E commerce analysis

Prepared for: Dr. Khalid Mohammed.

Prepared by:

Yousef Ahmed Abdelkader 202000016.

Abd El Rahman Yasser Ghonaim 202001502.

Youssef Ayman Abozeid 202001324.

Mohamed Ismail 202003000.

Mohamed Hani Ramadan 202000165.

**Contents**

[**Introduction** 3](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491541)

[**Data Description** 3](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491542)

[**Data Cleaning** 4](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491543)

[**Exploratory Data Analysis (EDA):** 4](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491544)

[**Modelling:** 5](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491545)

[**Results:** 6](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491546)

[**Conclusion:** 6](file:///C:\Users\joyou\Downloads\CSCI322%20-%20Project%20Documetation%20Template.docx#_Toc132491547)

# **Introduction**

E-commerce has become an integral part of our lives, and its importance has grown significantly in recent years. The rise of online shopping has completely transformed the retail industry, with more and more consumers preferring to shop from the comfort of their homes. As a result, understanding the dynamics of e-commerce has become crucial for businesses looking to succeed in the modern world.

The problem we are trying to solve in this project is to analyse the performance of e-commerce websites and identify areas for improvement. Specifically, we aim to determine the factors that affect customer satisfaction and loyalty in the e-commerce industry. By doing so, we hope to provide insights that can help businesses improve their online shopping experience and increase customer retention.

The importance of this problem lies in the fact that customer satisfaction and loyalty are key drivers of business success in the e-commerce industry. A positive customer experience can lead to increased sales, repeat purchases, and positive word-of-mouth, while a negative experience can have the opposite effect. Thus, understanding what factors influence customer satisfaction and loyalty can help businesses optimize their online shopping experience and gain a competitive advantage.

To achieve our objectives, we analysed data from a survey that collected information on customer satisfaction and loyalty for various e-commerce websites. We used statistical methods to identify the factors that have the most significant impact on customer satisfaction and loyalty. Our findings provide insights that can help businesses improve their online shopping experience and increase customer retention.

# **Data Description**

In the data description section, you should provide a detailed description of the dataset you used in your analysis. This section should help the reader understand the nature and quality of the data, and it should include the following information:

* **Data source:** Provide information on where the data was obtained from, including any relevant metadata or documentation.
* **Size and format:** Describe the size and format of the dataset, including the number of rows and columns, and the types of variables.
* **Data type:** Specify the type of data that is included in the dataset, such as categorical, numerical, or textual.
* **Missing values:** Describe the percentage of missing values in the dataset and explain how you handled them.
* **Outliers:** Explain how you identified and handled outliers in the dataset, if applicable.
* **Data quality:** Discuss any issues with the quality of the data, such as data entry errors, duplicates, or inconsistencies.

It's important to provide a clear and detailed description of the dataset in this section, as it will help the reader understand the context and quality of your analysis. Providing this information will also help to establish the credibility of your results and enable others to replicate your analysis if necessary.

# **Data Cleaning**

In the data cleaning section of your project report, you should describe the steps you took to clean and pre-process the data before conducting your analysis. This section should include the following information:

* **Handling missing values:** Describe how you handled missing values in the dataset, including any imputation techniques or deletion methods used.
* **Handling duplicates:** Explain how you identified and handled duplicate data, if applicable.
* **Handling outliers:** Describe how you identified and handled outliers in the dataset, if applicable.
* **Data transformation:** Explain any data transformation techniques you used, such as scaling or normalization.
* **Data integration:** If you used multiple datasets, describe how you integrated them.
* **Data validation:** Explain how you ensured the accuracy and integrity of the data, such as by checking for errors or inconsistencies.
* **Data sampling:** If you used a subset of the data, explain how you selected it and the rationale behind your selection.

