STATISTIC & SPREADSHEET

LUXURA E-COMMERCE

BY MAZMUR CHRISADRIANTO TELAUMBANUA

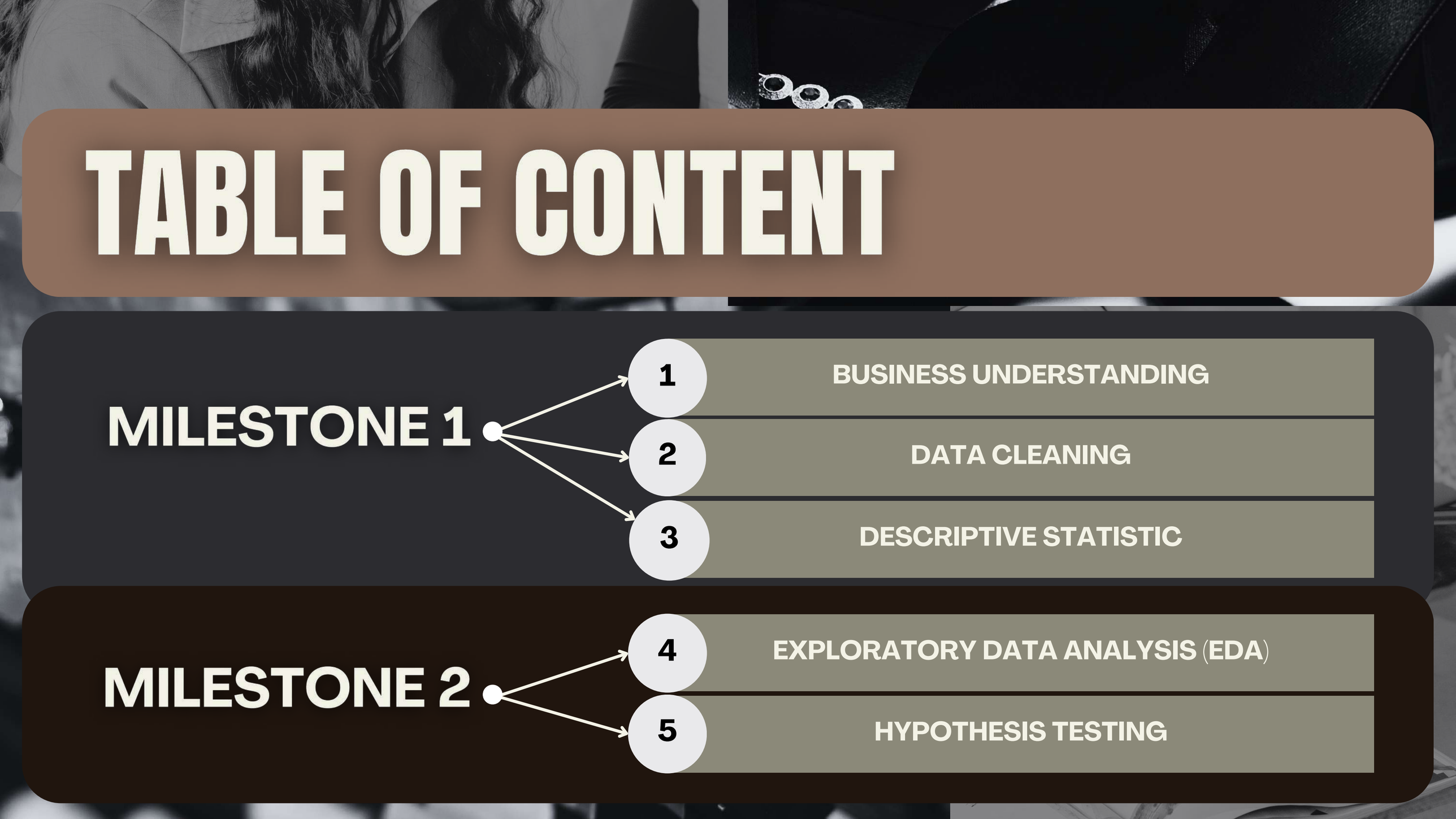


TABLE OF CONTENT

MILESTONE 1

MILESTONE 2

BUSINESS BACKGROUND

Luxura is an e-commerce company that focuses on selling **luxury fashion products**. The store exclusively sells several global-famous brands such as Adibi, Balena, and Celina.

