Sales Volume Forecasting

Zachary Corbett
Victor Dontsov
Sara Parveen
Set

Background

* Client: Product distributor who purchases office products from manufacturers, holds them and then resells them to its business customers

* Improve the inventory planning process for the product distributor

* Create a model that makes predictions about the sales volume for different product categories

Value of model

* Meet customer demand and ensure customer satisfaction

* Avoid having too much inventory which can lead to unnecessary storage costs, handling costs and cash-flow pressures

* Avoid stockouts which can result in loss of sales and/ or fines

* Maintain high profitability

Target Audience

* Product distributor whose data was analyzed

* All product sales planning, supply chain, and procurement professionals

Data Source

* The data used for this project comes in the form of CSV files obtained from the product distributor.

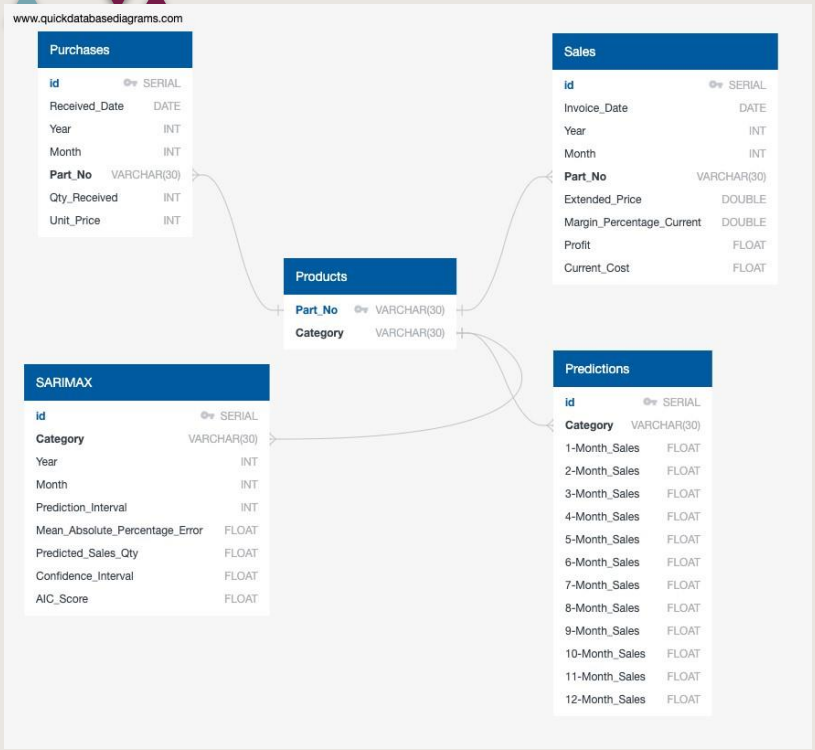
* The CSVs have 5-years worth of data (2018 to 2022) for Purchase, Sales and Product Details.

