BIA BOSTON INSTITUTE OF ANALYTICS ®

DEMAND FORECASTING & INVENTORY OPTIMIZATION

SOUMADEEP BHATTACHARYA

INTRODUCTION

- Demand Forecasting means estimating future customer demand for a product or service based on historical data and relevant factors. Inventory Optimization is the strategic management of inventory levels to ensure that the right amount of goods is available at the right time to meet customer demand while minimizing costs.
- Demand Forecasting involves predicting the quantity and pattern of customer orders, which is crucial for businesses to efficiently allocate resources, manage inventory, and plan production. Accurate demand forecasting enables companies to meet customer needs, avoid overstocking or understocking, and optimize their supply chain operations.
- Inventory Optimization aims to strike a balance between having sufficient stock to meet demand without carrying excess inventory that ties up capital and storage space. Effective inventory optimization helps businesses reduce carrying costs, improve cash flow, and enhance customer satisfaction.
- These concepts are especially relevant in retail, manufacturing, and distribution, where managing supply and demand dynamics is essential for profitability and customer satisfaction.

PROBLEM STATEMENT

- One of the largest retail chains in the world wants to use their vast data source to build an efficient forecasting model to predict the sales for each SKU in its portfolio at its 76 different stores using historical sales data for the past 3 years on a week-on-week basis. Sales and promotional information is also available for each week - product and store wise.

However, no other information regarding stores and products are available. Can you still forecast accurately the sales values for every such product/SKU-store combination for the next 12 weeks accurately?

- (**ADDITIONAL & EXTRA):**

1. There is no information about these product's inventory. Can you create inventory of those products individually? If Yes, then create it & optimize it & give some insights if possible.
2. I create a streamlit app framework where we can select the product id & return it's corresponding 12 week forecast values.

TABLE OF CONTENTS :– DEMAND FORECASTING

TABLE OF CONTENTS :– INVENTORY OPTIMIZATION

DEMAND FORECASTING :- DATASET DESCRIPTION

Data Description

Variable	Definition
record_ID	Unique ID for each week store sku combination
week	Starting Date of the week
store_id	Unique ID for each store (no numerical order to be assumed)
sku_id	Unique ID for each product (no numerical order to be assumed)
total_price	Sales Price of the product
base_price	Base price of the product
is_featured_sku	Was part of the featured item of the week
is_display_sku	Product was on display at a prominent place at the store
units_sold	(Target) Total Units sold for that week-store-sku combination

	record_ID	week	store_id	sku_id	total_price	base_price	is_featured_sku	is_display_sku	units_sold
0	1	17/01/11	8091	216418	99.0375	111.8625	0	0	20
1	2	17/01/11	8091	216419	99.0375	99.0375	0	0	28
2	3	17/01/11	8091	216425	133.9500	133.9500	0	0	19
3	4	17/01/11	8091	216233	133.9500	133.9500	0	0	44
4	5	17/01/11	8091	217390	141.0750	141.0750	0	0	52
...	...	...	...	...	...	...	...	...	...
150145	212638	09/07/13	9984	223245	235.8375	235.8375	0	0	38
150146	212639	09/07/13	9984	223153	235.8375	235.8375	0	0	30
150147	212642	09/07/13	9984	245338	357.6750	483.7875	1	1	31
150148	212643	09/07/13	9984	547934	141.7875	191.6625	0	1	12
150149	212644	09/07/13	9984	679023	234.4125	234.4125	0	0	15

DEMAND FORECASTING :- WHAT'S THE PLAN?

SO THE PLAN IS AS FOLLOWS:

1. AT FIRST WE JUST EXTRACT THE SALES DATA FOR THAT PARTICULER PRODUCT ID

2. AS THE DATA **HAS SO MANY MISSING DATES SO WE FILL THE MISSING DATES BY "ffill" METHOD**

3. **AS THAT METHOD CREATES A HIGH BIAS** SO WE HAVE **TO SMOOTH THE DATA** TO TACKLE THAT BIAS. FOR SMOOTHING WE USE HERE **ROLLING AVERAGE SMOOTHING**

4. NOW SEE THE DATA IS **STATIONARY** OR NOT BY **AUGMENTED DICKEY FULLER TEST**

5. IF p-VALUE < 0.05 THEN THE DATA IS STATIONARY & WE GO AHEAD. BUT IF p-VALUE>0.05 THEN THE DATA IS NON-STATIONARY & THEN **WE HAVE TO MAKE IT STATIONARY FIRST BY DIFFERENCING**.

