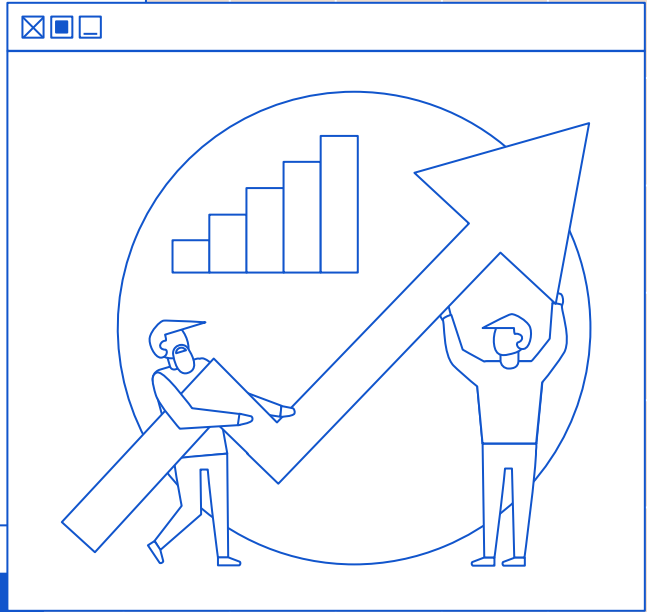


EDA USING AMAZON SALES DATA

MITHUN MEENAKSHI S, UMIP7194





BUSINESS OBJECTIVE

01

SHAPE AND STRUCTURE

To understand the framework of the dataset

02

EXPLORATION

To understand how the orders are distributed

03

EDA

To ask questions and attempt to find answers

04

CONCLUSION

To attempt to tell a story

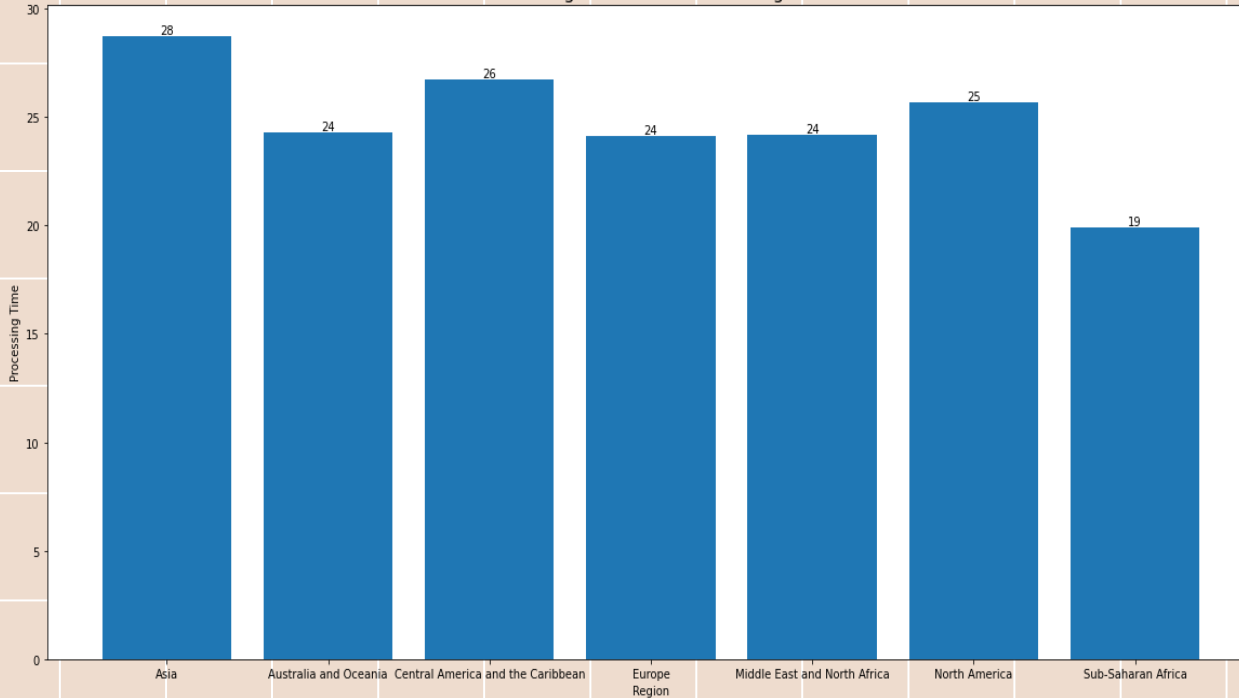


SHAPE & STRUCTURE

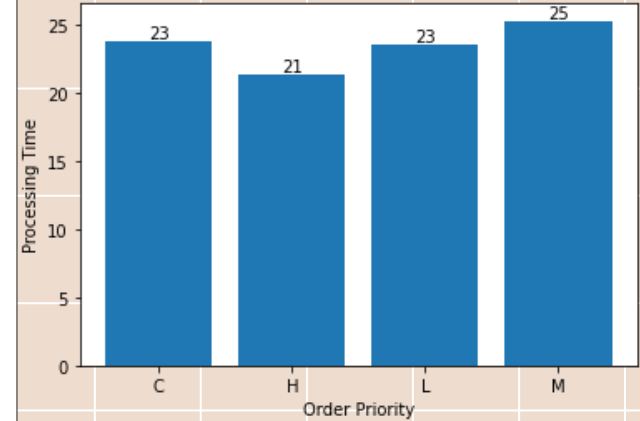
Necessary packages and modules were imported.
