Business Analytics using Python & Excel

Business Analytics with the Retail Sales Dataset analyzing using Python & Excel. This project is mainly focused on the Objective of the following targeted categories using Frameworks like pandas, plotly and Excel Workbook, so they are as follows,

- Sales Performance
- Customer Behavior
- Product Analysis
- Segmentation & Targeting

Using these four Analytics the **Actionable Recommendations** are provided increase the Profit in the Retail Business.

Libraries needed:-

Pandas & Plotly

[30]:
```
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
```

Load Dataset

[31]:
```
df = pd.read_csv(r"/kaggle/input/retail-sales-dataset/retail_sa
df.head()
```

The Above data is now ready to go for an Analysis.

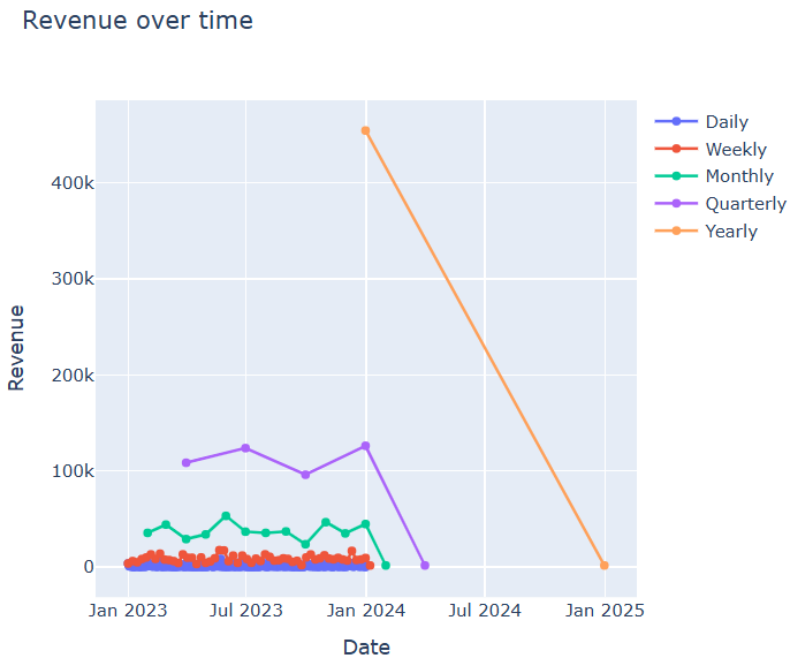
Let’s Deep Dive into the process to know more about Business Analytics and to know how Business Analytics works with Pandas and Excel. The step-by-step process is explained below.

