Predicting Customer Purchase Behavior using Customer Purchase Intention

By **Afida Sabrina**

Problem Definition

Coronavirus outbreak increase Year-over-year (YoY) trend in ecommerce

Increase 68% as of mid-April, surpassing an earlier peak of 49% at the beginning of 2020
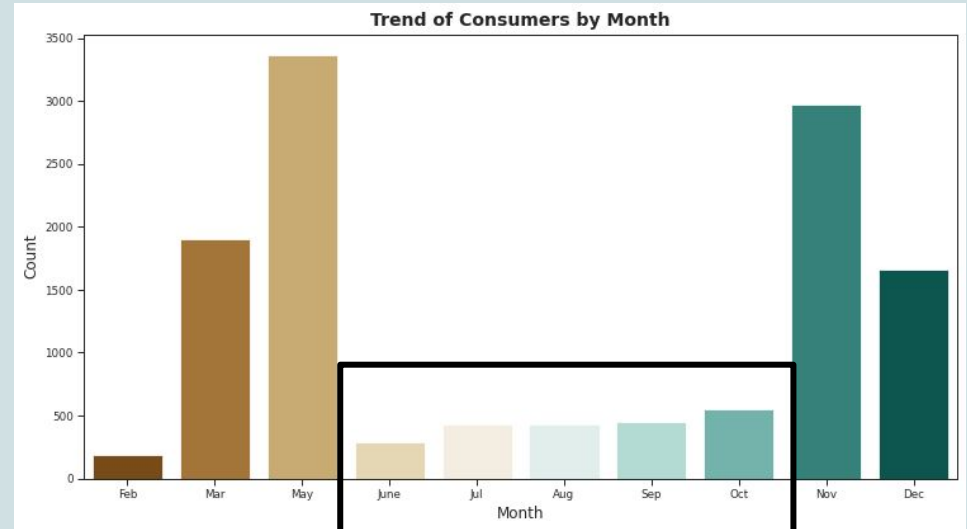
Optimize the Revenue📈

Barrier

Unnatural Decreased of Visitors

Key Outcome

Understand Consumers Behavior

Measure and Analyze

1. *Who's the visitors that only view your products or moving products to their cart?*
2. *How many customers checkout their cart?*
3. *How many sessions finish with transaction?*

Customers Shopping Intention

Predict

Customer Purchase Decision

Key outcome:
Understanding how customers move down the funnel. It helps identify what to fix in the process to increase the number of transactions

Customer Intention is the representation of human motivation to do an action

Observed behavior: Finding information about desired item, buying procedure, warrants, sellers, images, and reviews

Metric

Google Analytic Metric	Bounce Rates	Customers lands on page 1 of your site and hits their browser's back button.
	Exit Rates	Customers on page 2 and quit
	Page Values	Value calculation of the page before consumers complete the transaction
Page View and Duration	Informational Pages	Sample website: communication and address information of the shopping site
	Administrative Pages	Pages about account management
	Product Related Pages	

Data Preparing	Data Profiling	Understanding Dataset
	Exploratory Data Analysis	Find best predictor
	Data Cleaning	
Data Modelling	1. PCA 2. K-means 3. Evaluation	

