DSA Data Analysis Capstone Project

Project Instructions

You are required to select any 2 out of the 3 case studies shared with you for this project. For each selected case study, you will carry out a comprehensive Exploratory Data Analysis (EDA) using the appropriate analytical tools and techniques relevant to the context of each dataset.

📌 **Key Guidelines:**

- All three case studies carry equal marks—choose any two that interest you most or align with your strengths.
- You have 2 weeks from the date of assignment to complete your project.
- Your progress will be tracked and supported through your mentorship sessions on the LMS, where you are encouraged to discuss challenges, insights, and questions.
- Project submission will be done via the LMS by uploading your portfolio link.
- Your portfolio must include at least two separate repositories, one for each of the selected case studies.
- The dataset for each case study will be shared alongside the Project file on the LMS.

🎯 **Deliverables:**

1. Two separate reports are hosted in your GitHub portfolio.
2. Clear, well-documented code and visualizations.
3. Brief insights and interpretations based on your findings.

Use this as an opportunity to demonstrate your analytical thinking, technical skill, and storytelling through data. Good luck, and enjoy the exploration!

✨ Case Study 1: Amazon Product Review Analysis

1. **Company Overview**

You are working as a **Junior Data Analyst** at *RetailTech Insights*, a company that provides e-commerce analytics solutions to sellers on platforms like Amazon. Your team has been

tasked with analysing product and customer review data to generate insights that can guide product improvement, marketing strategies, and customer engagement.

2. Dataset Description

The dataset contains information scraped from Amazon product pages, including:

- **Product details: name, category, price, discount, and ratings**

- **Customer engagement: user reviews, titles, and content**

- **Each row represents a unique product, with aggregated reviewer data stored as comma-separated values**

Total Records: 1,465 rows
TotalFields: 16 columns

3. Analysis Tasks

Use pivot tables and calculated columns where necessary to answer the following:

1. What is the average discount percentage by product category?

2. How many products are listed under each category?

3. What is the total number of reviews per category?

4. Which products have the highest average ratings?

5. What is the average actual price vs the discounted price by category?

6. Which products have the highest number of reviews?

7. How many products have a discount of 50% or more?

8. What is the distribution of product ratings (e.g., how many products are rated 3.0, 4.0, etc.)?

9. What is the total potential revenue (actual_price × rating_count) by category?

10. What is the number of unique products per price range bucket (e.g., <₹200, ₹200−₹500, >₹500)?

11. How does the rating relate to the level of discount?

12. How many products have fewer than 1,000 reviews?

13. Which categories have products with the highest discounts?

14. Identify the top 5 products in terms of rating and number of reviews combined.

4. **Final Task: Dashboard Creation**

Using your cleaned dataset and pivot outputs, build an Excel dashboard. Unleash your Creativity

✨ Case Study 2: Kultra Mega Stores Inventory

1. **Company Overview**

Kultra Mega Stores (KMS), headquartered in Lagos, specialises in office supplies and furniture. Its customer base includes individual consumers, small businesses (retail), and large corporate clients (wholesale) across Lagos, Nigeria.

You have been engaged as a Business Intelligence Analyst to support the Abuja division of KMS. The Business Manager has shared an Excel file containing order data from 2009 to 2012 and has requested that you analyze the data and present your key insights and findings.

Apply your SQL skills from the DSA Data Analysis class and solve both case scenarios as shared in the document.

Case Scenario I

1. Which product category had the highest sales?

2. What are the Top 3 and Bottom 3 regions in terms of sales?

 3. What were the total sales of appliances in Ontario?

4. Advise the management of KMS on what to do to increase the revenue from the bottom 10 customers

5. KMS incurred the most shipping cost using which shipping method?

Case Scenario II

6. Who are the most valuable customers, and what products or services do they typically purchase?

7. Which small business customer had the highest sales?

8. Which Corporate Customer placed the most number of orders in 2009 – 2012?

9. Which consumer customer was the most profitable one?

10. Which customer returned items, and what segment do they belong to?

11. If the delivery truck is the most economical but the slowest shipping method and Express Air is the fastest but the most expensive one, do you think the company appropriately spent shipping costs based on the Order Priority? Explain your answer

✨ Case Study 3: Palmora Group HR Analysis

The Palmoria Group, a manufacturing company based in Nigeria, is embroiled in issues bordering on gender inequality in its 3 regions. Unfortunately, the media recently published the news with the headline "Palmoria, the Manufacturing Patriarchy". This doesn't look good for the owners of the business, based on their ambition to scale the business to other regions and even overseas. Cases like this can only spiral downwards, revealing other issues like the gender pay gap, amongst other possible issues.

The CEO of Palmoria, Mr Ayodeji Chukwuma, is keen to address these issues before they get out of hand. The CHRO, Mr Yunus Shofoluwe, has been assigned the task to identify key areas within the business that could give rise to issues and address them immediately. Mr Shofoluwe decided to recruit you as an HR Analytics expert to analyse the company's HR data and come up with recommendations for management's attention. "Now, the future of gender equality in Palmoria lies in your hands" – the exact words of Mr Shofoluwe before he handed the data to you.

CASE SCENARIO
- Analyse the company data and generate insights that the Palmoria management team would need to address

- Your analysis should be visualised using appropriate charts
- Your focus should be on gender-related issues within the organization and its regions
- The insights required are based on your discretion. However, Mr Gamma, as an insider, has offered to give you pointers on areas you need to pay attention to

Required:

- Generally, there are two genders in the organization. However, some employees refused to disclose their gender. You would need to assign a generic gender status to these employees

- Some employees are without a salary because they are no longer with the company. You will need to take those employees out

- Lastly, some departments are indicated as "NULL". These departments would also need to be taken out.

Pointers from Mr Gamma

1. What is the gender distribution in the organization? Distil to regions and departments
2. Show insights on ratings based on gender
3. Analyse the company's salary structure. Identify if there is a gender pay gap. If there is, identify the department and regions that should be the focus of management
4. A recent regulation was adopted which requires manufacturing companies to pay employees a minimum of $90,000
- Does Palmoria meet this requirement?
- Show the pay distribution of employees grouped by a band of $10,000. For example:
- How many employees fall into a band of $10,000 – $20,000, $20,000 – $30,000, etc.?
- Also visualize this by regions

Case Questions

5. Mr Gamma thought to himself that since you were already working on the employee data, you could help out with allocating the annual bonus pay to employees based on the

performance rating. He handed you another data set that contains rules for making bonus payments and asked you to:

- Calculate the amount to be paid as a bonus to individual employees
- Calculate the total amount to be paid to individual employees (salary inclusive of bonus)
- Total amount to be paid out per region and company–wide