



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.





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

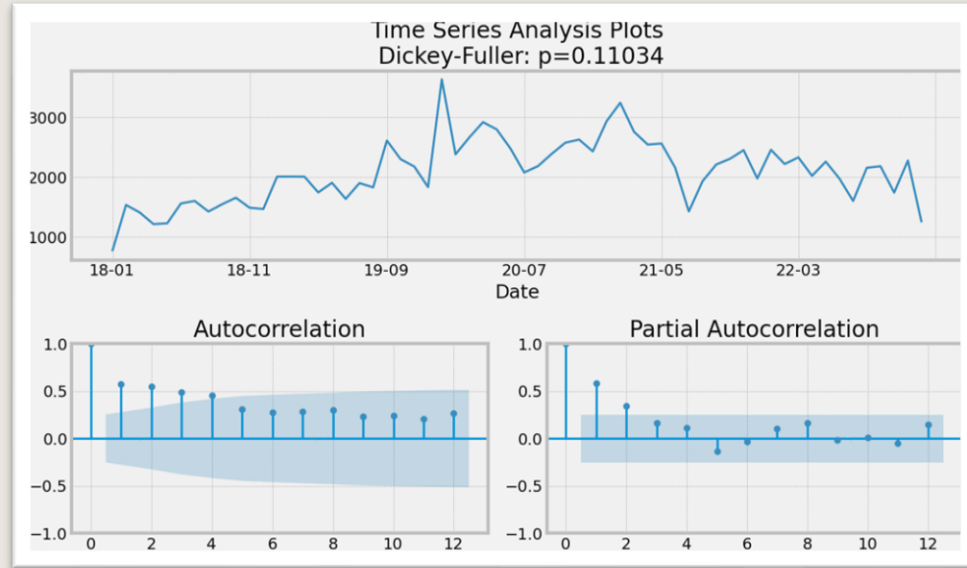
- * Time Series Analysis for Sales Volume predictions
- * Trained On 4-years (2018-2021)

* 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



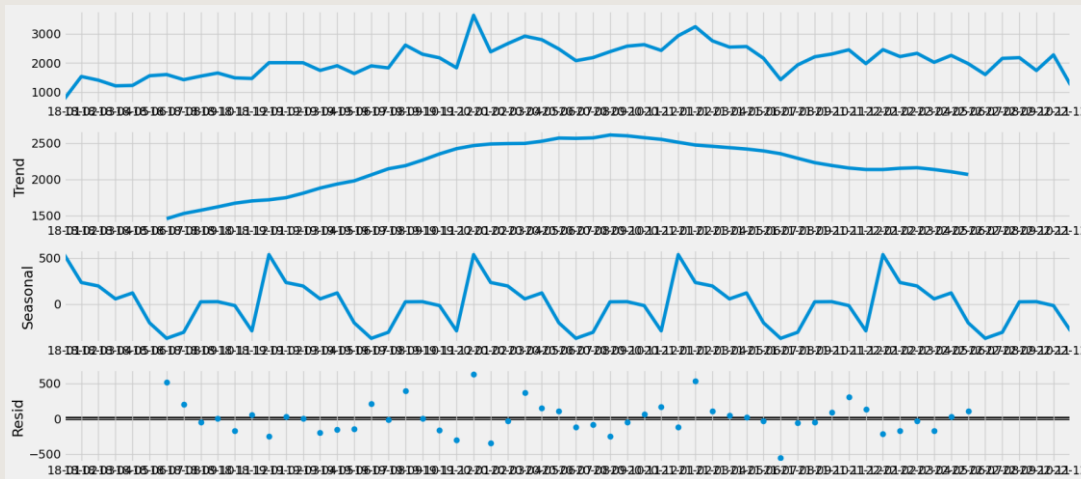
Step 1: Identified the stationarity of the time series




