PROJECT TITLE:

SUPERMARKET SALES ANALYSIS

SUBTITLE:

A DATA ANALYSIS AND VISUALIZATION PROJECT

AUTHOR:

OGUNFUYI MOTUNRAYO MONISOLA

DATE:

02/ 01/ 2025

TABLE OF CONTENT

PROJECT OVERVIEW

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating
2	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	01/05/2019	13:08	Ewallet	522.83	4.761904762	26.1415	9.1
3	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	03/08/2019	10:29	Cash	76.4	4.761904762	3.82	9.6
4	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	03/03/2019	13:23	Credit card	324.31	4.761904762	16.2155	7.4
5	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761904762	23.288	8.4
6	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	02/08/2019	10:37	Ewallet	604.17	4.761904762	30.2085	5.3
7	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761904762	29.8865	4.1
8	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761904762	20.652	5.8
9	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761904762	36.78	8
10	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	01/10/2019	17:15	Credit card	72.52	4.761904762	3.626	7.2
11	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2/20/2019	13:27	Credit card	164.52	4.761904762	8.226	5.9
12	351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	02/06/2019	18:07	Ewallet	57.92	4.761904762	2.896	4.5
13	529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	03/09/2019	17:03	Cash	102.04	4.761904762	5.102	6.8
14	365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	02/12/2019	10:25	Ewallet	234.75	4.761904762	11.7375	7.1
15	252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	02/07/2019	16:48	Ewallet	431.9	4.761904762	21.595	8.2
16	829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	3/29/2019	19:21	Cash	713.8	4.761904762	35.69	5.7
17	299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	1/15/2019	16:19	Cash	562.32	4.761904762	28.116	4.5
18	656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	03/11/2019	11:03	Credit card	482.51	4.761904762	24.1255	4.6
19	765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	01/01/2019	10:39	Credit card	435.66	4.761904762	21.783	6.9
20	329-62-1586	A	Yangon	Normal	Male	Food and beverages	54.67	3	8.2005	172.2105	1/21/2019	18:00	Credit card	164.01	4.761904762	8.2005	8.6
21	319-50-3348	B	Mandalay	Normal	Female	Home and lifestyle	40.3	2	4.03	84.63	03/11/2019	15:30	Ewallet	80.6	4.761904762	4.03	4.4
22	300-71-4605	C	Naypyitaw	Member	Male	Electronic accessories	86.04	5	21.51	451.71	2/25/2019	11:24	Ewallet	430.2	4.761904762	21.51	4.8
23	371-85-5789	B	Mandalay	Normal	Male	Health and beauty	87.98	3	13.197	277.137	03/05/2019	10:40	Ewallet	263.94	4.761904762	13.197	5.1
24	273-16-6619	B	Mandalay	Normal	Male	Home and lifestyle	33.2	2	3.32	69.72	3/15/2019	12:20	Credit card	66.4	4.761904762	3.32	4.4
25	636-48-8204	A	Yangon	Normal	Male	Electronic accessories	34.56	5	8.64	181.44	2/17/2019	11:15	Ewallet	172.8	4.761904762	8.64	9.9
26	549-59-1358	A	Yangon	Member	Male	Sports and travel	88.63	3	13.2945	279.1845	03/02/2019	17:36	Ewallet	265.89	4.761904762	13.2945	6
27	227-03-5010	A	Yangon	Member	Female	Home and lifestyle	52.59	8	21.036	441.756	3/22/2019	19:20	Credit card	420.72	4.761904762	21.036	8.5
28	649-29-6775	B	Mandalay	Normal	Male	Fashion accessories	33.52	1	1.676	35.196	02/08/2019	15:31	Cash	33.52	4.761904762	1.676	6.7
29	189-17-4241	A	Yangon	Normal	Female	Fashion accessories	87.67	2	8.767	184.107	03/10/2019	12:17	Credit card	175.34	4.761904762	8.767	7.7
30	145-94-9061	B	Mandalay	Normal	Female	Food and beverages	88.36	5	22.09	463.89	1/25/2019	19:48	Cash	441.8	4.761904762	22.09	9.6
31	848-62-7243	A	Yangon	Normal	Male	Health and beauty	24.89	9	11.2005	235.2105	3/15/2019	15:36	Cash	224.01	4.761904762	11.2005	7.4
32	871-79-8483	B	Mandalay	Normal	Male	Fashion accessories	94.13	5	23.5325	494.1825	2/25/2019	19:39	Credit card	470.65	4.761904762	23.5325	4.8
33	149-71-6266	B	Mandalay	Member	Male	Sports and travel	78.07	9	35.1315	737.7615	1/28/2019	12:43	Cash	702.63	4.761904762	35.1315	4.5
34	640-49-2076	B	Mandalay	Normal	Male	Sports and travel	83.78	8	33.512	703.752	01/10/2019	14:49	Cash	670.24	4.761904762	33.512	5.1
35	595-11-5460	A	Yangon	Normal	Male	Health and beauty	96.58	2	9.658	202.818	3/15/2019	10:12	Credit card	193.16	4.761904762	9.658	5.1
36	183-56-6882	C	Naypyitaw	Member	Female	Food and beverages	99.42	4	19.884	417.564	02/06/2019	10:42	Ewallet	397.68	4.761904762	19.884	7.5
37	232-16-2483	C	Naypyitaw	Member	Female	Sports and travel	68.12	1	3.406	71.526	01/07/2019	12:28	Ewallet	68.12	4.761904762	3.406	6.8
38	129-29-8530	A	Yangon	Member	Male	Sports and travel	62.62	5	15.655	328.755	03/10/2019	19:15	Ewallet	313.1	4.761904762	15.655	7
39	272-65-1806	A	Yangon	Normal	Female	Electronic accessories	60.88	9	27.396	575.316	1/15/2019	17:17	Ewallet	547.92	4.761904762	27.396	4.7
40	333-73-7901	C	Naypyitaw	Normal	Female	Health and beauty	54.92	8	21.968	461.328	3/23/2019	13:24	Ewallet	439.36	4.761904762	21.968	7.6
41	777-82-7220	B	Mandalay	Member	Male	Home and lifestyle	30.12	8	12.048	253.008	03/03/2019	13:01	Cash	240.96	4.761904762	12.048	7.7
42	280-35-5823	B	Mandalay	Member	Female	Home and lifestyle	86.72	1	4.336	91.056	1/17/2019	18:45	Ewallet	86.72	4.761904762	4.336	7.9
43	554-53-8700	C	Naypyitaw	Member	Male	Home and lifestyle	56.11	2	5.611	117.831	02/02/2019	10:11	Cash	112.22	4.761904762	5.611	6.3
44	354-25-5821	B	Mandalay	Member	Female	Sports and travel	69.12	6	20.736	435.456	02/08/2019	13:03	Cash	414.72	4.761904762	20.736	5.6
45	228-96-1411	C	Naypyitaw	Member	Female	Food and beverages	98.7	8	39.48	829.08	03/04/2019	20:39	Cash	789.6	4.761904762	39.48	7.6
46	617-15-4209	C	Naypyitaw	Member	Male	Health and beauty	15.37	2	1.537	32.277	3/16/2019	19:47	Cash	30.74	4.761904762	1.537	7.2
47	132-32-9879	B	Mandalay	Member	Female	Electronic accessories	93.96	4	18.792	394.632	03/09/2019	18:00	Cash	375.84	4.761904762	18.792	9.5
48	370-41-7321	B	Mandalay	Member	Male	Health and beauty	56.69	9	25.5105	535.7205	2/27/2019	17:24	Credit card	510.21	4.761904762	25.5105	8.4

