Course 2: Data Management and Analysis with MS Excel

Project Title: **KPMG Data Analysis using Excel**

Project Overview:

This project aims to analyze customer demographics, transactions, and new customer data to provide insights into business performance and customer behavior. The project consists of six tasks that involve data cleaning, data analysis, and visualization using Excel.

Dataset:KPMG Dataset

Dataset Description:

1. Customer Address Dataset

Description: Contains information about customers' addresses and property valuation.

Columns:

- **customer_id:** Unique identifier for each customer.
- **address:** The street address of the customer.
- **postcode:** Postal code for the customer's address.
- **state:** The state in which the customer resides (e.g., New South Wales, QLD).
- **country:** The country of residence, which is Australia for all entries.
- **property_valuation:** An integer value representing the valuation of the property where the customer resides.

2. Customer Demographic Dataset

Description: Contains demographic information about customers, including their personal details, job information, and purchase behavior.
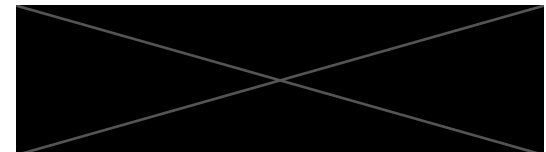
Columns:

- **customer_id:** Unique identifier for each customer.
- **first_name:** Customer's first name.
- **last_name:** Customer's last name.
- **gender:** Customer's gender.
- **past_3_years_bike_related_purchases:** Number of bike-related purchases made by the customer in the past three years.
- **DOB:** Date of birth of the customer.
- **job_title:** Job title of the customer.
- **job_industry_category:** Industry category of the customer's job.
- **wealth_segment:** Wealth segment classification of the customer (e.g., Mass Customer).
- **deceased_indicator:** Indicator if the customer is deceased (Y/N).
- **default:** Default status, including potential erroneous data.
- **owns_car:** Indicates if the customer owns a car (Yes/No).
- **tenure:** The tenure of the customer.

3. Transactions Dataset

Description: Contains details of transactions made by customers, including product details and transaction dates.

Columns:

- **transaction_id:** Unique identifier for each transaction.
- **product_id:** Unique identifier for each product involved in the transaction.
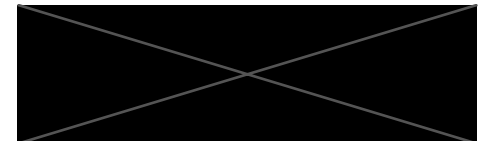
- **customer_id:** Unique identifier for the customer who made the transaction.
- **transaction_date:** The date when the transaction occurred.
- **online_order:** Indicates if the order was made online (TRUE/FALSE).
- **order_status:** Status of the order (e.g., Approved).
- **brand:** Brand of the product purchased.
- **product_line:** Product line (e.g., Standard).
- **product_class:** Product class (e.g., medium).
- **product_size:** Size of the product.
- **list_price:** Listed price of the product.
- **standard_cost:** Standard cost of the product.
- **product_first_sold_date:** The date when the product was first sold.

4. New Customer List Dataset

Description: Contains information about potential new customers, including their personal details, job information, and potential value.

Columns:

- **first_name:** First name of the potential new customer.
- **last_name:** Last name of the potential new customer.
- **gender:** Gender of the potential new customer.
- **past_3_years_bike_related_purchases:** Number of bike-related purchases made in the past three years.
- **DOB:** Date of birth of the potential new customer.
- **job_title:** Job title of the potential new customer.
- **job_industry_category:** Industry category of the potential new customer's job.
- **wealth_segment:** Wealth segment classification.

- **deceased_indicator:** Indicator if the potential new customer is deceased (Y/N).
- **owns_car:** Indicates if the potential new customer owns a car (Yes/No).
- **tenure:** Tenure of the potential new customer.
- **address:** Address of the potential new customer.
- **postcode:** Postal code for the address.
- **state:** State of residence.
- **country:** Country of residence.
- **property_valuation:** Valuation of the property.
- **Rank:** Rank based on some criteria.
- **Value:** Value of the potential new customer.

Task 1: Data Cleaning (12 Marks, 3-3 for each)

Objective: Prepare the datasets for analysis by cleaning and correcting any inconsistencies.
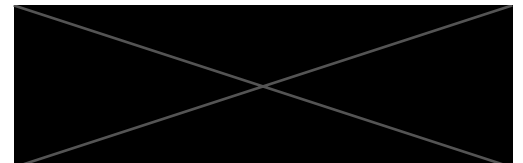
1. **Customer Address Data:**
   - Remove any duplicate records.
   - Ensure all state names are correctly formatted.
2. **Customer Demographic Data:**
   - Identify and correct any erroneous data entries (e.g., invalid characters in `default`).
   - Standardize the format for missing data entries.
   - Correct any anomalies in gender representation.
3. **Transaction Data:**

- Ensure that `transaction_date` is in a consistent date format.
- Remove any records with missing or incomplete information.

4. **New Customer Data:**
    - Standardize address formatting.
    - Ensure consistent gender representation.
    - Correct any anomalies in `job_title` and `job_industry_category`.

Task 2: Customer Segmentation (12 Marks, 4-4 for each)

Objective: Segment customers based on demographic and transaction data to identify key customer groups.