Firstly, Let’s See about the Sales Performance,

Sales Performance

1. What is the total revenue generated over a specific period (e.g., monthly, quarterly)?

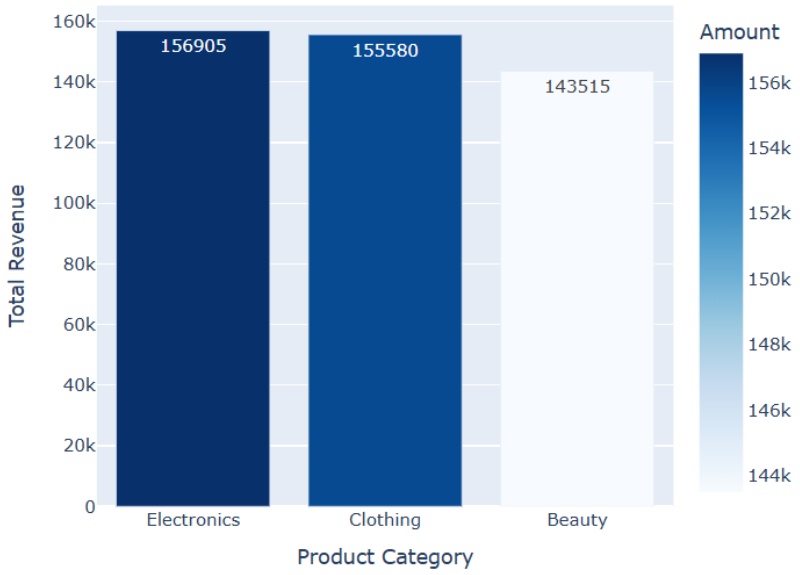
Plotly



2.*Which product category generates the highest revenue?*

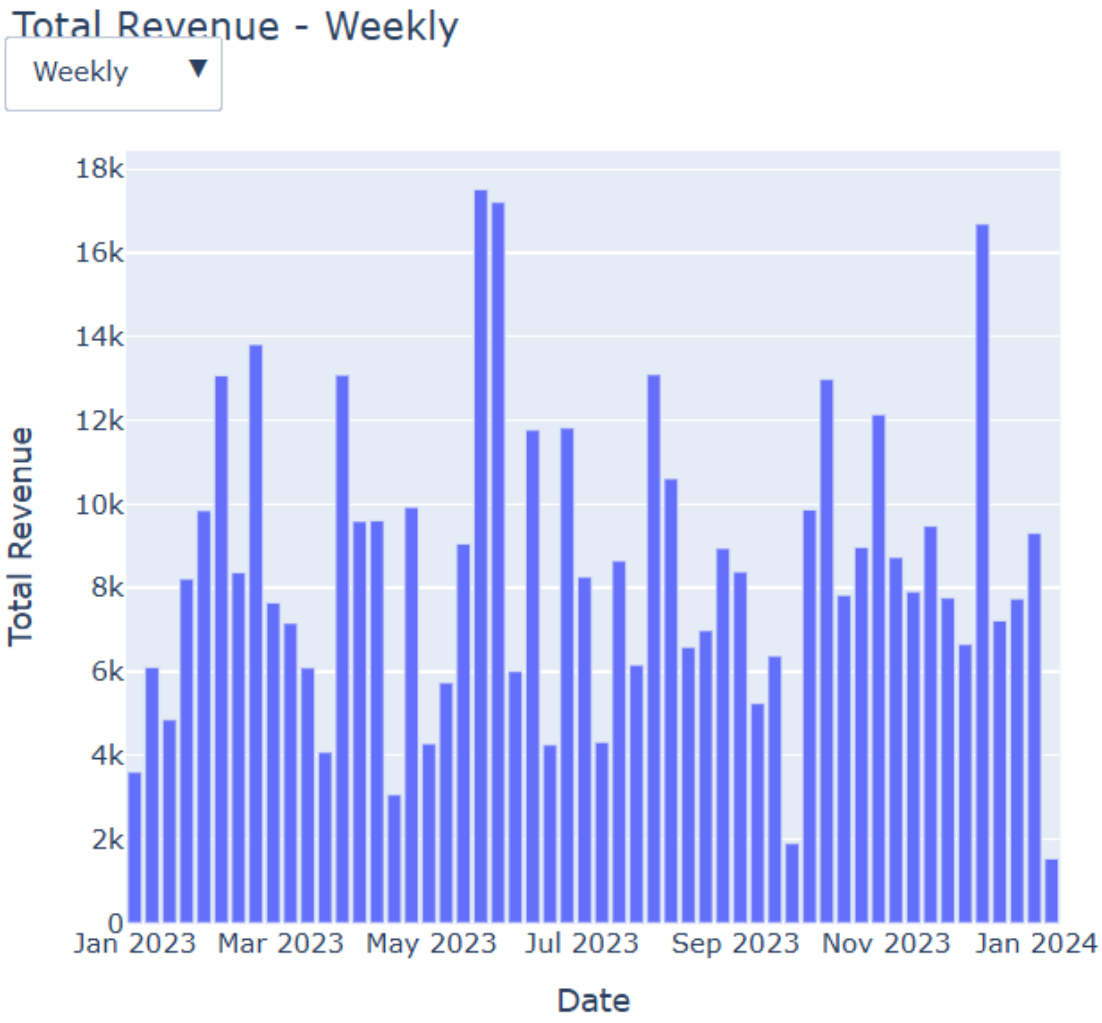
Pandas & Plotly

	Products	Amount
0	Electronics	156905
1	Clothing	155580
2	Beauty	143515



3.*What are the top-selling products by quantity sold?*

4.*How has the total amount of transactions changed over time?*

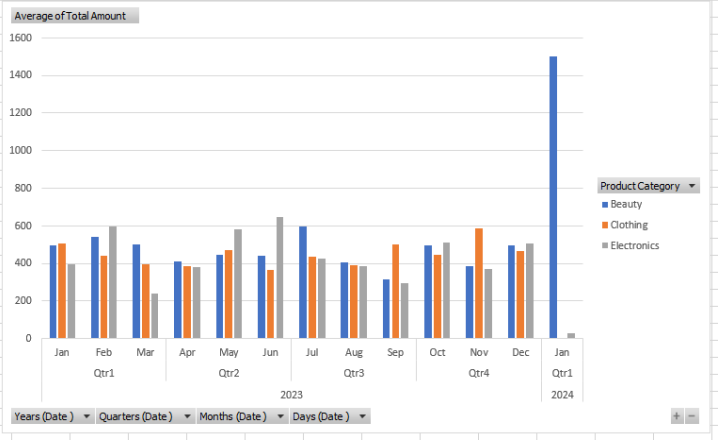


The above drop-down is used to change the Options like Weekly, Monthly, Daily, Quarterly and Yearly Revenue over Time.

Excel

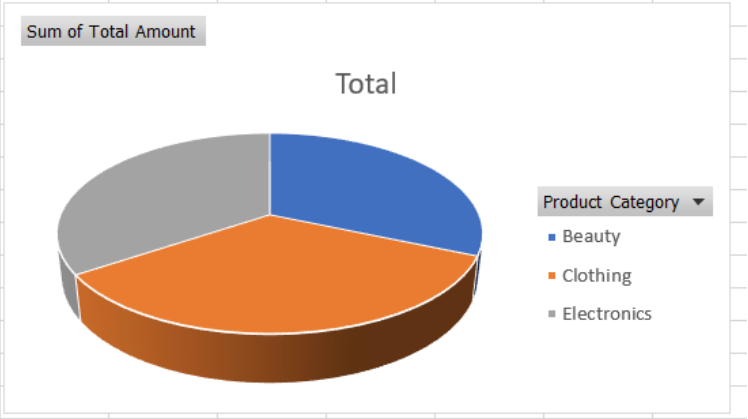
1.What is the total revenue generated over a specific period (e.g., monthly, quarterly)?

Average of Total Amount	Column Labels			
Row Labels	Beauty	Clothing	Electronics	Grand Total
2023	464.1013072	443.2478632	460.0439883	455.3807615
Qtr1	514.0277778	440.7216495	442.1538462	463.6752137
Jan	497.2	504.8076923	395.8	466.4473684
Feb	539.8076923	441.2121212	594.8076923	518.3529412
Mar	502.1428571	396.4473684	241.4285714	397.1232877
Qtr2	431.097561	411.5346535	550.8235294	461.6977612
Apr	410.5172414	387.2222222	382.1428571	393.8372093
May	444.6428571	471.7567568	581.125	506.1904762
Jun	439.8	363.2142857	647.9166667	476.8181818
Qtr3	453.5211268	432.1126761	372.6404494	415.7792208
Jul	595.9259259	434.2105263	427.8846154	492.5694444
Aug	407.9166667	389.21875	387.2368421	393.1914894
Sep	316	498.75	293	363.3846154
Qtr4	462.4074074	494.9390244	472.0588235	476.1886792
Oct	495.3225806	443.8333333	511.7142857	485.2083333
Nov	388	584.6153846	371.1111111	447.6923077
Dec	496	464.2307692	505.5	491.0989011
2024	1500		30	765
Qtr1	1500		30	765
Jan	1500		30	765
Grand Total	467.47557	443.2478632	458.7865497	456



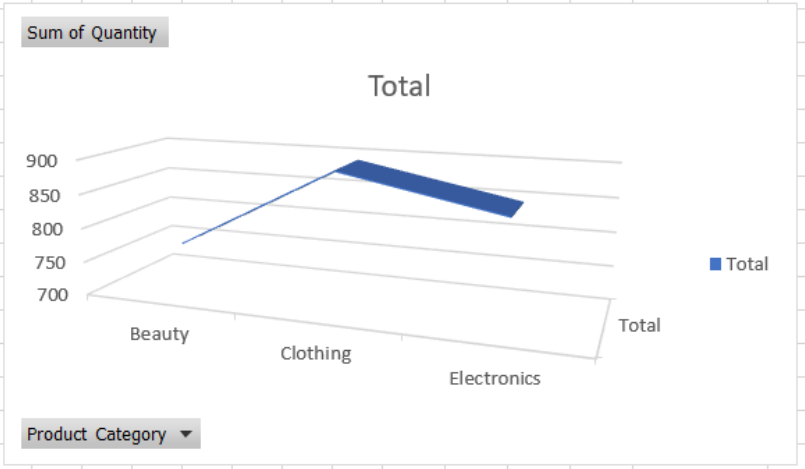
2.Which product category generates the highest revenue?

Row Labels	Sum of Total Amount
Beauty	143515
Clothing	155580
Electronics	156905
Grand Total	456000



3.What are the top-selling products by quantity sold?

Row Labels	Sum of Quantity
Beauty	771
Clothing	894
Electronics	849
Grand Total	**2514**



4. How has the total amount of transactions changed over time?

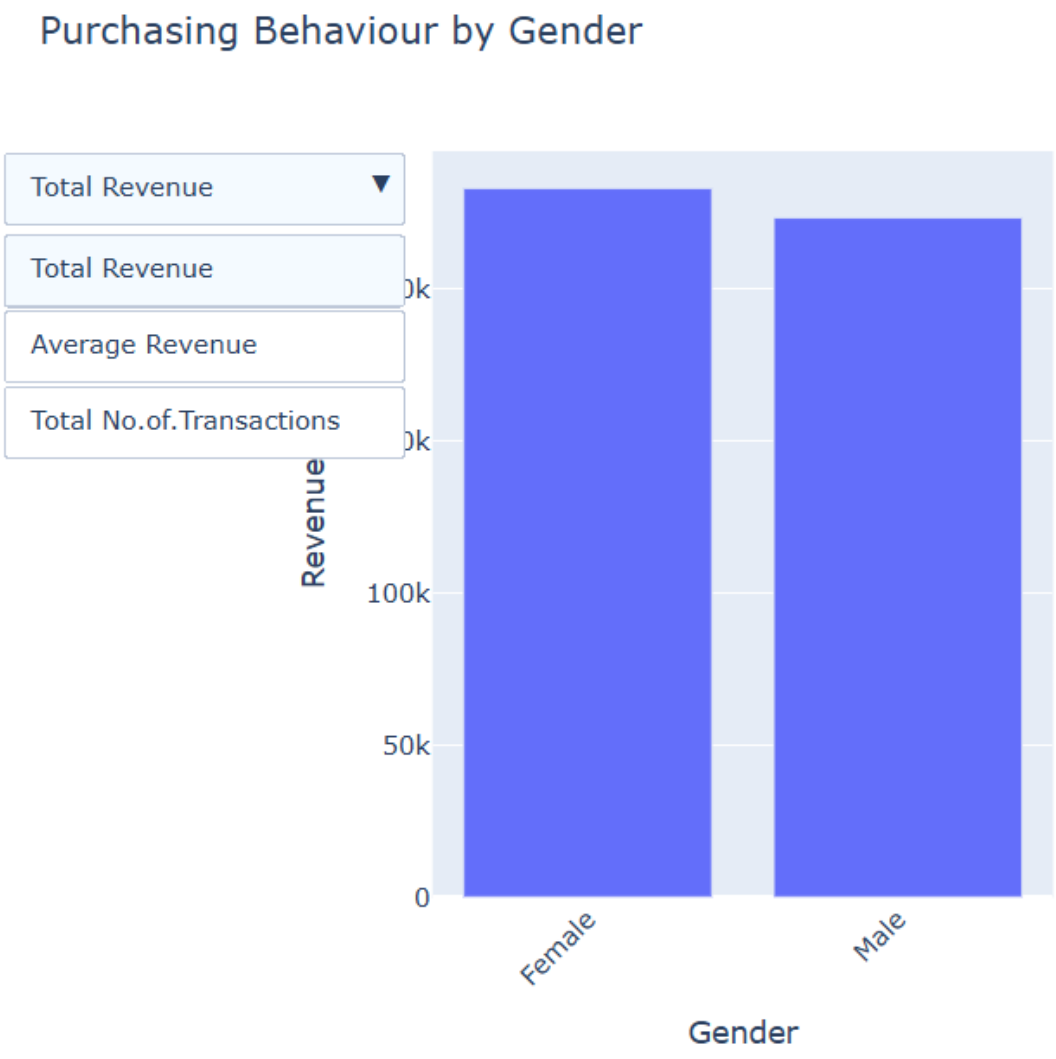
Sum of Total Amo Colum				
Row Labels	Beauty	Clothing	Electronics	Grand Total
2023				
Qtr1	37010	42750	28740	108500
Qtr2	35350	41565	46820	123735
Qtr3	32200	30680	33165	96045
Qtr4	37455	40585	48150	126190
2024	1500		30	1530
Grand Total	143515	155580	156905	456000