Luxura has the mission to **prioritize ther marketing on one of the brands above.** Our Head data asks us to analyze on this matter based on the data that has been provided, **Based on partnership fee data Adibi paid the highest fee from Luxura.**

MILESTONE 1

BUSINESS UNDERSTANDING

Some of action we could see considering the data we have :

- To understand **purchased power and the buying enthusiasm** of each customer we can analyzing customers **income**
- To see **the revenue of each brand** we can count **the amount of users spent of each brand**
- To know **the total number of product purchased** by each brand, we can see total **quantity of order** by each brand
- To understand **customer behaviour**, we could dianalyze **gender, marital status, age, education, occupation** and **settlement type** of customer

DATA CLEANING

- In the **brand preference** column, delete the column as there is a lot of **empty data** and delete the rows for the empty rows.
- Delete the **duplicated data**.
- Change the **currency type** of income column.
- **Lookup** range and find its description in **dictionary data**.

BEFORE

	A	B	C	D	E	F
1	ID	Sex	Marital status	Age	Education	Income
2	200000001	0	0	47	1	$110,866
3	200000002	1	1	52	2	$159,052
4	200000003	0	0	28	0	$113,212
5	200000004	0	0	37	1	$119,602
6	200000005	0	0	49	1	$121,466
7	200000006	0	0	39	1	$112,347
8	200000007	0	0	27	0	$125,190
9	200000008	0	0	26	0	$131,122
10	200000009	0	0	25	0	$108,960
11	200000011	0	0	25	0	$108,960
12	200000012	0	0	44	1	$181,262
13	200000013	0	0	50	1	$122,830
14	200000014	0	0	34	1	$116,525
15	200000016	0	0	40	1	$76,881
16	200000017	1	0	47	1	$110,287
17	200000018	0	0	29	0	$102,723
18	200000019	0	0	25	0	$104,505
19	200000020	0	1	26	1	$72,429
20	200000021	1	0	30	1	$58,207
21	200000023	0	0	63	2	$135,754
22	200000024	0	0	48	1	$94,868
23	200000025	1	1	32	1	$88,428
24	200000026	0	0	24	0	$107,633

AFTER

	A	B	C	D	E	F	G
1	ID	Sex	Marital status	Age	Age Group	Education	Income
2	200000001	male	single	47	Middle	high school	$110,866.00
3	200000002	female	arried / widowed)	52	Middle	university	$159,052.00
4	200000003	male	single	28	Middle	other / unknown	$113,212.00
5	200000004	male	single	37	Middle	high school	$119,602.00
6	200000005	male	single	49	Middle	high school	$121,466.00
7	200000006	male	single	39	Middle	high school	$112,347.00
8	200000007	male	single	27	Middle	other / unknown	$125,190.00
9	200000008	male	single	26	Middle	other / unknown	$131,122.00
10	200000009	male	single	25	Middle	other / unknown	$108,960.00
11	200000011	male	single	44	Middle	high school	$181,262.00
12	200000012	male	single	50	Middle	high school	$122,830.00
13	200000013	male	single	34	Middle	high school	$116,525.00
14	200000014	male	single	40	Middle	high school	$76,881.00
15	200000017	male	single	29	Middle	other / unknown	$102,723.00
16	200000018	male	single	25	Middle	other / unknown	$104,505.00
17	200000019	male	arried / widowed)	26	Middle	high school	$72,429.00
18	200000020	female	single	30	Middle	high school	$58,207.00
19	200000021	male	single	63	Senior	university	$135,754.00
20	200000023	male	single	48	Middle	high school	$94,868.00
21	200000024	female	arried / widowed)	32	Middle	high school	$88,428.00
22	200000025	male	single	24	Junior	other / unknown	$107,633.00
23	200000026	male	single	37	Middle	high school	$170,146.00
24	200000027	female	arried / widowed)	33	Middle	high school	$112,422.00

Datasets

DESCRIPTIVE ANALITYCS

Insight

- A Positive Skewness Score on Total Number of Orders indicates the presence of mild outliers.
- A High Positive Skewness Score on Total Quantity of Orders indicates the presence of very extreme outliers.

