

Codebasics Resume Project Challenge#6

Providing Insights to the Marketing Team in Food and Beverage Industry



<https://codebasics.io/challenge/codebasics-resume-project-challenge>

By May Thu Han

July 2023

Introduction

CodeX is a German beverage company that is aiming to make its mark in the Indian market. A few months ago, they launched their energy drink in 10 cities in India.

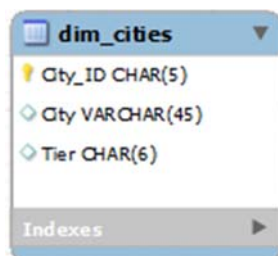
Their Marketing team is responsible for increasing brand awareness, market share, and product development. They conducted a survey in those 10 cities and received results from **10k** respondents.

Problem Statement

As a marketing data analyst, the task is to convert these survey results into meaningful insights that the team can use to drive actions.

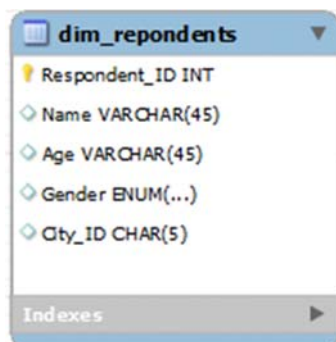
Tables

1. dim_cities table with three columns



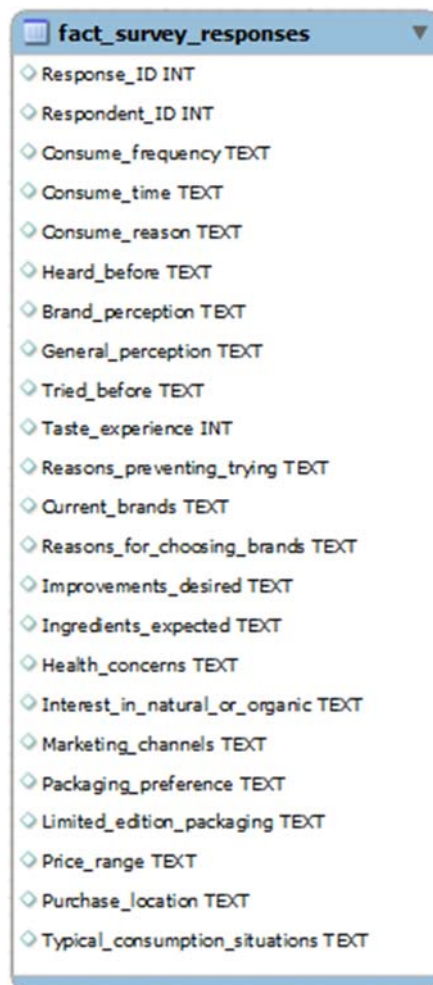
dim_cities	
City_ID	CHAR(5)
City	VARCHAR(45)
Tier	CHAR(6)
Indexes	

2. dim_repondents table with five columns



dim_repondents	
Respondent_ID	INT
Name	VARCHAR(45)
Age	VARCHAR(45)
Gender	ENUM(...)
City_ID	CHAR(5)
Indexes	

3. fact_survey_responses table with 23 columns



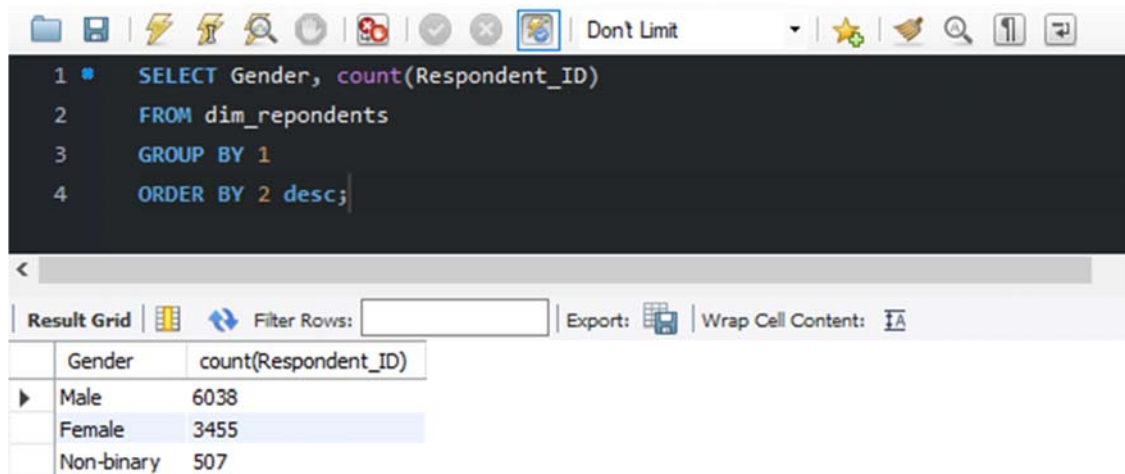
The screenshot shows a database interface with a table named 'fact_survey_responses'. The table has 23 columns, each with a diamond icon to its left. The columns are listed as follows:

Column Name	Data Type
Response_ID	INT
Respondent_ID	INT
Consume_frequency	TEXT
Consume_time	TEXT
Consume_reason	TEXT
Heard_before	TEXT
Brand_perception	TEXT
General_perception	TEXT
Tried_before	TEXT
Taste_experience	INT
Reasons_preventing_trying	TEXT
Current_brands	TEXT
Reasons_for_choosing_brands	TEXT
Improvements_desired	TEXT
Ingredients_expected	TEXT
Health_concerns	TEXT
Interest_in_natural_or_organic	TEXT
Marketing_channels	TEXT
Packaging_preference	TEXT
Limited_edition_packaging	TEXT
Price_range	TEXT
Purchase_location	TEXT
Typical_consumption_situations	TEXT

Questions along with Insights

Demographic Insights

1. Who prefers energy drinks more? (Male/Female/Non-binary)

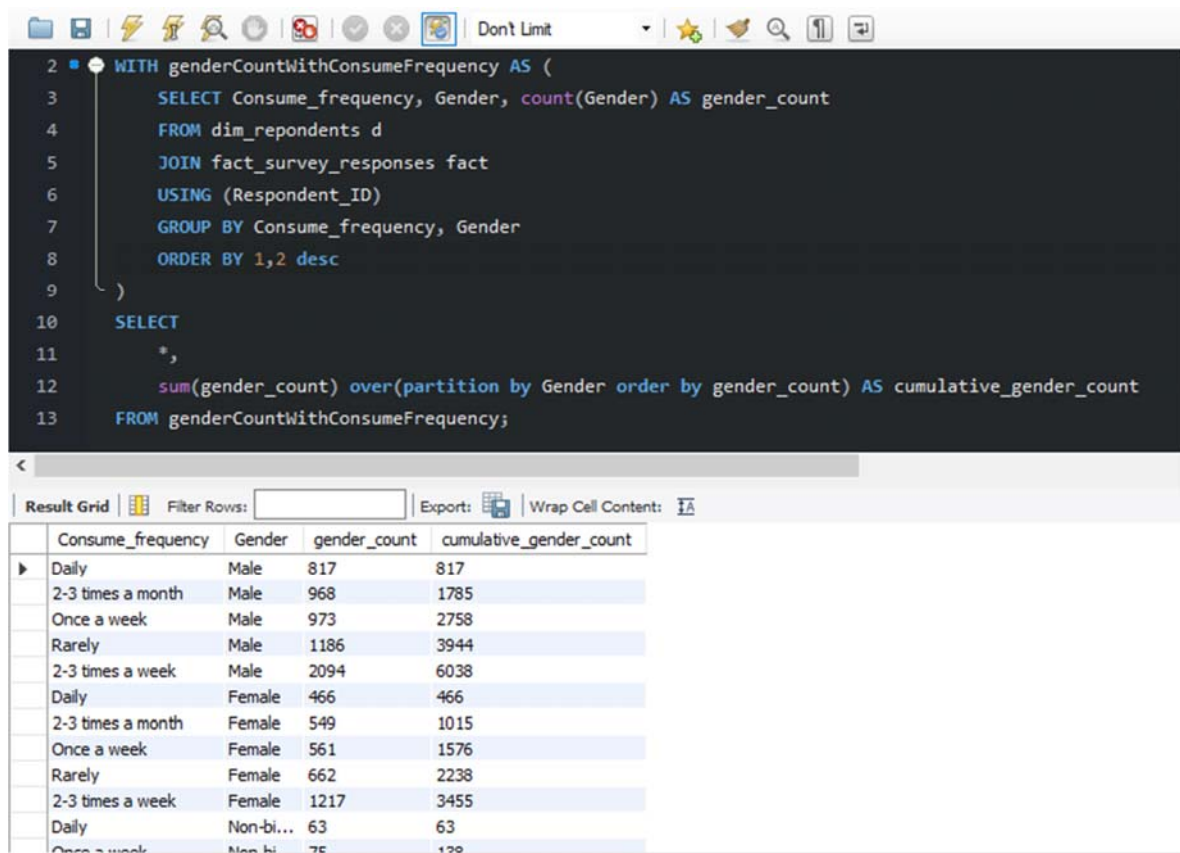


```
1 * SELECT Gender, count(Respondent_ID)
2 FROM dim_repondents
3 GROUP BY 1
4 ORDER BY 2 desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Gender	count(Respondent_ID)
▶	Male	6038
	Female	3455
	Non-binary	507

Doing a more detailed analysis by including consumption frequency by Gender

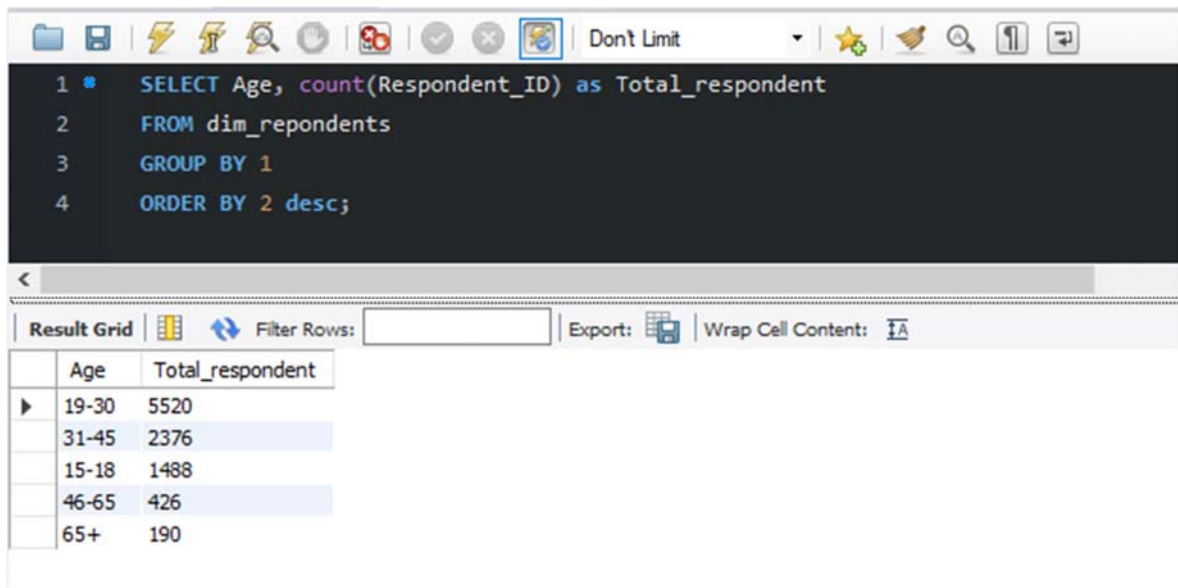


