**Business Understanding**

Al Dubai Fragrances is a privately held luxury goods retailer and distributor, headquartered in Dubai, UAE. Al Dubai Group is the largest retail operator in the Middle East. It played a crucial role in developing the luxury sector in the region. The company has more than 12,000 employees, in 14 countries.

Al Dubai operates in two major modes, one is through its mobile application and the second is through its various offline stores that they have across Dubai and the middle east. Under the brand name Al Dubai Fragrances, they have a wide variety of products for both Males and Females but their major distribution is across fashion products and accessories among female customers.

**Problem Statement**

In recent years, Al Dubai Fragrances has faced a certain dip in their revenue and they are now looking to go through a major transformation in the following two sectors.

1. They are looking to revamp their marketing strategy by targeting the right set of customers.
2. They are looking to promote their application among customers so that they may enter into the “E-Commerce” sector.

You are working as an Associate Data Analyst with Al Dubai Fragrances and you are now required to help the marketing team to identify the various segments of customers that can be found in the data so that the marketing team can accordingly roll out offers to enhance customer engagement and experience with their products.

**RFM Scoring Technique**

RFM Stands for recency, frequency, and monetary value. It’s a very popular scoring technique to identify the best-performing and least-performing customer groups.

Refer to the below demonstration to understand RFM Scoring Technique.

| CUSTOMER ID | RECENCY (DAY) | FREQUENY (NUMBER) | MONETARY (TOTAL) |
| --- | --- | --- | --- |
| 1 | 4 | 6 | 540 |
| 2 | 6 | 11 | 940 |
| 3 | 46 | 1 | 35 |
| 4 | 23 | 3 | 65 |
| 5 | 15 | 4 | 179 |
| 6 | 32 | 2 | 56 |
| 7 | 7 | 3 | 140 |
| 8 | 50 | 1 | 950 |
| 9 | 34 | 15 | 2630 |
| 10 | 10 | 5 | 191 |
| 11 | 3 | 8 | 845 |
| 12 | 1 | 10 | 1510 |
| 13 | 27 | 3 | 54 |
| 14 | 18 | 2 | 40 |
| 15 | 5 | 1 | 25 |

Given is the data of 15 customers with the details as follows:

1. Recency: It’s the number of days since the customers has made the last purchase. You may derive the same by subtracting the current date from the last purchase date. Lower the value, the better the customer.
2. Frequency: It’s the number of transactions made by a particular customer. Higher the value, the better the customer.
3. Monetary: It’s the total amount spent by a customer in making various purchases. Higher the value, the better the customer.

Now, In order to generate RFM Score, we will first be generating the individual scores for each the columns. Refer to the below tables for more details.

| CUSTOMER ID | FREQUENCY | F SCORE |
| --- | --- | --- |
| 9 | 15 | 5 |
| 2 | 11 | 5 |
| 12 | 10 | 5 |
| 11 | 8 | 4 |
| 1 | 6 | 4 |
| 10 | 5 | 4 |
| 5 | 4 | 3 |
| 13 | 3 | 3 |
| 7 | 3 | 3 |
| 4 | 3 | 2 |
| 14 | 2 | 2 |
| 6 | 2 | 2 |
| 15 | 1 | 1 |
| 8 | 1 | 1 |
| 3 | 1 | 1 |

In the above table, a customer with the highest frequency has been given the F-Score of 5 and so on the other customers are scored.

| CUSTOMER ID | MONETARY | M SCORE |
| --- | --- | --- |
| 9 | 2630 | 5 |
| 12 | 1510 | 5 |
| 8 | 950 | 5 |
| 2 | 940 | 4 |
| 11 | 845 | 4 |
| 1 | 540 | 4 |
| 10 | 191 | 3 |
| 5 | 179 | 3 |
| 7 | 140 | 3 |
| 4 | 65 | 2 |
| 6 | 56 | 2 |
| 13 | 54 | 2 |
| 14 | 40 | 1 |
| 3 | 35 | 1 |
| 15 | 25 | 1 |

In the above table, a customer with highest monetary value has been scored with the score of 5 and so on the other customers are scored.

| CUSTOMER ID | RECENCY | R SCORE |
| --- | --- | --- |
| 12 | 1 | 5 |
| 11 | 3 | 5 |
| 1 | 4 | 5 |
| 15 | 5 | 4 |
| 2 | 6 | 4 |
| 7 | 7 | 4 |
| 10 | 10 | 3 |
| 5 | 15 | 3 |
| 14 | 18 | 3 |
| 4 | 23 | 2 |
| 13 | 27 | 2 |
| 6 | 32 | 2 |
| 9 | 34 | 1 |
| 3 | 46 | 1 |
| 8 | 50 | 1 |

