Convolve PS1

NOOB CODERS

ANKITA ANAND
NAVIN PATWARI

Cisco Inventory Management and demand prediction

OUR APPROCH

We went through a number of processes and this is how we made to the conclusion

PLANNING

Clearing what are the objectives we should be focused on.

EXPLORING DATA

Analyzing data to bring out valuable insights..

MODEL TRAINING

Training various models on our data selected via analyzing literature work on the topic

DRAWING CONCLUSIONS

Predicting the value using the machine learning models.

Objective

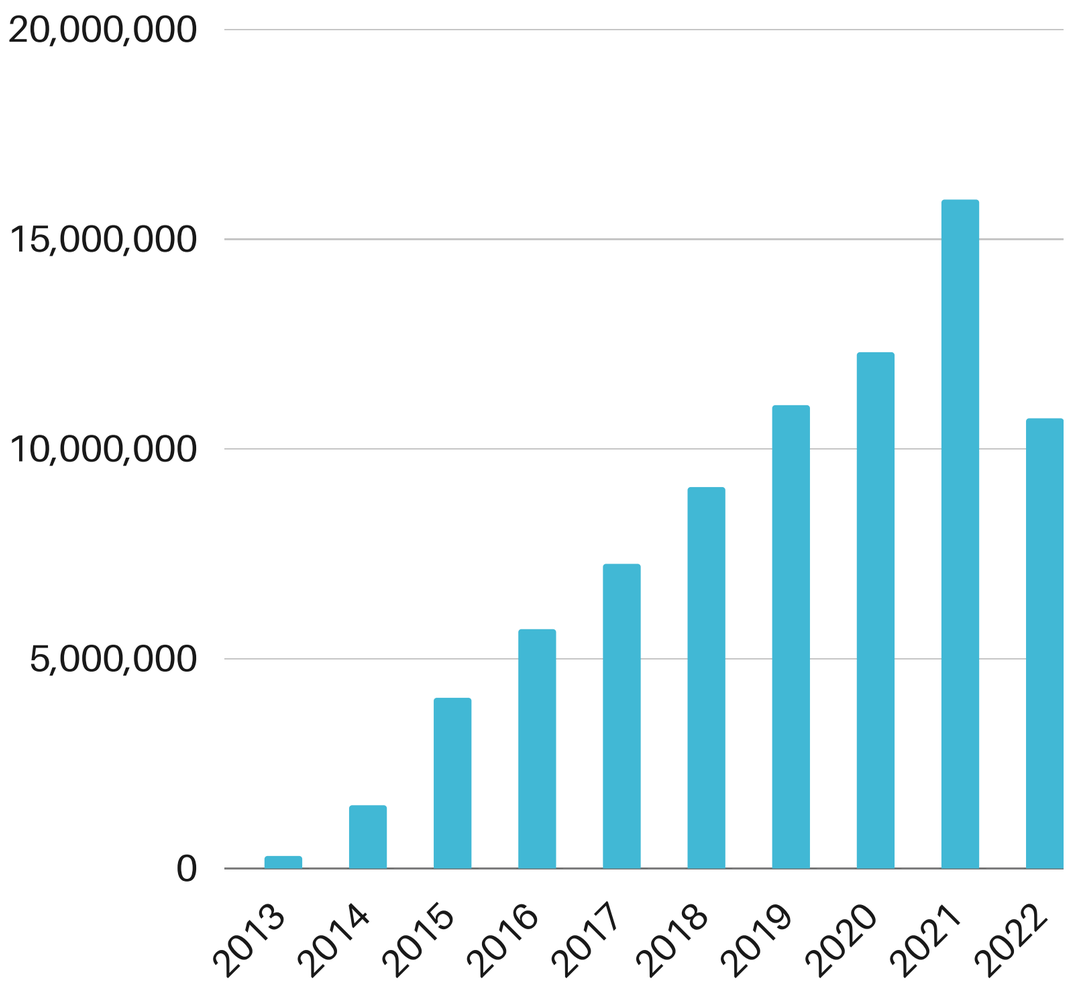
Our goal is to predict the demand for the given 'Product ID'~PLID in the upcoming quarters using the past year's sales data. This comes under demand forecasting.

