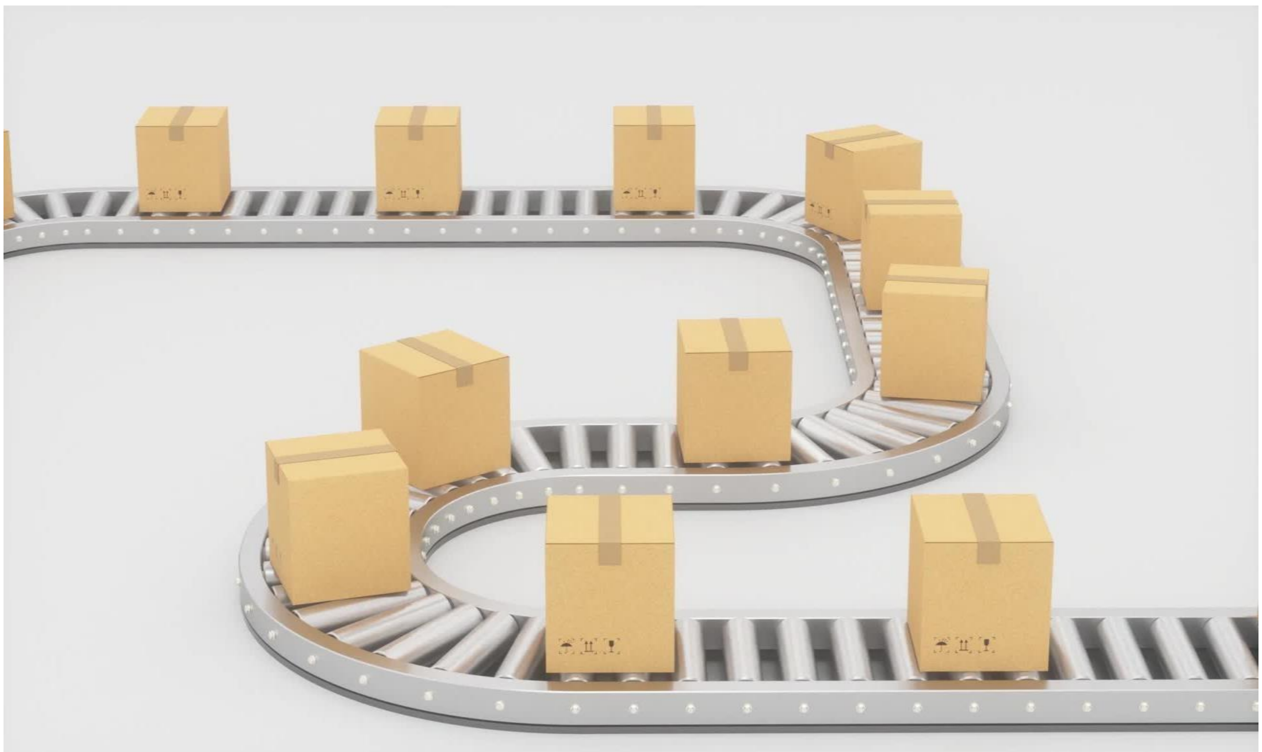
Product Distribution



TOPIC INTRODUCTION:

What is Product Distribution?

Product distribution is a plan of action that helps decide which approach makes the most strategic sense to provide products to consumers. A strategy for product distribution combines all the processes and methods a business institutes to execute their approach. Product distribution plays an essential role in a company's operations. Having the capability and resources to analyze and build on relationships between consumers and manufacturers is key. After the assessment of the operations, it is vital to have feedback loops that allow room for continuous developments. This project report determines an overview of the best ways to sell and deliver products and services to the consumer.

Why this project?

This project allows us to understand and analyze product distribution which will help our team and peers gain knowledge about the importance of improving distribution to ensure customers are overall satisfied with your distribution channels.