1. **Segmentation by Wealth Segment:**
    - Show the number of customers in each `wealth_segment`.
    - Calculate the average tenure for each `wealth_segment`.
2. **Segmentation by Gender:**
    - Showing the number of customers by gender.
    - Calculate the average `past_3_years_bike_related_purchases` for each gender.
3. **Segmentation by Job Industry:**
    - Showing the number of customers in each `job_industry_category`.
    - Analyze the distribution of `wealth_segment` within each industry.
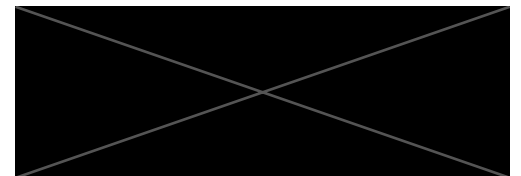
Task 3: Transaction Analysis (12 Marks, 4-4 for each)

Objective: Analyze transaction data to identify trends and patterns.

1. **Sales Trend Analysis:**
    - Create a chart showing the total sales per month.
    - Identify any seasonal trends or significant spikes in sales.
2. **Product Performance Analysis:**

- o Show the total sales for each `brand`.
- o Calculate the total sales and average `list_price` for each `product_line`.

3. **Customer Purchase Behavior:**
   - o Identify the top 10 customers based on total transaction value.
   - o Calculate the average number of purchases per customer.

Task 4: New Customer Insights (12 Marks, 4-4 for each)

Objective: Analyze the new customer dataset to provide insights into potential new customer behavior and value.

1. **New Customer Demographics:**
   - o Show the distribution of new customers by `wealth_segment` and `job_industry_category`.
   - o Calculate the average `past_3_years_bike_related_purchases` for new customers.
2. **New Customer Location Analysis:**
   - o Create a map or chart showing the distribution of new customers by state.
   - o Analyze the correlation between `property_valuation` and customer `wealth_segment`.
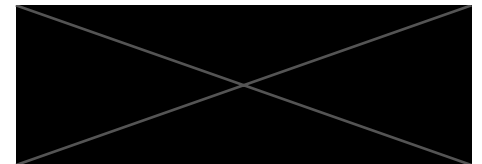3. **Potential Revenue from New Customers:**
   - o Estimate potential revenue based on `past_3_years_bike_related_purchases` and `value`.

Task 5: Customer Lifetime Value (CLV) Analysis (16 Marks)

Objective: Calculate and analyze the customer lifetime value to identify the most valuable customers.

1. **CLV Calculation:**
   - o Use the formula
   - o Calculate CLV for each customer using transaction data.

Formula

CLV=(Average Purchase Value×Purchase Frequency)×Customer Lifespan

Average Purchase Value (APV):

- This is the average amount of money a customer spends in a single purchase.

  APV=Total Revenue/ Number of purchases

- **Total Revenue** is the sum of all revenues generated from all purchases. You can get the total revenue by using Transactional data.
- **Number of Purchases** is the total count of all transactions of that customer.

Purchase Frequency (PF):

- This is the average number of times a customer makes a purchase in a given period.

  PF= Total Number of Transactions / Number of Unique Customers

Customer Lifespan (CL):

- This represents the average number of years a customer remains active.
- In your dataset, this is represented by the `tenure` column in the Customer Demographic dataset.
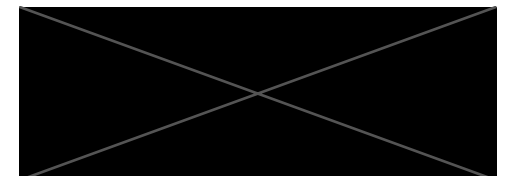
2. **Segment CLV Analysis:**
   - Show average CLV by `wealth_segment`.
   - Analyze the relationship between CLV and customer demographics (e.g., gender, job industry).

Task 6: Executive Summary and Recommendations (16 Marks)

Objective: Summarize findings and provide actionable recommendations for business strategies.

1. **Summary of Key Insights:**

  ○ Highlight key findings from customer segmentation, transaction analysis, new customer insights, and CLV analysis.

2. **Recommendations:**
  ○ Provide recommendations for marketing strategies targeting high-value customer segments.
  ○ Suggest potential areas for business expansion based on new customer location analysis.
  ○ Recommend improvements in product offerings based on transaction analysis.

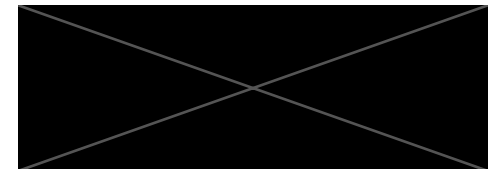**Video explanation: 20 Marks**

Important:

- Please note that, if the student sets their own criteria, they need to clearly mention and explain it. Marks will be given according to the specified criteria if they are acceptable.

- Please do all the tasks in different workbooks for clarity and visibility.

Submission Guidelines:

Save the worksheets in a folder and then convert it into a zipped (.zip) folder.(Please note, the drivelink for the video created should also be added in the sheet 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/sheets you want to compress (or just one) 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).

Note:

● Plagiarism will result in a penalty, including possible project disqualification.

● The project will be evaluated based on the quality of analysis, depth of insights, and feasibility of recommendations.