IMPORTING LIBRARIES

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

DATA LOADING

```
data = pd.read_csv("supermart_data.csv")
```

DATA EXPLORATION (Pandas)

```
data.head()
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
0	OD1	Harish	Oil & Masala	Masalas	Vellore	11-08-2017	North	1254	0.12	401.28	Tamil Nadu
1	OD2	Sudha	Beverages	Health Drinks	Krishnagiri	11-08-2017	South	749	0.18	149.80	Tamil Nadu
2	OD3	Hussain	Food Grains	Atta & Flour	Perambalur	06-12-2017	West	2360	0.21	165.20	Tamil Nadu
3	OD4	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	10-11-2016	South	896	0.25	89.60	Tamil Nadu
4	OD5	Ridhesh	Food Grains	Organic Staples	Ooty	10-11-2016	South	2355	0.26	918.45	Tamil Nadu

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 11 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Order ID       9994 non-null   object
 1   Customer Name  9994 non-null   object
 2   Category       9994 non-null   object
 3   Sub Category   9994 non-null   object
 4   City           9994 non-null   object
 5   Order Date     9994 non-null   object
 6   Region         9994 non-null   object
 7   Sales          9994 non-null   int64
 8   Discount       9994 non-null   float64
 9   Profit         9994 non-null   float64
 10  State          9994 non-null   object
dtypes: float64(2), int64(1), object(8)
memory usage: 859.0+ KB
```

```
data.describe()
```

	Sales	Discount	Profit
count	9994.000000	9994.000000	9994.000000
mean	1496.596158	0.226817	374.937082
std	577.559036	0.074636	239.932881
min	500.000000	0.100000	25.250000
25%	1000.000000	0.160000	180.022500
50%	1498.000000	0.230000	320.780000
75%	1994.750000	0.290000	525.627500
max	2500.000000	0.350000	1120.950000

```
data.shape
```

```
(9994, 11)
```

```
data.columns
```

```
Index(['Order ID', 'Customer Name', 'Category', 'Sub Category', 'City',
       'Order Date', 'Region', 'Sales', 'Discount', 'Profit', 'State'],
      dtype='object')
```

```
data.dtypes
```

```
Order ID          object
Customer Name     object
```

```
Category          object
Sub Category      object
City              object
Order Date        object
Region            object
Sales              int64
Discount         float64
Profit           float64
State             object
dtype: object
```

DATA CLEANING

```
data.isnull().sum()
```

```
Order ID        0
Customer Name   0
Category        0
Sub Category    0
City            0
Order Date      0
Region          0
Sales           0
Discount        0
Profit          0
State           0
dtype: int64
```

```
data.dropna()
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
0	OD1	Harish	Oil & Masala	Masalas	Vellore	11-08-2017	North	1254	0.12	401.28	Tamil Nadu
1	OD2	Sudha	Beverages	Health Drinks	Krishnagiri	11-08-2017	South	749	0.18	149.80	Tamil Nadu
2	OD3	Hussain	Food Grains	Atta & Flour	Perambalur	06-12-2017	West	2360	0.21	165.20	Tamil Nadu
3	OD4	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	10-11-2016	South	896	0.25	89.60	Tamil Nadu
4	OD5	Ridhesh	Food Grains	Organic Staples	Ooty	10-11-2016	South	2355	0.26	918.45	Tamil Nadu
...	...	...	...	...	...	...	...	...	...	...	...
9989	OD9990	Sudeep	Eggs, Meat & Fish	Eggs	Madurai	12/24/2015	West	945	0.16	359.10	Tamil Nadu
9990	OD9991	Alan	Bakery	Biscuits	Kanyakumari	07-12-2015	West	1195	0.26	71.70	Tamil Nadu
9991	OD9992	Ravi	Food Grains	Rice	Bodi	06-06-2017	West	1567	0.16	501.44	Tamil Nadu
9992	OD9993	Peer	Oil & Masala	Spices	Pudukottai	10/16/2018	West	1659	0.15	597.24	Tamil Nadu
9993	OD9994	Ganesh	Food Grains	Atta & Flour	Tirunelveli	4/17/2018	West	1034	0.28	165.44	Tamil Nadu

9994 rows × 11 columns

```
data.drop_duplicates()
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
0	OD1	Harish	Oil & Masala	Masalas	Vellore	11-08-2017	North	1254	0.12	401.28	Tamil Nadu
1	OD2	Sudha	Beverages	Health Drinks	Krishnagiri	11-08-2017	South	749	0.18	149.80	Tamil Nadu
2	OD3	Hussain	Food Grains	Atta & Flour	Perambalur	06-12-2017	West	2360	0.21	165.20	Tamil Nadu
3	OD4	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	10-11-2016	South	896	0.25	89.60	Tamil Nadu
4	OD5	Ridhesh	Food Grains	Organic Staples	Ooty	10-11-2016	South	2355	0.26	918.45	Tamil Nadu
...	...	...	...	...	...	...	...	...	...	...	...
9989	OD9990	Sudeep	Eggs, Meat & Fish	Eggs	Madurai	12/24/2015	West	945	0.16	359.10	Tamil Nadu
9990	OD9991	Alan	Bakery	Biscuits	Kanyakumari	07-12-2015	West	1195	0.26	71.70	Tamil Nadu
9991	OD9992	Ravi	Food Grains	Rice	Bodi	06-06-2017	West	1567	0.16	501.44	Tamil Nadu
9992	OD9993	Peer	Oil & Masala	Spices	Pudukottai	10/16/2018	West	1659	0.15	597.24	Tamil Nadu
9993	OD9994	Ganesh	Food Grains	Atta & Flour	Tirunelveli	4/17/2018	West	1034	0.28	165.44	Tamil Nadu