Secondly, Let's See about the Customer Behavior that is how the customers behaves at the time of Purchase and to notice some Trends like Gender & Age Group.

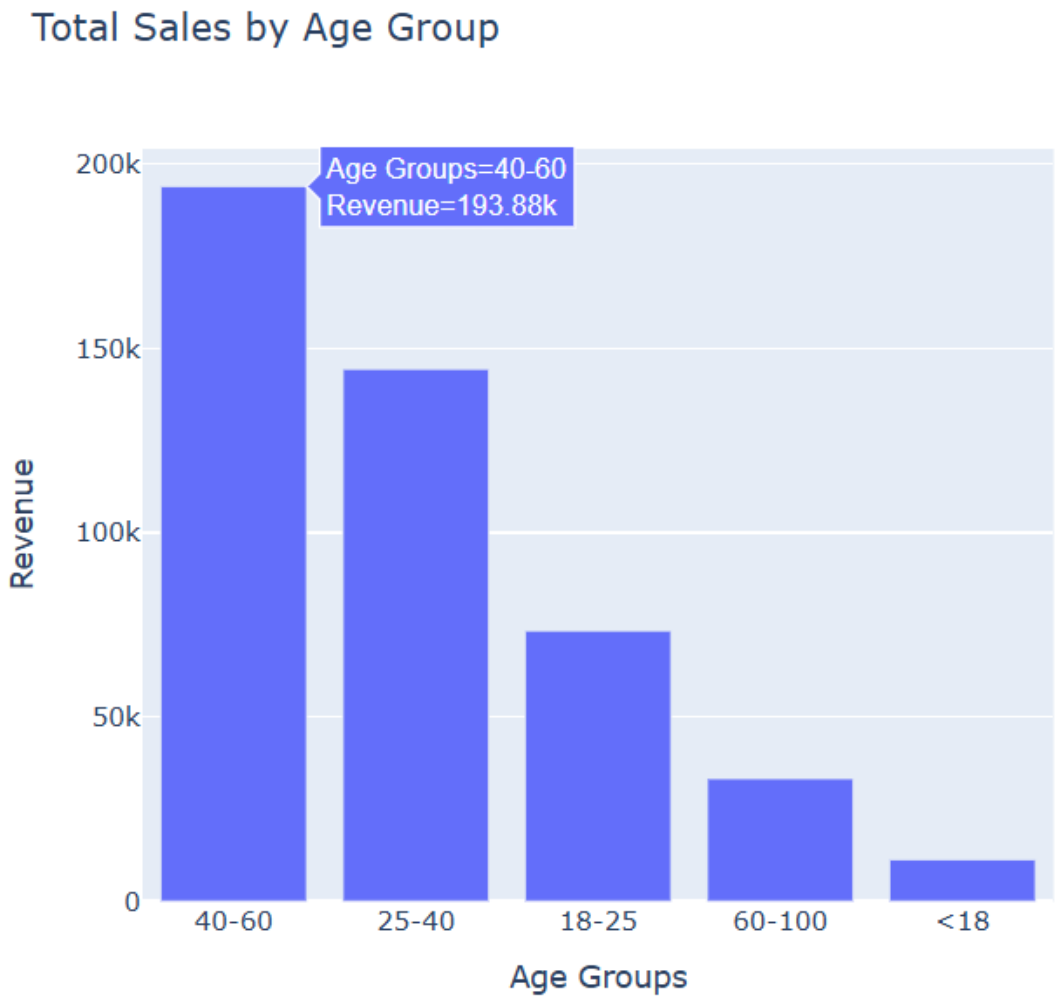
Customer Behavior

Using Plotly

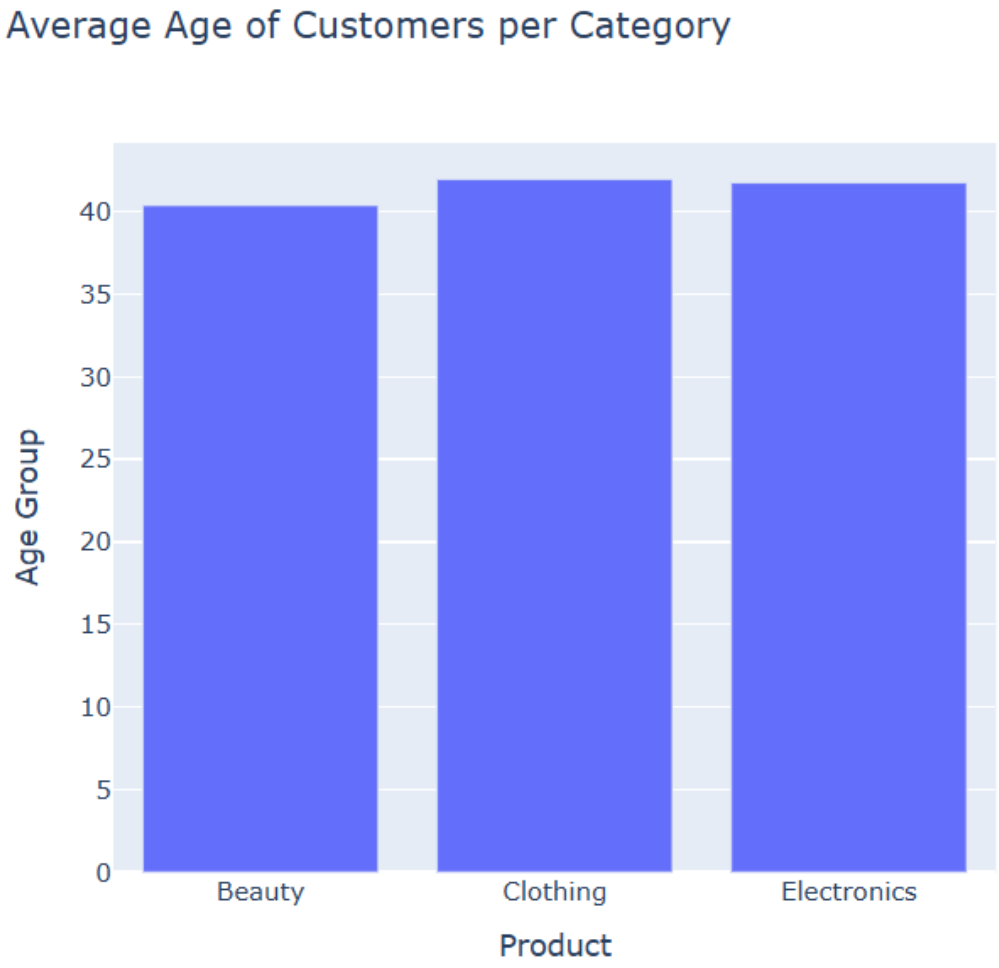
1. *How does purchasing behavior differ between genders?*