6. WE PLOT **THE ACF & PACF** ON THE STATIONARY DATA & WATCH FOR ACF PLOT TO SEE ANY WAVE LIKE PATTERN. IF THERE IS A WAVE PATTERN THEN WE CONCLUDE THAT THERE IS A **SEASONAL COMPONENT** PRESENT IN OUR DATA & WE HAVE TO FIT **SARIMA MODEL**. OTHERWISE WE GO AHEAD WITH **ARIMA MODEL**.

7. LASTLY WE FORECAST USING THAT ARIMA/SARIMA MODEL FOR NEXT 12 WEEKS & SAVE THEM AS A DATAFRAME & ALSO PLOT THEM TO SEE THE BEHAVIOUR

DEMAND FORECASTING :- BASIC TERMINOLOGY

I. <u>ffill</u> : This function in Pandas is used to fill missing values in a dataframe. It stands for forward fill & it propagates the last valid observation forward.

II. <u>Rolling Average Smoothing Method</u> : The rolling average smoothing method is used to smooth out short-term fluctuations and highlight longer-term trends or cycles in time series data .The method works by calculating the mean of a fixed subset of the number series, then modifying the subset by "shifting forward" and recalculating the mean. This process is repeated until all data points have been included in at least one subset.

III. **<u>Dickey-Fuller Test</u>** : The Dickey-Fuller test is a statistical test used in time-series analysis to test the null hypothesis that a unit root is present in an autoregressive (AR) time series model. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity.

DEMAND FORECASTING :- BASIC TERMINOLOGY[CONTD.]

IV. Autocorrelation : Autocorrelation is a statistical method used to measure the similarity between a time series and a lagged version of itself over successive time intervals. It is used to identify patterns and trends in time series data that would otherwise go unnoticed.

V. **Partial Autocorrelation** : Partial autocorrelation is a statistical method used to measure the correlation between a time series and a lagged version of itself after controlling for the effects of the other lags.

VI. **Akaike Information Criterion** : The Akaike Information Criterion (AIC) lets you test how well your model fits the data set without over-fitting it. The AIC score rewards models that achieve a high goodness-of-fit score and penalizes them if they become overly complex . By itself, the AIC score is not of much use unless it is compared with the AIC score of a competing model . The model with the lower AIC score is expected to strike a superior balance between its ability to fit the data set and its ability to avoid over-fitting the data set.

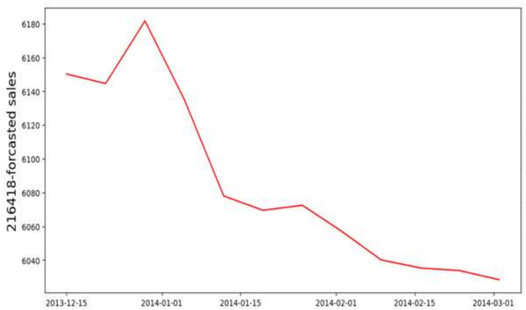
DEMAND FORECASTING :- BASIC TERMINOLOGY[CONTD.]

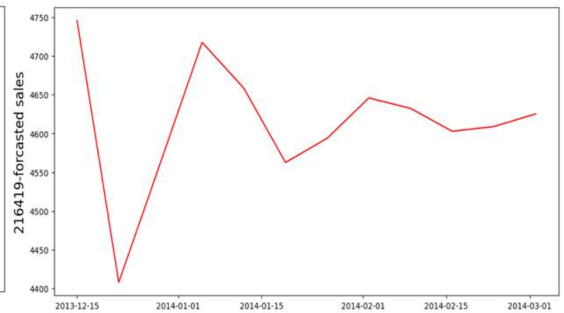
VII. Bayesian Information Criterion : Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike information criterion (AIC). When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC. The models can be tested using corresponding BIC values. Lower BIC value indicates lower penalty terms hence a better model.