1	To predict the demand for the given 'Product ID'~PLID in the upcoming quarters using the past year's sales data.
2	To ascertain which product should be stocked in which storage facility.
3	Plot the predicted values using appropriate graphs.

Data Analysis

Visualizing Data through Graphs

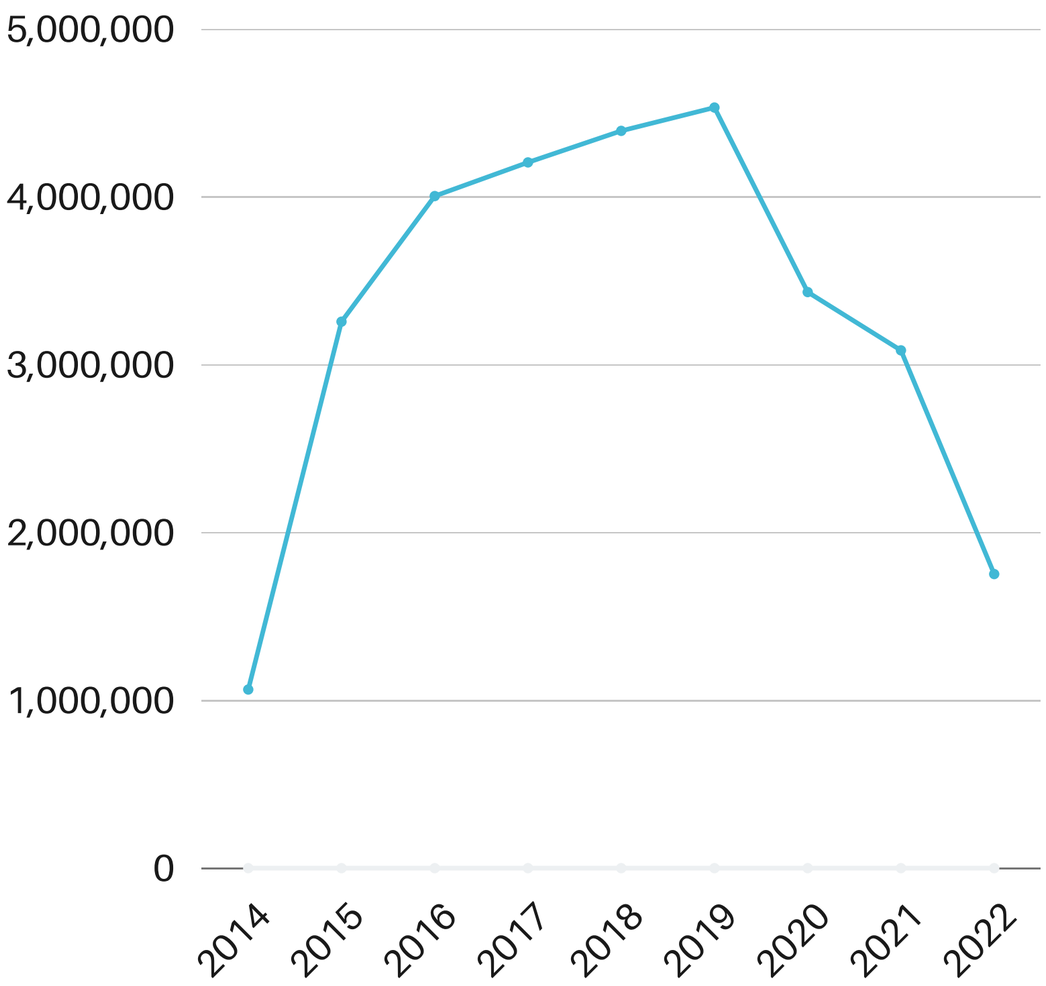
Cleaning the data and using relevant graphs for bringing deep insights from data.