Excerpt from the dataset.

The **Supermarket Sales Analysis** project aimed to extract actionable insights from supermarket sales data to improve operational efficiency, enhance customer experience, and drive revenue growth. This project involved the use of advanced analytical tools and machine learning techniques to uncover trends, patterns, and anomalies in the dataset, which included key features such as sales, product lines, customer demographics, and transaction details.

Objectives

- To analyze the performance of different product lines, branches, and customer segments.
- To predict customer churn and identify key factors contributing to it.
- To visualize sales trends and patterns using interactive dashboards.
- To provide data-driven recommendations for business strategy optimization.

Tools and Technologies Used

- **Programming Languages:** Python (Pandas, NumPy, Seaborn, matplotlib, Plotly)
- **Visualization Tools:** Tableau

- **Machine Learning:** Random Forest, Support Vector Machine
- **Libraries:** Scikit-learn

Data Source:

Kaggle dataset containing sales data for three months across multiple branches.

Dataset Details:

- **Features:**
  - *Invoice ID*
  - *Branch* (A, B, C): Represents the store branch.
  - *City* (Yangon, Naypyitaw, Mandalay)
  - *Customer type* (Normal, Member)
  - *Gender*: Male/Female customers.
  - *Product Line*: Category of products.
  - *Unit price*
  - *Quantity*
  - *Tax 5%*
  - *Total*: Total transaction amount.
  - *Date*: Date of the transaction.
  - *Payment*: Payment method (Cash, Credit, etc.).
  - *Cogs* (Cost of Goods Sold)
  - *gross margin percentage* (it's a profitability metric that shows the percentage of a company's revenue that remains after subtracting the costs of producing or selling its goods and services)
  - *gross income*
  - *Rating*

Data Preprocessing

Handling Date and Month Information

In the preprocessing phase, the dataset was enriched with additional time-based information derived from the existing Date column. The following steps were undertaken:

1. **Date Conversion:**
    - The Date column was converted into a datetime object using Python's pandas.to_datetime function. This ensured that the date values were properly formatted and could be manipulated for further analysis.
2. **Extracting Month Information:**
    - A new column, Month, was created by extracting the month component from the Date column.
    - To make the data more readable, the numerical month values were mapped to their corresponding names (e.g., 1 to "January", 2 to "February").
3. **Exporting the Preprocessed Data:**
    - The updated dataset was exported to an Excel file (new_table.xlsx) to allow for easy sharing and further analysis.
    - The openpyxl library was used to handle the Excel export process.
4. **Validation:**
    - The os.getcwd() function was used to confirm the working directory and ensure the file was saved in the intended location.

Output:

The resulting Excel file contains the original dataset along with a new Month column, which lists the month names derived from the Date column. This transformation improves the dataset's usability and readability, particularly for time-based analyses such as monthly sales trends or seasonal patterns.

Churn Definition and Labeling

In addition to time-based preprocessing, a key feature, **Churn**, was engineered based on customer ratings. This feature provides insights into customer retention by identifying customers likely to churn.

