

Mini Project – Amazon Sales Analysis

Batch: Data Analytics & Data Science

Estimated Duration: 2 Days

Problem Statement:

1. What are the first five rows of the dataset?
2. Describe the categorical data in the dataset?
3. What are the column names and their datatypes?
4. Are there any missing values in the dataset? If so, which columns have missing values?
5. How many unique regions are there in the dataset?
6. How many sales transactions occurred in each region?
7. What are the unique item types sold?
8. What is the distribution of sales channels (Online vs Offline)?
9. What is the total number of sales transactions for each order priority (High, Medium, Low, Critical)?
10. How many unique countries are represented in the dataset?
11. What is the total revenue generated?
12. What are the average, median, and standard deviation for the following?
 - i. units sold
 - ii. unit price
 - iii. unit cost
 - iv. total cost
 - v. total revenue
 - vi. total profit
13. What is the total profit made.?
14. What is the total revenue generated in each region?
15. What is the total profit made in each region?
16. Which country has the highest total revenue?
17. What is the most common item type sold?
18. What is the least common item type sold?
19. What is the average unit price for each item type?
20. What is the average unit cost for each item type?

21. Which sales channel generates more revenue: Online or Offline?
22. What is the distribution of total profit by order priority?
23. How does the total revenue compare across different regions?
24. How does the total profit compare across different regions?
25. What is the trend of total revenue over time (monthly or yearly)?
26. What is the trend of total profit over time (monthly or yearly)?
27. Are there any outliers in units sold, unit price, or unit cost?
28. What is the relationship between units sold and total revenue?
29. What is the relationship between units sold and total profit?
30. What is the relationship between unit price and total revenue?
31. Which factors (region, country, item type, sales channel, order priority) have the most significant impact on total profit?
32. Create a pivot table showing the average total revenue and total profit for each combination of region and item type.
33. Perform a time series analysis on the total revenue to identify any seasonal trends or patterns.
34. Perform a time series analysis on the total profit to identify any seasonal trends or patterns.
35. Which item type has the highest profit margin (average total profit per unit sold)?
36. Which region has the highest profit margin (average total profit per unit sold)?
37. How does the sales performance of different item types vary across different regions?
38. Identify any correlation between the numerical variables (units sold, unit price, unit cost, total revenue, total profit).
39. Visualize the distribution of total revenue and total profit for each region using box plots.
40. Create a heatmap to visualize the correlation between different numerical variables.

Points:

Successful executable code: 20 Marks.

Presentation (online/offline): Additional 10 Marks.

Submission: Code in Jupiter Notebook, Reports.