TOTAL SALES QUANTITY

A brief history of total sales growth over years

From the data here, it is clearly visible how cisco's sales quantity over time has grown. The year 2022 has lower sales quantity because we don't have data for all the months of this year.
Also, it should be noted that we have considered booking dates' year because it seems that there is some human error in the other columns.



SALES OF PHONVOC

chart of the most popular product family by booked quantity

The most sold product family by booked quantity over years had its peak in 2019. It may be possible that covid-19 has caused this sudden drop or we launched a better version of the product causing a decline in sales of outdated versions.

OVERVIEW OF ML MODELS

Models used in demand forecasting

Ridge Regression

a method of estimating the coefficients of multiple-regression models where the independent variables are highly correlated.

Gradient Boosting

calculates the difference between the current prediction and the known correct target value.

XGB Regressor

Expects to have the base learners which are uniformly bad so that when all the predictions are combined, bad predictions cancel out

Lightgbm

extends the gradient boosting algorithm by adding a type of automatic feature selection as well as focusing on boosting gradients.

MODEL EVOLUTION

1

Linear Regression

Trained model after basic preprocessing of the given data.
Accuracy: 56%

2

Ridge Regression

Tried to enhance the performance by cleaning data and using a more sophisticated model.
Accuracy: 69%

3

Feature Engineering

Derived new columns e.g., Average value of sales quantity by product id and sale year.
Accuracy: 72%

4

Gradient Boosting

Improved performance through better models and more thorough data analysis.
Accuracy: 82.36%

5

LighGBM Regressor

Regressor through data filtering and advancement.
Accuracy: 85%

6

Final Result

Took the average of models so as to improve the results and used it to predict the results.
Accuracy: 86%

Feature Engineering

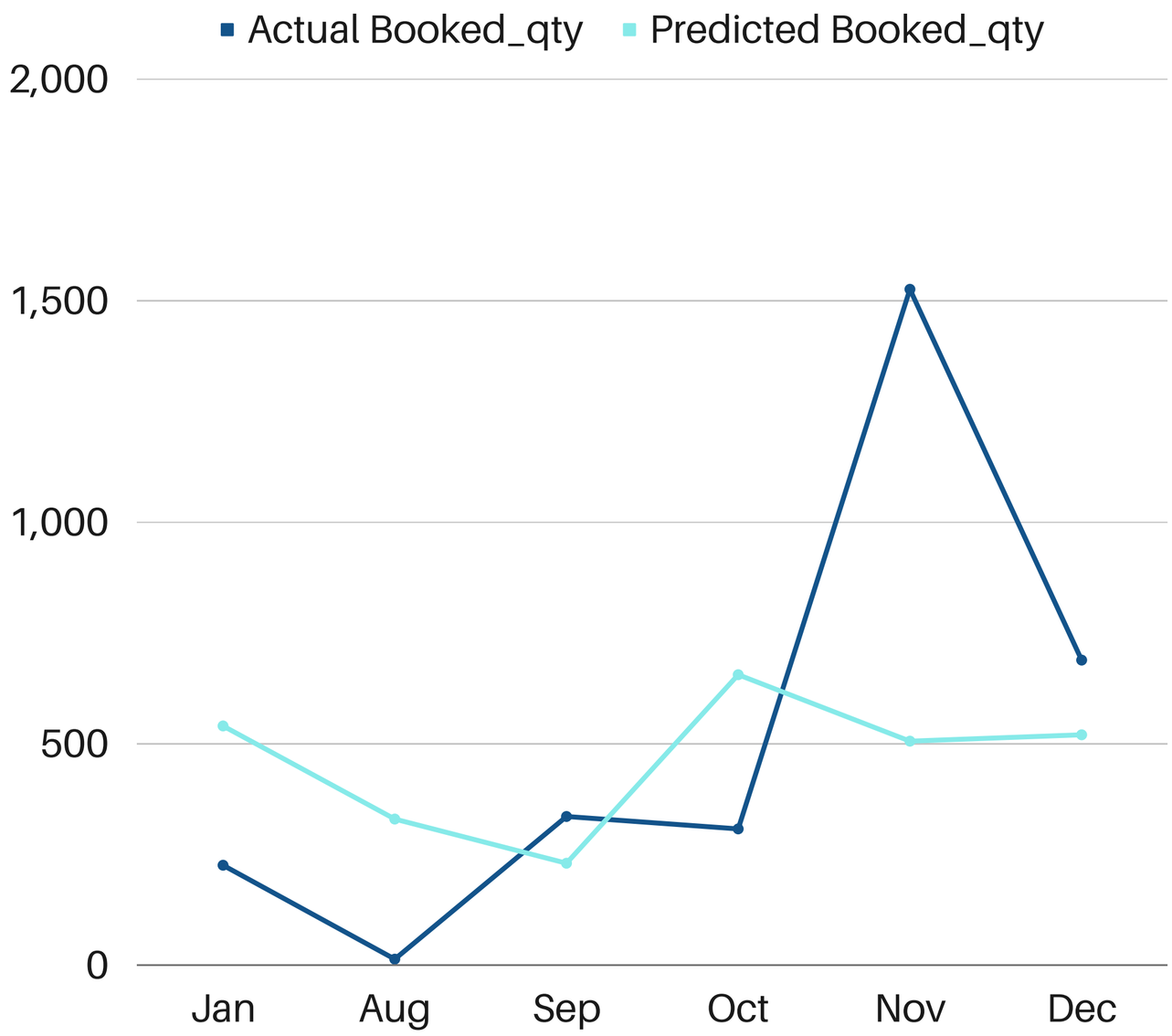
We created two extra features to better capture the flow of the data .
We added namely "qty_lag" and "mean_booked_qty" columns into the dataset.
Since the data given to us will not necessarily follow all the points in the time series, when we analyzed the data we found it to be not seasonal that's why to ensure our model doesn't fail to recognize the impact of previous sales we additionally added two features from our side.