P-Value > 0.05 implies data is not stationary

Step 2: Suggested the initial parameters

Applied differencing analysis if needed





	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

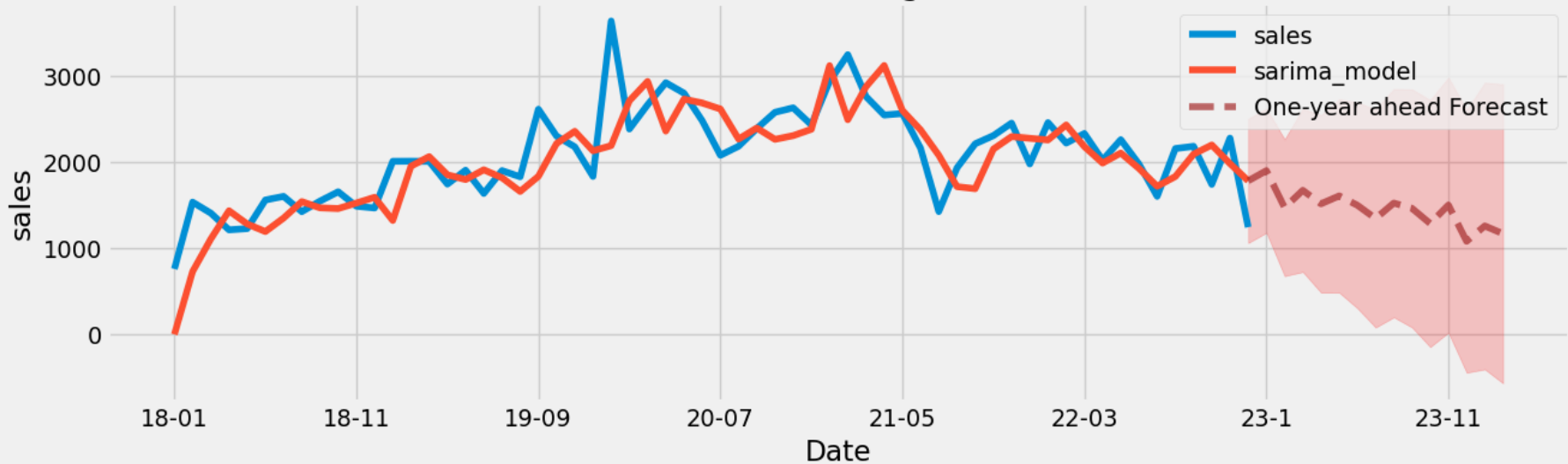
Step 3: Generated the final parameters for the model

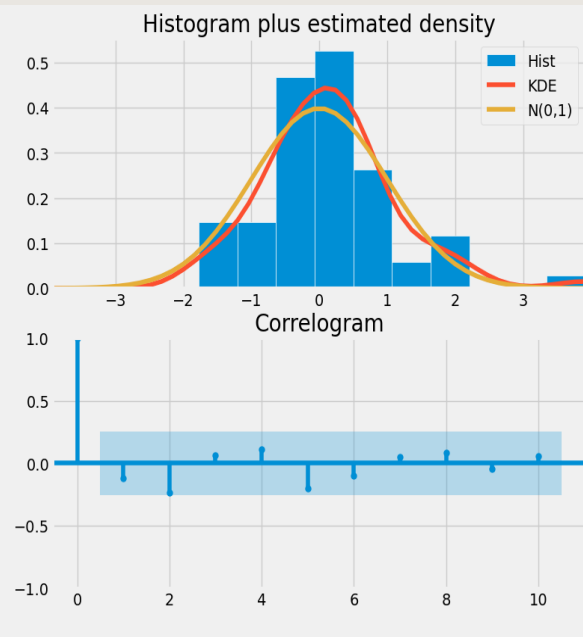
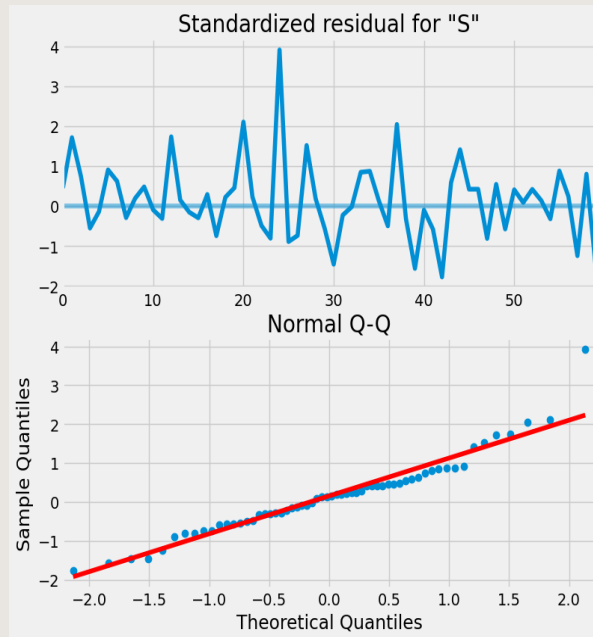
Selected parameters based on AIC Scores



Step 4: Ran the SARIMAX model

Mean Absolute Percentage Error: 12.72%





Error Analysis





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%

Attempt to Optimize the Model





Visualizations and Dashboards

- * Used Databricks Dashboards
- * Created filters in the Databricks Notebooks



Financial Dashboard (2019-2022)

Select Category:
 Select Year:

Sales Volume by Product Categories

Word cloud visualization showing product categories. The most prominent words are PVC Chair Mat, PVC Chair Pad, PVC Chair, and PVC Chair Mat.