Data Hosting

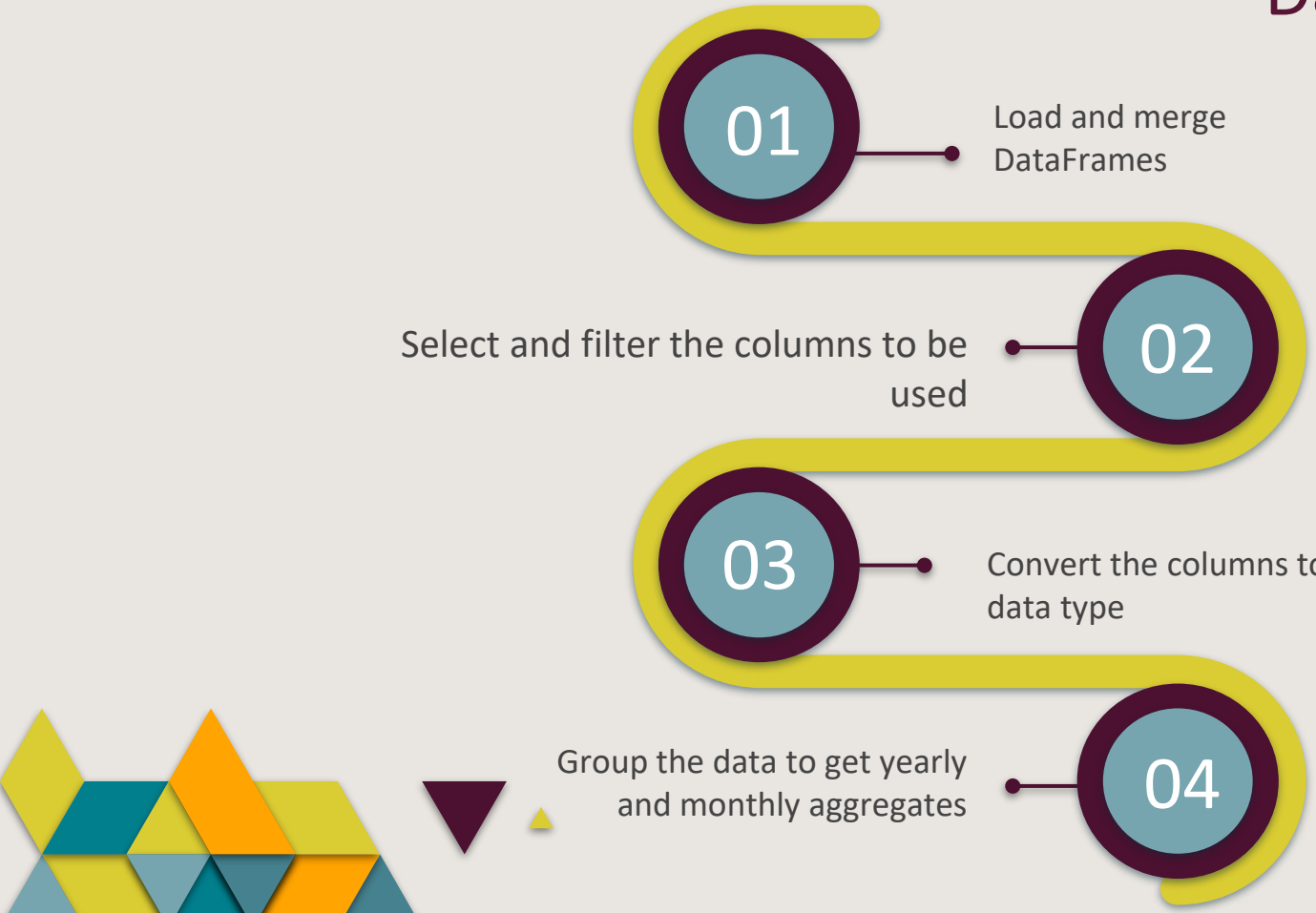
* The raw CSV data files are hosted on an S3 Bucket through Amazon AWS.

* The database schema was stored in the Databricks File System (DBFS) through a Databricks Community Edition Account. This file system is ultimately hosted on AWS without charges for computing.