"QTY_LAG"	This feature captures the previous month's booking quantity of the product.
"MEAN_BOOKED_QTY"	This feature captures the annual product wise mean of booked qty. While training ,our model used " { (**previous mean * 11)+"lag_qty"}/12 "** this expression to move forward.

Hyperparameter Tuning

Randomized SearchCV

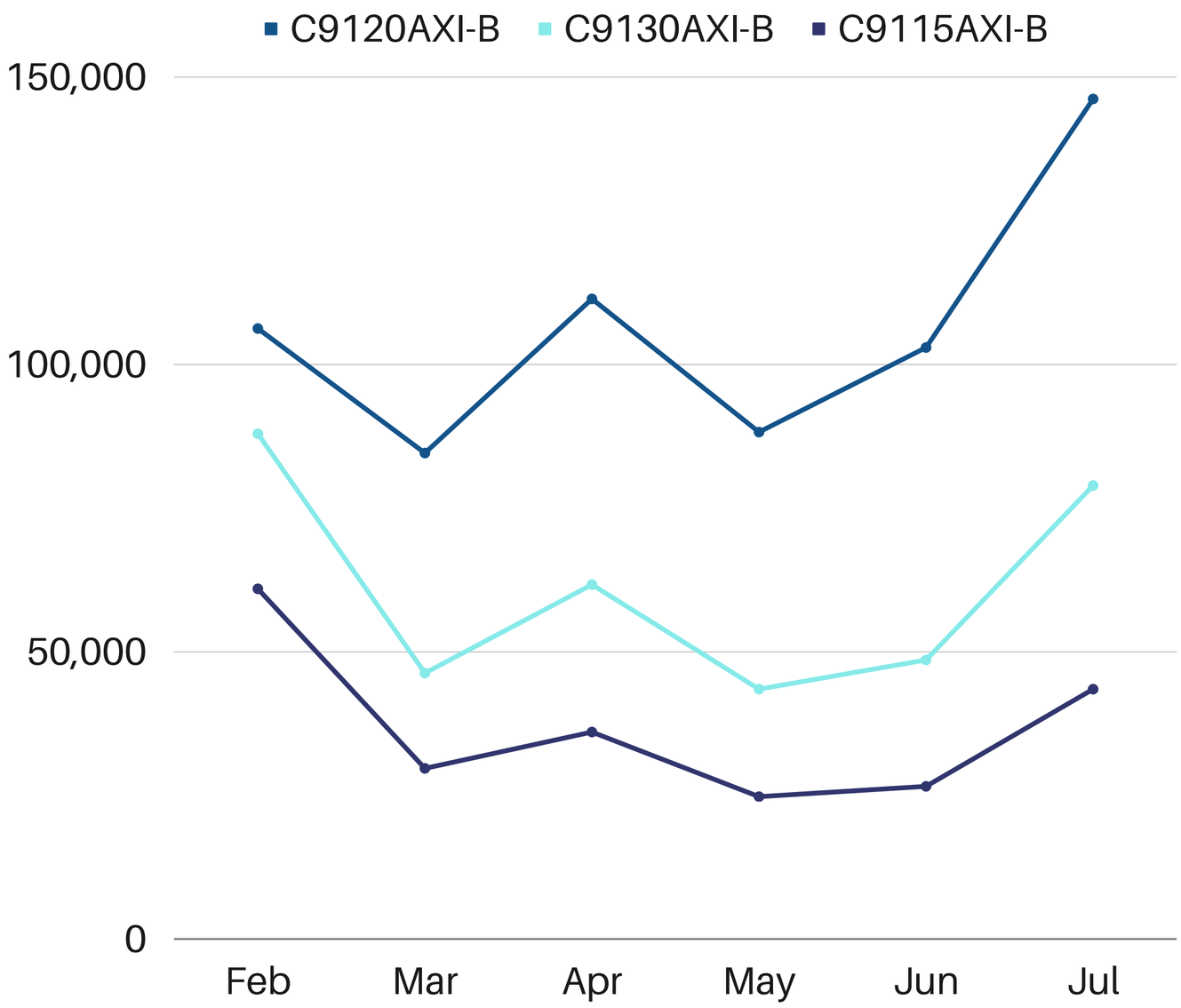
RandomizedSearchCV randomly passes the set of hyperparameters and calculate the score and gives the best set of hyperparameters which gives the best score as an output. So this is the recipe on How we can find optimal parameters using RandomizedSearchCV for Regression.



COMPARISON OF ACTUAL BOOKED QUANTITY AND MODEL GENERATED PREDICTED BOOKED QUANTITY FOR THE PLID "UCS-S3260-HDW18T" IN 2022.