Sales Volume by Month

Bar chart showing sales volume by month. The y-axis is Sales Volume (Units) ranging from 0 to 1000. The x-axis is Year (2019, 2020, 2021, 2022). The bars are colored by month: Jan (red), Feb (blue), Mar (green), Apr (orange), May (purple), Jun (brown), Jul (pink), Aug (grey), Sep (light blue), Oct (light green), Nov (light orange), Dec (light purple).

Revenue Breakdown

Stacked bar chart showing revenue breakdown by month. The y-axis is Revenue (USD) ranging from 0 to 1000. The x-axis is Year (2019, 2020, 2021, 2022). The bars are colored by month: Jan (red), Feb (blue), Mar (green), Apr (orange), May (purple), Jun (brown), Jul (pink), Aug (grey), Sep (light blue), Oct (light green), Nov (light orange), Dec (light purple).

Total Profit by Month

Stacked bar chart showing total profit by month. The y-axis is Profit (USD) ranging from 0 to 1000. The x-axis is Year (2019, 2020, 2021, 2022). The bars are colored by month: Jan (red), Feb (blue), Mar (green), Apr (orange), May (purple), Jun (brown), Jul (pink), Aug (grey), Sep (light blue), Oct (light green), Nov (light orange), Dec (light purple).

Total Profit (USD)

Category	Year											
	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Chair Mat	1000	1200	1500	1800	2000	2200	2400	2600	2800	3000	3200	3400
Chair Pad	800	1000	1200	1400	1600	1800	2000	2200	2400	2600	2800	3000
Chair	600	800	1000	1200	1400	1600	1800	2000	2200	2400	2600	2800
Chair Mat	400	600	800	1000	1200	1400	1600	1800	2000	2200	2400	2600
Chair Pad	200	400	600	800	1000	1200	1400	1600	1800	2000	2200	2400
Chair	100	200	300	400	500	600	700	800	900	1000	1100	1200
Chair Mat	50	100	150	200	250	300	350	400	450	500	550	600
Chair Pad	25	50	75	100	125	150	175	200	225	250	275	300
Chair	12.5	25	37.5	50	62.5	75	87.5	100	112.5	125	137.5	150
Chair Mat	6.25	12.5	18.75	25	31.25	37.5	43.75	50	56.25	62.5	68.75	75
Chair Pad	3.125	6.25	9.375	12.5	15.625	18.75	21.875	25	28.125	31.25	34.375	37.5
Chair	1.5625	3.125	4.6875	6.25	7.8125	9.375	10.9375	12.5	14.0625	15.625	17.1875	18.75
Chair Mat	0.78125	1.5625	2.34375	3.125	3.90625	4.6875	5.46875	6.25	7.03125	7.8125	8.59375	9.375
Chair Pad	0.390625	0.78125	1.171875	1.5625	1.953125	2.34375	2.734375	3.125	3.515625	3.90625	4.296875	4.6875
Chair	0.1953125	0.390625	0.5859375	0.78125	0.9765625	1.171875	1.3671875	1.5625	1.7578125	1.953125	2.1484375	2.34375
Chair Mat	0.09765625	0.1953125	0.29296875	0.390625	0.48828125	0.5859375	0.68359375	0.78125	0.87890625	0.9765625	1.07421875	1.171875
Chair Pad	0.048828125	0.09765625	0.146484375	0.1953125	0.244140625	0.29296875	0.341796875	0.390625	0.439453125	0.48828125	0.537109375	0.5859375
Chair	0.0244140625	0.048828125	0.0732421875	0.09765625	0.1220703125	0.146484375	0.1708984375	0.1953125	0.2197265625	0.244140625	0.2685546875	0.29296875
Chair Mat	0.01220703125	0.0244140625	0.03662109375	0.048828125	0.0610351							