Data Profiling

Revenue Analysis
Most Cosumers Choose to Not to Finish Their Transaction
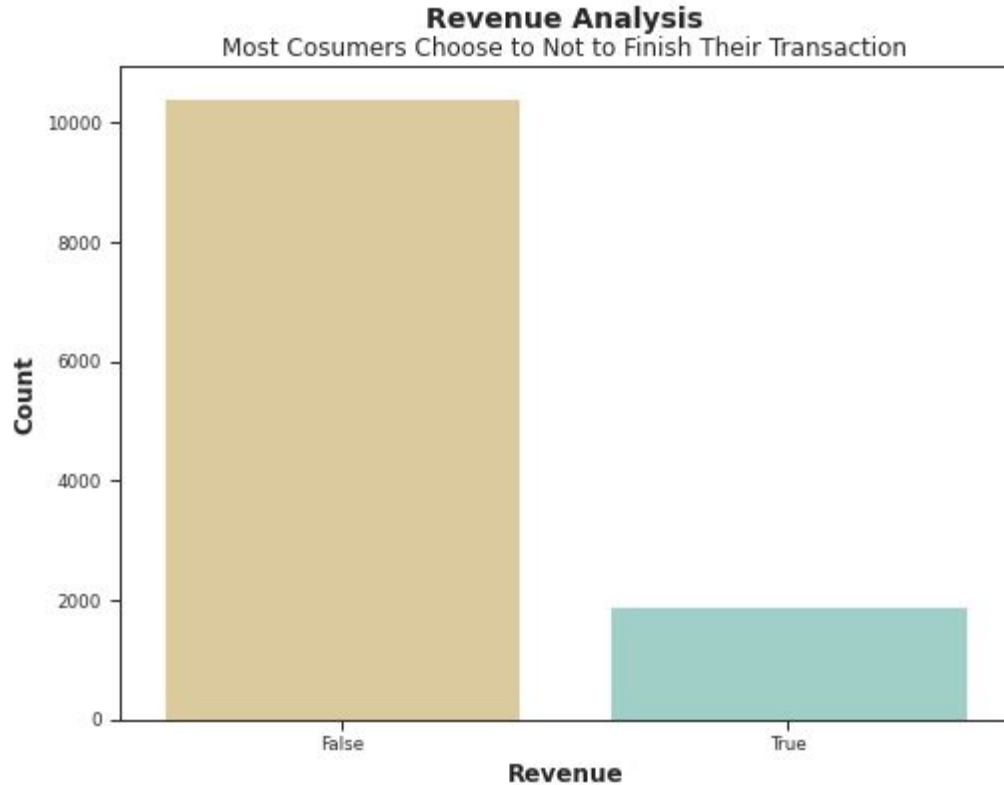
Conversion rate: 0.12

