Find a Dataset

Mukund Dhar
UCF ID: 5499369

Dataset: Supermarket sales
Link: https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales

This dataset includes information about supermarket sales from a retail chain in different stores and for different products. This dataset is interesting because it allows me to explore the factors that influence supermarket sales, which is a key business metric for retail chains. By analyzing this dataset, I can identify which factors are most strongly associated with sales amounts and make recommendations for marketing strategies that could improve sales. As someone who is interested in predictive and business analytics, this dataset aligns with my professional interests. By analyzing this dataset, I hope to gain insights into which factors have the greatest impact on supermarket sales and identify various patterns and trends to understand consumer behavior.

List of features of the dataset and the response variable:

Response variable: Total (Ratio/Interval)

Features:

Invoice ID (Nominal)
Branch (Nominal)
City (Nominal)
Product Line (Nominal)
Gross Margin Percentage (Ratio/Interval)
Gross Income (Ratio/Interval)
Unit Price (Ratio/Interval)
Quantity (Ratio/Interval)
Tax (Ratio/Interval)
Date (Ratio/Interval)
Time (Ratio/Interval)
Payment (Nominal)
COGS (Ratio/Interval)
Customer Type (Nominal)
Rating (Ratio/Interval)
Gender (Nominal)

Descriptive Information:

Invoice id: Computer generated sales slip invoice identification number.
Branch: Branch of supercenter (3 branches are available identified by A, B and C).
City: Location of supercenters
Customer type: Type of customers, recorded by Members for customers using member card and Normal for without member card.
Gender: Gender type of customer
Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel
Unit price: Price of each product in $
Quantity: Number of products purchased by customer
Tax: 5% tax fee for customer buying
Date: Date of purchase (Record available from January 2019 to March 2019)
Time: Purchase time (10am to 9pm)
Payment: Payment used by customer for purchase (3 methods are available – Cash, Credit card and Ewallet)
COGS: Cost of goods sold
Gross margin percentage: Gross margin percentage
Gross income: Gross income
Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

Total: Total price including tax

In [2]:
```
import pandas as pd
```

In [4]:
```
df = pd.read_csv('supermarket_sales.csv')
```

In [6]:
```
df.shape
```

Out[6]:
```
(1000, 17)
```

In [7]:
```
df.head(10)
```

Out[7]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.2200	3/
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.0480	1/2
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/
5	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/2
6	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.6520	433.6920	2/2
7	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.7800	772.3800	2/2
8	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.6260	76.1460	1/1
9	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.2260	172.7460	2/2

In [8]:
```
df.tail(10)
```

Out[8]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total
990	886-18-2897	A	Yangon	Normal	Female	Food and beverages	56.56	5	14.1400	296.9400
991	602-16-6955	B	Mandalay	Normal	Female	Sports and travel	76.60	10	38.3000	804.3000
992	745-74-0715	A	Yangon	Normal	Male	Electronic accessories	58.03	2	5.8030	121.8630
993	690-01-6631	B	Mandalay	Normal	Male	Fashion accessories	17.49	10	8.7450	183.6450
994	652-49-6720	C	Naypyitaw	Member	Female	Electronic accessories	60.95	1	3.0475	63.9975
995	233-67-5758	C	Naypyitaw	Normal	Male	Health and beauty	40.35	1	2.0175	42.3675
996	303-96-2227	B	Mandalay	Normal	Female	Home and lifestyle	97.38	10	48.6900	1022.4900
997	727-02-1313	A	Yangon	Member	Male	Food and beverages	31.84	1	1.5920	33.4320
998	347-56-2442	A	Yangon	Normal	Male	Home and lifestyle	65.82	1	3.2910	69.1110
999	849-09-3807	A	Yangon	Member	Female	Fashion accessories	88.34	7	30.9190	649.2990

In [10]:
```
df.describe()
```

Out[10]:

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	gross income	
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.00000	1000.000000	1000.000000	100
mean	55.672130	5.510000	15.379369	322.966749	307.58738	4.761905	15.379369	
std	26.494628	2.923431	11.708825	245.885335	234.17651	0.000000	11.708825	
min	10.080000	1.000000	0.508500	10.678500	10.17000	4.761905	0.508500	
25%	32.875000	3.000000	5.924875	124.422375	118.49750	4.761905	5.924875	
50%	55.230000	5.000000	12.088000	253.848000	241.76000	4.761905	12.088000	
75%	77.935000	8.000000	22.445250	471.350250	448.90500	4.761905	22.445250	
max	99.960000	10.000000	49.650000	1042.650000	993.00000	4.761905	49.650000	1

In [11]:
```
print(f'Number of rows with no missing values: {df.dropna().shape[0]}')
```

```
Number of rows with no missing values: 1000
```

In [12]:
```
print('Name: Mukund Dhar')
print('UCF ID: 5499369')
```

```
Name: Mukund Dhar
UCF ID: 5499369
```