Expected Outcomes:

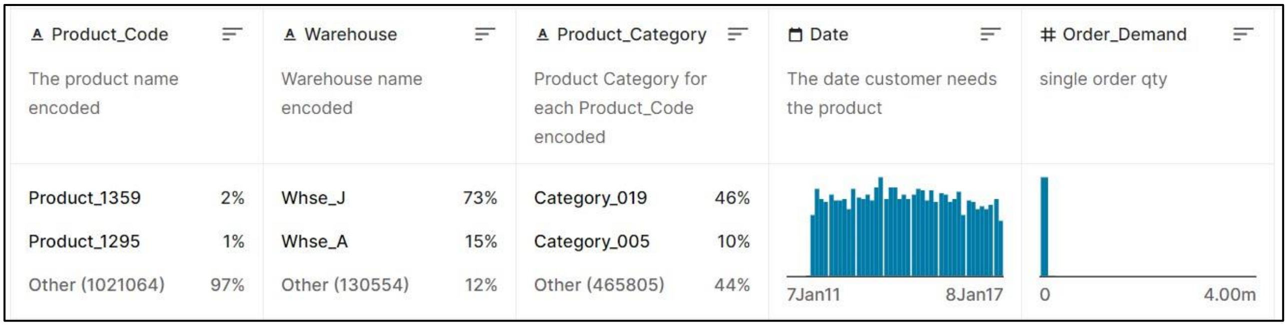
With respect to our Kaggle dataset, the company provides thousands of products within dozens of product categories. There are four central warehouses to ship products within the region it is responsible for. Since the products are manufactured in different locations all over the world, it normally takes more than one month to ship products via ocean to different central warehouses. If forecasts for each product with reasonable accuracy for the weekly demand for the week after next can be achieved, it would be beneficial to the company. This dataset is all real-life data and products/warehouse and category information encoded.

Throughout this project, we will use predictive analysis to forecast the weekly demand for one selected product in the next year. **Time series Auto Regression** will help us select one product with high-quality available data. Additionally, through prescriptive analytics, we will determine the optimal weekly order amount of the selected product to meet as many demands as possible yet without many surpluses. To execute these demands, we'll utilize **Simulation**.

DATA COLLECTION:

Data Source:

All the data for the dataset 'Historical product demand for a manufacturing company' from Kaggle website. CSV data file containing product demand for encoded product ID's. Here is a snap of the contents of the csv file.



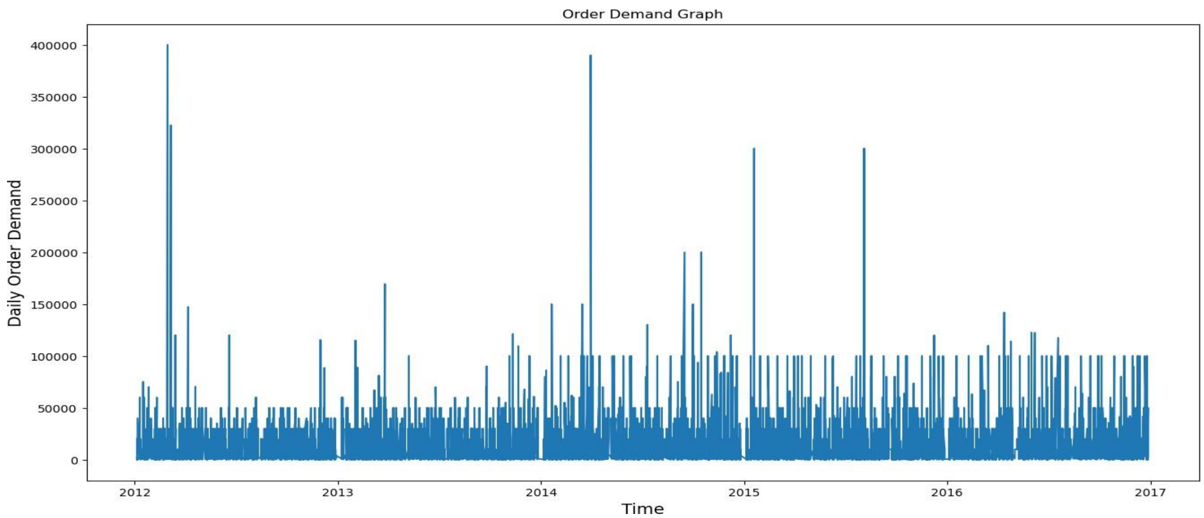
Data Cleaning and Preprocessing:

We started the Data Cleaning process by converting all columns into necessary data types for analysis. For instance, Object type 'Date' was converted into 'datetime' datatype, 'Order_Demand' into 'int64' datatype. There were special characters '(' and ')' in the column 'Order_Demand' that were removed before type conversion.