1. **Defining Churn:**
   - A **Churn** column was created based on customer ratings.
   - A threshold of 5.0 was defined. Ratings below this threshold were labeled as 'Yes' (indicating churn), and ratings equal to or above the threshold were labeled as 'No'.
   - This binary classification helps in subsequent analysis and predictive modeling.
2. **Dataset Enrichment:**
   - This step added the Churn column to the dataset, providing a categorical indicator for customer behavior.
   - The new feature was used extensively in exploratory data analysis and predictive modeling to identify trends and patterns among churned customers.
3. **Validation:**
   - A preview of the updated dataset was obtained to verify the transformation.

Output:

The dataset was enhanced with a new Churn column, clearly identifying customers at risk of churn. This preprocessing step laid the foundation for effective churn analysis and prediction.

Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	ross margin percentag	gross income	Rating	Month	Churn
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2019-01-05 00:00:00	13:08	Ewallet	522.83	4.761904762	26.1415	9.1	January	No
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2019-03-08 00:00:00	10:29	Cash	76.4	4.761904762	3.82	9.6	March	No
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2019-03-03 00:00:00	13:23	Credit card	324.31	4.761904762	16.2155	7.4	March	No
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2019-01-27 00:00:00	20:33	Ewallet	465.76	4.761904762	23.288	8.4	January	No
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2019-02-08 00:00:00	10:37	Ewallet	604.17	4.761904762	30.2085	5.3	February	No
699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	2019-03-25 00:00:00	18:30	Ewallet	597.73	4.761904762	29.8865	4.1	March	Yes
355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2019-02-25 00:00:00	14:36	Ewallet	413.04	4.761904762	20.652	5.8	February	No
315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2019-02-24 00:00:00	11:38	Ewallet	735.6	4.761904762	36.78	8	February	No
665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	2019-01-10 00:00:00	17:15	Credit card	72.52	4.761904762	3.626	7.2	January	No
692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2019-02-20 00:00:00	13:27	Credit card	164.52	4.761904762	8.226	5.9	February	No
351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	2019-02-06 00:00:00	18:07	Ewallet	57.92	4.761904762	2.896	4.3	February	Yes
529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	2019-03-09 00:00:00	17:03	Cash	102.04	4.761904762	5.102	6.8	March	No
365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	2019-02-12 00:00:00	10:25	Ewallet	234.75	4.761904762	11.7375	7.1	February	No
252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	2019-02-07 00:00:00	16:48	Ewallet	431.9	4.761904762	21.595	8.2	February	No
829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	2019-03-29 00:00:00	19:21	Cash	713.8	4.761904762	35.69	5.7	March	No
299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	2019-01-15 00:00:00	16:19	Cash	562.32	4.761904762	28.116	4.5	January	Yes
656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	2019-03-11 00:00:00	11:03	Credit card	482.51	4.761904762	24.1255	4.6	March	Yes
765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	2019-01-01 00:00:00	10:39	Credit card	435.66	4.761904762	21.783	6.9	January	No
329-62-1586	A	Yangon	Normal	Male	Food and beverages	54.67	3	8.2005	172.2105	2019-01-21 00:00:00	18:00	Credit card	164.01	4.761904762	8.2005	8.6	January	No
319-50-3348	B	Mandalay	Normal	Female	Home and lifestyle	40.3	2	4.03	84.63	2019-03-11 00:00:00	15:30	Ewallet	80.6	4.761904762	4.03	4.4	March	Yes
300-71-4605	C	Naypyitaw	Member	Male	Electronic accessories	86.04	5	21.51	451.71	2019-02-25 00:00:00	11:24	Ewallet	430.2	4.761904762	21.51	4.8	February	Yes
371-85-5789	B	Mandalay	Normal	Male	Health and beauty	87.98	3	13.197	277.137	2019-03-03 00:00:00	10:40	Ewallet	263.94	4.761904762	13.197	5.1	March	No
273-16-6619	B	Mandalay	Normal	Male	Home and lifestyle	33.2	2	3.32	69.72	2019-03-13 00:00:00	12:20	Credit card	66.4	4.761904762	3.32	4.4	March	Yes
636-48-8704	A	Yangon	Normal	Male	Electronic accessories	34.56	5	8.64	181.44	2019-02-17 00:00:00	11:15	Ewallet	172.8	4.761904762	8.64	9.9	February	No
549-59-1358	A	Yangon	Member	Male	Sports and travel	88.63	3	13.2945	279.1845	2019-03-02 00:00:00	17:36	Ewallet	265.89	4.761904762	13.2945	6	March	No
227-03-5010	A	Yangon	Member	Female	Home and lifestyle	52.59	8	21.036	441.756	2019-03-22 00:00:00	19:20	Credit card	420.72	4.761904762	21.036	8.5	March	No
649-29-6775	B	Mandalay	Normal	Male	Fashion accessories	33.52	1	1.676	35.196	2019-02-08 00:00:00	15:31	Cash	33.52	4.761904762	1.676	6.7	February	No
189-17-4241	A	Yangon	Normal	Female	Fashion accessories	87.67	2	8.767	184.107	2019-03-10 00:00:00	12:17	Credit card	175.34	4.761904762	8.767	7.7	March	No
145-94-9061	B	Mandalay	Normal	Female	Food and beverages	88.36	5	22.09	463.89	2019-01-25 00:00:00	19:48	Cash	441.8	4.761904762	22.09	9.6	January	No
848-62-7342	A	Yangon	Normal	Male	Health and beauty	24.89	9	11.2005	235.2105	2019-03-15 00:00:00	15:36	Cash	224.01	4.761904762	11.2005	7.4	March	No
871-79-8483	B	Mandalay	Normal	Male	Fashion accessories	94.13	5	23.5325	494.1825	2019-02-25 00:00:00	19:39	Credit card	470.65	4.761904762	23.5325	4.8	February	Yes
149-71-6266	B	Mandalay	Member	Male	Sports and travel	78.07	9	35.1315	737.7615	2019-01-28 00:00:00	12:43	Cash	702.63	4.761904762	35.1315	4.5	January	Yes
640-49-2076	B	Mandalay	Normal	Male	Sports and travel	83.78	8	33.512	703.752	2019-01-10 00:00:00	14:49	Cash	670.24	4.761904762	33.512	5.1	January	No
595-11-5460	A	Yangon	Normal	Male	Health and beauty	96.38	2	9.638	202.818	2019-03-13 00:00:00	10:12	Credit card	193.16	4.761904762	9.638	5.1	March	No
183-56-6882	C	Naypyitaw	Member	Female	Food and beverages	99.42	4	19.884	417.564	2019-02-06 00:00:00	10:42	Ewallet	397.68	4.761904762	19.884	7.5	February	No
232-16-2483	C	Naypyitaw	Member	Female	Sports and travel	68.12	1	3.406	71.526	2019-01-07 00:00:00	12:28	Ewallet	68.12	4.761904762	3.406	6.8	January	No
129-29-8530	A	Yangon	Member	Male	Sports and travel	62.62	5	15.655	328.755	2019-03-10 00:00:00	19:13	Ewallet	313.1	4.761904762	15.655	7	March	No
272-65-1806	A	Yangon	Normal	Female	Electronic accessories	60.88	9	27.396	575.316	2019-01-15 00:00:00	17:17	Ewallet	547.92	4.761904762	27.396	4.7	January	Yes
333-73-7901	C	Naypyitaw	Normal	Female	Health and beauty	54.92	8	21.968	461.328	2019-03-23 00:00:00	13:24	Ewallet	439.36	4.761904762	21.968	7.6	March	No
777-82-7220	B	Mandalay	Member	Male	Home and lifestyle	30.12	8	12.048	253.008	2019-03-03 00:00:00	13:01	Cash	240.96	4.761904762	12.048	7.7	March	No
280-35-5823	B	Mandalay	Member	Female	Home and lifestyle	86.72	1	4.336	91.056	2019-01-17 00:00:00	18:45	Ewallet	86.72	4.761904762	4.336	7.9	January	No
554-53-8700	C	Naypyitaw	Member	Male	Home and lifestyle	56.11	2	5.611	117.831	2019-02-03 00:00:00	10:11	Cash	112.22	4.761904762	5.611	6.3	February	No
354-25-5821	B	Mandalay	Member	Female	Sports and travel	69.12	6	20.736	435.456	2019-02-08 00:00:00	13:03	Cash	414.72	4.761904762	20.736	5.6	February	No
238-96-1411	C	Naypyitaw	Member	Female	Food and beverages	98.7	8	39.48	829.08	2019-03-04 00:00:00	20:39	Cash	789.6	4.761904762	39.48	7.6	March	No
617-15-4209	C	Naypyitaw	Member	Male	Health and beauty	15.37	2	1.537	32.277	2019-03-16 00:00:00	19:47	Cash	30.74	4.761904762	1.537	7.2	March	No
132-32-9879	B	Mandalay	Member	Female	Electronic accessories	93.96	4	18.792	394.632	2019-03-09 00:00:00	18:00	Cash	375.84	4.761904762	18.792	9.5	March	No
370-41-7321	B	Mandalay	Member	Male	Health and beauty	56.69	9	25.5105	535.7205	2019-02-27 00:00:00	17:24	Credit card	510.21	4.761904762	25.5105	8.4	February	No