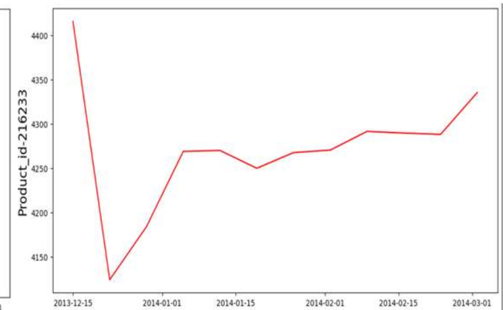
VIII. Arima/Sarima Model : An Autoregressive Integrated Moving Average (ARIMA) model is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. It's a generalization of an autoregressive moving average (ARMA) model. ARIMA models are widely used in technical analysis to forecast future security prices. However, they implicitly assume that the future will resemble the past. Therefore, they can prove inaccurate under certain market conditions, such as financial crises or periods of rapid technological change. Now if there was a seasonal component then we have to use Sarima Model where it takes into account of seasonal components. In other words, it is a seasonal arima model.

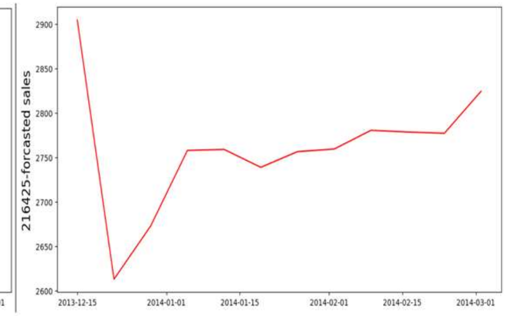
BIA BOSTON INSTITUTE OF ANALYTICS

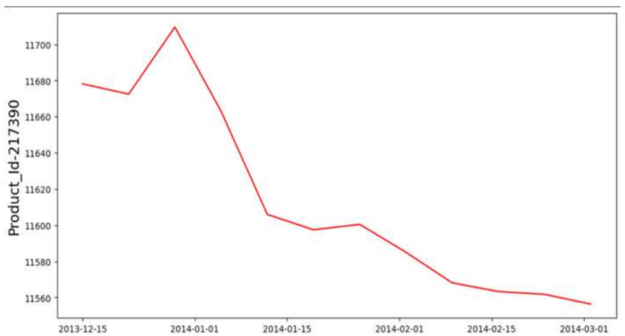
DEMAND FORECASTING :- GRAPHS , GRAPHS & MORE GRAPHS!