Database Creation



Data Cleaning



Data Model Description

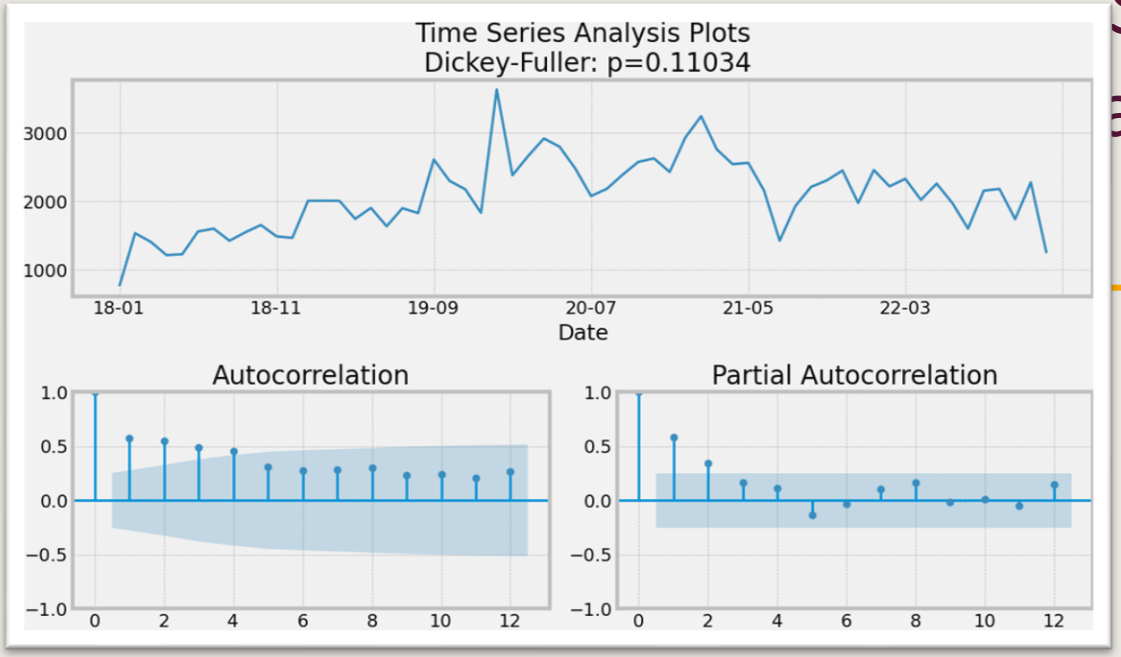
SARIMAX Model

* Time Series Analysis for Sales Volume predictions
* Trained On 4-years (2018-2021)

* Parameters:

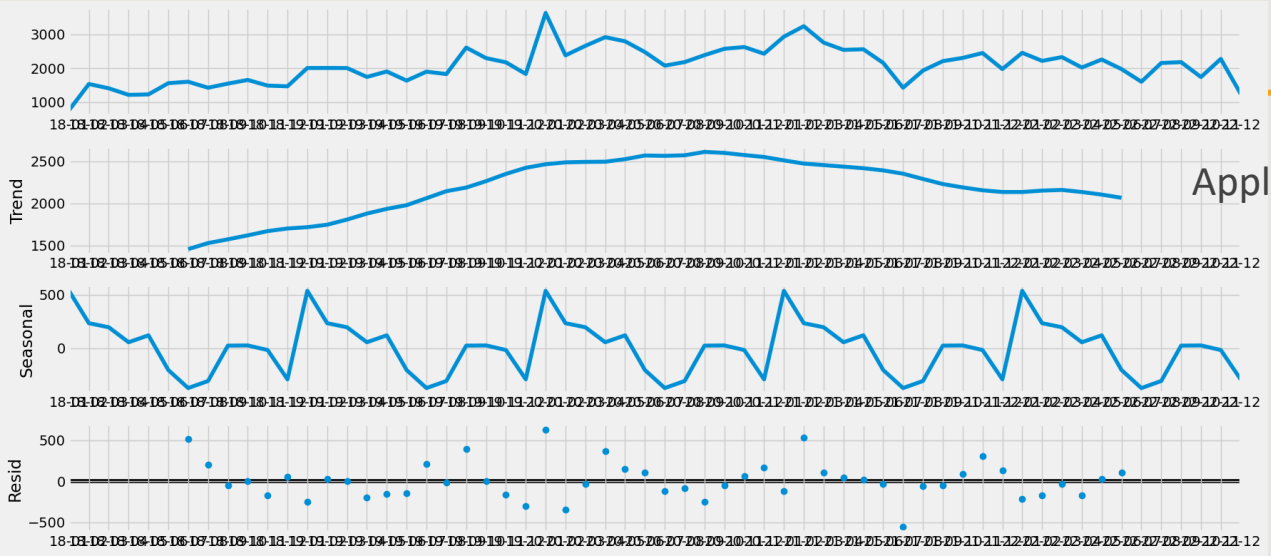
p - order of the autoregressive part
d - degree of first differencing involved
q - order of the moving average part
P, D, Q - all previous characteristics with seasonal factors
s - seasonal length in the data