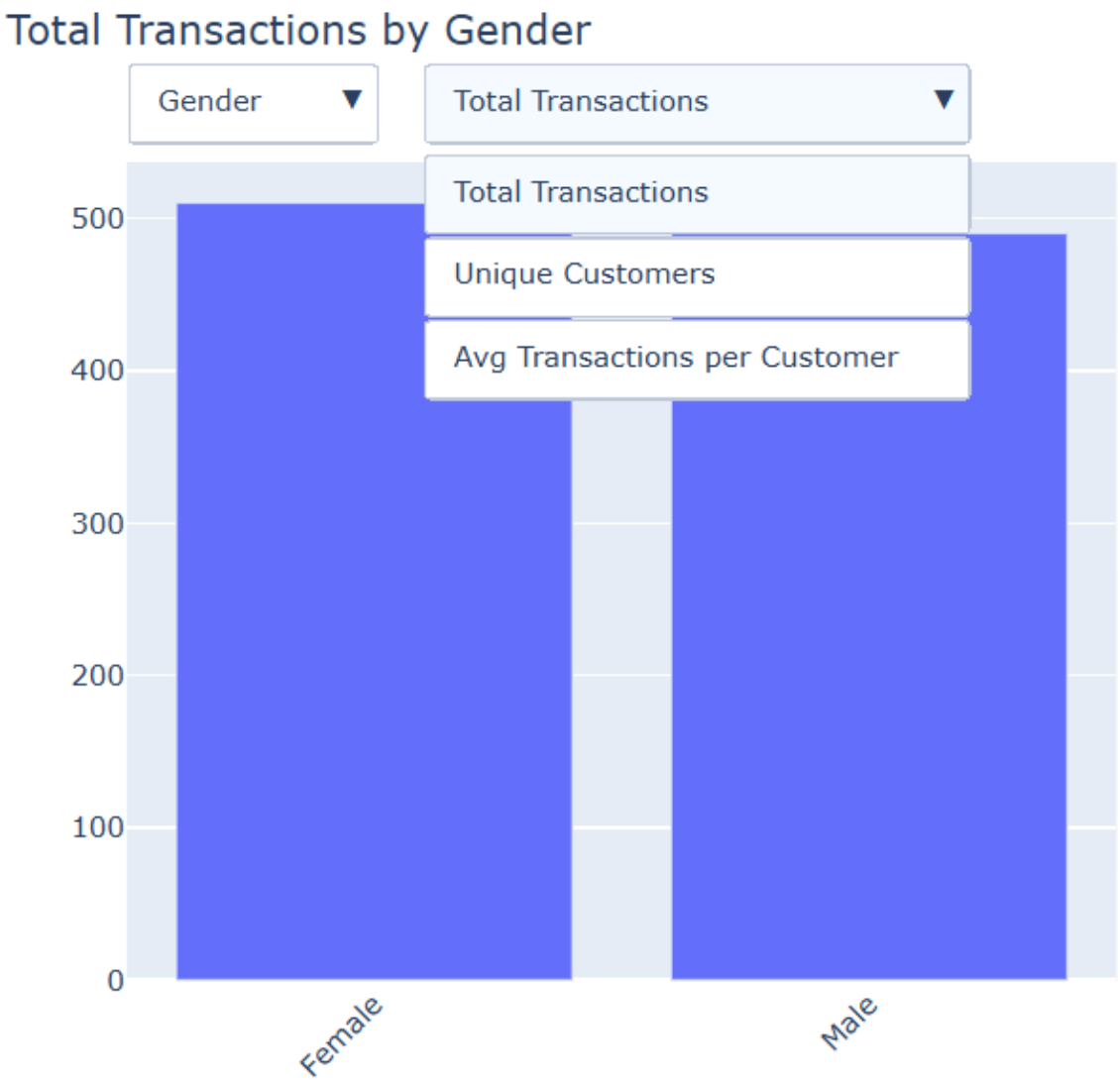
2.*Which age group contributes the most to total sales?*



3.*What is the average age of customers, and how does it vary by product category?*



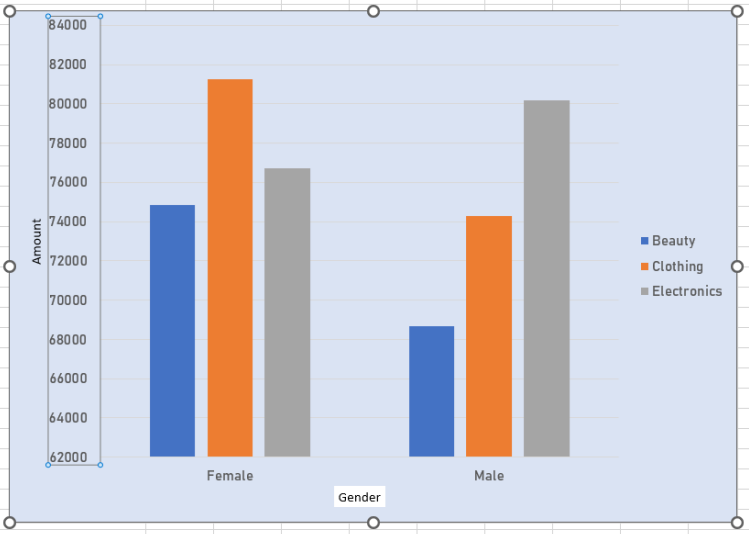
4.*Are there any trends in purchase frequency by gender or age?*



Using Excel,

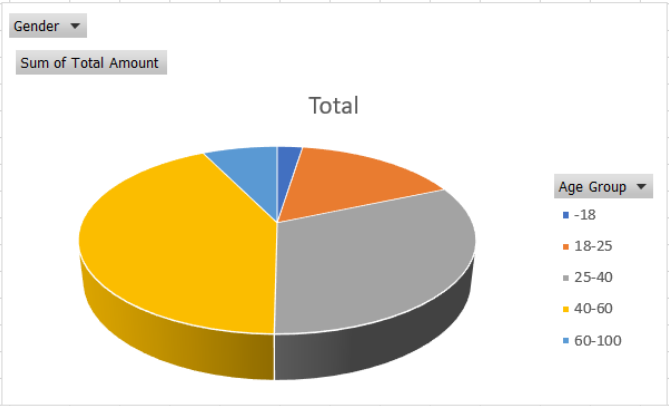
How does purchasing behavior differ between genders?

Sum of Tot	Column L			
Row Lab	Beauty	Clothing	Electronics	Grand Total
Female	74830	81275	76735	232840
Male	68685	74305	80170	223160
Grand Tota	**143515**	**155580**	**156905**	**456000**



2.Which age group contributes the most to total sales?

Gender	(All)
Row Labels	**Sum of Total Amount**
-18	11215
18-25	73335
25-40	144345
40-60	193880
60-100	33225
Grand Total	**456000**

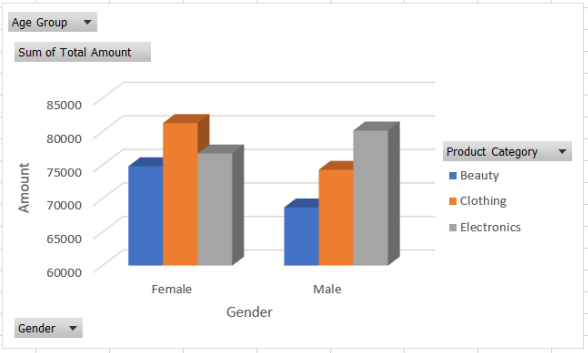


3.What is the average age of customers, and how does it vary by product category?