9994 rows × 11 columns

```
data['Order Date'] = pd.to_datetime(data['Order Date'], format='mixed')
```

```
data['Category'].value_counts()
```

```
Category
Snacks              1514
Eggs, Meat & Fish   1490
Fruits & Veggies    1418
```

```
Bakery              1413
Beverages           1400
Food Grains         1398
Oil & Masala        1361
Name: count, dtype: int64
```

```
data['Sub Category'].value_counts()
```

```
Sub Category
Health Drinks         719
Soft Drinks           681
Cookies               520
Breads & Buns         502
Chocolates            499
Noodles               495
Masalas               463
Biscuits              459
Cakes                 452
Edible Oil & Ghee     451
Spices                447
Mutton                394
Eggs                  379
Organic Staples       372
Fresh Fruits          369
Fish                  369
Fresh Vegetables      354
Atta & Flour          353
Organic Fruits        348
Chicken               348
Organic Vegetables    347
Dals & Pulses         343
Rice                  330
Name: count, dtype: int64
```

DATA MANIPULATION

```
data['Sales'] + data['Discount']
```

```
0       1254.12
1        749.18
2       2360.21
3        896.25
4       2355.26
         ...
9989     945.16
9990    1195.26
9991    1567.16
9992    1659.15
9993    1034.28
Length: 9994, dtype: float64
```

```
data.sort_values(by = 'Sales', ascending=False)
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
9851	OD9852	Sundar	Beverages	Health Drinks	Cumbum	2018-10-12	Central	2500	0.25	325.0	Tamil Nadu
8249	OD8250	Komal	Bakery	Cakes	Trichy	2016-07-31	South	2500	0.13	225.0	Tamil Nadu
1412	OD1413	Peer	Fruits & Veggies	Organic Vegetables	Kanyakumari	2015-12-12	West	2500	0.14	1025.0	Tamil Nadu
9972	OD9973	Vinne	Snacks	Chocolates	Perambalur	2018-02-20	West	2500	0.19	325.0	Tamil Nadu
2443	OD2444	Ravi	Snacks	Chocolates	Dindigul	2018-10-30	Central	2500	0.33	300.0	Tamil Nadu
...	...	...	...	...	...	...	...	...	...	...	...
7931	OD7932	Willams	Oil & Masala	Edible Oil & Ghee	Karur	2017-12-25	West	501	0.14	200.4	Tamil Nadu
7525	OD7526	Anu	Fruits & Veggies	Organic Vegetables	Viluppuram	2017-04-12	South	500	0.17	225.0	Tamil Nadu
1376	OD1377	Rumaiza	Oil & Masala	Edible Oil & Ghee	Dharmapuri	2015-07-25	West	500	0.11	220.0	Tamil Nadu
3696	OD3697	Suresh	Eggs, Meat & Fish	Chicken	Perambalur	2018-12-11	East	500	0.29	45.0	Tamil Nadu
6669	OD6670	Sudeep	Snacks	Noodles	Madurai	2016-05-31	East	500	0.11	95.0	Tamil Nadu

9994 rows × 11 columns

```
data.groupby('Category')['Sales'].sum()
```

```
Category
Bakery              2112281
Beverages           2085313
Eggs, Meat & Fish   2267401
Food Grains         2115272
Fruits & Veggies    2100727
Oil & Masala        2038442
Snacks              2237546
Name: Sales, dtype: int64
```

```
data.groupby('Category')['Discount'].sum()
```

```
Category
Bakery               318.53
Beverages            322.46
Eggs, Meat & Fish    339.41
Food Grains          319.51
Fruits & Veggies     325.14
Oil & Masala         305.58
Snacks               336.18
Name: Discount, dtype: float64
```

DATA FILTERING