Excerpt from Excel 'Supermarket Sales Analysis New Table'

Exploratory Data Analysis (EDA)

Branch Performance

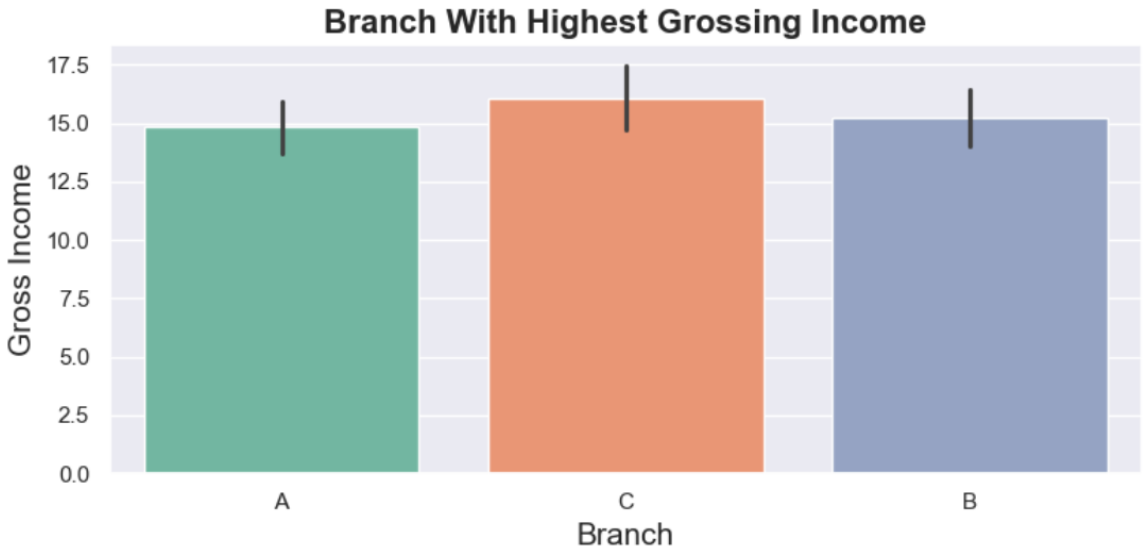
Visualization: Branch with Highest Grossing Income