Age Group	(All)

Sum of Total Amount	Column Labels			
Row Labels	Beauty	Clothing	Electronics	Grand Total
Female	74830	81275	76735	232840
Male	68685	74305	80170	223160
Grand Total	**143515**	**155580**	**156905**	**456000**

AVERAGE AGE OF CUSTOMERS
41.392

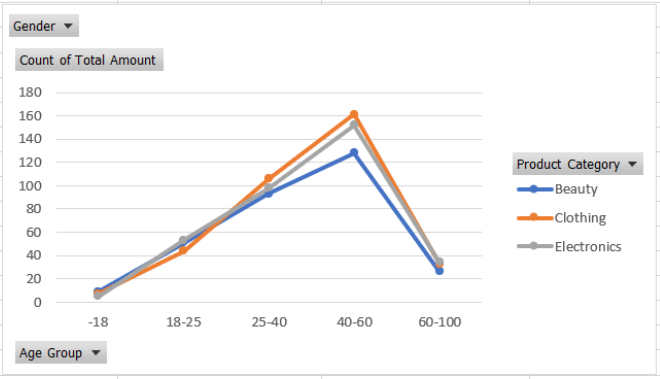


4.Are there any trends in purchase frequency by gender or age?

Total Transaction

Gender	(All)

Count of Total Amount	Column Labels			
Row Labels	Beauty	Cloth	Elec	Grand Total
-18	9	7	5	21
18-25	51	44	53	148
25-40	93	106	98	297
40-60	128	161	152	441
60-100	26	33	34	93
Grand Total	**307**	**351**	**342**	**1000**



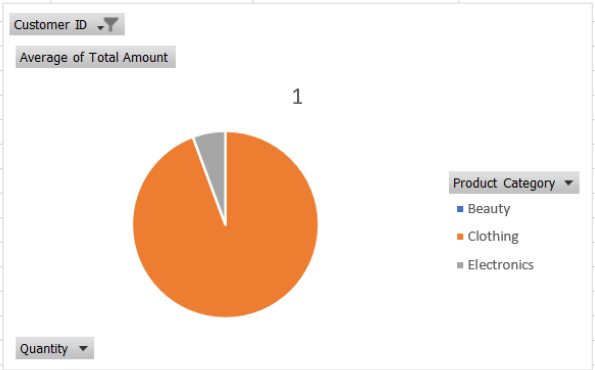
Unique Customers

Unique Customers = 1001	

Average Transaction per Customer

Customer ID	(Multiple Items)

Average of Total Amount	Column Labels			
Row Labels	1	2	3	Grand Total
Beauty		100	150	125
Clothing	500	1000		750
Electronics	30			30
Grand Total	**265**	**550**	**150**	**356**

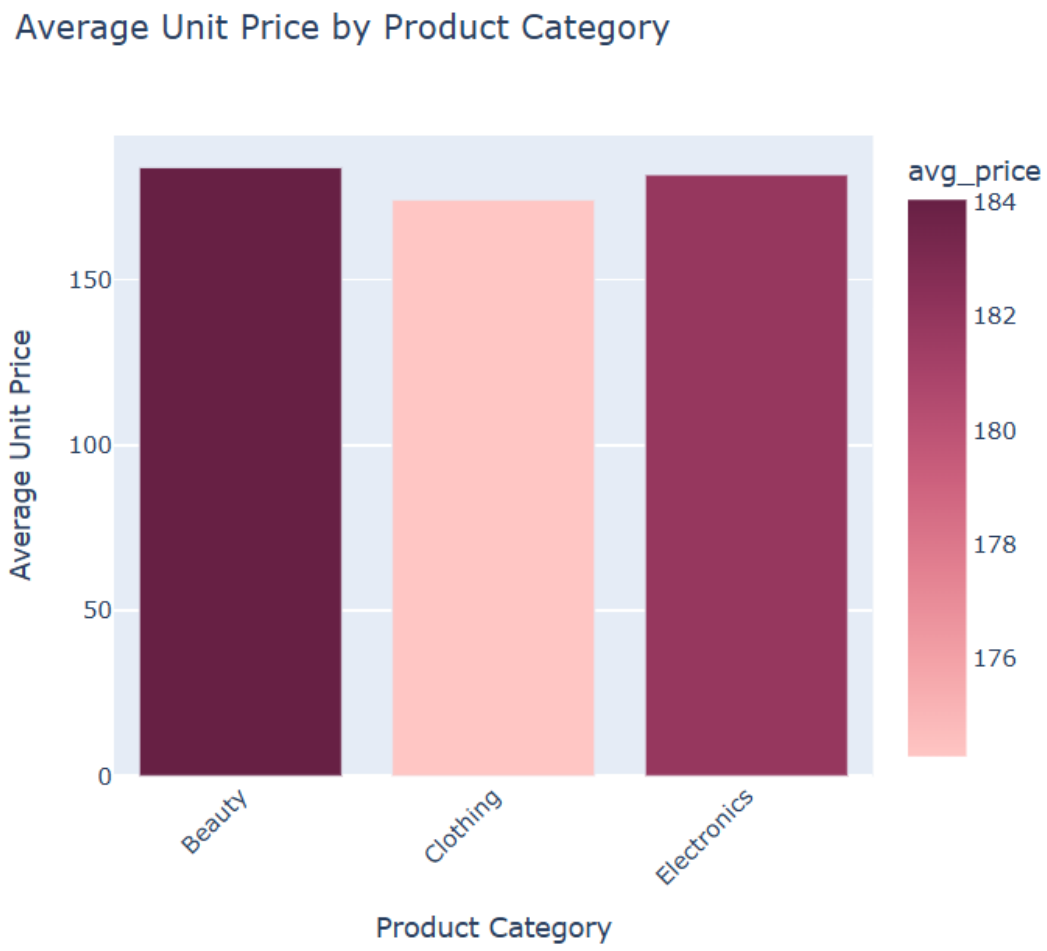


Thirdly, Let's See about the Product Analysis, To analyse the Product Category like Clothing, Electronics and Beauty in detail.

Product Analysis

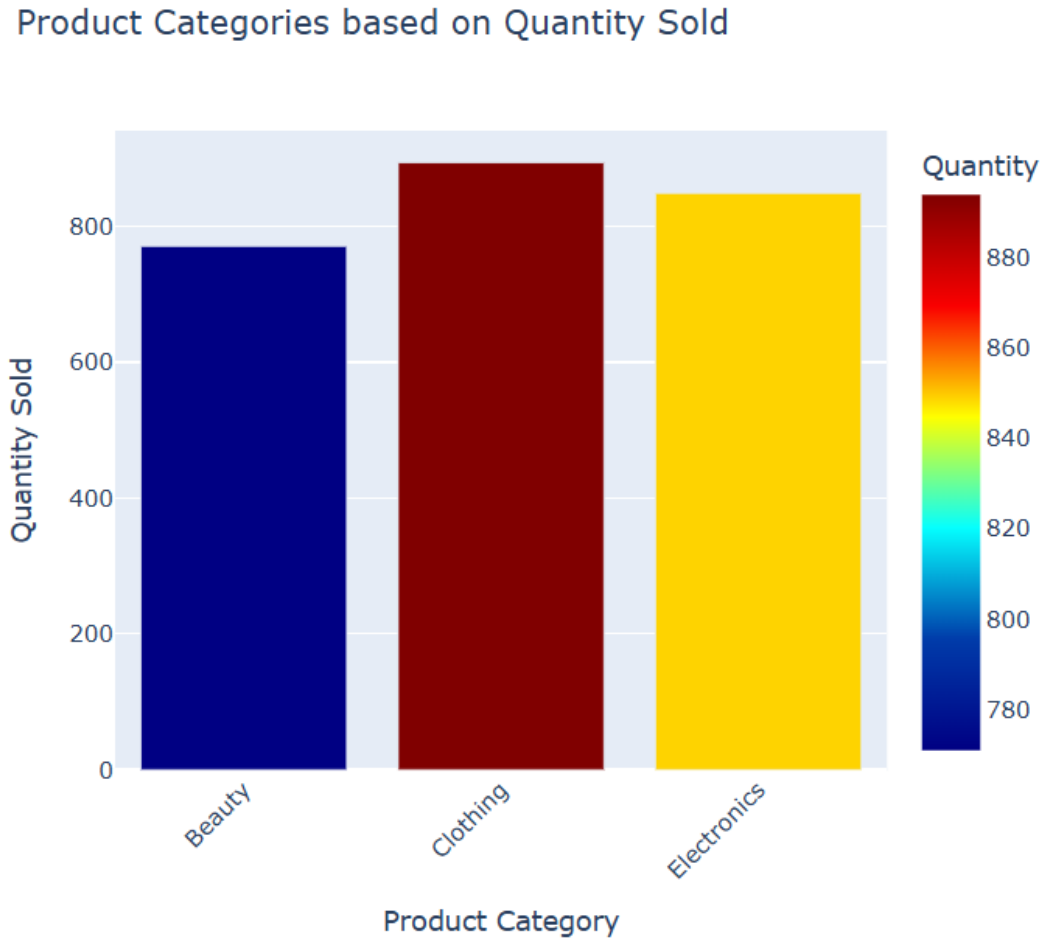
Plotly,

1.*Which product category has the highest average price per unit?*



	Product Category	avg_price
0	Beauty	184.055375
1	Clothing	174.287749
2	Electronics	181.900585

