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
;
```

Customer_ID	Total_Spending
1468	9964
1160	9716
1548	9716
968	9216
1434	9216
1033	9048
1577	8976
1724	8752
1967	8676
72	8632
1420	8628
1240	8612
1269	8544
924	8520
1281	8520
627	8420
408	8392
344	8368
1928	8368

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
    CASE
        WHEN education_Unknown = 1 THEN 'Unknown'
        WHEN education_Basic = 1 THEN 'Basic'
        WHEN education_Graduation = 1 THEN 'Graduation'
        WHEN education_Master = 1 THEN 'Master'
        WHEN education_PhD = 1 THEN 'PhD'
    END AS education_level,
    AVG(Income) AS Avg_income
FROM
    mkt_data
GROUP BY
    education_level
ORDER BY Avg_income;
```

education_level	Avg_income
Basic	20306.2593
Unknown	47625.3333
Graduation	51984.0081
Master	52629.5797
PhD	55220.5819

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?

```
SELECT
CONCAT((COUNT(CASE WHEN Accptd_CampOverall >= 1 THEN 1 END)*100) / COUNT(Accptd_CampOverall),'%') AS Percentage
FROM mkt_data;
```

Percentage
20.7710%

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?

```
SELECT
  CASE
    WHEN marital_Divorced = 1 THEN 'Divorced'
    WHEN marital_Married = 1 THEN 'Married'
    WHEN marital_Single = 1 THEN 'Single'
    WHEN marital_Together = 1 THEN 'Together'
    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);
```

Marital_Status	Avg_Spending
Married	4373.1991
Together	4514.8028
Divorced	4552.4174
Single	4554.1971
Widow	5376.8421

Similar to the 2nd 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
  CASE
    WHEN marital_Divorced = 1 THEN 'Divorced'
    WHEN marital_Married = 1 THEN 'Married'
    WHEN marital_Single = 1 THEN 'Single'
    WHEN marital_Together = 1 THEN 'Together'
    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);
```

Marital_Status	Avg_income
Single	51009.6226
Married	51350.5117
Together	51553.1989
Divorced	52465.1174
Widow	56481.5526

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
  CASE
    WHEN marital_Divorced = 1 THEN 'Divorced'
    WHEN marital_Married = 1 THEN 'Married'
    WHEN marital_Single = 1 THEN 'Single'
    WHEN marital_Together = 1 THEN 'Together'
    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);
```

Marital_Status	Avg_income	Avg_Spending
Single	51.01 k	4.55 k
Married	51.35 k	4.37 k
Together	51.55 k	4.51 k
Divorced	52.47 k	4.55 k
Widow	56.48 k	5.38 k

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 2nd 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?

```sql
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?

```sql
SELECT MAX(Age) FROM mkt_data; -- 80
SELECT MIN(Age) FROM mkt_data; -- 24
SELECT AVG(Age) FROM mkt_data; -- 51

/* ALTER TABLE mkt_data
ADD Age_Group VARCHAR(20) AFTER Age;

UPDATE mkt_data
SET Age_Group =
  CASE
    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
FROM mkt_data
GROUP BY Age_Group
ORDER BY ((COUNT(Age_Group)/ (SELECT COUNT(*) FROM mkt_data))*100) DESC;
```

Age_Group	Count	Percentage
Middle-aged	926	42.00%
Older Adult	748	33.92%
Senior	312	14.15%
Young Adult	219	9.93%

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?

```
SELECT
SUM(Wine_Spending) AS 'Wine Total',
SUM(Fruits_Spending) AS 'Fruits Total',
SUM(Meat_Spending) AS 'Meat Total',
SUM(Fish_Spending) AS 'Fish Total',
SUM(Sweet_Spending) AS 'Sweet Total',
SUM(Gold_Spending) AS 'Gold Total',
SUM(Regular_Prod_Spending) AS 'Regular Products Total'
FROM mkt_data;
```

Wine Total	Fruits Total	Meat Total	Fish Total	Sweet Total	Gold Total	Regular Products Total
675093	58219	364513	83253	59818	97146	1143750

Or

```
SELECT
    'Wine' AS 'Product Category', SUM(Wine_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Fruits' AS 'Product Category', SUM(Fruits_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Meat' AS 'Product Category', SUM(Meat_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Fish' AS 'Product Category', SUM(Fish_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Sweets' AS 'Product Category', SUM(Sweet_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Gold' AS 'Product Category', SUM(Gold_Spending) AS Total FROM mkt_data
UNION ALL
SELECT
    'Regular Products' AS 'Product Category', SUM(Regular_Prod_Spending) AS Total FROM mkt_data
ORDER BY Total DESC;
```

Product Category	Total
Regular Products	1143750
Wine	675093
Meat	364513
Gold	97146
Fish	83253
Sweets	59818
Fruits	58219

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?

