#### IFOOD ANALYSIS IN SQL

1. Which customers have spent the most across all product categories?

```
SELECT Customer_ID, Total_Spending FROM mkt_data

ORDER BY Total_Spending DESC

1468 9946
1150 9716
1548 9716
1684 9216
1033 9049
1033 9049
1033 9049
10724 8752
1067 8676
72 8632
120 8632
1240 8612
1269 8544
924 8550
1261 8550
1274 8550
1274 8612
1269 8544
924 8550
1261 8550
1274 8550
1274 8612
1269 8544
924 8550
1261 8550
1274 8550
1274 8612
1269 8544
924 8550
1261 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8550
1274 8568
```

The above returns a list of customers who have spent the most to least across all product categories. This returns all data for all customers but if we want to know the top 100 customers who have spent the most overall, the 'LIMIT 100' statement can be used.

2. What is the average income for each education level?

```
SELECT
                                                                                 education_level Avg_income
   CASE
                                                                                                  20306.2593
                                                                                Basic
       WHEN education_Unknown = 1 THEN 'Unkn
                                                                                                 47625,3333
                                                                                Unknown
      WHEN education_Basic = 1 THEN 'Basic'
                                                                                Graduation
                                                                                                  51984.0081
      WHEN education_Graduation = 1 THEN 'Graduation'
                                                                                Master 52629.5797
      WHEN education Master = 1 THEN 'Master'
                                                                                                  55220.5819
       WHEN education PhD = 1 THEN 'PhD'
   END AS education_level,
   AVG(Income) AS Avg incom
   mkt_data
GROUP BY
   education_level
ORDER BY Avg_income;
```

Since 1 represents yes and 0 represents no, the line of code filters the education status based on the 1s to calculate the average income. The 'GROUP BY' statement categorises all those data into five distinct categories as shown above and the 'ORDER BY' statement shows the average income in ascending order.

The result corresponds with what we would expect in a real-life scenario where higher education level implies higher income.

3. What percentage of customers accepted at least one campaign offer?



Here the query is a simple mathematical formula of counting how many customers accepted the campaign overall and then dividing it by the total customers to find the required percentage.

A 21% does not seem like a big number but that would ultimately depend on the marketing goals of the company.

Note: The COUNT statement for any column in this case would return the total amount of customers there are.

## 4. How does spending vary by marital status?

```
Marital_Status Avg_Spending
                                                                                   Married
                                                                                                   4373,1991
   WHEN marital_Divorced = 1 THEN 'Divorced'
                                                                                   Together 4514.8028
   WHEN marital_Married = 1 THEN 'Married'
                                                                                   Divorced 4552.4174
Single 4554.1971
   WHEN marital_Single = 1 THEN 'Single'
   WHEN marital_Together = 1 THEN 'Together'
                                                                                   Widow
                                                                                                   5376.8421
   WHEN marital Widow = 1 THEN 'Widow'
END AS Marital Status
AVG(Total_Spending) AS Avg_Spending
FROM mkt data
GROUP BY Marital Status
ORDER BY AVG(Total_Spending);
```

Similar to the 2<sup>nd</sup> question's query, marital status is returned based on average spending. People who are married seems to be spending the least (4k), together, divorced and single seems to be spending approximately the same (5k) and widows seems to be spending the most (5.38k).

#### 5. How is income distributed across different marital statuses?

```
SELECT
                                                                                 Marital_Status Avg_income
 CASE
                                                                                Single
                                                                                                51009,6226
    WHEN marital_Divorced = 1 THEN 'Divorced'
                                                                                Married 51350.5117
   WHEN marital_Married = 1 THEN 'Married'
                                                                                Together
                                                                                                51553.1989
   WHEN marital_Single = 1 THEN 'Single'
                                                                               Divorced 52465.1174
   WHEN marital_Together = 1 THEN 'Together
                                                                               Widow
                                                                                               56481.5526
   WHEN marital Widow = 1 THEN 'Widow'
END AS Marital Status.
AVG(Income) AS Avg income
FROM mkt_data
GROUP BY Marital_Status
ORDER BY AVG(Income);
```

Single and married people seem to be earning around the least (51k), people who are in a relationship (Together) and divorced earn a bit more (52k) and widows are earning the most (56k).

