

Superstore Analytics Dashboard: Business Intelligence, Forecasting, and Customer Segmentation

Seminar Paper

“Business Intelligence Dashboard implementation”

“In the context of Advanced Analytics”

Winter semester 26/27

Supervisor: Prof. Dr. Christian Leubner

Course: Informatics Business (M.Sc.)

Submitted By: Md Abdullah Al Noman Mridha
Matriculation Number: 30494592
Haydnstraße, 63
44145, Dortmund
Nomanmridha.mdabdullahal@fh-swf.de
<https://github.com/nomanmridha/Superstore-data-Analysis>

Deadline: 2027-01-31

Table of Contents

List of Abbreviations	2
List of Symbols	3
List of Figures.....	4
Introduction.....	5
Theoretical Background.....	6
Dataset Preparation and Exploratory Analysis	7
Raw Dataset Overview.....	7
Data Cleaning.....	7
Exploratory Analysis	8
Descriptive Dashboard Analytics	9
Category-Level Performance	9
Discount Impact	10
Sales vs Profit Overtime	11
Forecasting Results	11
Six-Month Sales Forecast	11
Segment-Specific Forecasts	12
Segment-level Forecasting Evaluation	13
Customer Segmentation (RFM).....	14
Optimal Clustering	14
Cluster Distribution	15
Behavioral Visualization.....	15
Final Dashboard.....	16
Discussion	18
Limitations.....	20
Conclusion	20
References.....	21

List of Abbreviations

Abbreviation	Full Form	Definition
BI	Business Intelligence	Technologies and processes used to collect, analyse and visualize business data for decision-making.
EDA	Exploratory Data Analysis	Initial data investigation process used to summarize patterns, detect anomalies and validate assumption.
KPI	Key Performance Indicator	A measurable metric used to evaluate business performance (e.g.. Sales, Profit, margin)
RFM	Recency, Frequency, Monetary	A customer segmentation framework evaluating purchase recency, purchase frequency, and total spending.
MAE	Mean Absolute Error	Average of absolute differences between actual and predicted values.
RMSE	Root Mean Squared Error	Square root of average squared forecast errors; penalizes large deviations.
MAPE	Mean Absolute Percentage Error	Percentage-based forecasting accuracy metric.
SARIMA	Seasonal Autoregressive Integrated Moving average	Time series model capturing trend, seasonality and autoregressive behaviour.
ARIMA	Autoregressive integrated moving average	Statistical forecasting model combining AR and MA components with differencing.
K-means	K-Means Clustering	Unsupervised learning algorithm used to group customers based on similarity.
Dash	Plotly Dash Framework	Python-based web framework for building analytical dashboards.
Prophet	Facebook Prophet	Additive time-series forecasting model handling trend and seasonality.

List of Symbols

Symbol	Name	Definition / Use in Project
μ	Mean (Average)	Statistical average value of a dataset.
σ	Standard Deviation	Measure of variability in a dataset.
Σ	Summation	Represents total aggregation (e.g. Total Sales, total monetary value).
Δt	Time Difference	Difference in time used to calculate recency in RFM analysis.
\geq	Greater Than or Equal	Used in filtering conditions (e.g. Sales thresholds)
\leq	Less than or Equal to	Used in conditional data filtering.
%	Percentage	Used in Profit margin and MAPE calculations.
K	Number of Clusters	Represents number of clusters in K-means algorithm.
y_t	Actual Value at Time t	Observed sales value in time-series forecasting.
\hat{y}_t	Predicted Value at Time t	Forecasted sales value in time-series value.
e_t	Forecast Error	Difference between actual and predicted values.
R	Recency	Time since last customer purchase
F	Frequency	Number of customer purchase.
M	Monetary	Total revenue generated by a customer.

List of Figures

FIGURE 1: RAW DATASET OVERVIEW	7
FIGURE 2: CLEANED DATASET.....	7
FIGURE 3: MONTHLY SALES TREND (2014-2017).....	8
FIGURE 4: MONTHLY PROFIT TREND (2014-2017)	8
FIGURE 5: PROFIT BY CATEGORY	9
FIGURE 6: SALES BY CATEGORY	9
FIGURE 7: DISCOUNT-PROFIT RELATIONSHIP	10
FIGURE 8: OVERALL DISCOUNT-PROFIT RELATIONSHIP.....	10
FIGURE 9: SAMPLED DASHBOARD VIEW	11
FIGURE 10: SIX MONTH SALES FORECAST	11
FIGURE 11: CHAMPION SEGMENT FORECAST	12
FIGURE 12: LOYAL SEGMENT FORECASTS.....	12
FIGURE 13: PROPHET-BASED FORECASTING	13
FIGURE 14: CATEGORY-LEVEL FORECAST EVALUATION	13
FIGURE 15: REGION-LEVEL EVALUATION	14
FIGURE 16: SEGMENT-LEVEL EVALUATION	14
FIGURE 17: ELBOW METHOD CLUSTERING.....	14
FIGURE 18: CLUSTER SIZE DISTRIBUTION	15
FIGURE 19: RECENCY VS FREQUENCY	15
FIGURE 20: FREQUENCY VS MONETARY	16
FIGURE 21: FINAL DASHBOARD PREVIEW	17

Introduction

The rapid digitalization of retail operations has transformed how organizations capture and utilize transactional data. Modern retail enterprises no longer rely solely on descriptive summaries of past sales; instead, they increasingly integrate predictive modelling and customer intelligence to guide strategic decisions. Business Intelligence (BI) systems serve as the backbone of this transformation, enabling interactive visualization, forecasting, and segmentation in a unified analytical environment. (Sharda, 2018)

The Superstore dataset, obtained from Kaggle, provides a structured representation of retail transactions across multiple regions, categories, and customer segments. While often used for introductory analytics exercises, this project approaches the dataset from a decision-support perspective. Rather than performing isolated statistical analyses, the objective is to design and implement a fully functional analytics dashboard that integrates descriptive analytics, forecasting, and RFM-based customer segmentation.

The research questions guiding this project are:

1. How can transactional retail data be transformed into an interactive Business Intelligence dashboard?
2. To what extent can future sales be forecasted reliably using time-series method?
3. How effectively can RFM segmentation reveal actionable customer groups?
4. How can these analytical outputs inform managerial decision-making?

The report proceeds by presenting theoretical foundations, methodology, empirical findings and business implications.

Theoretical Background

Business Intelligence in Retail

Business Intelligence encompasses tools and processes used to transform raw data into meaningful information for decision-making (Sharda, 2018). In retail environments, BI dashboards typically present KPIs such as total sales, profit margins, regional performance, and customer behaviour metrics. Effective BI systems are interactive, allowing users to filter across time periods, product categories, and geographic regions.

Time Series Forecasting

Time-series forecasting models aim to capture temporal patterns such as trend and seasonality. The ARIMA framework developed by (Box, 1976) remains foundational in modelling linear dependencies in time-series data. Seasonal (SARIMA) extends this approach by explicitly incorporating periodic patterns (Hyndman, 2018).

Forecast accuracy is evaluated using metrics such as MAE, RMSE, and MAPE. While MAPE provides interpretability in percentage terms, it may exaggerate errors when actual values are small (Hyndman, 2018).

RFM Customer Segmentation

RFM analysis evaluates customers based on recency, frequency, and monetary value (Fader, 2005). This behavioural segmentation approach is widely used in retail marketing to identify high-value customers and those at risk of churn. When combined with clustering algorithms such as K-means (MacQueen, 1967), RFM enables automated grouping of customers into actionable segments.