To better understand the financial performance across different branches, a bar plot was created to compare the **gross income** generated by each branch. This visualization highlights the branch with the highest grossing income, aiding in branch-specific strategy formulation.

Key Details:

1. **Visualization Details:**
    - A bar plot was created using **Seaborn** for a clean and aesthetic representation.
    - The Branch column was used for the x-axis, while gross income represented the y-axis.
    - Each branch was uniquely colored using the Set2 palette for clarity.
2. **Insights:**

- The bar plot clearly shows the gross income performance for each branch, allowing stakeholders to identify the highest-performing branch.
- This visualization provides actionable insights for optimizing branch operations and allocating resources efficiently.



Visualization using Seaborn

Product Line Performance

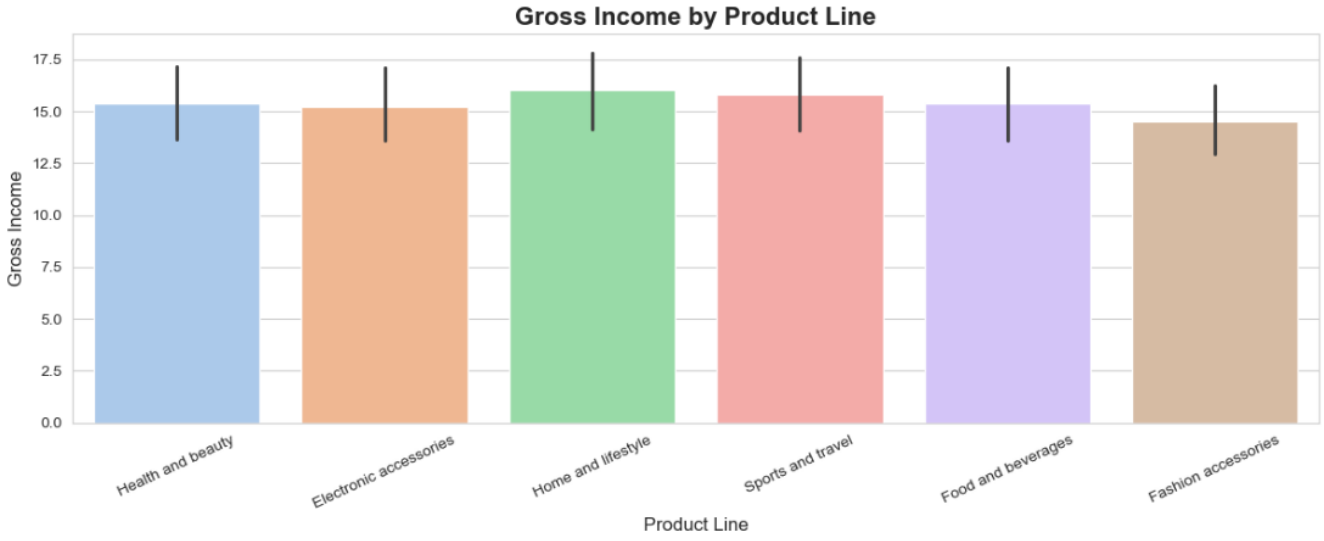
Visualization: Gross Income by Product Line

To identify the **Product line** generating the highest gross income, a bar chart was created using **Seaborn**. This visualization helps in understanding which product categories are the most profitable and can guide decisions on inventory and promotional strategies.

Key Details:

1. **Visualization Details:**
    - The x-axis represents the Product line, while the y-axis represents the gross income.
    - The hue parameter was utilized to assign unique pastel colors to each product line for better differentiation.
2. **Insights:**

- The bar chart highlights the product line with the highest gross income, allowing targeted focus on high-performing categories.
- This insight supports resource allocation and inventory management.



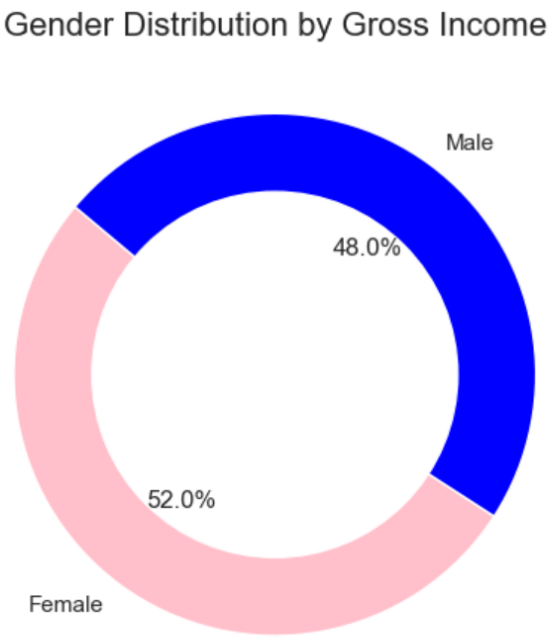
Excerpt from jupyter notebook

Gender Distribution by Gross Income

To analyze the contribution of **gender** to gross income, a doughnut chart was created, showcasing the percentage distribution. This visualization provides an overview of gender-based purchasing behavior.