```
data[data['Sales'] >= 2500]
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
267	OD268	Arvind	Oil & Masala	Masalas	Virudhunagar	2017-01-22	East	2500	0.33	425.0	Tamil Nadu
1412	OD1413	Peer	Fruits & Veggies	Organic Vegetables	Kanyakumari	2015-12-12	West	2500	0.14	1025.0	Tamil Nadu
1877	OD1878	Mathew	Oil & Masala	Masalas	Dharmapuri	2016-01-30	West	2500	0.26	400.0	Tamil Nadu
2443	OD2444	Ravi	Snacks	Chocolates	Dindigul	2018-10-30	Central	2500	0.33	300.0	Tamil Nadu
2484	OD2485	Jonas	Beverages	Soft Drinks	Salem	2016-07-26	East	2500	0.34	175.0	Tamil Nadu
8249	OD8250	Komal	Bakery	Cakes	Trichy	2016-07-31	South	2500	0.13	225.0	Tamil Nadu
9851	OD9852	Sundar	Beverages	Health Drinks	Cumbum	2018-10-12	Central	2500	0.25	325.0	Tamil Nadu
9972	OD9973	Vinne	Snacks	Chocolates	Perambalur	2018-02-20	West	2500	0.19	325.0	Tamil Nadu

```
data[data['Sales'] <= 500]
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
1376	OD1377	Rumaiza	Oil & Masala	Edible Oil & Ghee	Dharmapuri	2015-07-25	West	500	0.11	220.0	Tamil Nadu
3696	OD3697	Suresh	Eggs, Meat & Fish	Chicken	Perambalur	2018-12-11	East	500	0.29	45.0	Tamil Nadu
6669	OD6670	Sudeep	Snacks	Noodles	Madurai	2016-05-31	East	500	0.11	95.0	Tamil Nadu
7525	OD7526	Anu	Fruits & Veggies	Organic Vegetables	Viluppuram	2017-04-12	South	500	0.17	225.0	Tamil Nadu

```
data[['Category','Sub Category']]
```

	Category	Sub Category
0	Oil & Masala	Masalas
1	Beverages	Health Drinks
2	Food Grains	Atta & Flour
3	Fruits & Veggies	Fresh Vegetables
4	Food Grains	Organic Staples
...	...	...
9989	Eggs, Meat & Fish	Eggs
9990	Bakery	Biscuits
9991	Food Grains	Rice
9992	Oil & Masala	Spices
9993	Food Grains	Atta & Flour

9994 rows × 2 columns

```
data[(data['Sub Category'] == 'Chocolates') & (data['Sales'] >= 2480)]
```

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
947	OD948	Amy	Snacks	Chocolates	Tenkasi	2018-11-28	East	2480	0.19	248.00	Tamil Nadu
2443	OD2444	Ravi	Snacks	Chocolates	Dindigul	2018-10-30	Central	2500	0.33	300.00	Tamil Nadu
3909	OD3910	Shree	Snacks	Chocolates	Ooty	2018-04-10	East	2481	0.11	669.87	Tamil Nadu
8996	OD8997	Sheeba	Snacks	Chocolates	Vellore	2018-12-10	South	2486	0.34	969.54	Tamil Nadu
9972	OD9973	Vinne	Snacks	Chocolates	Perambalur	2018-02-20	West	2500	0.19	325.00	Tamil Nadu