```
2 * WITH genderCountWithConsumeFrequency AS (
3     SELECT Consume_frequency, Gender, count(Gender) AS gender_count
4     FROM dim_repondents d
5     JOIN fact_survey_responses fact
6     USING (Respondent_ID)
7     GROUP BY Consume_frequency, Gender
8     ORDER BY 1,2 desc
9 )
10 SELECT
11     *,
12     sum(gender_count) over(partition by Gender order by gender_count) AS cumulative_gender_count
13 FROM genderCountWithConsumeFrequency;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Consume_frequency	Gender	gender_count	cumulative_gender_count
▶	Daily	Male	817	817
	2-3 times a month	Male	968	1785
	Once a week	Male	973	2758
	Rarely	Male	1186	3944
	2-3 times a week	Male	2094	6038
	Daily	Female	466	466
	2-3 times a month	Female	549	1015
	Once a week	Female	561	1576
	Rarely	Female	662	2238
	2-3 times a week	Female	1217	3455
	Daily	Non-bi...	63	63
	Once a week	Non-bi...	75	138

2. Which age group prefers energy drinks more?

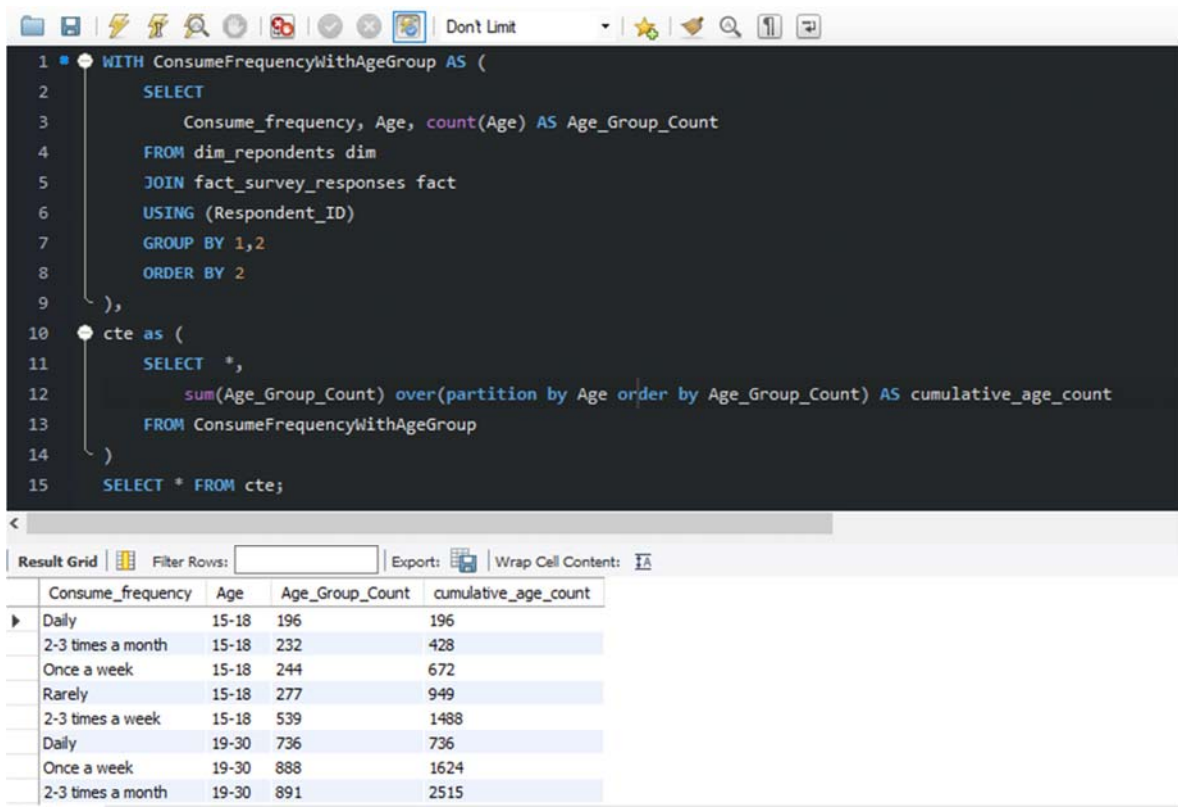


```
1 SELECT Age, count(Respondent_ID) as Total_respondent
2 FROM dim_repondents
3 GROUP BY 1
4 ORDER BY 2 desc;
```

Result Grid

	Age	Total_respondent
▶	19-30	5520
	31-45	2376
	15-18	1488
	46-65	426
	65+	190

By doing more analysis and deciding on consumption frequency,

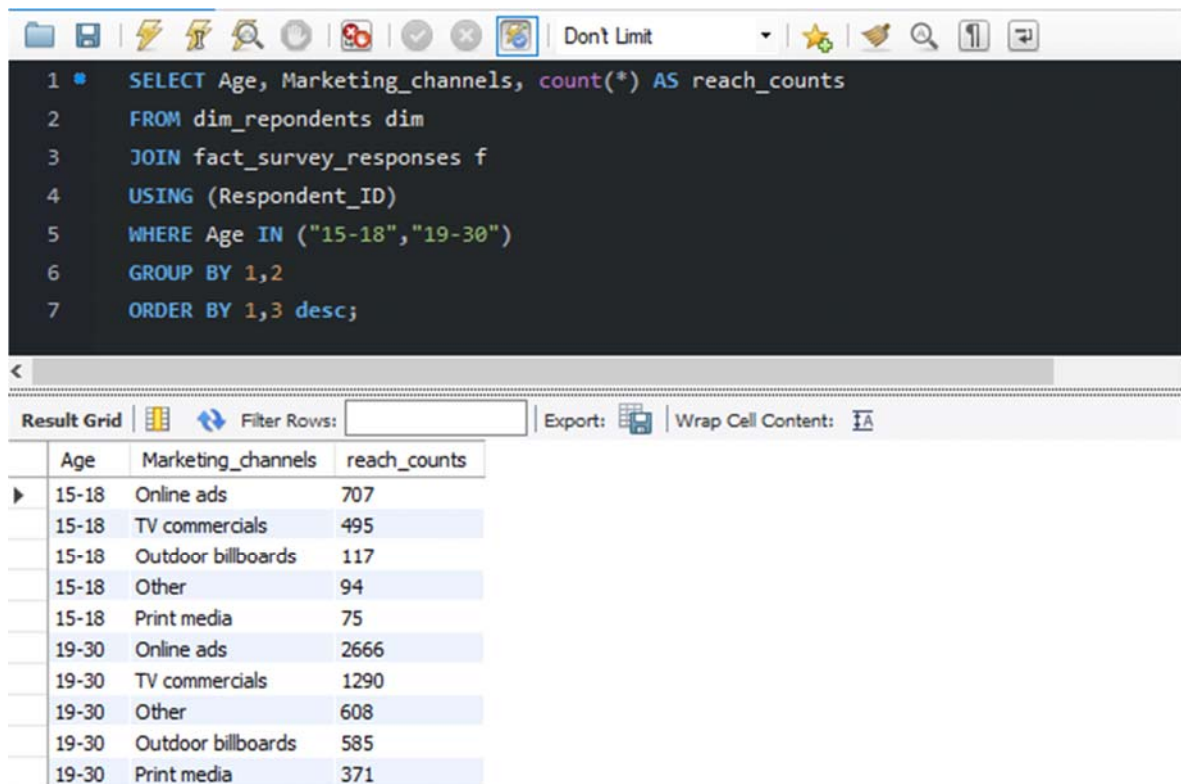


```
1 WITH ConsumeFrequencyWithAgeGroup AS (
2     SELECT
3         Consume_frequency, Age, count(Age) AS Age_Group_Count
4     FROM dim_repondents dim
5     JOIN fact_survey_responses fact
6     USING (Respondent_ID)
7     GROUP BY 1,2
8     ORDER BY 2
9 ),
10 cte as (
11     SELECT *,
12         sum(Age_Group_Count) over(partition by Age order by Age_Group_Count) AS cumulative_age_count
13     FROM ConsumeFrequencyWithAgeGroup
14 )
15 SELECT * FROM cte;
```

