Exploratory Data Analysis (EDA) Assignment

Total Marks: 50

Instructions

- Use the dataset provided to perform Exploratory Data Analysis.
- All work must be done in a **Google Colab notebook**.
- Include your code, visualizations (if any), and **markdown cells for insights and observations**.
- Submit your final work as a **Colab notebook link**.

Section A: Data Overview (10 Marks)

1. What is the shape of the dataset?
2. Identify the data types of all columns.
3. List the numerical and categorical features.
4. Are there any duplicated rows or columns?
5. How many unique entries are there in categorical fields like city or gender?

Section B: Data Cleaning (10 Marks)

6. Identify any missing values in the dataset. Which columns are affected and what are their proportions?

7. Suggest and apply appropriate strategies to deal with missing values.

8. Explore the dataset for any outliers. Which columns show unusual values?

9. What approach do you take to handle outliers, and why?

10. Briefly explain any assumptions you made during cleaning.

Section C: Descriptive Statistics (10 Marks)

11. Use `.describe()` to summarize the dataset. What do you observe about central tendency and spread?

12. What are the minimum, maximum, and average values of income and age?

13. How is the `purchases` column distributed overall?

14. What percentage of users made zero purchases?

15. Comment on the range of values in `days_on_platform`.

Section D: Grouping & Binning (10 Marks)

16. Group the data by `gender`. Compare average income and number of purchases.

17. Create age bins (e.g. 10–20, 21–30, etc.). Analyze average purchases per age group.

18. Group by `city`. Which city has the most users and the highest average `days_on_platform`?

19. Group the data by both `city` and `gender`. What patterns do you observe?

20. Identify which age bin and gender combination shows the highest purchase activity.

Section E: Insights & Interpretation (10 Marks)

21. What do you think are the top 3 factors influencing purchases in this dataset? Justify your answer using group comparisons or summaries.

22. Do users with higher income consistently make more purchases?

23. Does time spent on the platform relate to purchase behavior?

24. Are there any noticeable patterns based on gender or location?

25. Based on your analysis, suggest one strategy the platform can use to increase purchases.