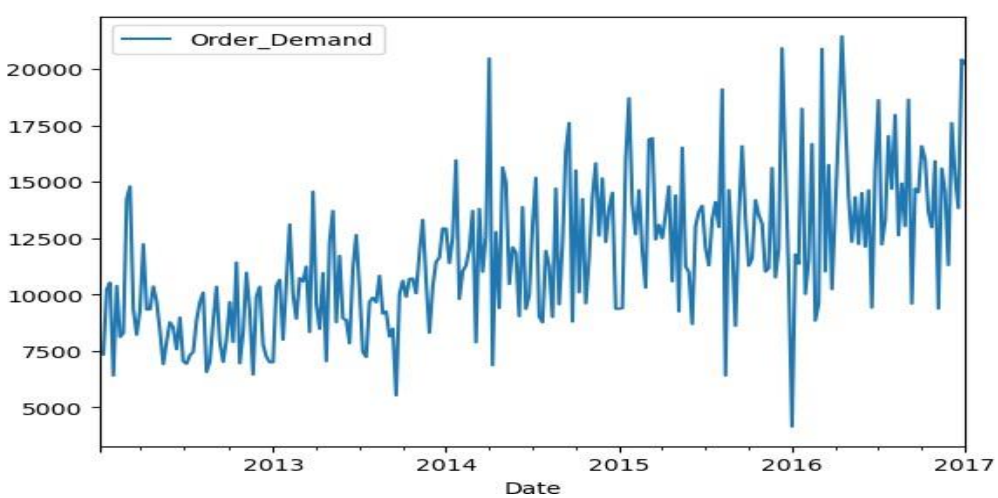
We sorted the whole dataset on the column 'Date' and found many records without any values in the 'Date' column and they were removed. We had to select one product with high-quality data from the dataset. We took many factors into consideration like – Consistancy, Continuity of the data, range of the Order_Demand, Outlier values and narraowed it down to **Product_1295**.

VISUALIZATION:

We found the data ranging between 5th Jan 2012 to 28th Dec 2016 for the Product_1295. Here is a plot of the Date vs Order_Demand for the Product_1295. The Order_Demand ranges between 0 to 400,000 in this plot.



In order reduce noise, we did resampling of the data on the weekly mean. We have 260 weeks of data with values of Order_Demand ranging between 4,142 to 21,447.