Dataset Preparation and Exploratory Analysis

Raw Dataset Overview

Figure 01 presents a preview of the raw Superstore dataset. It contains order-level transactional data, including order date, ship mode, region, category, sales, profit, and discounts.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Price	Quantity	Discount	Profit	Profit_raw	
1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-B0-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96	2.0	0.0	0.5	53795918367346	419.136
2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.94	3.0	0.0	0.44	81265306122449	219.587
3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	OV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10002040	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14.62	2.0	0.0	0.14	023265306122447	68.734
4	US-2015-108966	2015-10-11	2015-10-16	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table						
5	US-2015-108966	2015-10-11	2015-10-16	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10007600	Office Supplies	Storage	Eldon Fold 'N Roll Cart System						
6	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48.86	7.0	0.0	0.28	91742857142853	141.694
7	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48.86	7.0	0.0	0.28	91742857142853	141.694
8	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TFC-PH-10002275	Technology	Phones	Mitel 5320 IP Phone VoIP Phone	307.152	6.0	0.0	0.185	13306122449	807.152
9	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18.504	3.0	0.0	0.21	80120408163264	57.825
10	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	Beklin FSC206VTEL 6 Outlet Surge	114.55	0.0	0.0	0.7	03496387751013	34.47
11	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TFC-PH-10002033	Technology	Phones	Konftel 250 Conference phone - Charcoal black	911.424	4.0	0.0	0.139	50367346938773	683.568
12	CA-2017-114412	2017-04-15	2017-04-20	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-ST-10004186	Office Supplies	Storage	Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.88	6.0	0.0	0.27	17877551020406	7.133.176
13	US-2015-118983	2015-11-22	2015-11-26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000756	Office Supplies	Binders	Stores DuraTech Recycled Plastic Frosted Binders	2.544	3.0	0.0	0.8	0.7787755102040815	3.816
14	US-2015-118983	2015-11-22	2015-11-26	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000756	Office Supplies	Binders	Stores DuraTech Recycled Plastic Frosted Binders	2.544	3.0	0.0	0.8	0.7787755102040815	3.816
15	CA-2014-105893	2014-11-11	2014-11-16	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004186	Office Supplies	Storage	Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.88	6.0	0.0	0.27	17877551020406	7.133.176
16	CA-2014-105893	2014-11-11	2014-11-16	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004186	Office Supplies	Storage	Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.88	6.0	0.0	0.27	17877551020406	7.133.176
17	CA-2014-143336	2014-08-27	2014-09-01	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10003056	Office Supplies	Binders	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
18	CA-2014-143336	2014-08-27	2014-09-01	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10003056	Office Supplies	Binders	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
19	CA-2014-143336	2014-08-27	2014-09-01	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10003056	Office Supplies	Binders	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
20	CA-2014-143336	2014-08-27	2014-09-01	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10003056	Office Supplies	Binders	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
21	CA-2016-173730	2016-12-09	2016-12-13	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10001492	Office Supplies	Appliances	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
22	CA-2016-173730	2016-12-09	2016-12-13	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10001492	Office Supplies	Appliances	Acco Six-Outlet Power Strip, 4 Cord Length	60.34	7.0	0.0	0.32	01742857142853	25.884
23	US-2017-156909	2017-07-16	2017-07-18	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture	Chairs	Global Deluxe Stacking Chair, Gray	71.372	2.0	0.0	0.2	080816326530612	10.196
24	US-2017-156909	2017-07-16	2017-07-18	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture	Chairs	Global Deluxe Stacking Chair, Gray	71.372	2.0	0.0	0.2	080816326530612	10.196
25	CA-2016-121755	2016-01-16	2016-01-20	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West	OFF-BI-10001634	Office Supplies	Binders	Wilson Jones Active Use Binders	11.648	2.0	0.0	0.2	8.61742857142856	42.224
26	CA-2016-121755	2016-01-16	2016-01-20	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West	TFC-AC-10003027	Technology	Accessories	Imation 8GB Mini TravelDrive USB 2.0 Flash Drive	30.57	3.0	0.0	0.24	03877551020408	117.741
27	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
28	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
29	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
30	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
31	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
32	US-2015-150630	2015-09-17	2015-09-21	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10003056	Office Supplies	Binders	Avery Recycled Flexi-Vinyl Covers for Binding Systems	9.618	2.0	0.0	0.27	1394285714285713	70.532
33	CA-2017-107727	2017-10-19	2017-10-23	Second Class	MA-17560	Matt Abelman	Home Office	United States	Houston	Texas	77095	Central	OFF-PA-10000249	Office Supplies	Paper	Easy-staple paper	29.472	3.0	0.0	0.2	2999318373469	99.468
34	CA-2016-117950	2016-12-08	2016-12-10	First Class	GH-14485	Gene Hale	Corporate	United States	Richardson	Texas	75080	Central	FUR-FU-10000464	Furniture	Furnishings	Electric Architect's Clamp-On Swing Arm Lamp, Black	150.92	5.0	0.0	0.6	19653061224496	147.963
35	CA-2015-117415	2015-12-27	2015-12-31	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Houston	Texas	77041	Central	OFF-EN-10002986	Office Supplies	Envelopes	110-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes	113.328	9.0	0.0	0.27	1394285714285713	70.532
36	CA-2015-117415	2015-12-27	2015-12-31	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Houston	Texas	77041	Central	OFF-EN-10002986	Office Supplies	Envelopes	110-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes	113.328	9.0	0.0	0.27	1394285714285713	70.532
37	CA-2015-117415	2015-12-27	2015-12-31	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Houston	Texas	77041	Central	OFF-EN-10002986	Office Supplies	Envelopes	110-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes	113.328	9.0	0.0	0.27	1394285714285713	70.532
38	CA-2017-120999	2017-09-19	2017-09-19	Standard Class	LC-16930	Linda Cazamas	Corporate	United States	Naperville	Illinois	60540	Central	TFC-PH-10004091	Technology	Phones	Panasonic Kx-T5550	147.168	4.0	0.0	0.33	7885742857142857	165.564
39	CA-2016-101343	2016-07-17	2016-07-22	Standard Class	RA-19885	Ruben Ausman	Corporate	United States	Los Angeles	California	90045	West	OFF-ST-10003479	Office Supplies	Storage	Eldon Base for Stackable Storage Shelf, Platinum	77.88	2.0	0.0	0.6	7946938775510204	3.894
40	CA-2017-139619	2017-09-19	2017-09-19	Standard Class	ES-14080	Eric Smith	Corporate	United States	Melbourne	Florida	32935	South	OFF-ST-10002382	Office Supplies	Storage	Advantus 10 Drawer Portable Organizer, Chrome Metal Frame, Smoke Drawers	55.616	2.0	0.0	0.2	19.513469387755102	95.616
41	CA-2016-118255	2016-03-11	2016-03-11	First Class	ON-18715	Odella Nelson	Corporate	United States	Eagan	Minnesota	55122	Central	TFC-AC-10000171	Technology	Accessories	Verbatim 25 GB Blu-ray Single Layer Recordable Disc, 25/Pack	45.98	2.0	0.0	0.4	34979591836734	197.714
42	CA-2016-118255	2016-03-11	2016-03-11	First Class	ON-18715	Odella Nelson	Corporate	United States	Eagan	Minnesota	55122	Central	OFF-BI-10003291	Office Supplies	Binders	Wilson Jones Leather-Like Binders with Durable Lock Rings	17.46	2.0	0.0	0.16	747346938775507	82.062

FIGURE 1: RAW DATASET OVERVIEW

The dataset spans 2014-2017 and includes over 9,000 transactions.