Step 1: Identified the ationarity of the time series



P-Value > 0.05 implies data is not stationary

Step 2: Suggested the initial parameters



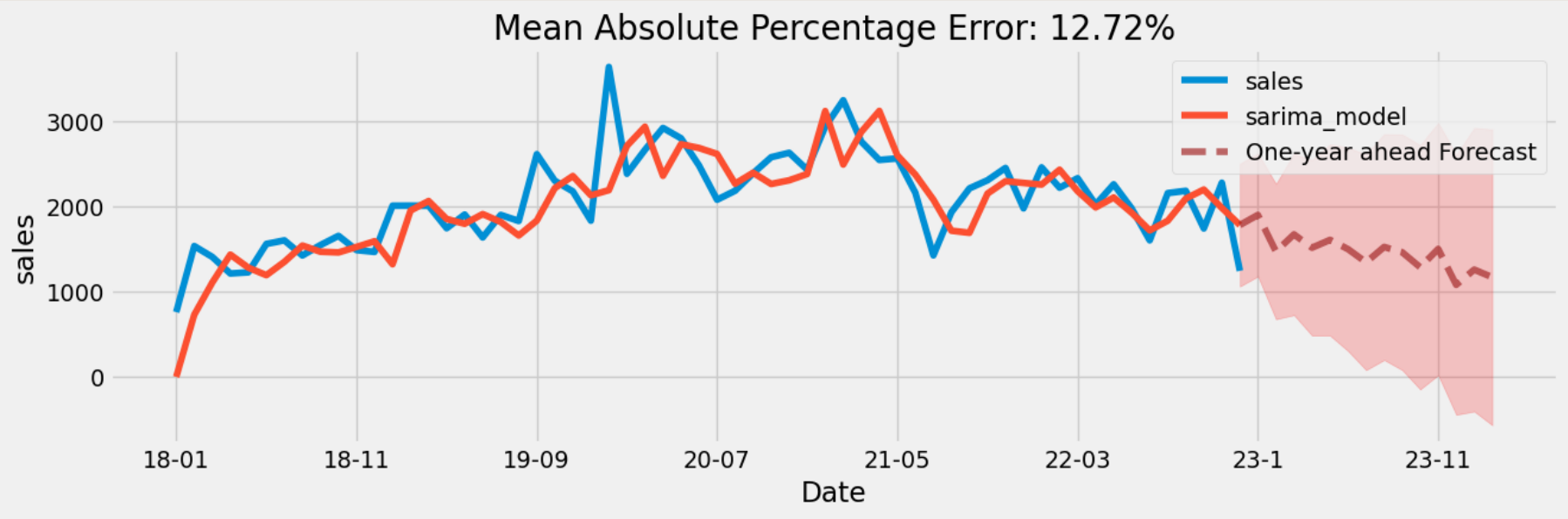
Applied differencing analysis if needed

Step 3: Generated the final parameters for the model

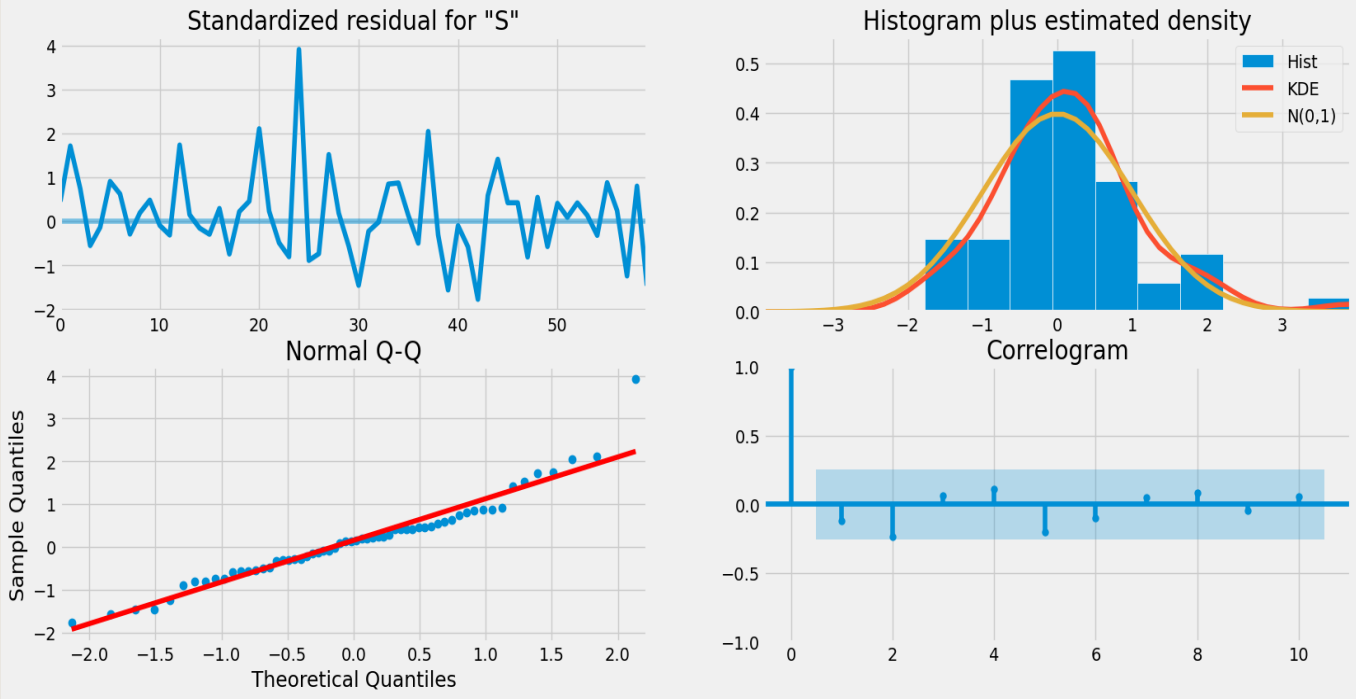
	parameters	aic
0	(2, 0, 0, 2, 0, 1)	182.932220
1	(0, 1, 1, 0, 0, 1)	697.156454
2	(0, 1, 1, 1, 0, 0)	697.705893
3	(0, 1, 1, 2, 0, 0)	699.072669
4	(1, 1, 1, 0, 0, 1)	699.142650
5	(0, 1, 2, 0, 0, 1)	699.143229
6	(0, 1, 1, 1, 0, 1)	699.156443
7	(0, 1, 2, 1, 0, 0)	699.699439
8	(1, 1, 1, 1, 0, 0)	699.699695
9	(2, 1, 0, 0, 0, 1)	700.852337

