




# 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





\* Product distributor whose data was analyzed

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

# Target Audience

---





\* 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 Source

---





# 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.



www.quickdatabasediagrams.com

#### Purchases

<b>id</b>	SERIAL
Received_Date	DATE
Year	INT
Month	INT
<b>Part_No</b>	VARCHAR(30)
Qty_Received	INT
Unit_Price	INT

#### Sales

<b>id</b>	SERIAL
Invoice_Date	DATE
Year	INT
Month	INT
<b>Part_No</b>	VARCHAR(30)
Extended_Price	DOUBLE
Margin_Percentage_Current	DOUBLE
Profit	FLOAT
Current_Cost	FLOAT

#### Products

<b>Part_No</b>	VARCHAR(30)
<b>Category</b>	VARCHAR(30)

#### SARIMAX

<b>id</b>	SERIAL
<b>Category</b>	VARCHAR(30)
Year	INT
Month	INT
Prediction_Interval	INT
Mean_Absolute_Percentage_Error	FLOAT
Predicted_Sales_Qty	FLOAT
Confidence_Interval	FLOAT
AIC_Score	FLOAT

#### Predictions

<b>id</b>	SERIAL
<b>Category</b>	VARCHAR(30)
1-Month_Sales	FLOAT
2-Month_Sales	FLOAT
3-Month_Sales	FLOAT
4-Month_Sales	FLOAT
5-Month_Sales	FLOAT
6-Month_Sales	FLOAT
7-Month_Sales	FLOAT
8-Month_Sales	FLOAT
9-Month_Sales	FLOAT
10-Month_Sales	FLOAT
11-Month_Sales	FLOAT
12-Month_Sales	FLOAT

# Database Creation

# Data Cleaning

01

Load and merge  
DataFrames

Select and filter the columns to be  
used

02

03

Convert the columns to the appropriate  
data type

Group the data to get yearly  
and monthly aggregates

04





# Data Model Description



## SARIMAX Model

---

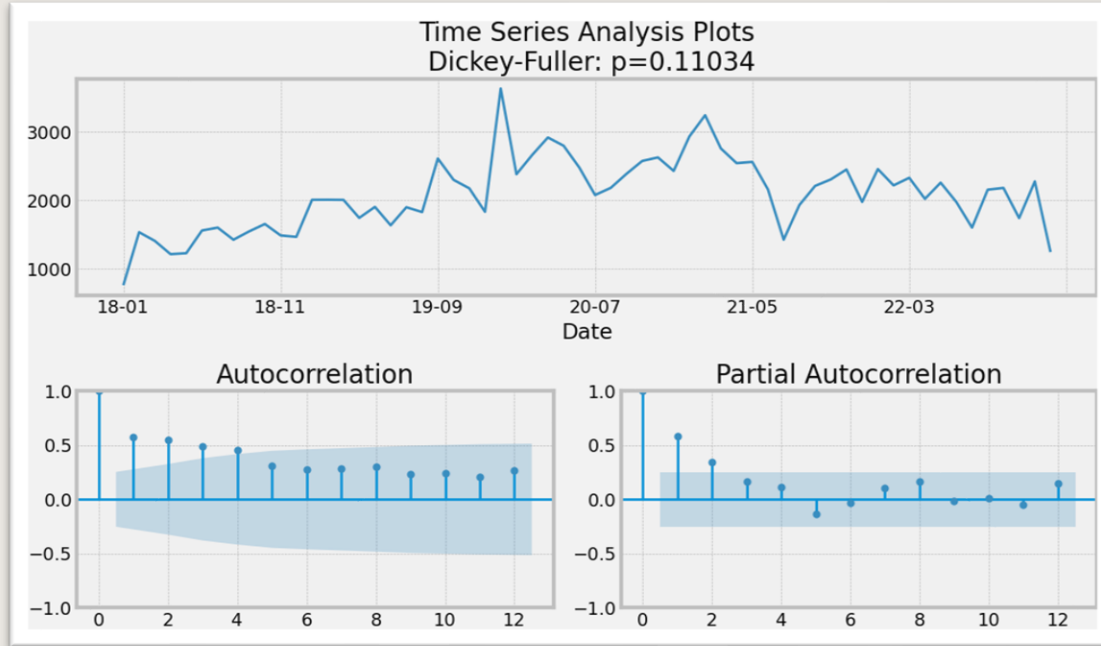
- \* Time Series Analysis for Sales Volume predictions
- \* Trained On 4-years (2018-2021)

