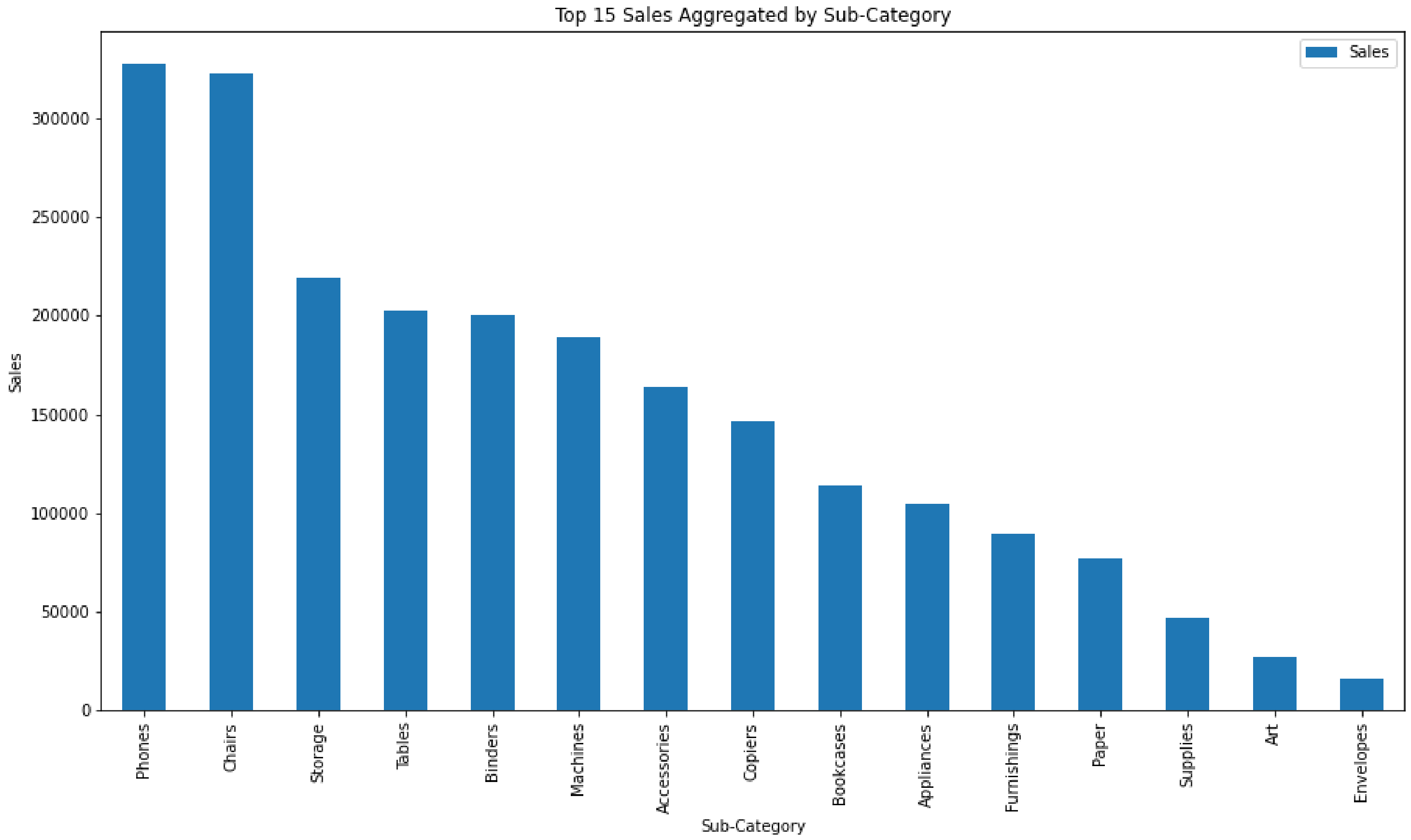


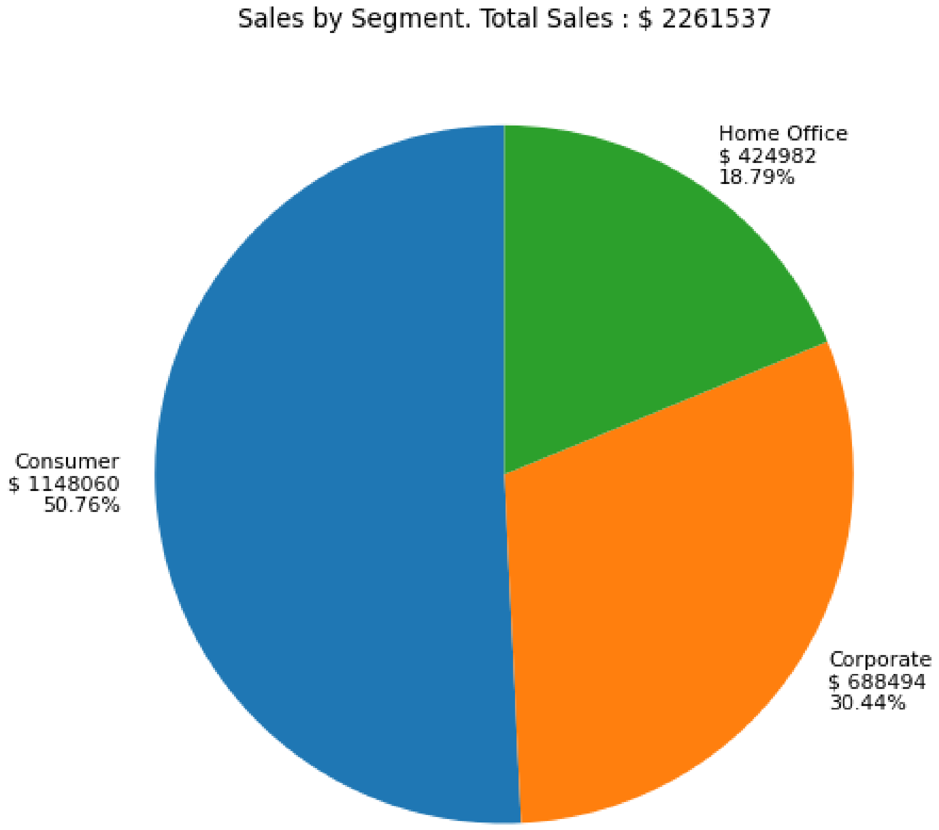
Features Analysis and Methods (continued)