In the above table, we have scored the customers based on their recency. Now, if a customer has recently made a purchase, that customer is a good customer for us and vice versa. In the table, a customer with recency of 1 has been scored with a score of 5 and a customer with a recency of 50, has been scored with a score of 1.

| CUSTOMER ID | R SCORE | F SCORE | M SCORE | RFM Score |
| --- | --- | --- | --- | --- |
| 12 | 5 | 5 | 2 | 4.00 |
| 11 | 5 | 4 | 2 | 3.67 |
| 1 | 5 | 4 | 4 | 4.33 |
| 15 | 4 | 1 | 1 | 2.00 |
| 2 | 4 | 5 | 4 | 4.33 |
| 7 | 4 | 3 | 2 | 3.00 |
| 10 | 3 | 4 | 2 | 3.00 |
| 5 | 3 | 3 | 2 | 2.67 |
| 14 | 3 | 2 | 1 | 2.00 |
| 4 | 2 | 2 | 4 | 2.67 |
| 13 | 2 | 3 | 1 | 2.00 |
| 6 | 2 | 2 | 2 | 2.00 |
| 9 | 1 | 5 | 2 | 2.67 |
| 3 | 1 | 1 | 4 | 2.00 |
| 8 | 1 | 1 | 2 | 1.33 |

In the above table, we have taken the average of R-Score, F-Score and M-Score to generate RFM Score. Now, the higher the score, the better the customer, the lower the score, the worst the customer.

**Tasks-Excel**

You are given the excel file with the data consisting of the details of various customers. Perform the following tasks using excel/google sheets to answer the questions.

1. Use the RFM technique defined above to find the top-performing customers and the bottom-performing customers.
   1. Filter the data and create two sheets, one with those customers having an RFM score greater than 4.5. Name this sheet as Top Performing Group
   2. Create another sheet with those customers having RFM Score less than 2. Name this sheet as Worst Performing Group.
   3. Using both sheets and answer the following questions.(Using Pivot Table)
      1. What is the proportion of males and females?
      2. Find the average sales
      3. Find the number of app and non-app users.
2. Create a new column(sales level) using the following conditions(Use if-else).
   1. Total sales<1000: ‘L5’
   2. Total sales>=1000 and total sales<10000: ‘L4’
   3. Total sales>=10000 and total sales<20000: ‘L3’
   4. Total sales>=20000 and total sales<50000: ‘L2’
   5. Total sales>50000: ‘L1’
3. Find the number of customers and the average number of transactions under each sales level.
4. Use V-Lookup to achieve the following result.
   1. You are given the category and the number of transactions ranges in the below table.

| A | 1-10 |
| --- | --- |
| B | 10-50 |
| C | 50-100 |
| D | >100 |

* 1. Using v-lookup append the category column to the data
  2. Find the category-wise count of orders placed.

1. Find the top 5 customers according to the percentage of total sales. (Use pivot table)

**Tasks-SQL**

Al Dubai Fragrances wants to expand its business to the online web market and for that expansion, they are looking to improve its team. The company has the data of its various employees. You are required to answer the following questions to help the Al Dubai Fragrances HR team better understand its employees.

**Tasks:**

1. Load the tables in the database and name them as hr.
2. Find the number of employees working in the ‘Sales’ department. (Use sub-query).
3. Join the 3 tables and find the country-id wise count of employees and the avg salary. Which country has the maximum number of employees and which country has the maximum average salary?
4. Find the top 5 managers according to their salary.
5. Find the department name-wise percentage of employees working under each department. Which department is having the maximum percentage of employees?

**Results:**

Present your result in a business presentation and submit three files.

1. Excel file with all the results(each in a separate sheet)
2. SQL file with all the queries.
3. Presentation explanation with the insights derived from the analysis.

**Important Notes:**

1. Please note that since this is an EDA Case Study, your analysis should not be just limited to the questions asked above. You are required to take extra steps in order to find out more interesting insights from the data.
2. If you are generating plots, please make sure that you have also added the inference that you may derive from the plot. Any plot generated without your observation/inference will not be considered for evaluation.
3. Add details regarding any assumptions that you make while working on the case study. You are free to make assumptions or add more insights to the analysis by adding data from various sources but any such steps taken by you need to be mentioned in your solution file.