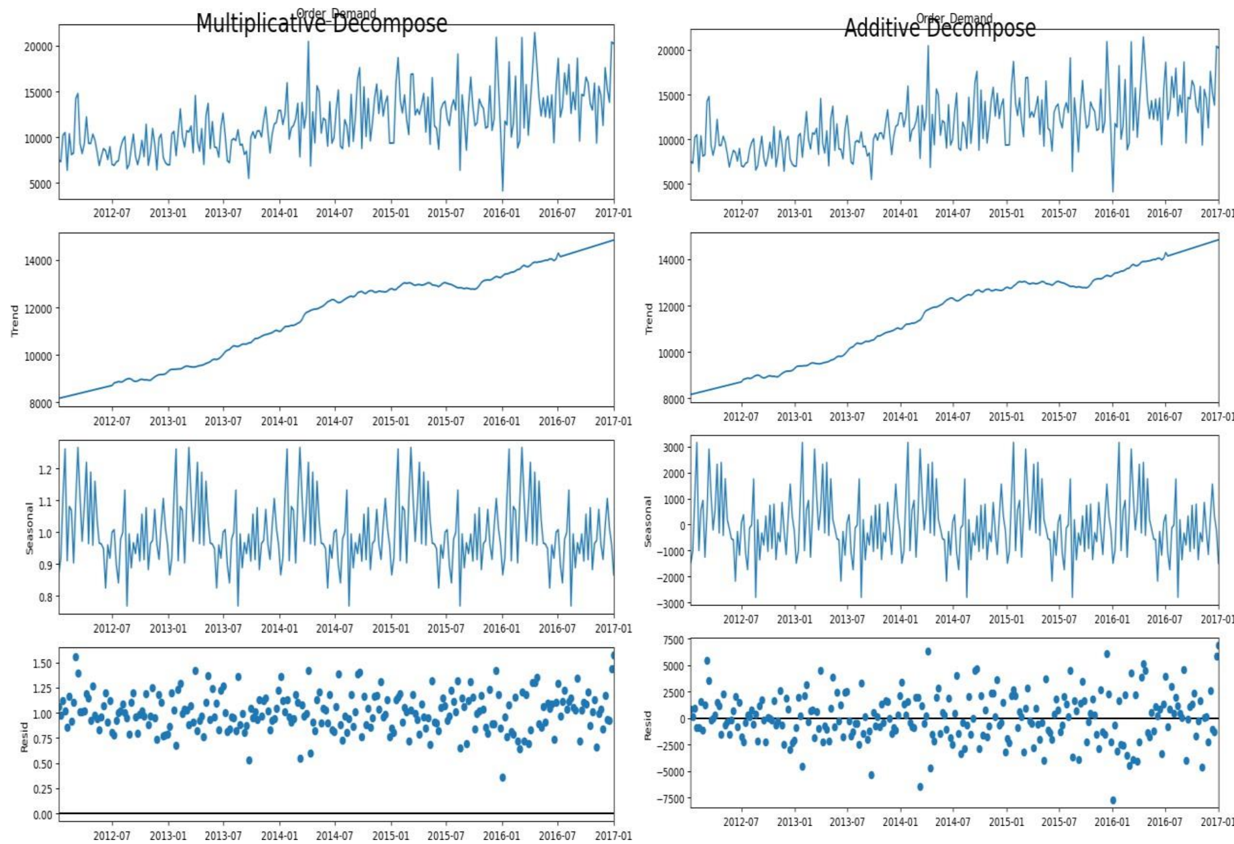


PREDICTIVE ANALYTICS:

Seasonal Decomposition:

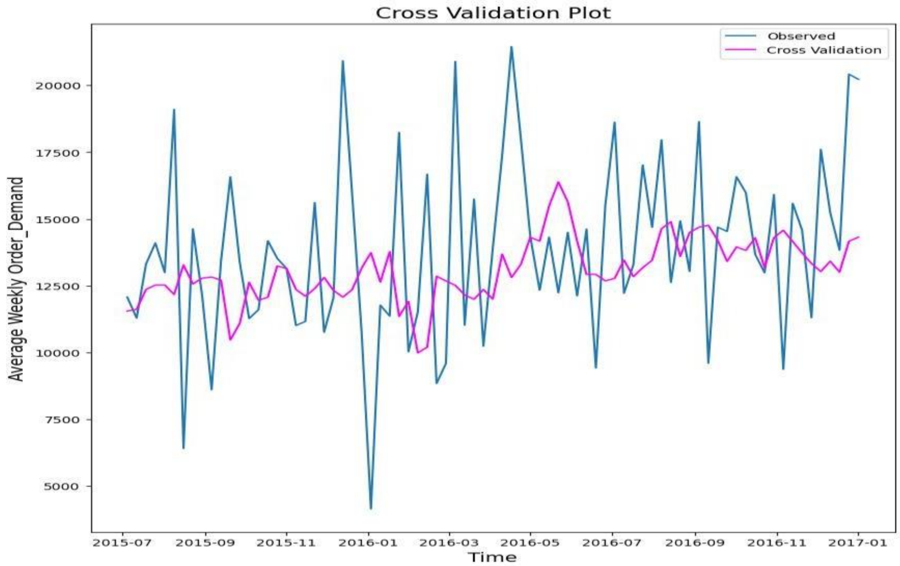
In order to find our seasonal decomposition, we used **statsmodels.tsa.seasonal** from the library **seasonal_decompose**. There are four interpretations in both Additive and Multiplicative decomposition plots.