Result Grid

	Consume_frequency	Age	Age_Group_Count	cumulative_age_count
▶	Daily	15-18	196	196
	2-3 times a month	15-18	232	428
	Once a week	15-18	244	672
	Rarely	15-18	277	949
	2-3 times a week	15-18	539	1488
	Daily	19-30	736	736
	Once a week	19-30	888	1624
	2-3 times a month	19-30	891	2515

3. Which type of marketing reaches the most Youth (15-30)?



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

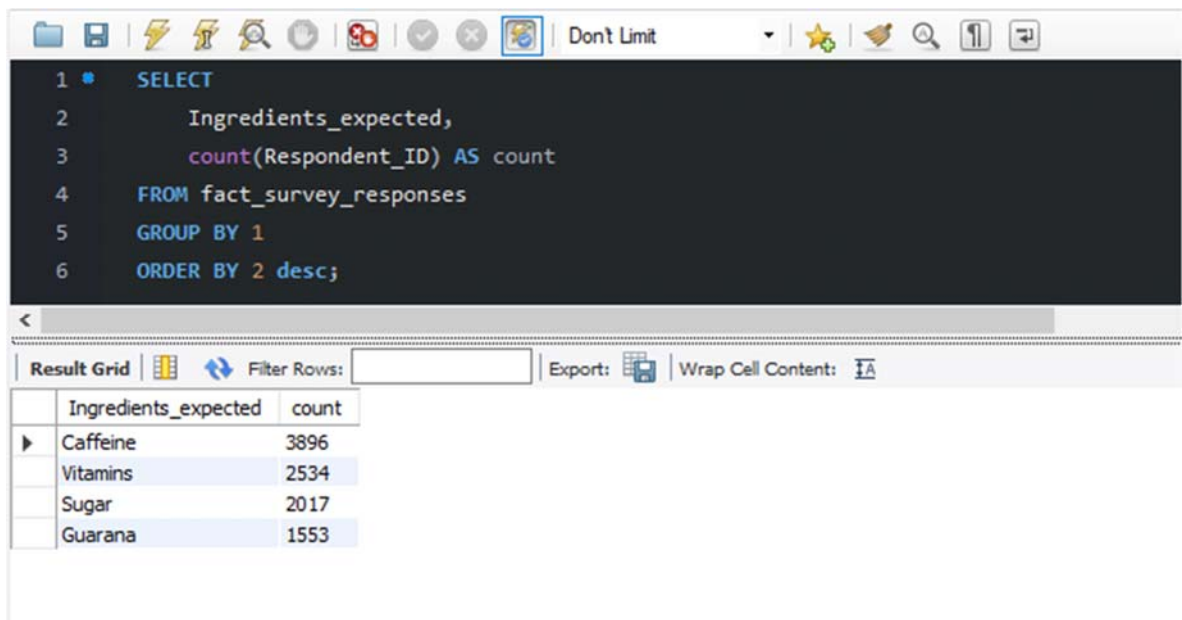
```
1 • SELECT Age, Marketing_channels, count(*) AS reach_counts
2 FROM dim_repondents dim
3 JOIN fact_survey_responses f
4 USING (Respondent_ID)
5 WHERE Age IN ("15-18","19-30")
6 GROUP BY 1,2
7 ORDER BY 1,3 desc;
```

Below the query editor is a "Result Grid" with the following data:

	Age	Marketing_channels	reach_counts
▶	15-18	Online ads	707
	15-18	TV commercials	495
	15-18	Outdoor billboards	117
	15-18	Other	94
	15-18	Print media	75
	19-30	Online ads	2666
	19-30	TV commercials	1290
	19-30	Other	608
	19-30	Outdoor billboards	585
	19-30	Print media	371