Data Cleaning

Figure 2 shows the cleaned dataset structure after pre-processing.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name
1	CA-2016-152156	11/08/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-B0-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase
2	CA-2016-152156	11/08/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
3	CA-2016-138688	06/12/2016	06/16/2016	Second Class	OV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10002040	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal
4	US-2015-108966	10/11/2015	10/16/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table
5	US-2015-108966	10/11/2015	10/16/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10007600	Office Supplies	Storage	Eldon Fold 'N Roll Cart System
6	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001187	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, C1
7	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10002823	Office Supplies	Art	Newell 322
8	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-Ph-10002073	Technology	Phones	Motor S210 IP Phone VoIP Phone
9	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BL-10003919	Office Supplies	Binders	DLS Angel Ring Binders with Locking Rings by Samill
10	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	Belkin FC5200V1, 6 Outlet Surge
11	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10001339	Furniture	Tables	Chromcraft Rectangular Conference Table
12	CA-2016-115812	06/09/2014	06/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-Ph-10002073	Technology	Phones	Konfelz 220 Conference Phone - Charcoal black
13	CA-2016-161389	12/15/2016	12/16/2016	Standard Class	HP-14815	Irene Maddox	Consumer	United States	San Francisco	California	94109	West	OFF-PA-10001625	Office Supplies	Paper	Post-it Super Sticky Notes
14	CA-2016-161389	12/15/2016	12/16/2016	Standard Class	HP-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10003656	Office Supplies	Binders	Hofner PB200 Plastic Comb Binding Machine
15	US-2015-118983	12/22/2015	12/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-PA-10001625	Office Supplies	Appliances	Holmes Replacement Filter for HEPA Air Cleaner, Very La
16	US-2015-118983	12/22/2015	12/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000716	Office Supplies	Binders	Stores Duracraft Recycled Plastic Frosted Binders
17	CA-2016-105893	11/11/2014	11/14/2014	Standard Class	PK-16075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004008	Office Supplies	Storage	Stur-D Stor-Shelving, Vertical-5-Shell 72" Hx 36" Wx 18.1
18	CA-2016-143136	07/14/2014	07/16/2014	Second Class	ZD-21925	Zuchitsus Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-PA-10001625	Office Supplies	Paper	Post-it Super Sticky Notes
19	CA-2016-143136	07/14/2014	09/01/2014	Second Class	ZD-21925	Zuchitsus Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10003056	Office Supplies	Art	Newell 341
20	CA-2016-143136	07/14/2014	09/01/2014	Second Class	ZD-21925	Zuchitsus Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-Ph-10011949	Technology	Phones	Cisco SPA 501G IP Phone
21	CA-2016-143136	07/14/2014	09/01/2014	Second Class	ZD-21925	Zuchitsus Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10002215	Office Supplies	Binders	Wilson Jones Hanging View Binder, White, 1"
22	CA-2016-137730	12/09/2016	12/13/2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Freemont	Nebraska	68025	Central	OFF-AR-10001492	Office Supplies	Appliances	Acco 6x Outlet Power Strip, 4 Cord Length
23	CA-2016-156909	7/18/2017	7/18/2017	Second Class	SV-20605	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture	Chairs	Globe Deluxe Stacking Chair, Gray
24	CA-2015-106320	9/29/2015	9/30/2015	Standard Class	EB-13840	Emily Burns	Consumer	United States	Orem	Utah	84057	West	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table
25	CA-2015-106320	9/29/2015	9/30/2015	Standard Class	EB-13840	Emily Burns	Consumer	United States	Orem	Utah	84057	West	OFF-AR-10001625	Office Supplies	Binders	Wilson Jones Hanging View Binder, White, 1"
26	CA-2015-122755	4/7/2016	4/7/2016	Standard Class	EH-13945	Steve Hoffmann	Consumer	United States	Philadelphia	Pennsylvania	19140	East	TEC-Ph-10002073	Technology	Phones	Unimate 6000 Series Slim 2.0 Flat Drive
27	CA-2015-122755	4/7/2016	4/7/2016	Standard Class	EH-13945	Steve Hoffmann	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-B0-10000434	Furniture	Bookcases	Riverside Palais Royal Layovers Bookcase, Royale Cherry
28	US-2015-150630	9/17/2015	9/21/2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10000474	Office Supplies	Binders	Avery Recycled Flexi-View Covers for Binding Systems
29	US-2015-150630	9/17/2015	9/21/2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-FU-10000448	Furniture	Furnishings	Howard Mill 13-1/4" Diameter Brushed Chrome Shoppers
30	US-2015-150630	9/17/2015	9/21/2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-CH-10001625	Office Supplies	Chairs	Belmont Model 1800 Electric Pen Sharpener, Putty/W
31	US-2015-150630	9/17/2015	9/21/2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-10001255	Office Supplies	Binders	Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11"
32	US-2015-150630	9/17/2015	9/21/2015	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-10001683	Office Supplies	Art	Lumber Cards
33	CA-2016-127627	10/27/2016	10/28/2016	First Class	GH-14485	Gene Hale	Corporate	United States	Richmond	Texas	75080	Central	OFF-PA-10001625	Office Supplies	Paper	Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11"
34	CA-2016-117590	12/08/2016	12/10/2016	First Class	GH-14485	Gene Hale	Corporate	United States	Richmond	Texas	75080	Central	TEC-Ph-10001497	Technology	Phones	GE 3024SEA2
35	CA-2016-117590	12/08/2016	12/10/2016	First Class	GH-14485	Gene Hale	Corporate	United States	Richmond	Texas	75080	Central	FUR-FU-10003664	Furniture	Furnishings	Electrictx Archetype's Clamp-On Swing Arm Lamp, Black
36	CA-2016-117415	12/27/2015	12/31/2015	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Houston	Texas	77041	Central	OFF-EN-10002986	Office Supplies	Envelopes	MID-A12 7" x 9 1/2" Premium Diagonal Seal Envelopes
37	CA-2015-117415	12/27/2015	12/31/2015	Standard Class	SN-20710	Steve Nguyen	Home Office	United States	Houston	Texas	77041	Central	FUR-B0-10000254	Furniture	Bookcases	Atlantic Metals Mobile 3-Shell Bookcases, Custom Col
Sample data generated for testing purposes.																

- Verifying discount consistency.
- Aggregating monthly sales.

This step ensured that forecasting and segmentation would rely on consistent, reliable data.

Exploratory Analysis

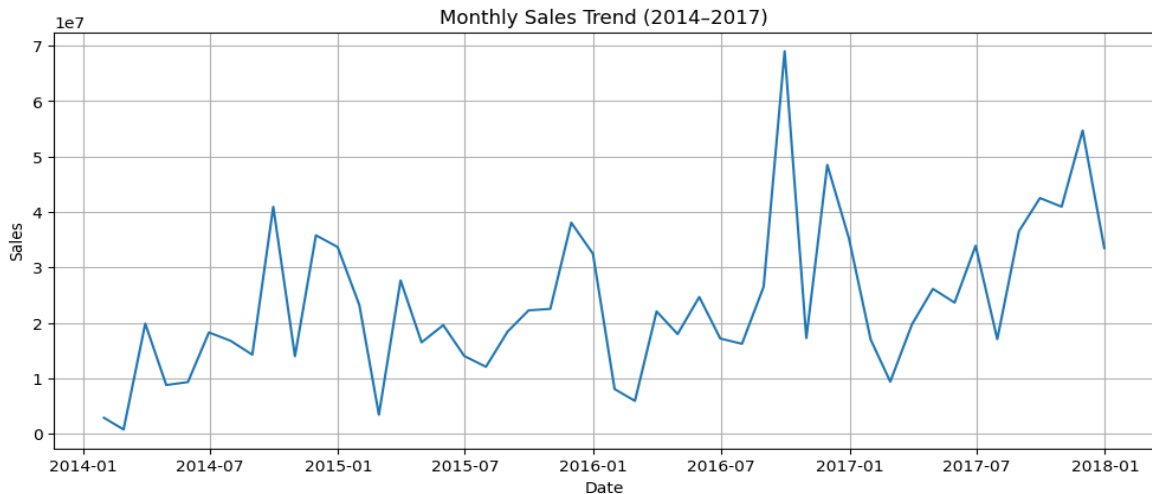


FIGURE 3: MONTHLY SALES TREND (2014-2017)

Sales show clear seasonality, with peaks near year-end periods. This pattern suggests inventory planning must account for seasonal demand spikes.

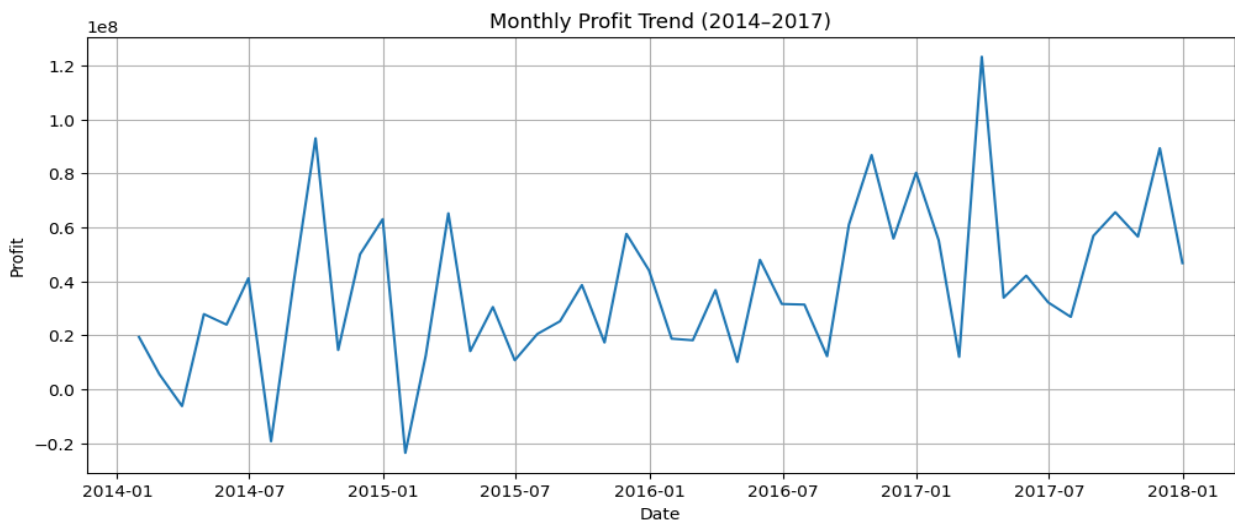


FIGURE 4: MONTHLY PROFIT TREND (2014-2017)

Profit does not increase proportionally with sales, indicating that discounting and cost structures influence margins.

