



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

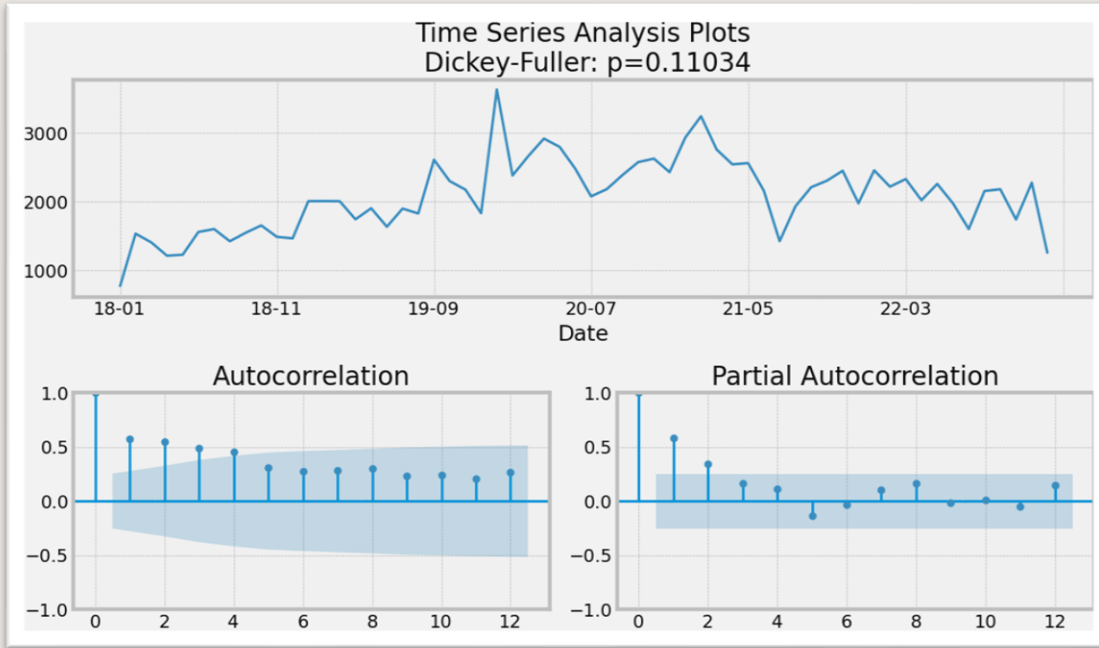
- * Time Series Analysis for Sales Volume predictions
- * Trained On 4-years (2018-2021)

* 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



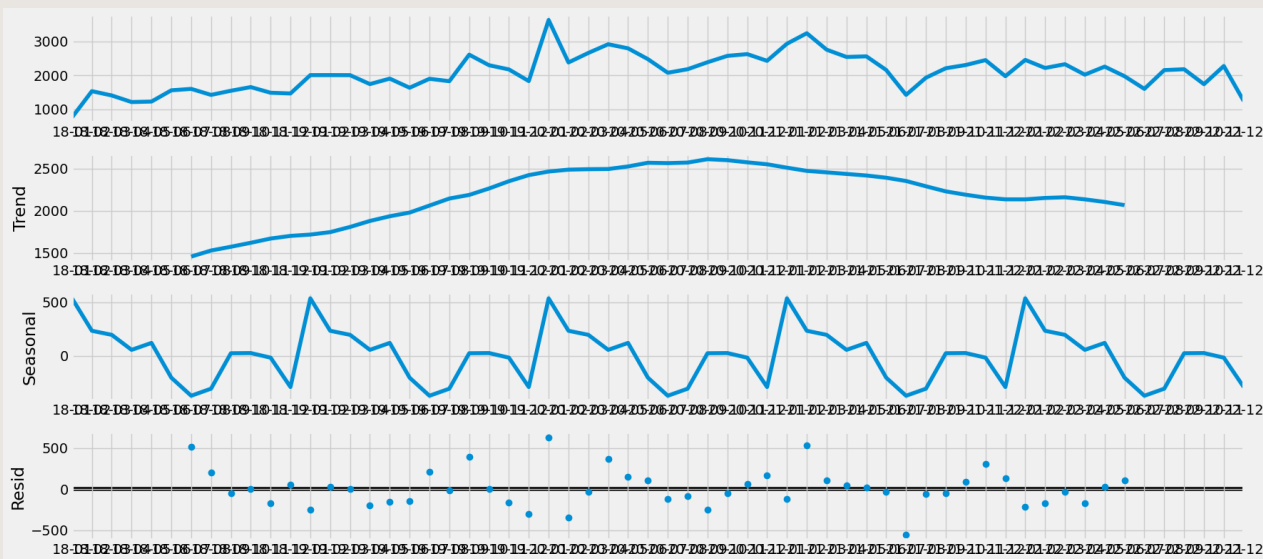
Step 1: Identified the stationarity of the time series