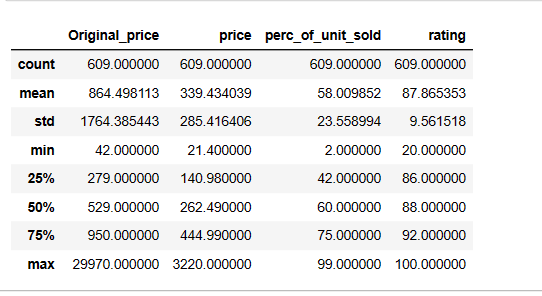
It's important to provide a clear and detailed description of the data cleaning process in this section, as it will help the reader understand the quality of the data used in your analysis. This information will also help others to replicate your analysis and build upon your work. By describing your data cleaning process, you can also provide transparency to your analysis, which is important for establishing the credibility of your results.

***Exploratory Data Analysis (EDA):***

*In this section, we will explore the dataset to gain a better understanding of its characteristics and relationships. The dataset contains information about customer orders for an E-commerce.*

* **Summary statistics**

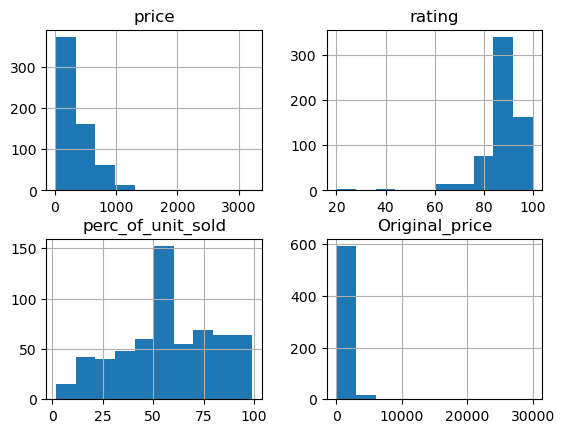
This is a summary of the main descriptive statistics of the dataset, including mean, median, standard deviation, minimum and maximum values, and any outliers.

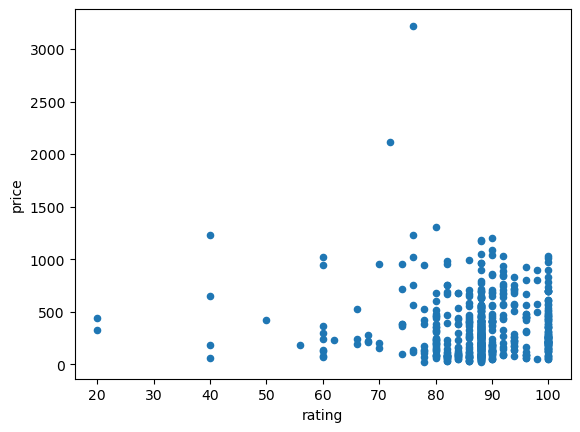
**

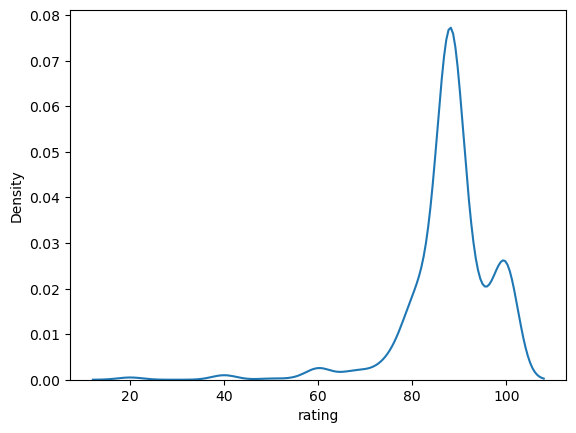
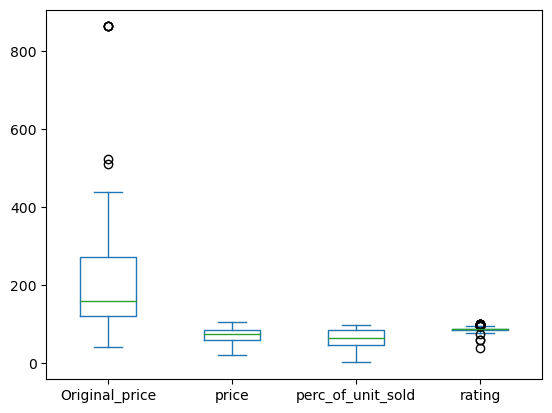
* **Visualizations**

We Created visualizations such as histograms, box plots, scatter plots, heat map, to help illustrate the distributions, relationships, and patterns in the data. Here are some of the data Visualizations.