Descriptive Dashboard Analytics

Category-Level Performance

Office Supplies generate the highest profit despite moderate sales volume. Furniture exhibits higher revenue but weaker profitability.

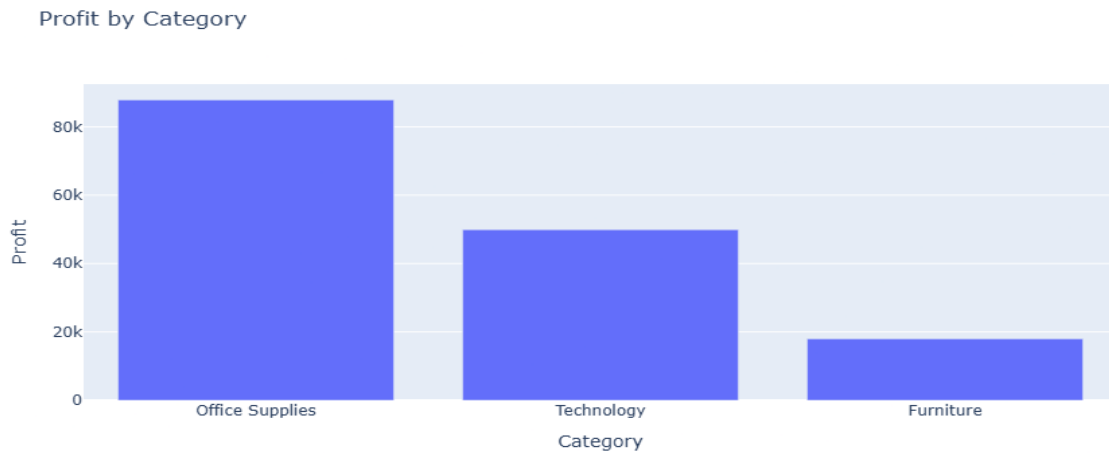


FIGURE 5: PROFIT BY CATEGORY

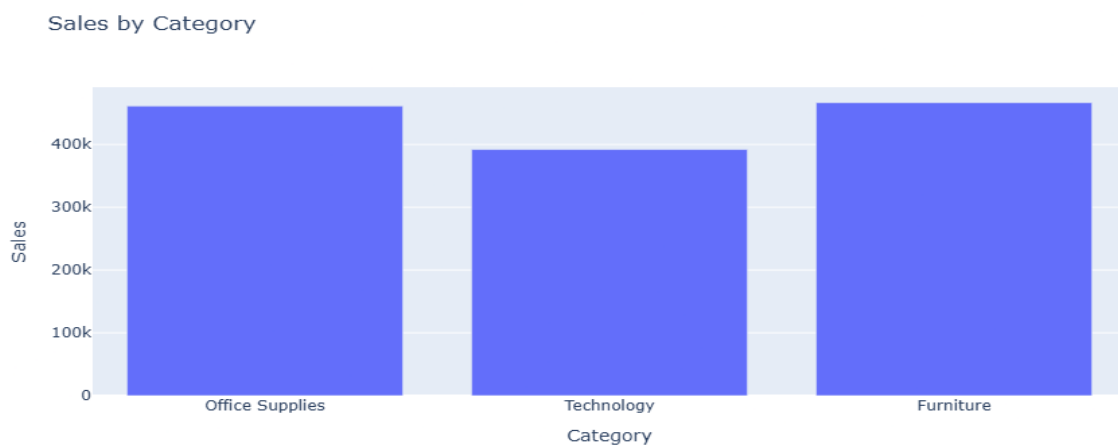


FIGURE 6: SALES BY CATEGORY

This divergence between revenue and profitability highlights the importance of margin-focused strategy.

Discount Impact

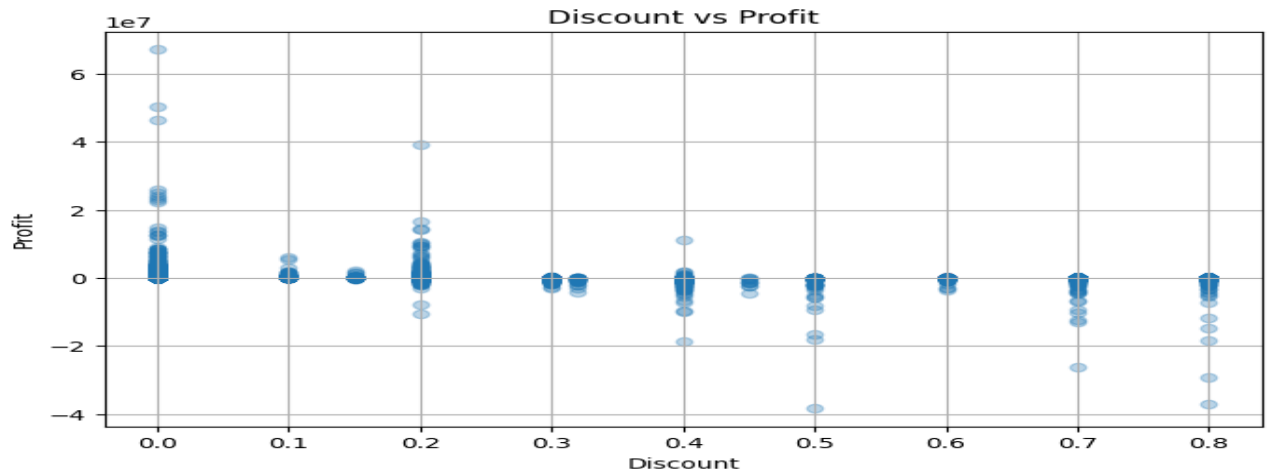


FIGURE 7: DISCOUNT-PROFIT RELATIONSHIP

A negative correlation (~ -0.38) indicates that heavy discounting reduces profitability.

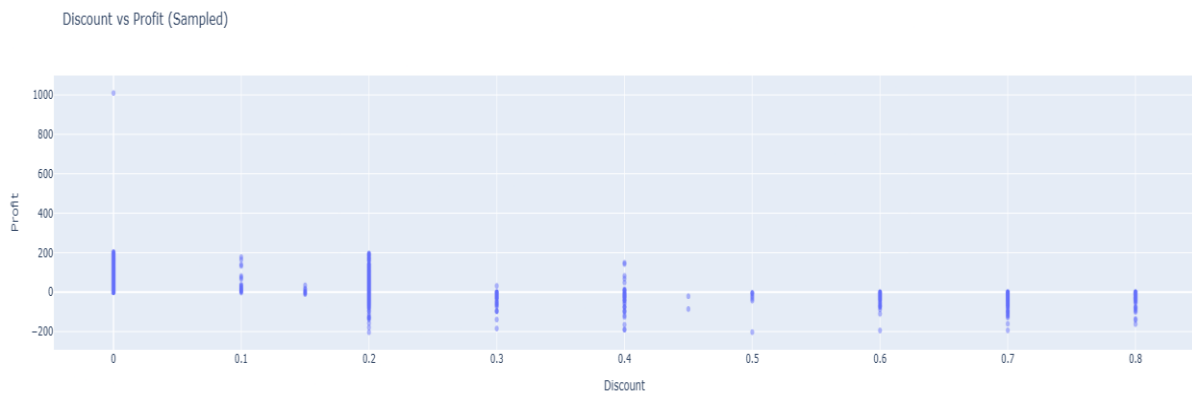


FIGURE 8: OVERALL DISCOUNT-PROFIT RELATIONSHIP

The scatterplot reinforces the nonlinear relationship between discount levels and profit outcomes.