The data set was explored using basic functions like `info()`, `describe()`, `isna()` etc..
Shape: (100, 14)
Size: 1400

EXPLORATION

Processing time across various regions



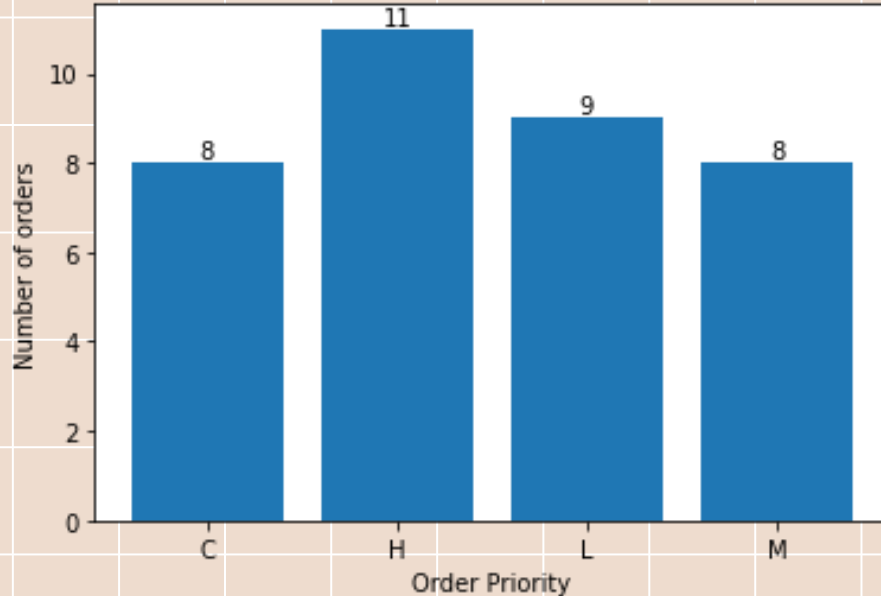
Processing time across various order priorities



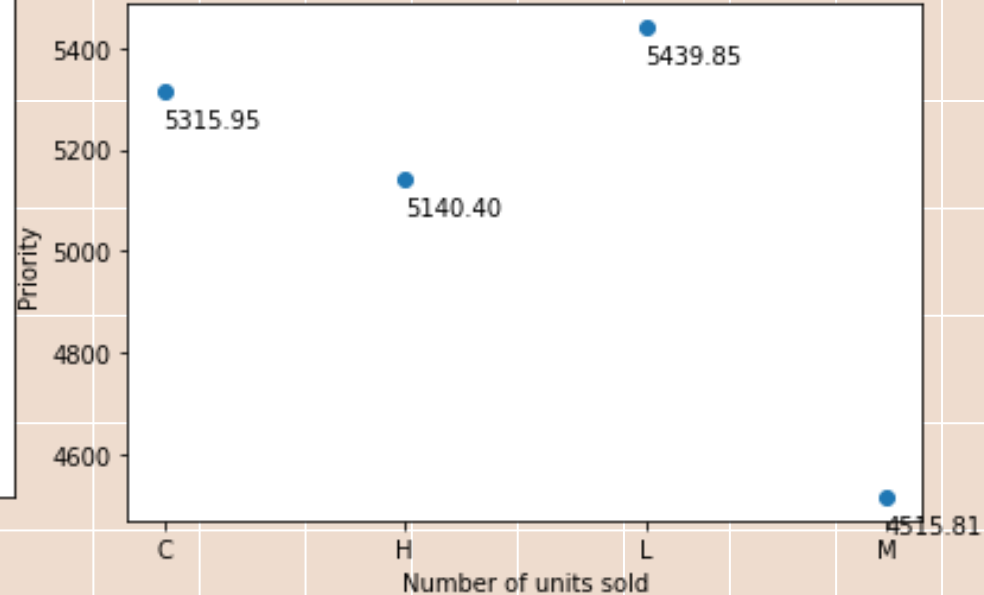
Sub Saharan Africa has the highest amount of orders. Processing time remains similar across order priority. Asia has the highest processing time, and Sub-Saharan Africa has the lowest processing time.