X-Axis: Month

Y-Axis: Booked Quantity



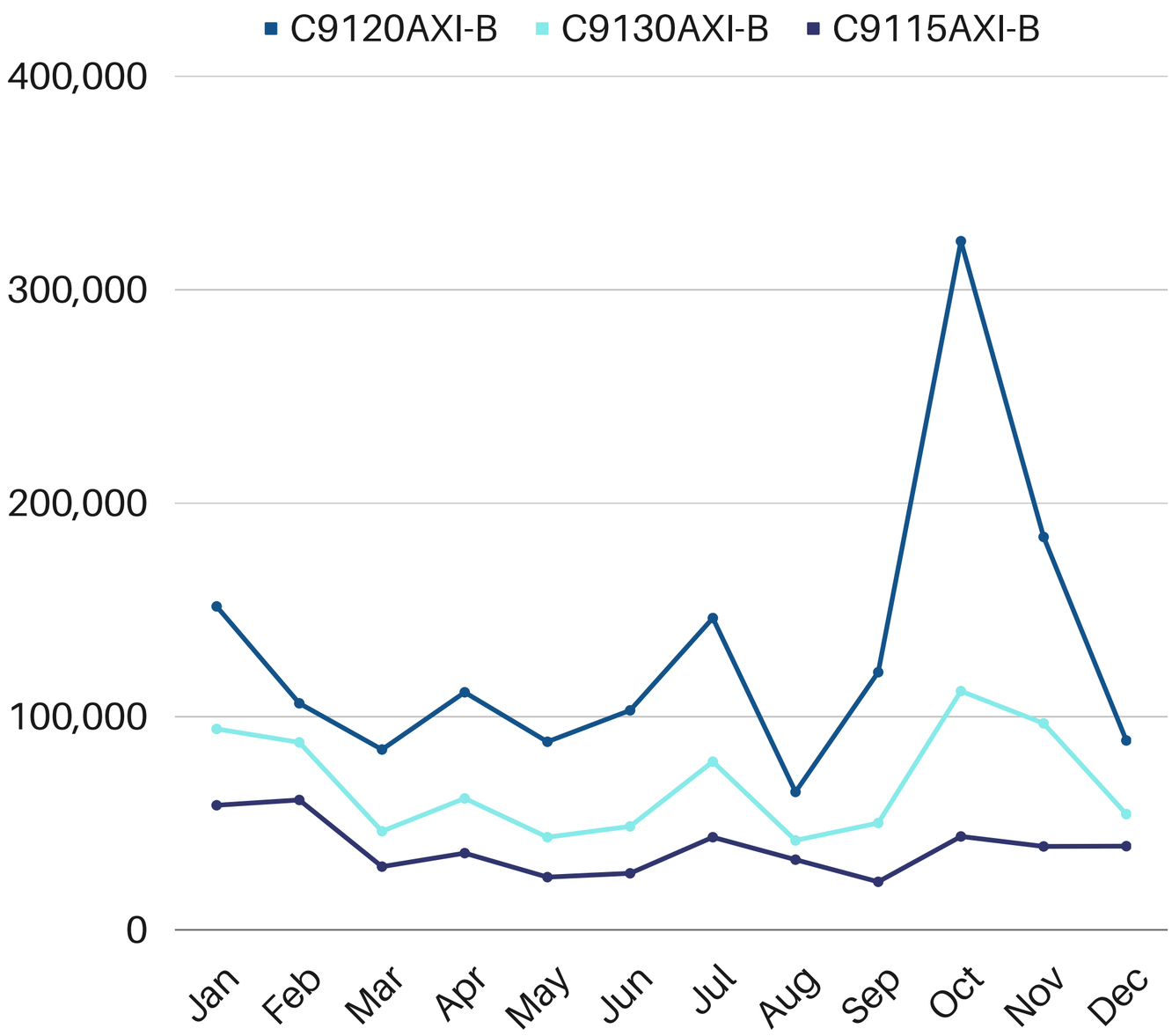
COMPARISON OF TOP 3 PRODUCTS BASED ON PREDICTED SALES DATA IN 2022.

ALL OF THEM ARE WIRELESS ACCESS POINTS.

(PRODUCT FAMILY- PHONVOC)