Sales vs Profit Overtime

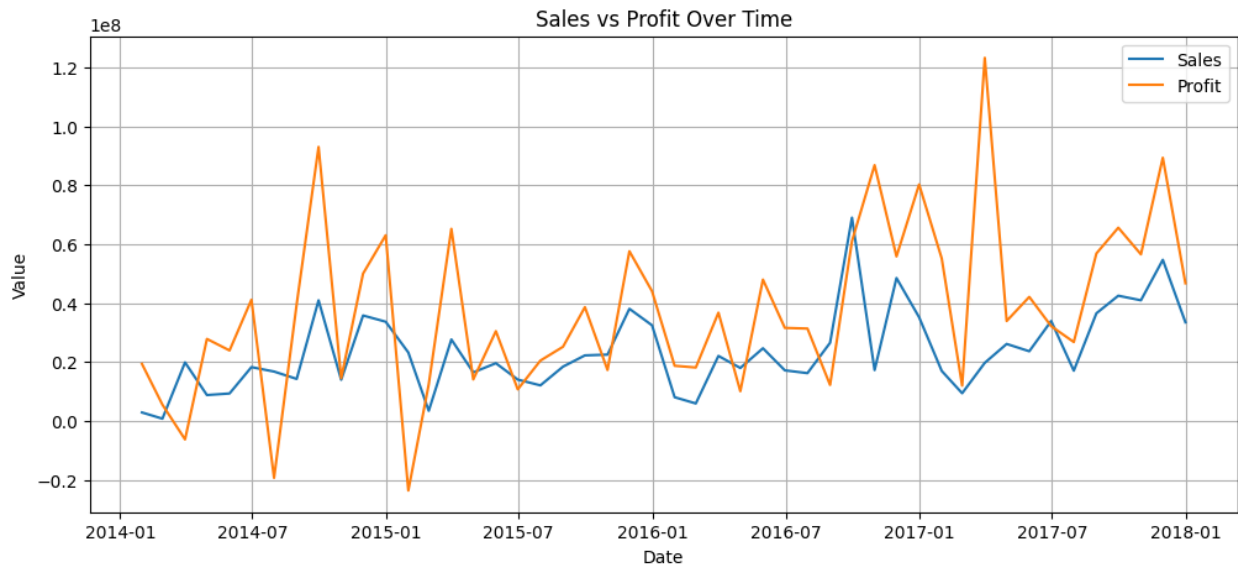


FIGURE 9: SAMPLED DASHBOARD VIEW

Periods of high sales do not always align with high profitability, suggesting promotional strategies require optimization.

Forecasting Results

Six-Month Sales Forecast

Sales Forecast (Next 6 Months) | MAPE: 28.88%

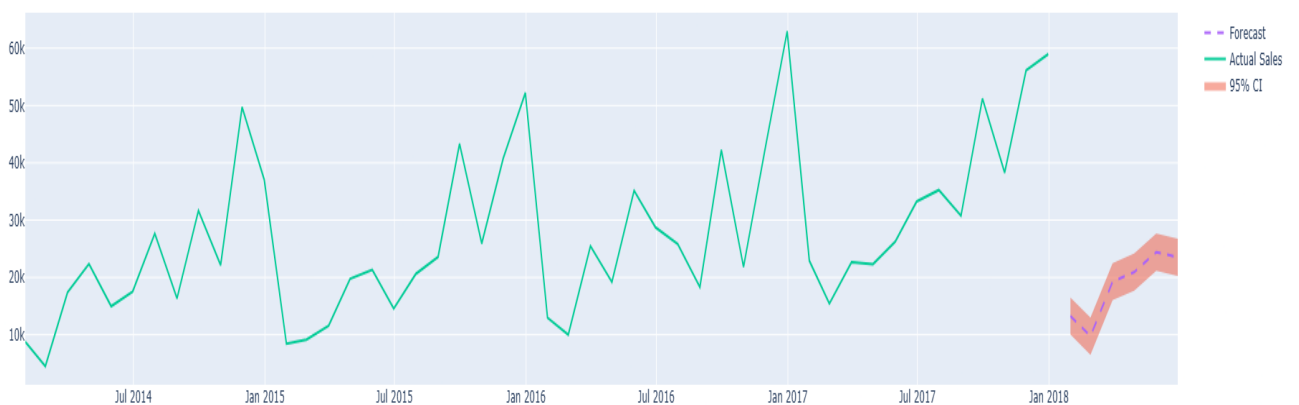


FIGURE 10: SIX MONTH SALES FORECAST

The model achieved a MAPE of 28.88%. While moderate, this accuracy level is acceptable for retail forecasting (Makridakis, 2020). Forecasts indicate short-term stabilization rather than aggressive growth.

Segment-Specific Forecasts

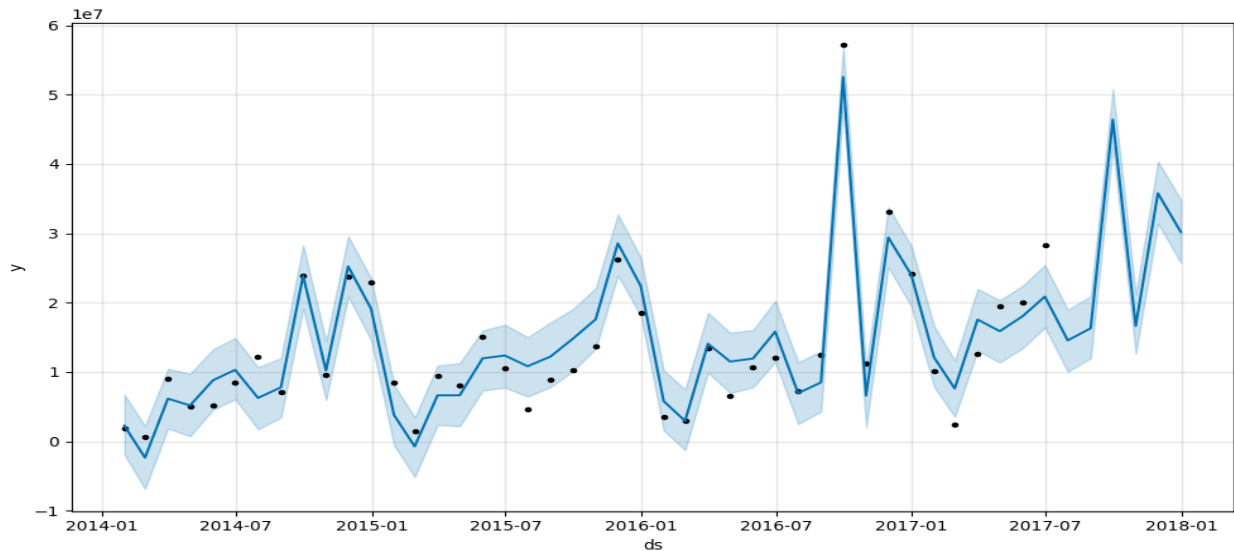


FIGURE 11: CHAMPION SEGMENT FORECAST

Champions show stable purchasing behaviour with consistent projected sales.

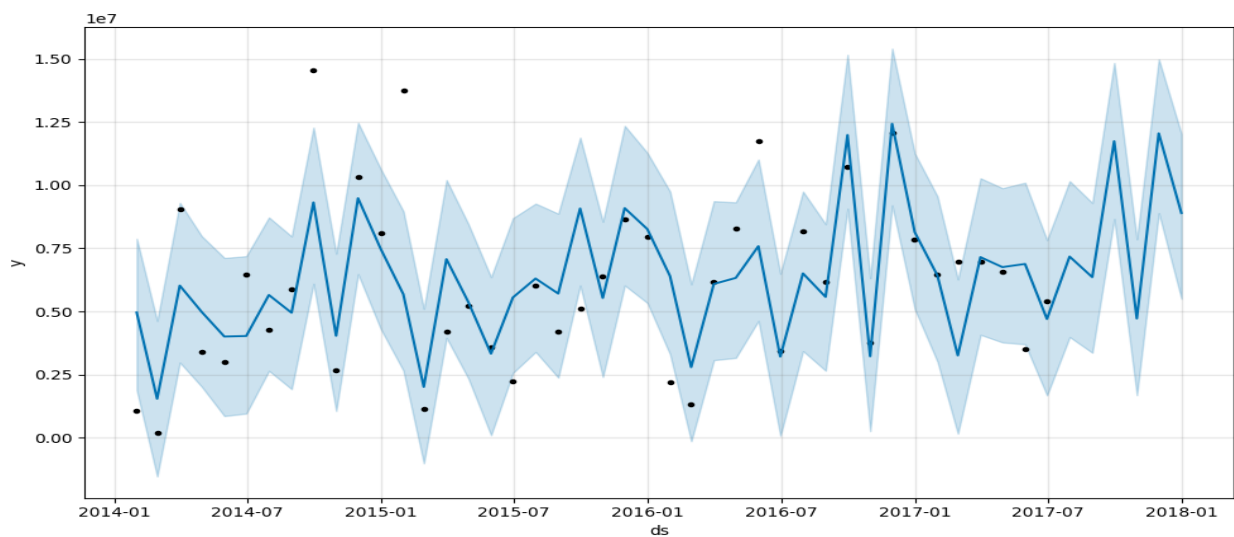


FIGURE 12: LOYAL SEGMENT FORECASTS

Loyal customers demonstrate moderate but predictable revenue patterns.