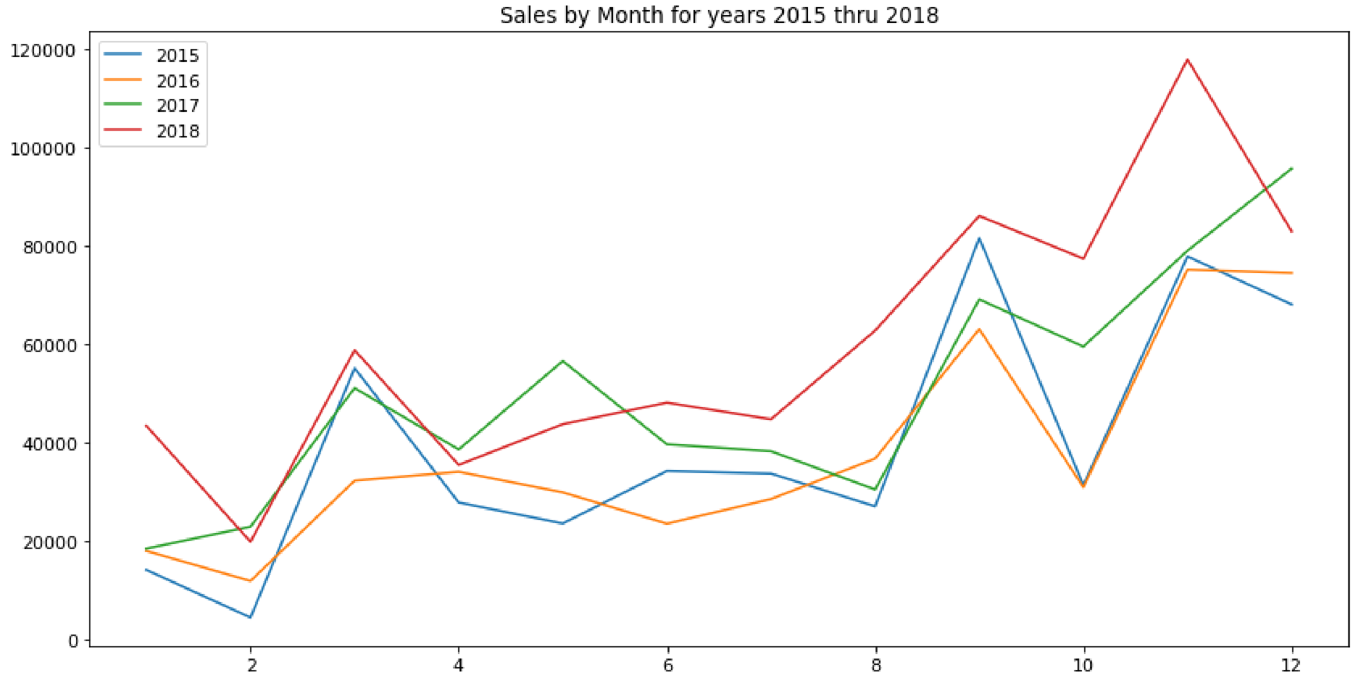


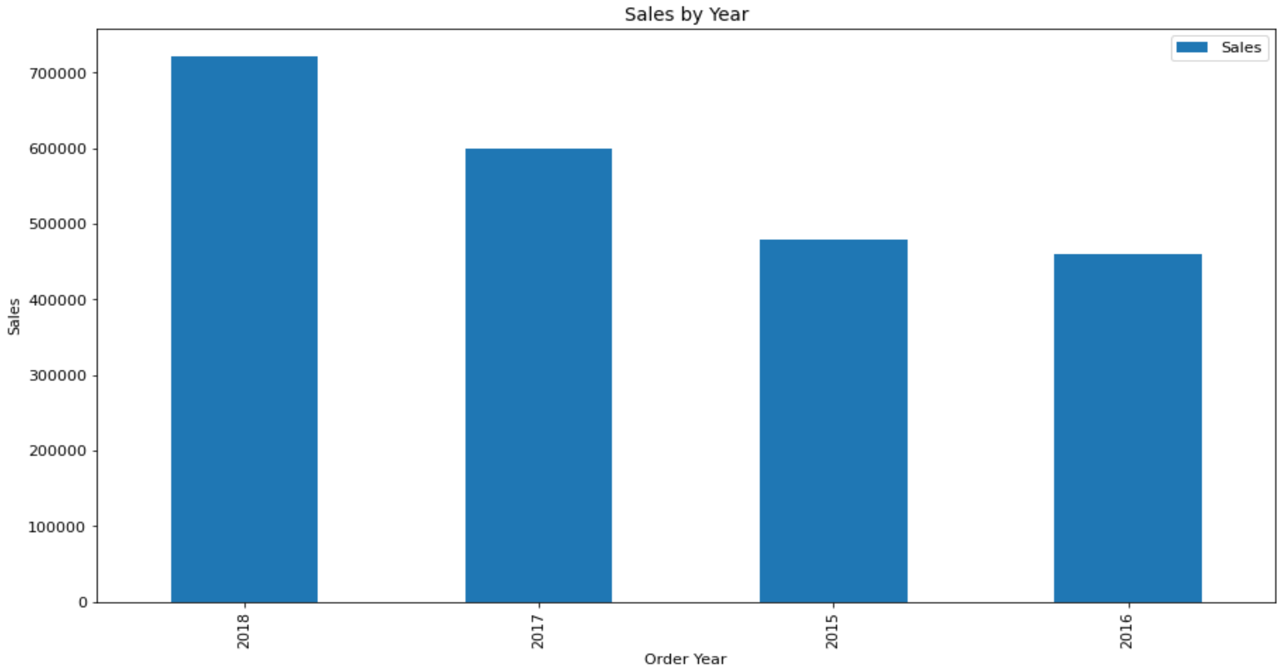
Features Analysis and Methods (continued)