Consumer Preferences

1. What are the preferred ingredients of energy drinks among respondents?



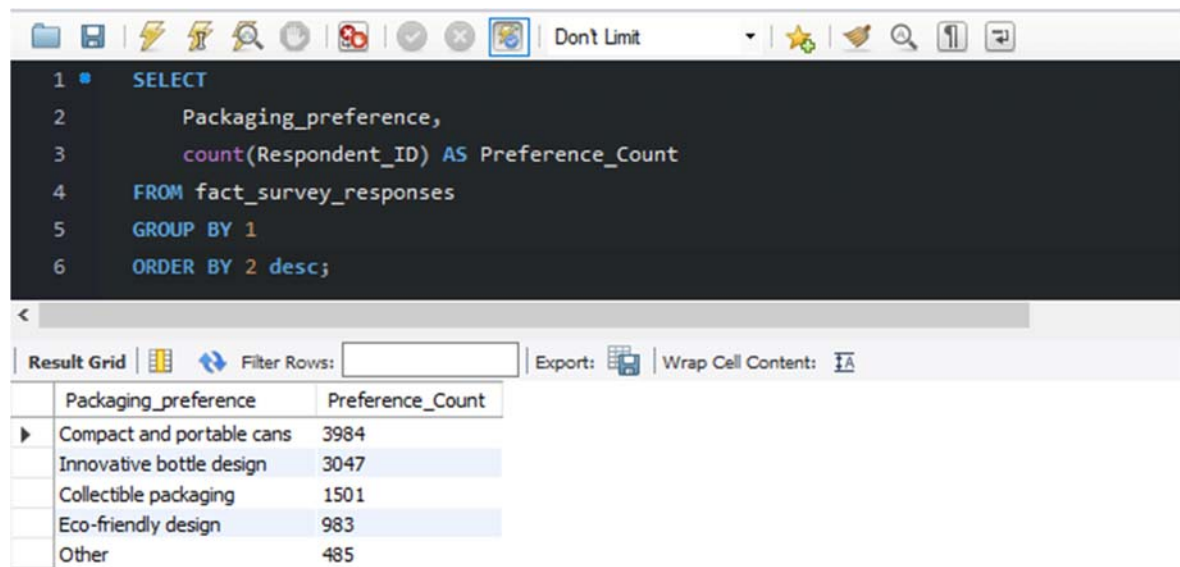
The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
1 • SELECT
2     Ingredients_expected,
3     count(Respondent_ID) AS count
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Below the query editor is a "Result Grid" with the following data:

	Ingredients_expected	count
▶	Caffeine	3896
	Vitamins	2534
	Sugar	2017
	Guarana	1553