We can use **median** when there are **extreme outliers.**

Before outliers removed

	Total Number of Order		Total Quantity of Order
Mean	2.982	Mean	57.62002
Standard Error	0.06433996629	Standard Error	24.73328799
Median	3	Median	16
Mode	1	Mode	15
Standard Deviation	1.438685383	Standard Deviation	553.0531325
Sample Variance	2.069815631	Sample Variance	305867.7674
Kurtosis	-1.329894333	Kurtosis	219.0621571
Skewness	0.02755174005	Skewness	14.43264132
Range	4	Range	9534.6
Minimum	1	Minimum	3
Maximum	5	Maximum	9537.6
Sum	1491	Sum	28810.01
Count	500	Count	500
Largest(1)	5	Largest(1)	9537.6
Smallest(1)	1	Smallest(1)	3
Confidence Level(95%)	0.1264106174	Confiden	
Q1	2	Q1	
Q3	4	Q3	
IQR	2	IQR	

Datasets

DESCRIPTIVE ANALITYCS

Based on the **Lower Inner Fence** (**LIF**) and **Upper Inner Fence** (**UIF**), we get outlier results on the **Total Number of Orders** and **Total Quantity of Orders data.**

In **Total Number of Order**, there are no outliers according to the LIF or UIF.

In **Total Quantity of Order** there are several outliers based on UIF and LIF so they can be removed for the data used.

Total Number of Order	
Q1	2
Q3	4
IQR	2
LIF	-1
UIF	7

Total Quantity of Order	
Q1	13
Q3	20
IQR	7
LIF	2.5
UIF	30.5

ID	Total Number of Order	Total Quantity of Order	Amount Spent on Adibi	Amount Spent on Balena	Amount Spent on Celinna
200000016	5	215	$3,148.63	$3,189.47	$211.96
200000160	5	207	$3,145.17	$3,184.17	$3,208.26
200000194	5	325	$3,135.23	$172.80	$1,194.32

Datasets

DESCRIPTIVE ANALITYCS

Based on the **Lower Inner Fence** (**LIF**) and **Upper Inner Fence** (**UIF**), we get outlier results on the **Total Number of Orders** and **Total Quantity of Orders data.**

Total Number of Order	
Q1	2
Q3	4
IQR	2
LIF	-1
UIF	7

Total Quantity of Order	
Q1	13
Q3	20
IQR	7
LIF	2.5
UIF	30.5

In **Total Number of Order**, there are no outliers according to the LIF or UIF.

In **Total Quantity of Order** there are several outliers based on UIF and LIF so they can be removed for the data used.

ID	Total Number of Order	Total Quantity of Order	Amount Spent on Adibi	Amount Spent on Balena	Amount Spent on Celinna
200000016	5	215	$3,148.63	$3,189.47	$211.96
200000160	5	207	$3,145.17	$3,184.17	$3,208.26
200000194	5	325	$3,135.23	$172.80	$1,194.32

Datasets

DESCRIPTIVE ANALITYCS

Based on the Interquartile Range (**IQR**) **method**, and search LIF ,UIF to find outliers **then we can delete the row**, because at this stage of data cleaning there is no outlier data.

From 500 rows to 497 rows after data outlier removed.

	A	B	C	D	
1	ID	Sex	Marital status	Age	A
487	200000497	male	single	60	
488	200000499	female	arried / widowed)	26	
489	200000500	male	single	60	
490	200000501	female	arried / widowed)	29	
491	200000502	female	single	35	
492	200000503	male	single	30	
493	200000504	male	single	38	
494	200000505	male	single	60	
495	200000506	male	single	25	
496	200000507	male	single	35	
497	200000508	female	arried / widowed)	30	
498	200000509	female	arried / widowed)	35	
499					

Datasets

DESCRIPTIVE ANALITYCS

- The Total Quantity of Order column has a wide range.The Total Quantity of Order ranges from 3 to 28.