A picture containing text, plot, diagram, line

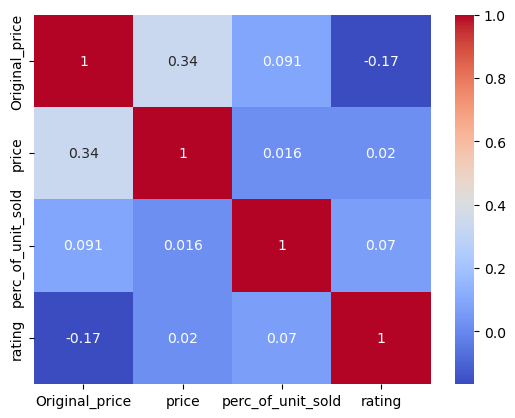
Description automatically generatedA picture containing text, screenshot

Description automatically generated



* **correlation**

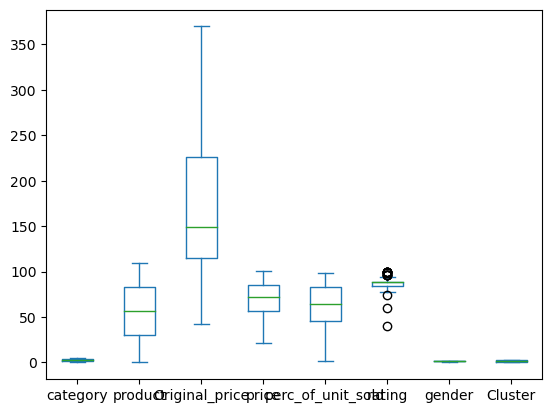
A picture containing text, screenshot, font, number

Description automatically generated****The correlation coefficient calculate the intensity in the linear relationship between the variables involved in a correlation analysis, is usually a value without units which is located between 1 and -1, here is The correlation in this dataset.

* **Outliers and missing data:**

Outliers and missing data are common issues that can affect the validity of statistical analysis. First, we need to Identify any outliers and missing data in the dataset then handled them during the data cleaning process. To handle the missing value in the dataset we change the missing values using statistical methods (mean, median).

To handle the outliers, we removed it to clean the data.

A picture containing text, diagram, screenshot, line

Description automatically generated

# **Modelling:**

In the modelling section of your project report, you should describe the models you developed to analyse your data. This section should include details on the types of models you used (e.g., linear regression, decision trees, neural networks).

You should also present the results of your modelling, including any metrics or evaluation criteria you used to assess the performance of your models. This may include measures such as accuracy or other relevant metrics depending on the type of model you used.

In addition, you can discuss any limitations or assumptions of your models, and any potential sources of error or bias in your analysis.

Overall, the modelling section of your report should provide a clear and detailed description of your modelling approach and the results of your analysis. It should demonstrate your understanding of the modelling techniques.

# **Results:**

In the results section of the project report, you should present the findings of your analysis.

Some of the key elements you can include in the results section are:

* **Model results:** provide the coefficients, standard errors, and goodness-of-fit measures such as R-squared or AIC.
* **Visualizations:** Include visualizations such as graphs, charts, or tables that illustrate the main findings of your analysis.
* **Key results:** Summarize the key results of your analysis and highlight the most important findings.

It is important to be clear and concise in your presentation of results, using plain language and avoiding jargon. You should also interpret the results in the context of the research question and discuss any limitations or future work that could be done.

# **Conclusion:**

The conclusion section is the final part of your project report and should summarize your main findings.

In this section, you should briefly review the research questions or objectives you set out to answer at the beginning of the project and summarize how you addressed them through your data analysis. You can also discuss the implications of your findings.

Finally, you can provide suggestions for further research or improvements that could be made to your analysis, highlighting the potential areas for future work. It is important to make your conclusion concise and to the point, while still conveying the main findings and contributions of your project.