

Statistics & Spreadsheets Intermediate Answer

Syihab Agung Satriotomo
Berlin - Team 4

Milestone 1

1. Business Understanding

Question

A. What is the core business problem faced by Luxura?

Answer

Based on Background brief, that we have 2 Problem in Luxura Business that need to analyze, these are the problems :

1. From Information that I understand we could insight on **which brand that should be Prioritize Adibi, Balena, or Celinna** so we can give company recommendations
2. To test on the **Purchase Performance based average product revenue on each brand that consist purchase information and identity of each user** who has made a purchase to Luxura in the last quarter of 2022

1. Business Understanding

Question

B. Considering the data that you have, how do you think the data can help you answer the business problem?

Answer

1. To Understand customer purchase performance by analyzing **purchase data by total quantity of order, amount spent on brand, and etc.**
2. Analyzing data by comparing X and Y on **Campaign Data** we can **compare which brand that dominating** in the last quarter of 2022.
3. By knowing **Age, Income, Settlement type Total Number of Order, Total quantity of order, and each brand amount spent** we could **Prioritize** which **brand or group** we can **focus** on.

1. Business Understanding

Question

C. What is/are the statistics that you think is/are best to answer business problems in point A?

Answer

1. Business Problem: we could insight on **which brand that should be Prioritize Adibi, Balena, or Celinna** so we can give company recommendations
 - a. Proposed statistics: We may use total average, sum or median on **customer purchase identity (amount on spent, Age, Settlement type, etc)**. Then we could do compare on descriptive statistics to understand **which brand differences predominate**.
2. Business Problem: To test on the **Purchase Performance based average product revenue on each brand that consist purchase information and identity of each user** who has made a purchase to Luxura in the last quarter of 2022
 - a. Proposed statistics: We could **compare the average product revenue of total amount of spent** each brand by campaign

2. Clean Up The Data

Question

A. Remove rows and column with missing data

Answer

1. Removing **Brand Preference column** because the missing data is reaching about 90% of the total data
2. In **sex** column we remove 3 rows, delete 3 rows
3. In **Marital status** column we remove 2 rows, delete 2 rows
4. In **Education** column we remove 1 rows, delete it since it has no effect
5. In **settlement type** column we remove 1 rows
6. In **Income** column we remove 3 rows
7. In **Total Number of Order** column we remove 2 rows
8. In **Total quantity of order** column we remove 2 rows
9. In **Adibi Quantity of Order** column we remove 2 rows
10. In **Balena Quantity of Order** column we remove 1 rows
11. In **Celina Quantity of Order** column we remove 1 rows
12. We Remove **Education column** because it doesn't effect on Total quantity of order and Total number of order

All of above Column that has missing data is consist on rows 152, 315, and 411 so we only delete 3 rows.

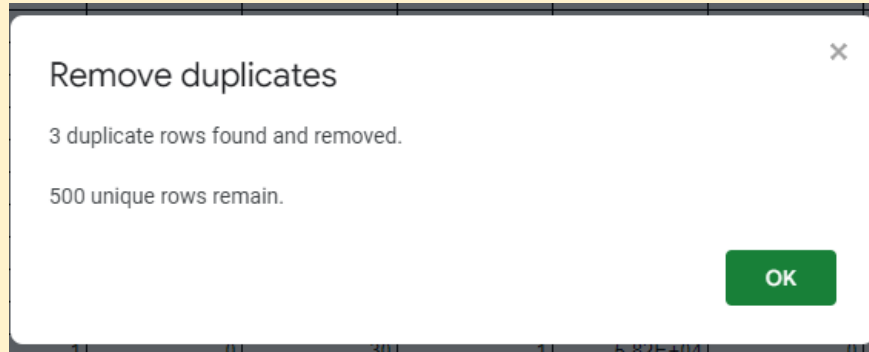
2. Clean Up The Data

Question

B. Remove rows with duplicated price

Answer

Duplicate data has been remove by using Data > Data Cleanup > Remove Duplicates



2. Clean Up The Data

Question

C. Change data type of columns that is not suitable

Answer

1. In **Income column** we formatting the data from scientific to US dollars currency
2. In **Amount Spent on Adibi, Balena, and Celinna column** we formatting from numbers into thousand of dollar currency
3. In **Promotion Adibi, Balena and Celinna** column we change it to Percentage(%) format

Reasoning

- Since we want to calculate each column we change the format each column into number so it would be easier to read and calculate.
- In **Promotion Adibi, Balena and Celinna** we used Number format so it could easier to read and calculate.