Exploratory Data Analysis
Trend of Revenue by Month

Revenue by Month
Plot is sorted from the month with the highest number of consumers
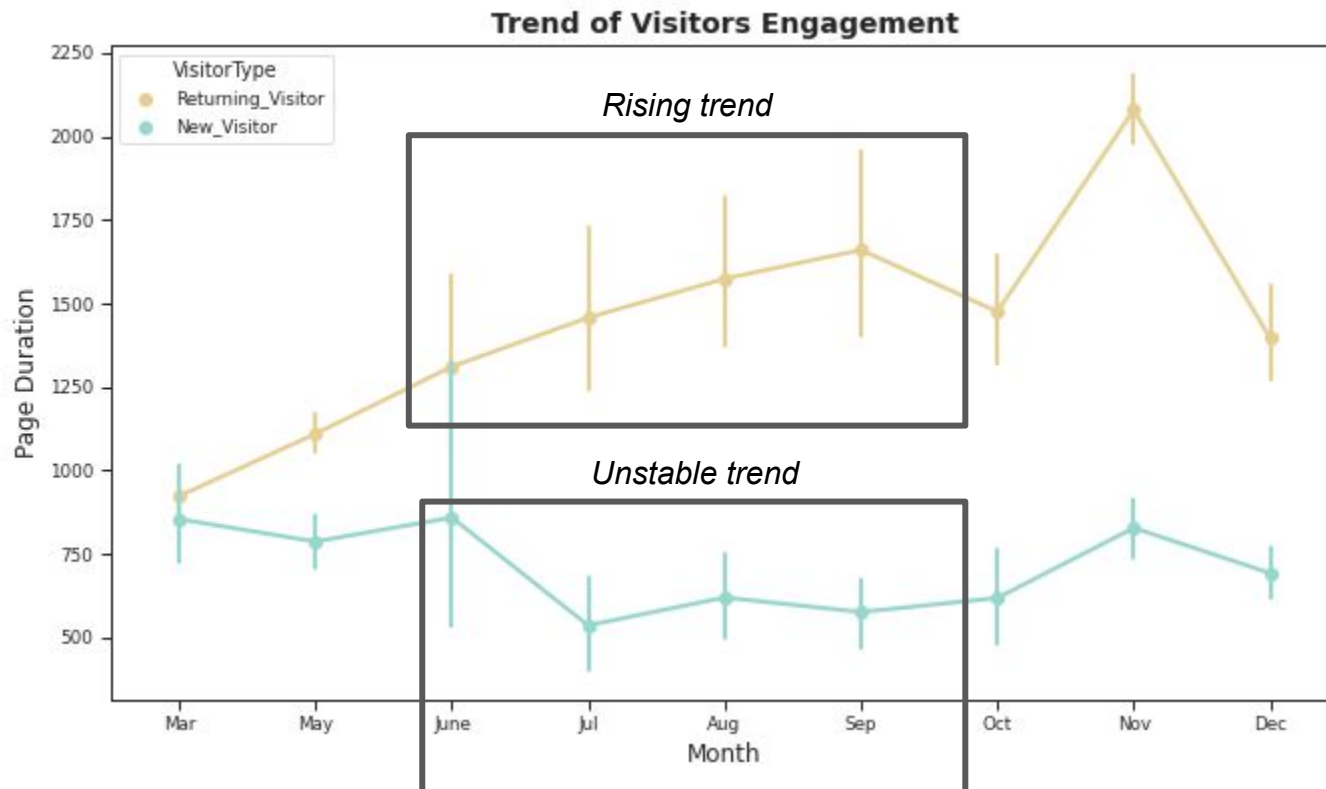
Customers Engagement (Page View and Page Duration) by Month
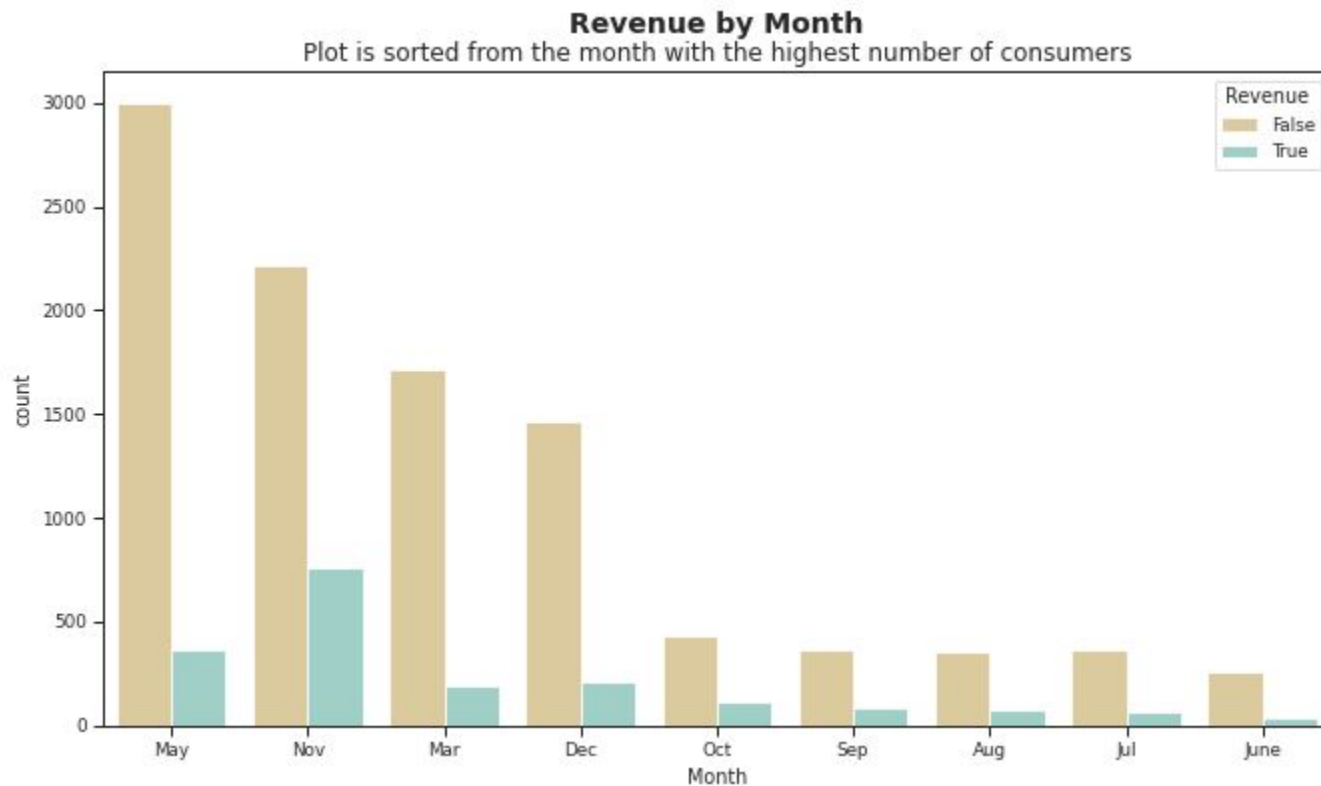
Exploratory Data Analysis
Customers Engagement (Page View and Page Duration) by Visitor's Type

Trend of Visitors Engagement

Rising trend

Unstable trend

VisitorType
- Returning_Visitor
- New_Visitor

Page Duration

Month

Is Engagement Performance Related to the Operating System?