Key Details:

1. **Visualization Details:**
    - The Gender column was grouped, and the gross income was summed for each category.
    - A **doughnut chart** (a modified pie chart) was created to visualize the proportional contribution.
    - Distinct colors (pink and blue) were used to represent genders for clear differentiation.
2. **Insights:**
    - The doughnut chart reveals the proportional contribution of each gender to gross income.
    - This information can be used to design gender-specific marketing campaigns and promotions.



Excerpt from Jupyter notebook using

Churn Analysis by Product Line and Branch

Visualization: Gross Income Analysis by Product Line and Branch with Churn

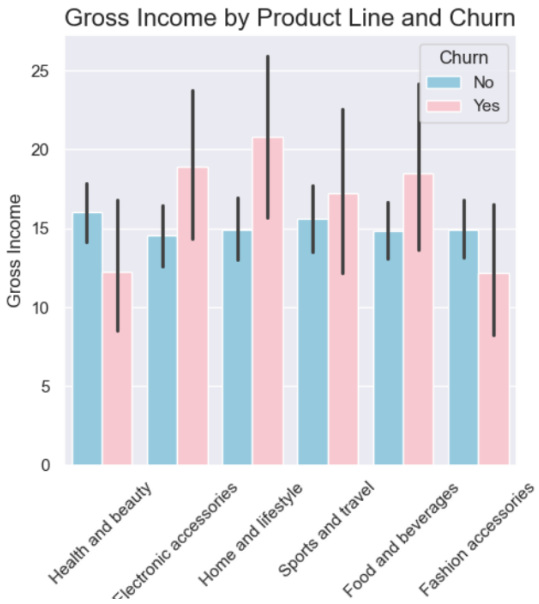
To understand the relationship between **gross income**, **product lines**, **branches**, and **churn**, dual bar plots were created. These visualizations help identify patterns in churn across different categories, providing actionable insights into customer retention strategies.

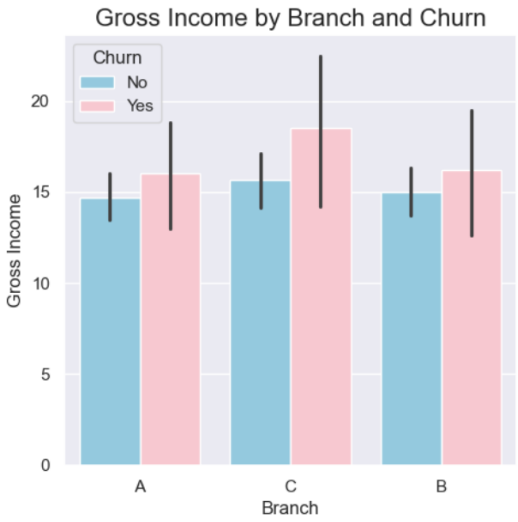
Key Details:

1. **Churn by Product Line:**
    - This bar chart visualizes the gross income across different product lines, categorized by churn status (Yes or No).
    - It provides insights into which product lines are more susceptible to churn.
2. **Churn by Branch:**
    - This bar chart showcases gross income by branch, with churn status highlighted.
    - It helps pinpoint branches with higher churn rates and assess their impact on gross income.

Insights:

1. **Gross Income by Product Line:**
    - Certain product lines may have higher churn rates (Yes), impacting overall gross income.
    - The comparison highlights opportunities to improve retention within specific product categories.
2. **Gross Income by Branch:**
    - Branches with higher churn rates can be targeted with customer engagement strategies to reduce attrition.
    - Branch-level analysis helps in understanding regional or location-specific challenges.


Gross Income by Product Line and Churn
Churn
No
Yes
Gross Income
0
5
10
15
20
25
Health and beauty
Electronic accessories
Home and lifestyle
Sports and travel
Food and beverages
Fashion accessories


Gross Income by Branch and Churn
Churn
No
Yes
Gross Income
0
5
10
15
20
A
C
B
Branch

Dataset Overview

1. Statistical Summary of the Data

A statistical summary was computed using the describe() function to provide an overview of the numerical features in the dataset. This includes metrics like count, mean, standard deviation, and percentiles.

Key Observations:

- **Unit Price**: The average price of a product is **55.67**, with prices ranging from **10.08** to **99.96**.
- **Quantity**: Customers typically purchase an average of **5.51 items**, with a maximum quantity of **10 items**.
- **Gross Income**: The average gross income per transaction is **15.38**, with a maximum of **49.65**.
- **Rating**: Customer ratings range from **4.0** to **10.0**, with an average of **6.97**.

2. Dataset Information

The info() function was used to examine the structure and data types of the dataset.

Key Details:

- The dataset consists of **1,000 entries** across **17 columns**.
- Columns include categorical variables (e.g., Branch, Gender) and numerical variables (e.g., Unit price, Gross Income).
- All columns have complete data (no missing values).