2.*How does the quantity sold vary across different product categories*



3.*Are there any product categories with unusually high or low total amounts?*

No unusuals detected as the Z-Score is between -1 to +1

Code

```
#To find outliers to calculate Z-score = (Total + Mean) / SD
category_tot = df.groupby('Product Category')['Revenue'].sum().reset_index()
category_tot

#Mean
mean_tot = category_tot['Revenue'].mean()

#Standard Deviation
std_tot = category_tot['Revenue'].std()

#Z-Score
category_tot['z_score'] = (category_tot['Revenue'] - mean_tot) / std_tot

#Check Unusuals
category_tot['Unusuals'] = category_tot['z_score'].apply(
    lambda x: 'High' if x > 2 else ('Low' if x<-2 else 'Normal')
)

category_tot.drop(columns=['z_score'], errors='ignore')
```

Pandas

	Product Category	Revenue	Unusuals
0	Beauty	143515	Normal
1	Clothing	155580	Normal
2	Electronics	156905	Normal

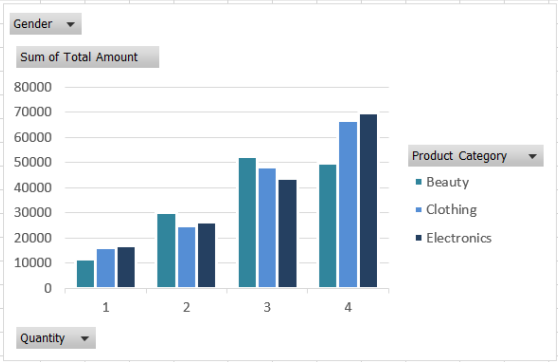
Excel

1.Which product category has the highest average price per unit?

Row Labels	Average of Price per Unit
Beauty	184.0553746
Clothing	174.2877493
Electronics	181.9005848
Grand Total	**179.89**

2.How does the quantity sold vary across different product categories?

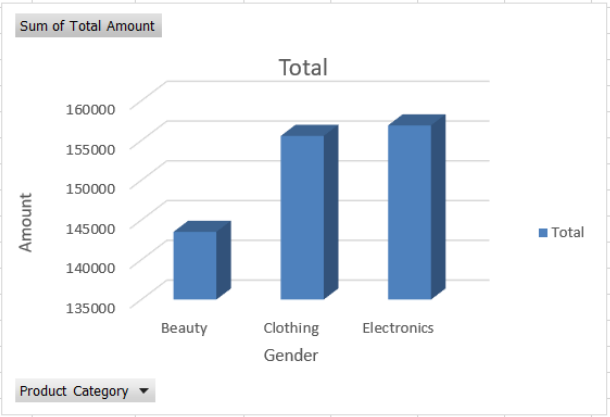
Gender	(All)			
Sum of Total Amount	**Column Labels**			
Row Labels	**Beauty**	**Clothing**	**Electronics**	**Grand Total**
1	11730	16115	16960	44805
2	29930	24670	26450	81050
3	52155	48315	43815	144285
4	49700	66480	69680	185860
Grand Total	**143515**	**155580**	**156905**	**456000**



3.Are there any product categories with unusually high or low total amounts?,

Row Labels	Sum of Total Amount
Beauty	143515
Clothing	155580
Electronics	156905
Grand Total	**456000**

Z - SCORE
Normal
No Outliers Detected



Refer Table Sheet for z_score calculation

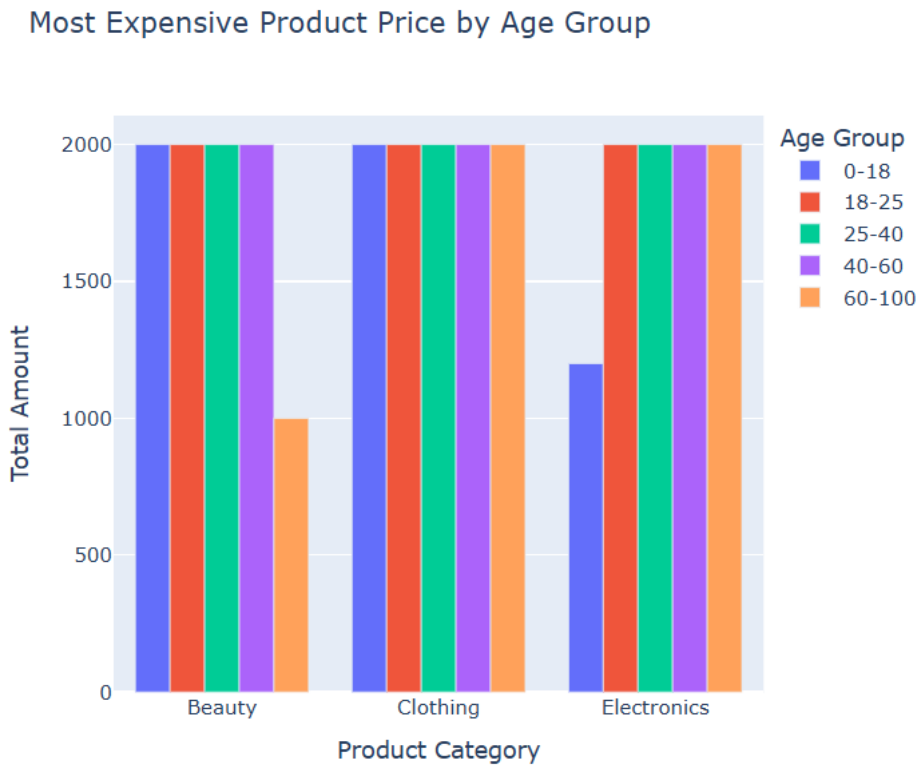
[=(I2 - AVERAGE($I$2:$I$1001)) / STDEV.S($I$2:$I$1001)]