2. What packaging preferences do respondents have for energy drinks?

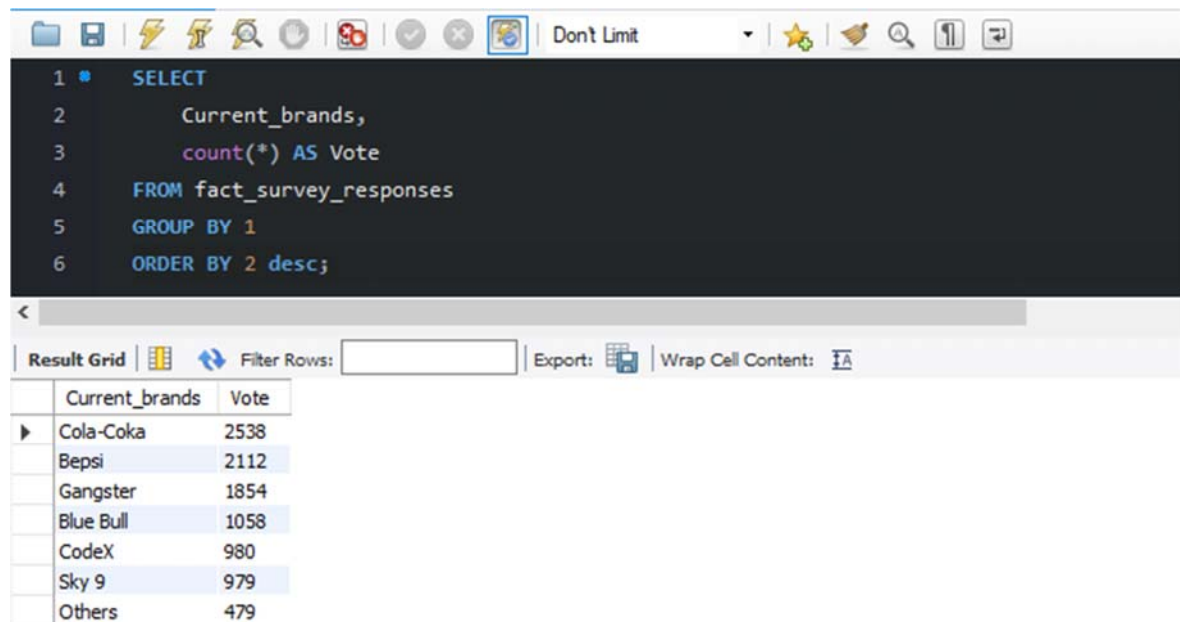


```
1 SELECT
2     Packaging_preference,
3     count(Respondent_ID) AS Preference_Count
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Packaging_preference	Preference_Count
Compact and portable cans	3984
Innovative bottle design	3047
Collectible packaging	1501
Eco-friendly design	983
Other	485

Competition Analysis

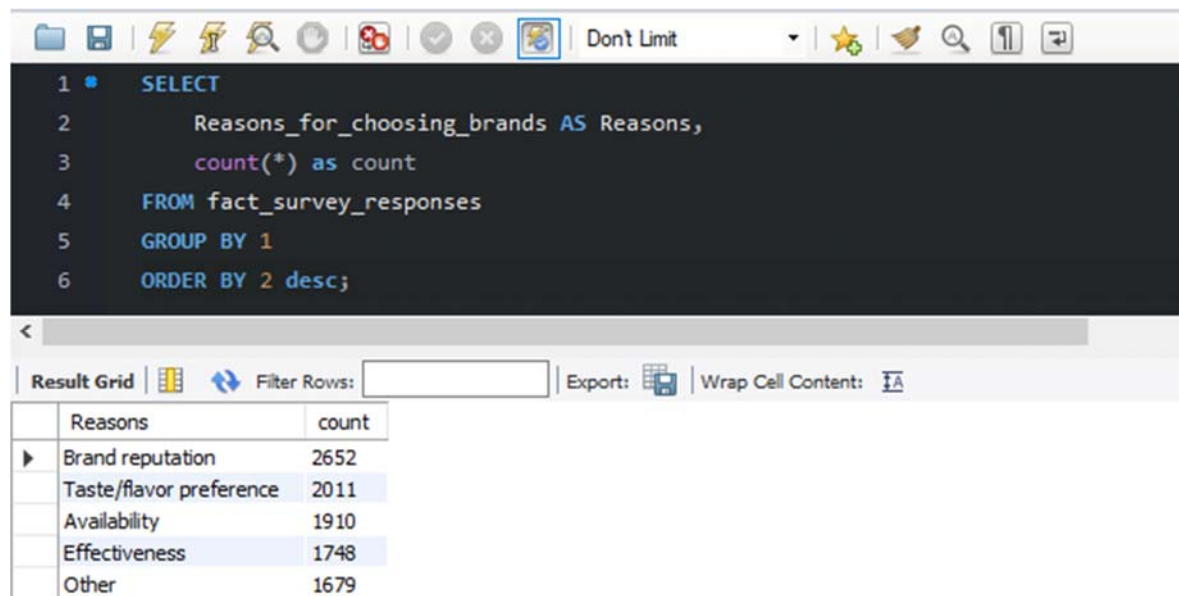
1. Who are the current market leaders?



```
1 SELECT
2     Current_brands,
3     count(*) AS Vote
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Current_brands	Vote
Cola-Coka	2538
Bepsi	2112
Gangster	1854
Blue Bull	1058
CodeX	980
Sky 9	979
Others	479