### \* 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



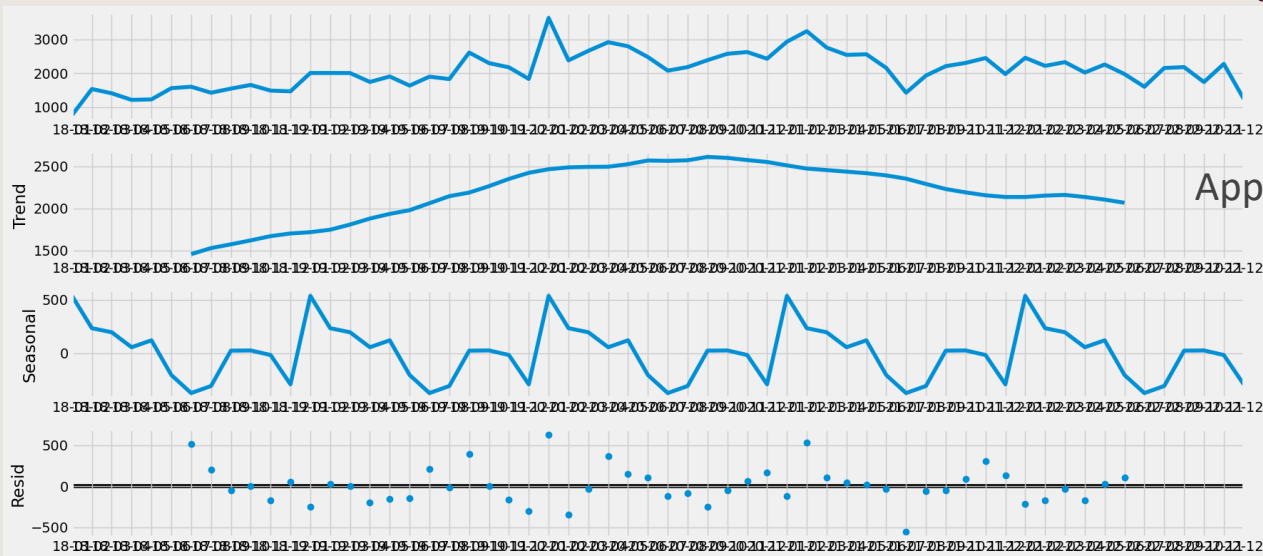
# Step 1: Identified the non-stationarity of the time series




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

---

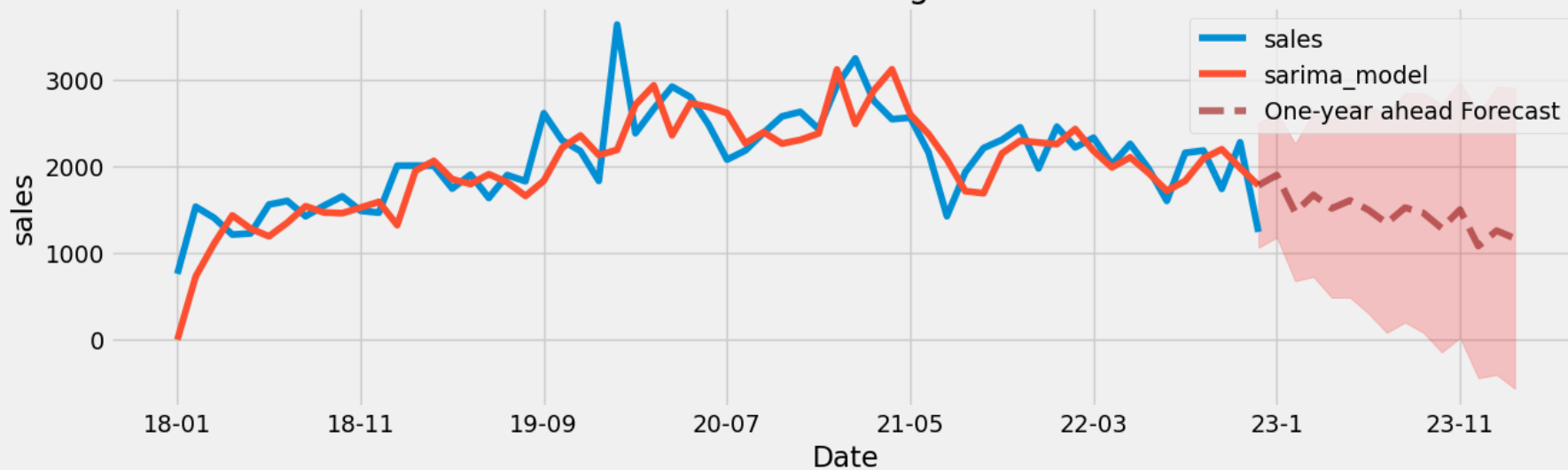
	parameters	aic
0	(2, 0, 0, 2, 0, 1)	182.932220
1	(0, 1, 1, 0, 0, 1)	697.156454
2	(0, 1, 1, 1, 0, 0)	697.705893
3	(0, 1, 1, 2, 0, 0)	699.072669
4	(1, 1, 1, 0, 0, 1)	699.142650
5	(0, 1, 2, 0, 0, 1)	699.143229
6	(0, 1, 1, 1, 0, 1)	699.156443
7	(0, 1, 2, 1, 0, 0)	699.699439
8	(1, 1, 1, 1, 0, 0)	699.699695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