### 6. Income vs Spending based on marital status

```
SELECT
                                                                                    Marital_Status Avg_income Avg_Spending
 CASE
                                                                                    Single
                                                                                                    51.01 k
                                                                                                                  4.55 k
    WHEN marital_Divorced = 1 THEN 'Divorced'
                                                                                    Married 51.35 k 4.37 k
   WHEN marital_Married = 1 THEN 'Married'
                                                                                   Together 51.55 k 4.51 k
Divorced 52.47 k 4.55 k
   WHEN marital_Single = 1 THEN 'Single'
   WHEN marital_Together = 1 THEN 'Together
                                                                                   Widow
                                                                                                   56.48 k 5.38 k
   WHEN marital_Widow = 1 THEN 'Widow
END AS Marital_Status,
CONCAT(ROUND(AVG(Income)/1000, 2), 'k') AS Avg_income,
CONCAT(ROUND(AVG(Total_Spending)/1000,2),' k') AS Avg_Spending
FROM mkt_data
GROUP BY Marital Status
ORDER BY AVG(Income);
```

It seems that widows are earning and spending the most (56k vs 5.38k).

Single people earn the least but are second to widows in terms of spending (51k vs 4.55k).

Divorcees are the 2<sup>nd</sup> in terms of earnings and are spending approximately the same as singles (52k vs 4.55k).

Married people are making approximately the same as single people but are spending the least (51k vs 4.37k). Lastly people who are together are in the middle both in terms of earning and spending (52k vs 4.51k).

7. What is the average age of customers who have accepted at least one campaign offer?

```
SELECT AVG(Age) AS Avg_age

FROM mkt_data

WHERE Accptd_CampOverall >= 1;

Avg_age

> 51.7860
```

On average, people who are accepting campaign offers are 51 years old.

8. What is the proportion of customers in each age group?

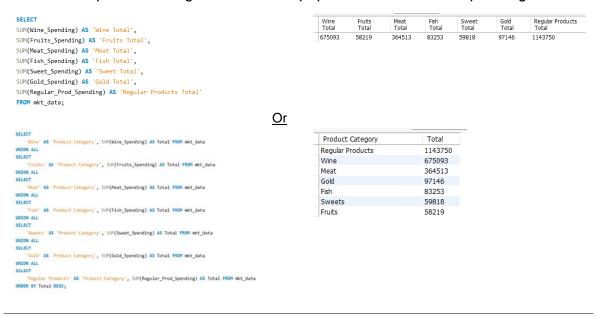
```
SELECT MAX(Age) FROM mkt_data; -- 80
                                                                                           Age_Group Count Percentage
 SELECT MIN(Age) FROM mkt_data; -- 24
                                                                                           Middle-aged 926 42.00%
Older Adult 748 33.92%
 SELECT AVG(Age) FROM mkt_data; -- 51
                                                                                                           312
                                                                                                                  14.15%
@ /* ALTER TABLE mkt_data
                                                                                           Senior
 ADD Age_Group VARCHAR(20) AFTER Age;
                                                                                           Young Adult 219 9.93%
 UPDATE mkt_data
 SET Age_Group =
     WHEN Age BETWEEN 24 AND 35 THEN 'Young Adult'
     WHEN Age BETWEEN 36 AND 50 THEN 'Middle-aged'
     WHEN Age BETWEEN 51 AND 65 THEN 'Older Adult'
     WHEN Age BETWEEN 66 AND 80 THEN 'Senior'
     ELSE 'Unknown'
  END;
SELECT
Age_Group, COUNT(Age_group) As Count,
CONCAT(ROUND((COUNT(Age_Group)/ (SELECT COUNT(*) FROM mkt_data))*100, 2), "%")
AS Percentage
 ORDER BY ((COUNT(Age_Group)/ (SELECT COUNT(*) FROM mkt_data))*100) DESC;
```

The commented code creates a new column 'Age\_Group' which breaks the ages in five categories:

- Young adult: customers aged between 24 and 35
- Middle-aged: customers aged between 36 and 50
- Older Adult: customers aged between 51 and 65
- Senior: customers aged between 66 and 80
- Unknown: customers who have no age assigned to them

The query then returns the total number of customers based on their age group and what percentage of the total customers they represent. The output shows that almost half the customers are middle aged (42%) while the rest of the customers are distributed between Older Adult, Senior and Young Adult with the latter representing the least customers at 9.93%.

9. Which product categories are most popular based on total spending?



Both queries return the sum of spending based on different products. However, the second query makes it easier to distinguish on which product category customers spent most, which is regular products, and the least, which is fruits.