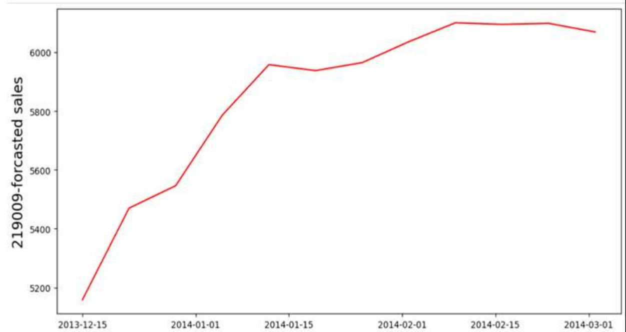


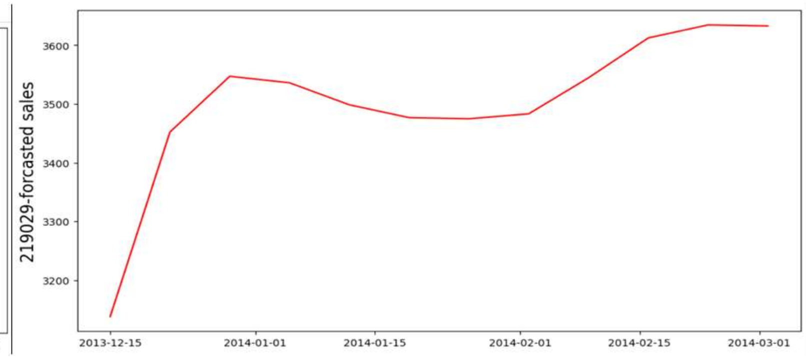




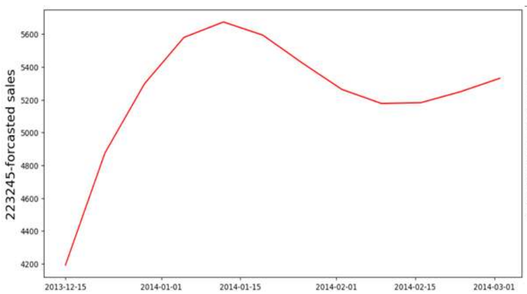


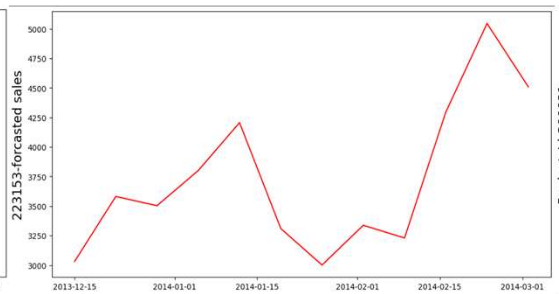


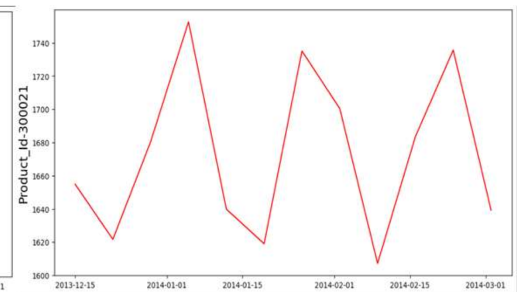


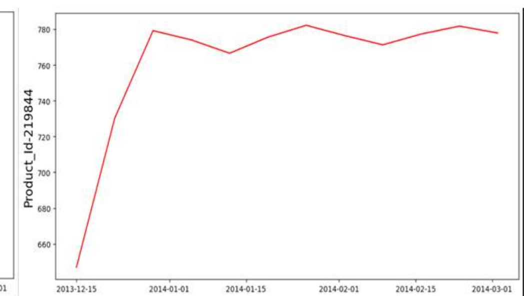


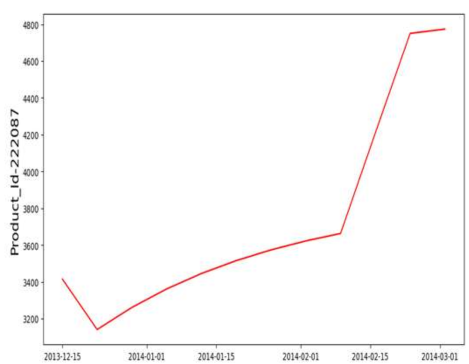
DEMAND FORECASTING :-
GRAPHS , GRAPHS & MORE GRAPHS!