- We can identify 2 as the central tendency in representing the character of Total Quantity of Order column. Because the Mean, Median, and Mode are all in close proximity (16),

- Total Quantity of Order column skewness is symmetrical (Mean = Median).

BEFORE

Total Quantity of Order	
Mean	57.62002
Standard Error	24.73328799
Median	16
Mode	15
Standard Deviation	553.0531325
Sample Variance	305867.7674
Kurtosis	219.0621571
Skewness	14.43264132
Range	9534.6
Minimum	3
Maximum	9537.6
Sum	28810.01
Count	500
Largest(1)	9537.6
Smallest(1)	3
Confidence Level(95%)	48.59421578
Q1	13
Q3	20
IQR	7
LIF	2.5
UIF	30.5

AFTER

Total Quantity of Order	
Mean	57.62002
Standard Error	24.73328799
Median	16
Mode	15
Standard Deviation	553.0531325
Sample Variance	305867.7674
Kurtosis	219.0621571
Skewness	14.43264132
Range	9534.6
Minimum	3
Maximum	9537.6
Sum	28810.01
Count	500
Largest(1)	9537.6
Smallest(1)	3
Confidence Level(95%)	48.59421578
Q1	13
Q3	20
IQR	7
LIF	2.5
UIF	30.5

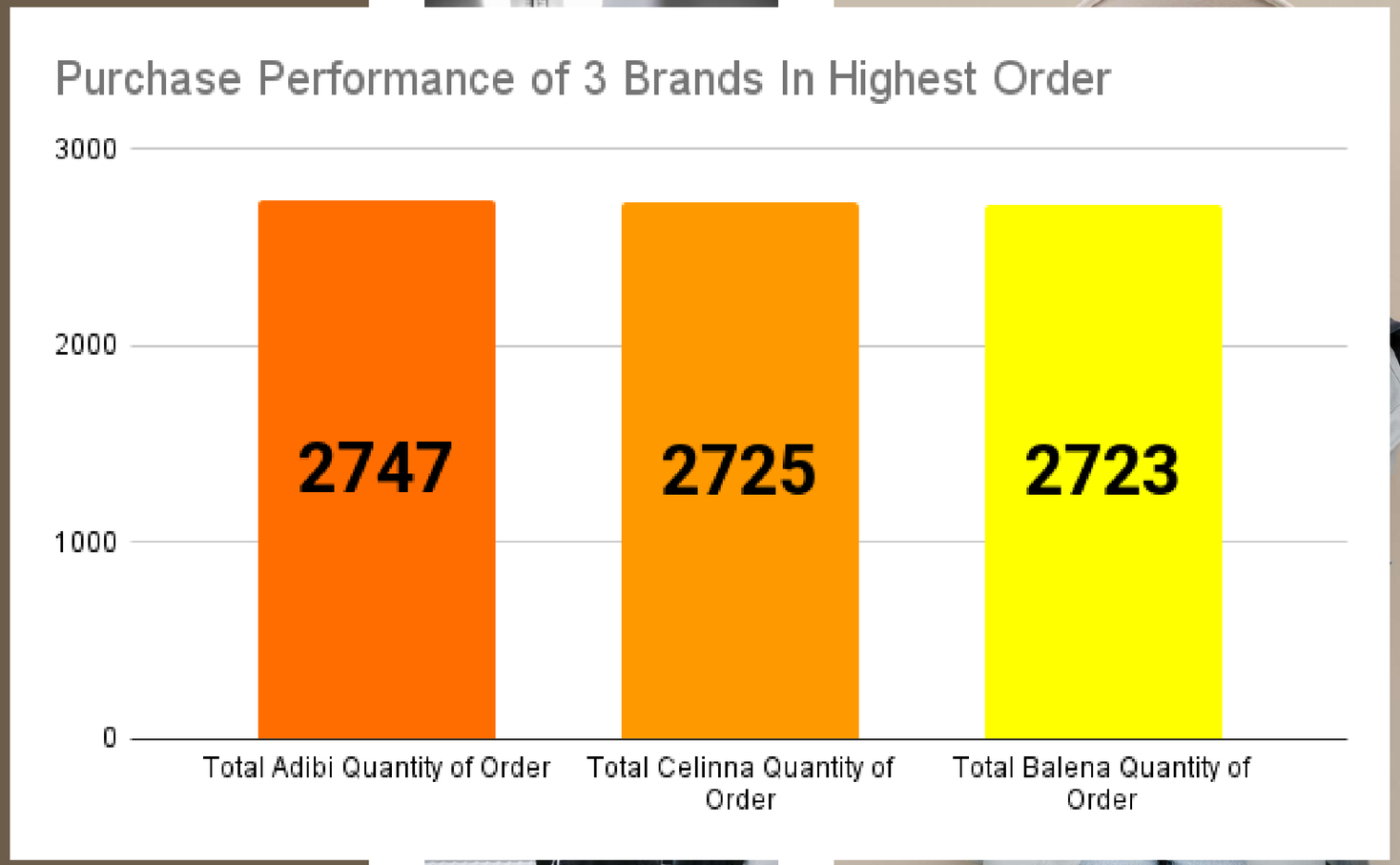
Datasets

MILESTONE 2

EXPLORATORY DATA ANALYTIC

Purchase Perfromance

Based on total product orders, it seems that the distance between **Adibi, Balena, and Celina** is **not that significant** and it can be seen that the total number of orders for the Adibi brand is higher than for other brands.

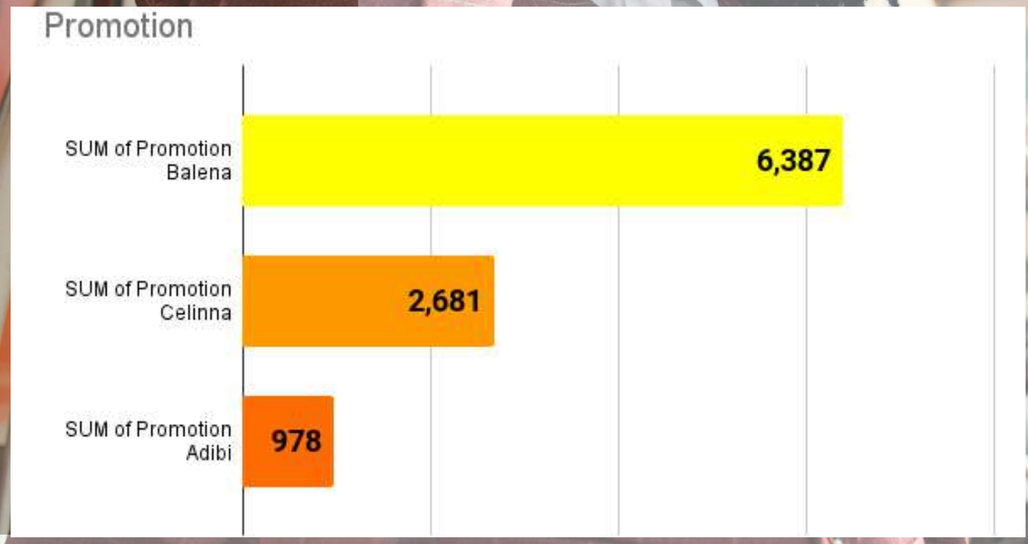


EXPLORATORY DATA ANALYTIC

Purchase Perfromance

Celinna has the highest amount spent products by customers with total of **$117,150,20** money customers spent on this brand.

Based on the total of promotions we can assume **promotion help raise Balena and Celinna quantity order.** That means customers might only purchase Balena because there is a promotion on the product seen by the purchase user with promotion.