```
data.to_excel('cleaned_data.xlsx')
```

STATISTICAL ANALYSIS

```
# round(np.mean(data['Sales']),2)
np.mean(data['Sales'])
```

```
1496.5961576946167
```

```
np.median(data['Sales'])
```

```
1498.0
```

```
np.mean(data['Profit'])
```

```
374.9370822493496
```

```
np.median(data['Profit'])
```

```
320.78
```

```
np.std(data['Sales'])
```

```
577.5301402065124
```

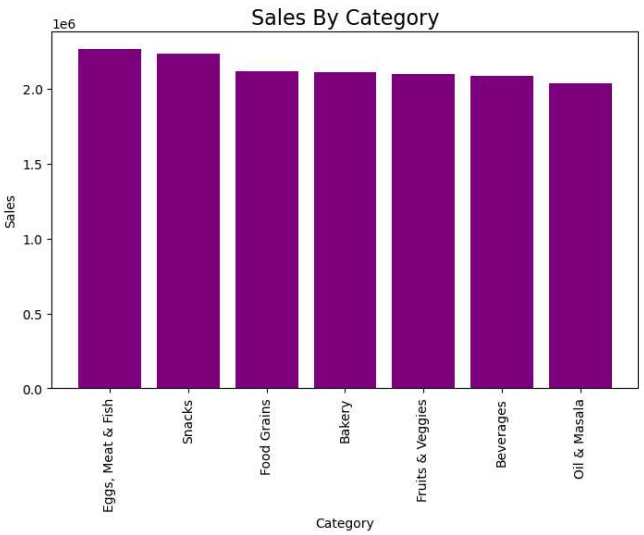
```
np.std(data['Profit'])
```

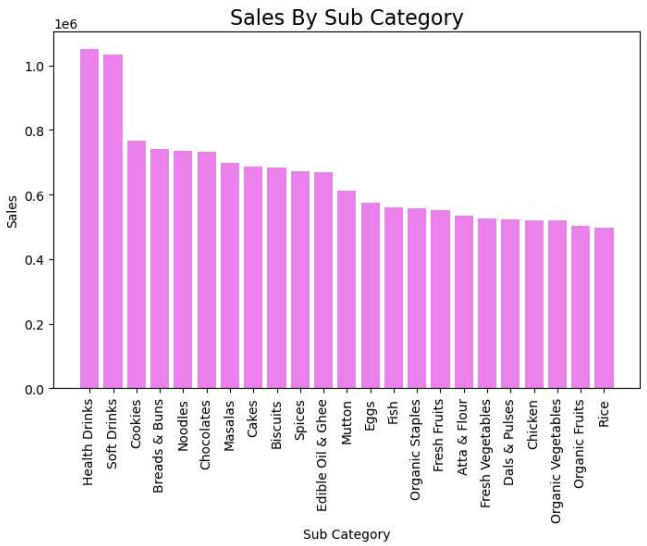
```
239.92087717740438
```

DATA VISUALIZATION

1. Sales By Category and Sub Category

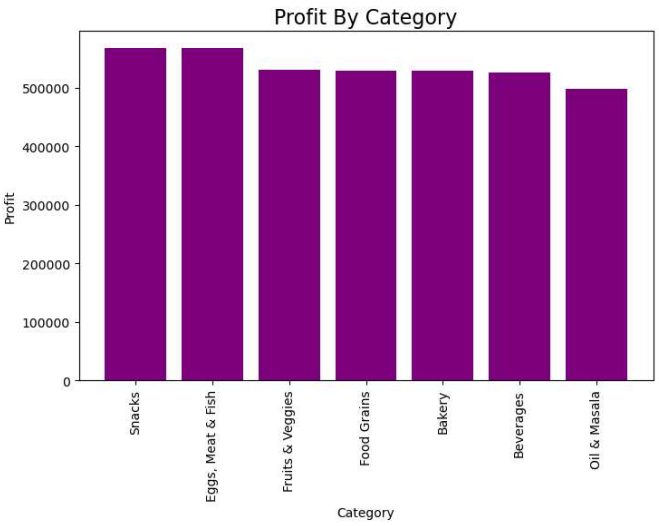
```
sales_category = data.groupby('Category')['Sales'].sum().reset_index().sort_values(by= 'Sales', ascending=False)
sales_sub_category = data.groupby('Sub Category')['Sales'].sum().reset_index().sort_values(by= 'Sales', ascending=False)

plt.figure(figsize = (18,5))

plt.subplot(1,2,1)
plt.bar(sales_category['Category'], sales_category['Sales'], color = 'purple')
plt.title('Sales By Category', fontsize=16)
plt.xlabel('Category')
plt.ylabel('Sales')
plt.xticks(rotation = 90)

plt.subplot(1,2,2)
plt.bar(sales_sub_category['Sub Category'], sales_sub_category['Sales'], color = 'violet')
plt.title('Sales By Sub Category', fontsize=16)
plt.xlabel('Sub Category')
plt.ylabel('Sales')
plt.xticks(rotation = 90)

plt.show()
```

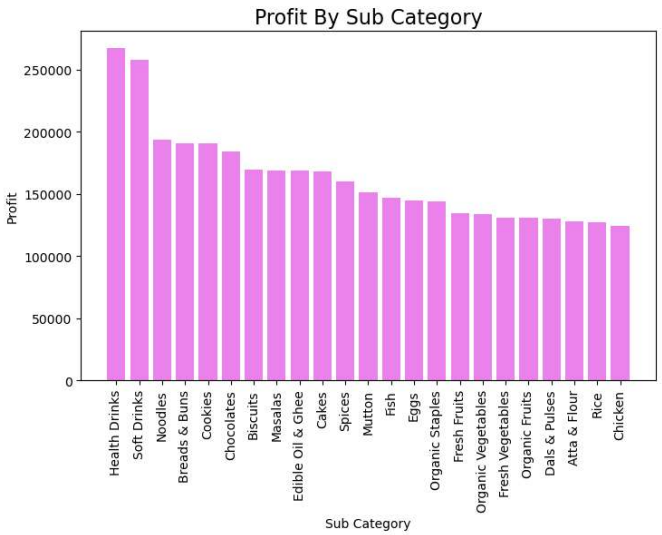




2. Profit By Category and Sub Category

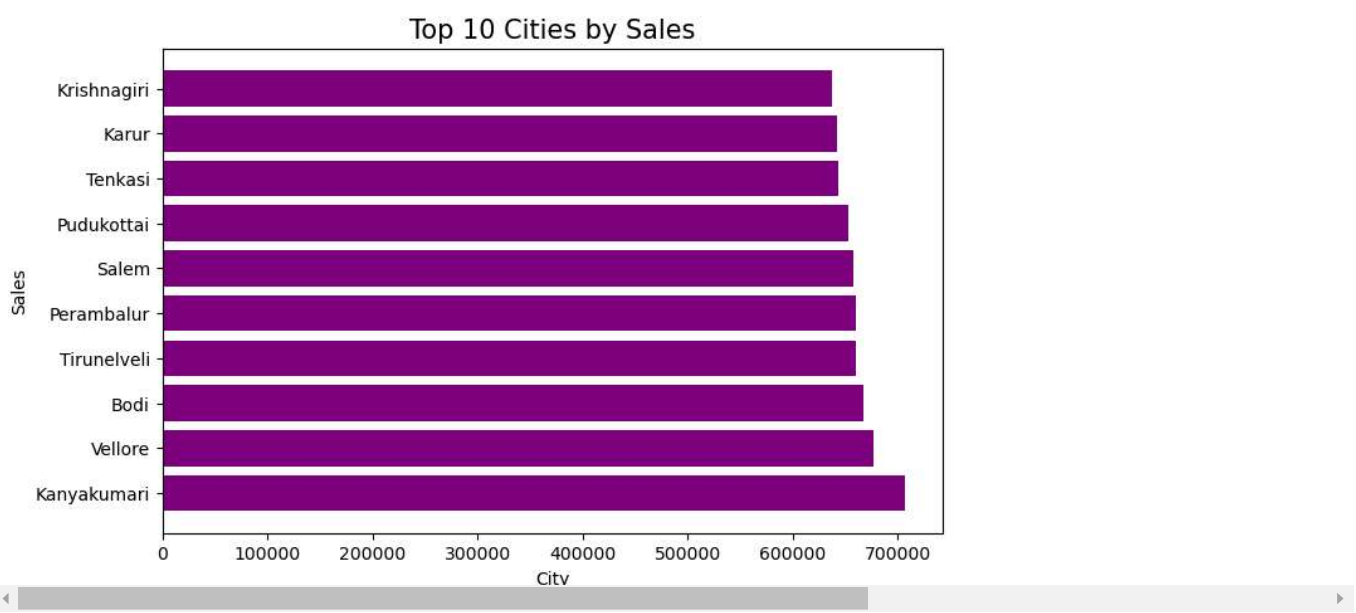
```
profit_category = data.groupby('Category')['Profit'].sum().reset_index().sort_values(by= 'Profit', ascending=False)
profit_sub_category = data.groupby('Sub Category')['Profit'].sum().reset_index().sort_values(by= 'Profit', ascending=False)