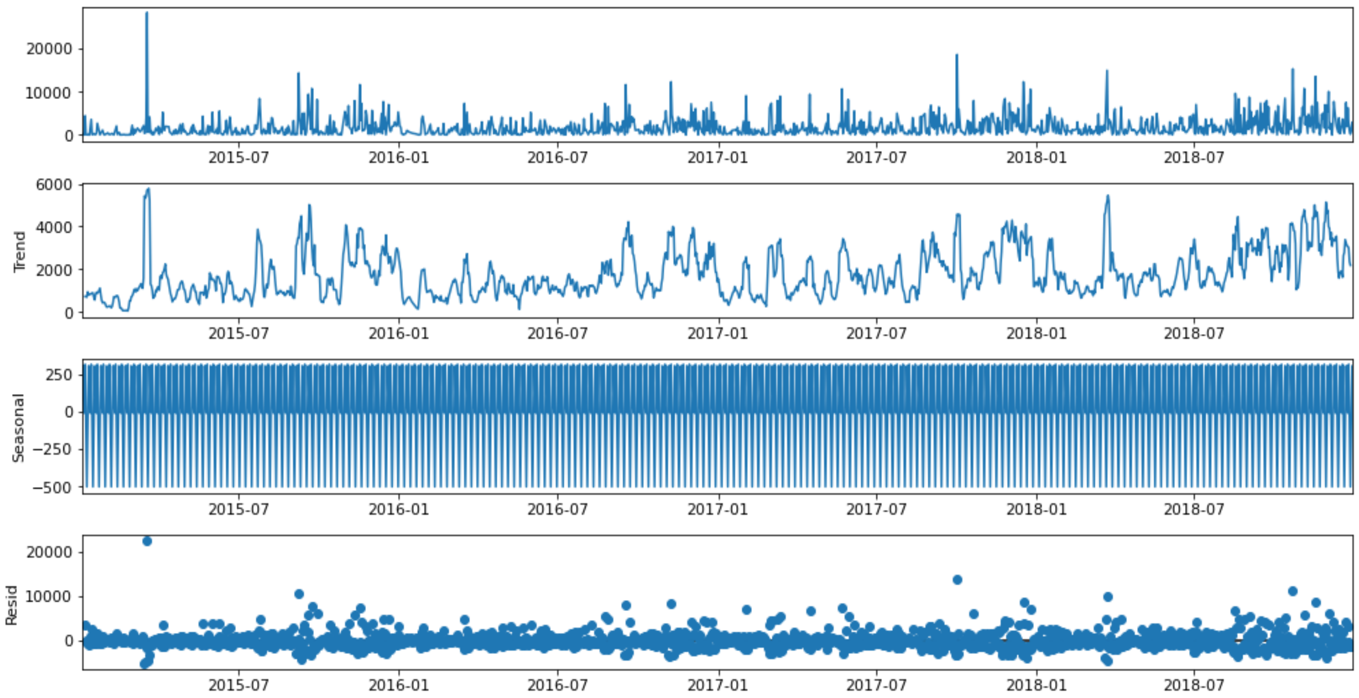
Features Analysis and Methods





Features Analysis and Methods

(Seasonality / Stationary test)