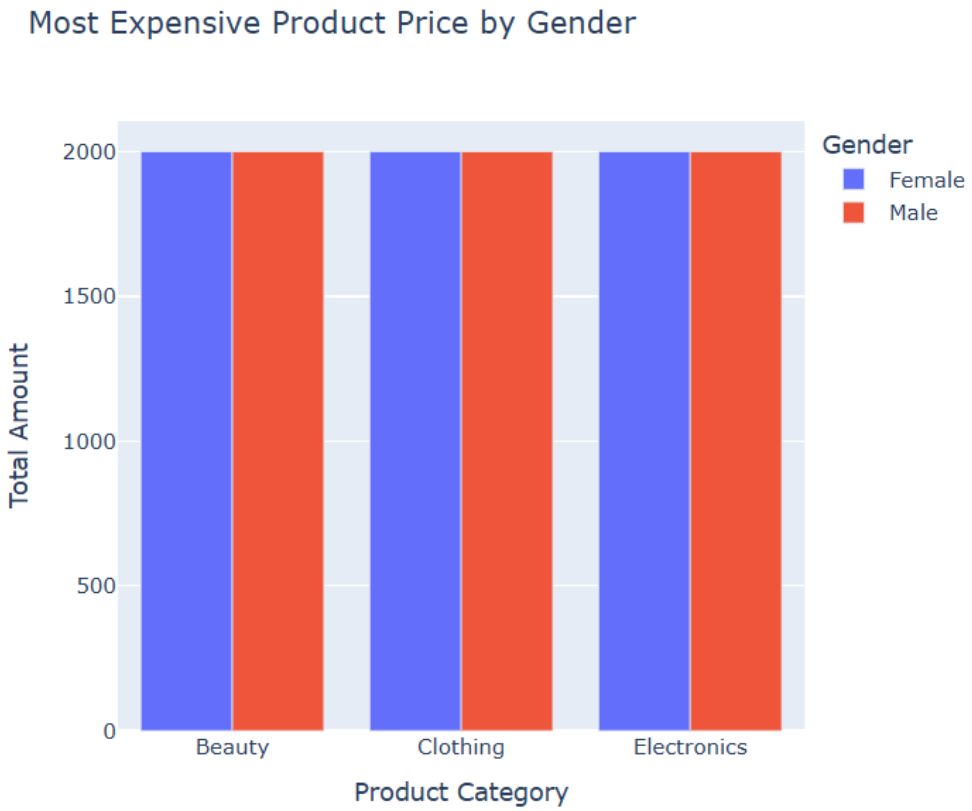
Fourly, Let's See about the Segmentation & Targeting, This involves Segmenting the Column Names according to our analysis and targeting on specific Trends for growth of the Business.

Segmentation & Targeting

Plotly,

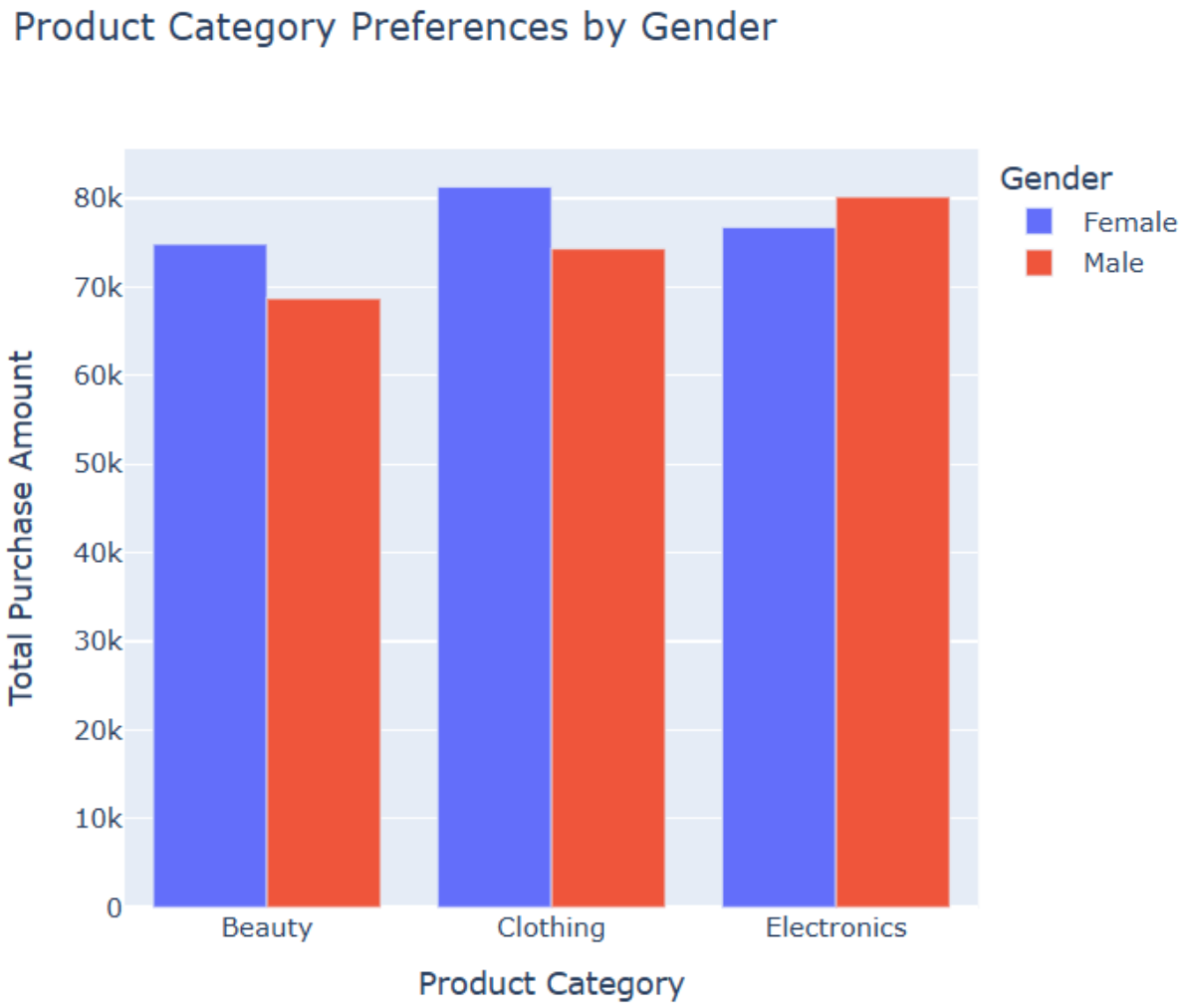
1.*Which combination of gender and age group buys the most expensive products?*



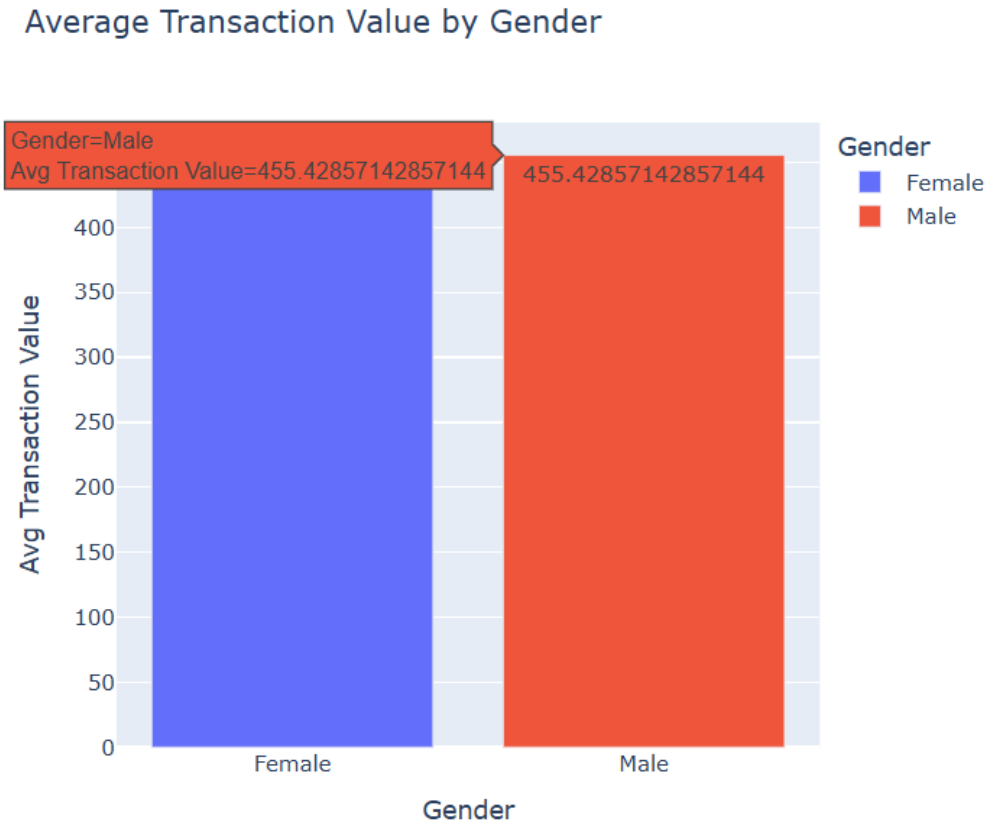


The above tells us Both the Male and Female brought Expensive Gift

2.*Are there specific product categories preferred by certain demographic segments?*



