Market Basket Analysis with Python- Amazon

Project overview:

You are hired as a **Customer Insights Analyst** for Amazon, a leading global e-commerce platform. With millions of customers making purchases across diverse categories, the leadership team wants to understand customer purchasing behavior, product affinities, and engagement trends to enhance personalized shopping experiences and optimize inventory and marketing strategies.

The leadership is particularly interested in understanding:

- What products or categories are frequently purchased together, indicating cross-selling opportunities?
- How do customer purchasing patterns vary across demographics or regions?
- Can we segment customers based on buying behavior to target personalized promotions?
- Which products drive repeat purchases or high customer loyalty?

Objective: Build a comprehensive Customer Purchasing Behavior and Product Recommendation Report using Python for data analysis. This report should clean and consolidate Amazon's transaction and customer data, apply techniques such as Market Basket Analysis, clustering, and association rule mining, and provide actionable insights for personalized product recommendations, promotional strategies, and inventory optimization.

(If you're unfamiliar with concepts, or domain-specific terms, feel free to take the help of Google or AI tools like ChatGPT or Gemini to understand them better and apply them effectively in your project.)

Dataset: **Amazon Customer Behaviour Survey**

Dataset Overview: This dataset captures Amazon customers' online shopping behavior, including demographics, purchase frequency, browsing habits, product search preferences, review engagement, satisfaction levels, and areas for service improvement.

Column Name	Description
Timestamp	Date and time when the survey response was recorded.
age	Age of the respondent.
Gender	Gender of the respondent (Male, Female, Prefer not to say, etc.).
Purchase_Frequency	How often the respondent makes purchases on Amazon (e.g., once a month, few times a week).
Purchase_Categories	Product categories most frequently purchased by the respondent (e.g., Clothing, Electronics, Beauty)
Personalized_Recommendation_Frequency	How often respondents receive personalized product recommendations.
Browsing_Frequency	Frequency of browsing Amazon without necessarily purchasing.
Product_Search_Method	How users search for products (e.g., keyword, filters, browsing categories).
Search_Result_Exploration	How deep respondents go into search results (e.g., first page only, multiple pages).
Customer_Reviews_Importance	Importance rating (scale 1–5) of customer reviews during purchase decisions.
Add_to_Cart_Browsing	Whether users frequently add items to the cart while browsing.
Cart_Completion_Frequency	How often users complete purchases after adding items to their cart.
Cart_Abandonment_Factors	Main reasons for abandoning the cart (e.g., price, shipping, indecision).
Saveforlater_Frequency	How often users save items for later.
Review_Left	Whether the respondent leaves product reviews (Yes/No).
Review_Reliability	How much respondents rely on reviews (Occasionally, Heavily, etc.).
Review_Helpfulness	Whether respondents find reviews helpful (Yes/No).
Personalized_Recommendation_Frequency (duplicate column)	Duplicate version of the earlier column — may need cleaning.
Recommendation_Helpfulness	Whether respondents find recommendations helpful (Yes/No/Sometimes).
Rating_Accuracy	How accurately product ratings reflect actual experience (scale 1–5).
Shopping_Satisfaction	Overall satisfaction level with Amazon shopping experience (scale 1–5).
Service_Appreciation	Aspects of Amazon's service that customers appreciate (e.g., delivery speed, prices).
Improvement_Areas	Areas where customers think Amazon can improve (e.g., packaging, product quality).

Tasks to be performed:

(Please refrain from using AI to perform the tasks mentioned below, as it will only provide generic solutions.)

Task 1: Data Cleaning and Preparation (10 Marks)

- Remove duplicate or inconsistent survey responses.
- Standardize categorical entries (e.g., frequency levels, gender, recommendation responses).
- Handle missing values and inconsistent formats in Product_Search_Method and other fields.
- Rename duplicate or misformatted columns (e.g., remove trailing spaces in Rating_Accuracy).
- Convert numerical rating columns (e.g., Customer_Reviews_Importance, Shopping_Satisfaction) to appropriate numeric types for analysis.

Task 2: Descriptive Behavior Analysis (20 Marks)

- Summarize customer demographics (age, gender distribution).
- Analyze overall purchase frequency and most popular product categories.
- Identify top browsing methods and most common cart abandonment factors.
- Calculate mean and median satisfaction, recommendation helpfulness, and rating accuracy.
- Generate summary statistics and visualizations for key behavioral variables.

Task 3: Customer Segmentation and Profiling (20 Marks)

- Segment customers based on purchase frequency and shopping satisfaction levels.
- Create profiles such as:
    - Frequent Buyers: High purchase frequency, high satisfaction.
    - Occasional Shoppers: Medium frequency, moderate satisfaction.
    - At-Risk Customers: Low satisfaction or frequent cart abandonment.
- Analyze demographic or behavioral differences across these segments.
- Use clustering (e.g., K-Means) for behavioral grouping based on survey responses.

Task 4: Recommendation and Review Insights (10 Marks)

- Examine the relationship between recommendation helpfulness and shopping satisfaction.
- Evaluate how review reliability and helpfulness impact overall ratings.
- Identify trends in how often customers engage with or trust personalized recommendations.
- Suggest actionable insights for improving Amazon's recommendation system.

Task 5: Visualization and Reporting (10 Marks)

- Create attractive visualizations (bar charts, heatmaps, pie charts) for:
    - Purchase categories
    - Browsing frequency distribution
    - Satisfaction levels
    - Correlation between recommendation usefulness and satisfaction
- Summarize findings in a clear and visually appealing dashboard or report format.