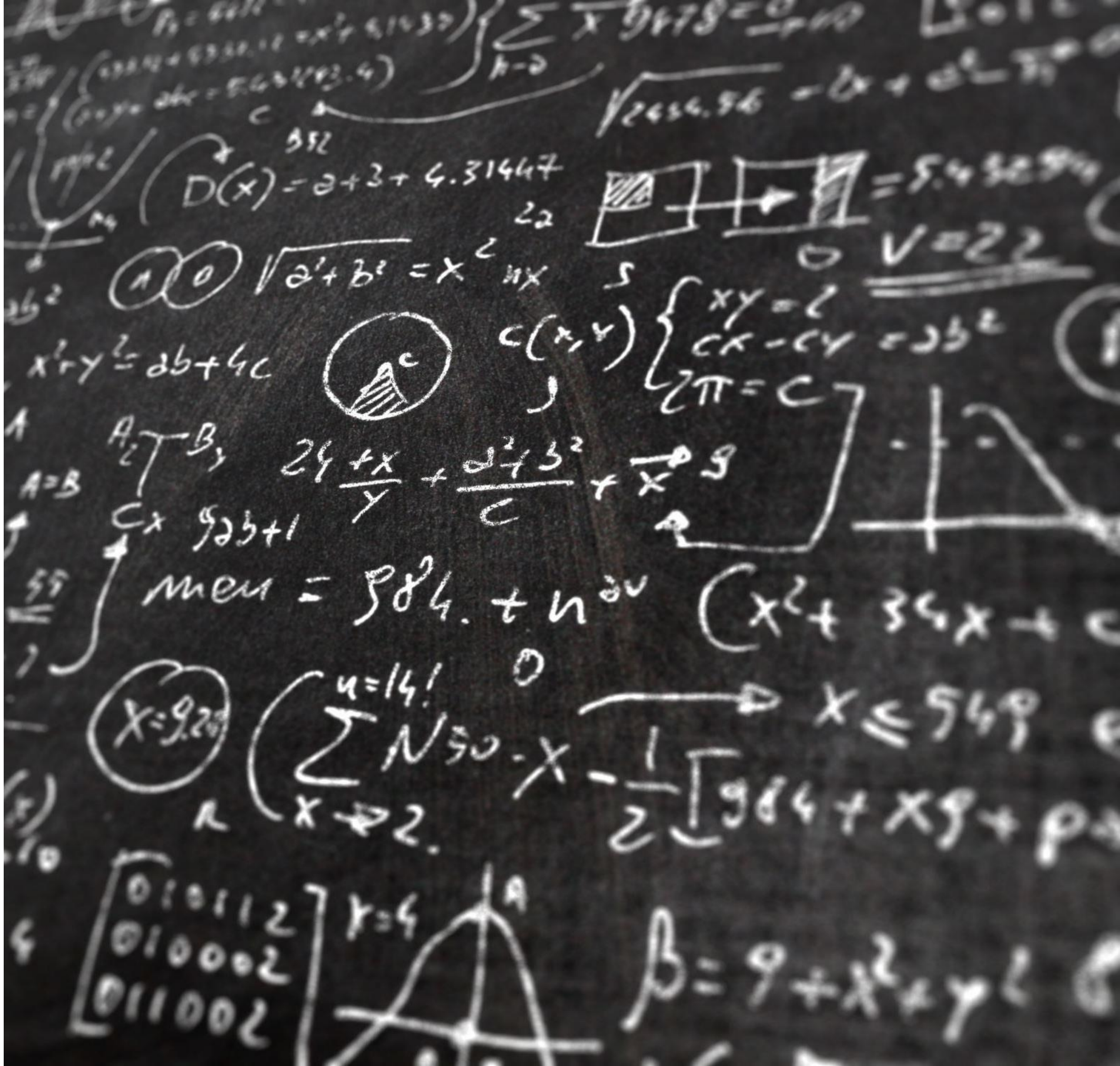

```
adf results:  (-6.444493081294046, 1.5781789978043213e-08, 15, 1442, {'1%': -3.4348929812602784, '5%': -2.863546418485167,
'10%': -2.5678382024888378}, 26015.358001313038)

ADF Statistic: -6.444493081294046
p-value: 1.5781789978043213e-08
```