First plot represents Weekly Order_Demand vs Time. Second plot represents <u>Trend</u>, which is long term movement of the data; we see a gradual increase in the trend. Third is the <u>Seasonality</u> component that represents repeating patterns in the data that occur at regular intervals. Fourth is the <u>Residual</u> component that represents variation in the data that cannot be explained by trend or seasonality. In this dataset, Residuals are randomly distributed around zero with no discernible patterns in the additive decomposition; thus we conclude that this particular dataset falls under Additive Decompose.



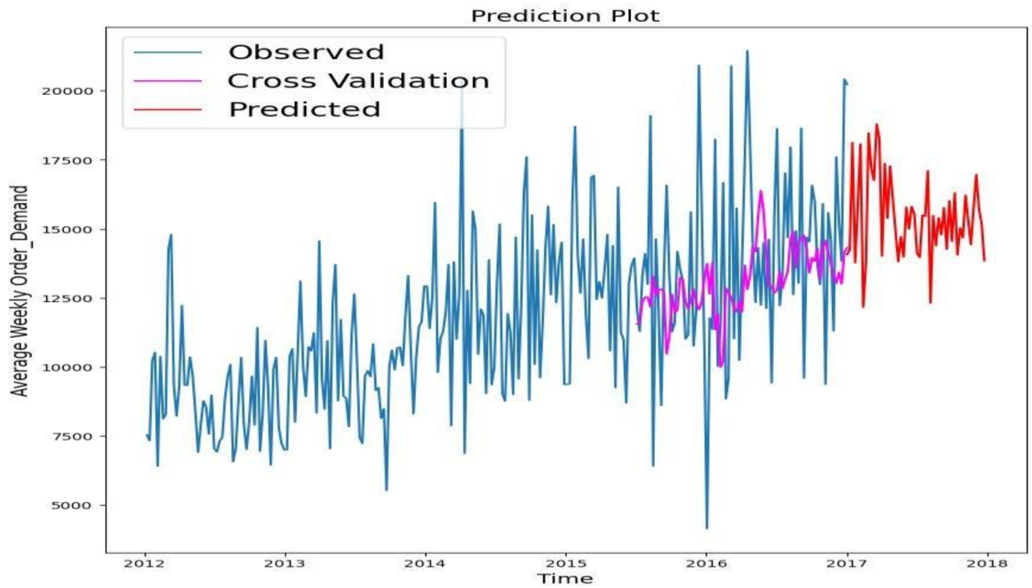
Cross Validation:

We used Auto-Regression method for Predictive Analytics. We took first 70% of the data as '*Training data*' and rest 30% as '*Test Data*'. We did cross-validation for about 70 datapoints and below is the plot. The forecast accuracy parameters were calculated and it was least for product_1295; that was one more reason to choose this product for our analysis.



Prediction:

Here is the forecast plot, we are predicting Order_Demand for next 52 weeks (for year 2017) for Product_1295.



Analysis and Conclusion:

Predicting values for a dataset using time series analysis can be a powerful tool for forecasting future trends and making informed decisions. The autoregression model is based on the assumption that past values are a good predictor of future values, which may not always hold true if there are significant changes in the underlying data generating process. Factors such as seasonality, trends, and external events may all influence the data, and should be carefully considered when interpreting the predictions. Finally, predictions are only estimates, and actual values may vary from the predicted values, but it is an important tool for making informed decisions. We have used the predicted data to perform *Prescriptive Analytics* using Simulation method for the observations from Predicted data.

PRESCRIPTIVE ANALYTICS:

Simulation:

Simulation is implemented based on the predicted data for the year 2017 i.e.,52 weeks and we are assuming cost price, selling price and salvage value as there is no cost/price provided in the dataset.
Cost Price = $2.5
Selling Price = $5.5
Salvage Value = $1.00

From the predicted data, we analyzed and performed **7 simulations** as the predicted values ranges from 12,000 to 18,000, namely *Simulation1* to *Simulation7*.

Here are the definition of each column being used in each of the Simulation steps:

- Forecast Demand - predicted values
- Inventory **-** values ranging from 12000 to 18000 (these are the values from predicted data)
- Underage – the amount of stock less than the actual stock
- Overage – the amount of stock more than the actual stock
- Profit – the average amount which comes from sold stocks
- Cost of Lost Sales – the average amount of understock sales

How are calculations performed?