plt.figure(figsize = (18,5))

plt.subplot(1,2,1)
plt.bar(profit_category['Category'], profit_category['Profit'], color = 'purple')
plt.title('Profit By Category', fontsize=16)
plt.xlabel('Category')
plt.ylabel('Profit')
plt.xticks(rotation = 90)

plt.subplot(1,2,2)
plt.bar(profit_sub_category['Sub Category'], profit_sub_category['Profit'], color = 'violet')
plt.title('Profit By Sub Category', fontsize=16)
plt.xlabel('Sub Category')
plt.ylabel('Profit')
plt.xticks(rotation = 90)

plt.show()
```





3. Top Cities by Sales

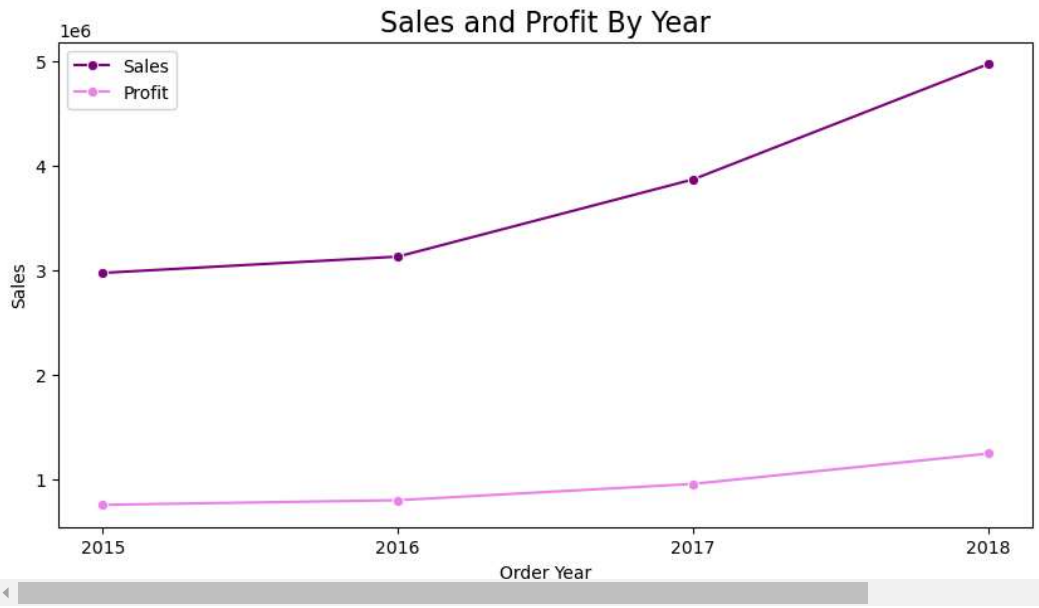
```
city_sales = data.groupby('City')['Sales'].sum().reset_index()             # calculating total sales per city
top_cities = city_sales.sort_values(by = 'Sales', ascending = False)[:10]   # selecting top 10 cities