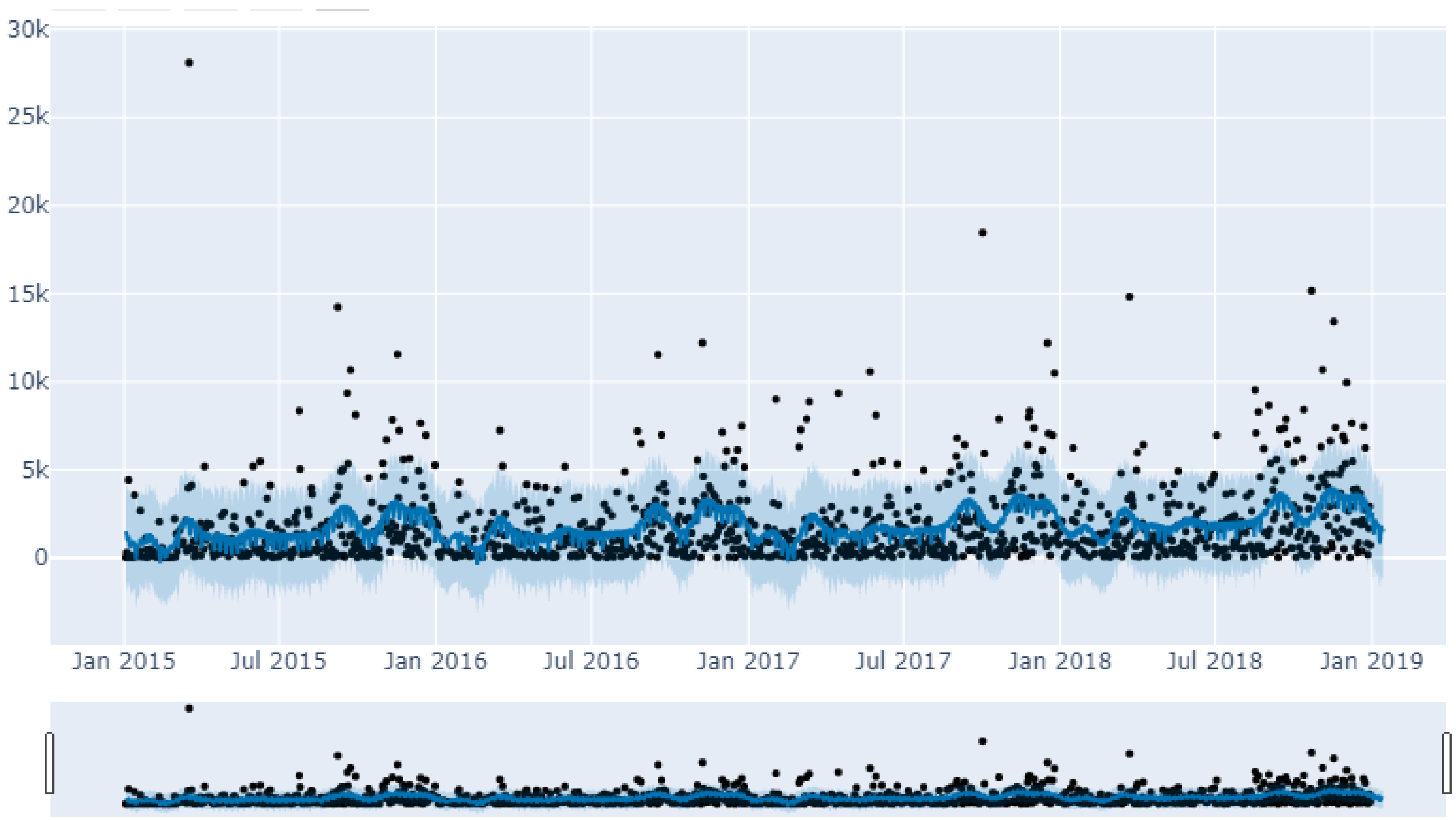
Since the p-value is much lesser than 0.05 (very close to 0.0 in fact), we can conclude that this Sales data time series is Stationary series