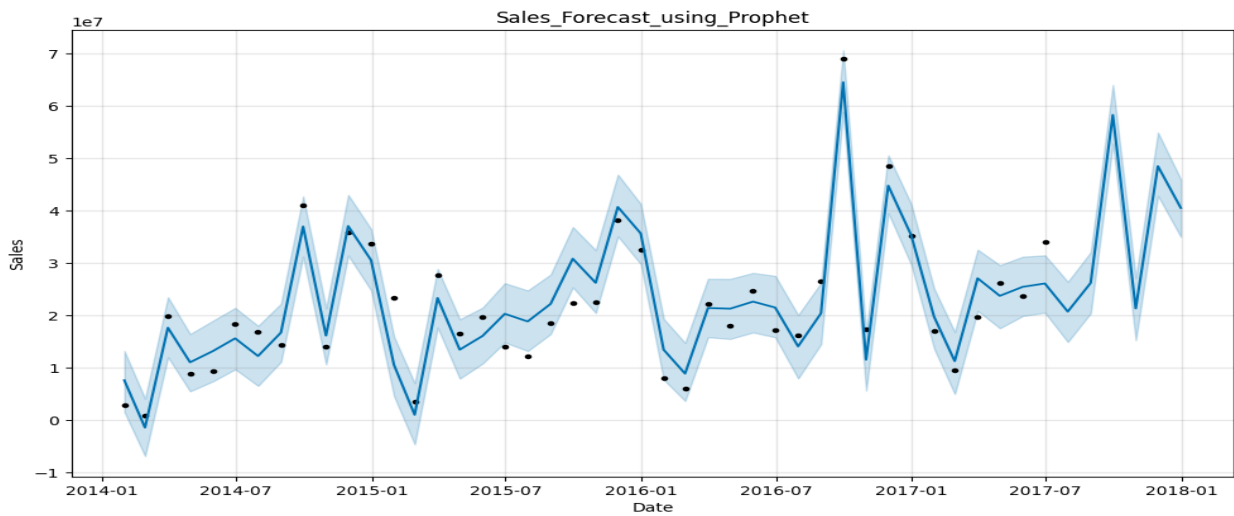


FIGURE 13: PROPHET-BASED FORECASTING

The Prophet model captures trend components similarly but introduces smoother seasonal adjustments.

Segment-level Forecasting Evaluation

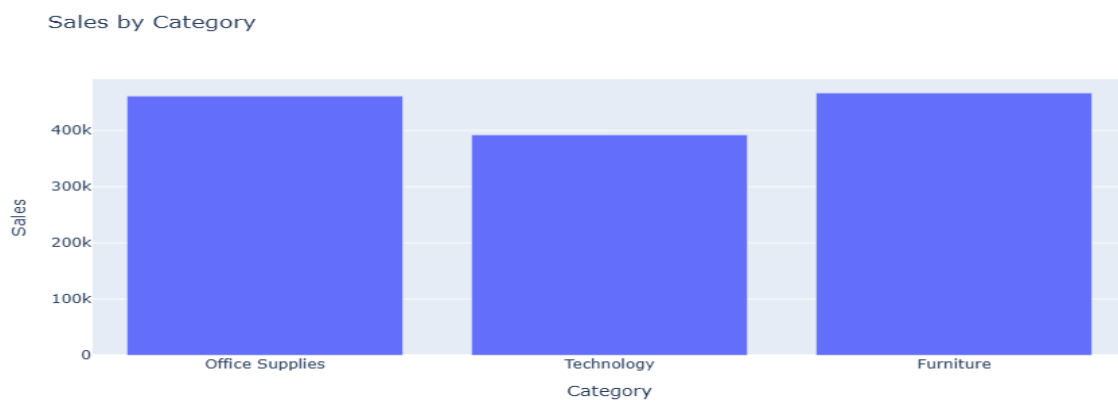


FIGURE 14: CATEGORY-LEVEL FORECAST EVALUATION

Technology category demonstrates higher volatility.

Segment evaluation (forecast KPIs)

Region	months	forecast_sum	last6_actual_sum	growth_pct	mape	rmse	status
West	48	36625.92	96452.87	-62.03	28.19	4147.01	ok
East	48	26224.59	82288.23	-68.1	24.64	3988.97	ok
Central	48	25137.72	53664.69	-53.16	14.91	1328.14	ok
South	48	23183.34	38389.75	-39.48	18.42	2258.75	ok

FIGURE 15: REGION-LEVEL EVALUATION

Central region shows lowest MAPE (~14.09%), indicating greater predictability compared to East (~24.6%).

Segment evaluation (forecast KPIs)

Segment	months	forecast_sum	last6_actual_sum	growth_pct	mape	rmse	status
Consumer	48	55899.77	132316.69	-57.72	13.21	3994.39	ok
Corporate	48	33972.1	88832.86	-57.97	19.94	3219.61	ok
Home Office	48	21259.69	57486	-63.82	36.21	3945.71	ok

FIGURE 16: SEGMENT-LEVEL EVALUATION

Consumer segment exhibits lower forecast error than home office.

These variations support differentiated regional inventory strategies.

Customer Segmentation (RFM)

Optimal Clustering

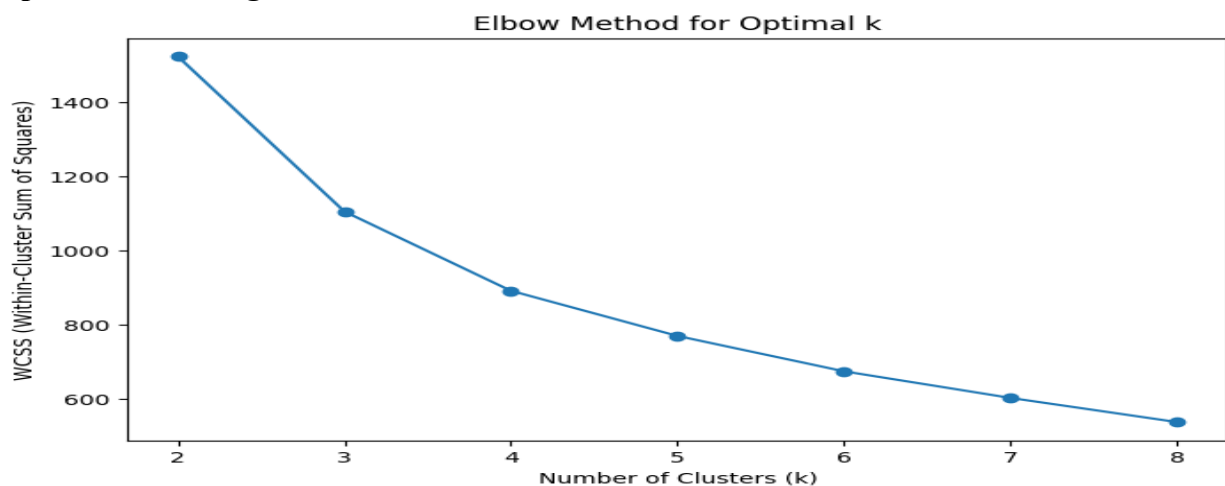


FIGURE 17: ELBOW METHOD CLUSTERING

The inflection point suggests k = 4 clusters.

Cluster Distribution

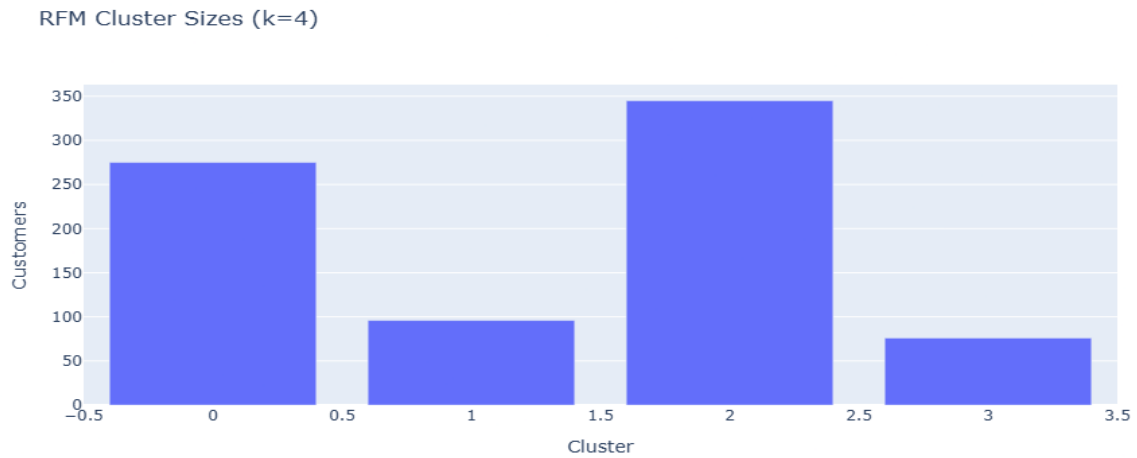


FIGURE 18: CLUSTER SIZE DISTRIBUTION

Cluster sizes reveal a concentration of revenue in high-value segments.