3. Memory Usage

The dataset uses approximately **132.9 KB** of memory.

This analysis ensures a comprehensive understanding of the data structure and its characteristics before proceeding with further exploration and modeling.

MACHINE LEARNING MODELS

Gross Income Prediction

1. Modeling Overview

The goal was to predict **gross income** using the **Tax 5%** feature as the independent variable. Two models were implemented and evaluated:

- **Random Forest Regressor**
- **Support Vector Machine (SVM) Regressor**

2. Data Splitting

The data was split into training and testing sets using a 70-30 split ratio with a fixed random state for reproducibility:

- **Training Set**: 70%
- **Testing Set**: 30%

3. Model Evaluation Metrics

- **Mean Absolute Error (MAE)**: Measures the average absolute difference between actual and predicted values.
- **R-squared (R^2)**: Indicates the proportion of variance explained by the model.

Results:

Model	MAE	R^2
Random Forest	0.0276	0.99997
SVM	0.1922	0.99575

4. Visualizing Model Predictions

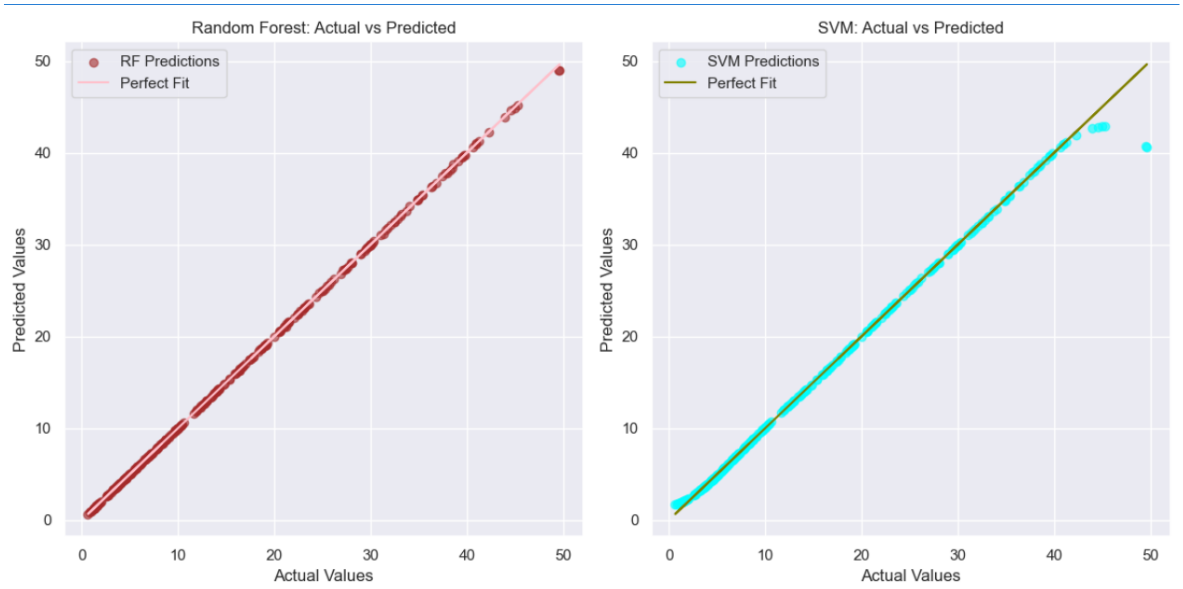
The relationship between actual and predicted values was visualized for both models to understand their performance.

- **Random Forest**:
    - The predictions closely align with the actual values, as shown by the scatter plot where points cluster near the diagonal line.

- **SVM**:
    - While the SVM model performs well, the points exhibit slightly more deviation from the diagonal line compared to Random Forest.

5. Conclusion

- The **Random Forest Regressor** outperformed the **SVM Regressor** in terms of both **MAE** and **R^2**.
- The superior performance of the Random Forest model suggests it is better suited for predicting **gross income** in this dataset.