Machine Learning Models and Evaluation

- SARIMA
- Prophet (fbprophet)
- Neural Prophet
- LSTM

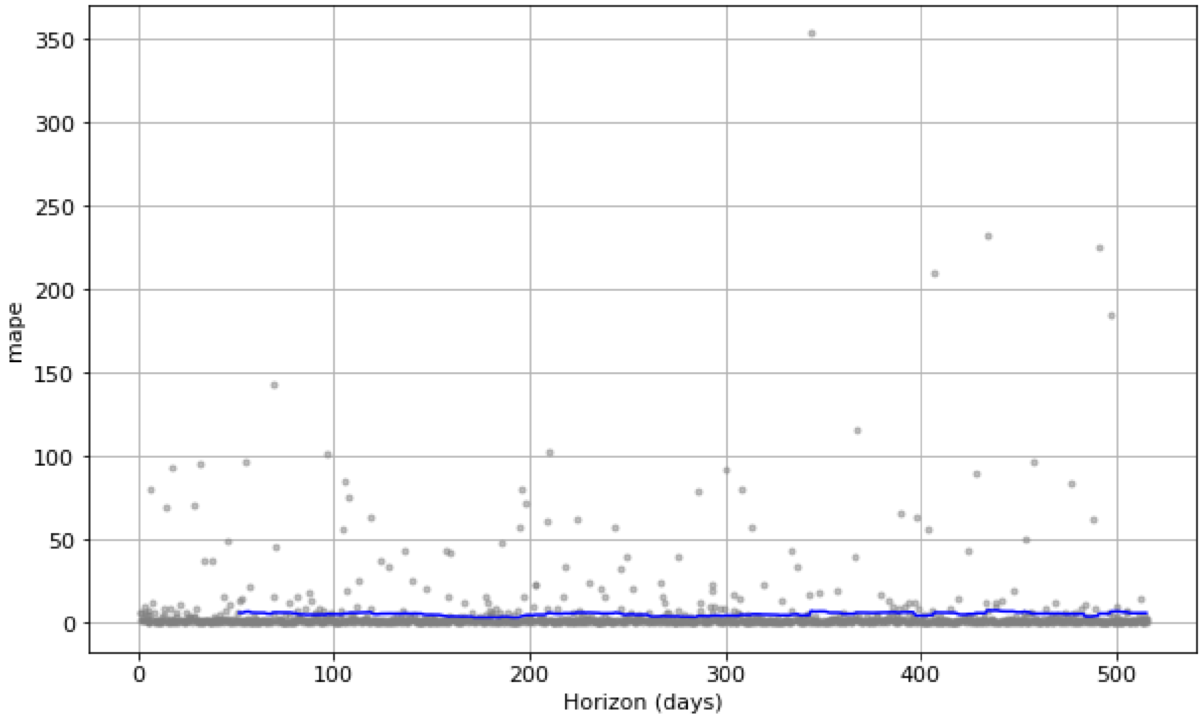


Machine Learning Models and Evaluation (fbprophet)



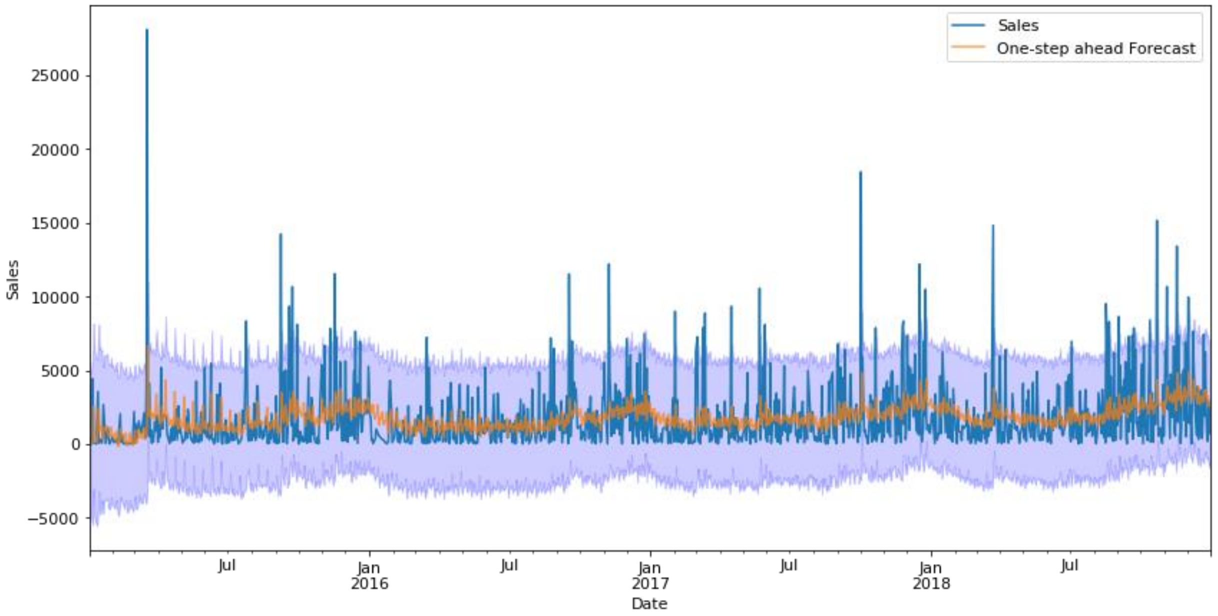
ds	yhat
2018-12-31	2314.560165
2019-01-01	2559.996785
2019-01-02	1986.735887
2019-01-03	1310.160220
2019-01-04	1831.587080
2019-01-05	2179.070029
2019-01-06	1858.885180
2019-01-07	1675.237461
2019-01-08	1962.377954
2019-01-09	1438.779767
2019-01-10	818.861687