Top 3 of most frequently used Operating Systems by consumers

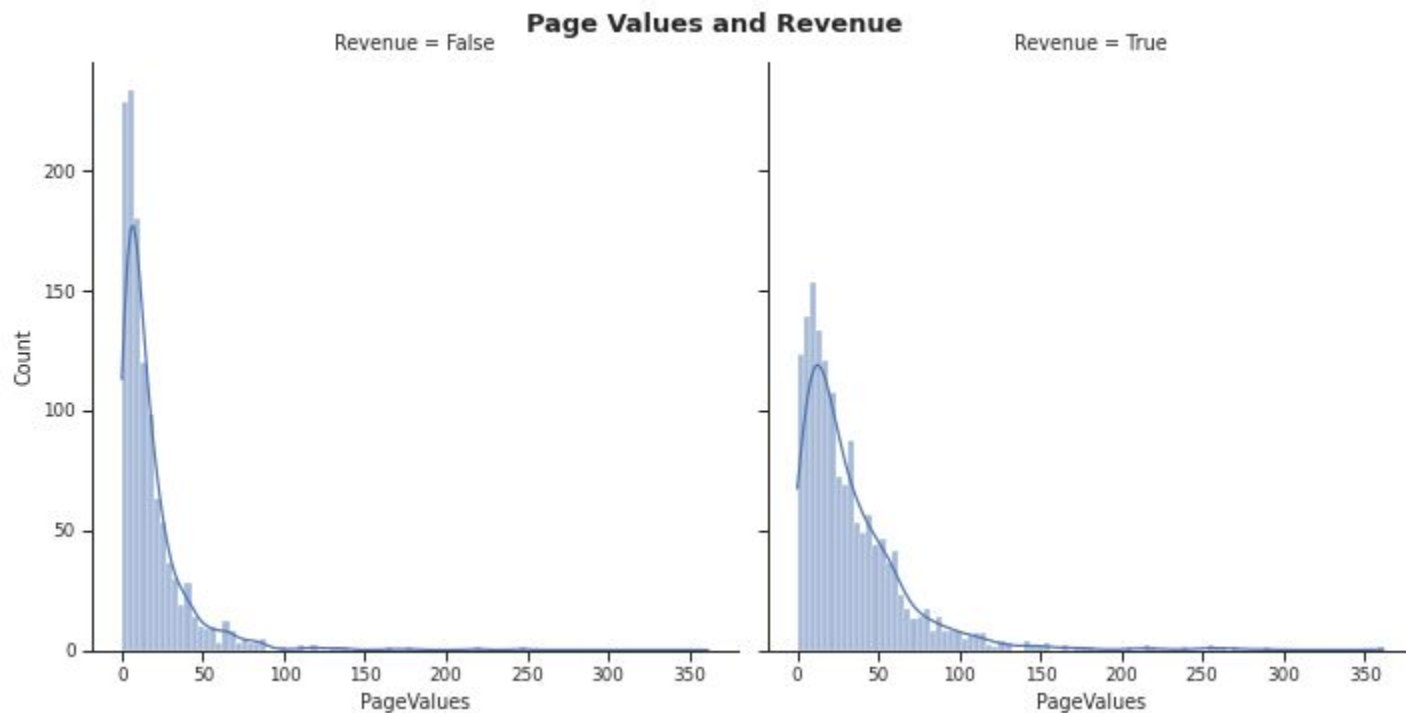Top 3 Operating Systems with the most revenue

Operating System Type	New Visitor	Return Visitor	All
2	935	5567	6502
3	272	2247	2519
1	412	2103	2515