Behavioral Visualization

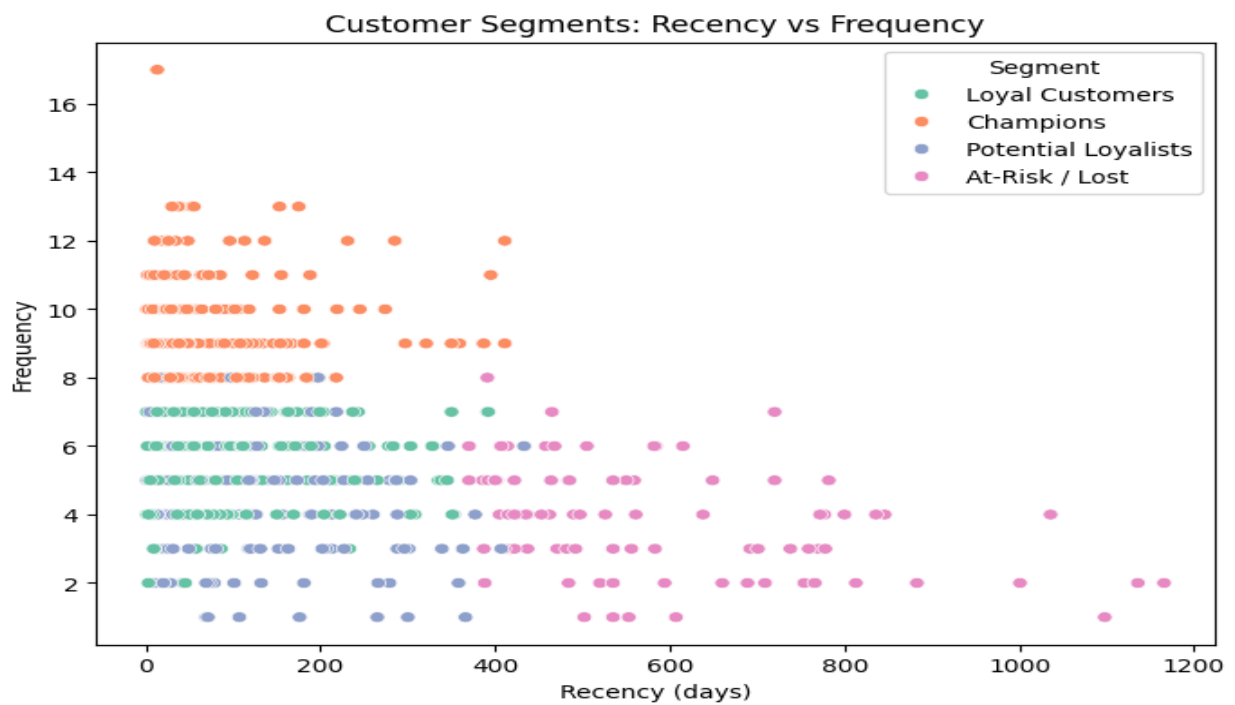


FIGURE 19: RECENCY VS FREQUENCY

Customers with low recency and high frequency represent Champions.

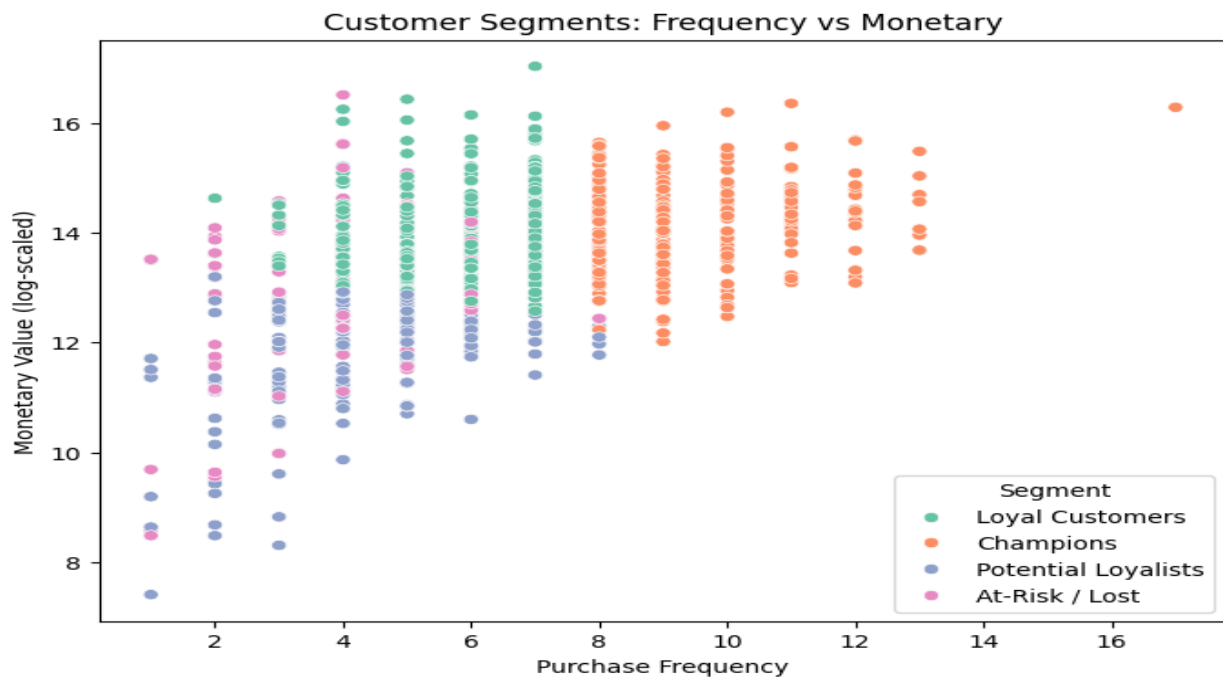


FIGURE 20: FREQUENCY VS MONETARY

High-monetary clusters represent big spenders.

Final Dashboard

The dashboard integrates:

- KPIs
- Forecast visualization
- RFM Clustering
- Segment evaluation
- Interactive filters

It functions as a unified retail decision-support tool.

Superstore Analytics Dashboard

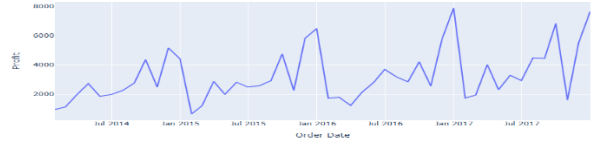
BI - Customer Segmentation - Forecasting (Analysis-ready dataset)



Monthly Sales Trend



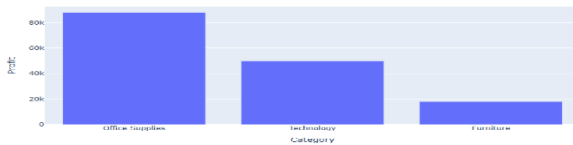
Monthly Profit Trend



Sales Forecast (Next 6 Months) | MAPE: 28.88%



Profit by Category



Sales by Category



Discount vs Profit (Sampled)

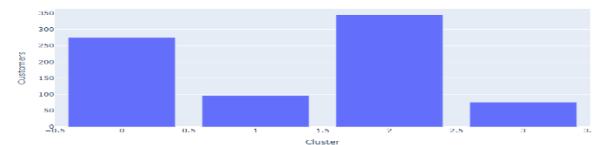


Business Insights (filtered view)

- Total Sales: 1,319,773 | Total Profit: 155,705 | Margin: 11.80%
- Top Category (Sales): Furniture | Top Category (Profit): Office Supplies
- Most Profitable Region: West
- Discount vs Profit Relationship: negative (corr = -0.38)
- Use these insights to prioritize profitable categories/regions and monitor discounting impacts

Customer Segmentation (RFM)

RFM Cluster Sizes (k=4)



Cluster	Recency	Frequency	Monetary	Customers	Segment
0	76.53	8.74	2676.69	275	Loss / At Risk
1	577.4	3.78	1168.9	86	Low1
2	182.87	5.18	1315.2	345	Low1
3	117.83	1.16	733.2	70	Champions

Interpretation: lower Recency + higher Frequency/Monetary typically indicates higher-value customers (e.g., Champions).