3. Descriptive Analytics

Question

A. Create Statistical measurement to know the data distribution

(Insert Google Sheet Link specific page here)

Answer

- <https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRI1Sw3Hioik5ECUVYWYgl7RULA/edit#gid=1241161943>
- <https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRI1Sw3Hioik5ECUVYWYgl7RULA/edit#gid=596024240>
- <https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRI1Sw3Hioik5ECUVYWYgl7RULA/edit#gid=95839343>

3. Descriptive Analytics

Question

B. Eliminate outliers data based on column “Total Number of Order” and “Total Quantity of Order”

Answer

- Based on the data outliers on “**Total Number of Order**” column for the **lower outliers was -1** and **upper outliers was 7**, So there is no lower and upper outliers data in this column (**check at the Sheet Descriptive Statistics1, and check at sheet cleaned with eliminate outliers**).
- Based on the data outliers on “**Total Quantity of Order**” column for the **lower outliers was 2.5** and **upper outliers was 30.5**, So there is a data that has Total Quantity of Order **207, 215, 325** data and we eliminate that data because it can cause effect on “Total Quantity of Order (**check at the Sheet Descriptive Statistics1, and check at sheet cleaned with eliminate outliers**)

milestone 2

4. EDA

Question

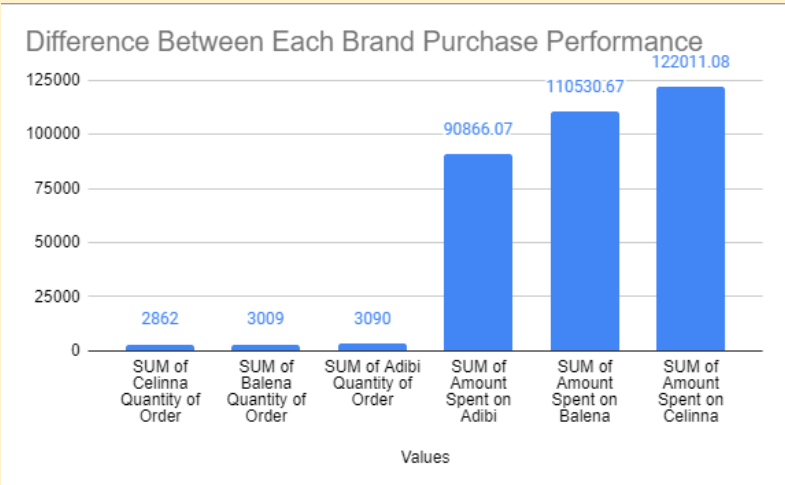
- A. Is there any difference in purchase performance between Adibi, Balena, & Celinna?
Explain your insight based on the data.

(Insert Google Sheet Link specific page here)

Answer

We are using SUM to each brand of **total amount spent** and **total quantity order** using Pivot Table from that we know Purchase performance from each brand that :

- Adibi has highest total quantity of Order
- Celinna has lowest of total quantity of Order
- Celinna has highest amount spent from each customer
- Adibi has lowest amount Spent from each customer



4. EDA

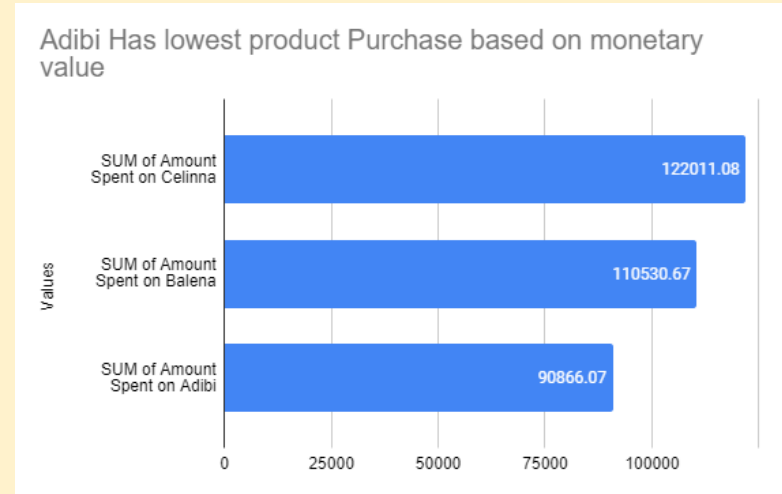
Question

B. Between Adibi, Balena, and Celinna, which brands have relatively lowest product purchased based on monetary value? (Hint: monetary value = value of order)