X-Axis: Month

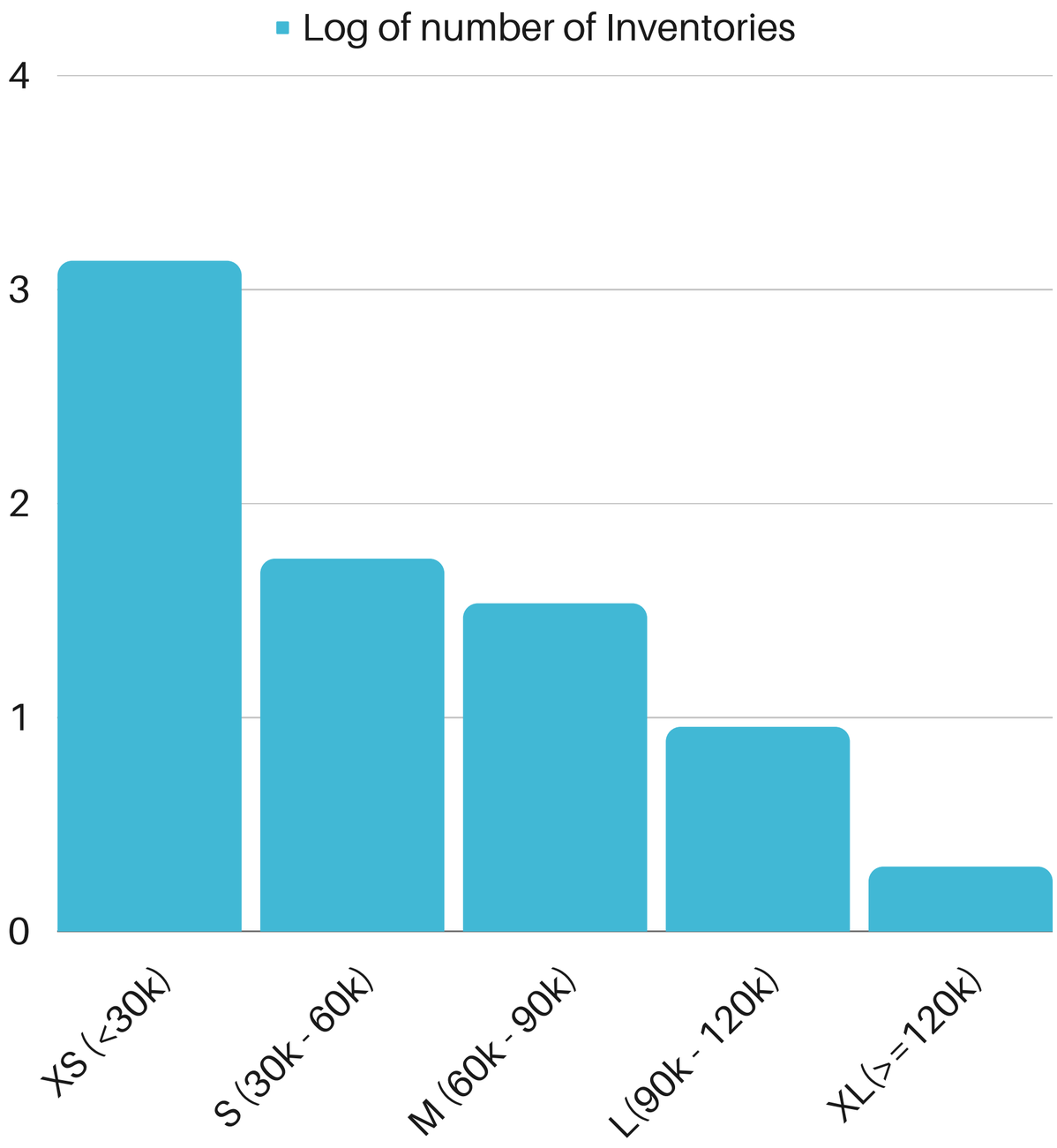
Y-Axis: Predicted Booked Quantity



COMPARISON OF TOP 3 PRODUCTS BASED ON SALES DATA FOR YEAR 2022.
ALL OF THEM ARE WIRELESS ACCESS POINTS.
(PRODUCT FAMILY- PHONVOC)

X-Axis: Month

Y-Axis: Predicted Booked Quantity



LOG OF NUMBER OF INVENTORIES REQUIRED IS TAKEN IN LINEAR INTERVALS

X-Axis: Size of Inventory
Y-Axis: Log of number of Inventories

Conclusion

We used the models listed and then calculated an average to arrive at our results. Our model can predict the results with an accuracy of around 86 %.

Link to the dataset and Problem statement.

- cs.co/PS1

Link to our dashboard

- https://drive.google.com/file/d/1UKO8SRnTvQxGwaHI-QTmnEG8QRTUIH-T/view

Link to Jupyter notebook

- https://drive.google.com/file/d/1cxIdmwmKm93Verqqzx5yqLbgOVOIs83-/view?usp=sharing