Selected parameters based on AIC Scores

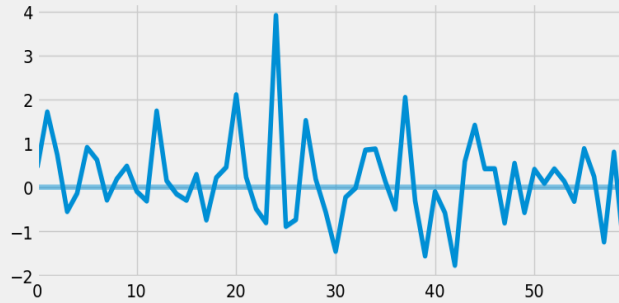


## Step 4: Ran the SARIMAX model

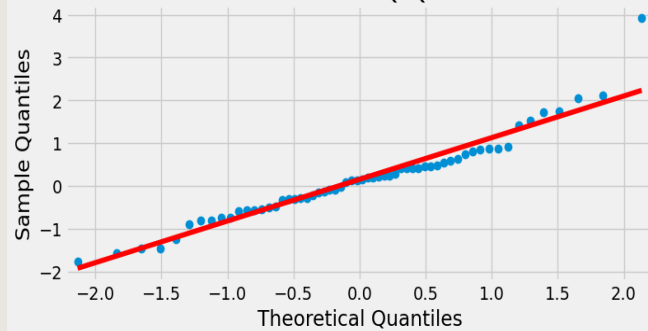
Mean Absolute Percentage Error: 12.72%



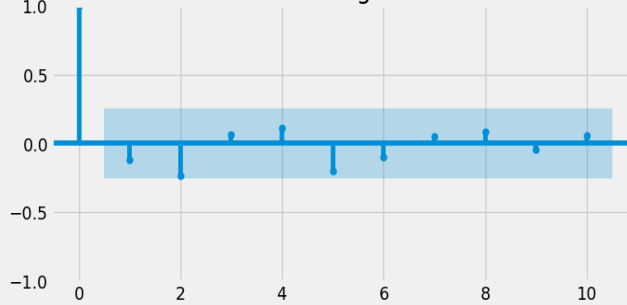
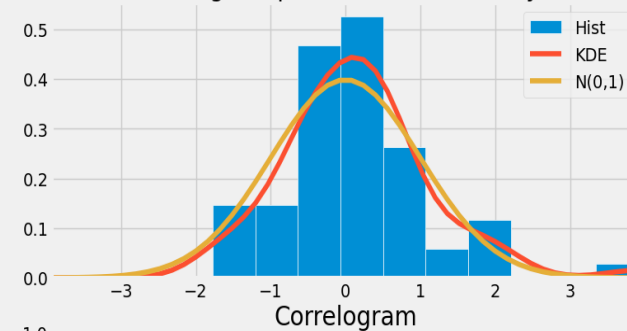
Standardized residual for "S"




Normal Q-Q



Histogram plus estimated density




# 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







## 2023 Predictions Dashboard

Select Product Category

TOTAL - ALL CATEGORIES

### Predicted Sales Quantities

Category	Months												Totals
	1	2	3	4	5	6	7	8	9	10	11	12	
All Categories	1661	1677	1661	1776	1660	1660	1716	1660	1660	1660	1660	1660	21176
Anti-Fatigue Mat	104	112	105	97	100	108	104	101	105	104	102	102	1205
Desk Pad	353	353	353	353	353	353	353	353	353	353	353	353	4605
Entrance Mat	15	16	6	2	4	7	1	1	1	9	20	4	82
Polycarbonate Chair Mat	186	186	186	186	186	186	186	186	186	186	186	186	2232
Porcelain Whiteboard	58	36	31	39	30	29	30	30	30	30	30	30	361
PVC Chair Mat	372	463	570	509	598	558	501	559	599	590	559	525	6717
Recycled Chair Mat	288	198	259	172	214	187	204	191	209	195	199	197	2425
Steel Whiteboard	99	61	84	48	55	61	44	67	64	46	55	67	671
Tempered Glass Chair Mat	6	6	6	6	6	6	6	6	6	6	6	6	75
Tempered Glass Whiteboard	11	12	12	12	12	12	12	12	12	12	12	12	139
Totals	8871	9231	9670	9502	9609	9551	9368	9391	9488	9284	9369	9196	41911