(Insert Google Sheet Link specific page here)

Answer

Between Adibi, Balena, and Celinna we use total SUM of Amount Spent on each brand base by customer and make a Pivot table and Bar Plot that we can conclude that **Adibi has the lowest product purchased based by monetary value**



4. EDA

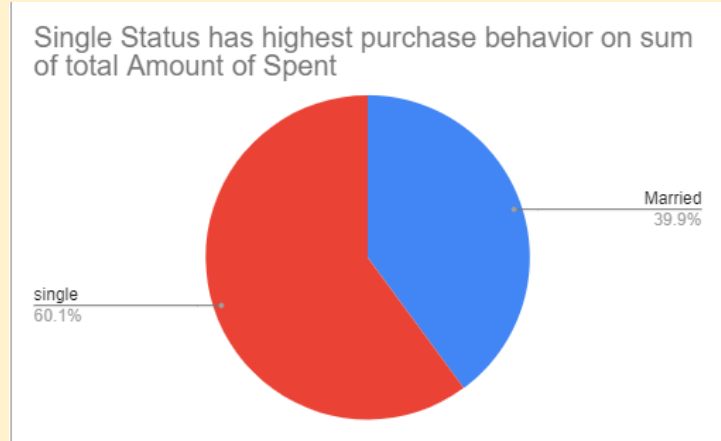
Question

C. What is the difference of purchase behavior between users who are married vs not married?
(Hint: consider Pivot Table or XLMiner)

(Insert Google Sheet Link specific page here)

Answer

From data that we use **Sum of Total amount spent** by Married and single behavior using **Pivot Table** and **Pie Chart** between users who are Married and Not Married, we can conclude that highest buyer is from **Single marital status** which had **60.1%** and **39.9%** are married



4. EDA

Question

D. What is the difference of purchase behavior between users based on their purchasing power? (Hint: use Income as grouping for your preference. For example: Income between 10k - 15k are low income, 15-25k are middle, etc.)
(Insert Google Sheet Link specific page here)

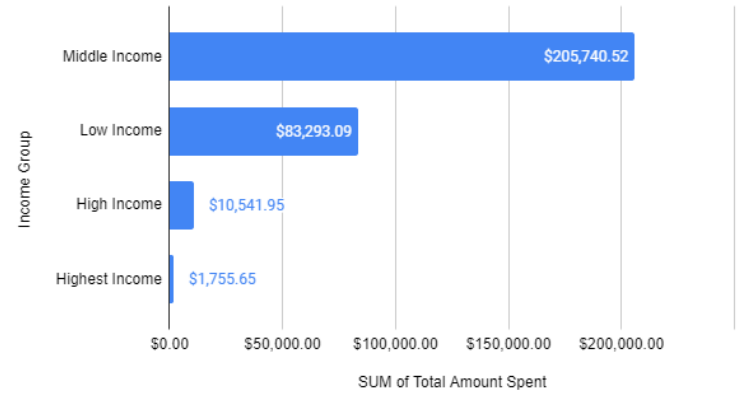
Answer

By Income group that we know

- <100.000 are **low Income**
- 100.000-200.000 are **middle income**
- 200.000-300.000 are **high income**
- >300.000 are **highest income**

We can conclude that **Middle Income** people are more tend to buy our product on our ecommerce based by **Total Amount Spent**

Middle Income group has highest purchase behavior



4. EDA

Question

E. Based on point A-D what recommendation(s) can you provide? (Hint: you can connect it with previous business problem you stated in Milestone 1)

(Insert Google Sheet Link specific page here)

Answer

Based on information from EDA, we found insight that **we need to focus on Celinna Brand campaign** because **Celinna has highest total Amount of Spent rather than Adibi and Balena**, even though **Adibi has paid us 80 million US dollars** which highest fee ever paid to Luxura, but still our customer **still choose Celinna as their favorite Brand** based on **Total Amount of Spent in last quarter of 2022**.

We may also **create some Couple product** that may increase revenue and user growth from **those who are married** and we might focus on **High Income and highest income Group** as we **should create Premium class Product** that very limited as related to Adibi has paid huge fee to luxura so it would **overcome Adibi increase Purchase Value** and releasing Adibi new **Premium product** so we could increase our product value from those group.