EXPLORATION

Number of orders across various priority levels in Sub-Saharan Africa

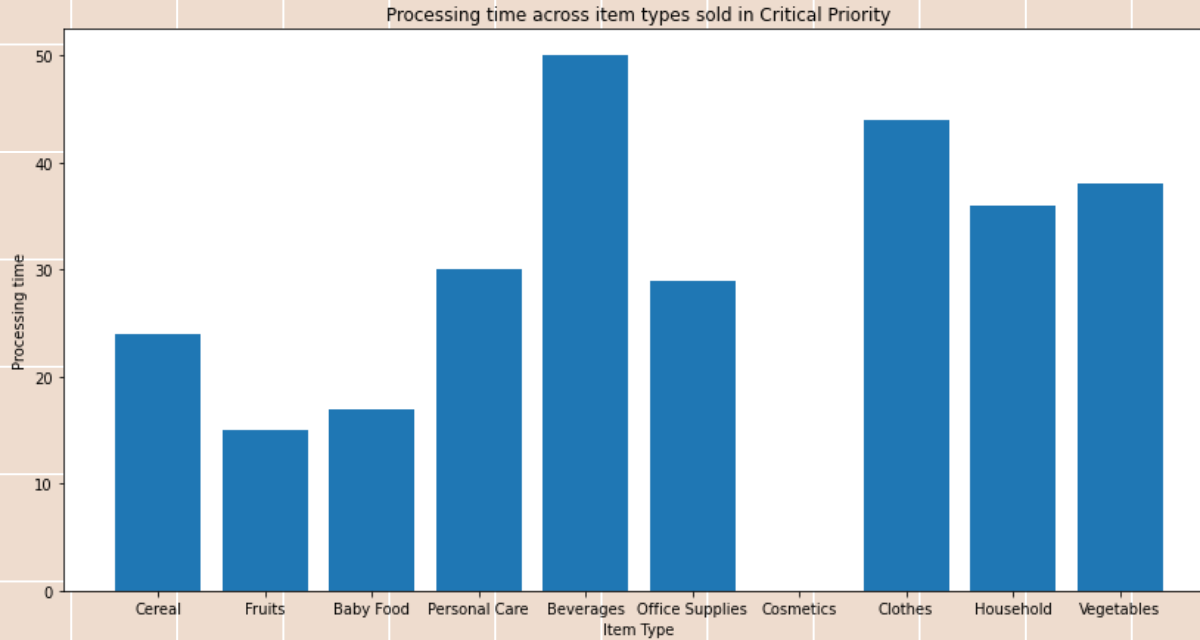


Priority across number of units sold



Sub-Saharan Africa has placed the highest amount of orders in C and H priority. Most of the orders placed were either in C or H priority mode, hence the least processing time. This leads us to question whether number of units has any impact on processing time.

EDA



Statistical testing proves that number of units has an impact on processing time. Further analysis of processing time revealed that a mid-sized order for beverages took 50 days to process in C-priority. Outlier testing revealed that the data point is not an anomaly



EDA



- A two-sample statistical test was conducted to check whether the ordering mode (Online or Offline) impacted the processing time, which revealed that the ordering mode doesn't impact processing time.
- A hypothesis test was conducted to check whether processing time had any impact on total profit, and the result affirmed the same.
- There is not enough data to determine the reason for the one order which took 50 days to process, which in turn impacted the mean processing times.



CONCLUSION



The shape and structure of the dataset was examined.



One mid-sized order took 50 days to process in C priority



EDA was conducted to reveal the factors affecting processing time



Statistical tests were conducted to analyse impact.



Outlier test revealed that the data point is not an outlier



EDA concluded that there is not enough data to give valid reason