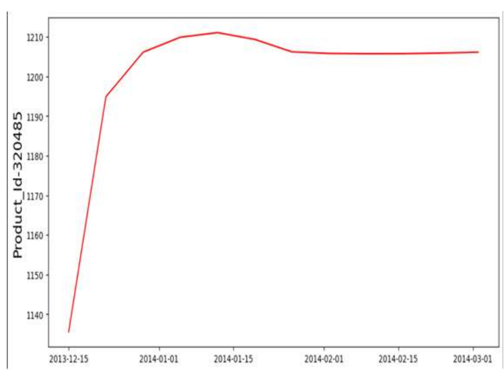


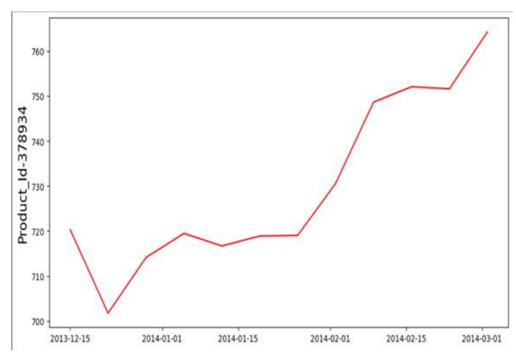






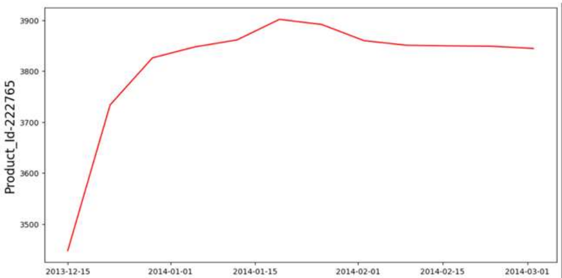


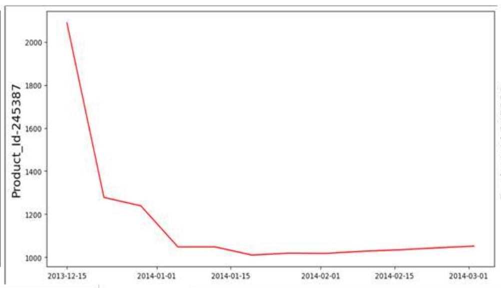


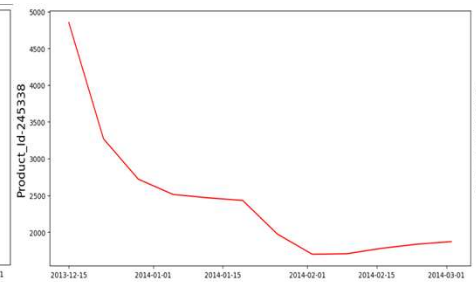


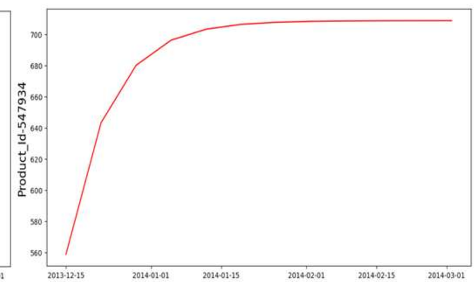
BIA BOSTON INSTITUTE OF ANALYTICS

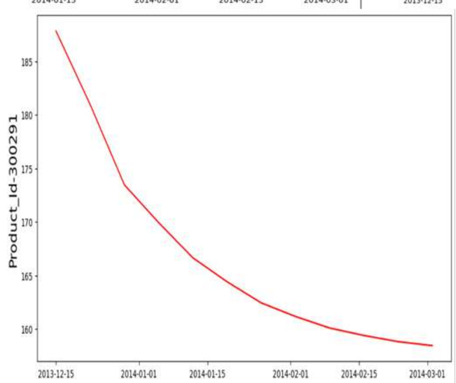
DEMAND FORECASTING :- GRAPHS , GRAPHS & MORE GRAPHS!