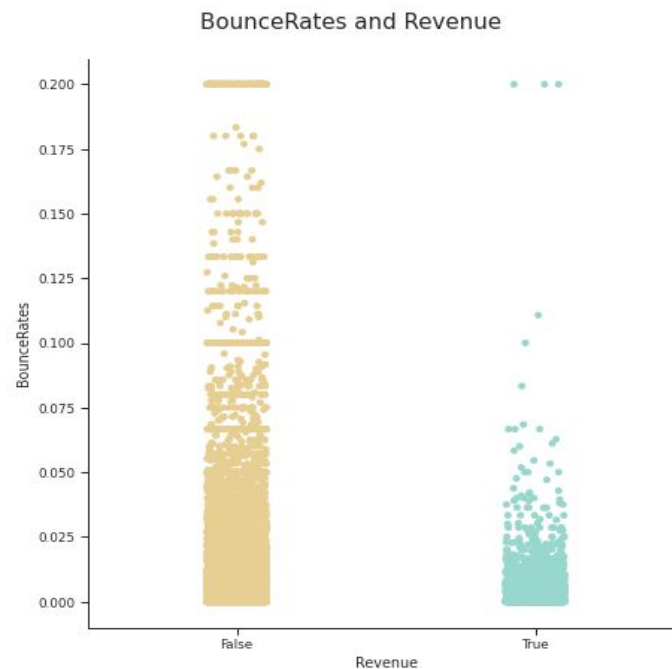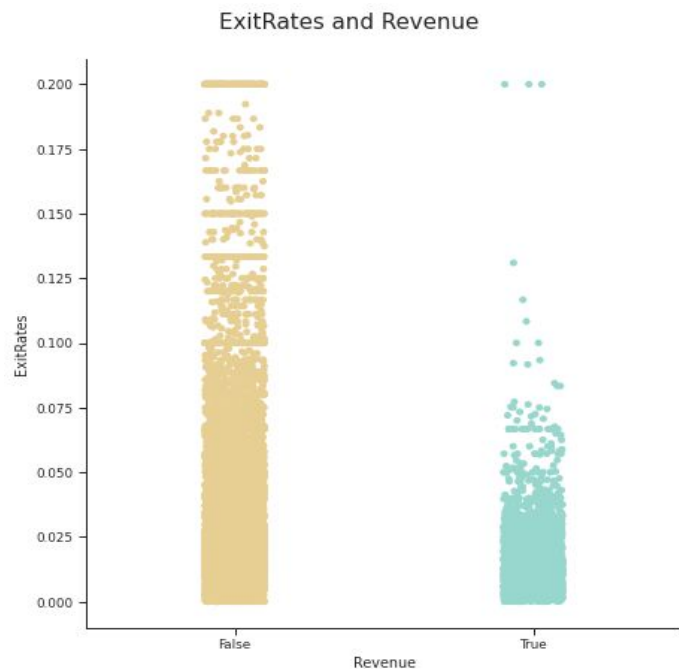| Operating System Type	Revenue
2	1151
1	378
3	268

Page Value and Revenue

Exploratory Data Analysis
Bounce Rates, Exit Rates, and Revenue

Best predictor

Data Modelling

Product Related
Product Related Duration
Administrative
Administrative Duration
Informational
Informational Duration
Bounce Rates
Exit Rates
Page Values

Sorted variable

PCA + K-Means →

Three Clusters* of Consumers Segmentation

*Optimal number of cluster based on silhouette method
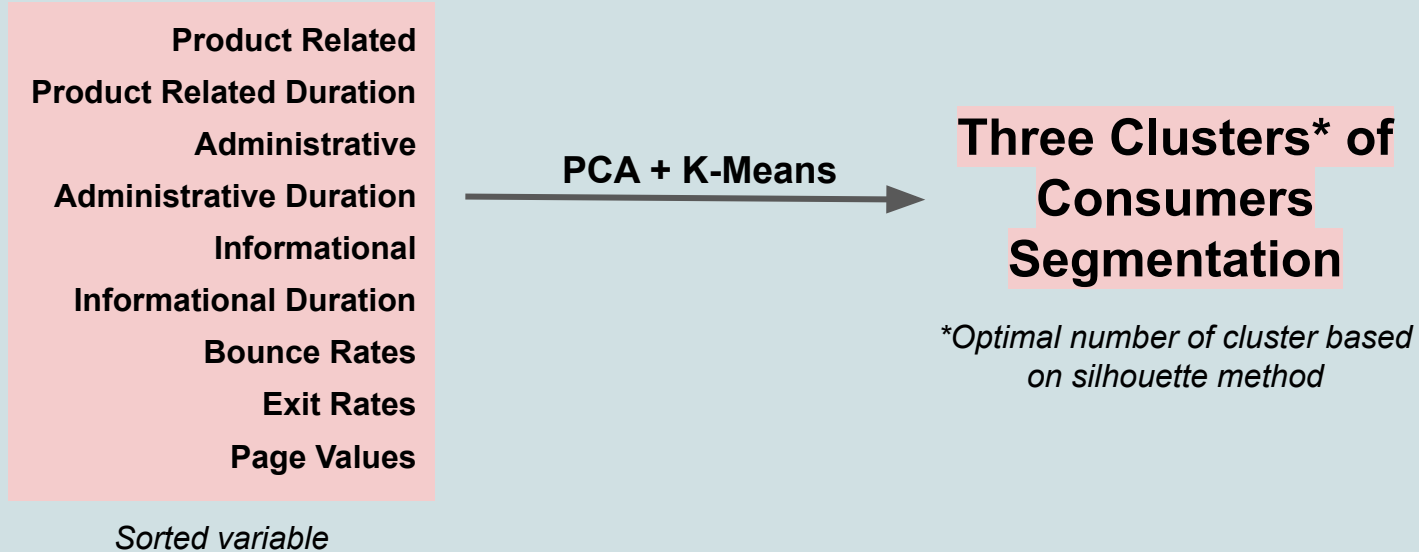
Data Modelling Output