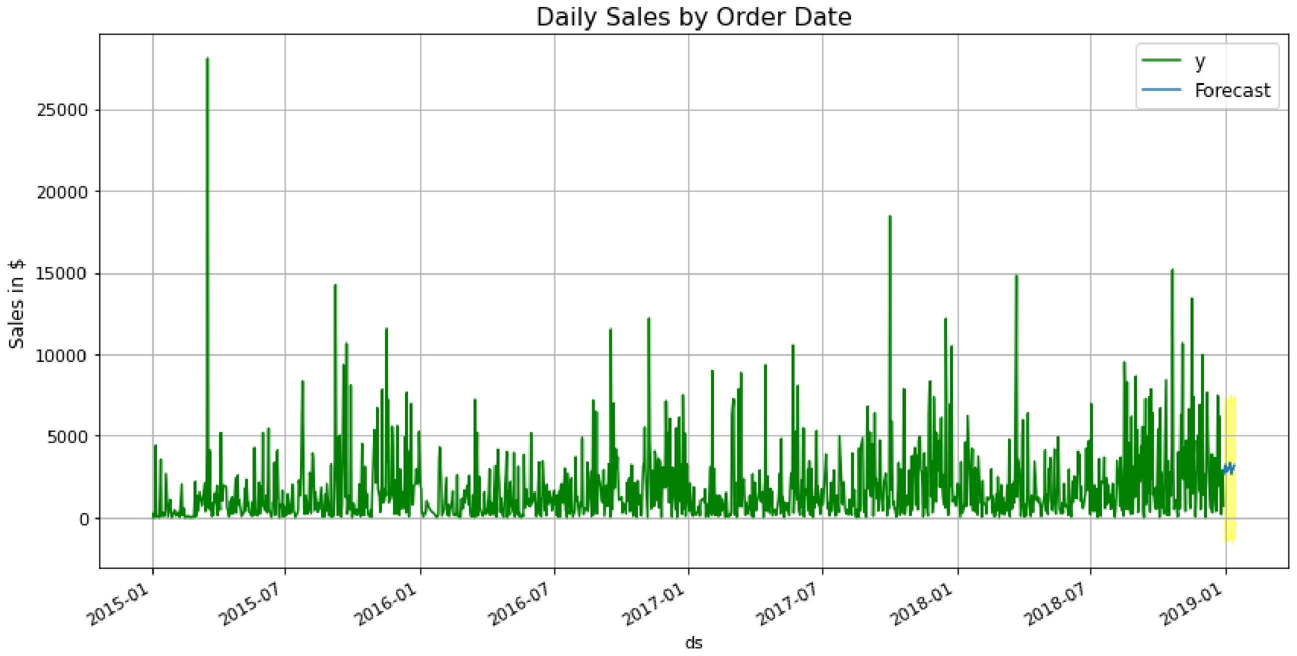
Machine Learning Models and Evaluation (fbprophet)



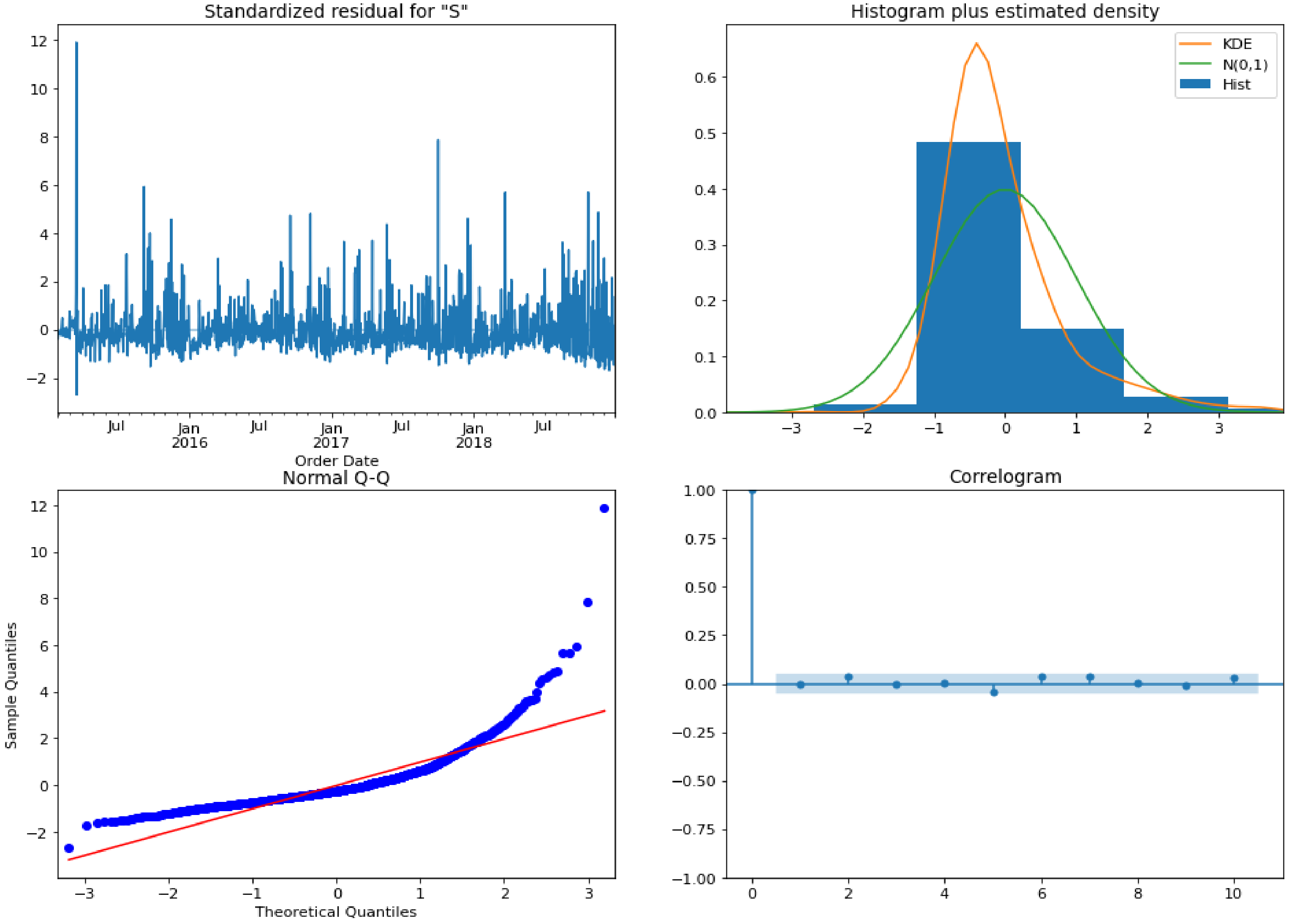
horizon	mse	rmse	mae	mape	mdape	coverage
511 days	6.879071e+06	2622.798236	1987.559586	5.580470	0.671296	0.742647
512 days	6.956558e+06	2637.528676	2011.011200	5.640538	0.687591	0.742647
513 days	6.903811e+06	2627.510453	1987.757728	5.631451	0.671296	0.742647
514 days	6.792085e+06	2606.162840	1960.255295	5.661361	0.674211	0.750000
515 days	7.019155e+06	2649.368852	1993.529684	5.664362	0.674211	0.742647