plt.figure(figsize=(8,5))
plt.barh(top_cities['City'], top_cities['Sales'], color = 'purple')
plt.title('Top 10 Cities by Sales', fontsize=15)
plt.xlabel('City')
plt.ylabel('Sales')
plt.show()
```



4. Sales and Profit Over Years

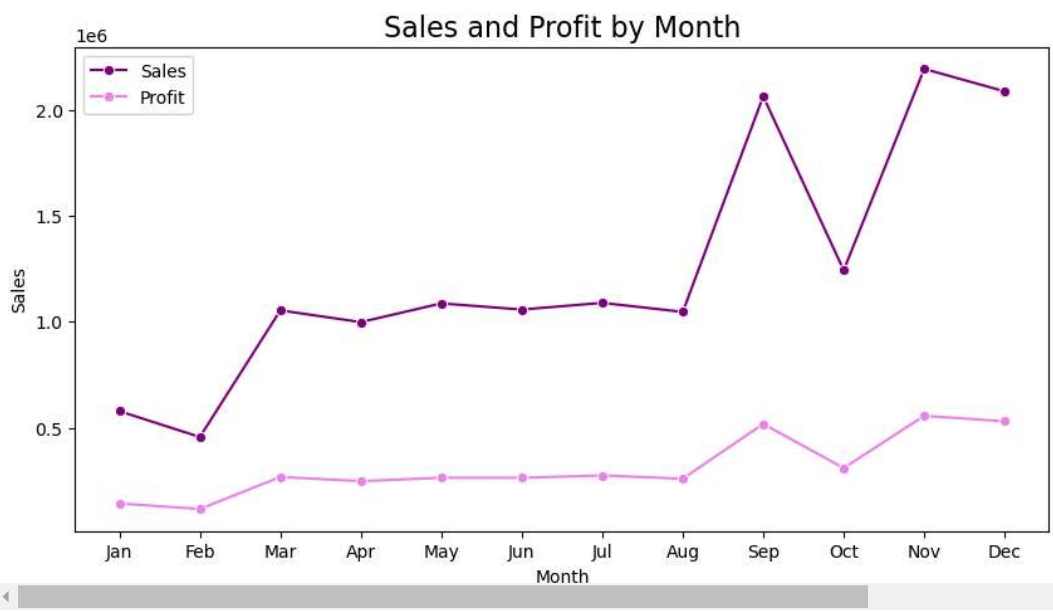
```
data['Order Year'] = data['Order Date'].dt.year
yearly_sales = data.groupby('Order Year')['Sales'].sum()
yearly_profit = data.groupby('Order Year')['Profit'].sum()

plt.figure(figsize = (10,5))
plt.title('Sales and Profit By Year', fontsize = 16)

sns.lineplot(yearly_sales, label ='Sales',marker= 'o', markerfacecolor = 'purple', color='purple')
sns.lineplot(yearly_profit,  label='Profit', marker= 'o', markerfacecolor = 'violet', color='violet')
plt.xticks(ticks=yearly_profit.index, labels=yearly_profit.index, fontsize=10)
plt.legend()
```

```
plt.show()
```



5. Sales and Profit Over Months

```
data['Order Month'] = data['Order Date'].dt.month
monthly_sales = data.groupby('Order Month')['Sales'].sum().reset_index()   # Sum up sales by month
monthly_profit = data.groupby('Order Month')['Profit'].sum().reset_index()   # Sum up profit by month

# Create the line chart
plt.figure(figsize=(10, 5))
plt.title('Sales and Profit by Month', fontsize=16)

sns.lineplot(x = monthly_sales['Order Month'], y = monthly_sales['Sales'], marker= 'o', markerfacecolor = 'purple', color='purple', label='Sales
sns.lineplot(x = monthly_profit['Order Month'], y = monthly_profit['Profit'], marker= 'o', markerfacecolor = 'violet', color='violet', label='Pr
plt.xlabel('Month')
plt.ylabel('Sales')
plt.xticks(monthly_sales['Order Month'], ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec'])
plt.legend()