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

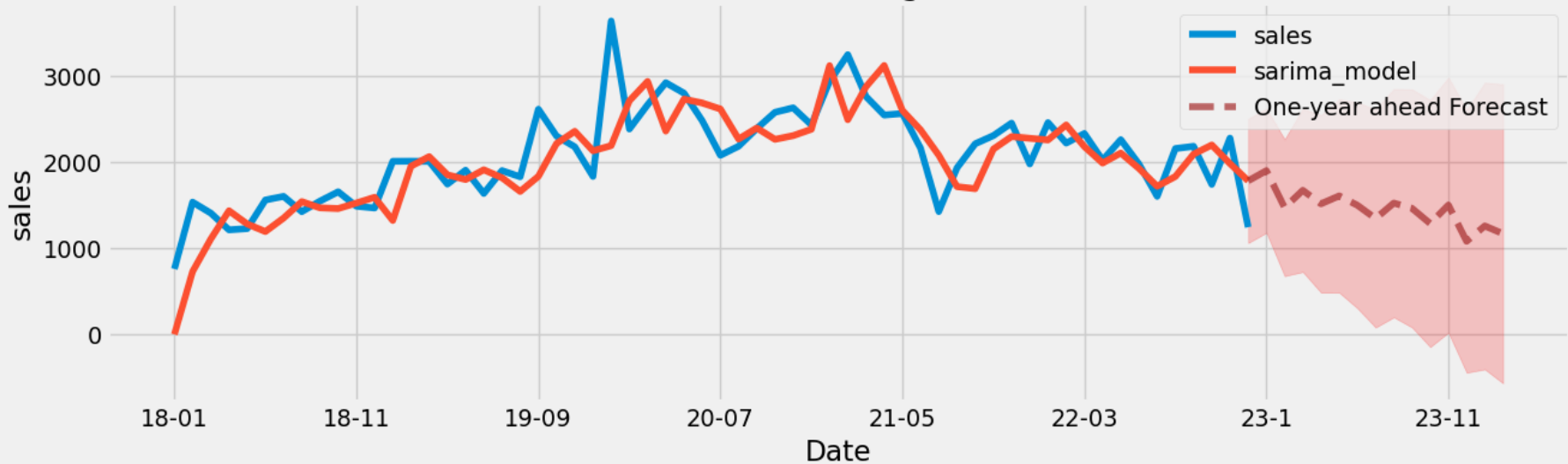
Step 3: Generated the final parameters for the model