EXPLORATORY DATA ANALYTIC

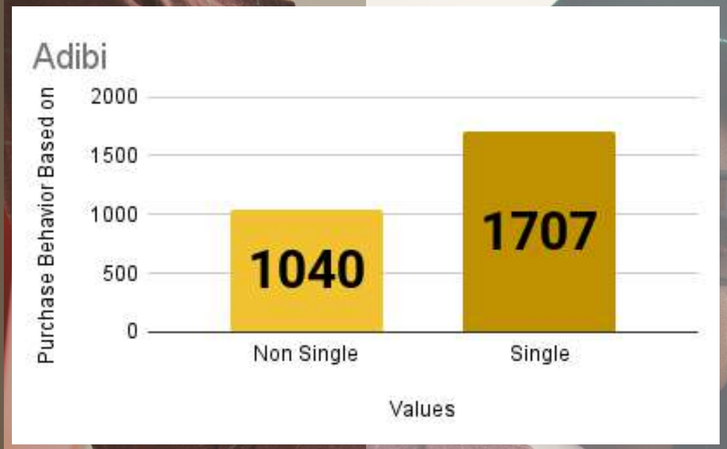
Purchase Perfromance Based on Marital Status

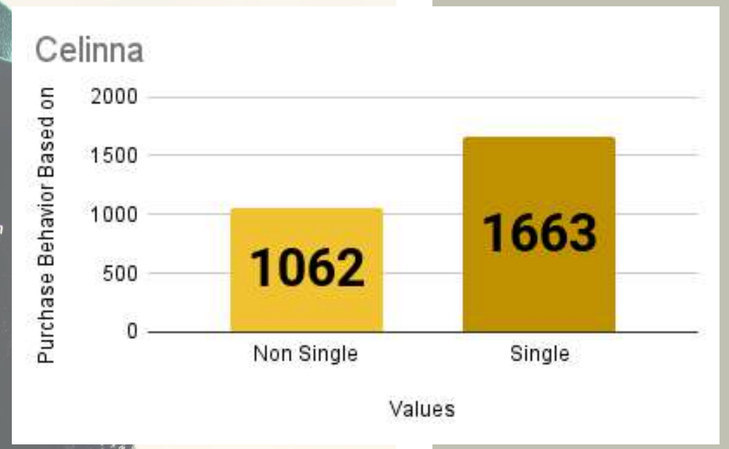
There's a difference in purchase performance of customers who are married and not married

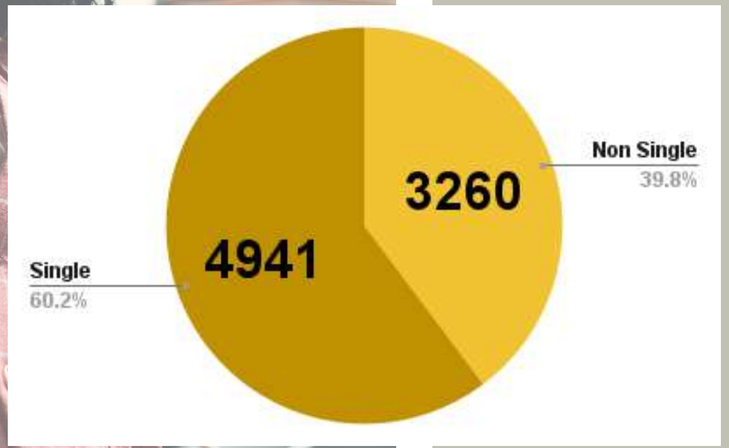
We can see that single (not married) customer purchased more than non-single (married) customer (**More than 50% of purchased products is done by single**)

There is different choice of product, the **single most purchased** product is **Adibi** and the **non-single most purchased** product is **Balena**

It could happen because when you are married and styles don't matter anymore so buying new clothes is unnecesarry