Selected parameters based on AIC Scores

Step 4: Ran the SARIMAX model



Error Analysis



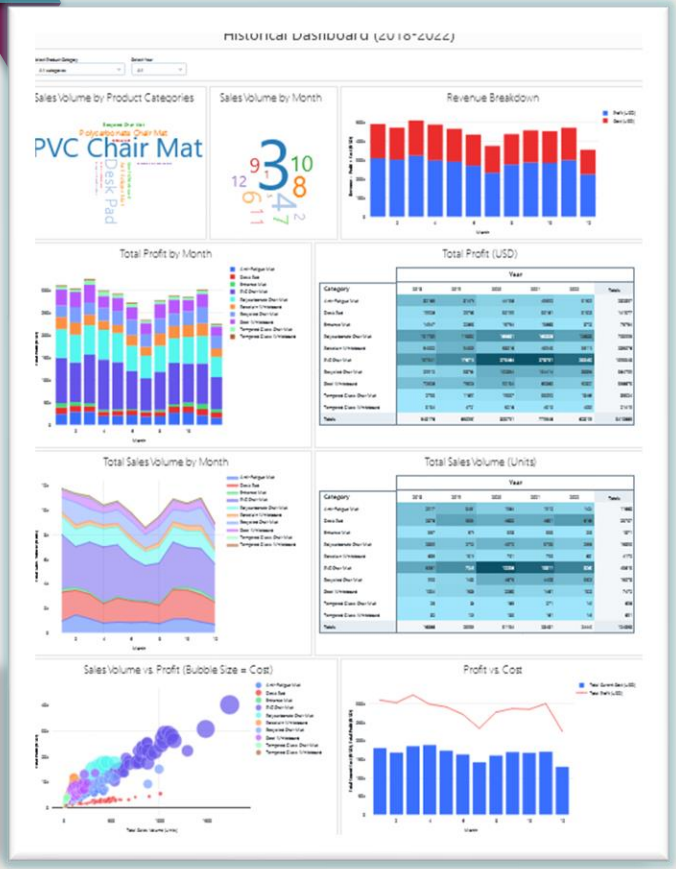
Attempt to Optimize the Model

Product Category	Mean Absolute Percentage Error (Attempt 1)	Mean Absolute Percentage Error (Attempt 2)
All categories	14.09%	12.28%
Anti-Fatigue Mat	39.86%	34.19%
Desk Pad	32.62%	37.37%
Entrance Mat	43.63%	42.40%
Polycarbonate Chair Mat	25.91%	21.24%
Porcelain Whiteboard	42.87%	48.14%
PVC Chair Mat	19.99%	20.31%
Recycled Chair Mat	41.50%	41.50%
Steel Whiteboard	31.49%	34.42%
Tempered Glass Chair Mat	41.95%	46.20%
Tempered Glass Whiteboard	44.48%	47.31%