plt.show()
```



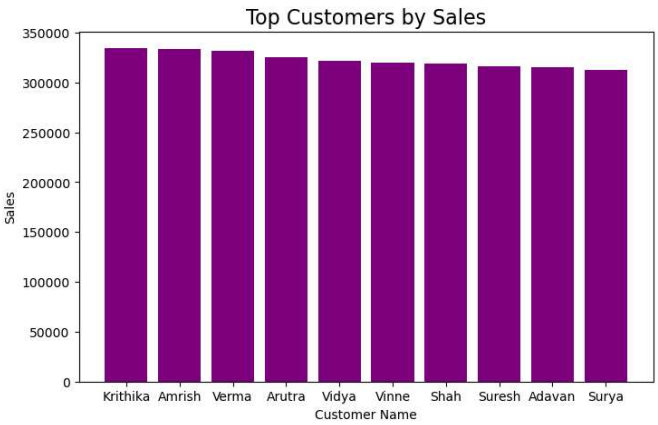
6. Sales, Profit, Discount Distribution over Region

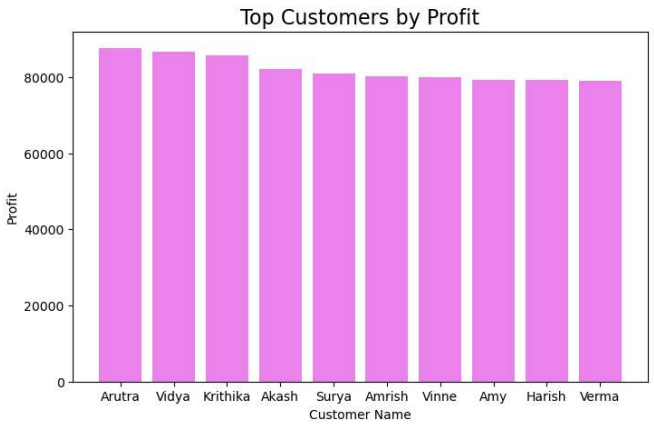
```
region_summary = data.groupby('Region')[['Sales', 'Profit', 'Discount']].sum()

region_summary.plot(
    kind='bar',
    stacked=True,
    figsize=(12, 6),
    color=['purple', 'violet', 'lightblue'],  # Colors for Sales, Profit, and Discount
    alpha=0.8
```

```
)

plt.title('Sales, Profit, and Discount Distribution by Region', fontsize=15)
plt.xlabel('Region', fontsize=12)
plt.ylabel('Amount', fontsize=12)
plt.legend(fontsize=10)
plt.xticks(rotation=45)

plt.show()
```

7. Top Customers by Sales and Profit

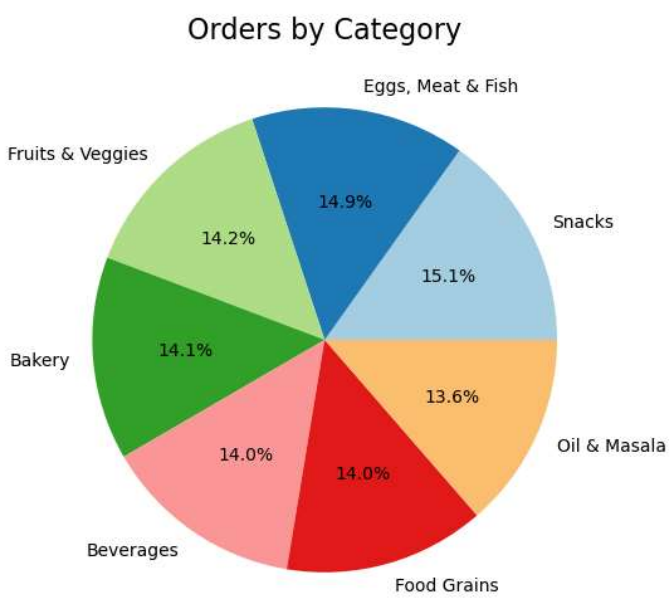
```
customer_sales = data.groupby('Customer Name')['Sales'].sum().reset_index().sort_values(by= 'Sales', ascending=False)[:10]
customer_profit = data.groupby('Customer Name')['Profit'].sum().reset_index().sort_values(by= 'Profit', ascending=False)[:10]

plt.figure(figsize = (18,5))

plt.subplot(1,2,1)
plt.bar(customer_sales['Customer Name'], customer_sales['Sales'], color = 'purple')
plt.title('Top Customers by Sales', fontsize=16)
plt.xlabel('Customer Name')
plt.ylabel('Sales')
plt.xticks()

plt.subplot(1,2,2)
plt.bar(customer_profit['Customer Name'], customer_profit['Profit'], color = 'violet')
plt.title('Top Customers by Profit', fontsize=16)
plt.xlabel('Customer Name')
plt.ylabel('Profit')
plt.xticks()

plt.show()
```





8. Orders by Category

```
category_counts = data['Category'].value_counts()

plt.figure(figsize=(6,6))
plt.title('Orders by Category', fontsize=16)

plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', colors=plt.cm.Paired.colors)
plt.show()
```

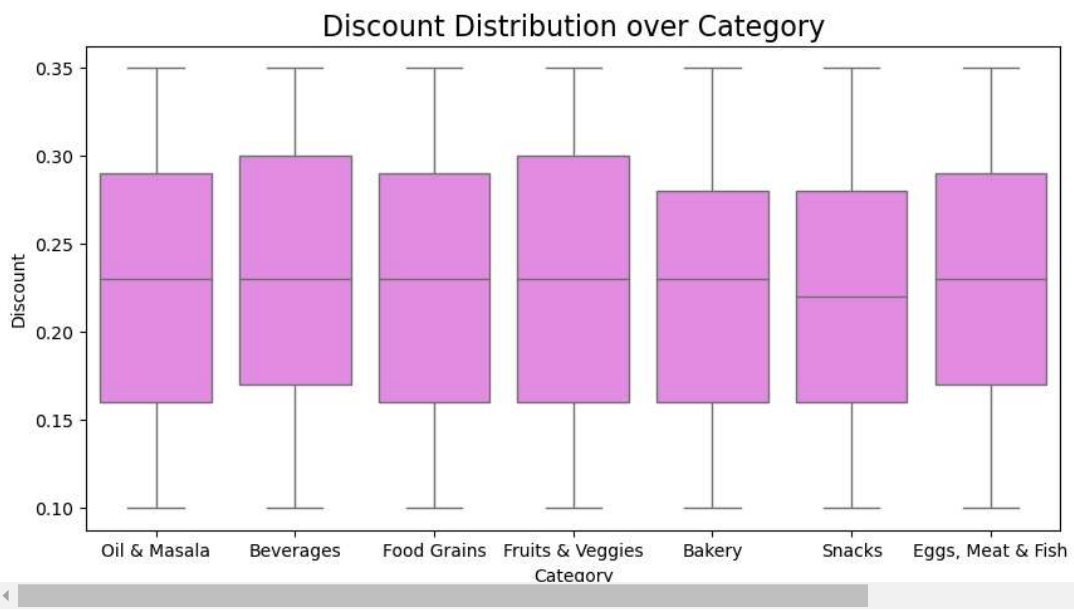


9. Correlation between Discount and Profit

```
plt.figure(figsize=(10, 6))
sns.scatterplot(
    x=data['Discount'],
    y=data['Profit'],
    alpha=0.2,
    color='purple'
)

plt.title('Correlation Between Discount and Profit', fontsize=16)
plt.xlabel('Discount', fontsize=12)
plt.ylabel('Profit', fontsize=12)


plt.show()
```

10. Distribution of Discounts offered across Categories

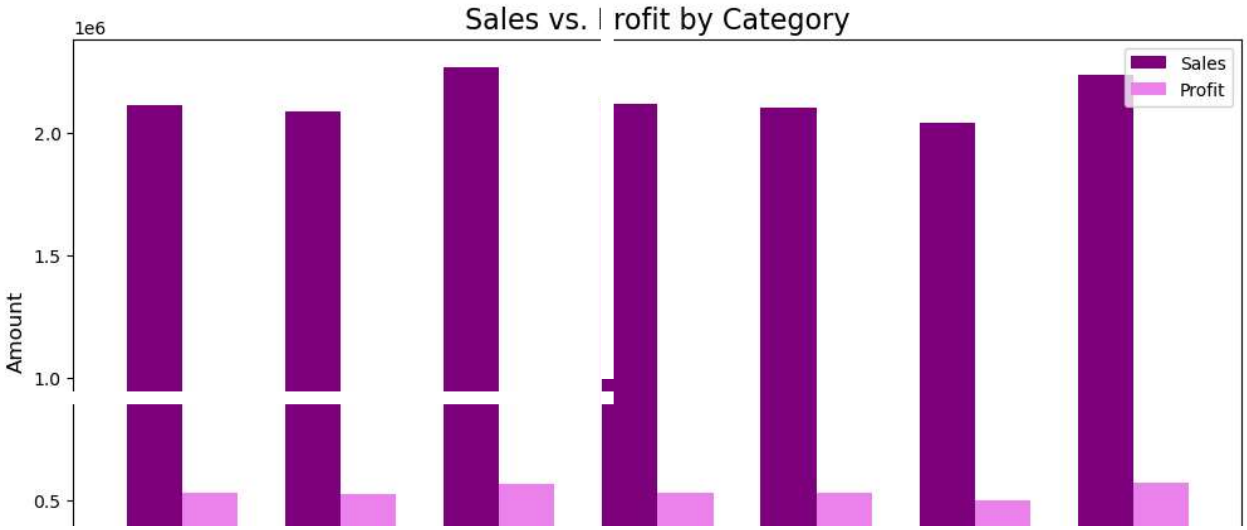
```
plt.figure(figsize=(10,5))
plt.title('Discount Distribution over Category', fontsize=16)

sns.boxplot(data=data, x = 'Category', y= 'Discount', color='violet')
plt.show()
```



11. Sales vs. Profit across Categories

```
category_summary = data.groupby('Category')[['Sales','Profit']].sum(
categories = category_summary.index
sales = category_summary['Sales']
profits = category_summary['Profit']

x = np.arange(len(categories))
width = 0.35

plt.figure(figsize=(12, 6))
plt.bar(x - width/2, sales, width, label='Sales', color='purple')
plt.bar(x + width/2, profits, width, label='Profit', color='violet')

plt.title('Sales vs. Profit by Category', fontsize=16)
plt.xlabel('Category', fontsize=12)
plt.ylabel('Amount', fontsize=12)
plt.xticks( x, categories, fontsize=10)
plt.legend()
```

```
plt.show()
```


Sales vs. Profit by Category
1e6
Amount
2.0
1.5
1.0
0.5
Sales
Profit