**Case Study-1**

An emerging sustainable clothing brand, EcoWear, is trying to make its mark in the fashion industry. The company's unique selling proposition is to provide high-quality, eco-friendly apparel options at affordable prices.

Being a data-driven company, EcoWear utilises various data sources to inform their strategies. Recently, the company decided to delve deeper into understanding their customers' behaviour and preferences to tailor their product offering and marketing strategy. They conducted a customer survey and collected a considerable amount of data: age, gender, geographical location, shopping habits, and preferences related to the type, colour and fabric of clothing they prefer. The data also included information on customers' willingness to pay more for sustainable products, their online shopping frequency, and the factors influencing their purchasing decisions.

To help navigate this challenge, EcoWear has hired an external data analyst who is a student.

The student was given the data and a clear business objective: to identify key customer segments and predict the types of products that will be well received by these customer segments. The student was also asked to provide recommendations for strategies that EcoWear can adopt to effectively reach and engage these customer segments, with the ultimate goal of improving sales and customer satisfaction.

After weeks of detailed analysis, the student uncovered key insights and patterns from the data, and they were ready to present their findings to the EcoWear management.

1. Describe how the student might have processed and analysed the data. What methodologies or models could they have used to identify key customer segments and predict their preferences? (100 words)
2. Discuss the significance of the insights gained from this data analysis for EcoWear's strategy. How can these insights be used to inform decisions related to product development, marketing and customer engagement? (100 words)
3. Reflect on the role of data-driven decision-making in today's business world. How does data analysis contribute to the success and competitive advantage of an organisation? (200 words)

**Module 1: Describing and Summarizing Data**

**Case Study: Sales Performance of Regional Branches**

**Scenario**:  
You are a business analyst at a retail chain. The company wants to evaluate the sales performance of its 8 regional branches last quarter.

**Tasks**:

1. Import the dataset (branch-wise sales, customer footfall, returns).
2. Calculate:
   * Mean, median, mode, and standard deviation for sales.
   * Identify outlier branches using boxplots.
3. Determine the measures of central tendency.
4. Determine the variance and standard deviation.
5. Graph the distribution of the data and determine whether it is normal or skewed.
6. Create a pivot table using the dataset provided.
7. Create a pivot chart using the same dataset that you used above.
8. Try to use your pivot table and pivot chart to extract some insight from the dataset.
9. Create:
   * Histograms for sales
   * Scatter plots between customer footfall and sales
10. Analyze:
    * Correlation between footfall and sales.

**Learning Outcome**:

* Understand data distribution
* Use Excel to summarize and visualize data effectively

**Pivot Table and Pivot Charts  
The insurance dataset is attached. Please create the pivot table and pivot chart for the following queries: -**

Q1. Create a pivot table that shows the average insurance charges by smoker status and gender.

Q2.Generate a pivot table showing the total charges by region and number of children.

Q3. Build a pivot table to show the count of policyholders by region and smoker status.

Q4. Create a pivot table that displays the average and maximum insurance charges by number of children and gender.

Q5. What is the maximum BMI recorded in each region?

Q6.What is the average charge for each combination of region and smoker status?

Q7.Which gender has higher average charges among smokers and non-smokers?

Q8. What is the maximum BMI recorded in each region?

Q9.  Average BMI grouped by age groups and smoker status

Q10.How does the number of children affect average BMI?

**Binomial distribution**  
Data of  100 policyholders. Each was asked if they filed a claim in the last year (1 = Yes, 0 = No).

Question:

* Calculate the probability of filing a claim (p) from the dataset.
* Assume each policyholder is an independent trial. Plot a binomial distribution showing the probability of k claims out of 100  policyholders.
* Overlay actual data on the same plot (bar chart of actual claim counts).

**Module 2: Sampling and Estimation**

**Case Study: Amazon's Customer Satisfaction Survey (Dataset Attached)**

**Scenario**:  
Amazon wants to estimate the average customer satisfaction score for a new delivery model using a random sample.

**Tasks**:

1. Generate a random sample of 50 responses from a 1,000-customer dataset.
2. Calculate:
   * Sample mean and standard deviation
   * 95% confidence interval for the mean satisfaction score
3. Design 5 effective survey questions
4. Compare results across different regions using stratified sampling

**Learning Outcome**:

* Apply sampling methods and confidence intervals in Excel
* Develop managerial insight from limited data

**Case Study: Amazon A/B Test on Website Redesign(Dataset Attached)**

**Scenario**:  
Amazon tested two versions of its homepage (A & B) to see which one leads to higher purchase conversion.

**Tasks**:

1. Analyze conversion rates from groups A (control) and B (treatment)
2. Formulate hypotheses:
   * H₀: No difference in conversion rate
   * H₁: A significant difference exists
3. Conduct a two-sample t-test in Excel
4. Interpret p-values and decision

**Learning Outcome**:

* Use hypothesis testing for business decisions
* Interpret statistical results using Excel

**Module 4: Single Variable Linear Regression**

**Case Study: Disney Box Office vs. DVD Sales (Dataset Attached)**

**Scenario**:  
Walt Disney Studios wants to predict DVD sales based on box office performance.

**Tasks**:

1. Load historical data for 20 movies
2. Plot a scatter plot between box office and DVD sales
3. Perform linear regression in Excel
4. Interpret:
   * Regression coefficients
   * R-squared
   * Forecast DVD sales for a new movie with $100M box office

**Learning Outcome**:

* Conduct and interpret simple linear regression
* Use Excel’s regression tools for forecasting

**Module 5: Multiple Regression**

**Case Study: Caesars Entertainment Hotel Staffing (Dataset Attached)**

**Scenario**:  
Caesars is trying to predict optimal staffing levels using variables like occupancy rate, day of the week, and event schedules.

**Tasks**:

1. Analyze a dataset with staffing level, occupancy, and event status
2. Perform multiple regression:
   * Use dummy variables for weekends/events
   * Include lagged variables if applicable
3. Evaluate model fit and predictive power
4. Recommend a staffing strategy based on the regression output

**Learning Outcome**:

* Apply and interpret multiple regression
* Improve managerial decisions using complex models in Excel