K-means Clustering
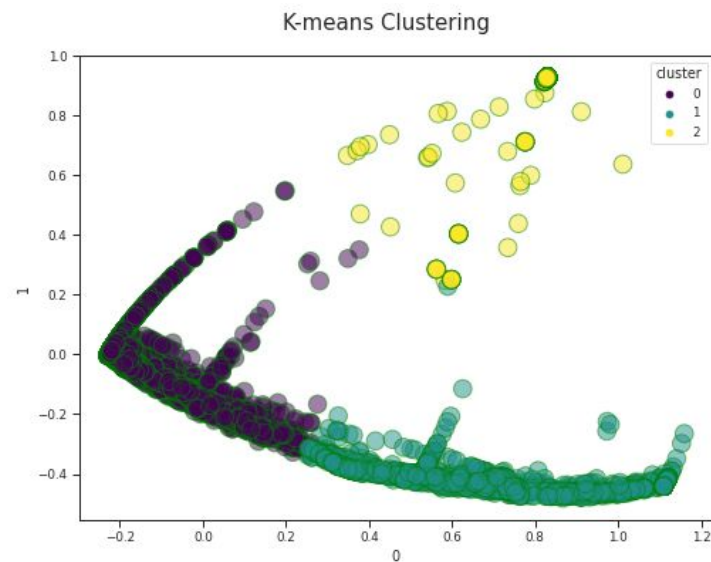
Cluster	Administrative	Administrative Duration	Informational	Information al Duration	Product Related	Product Related Duration	Bounce Rates	Exit Rates	Page Values
0	0	0.000000	0	0.0	14	394.787506	0.003295	0.030114	0.0
1	3	46.053243	0	0.0	2	18.901708	0.000000	0.040726	0.0
2	0	0.000000	0	0.0	1	0.000000	0.200000	0.200000	0.0

Analysis Output

Cluster 0

Consumer's Behavior
Consumers in this cluster spend more time on product related pages than other clusters. It also has higher Exit Rates than Bounce Rates.

Business recommendation
High value of product related shows there is an interest for users to buy the product. Intervention on the consumer can increase the probability of the user to completing the transaction. For example, by providing a discount.

Cluster 1

Consumer's Behavior

Consumers in this cluster have the highest administrative value, while the value of product-related is much lower than cluster 0. Its Bounce rates have lower values than Exit rates.

Business recommendation

This cluster has a value of zero in bounce rates. It indicates a technical problem on the website, for example, page load time is too long. High administrative value and low product-related value designate a bad interface design. In consequence, consumers explore the administrative product only.

Analysis Output

Cluster 2
Consumer's Behavior
Consumers in the cluster seem to have least interest in buying a product based on the low value of the product-related pages exploration and high value of bounce rates and exit rates.
Business recommendation
Most users choose to leave the website without doing any interaction. It makes the company must conduct an in-depth evaluation of the website's performance.

Conclusion

Analysis Conclusion

1. Based on several variables used in clustering, three variables have the same value. Therefore, they're ineffective to be used as a basis of behavioral analysis. These variables are Informational, Informational Duration, and Page Values.
2. Exit Rates and Bounce Rates have a high (0.91) correlation value based on heatmap analysis. It also happens in clustering analysis, the higher the bounce rates, the higher the exit rates.
3. Page View and Page Duration in each cluster show a correlation because the more consumers access a page, the more time they spend.

Business Recommendation Conclusion

Based on the analysis, the website's performance influences the amount of earned revenue. It can be improved by redesigning the user interface to create a better user experience.