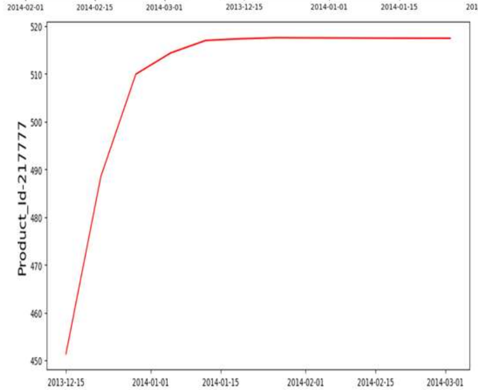




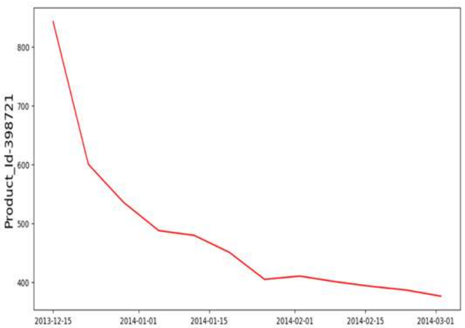


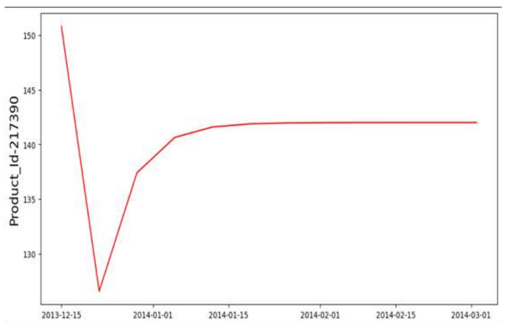


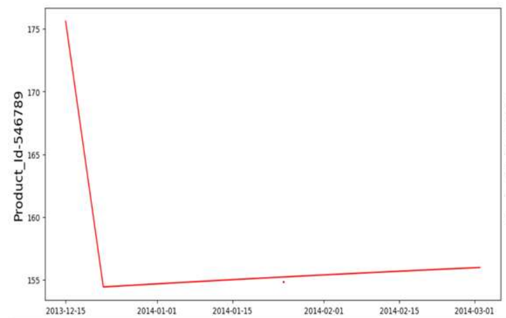


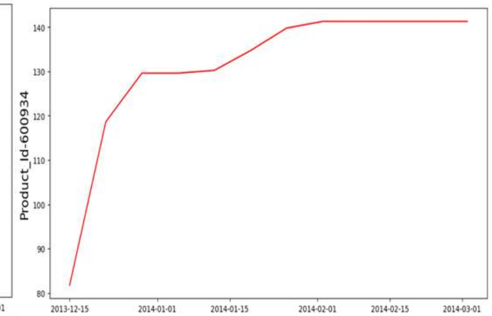


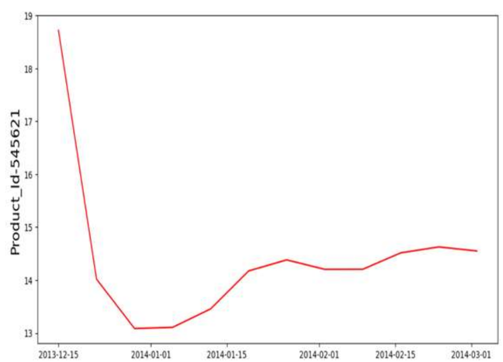
DEMAND FORECASTING :-
GRAPHS , GRAPHS & MORE GRAPHS!

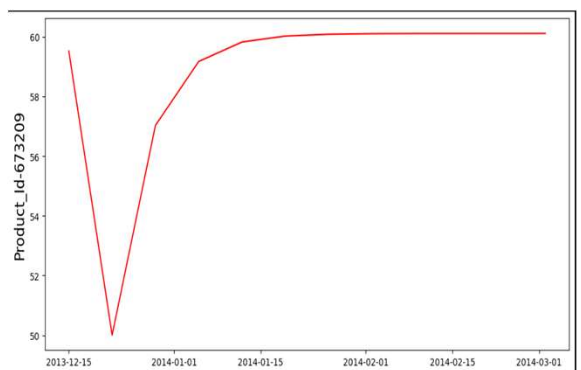


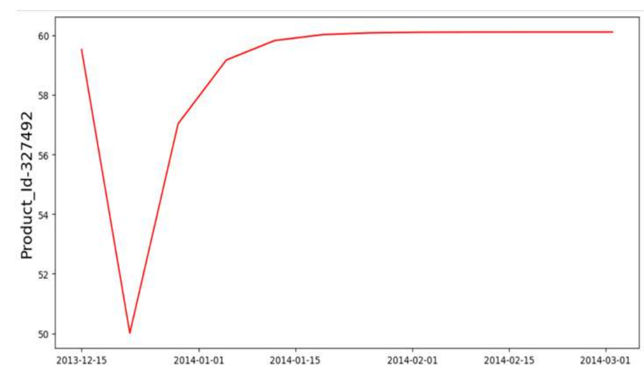












INVENTORY OPTIMIZATION :- SOME TERMINOLOGY

BIA BOSTON INSTITUTE OF ANALYTICS ®

1. Lead Time : Lead time in inventory management is the **amount of time between when an order is placed to replenish inventory and when the order is received.** It affects the amount of stock a company needs to hold at any point in time. Lead time includes the time it takes to place the order, receive the shipment, and process it into the company's inventory management system. Order lead times can vary between suppliers.