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	1885	1272	1881	1775	1882	1881	1775	1888	1881	1881	1881	1881	21176
Anti-Fatigue Mat	184	112	181	87	181	181	181	181	181	181	181	181	1260
Desk Pad	352	352	352	352	352	352	352	352	352	352	352	352	4608
Entrance Mat	15	16	6	2	4	7	5	1	7	9	20	4	52
Polycarbonate Chair Mat	186	186	186	186	186	186	186	186	186	186	186	186	2232
Porcelain Whiteboard	88	26	17	29	80	29	80	80	80	80	80	80	861
PVC Chair Mat	322	462	322	326	398	333	321	356	499	490	338	322	4717
Recycled Chair Mat	288	188	288	172	214	187	234	181	239	196	189	187	2423
Steel Whiteboard	99	61	84	48	95	91	44	67	64	46	93	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	189
Totals	6871	4251	6670	6502	6609	6551	6368	6991	8488	8284	6569	6186	41911

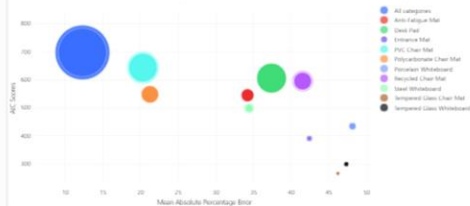
Margin Error

Category	Months												Totals
	1	2	3	4	5	6	7	8	9	10	11	12	
All categories	688	526	761	761	681	681	884	884	884	981	1081	1081	10467
Anti-Fatigue Mat	181	181	181	181	181	181	181	181	181	181	181	181	2182
Desk Pad	352	352	352	352	352	352	352	352	352	352	352	352	4287
Entrance Mat	22	26	27	62	34	96	91	47	65	45	47	46	498
Polycarbonate Chair Mat	142	182	182	184	171	178	184	191	197	202	208	214	2188
Porcelain Whiteboard	48	95	96	79	81	87	94	99	106	110	115	120	1064
PVC Chair Mat	302	492	329	398	318	338	376	576	586	609	625	641	6482
Recycled Chair Mat	288	271	309	362	402	422	450	482	509	531	551	575	5167
Steel Whiteboard	88	102	114	125	138	144	151	161	169	176	183	190	1789
Tempered Glass Chair Mat	16	16	16	18	19	20	21	22	23	24	25	26	243
Tempered Glass Whiteboard	11	11	11	11	11	11	11	11	11	11	11	11	154
Totals	2110	2285	2485	2578	2719	2842	2964	3075	3180	3288	3390	3489	34676

Mean Absolute Percentage Error

Category	Totals
All categories	12.28
Anti-Fatigue Mat	38.18
Desk Pad	37.27
Entrance Mat	46.40
Polycarbonate Chair Mat	11.28
Porcelain Whiteboard	48.14
PVC Chair Mat	10.81
Recycled Chair Mat	41.80
Steel Whiteboard	34.62
Tempered Glass Chair Mat	46.20
Tempered Glass Whiteboard	47.81
Totals	389.88

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

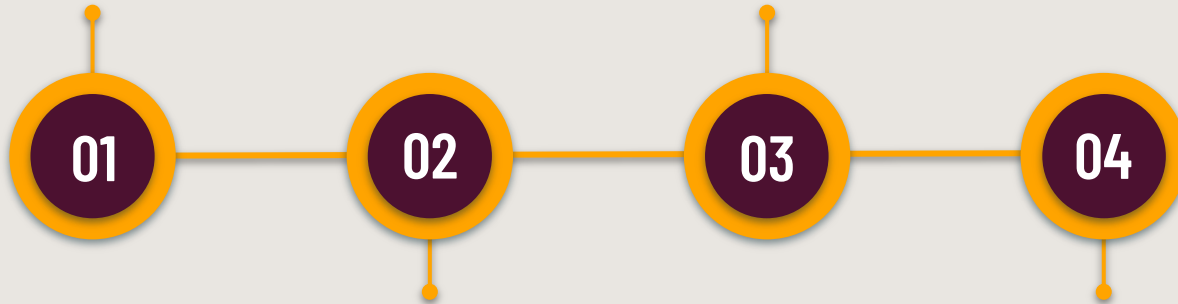


Predictions Dashboard (2023)

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





- * The clusters on Databricks Community Edition reset after 2 hours of inactivity.
- * 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 additional models incorporating 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 distributors.

Conclusions



Thank you for
listening!



Any Questions?