Selected parameters based on AIC Scores

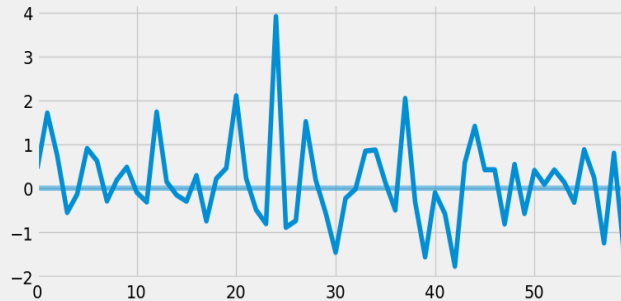


Step 4: Ran the SARIMAX model

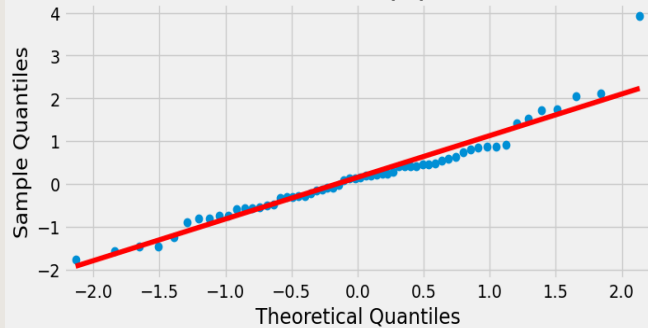
Mean Absolute Percentage Error: 12.72%



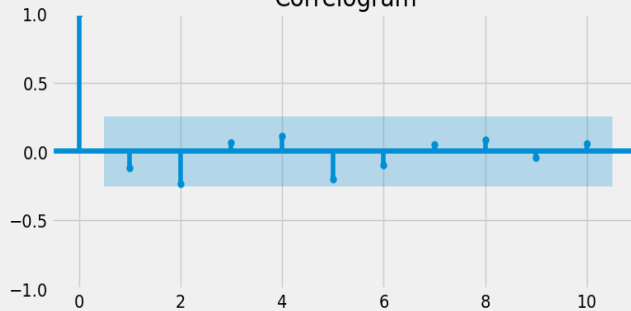
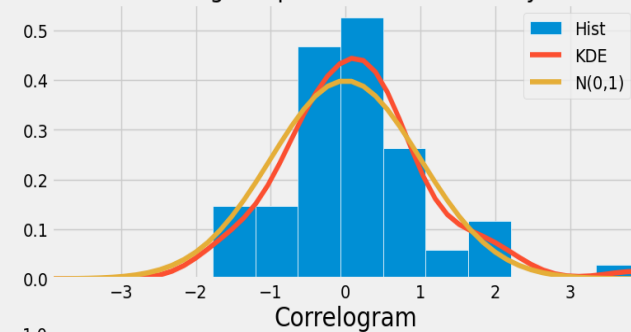
Standardized residual for "S"



Normal Q-Q



Histogram plus estimated density



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



[illegible]

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	1280
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	82
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	402	322	326	398	333	321	359	499	490	338	322	4717
Recycled Chair Mat	288	188	288	172	214	187	234	181	239	196	199	197	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	4231	6670	6502	6609	6531	6368	6931	8488	8284	6569	6186	41911

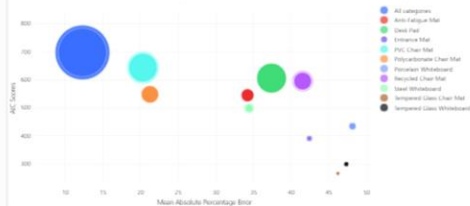
Margin Error

Category	Months												Totals
	1	2	3	4	5	6	7	8	9	10	11	12	
All categories	688	326	761	761	611	861	861	861	861	961	1061	1061	1061
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	192	192	164	171	178	164	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	292	279	299	319	338	359	379	399	419	439	459	4402
Recycled Chair Mat	298	271	299	262	292	322	352	382	412	442	472	502	5167
Steel Whiteboard	88	102	114	125	138	144	151	161	169	176	183	190	1799
Tempered Glass Chair Mat	16	16	16	18	19	20	21	22	23	24	25	26	240
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	38.19
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	41.90
Steel Whiteboard	34.62
Tempered Glass Chair Mat	46.20
Tempered Glass Whiteboard	47.81
Totals	389.86

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

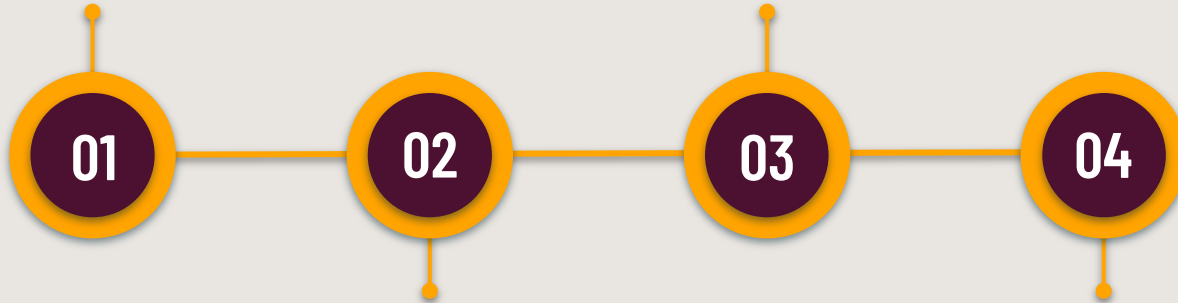


Predictions Dashboard (2023)

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





* Sales predictions show a slight decline in sales 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 further time period. This make it more appropriate for Just-In-Time distributor.

Conclusions



Thank you for
listening!



Any Questions?