Excerpt from Jupyter notebbok

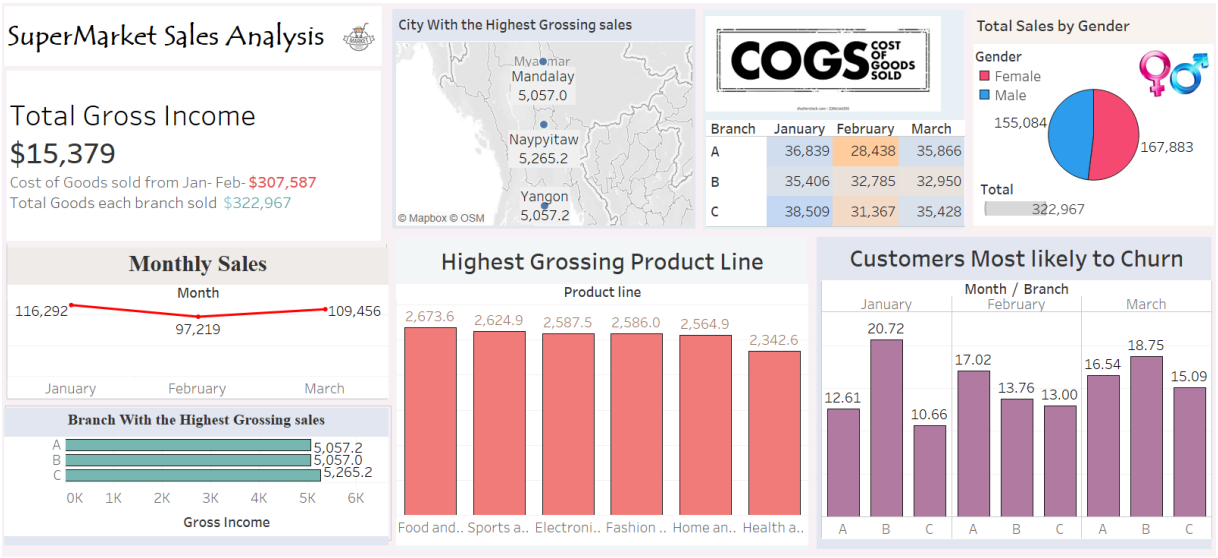
Tableau Dashboard

1. Overview

To provide an interactive and visually compelling analysis, a Tableau dashboard was created. This dashboard enables stakeholders to explore and derive insights from the supermarket sales data effectively. The visualizations focus on key metrics such as sales trends, product line performance, branch-wise analysis, and customer churn.

2. Key Features of the Dashboard

- **Gross Income by Product Line**
  A bar chart displaying the gross income generated by each product line, helping to identify the most profitable categories.
- **Branch Performance**
  A comparative analysis of gross income across branches to determine which branch generates the highest revenue.
- **Monthly Sales Trends**
  A line chart illustrating the sales performance over time, helping to uncover seasonal trends or patterns.
- **Customer Churn Analysis**
  A visual representation showing churn distribution across branches and product lines, enabling stakeholders to identify and address potential issues.



Supermarket sales analysis using Tableau

3. Insights Gained

- **Top Product Line**: Identified the product line with the highest gross income.
- **Branch Performance**: Determined the most profitable branch and factors influencing its success.
- **Customer Retention Challenges**: Highlighted the branches and product lines most affected by churn.
- **Seasonality Trends**: Observed periods with high or low sales activity.

6. Benefits of Tableau Visualization

- Easy-to-interpret graphs and charts.
- Enhanced decision-making capabilities for business stakeholders.

Results and Insights

Sales and Revenue Analysis

- **Highest Gross Income by Product Line:**
  The *Sports and travel* product line generated the highest gross income, making it the most profitable category.
- **Branch Performance:**
  Branch *C* had the highest gross income, outperforming other branches in overall sales.
- **Gender Contribution:**
  Female customers contributed slightly more to the gross income compared to male customers, as shown in the gender-based doughnut chart.
- **Monthly Sales Trends:**
  Sales peaked in **January**, suggesting a seasonal or promotional factor influencing customer purchases.

Customer Behavior and Churn Analysis

- **Churn Distribution by Product Line:**
  Customers purchasing *Home and lifestyle* products showed a higher likelihood of churn compared to other product lines.
- **Churn by Branch:**
  In *January,* Branch *B* experienced the highest churn rate, while in *February* and *March,* Branch *B experienced* the highest churning rate highlighting potential issues with customer satisfaction or service quality.

Machine Learning Model Results

- **Predicting Gross Income:**
  - The **Random Forest Regressor** achieved an exceptional R-squared value of **0.9999**, accurately predicting gross income with a mean absolute error of **0.0276**.
  - The **Support Vector Regressor** also performed well with an R-squared value of **0.9957** but was slightly less accurate compared to Random Forest.
- **Visualization of Predictions:**
  Scatter plots comparing actual versus predicted values demonstrated strong model performance, with points closely aligning to the diagonal *perfect fit* line.

Actionable Insights

- **Enhancing Product Lines:**
  The **Home and Lifestyle** product line generated the highest gross income, indicating its popularity and significant contribution to revenue. However, it also has the highest churn rate, signaling dissatisfaction among a segment of its customer base. This may be because of high customer volume, misaligned customer expectations or operational challenges. These may be resolved by conducting detailed surveys, improving quality and services or offering promotional services for At-risk customers.
- **Branch Strategy:**

  Branch C has the highest grossing income, demonstrating its role as a major revenue driver. However, it also experiences the highest churn rate, suggesting dissatisfaction among customers in this branch's region. This may be because of its high customer volume, difficulty in maintaining service quality, regional dynamics or operational challenges. The supermarket can engage in regular customer surveys, staff training and loyalty programs,

Tableau Insights

The Tableau dashboard provided an interactive platform to explore these trends, offering stakeholders a clear understanding of key metrics such as gross income distribution, churn rates, and monthly sales trends.

Conclusion

The supermarket sales analysis provided a comprehensive overview of key trends and patterns across product lines, branches, and customer behaviors. By leveraging advanced data preprocessing, exploratory data analysis, and machine learning models, uncovered actionable insights that can drive strategic decision-making.

Key findings include:

- **Home and Lifestyle** being the highest-grossing product line, yet exhibiting the highest churn rate.
- **Branch C**, despite generating the most revenue, also facing the highest customer churn.

These findings emphasize the need to balance revenue generation with customer retention by addressing underlying causes of dissatisfaction. Targeted interventions, such as improving service quality, optimizing inventory, and implementing customer loyalty programs, can mitigate churn while maintaining profitability.

Through this project, data-driven approaches highlighted the importance of understanding not only what contributes to revenue but also what risks the business faces. By taking proactive measures based on these insights, the supermarket can enhance customer satisfaction, increase retention, and sustain long-term growth.