```
SELECT
    CASE
        WHEN Kidhome >= 1 THEN 'Kids'
        ELSE 'No Kids'
    END AS Kid_Status,
    CASE
        WHEN Teenhome >= 1 THEN 'Teenagers'
        ELSE 'No Teenagers'
    END AS Teen_Status,
    SUM(Total_Spending) AS TotalSpending
FROM mkt_data
GROUP BY Kid_Status, Teen_Status
ORDER BY SUM(Total_Spending);
```

Kid_Status	Teen_Status	TotalSpending
Kids	Teenagers	663624
Kids	No Teenagers	675912
No Kids	Teenagers	3356608
No Kids	No Teenagers	5231024

Or

```
SELECT
    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?

```sql
WITH 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
WHERE rank_group = 3
UNION ALL
SELECT
    'Avg Income' AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
WHERE rank_group = 2
UNION ALL
SELECT
    'Low Income' AS Income_Rank, SUM(Discount_Purchases) AS Total_Discount_Purchases
FROM income_ranks
WHERE rank_group = 1;
```

Income_Rank	Total_Discount_Purchases
High Income	1235
Avg Income	2302
Low Income	1575

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?

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

Customer_ID	Days_Since_Purchase
23	0
62	0
92	0
150	0
331	0
375	0
520	0
535	0
730	0
739	0
791	0
803	0
832	0
847	0
1009	0
1026	0
1070	0
1128	0
1142	0

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?

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

Customer_ID	# of times customer visited website
10	20
755	20
1226	20
1023	19
1821	19
1308	17
962	14
1500	14
2182	13
45	10
589	10
1843	10
9	9
44	9
103	9
109	9
174	9
194	9
220	9

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?

(i)

```
SELECT
'Catalogue' AS `Method of Purchase`, SUM(Catalog_Purchases) AS Total
FROM mkt_data
UNION ALL
SELECT
'Website' AS `Method of Purchase`, SUM(Web_Purchases) AS Total
FROM mkt_data
UNION ALL
SELECT
'Store' AS `Method of Purchase`, SUM(InStore_Purchases) AS Total
FROM mkt_data
ORDER BY Total DESC;
```

Method of Purchase	Total
Store	12841
Website	9042
Catalogue	5833

(ii)

```
SELECT
'Catalogue' AS `Method of Purchase`, SUM(Catalog_Purchases) AS Total, Age_Group
FROM mkt_data
GROUP BY Age_Group
UNION ALL
SELECT
'Website' AS `Method of Purchase`, SUM(Web_Purchases) AS Total, Age_Group
FROM mkt_data
GROUP BY Age_Group
UNION ALL
SELECT
'Store' AS `Method of Purchase`, SUM(InStore_Purchases) AS Total, Age_Group
FROM mkt_data
GROUP BY Age_Group
ORDER BY Total DESC;
```

Method of Purchase	Total	Age_Group
Store	4990	Middle-aged
Store	4605	Older Adult
Website	3545	Middle-aged
Website	3327	Older Adult
Catalogue	2153	Older Adult
Catalogue	2061	Middle-aged
Store	2039	Senior
Website	1462	Senior
Store	1207	Young Adult
Catalogue	1067	Senior
Website	708	Young Adult
Catalogue	552	Young Adult

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?

```
SELECT Age_Group,
COUNT(CASE WHEN Accptd_CampOverall >= 1 THEN 1 END) AS Accepted_Campaign,
COUNT(CASE WHEN Accptd_CampOverall = 0 THEN 1 END) AS Declined_Campaign,
CONCAT(((COUNT(CASE WHEN Accptd_CampOverall >= 1 THEN 1 END)/COUNT(Accptd_CampOverall))*100),'%')
AS Percentage_Acceptance
FROM mkt_data
GROUP BY Age_Group
ORDER BY ((COUNT(CASE WHEN Accptd_CampOverall >= 1 THEN 1 END)/COUNT(Accptd_CampOverall))*100) DESC;
```

Age_Group	Accepted_Campaign	Declined_Campaign	Percentage_Acceptance
Senior	73	239	23.3974%
Young Adult	48	171	21.9178%
Older Adult	163	585	21.7914%
Middle-aged	174	752	18.7905%

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?

```
SELECT
CONCAT(((COUNT(CASE WHEN Complain =1 THEN 1 END)/COUNT(Complain))*100),'%')
AS Percentage_of_Complains
FROM mkt_data;
```

Percentage_of_Complains
0.9070%

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