Task 6: Video Presentation (20 Marks)

(The summary should be in your own words and must not be generated using AI. Please don't write a script and read it aloud. Use charts from your analysis as visual aids in the video.)

- Record a short video of 5 mins summarizing findings and insights.
- Highlight what drives customer transaction behavior and financial risk.
- Discuss data-backed recommendations for customer engagement or monitoring.

Note:

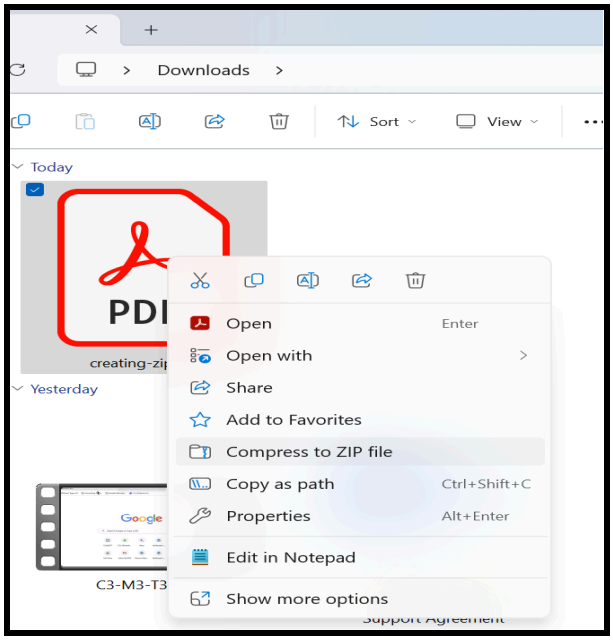
- Place all code files, and summary PDF in one folder and then zip that folder. Follow basic coding ethics (headlines, pointers for insights and comments wherever required)
- Plagiarism will result in a penalty, including possible project disqualification.
- The project will be evaluated based on the quality of analysis and visualizations, depth of insights, feasibility of recommendations, clarity of explanations, and adherence to instructions and deliverables.
- If you set your own criteria (e.g., for flagging risk years or segmenting periods), clearly mention them. Marks will be given according to the specified criteria if they are acceptable.
- Remember to keep the video length less than 5 minutes with your face clearly visible. Add the video link in the code file or in the document.

Submission Guidelines:

- Save the code file, and summary in a PDF and then convert it into a zipped (.zip) folder. **(Please note, the drivelink for the video created should also be added in the PDF itself.)**
- Upload the zipped folder on your respective dashboard.
- Failure to comply with submission guidelines will result in no grading/0 marks.

How to ZIP a PDF file:

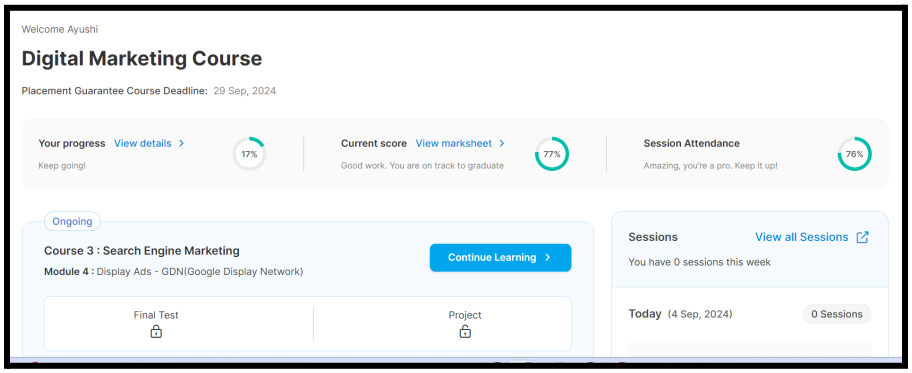
- Put all of the documents (or just one) you want to compress into a new folder.
- Right click on that folder.



- Select the “Compress to ZIP file” option and then click “Compressed (Zipped) folder.”
- A new .ZIP file will be created that contains your document(s).

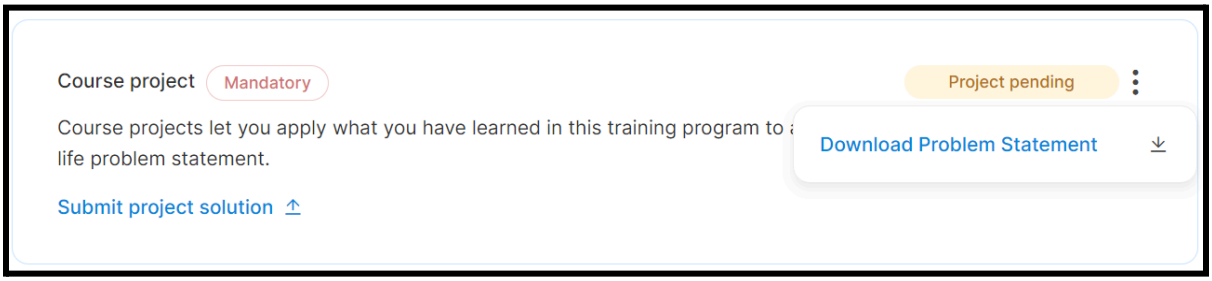
In order to submit the projects please follow the following steps:

1. Click on “ Your progress View details” after logging into your dashboard.



2. Next, click on the tab for the specific child course for which you want to download the problem statement. Then, scroll down to find the "**Course Project**" section.

3. Now, click on the three dots on the right-hand side of the "Course Project" tab to select **"Download Problem Statement."**



4. Please follow the guidelines (screenshot is shared below) provided in the project to ensure correct submissions. Then, click on "**Upload Project Solution**" to submit your work.