Segment evaluation (forecast KPIs)

Region	Months	Forecast_sum	last_actval_sum	growth_act	margin	resid	status
West	48	29929.02	29452.87	82.89	29.37	4347.03	OK
West	48	28224.02	82289.22	69.1	24.84	2028.07	OK
Central	48	23357.72	22884.89	57.39	34.93	3228.34	OK
South	48	22183.94	28389.75	-39.48	18.42	2258.75	OK

Filtered data preview (first 200 rows)

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City
1	CA-2015-132336	2015-11-09T00:00:00	2015-11-11T00:00:00	Second Class	C6-32929	Casidy Wells	Consumer	United States	Memph
2	CA-2015-132336	2015-11-09T00:00:00	2015-11-11T00:00:00	Second Class	C6-32929	Casidy Wells	Consumer	United States	Memph
3	CA-2015-132920	2015-09-12T00:00:00	2015-09-18T00:00:00	Second Class	OV-10969	Burton Van Hout	Consumer	United States	Los Ang
4	US-2015-138098	2015-10-11T00:00:00	2015-10-18T00:00:00	Standard Class	SO-28855	Sean O'Donnell	Consumer	United States	Port Lan
5	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
6	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
7	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
8	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
9	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
10	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
11	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang
12	CA-2014-115812	2014-06-09T00:00:00	2014-06-14T00:00:00	Standard Class	BH-11710	Bronnie Hoffman	Consumer	United States	Los Ang

Data SOURCE: F:\Academic\15th_Semester\Advanced_Analytics\data\superstore_data\Analysis\data\processed\superstore_processed.csv

FIGURE 21: FINAL DASHBOARD PREVIEW

Discussion

The results of this project demonstrate how combining descriptive analytics, forecasting, and customer segmentation in a single dashboard creates a more complete decision-support system than using each method independently. From a retail management perspective, the dashboard does not only describe “what happened”, but also supports answering “what is likely to happen next” and “which customers contribute most to value and risk”.

A key descriptive insight is the clear separation between sales volume and profitability. The category analysis (Figure 5 and 6) suggests that higher sales do not automatically imply higher profit. In practice, this is important because managers can over-prioritize revenue while overlooking margin efficiency. In this dataset, Office supplies appears to generate strong profit relative to its sales, whereas Furniture shows comparatively weaker profitability. This pattern is consistent with real retail environments, where product mix, discount sensitivity, and fulfilment costs can produce different margin structures even when sales are high. Therefore, category-level decisions (e.g. Promotion intensity, shelf focus, bundling) should incorporate profit contribution rather than sales alone.

Discounting emerges as another major lever. The negative discount–profit relationship (Figure 7 and Figure 8) indicates that discounting tends to erode profitability in this dataset. While discounts can drive short-term volume, the plots suggest diminishing or negative returns at higher discount levels. From a managerial standpoint, this implies the need for a more disciplined discount policy: discounts should be targeted to specific customer segments, time windows, or categories rather than applied broadly. The sampled dashboard view (Figure 9) also reinforces that even when sales are high, profitability may not follow—suggesting that promotional campaigns can inflate revenue at the expense of margin. A practical recommendation is to set guardrails such as maximum discount thresholds by category, and to monitor profitability impact immediately during campaign periods.

The forecasting component adds forward-looking capability. The six-month forecast (Figure 10) provides a short-term projection that can support inventory planning, staffing, and budgeting. The reported forecasting accuracy ($\text{MAPE} \approx 28.88\%$) indicates moderate performance. In retail forecasting, this level can be acceptable for high-level planning, especially when the underlying demand is volatile or affected by promotions. However, it also signals that forecast should be used as directional guidance rather than precise targets. This is exactly why the project includes holdout validation metrics and segment evaluation: the value is not only producing a forecast line, but helping decision-makers understand where the forecast is more reliable and where it is riskier.

The segment-level forecasting evaluation (Figures 14–16) is one of the most practically valuable parts of the dashboard because it reveals that forecast reliability is not uniform across business dimensions. For example, the regional evaluation shows Central as more predictable (lower MAPE) compared to East (higher MAPE). Operationally, this supports differentiated planning: regions with stable demand can operate with leaner safety stock and tighter replenishment cycles, while volatile regions require buffer inventory and more frequent forecast updates. Similarly, category-level differences imply that certain categories may require enhanced modelling or additional explanatory variables (e.g., promotional calendars), while others can be managed effectively with simpler models.

The segmentation component provides customer-level insight that complements the aggregate time-series view. The RFM analysis and clustering (Figures 17–20) identify meaningful behavioural groupings: customers who purchase recently and frequently (Champions) differ fundamentally from customers with long inactivity (At Risk/Lost). This segmentation is directly actionable. Champions represent the customers most likely to generate consistent future revenue, so retention strategies (loyalty offers, priority service, exclusive promotions) should focus there. At Risk customers should receive reactivation efforts (targeted discounting, win-back campaigns) but those efforts should be measured carefully given the earlier discount–profit findings. The Frequency vs Monetary visualization (Figure 20) also helps interpret “Big Spenders” patterns—customers who contribute high monetary value but may not purchase frequently. This can guide relationship strategies such as personalized outreach or premium bundles.

An important contribution of this project is the integration of these outputs into a unified interface (Figure 21). Instead of requiring separate notebooks for EDA, forecasting, and segmentation, the dashboard enables interactive exploration and decision support. Filters (date range, category, region, segment) allow users to test business questions dynamically, such as whether discount–profit behaviour changes by segment, or whether forecasting reliability shifts when focusing on a subset of the business.

At the same time, several methodological limitations should be considered when interpreting results. First, the forecasting approach is intentionally lightweight and emphasizes interpretability rather than maximum predictive accuracy. It does not include external regressors such as promotions, holidays, macroeconomic indicators, or competitor dynamics, which are known drivers of retail demand. Second, the dataset spans 2014–2017, which limits generalization beyond this period. Third, K-Means clustering assumes spherical clusters and relies on the fixed choice of $k=4$ for interpretability; alternative clustering methods (e.g., hierarchical or Gaussian mixture models) could produce different segmentation boundaries.

These limitations do not invalidate the results, but they define how the dashboard should be used: as a strong decision-support tool for exploration and short-term planning, not as a fully optimized enterprise forecasting system.

Overall, the project achieves its objective: it transforms transactional retail data into a structured, interactive analytics system that supports descriptive monitoring, predictive planning, and customer intelligence. The strongest practical contribution is not any single chart, but the combined workflow that enables managers to understand performance drivers (category/discount), anticipate near-term demand (forecasting), and prioritize customer strategy (RFM clusters) within one coherent framework.

Limitations

- Forecasting does not include external variables
- Dataset limited to 2014-2017
- Static clustering ($k = 4$)
- No demographic segmentation

Future research could integrate macroeconomic indicators and dynamic clustering technique.

Conclusion

The project successfully demonstrates the transformation of transactional retail data into a predictive segmentation-enabled Business Intelligence dashboard. The integration of forecasting and customer analytics elevates the system beyond descriptive reporting toward actionable strategic insight.

References

- Box, G. E. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Chen, H. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188. doi:<https://doi.org/10.2307/41703503>
- Fader, P. S. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415–430. doi:<https://doi.org/10.1509/jmkr.2005.42.4.415>
- Hyndman, R. J. (2018). *Forecasting: Principles and practice (2nd ed.)*. OTexts. doi:<https://doi.org/10.5072/zenodo.593231>
- James, G. W. (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.
- Kuhn, M. &. (2019). *Feature engineering and selection: A practical approach for predictive models*. Boca Raton, FL: CRC Press / Taylor & Francis.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, (pp. 281–297). Berkeley, CA.
- Makridakis, S. S. (2020). The M4 competition: Results, findings, and conclusions. *International Journal of Forecasting*, 36(1), 54–74. doi:<https://doi.org/10.1016/j.ijforecast.2019.04.014>
- Montgomery, D. C. (2015). *Introduction to time series analysis and forecasting (2nd ed.)*. Hoboken, NJ: Wiley. doi:<https://doi.org/10.1002/9781118745113>
- Moubray, J. (1997). *Reliability-centred maintenance*. Oxford: Butterworth-Heinemann.
- Sharda, R. D. (2018). *Business intelligence, analytics, and data science: A managerial perspective (4th ed.)*. New York, NY: Pearson.
- Wedel, M. &. (2000). *Market segmentation: Conceptual and methodological foundations (2nd ed.)*. New York, NY: Springer. doi:<https://doi.org/10.1007/978-1-4615-4651-1>