There are 2 scenarios that were considered while performing calculations:

1. Understock of inventory – it falls into 'cost of lost sales', calculated as
   (**demand*selling price – cost price)**
2. Overstock of inventory – it falls into 'profit' where overstock is sold on a discount price (salvage value), calculated as:
   (selling price *demand + salvage value*overstock) – (cost price * inventory)

Each simulation is performed based on the inventory value with the above calculations.

Simulations:

Simulation 1:

	A	B	C	D	E	F
1	Forecast Demand	Inventory	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	12000	2068.801099	0	66000	6206.403296
3	14236.63174	12000	2236.631743	0	66000	6709.895228
4	18124.64866	12000	6124.648658	0	66000	18373.94598
5	13768.18334	12000	1768.18334	0	66000	5304.55002
6	15861.91116	12000	3861.911156	0	66000	11585.73347
7	18070.23185	12000	6070.231849	0	66000	18210.69555
8	12158.1374	12000	158.1373982	0	66000	474.4121946
9	13713.92739	12000	1713.927388	0	66000	5141.782164
10	18466.36286	12000	6466.362859	0	66000	19399.08858

Simulation 2:

	A	B	C	D	E	F
1	Forecast Demand	Stock	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	13000	1068.801099	0	71500	3206.403296
3	14236.63174	13000	1236.631743	0	71500	3709.895228
4	18124.64866	13000	5124.648658	0	71500	15373.94598
5	13768.18334	13000	768.1833401	0	71500	2304.55002
6	15861.91116	13000	2861.911156	0	71500	8585.733467
7	18070.23185	13000	5070.231849	0	71500	15210.69555
8	12158.1374	13000	0	841.8626	35211.6183	0
9	13713.92739	13000	713.9273879	0	71500	2141.782164
10	18466.36286	13000	5466.362859	0	71500	16399.08858

Simulation 3:

	A	B	C	D	E	F
1	Forecast Demand	Stock	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	14000	68.8011	0	77000	206.4033
3	14236.63174	14000	236.63174	0	77000	709.89522
4	18124.64866	14000	4124.64866	0	77000	12373.94598
5	13768.18334	14000	0	231.81666	40956.82503	0
6	15861.91116	14000	1861.91116	0	77000	5585.73348
7	18070.23185	14000	4070.23185	0	77000	12210.69555
8	12158.1374	14000	0	1841.8626	33711.6183	0
9	13713.92739	14000	0	286.07261	40712.67326	0
10	18466.36286	14000	4466.36286	0	77000	13399.08858

Simulation 4:

	A	B	C	D	E	F
1	Forecast Demand	Inventory	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	15000	0	931.1989	40809.60495	0
3	14236.63174	15000	0	763.36826	41564.84283	0
4	18124.64866	15000	3124.64866	0	82500	9373.94598
5	13768.18334	15000	0	1231.81666	39456.82503	0
6	15861.91116	15000	861.91116	0	82500	2585.73348
7	18070.23185	15000	3070.23185	0	82500	9210.69555
8	12158.1374	15000	0	2841.8626	32211.6183	0
9	13713.92739	15000	0	1286.07261	39212.67326	0
10	18466.36286	15000	3466.36286	0	82500	10399.08858

Simulation 5:

	A	B	C	D	E	F
1	Forecast Demand	Stock	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	16000	0	1931.1989	39309.60495	0
3	14236.63174	16000	0	1763.36826	40064.84283	0
4	18124.64866	16000	2124.64866	0	88000	6373.94598
5	13768.18334	16000	0	2231.81666	37956.82503	0
6	15861.91116	16000	0	138.08884	47378.60022	0
7	18070.23185	16000	2070.23185	0	88000	6210.69555
8	12158.1374	16000	0	3841.8626	30711.6183	0
9	13713.92739	16000	0	2286.07261	37712.67326	0
10	18466.36286	16000	2466.36286	0	88000	7399.08858

Simulation 6:

	A	B	C	D	E	F
1	Forecast Demand	Inventory	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	17000	0	2931.1989	37809.60495	0
3	14236.6317	17000	0	2763.3683	38564.84265	0
4	18124.6487	17000	1124.6487	0	93500	3373.9461
5	13768.1833	17000	0	3231.8167	36456.82485	0
6	15861.9112	17000	0	1138.0888	45878.6004	0
7	18070.2319	17000	1070.2319	0	93500	3210.6957
8	12158.1374	17000	0	4841.8626	29211.6183	0
9	13713.9274	17000	0	3286.0726	54286.0726	0
10	18466.3629	17000	1466.3629	0	93500	4399.0887

Simulation 7:

	A	B	C	D	E	F
1	Forecast Demand	Inventory	Underage	Overage	Profit	Cost of Lost Sales
2	14068.8011	18000	0	3931.1989	36309.60495	0
3	14236.6317	18000	0	3763.3683	37064.84265	0
4	18124.6487	18000	124.6487	0	99000	373.9461
5	13768.1833	18000	0	4231.8167	34956.82485	0
6	15861.9112	18000	0	2138.0888	44378.6004	0
7	18070.2319	18000	70.2319	0	99000	210.6957
8	12158.1374	18000	0	5841.8626	27711.6183	0
9	13713.9274	18000	0	4286.0726	58286.0726	0
10	18466.3629	18000	466.3629	0	99000	1399.0887

Summary of Simulations:

From the Summary stats, it can be observed that the **highest average profit is $71,917.43** which comes from **optimal sales of 14,000**.

A	B	C	D	E	F
			Summary of Simulations		
	Average Inventory	Average Underage	Avearge Overage	Average Lost	Average Profit
	12000	3415.780389	0	10247.34117	66000
	13000	2445.155509	29.375121	7335.466526	70117.812
	14000	1501.326538	85.54615	4503.979615	71917.43447
	15000	805.0373856	389.2569969	2415.112157	65581.78628
	16000	364.7063723	948.9259837	1094.119117	55268.29461
	17000	146.9517769	1731.171377	440.8553308	50913.06341
	18000	32.81088846	2617.03049	98.43266538	47003.62048

Analysis and Conclusion:

Overall, the Simulation process for Prescriptive Analytics involves iteratively running scenarios, evaluating results and making data-driven decisions to optimize outcomes. It helps businesses and organizations make more informed choices by simulating and comparing various possible courses of action before implementing them in the real world.

CRITICAL THINKING:

We are evaluating the assumptions, results and methods of both Auto Regression and Simulation analysis. This is to ensure the robustness of the methods used, their reliability and validity.

1. **Data Quality:**
   The dataset under consideration is just a sample of the Product Distribution of one product. We are assuming that data is complete and accurate. However, the data is historic and might not be applicable for the current trend.

2. **Model Selection:**
   We used Auto Regression for Time Series Analysis. This might not be the ideal method for predicting the data if there are many outliers and missing values in the dataset. Other models like ARIMA, LSTM might yield better predictions.

3. **Assumptions:**
   We assumed the Cost Price, Selling Price and Salvage Value in the Simulation process. It might not reflect the actual scenario. Extreme violations of these assumptions can lead to inaccurate forecasts and unreliable results which should be taken into account.

4. **Validation:**
   In order to validate the Auto Regression analysis, we considered forecast accuracy parameters to compare predicted and actual values.

By applying critical thinking to these aspects, we can identify opportunities for improvement, optimize the distribution strategy, and enhance overall efficiency in Product Distribution. This will aid the company for a better data driven decision to enhance profits and customer satisfaction.

Data References

1. McClure, Nathan. "What Are Product Distribution Models?" *Definition and Overview*, 11 Feb. 2022, https://www.productplan.com/learn/product-distribution-models/.
2. Zhao, Felix. "Forecasts for Product Demand." *Kaggle*, June 2017, https://www.kaggle.com/datasets/felixzhao/productdemandforecasting.
3. Camm, D. J., J.J. Cochran. Business Analytics 4th edition. Cengage Learning: Boston MA (2020)
4. Downey, A. (2015). *Think Python: How to think like a computer scientist 2nd edition.* Green Tea Press: Needham, MA