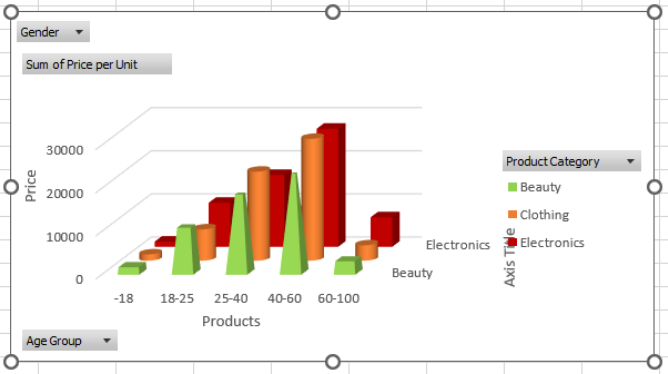
3.*What is the average transaction value per customer segment (e.g., by gender or age range)?*



Excel,

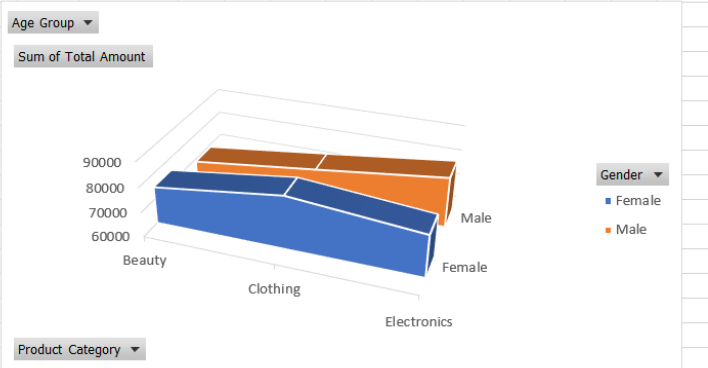
1.Which combination of gender and age group buys the most expensive products?

Gender	(All)			
Sum of Price pe	**Column Label**			
Row Labels	**Beauty**	**Clothing**	**Electronics**	**Grand Total**
-18	1690	1455	1180	4325
18-25	10660	7265	10185	28110
25-40	18185	20640	16655	55480
40-60	22930	28255	27350	78535
60-100	3040	3560	6840	13440
Grand Total	**56505**	**61175**	**62210**	**179890**



2.Are there specific product categories preferred by certain demographic segments?

Age Group	(All)		
Sum of Total Amount	**Column Labels**		
Row Labels	**Female**	**Male**	**Grand Total**
Beauty	74830	68685	143515
Clothing	81275	74305	155580
Electronics	76735	80170	156905
Grand Total	**232840**	**223160**	**456000**



3.What is the average transaction value per customer segment (e.g., by gender or age range)?

Gender	Female					
Average of Total Amount	**Column Labels**					
Row Labels	**-18**	**18-25**	**25-40**	**40-60**	**60-100**	**Grand Total**
Beauty	639	379.25926	579	415.94595	166.5	450.78313
Clothing	858.3333333	544.11765	542.2449	410.63953	398.15789	467.0977
Electronics	434	501.66667	453.75	457.15278	323.07692	451.38235
Grand Total	**610.7692308**	**463.67647**	**522.12903**	**426.76724**	**319.7619**	**456.54902**

AVERAGE TRANSACTION
Female = 455.43
Male =456.55

Therefore these are the four steps of analyses that are done by using ***Retail Sales Dataset*** and the helpful insights that we get from the analyses helps us Scale-up, Up-skill and Improve the Business or a particular organization. The Insights that we get from analyzing the dataset is very important and helps us to solve many Business Problems. The key note comes here by asking questions about Businesses, Why Customer Sales is very low? Why the particular product is dropping? And What Recommendations can be provided to increase the sales that profits the company? Let us see the final recommendations by the analyzed results that answers for many question in the business and solve many problems in Business Analytics, Python Libraries like Pandas and Plotly helps us to analyze and visualize data more efficiently and Excel plays a main role for a Quick and Easy analytics for Business.

Let us Deep Dive by Recommending some solutions by providing answer for the questions that are being asked in each analysis.

Answers that are provided by each analysis:

Sales Performance

1.What is the total revenue generated over a specific period (e.g., monthly, quarterly)?
Total Revenue generated over a period of time:
Daily Revenue - 1000-2000
Weekly Revenue - 7000-9000
Monthly Revenue -30.00k-50.00k
Yearly Revenue -100.5k-130.20k

2.Which product category generates the highest revenue?
The product category that generates highest revenue is ***Electronics*** & secondly Clothing

3.What is the top-selling products by quantity sold?
Clothing is the top-selling product in Quantities

4.How has the total amount of transactions changed over time?
Monthly Highest Sales is ***May***, secondly October, thirdly December and February and the Last is September.

Customer Behavior

1.How does purchasing behavior differ between genders?
Female Customers are increased compared to Male in all (Average Revenue, Total Revenue, Total No. of Transactions)

2.Which age group contributes the most to total sales?
The Age Group that contributes the most to total sales is ***40-60***

3.What is the average age of customers, and how does it vary by product category?
The average age of customers is ***40***, Their top purchases is ***Clothing.***

4.Are there any trends in purchase frequency by gender or age?
Gender-vise ***Female*** customers are high and Age-vise ***40-60*** age group is high (Unique Customers, Total Transaction, Average Transactions per Customer)

Product Analysis

1.Which product category has the highest average price per unit?
Beauty Product has the highest average per unit.

2.How does the quantity sold vary across different product categories?
Clothing is the Highly sold product in Quantities & ***Beauty*** is the Least Sold product in quantities.

3.Are there any product categories with unusually high or low total amounts?
No Unusual high or low amount, all tends to be normal (z-score meets -1 to +1).

Segmentation and Targeting

1.Which combination of gender and age group buys the most expensive products?

a. ***All Age Groups buy expensive products***, But ***60-100 shows low interest in Beauty*** *and* ***0-18 in Electronics.***
b. ***All Gender shows equal interest*** on buying the expensive products.

2.Are there specific product categories preferred by certain demographic segments?

Female show more interest in ***Clothing*** and Less interest in Beauty
Male show more interest in ***Electronics*** and less interest in Beauty
(total_purchase_amount)

3.What is the average transaction value per customer segment (e.g., by gender or age range)?

Average Transaction value per Customer Segment:
**** Female = 456.55***
**** Male = 455.43***

Therefore, these are the Analytics done by the Retail Sales Dataset. This Answers provide the most valuable and useful insights to change the Marketing campaign and Other things, The KEY of this Quick Summary of this analysis is,

a. ***Firstly, start with understanding the Retail Sales Dataset, Understanding the Column Names, whether it is a category, numerical or string***
b. ***Asking Questions for the Dataset like What? Why? When? and how? questions***
c. ***Secondly Choosing the right platform to analyze data like Excel, Pandas and Plotly(to visualize).***
d. ***Find Answers for these questions using the platform that you have chose to analyze data with...***

e. ***Research on that answer to find the best recommendations ideas***
f. ***Convert the Ideas to <u>Visualization.</u>***

Now let's deep dive into provide Actionable Recommendation to improve the Sales, Customer, Marketing Campaign etc..

Actionable Recommendations

Which product categories should be promoted based on sales performance?
Beauty products

Are there opportunities to adjust pricing for underperforming categories?
Yes, Beauty Products & Male

Should marketing campaigns target specific age groups or genders based on purchase trends?
Yes, 0-18 age = Electronics and 40-60 age = skincare(Beauty), Marketing campaigns should target male customers.