5. Hypothesis Testing

Question

A. Is there any significant difference between Adibi and Balena purchase value ?

(Insert Google Sheet Link specific page here)

Answer

From **T-testing Two-sample Assuming equal variance** that Between Adibi and Balena Purchase Value has Significant difference between those two brand by looking **Mean and because P-value < Alpha threshold there is difference between Adibi and Balena Purchase Value**

<https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRI1Sw3Hioik5ECUVYWYgl7RULA/edit#gid=1195953596>

<i>Difference between</i>	<i>Amount Spent on Adibi</i>	<i>Amount Spent on Balena</i>
Mean	164.0433603	209.4616194
Variance	1378.516722	2343.427186
Observations	494	494
Pooled Variance	1860.971954	
Hypothesized Mean	0	
df	986	
t Stat	-16.54661059	
P(T<=t) one-tail	0	
t Critical one-tail	1.646400433	
P(T<=t) two-tail	0	
t Critical two-tail	1.962372801	

5. Hypothesis Testing

Question

B. Is there any significant difference between Adibi and Celinna purchase value?

(Insert Google Sheet Link specific page here)

Answer

From **T-testing Two-sample Assuming equal variance** that Between Adibi and Celinna Purchase Value has Significant difference between those two brand by looking **Mean and because P-value < Alpha threshold there is huge difference between Adibi and Balena Purchase Value**

<https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRIISw3Hioik5ECUVYWYgl7RULA/edit#gid=1195953596>

	Amount Spent on Adibi	Amount Spent on Celinna
Mean	164.0433603	236.4772267
Variance	1378.516722	2852.750929
Observations	494	494
Pooled Variance	2115.633825	
Hypothesized Mean	0	
df	986	
t Stat	-24.74969739	
P(T<=t) one-tail	0	
t Critical one-tail	1.646400433	
P(T<=t) two-tail	0	
t Critical two-tail	1.962372801	

5. Hypothesis Testing

Question

C. Is there any significant difference between Balena and Celinna purchase value?

(Insert Google Sheet Link specific page here)

Answer

From **T-testing Two-sample Assuming equal variance** that Between Balena and Celinna Purchase Value has Significant difference between those two brand by looking **Mean and because P-value < Alpha threshold there is huge difference between Balena and Celinna Purchase Value**

<https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRI1Sw3Hioik5ECUVYWYgl7RULA/edit#gid=1195953596>

	<i>Amount Spent on Balena</i>	<i>Amount Spent on Celinna</i>
Mean	209.4616194	236.4772267
Variance	2343.427186	2852.750929
Observations	494	494
Pooled Variance	2598.089058	
Hypothesized Mean	0	
df	986	
t Stat	-8.329831076	
P(T<=t) one-tail	0	
t Critical one-tail	1.646400433	
P(T<=t) two-tail	0	
t Critical two-tail	1.962372801	

5. Hypothesis Testing

Question

D. Based on the result on point A-C, what insights could you get?

(Insert Google Sheet Link specific page here)

Answer

1. Based on observation each brand has same value so we using t-testing two-sample assuming equal variance even variance not really had huge difference between those three brand
2. Based on calculations **p-value is below 0.05 on each brand so each brand has significantly difference** and we accept H_1 hypotesting
3. **Adibi and Celinna has huge gap that is proof which those two brand has significant difference even Balena-Adibi and Balena-Celinna not that huge gap but still has significant difference.**

<https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRIISw3Hioik5ECUVYWYgI7RULA/edit#gid=1195953596>

5. Hypothesis Testing

Question

E. Based on point D, what business recommendations could you provide?

(Insert Google Sheet Link specific page here)

Answer

Suggestion :

1. We can **Overcome Adibi behind by making more Outfit Idea** on Adibi Product and Market strategy on Adibi product since Adibi has paid Luxura 80 million US dollar this idea could increase Adibi Purchase value
2. Adibi **should duplicate Celinna Idea and Strategy** because Celinna Excellency Purchase Value value
3. Even Balena and Celinna Purchase **value wasn't a huge gap but still Balena could Study from Celinna Excellency**

<https://docs.google.com/spreadsheets/d/1GBEL6Gwa-K5rHSN4oRIISw3Hioik5ECUVYWYgI7RULA/edit#gid=1195953596>