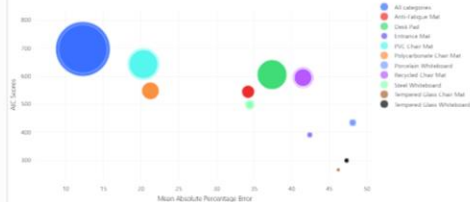
### Margin Error

Category	Months												Totals
	1	2	3	4	5	6	7	8	9	10	11	12	
All Categories	1661	1677	1661	1776	1660	1660	1716	1660	1660	1660	1660	1660	21176
Anti-Fatigue Mat	104	112	105	97	100	108	104	101	105	104	102	102	1205
Desk Pad	353	353	353	353	353	353	353	353	353	353	353	353	4605
Entrance Mat	22	25	27	12	14	16	19	17	15	17	47	46	438
Polycarbonate Chair Mat	142	150	157	164	171	178	184	191	197	202	208	214	2158
Porcelain Whiteboard	48	35	46	75	81	87	94	99	106	110	115	120	1054
PVC Chair Mat	402	493	570	509	598	558	501	559	599	590	559	525	6717
Recycled Chair Mat	288	198	259	172	214	187	204	191	209	195	199	197	2425
Steel Whiteboard	99	61	84	48	55	61	44	67	64	46	55	67	671
Tempered Glass Chair Mat	6	6	6	6	6	6	6	6	6	6	6	6	75
Tempered Glass Whiteboard	11	12	12	12	12	12	12	12	12	12	12	12	139
Totals	2110	2285	2455	2578	2719	2719	2842	2964	3075	3183	3288	3390	34676

### Mean Absolute Percentage Error

Category	Totals
All Categories	12.28
Anti-Fatigue Mat	36.19
Desk Pad	37.37
Entrance Mat	42.40
Polycarbonate Chair Mat	21.24
Porcelain Whiteboard	48.14
PVC Chair Mat	20.81
Recycled Chair Mat	41.90
Steel Whiteboard	56.62
Tempered Glass Chair Mat	46.20
Tempered Glass Whiteboard	47.81
Totals	385.36

### Error vs. AIC Scores (Bubble Size = Predicted Sales Quantity)

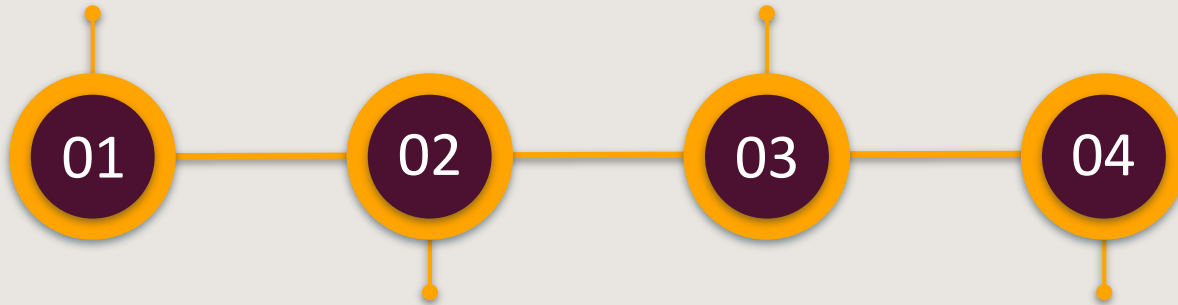


# Predictions Dashboard (2023)

# 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





- \* 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.

# Challenges

---





- \* 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.

# Conclusions

---



Thank you for  
listening!



The background features a complex, abstract geometric pattern composed of various colored triangles and hexagons. The colors include shades of teal, orange, yellow, and dark purple. These shapes are arranged in a way that creates a sense of depth and movement, with some elements appearing to overlap others. The overall effect is a vibrant, modern, and geometric aesthetic.

Any Questions?