2. Service Level : Service level is a **measure of how well an inventory system can avoid stock-outs and meet customer demand.** It is the probability that the inventory on hand during the lead time is sufficient to cover the expected demand. Service level reflects a trade-off between the cost of inventory and the cost of stock-outs, which can result in missed sales, lost opportunities, and customer dissatisfaction.

3. Holding Cost : **Holding cost is the cost associated with storing inventory that remains unsold.** It is one component of total inventory costs, along with ordering & shortage costs. Holding cost, also known as the carrying cost of inventory, refers to the cost that an entity incurs for handling & storing its unsold inventory during the accounting period.

INVENTORY OPTIMIZATION :- SOME TERMINOLOGY [contd.]

IV. Stockout Cost : Stockout cost is the lost income and expense associated with a shortage of inventory. received. A business can avoid stockout issues by maintaining a high level of inventory record accuracy and a reasonable safety stock level that is adjusted to match ongoing changes in customer demand.

V. Optimal Order Quantity : The optimal order quantity refers to the quantity of a product that should be ordered from suppliers when the inventory level reaches a certain point.

VI. Reorder Point : The reorder point is the inventory level at which a new order should be placed to replenish stock before it runs out.

VII. Safety Stock : Safety stock is the additional inventory kept on hand to account for uncertainties in demand and supply. It acts as a buffer against unexpected variations in demand or lead time.

VIII. Total Cost : The total cost represents the combined costs associated with inventory management.

INVENTORY OPTIMIZATION :- AN EXTRA TOPIC : NEWSVENDOR MODEL

BOSTON INSTITUTE OF ANALYTICS

- The newsvendor model, also known as the newsboy model, is a mathematical model in operations management and applied economics. It's used to determine the optimal inventory levels in scenarios where demand is uncertain. The model gets its name from the situation of a news vendor trying to decide how many newspapers to order for the day, given that demand can vary and unsold newspapers have no value the next day. The goal of the newsvendor model is to balance the cost of ordering too many items (which then go unsold) against the cost of ordering too few (and missing potential sales). It's particularly useful for perishable products or items with a limited selling period.
- The newsvendor formula can be used to calculate the optimal order quantity, which is given by the formula:

$Q^* = F(1 - (P - C)/F)$

where Q* is the optimal order quantity, F is the cumulative distribution function of demand, P is the selling price, and C is the unit cost

INVENTORY OPTIMIZATION :- WHAT'S THE PLAN?

THE PLAN IS AS FOLLOWS:

i. As we have no data available related to the inventory so we created it by using cumulative frequency of more than type. We take the rolling mean as our sold unit measure & determine cumulative frequency of more than type to get the inventory data.

ii. We take some assumptions like :
- Lead Time = 1
- Service Level = 0.95
- Holding Cost = 0.1
- Stockout Cost = 10

iii. Lastly we determine Optimal Order Quantity , Reorder Point , Safety Stock , Total Cost of every 28 product available in our data & combine it into a single dataframe to get some insights.

INVENTORY OPTIMIZATION : INFORMATION MINING

$ The average Optimal Order Quantity of the whole retail chain is just 5389.3 that means they order 5389 products on average from suppliers When the average inventory reaches a certain point.