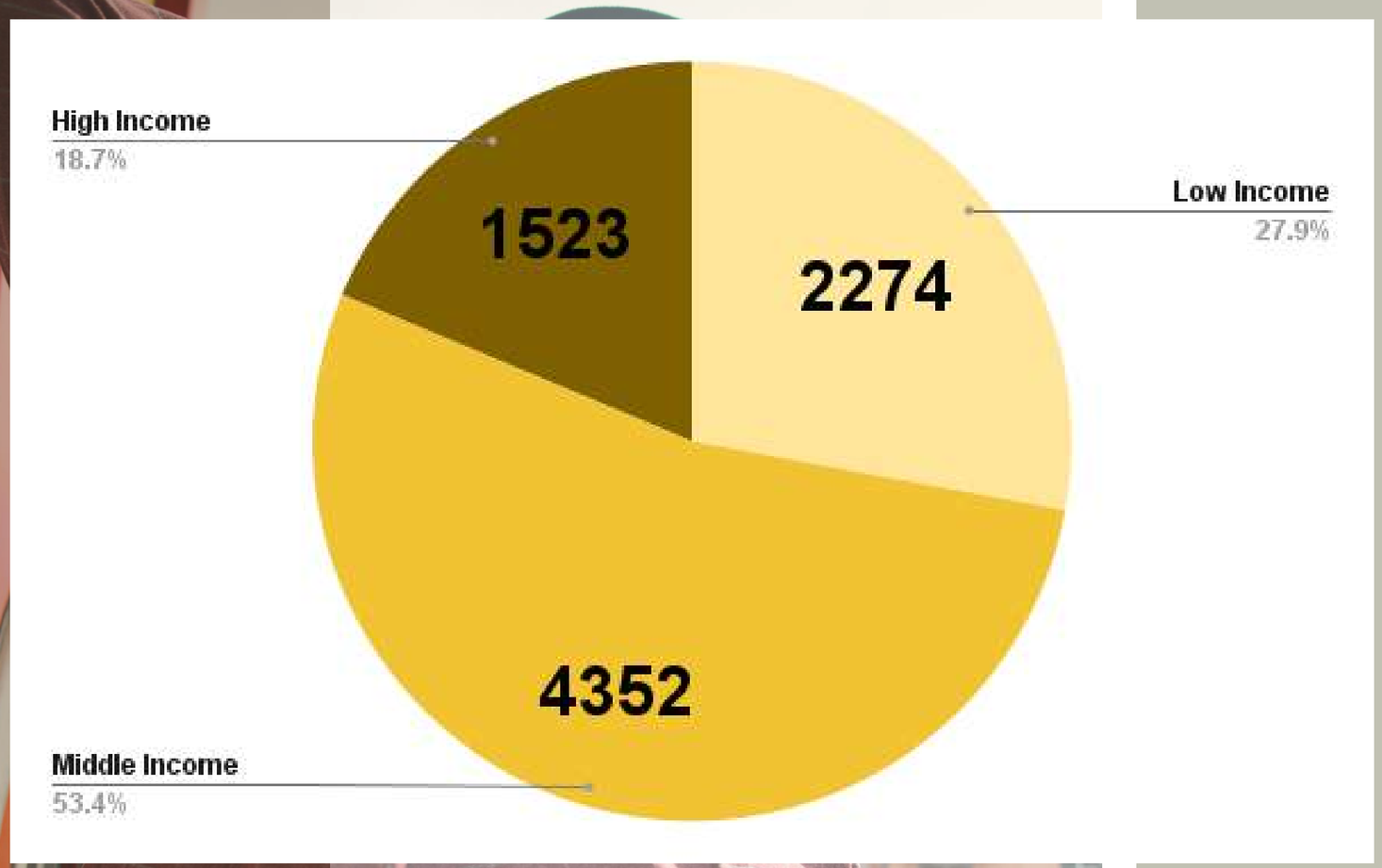




EXPLORATORY DATA ANALYTIC

Purchase Perfromance Based on Income Level

As we can see, customer with middle income level had highest quantity of order and by it means more than 50% of purchasing is done by customer with middle income level with range income > $100,000 and < $150,000. There is a unique condition. Celinna is the most purchased product among all brands if we compare the low income level of purchase behavior.



EXPLORATORY
Monetary of Value

Monetary Of Value	
Celinna Product Price	$42.99
Balena Product Price	$38.11
Adibi Product Price	$29.58

DATA ANALYTIC

Based on the Value of Order (the average amount of money each customer spends per transaction

Adibi has lower-priced products and this might become a causality of the high quantity of order of Adibi.
Adibi might be the best customer choice because of the attainable price and the quality of the product

Recommendation

- The fluctuation of products spending can be a consideration for marketing strategy or inventory management
- Company should give more attention to Adibi because it has the highest contribute of company's revenue by doing a online advertisement campaigns, social media promotions by create an engaging content to grab people attention.
- Always review the forecast the brands