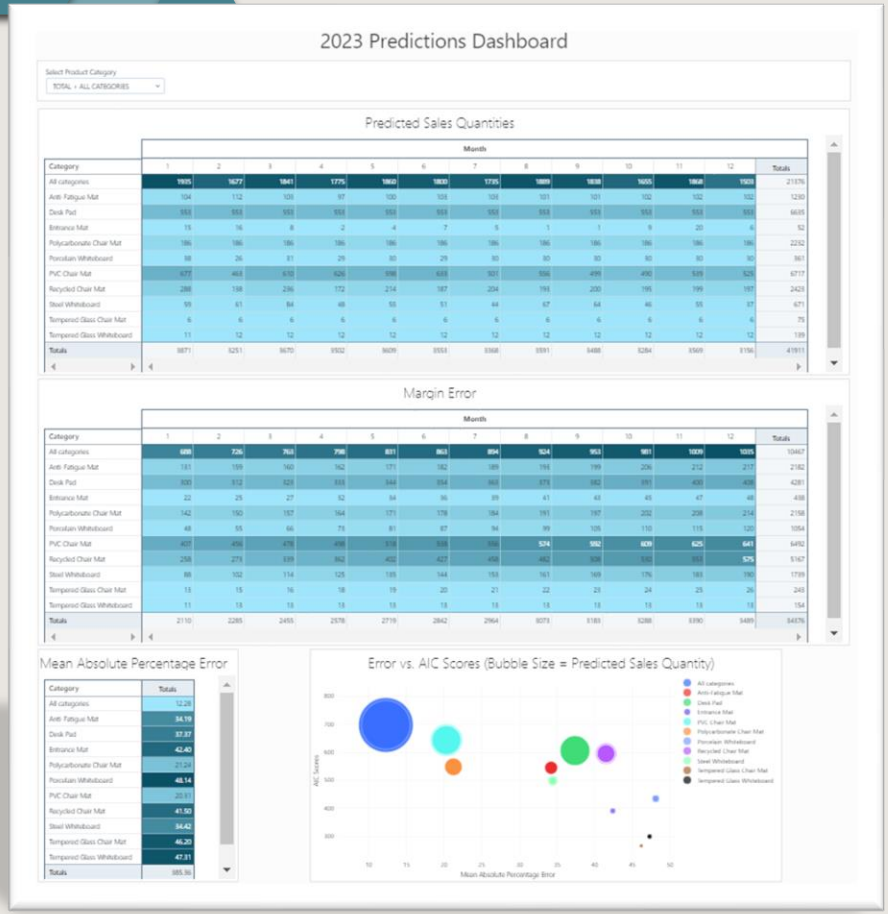
Visualizations and Dashboards

* Used Databricks Dashboards

* Created filters in the Databricks Notebooks



Historical Dashboard (2018-2022)

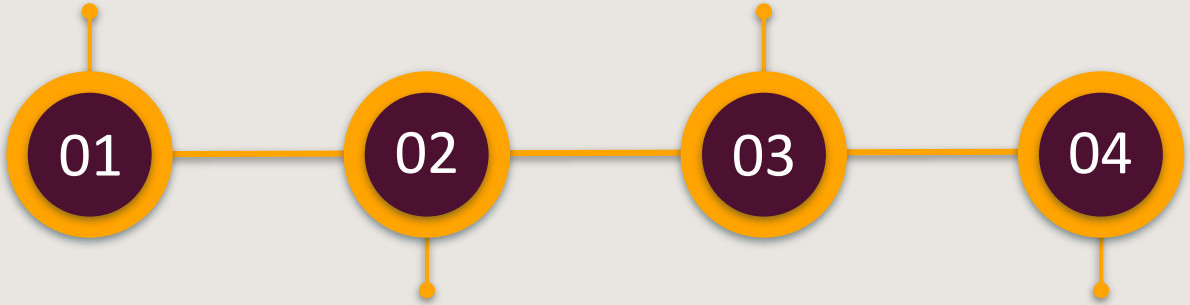


Predictions Dashboard (2023)

Limitations and Assumptions

134 different product SKUs within the original dataset

Unlimited resources and budget



Unlimited warehouse space to store the products

No minimum order quantities

Challenges

* ETL was processed in Databricks and it was truncating the data to 10,000 rows.

* The dashboards in Databricks do not have a default option for adding filters to visualizations.

* The filters on the Databricks dashboards do not carry over to the HTML file.

Conclusions

* Predictions show a slight decline in sales volume in the next year for overall sales but stable sales for some categories.

* Sales predictions are helpful but other models using special dimensions of warehouse, and budget constraints could help make more applicable predictions.

* Margin Error increases for longer time periods. This makes it more appropriate for Just-In-Time distributor.

Thank you for listening!

Any Questions?