$ The average Reorder point of the whole retail is 5382.6 that means they replenish the inventory when average inventory level reaches that mark.

```
optimization.describe().mean()
```

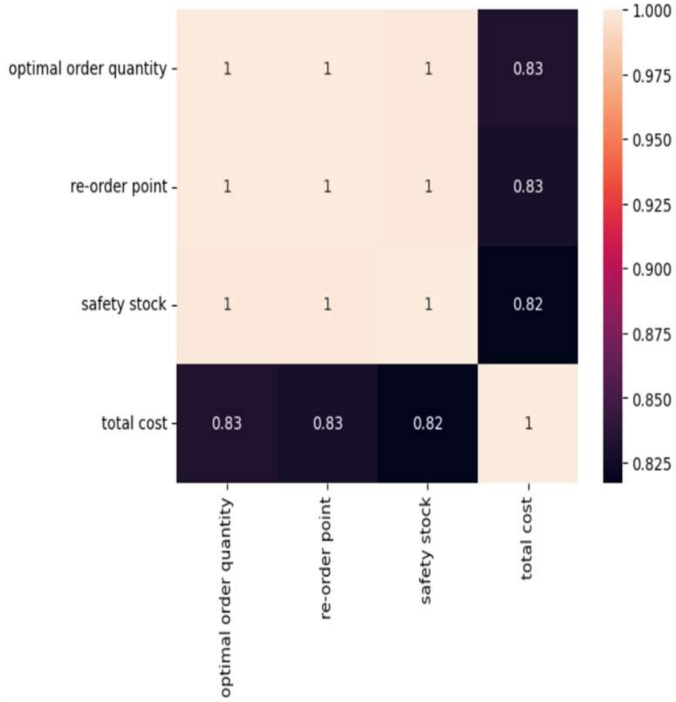
```
optimal order quantity      5389.297291
re-order point              5382.609753
safety stock                2649.355872
total cost                 32218.510540
dtype: float64
```

$ The average Safety Stock of the whole retail is 2649.3 that means they hold only 2649 products when any disruptions happens on supply or demand or on both.

By combining all these information we can clearly that the retail chain which has 76 unique stores & 28 unique products , are not on high demanding. These indices clearly indicating on this statement .

INVENTORY OPTIMIZATION : INFORMATION MINING[contd.]

From the heatmap stated beside we clearly can say that the 4 indices are highly correlated with other . That means if any one's value is changed then remaining 3 indices also will be affected



INVENTORY OPTIMIZATION : INFORMATION MINING[contd.]

There are some products like sku_id = 217390 , 216418 , 219009 , 223245 , 216419 have the high demand as their optimal order quantity is quite high , actually they are top 5 products according to that index .

But among these 5 , 4 products are there such that they creating also high cost .

There are no such products that the optimal order quantity is big enough & also the total cost is low enough & from which they can create a profit .

```
optimization['optimal order quantity'].nlargest(n=5)
```

```
sku_id
217390    23173
216418    12117
219009    11182
223245     9804
216419     9107
```

```
optimization['total cost'].nlargest(n=5)
```

```
sku_id
219009    109910.31
216418     92249.02
223245     81225.83
222087     78719.84
216419     72288.25
```

REFERENCES

- Demand Forecasting and Inventory Optimization using Python | Aman Kharwal (thecleverprogrammer.com)
- Python | Pandas dataframe.ffill() – GeeksforGeeks
- How to Calculate a Rolling Average (Mean) in Pandas • datagy
- Stack Overflow - Where Developers Learn, Share, & Build Careers
- Time Series forecasting using Auto ARIMA in python | by Sushmitha Pulagam | Towards Data Science
- Newsvendor Model - Tutorial (vskills.in)

ACKNOWLEDGEMENT

I would like to express my special thanks to BIA to give me this opportunity , to our project allocator Mr. Suvom Shaw & lastly my personal mentor Mr. Sandeep Banerjee, for yours time & efforts you're provided throughout this project . Yours useful advice & suggestions were really helpful to me during the project's completion. In this aspect, I'm eternally grateful to yours.

Also, I would like to take this opportunity to thank my family members & supporters , without them it could not have seen done effectively in such a short period of time. I can not forget their love & support.

A project is a bridge between theoretical & practical learning & with this thinking I worked on the project & made it successful due to timely support & efforts of all who helped me.

THANK YOU

BOSTON INSTITUTE OF ANALYTICS