10. How does having kids or teenagers affect the total amount spent by customers?

```
Kid_Status Teen_Status
                                                                                                                            TotalSpending
                                                                                             Kids
                                                                                                           Teenagers
                                                                                                                            663624
        WHEN Kidhome >= 1 THEN 'Kids'
        ELSE 'No Kids'
                                                                                             Kids
                                                                                                           No Teenagers 675912
     END AS Kid_Status,
                                                                                             No Kids
                                                                                                           Teenagers
                                                                                                                            3356608
                                                                                             No Kids No Teenagers 5231024
        MHEN Teenhome >= 1 THEN 'Teenagers
        ELSE 'No Teen
     END AS Teen_Status,
     SUM(Total_Spending) AS TotalSpending
 GROUP BY Kid_Status, Teen_Status
 ORDER BY SUM(Total_Spending);
                                                                                 Or
   IF(Kidhome >= 1, 'Kids', 'No Kids') AS Kid_Status
   IF(Teenhome >= 1, 'Teenagers', 'No Teenagers') AS Teen_Status,
   SUM(Total_Spending) AS TotalSpending
FROM mkt_data
GROUP BY Kid_Status, Teen_Status
ORDER BY SUM(Total Spending):
```

The case statement and if statement both return the same result which is that customers with kids and teenagers spend more that customers with no children.

11. Do higher-income customers use discounts more or less frequently?

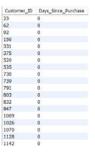
```
SHITH income_ranks AS (
SELECT Income,
NTILE(3) OVER (ORDER BY Income) AS rank_group, Discount_Purchases
FROM mkt_data
)
SELECT High Income: AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
NUMER rank_group = 3
UNION ALL
SELECT 'Avg Income' AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
NUMER rank_group = 2
UNION ALL
SELECT 'Low Income' AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
NUMER rank_group = 2
UNION ALL
SELECT 'Low Income' AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
```

The above query makes use of a subquery in which the NTILE(3) function equally divides the income into 3 categories: High Income, Avg Income and Low income. It is clear from the results above that high income earners make use of discounts the least.

Note: In order to change the number of categories, simply change the number 3 to any number and define the income status accordingly.

### 12. Who are the customers with the most recent purchases?

```
SELECT Customer_ID, Days_Since_Purchase FROM mkt_data
ORDER BY Days_Since_Purchase;
```



The above query simply returns a list of customers who have the most recent purchase history. Here however, due to there being a lack of data as to the exact time of purchase, customers are sorted by the smallest days since purchase then by the smallest ID. Customer with ID 23 is not necessarily the latest shopper.

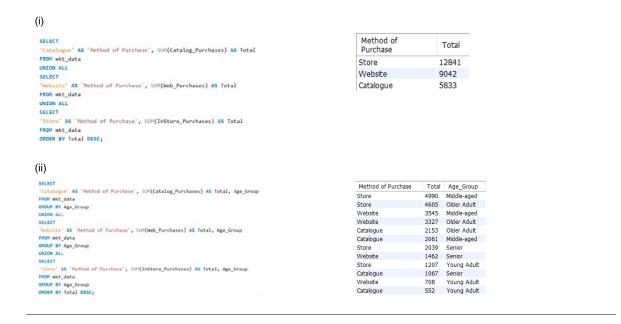
# 13. Which customers have visited the website the most in the last month?

```
SELECT Customer_ID,
Last_Month_Web_Visits AS '# of times customer visited website' FROM mkt_data
ORDER BY Last_Month_Web_Visits DESC;
```



In the last month, the customers having visited the website the most have IDs 10, 755 and 1226.

14. Which method of purchase (catalogue or web or store) is more popular across all customers/ different age groups?



It seems that Store followed by Website then Catalogue is the method that customers prefer to shop. Adding Age group to the table further shows that across all ages, the above is also true.

## 15. How does the success of marketing campaigns vary by age group?

```
Age_Group Accepted_Campaign
                                                                                                                                                                      Percentage_Acceptance
COUNT(CASE WHEN Acceptd_CampOverall >= 1 THEN 1 END) AS Accepted_Campaign,
                                                                                                                                                   239
                                                                                                                                                                     23.3974%
                                                                                                                   Young Adult 48
COUNT(CASE WHEN Acceptd CampOverall = 0 THEN 1 END) AS Declined Campaign,
CONCAT(((COUNT(CASE WHEN Acceptd_CampOverall >= 1 THEN 1 END)/COUNT(Acceptd_CampOverall))*100), '%')
                                                                                                                  Older Adult
                                                                                                                               163
                                                                                                                                                   585
                                                                                                                                                                     21.7914%
                                                                                                                  Middle-aged 174
AS Percentage_Acceptance
GROUP BY Age Group
ORDER BY ((COUNT(CASE WHEN Accetd_CampOverall >= 1 THEN 1 END)/COUNT(Accetd_CampOverall))*100) DESC;
```

Based on the count of customers who have accepted the campaign, Middle aged seems to have the highest count but this age group also has the highest decline count making its percentage the smallest. Seniors on the other hands have the second smallest acceptance count but have the highest acceptance percentage implying a more successful marketing campaign with seniors.

## 16. What is the proportion of customers who have made complaints in the last 2 years?



Less than 1% of customers complained during the last two years.