Machine Learning Models and Evaluation (SARIMA model)





Machine Learning Models and Evaluation (SARIMA model)



2018-12-31	2667.872280
2019-01-01	3200.266913
2019-01-02	3220.774239
2019-01-03	2796.500929
2019-01-04	2816.080403
2019-01-05	3045.837516
2019-01-06	2946.380407
2019-01-07	3008.356850
2019-01-08	3344.342459
2019-01-09	3066.914374
2019-01-10	2646.104611

The correlogram on the bottom right suggests that there is no autocorrelation in the residuals. Therefore, these residuals are not correlated, and the mean is close to zero.

Conclusion

- ❖ Looking at the Sales trends, we can see higher level of staff would be needed to support the operations during the Months October to January of following year and in the March
- ❖ West and East regions contribute to high level of Sales. So, inventories, operations and supply chains need to be fully ready to support the customer demands in these regions / states.
- ❖ The states of California, Washington, Texas, New York and Pennsylvania would see higher sales. So, maintain the sufficient inventories and staff levels in these places.
- ❖ Maintain sufficient inventory of the supplies in the Phones, Chairs, Storage Units, Tables as these contribute to higher level of Sales
- ❖ Customers in Consumer segments have high level of demands, which need to be supported
- ❖ Standard and Second-class Shipping modes are important for the business and hence shipping arrangements need to be made appropriately
- ❖ Expected Daily Sales levels would be in the range of $1000 - $2600 in the rolling 10 days time

Assumptions / Challenges / Future Usage

Assumption: The data has no outliers in terms of Sales figures. Most of the other information being text, has no issues.

Challenges / Limitations : Sales data having occasional spikes causes the sales predictions to be wide range bound and may deviate from actuals

Future Usage : The model can be partially utilized, or new system can be developed to forecast at the customer segments, sub-category, states, city level sales

References

- Data source : https://www.kaggle.com/rohitsahoo/sales-forecasting
- Sharma, Rishabh (2021) “How to build SARIMA model in Python” https://medium.com/mlearning-ai/how-to-build-sarima-model-in-python-7ae83b14c884
- Brownlee, Jason (2020) “Timeseries forecasting with Prophet in Python” https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/
- https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/
- https://facebook.github.io/prophet/docs/diagnostics.html
- https://nextjournal.com/fb-prophet/facebook-prophet-diagnostics
- Berk, Michael (2021) “Prophet vs. NeuralProphet” https://towardsdatascience.com/prophet-vs-neuralprophet-fc717ab7a9d8
- Alizadeh, Esmaeil (2020) “NeuralProphet: A Time-series modeling library based on Neural-Networks” https://towardsdatascience.com/neural-prophet-a-time-series-modeling-library-based-on-neural-networks-dd02dc8d868d

THANK YOU