2. What are the primary reasons consumers prefer those brands over ours?

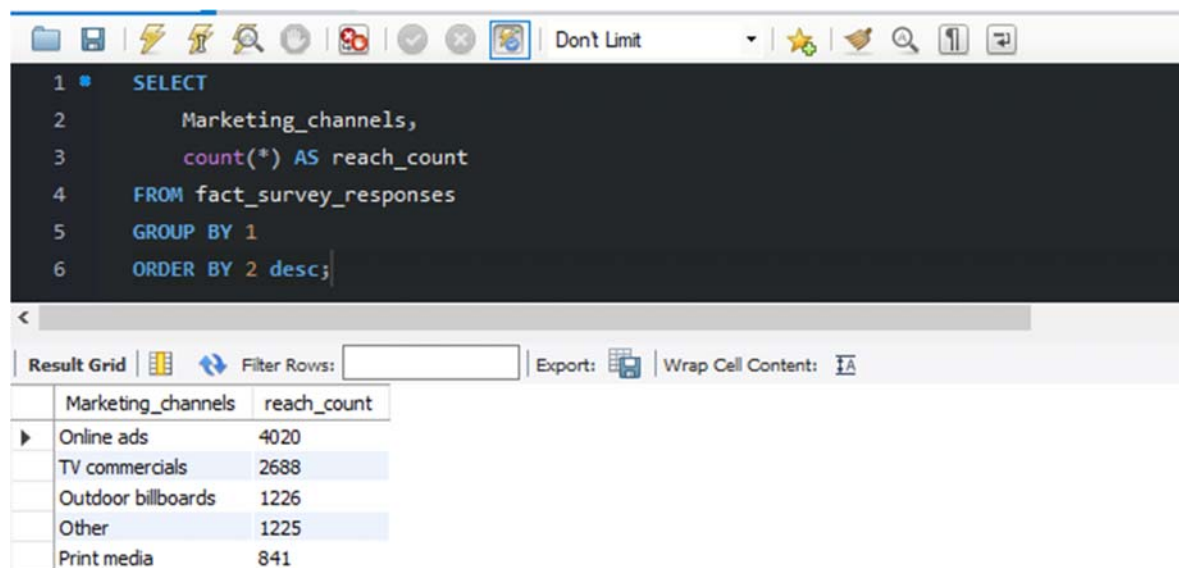


```
1 SELECT
2     Reasons_for_choosing_brands AS Reasons,
3     count(*) as count
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Reasons	count
Brand reputation	2652
Taste/flavor preference	2011
Availability	1910
Effectiveness	1748
Other	1679

Marketing Channels and Brand Awareness

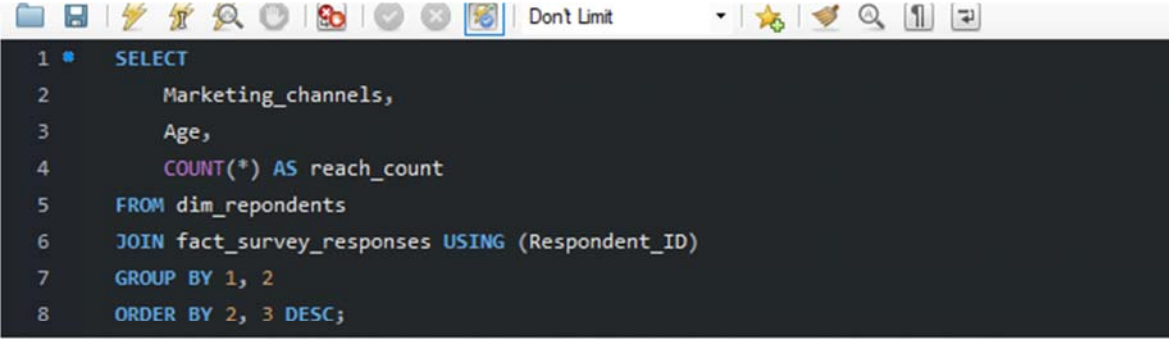
1. Which marketing channel can be used to reach more customers?



```
1 SELECT
2     Marketing_channels,
3     count(*) AS reach_count
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Marketing_channels	reach_count
Online ads	4020
TV commercials	2688
Outdoor billboards	1226
Other	1225
Print media	841

Doing detailed analysis with Age Groups,



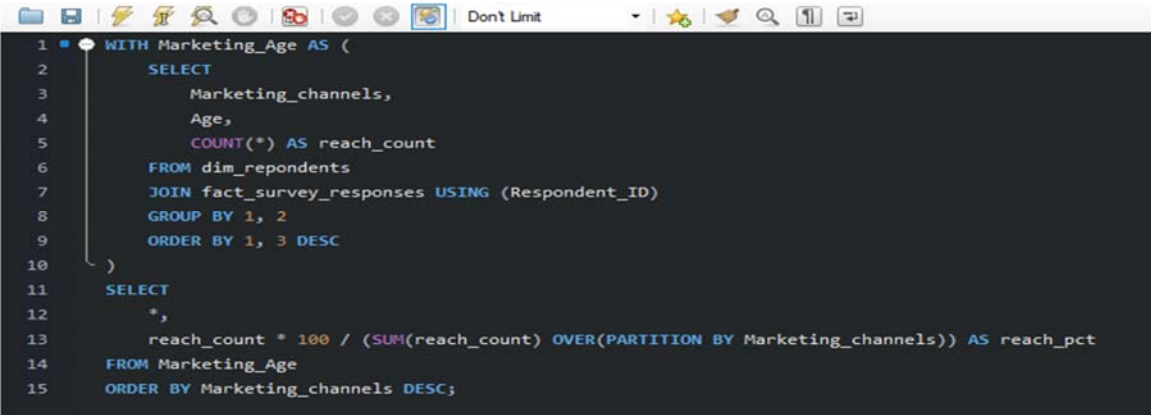
```
1 • SELECT
2     Marketing_channels,
3     Age,
4     COUNT(*) AS reach_count
5 FROM dim_repondents
6 JOIN fact_survey_responses USING (Respondent_ID)
7 GROUP BY 1, 2
8 ORDER BY 2, 3 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Marketing_channels	Age	reach_count
▶	Online ads	15-18	707
	TV commercials	15-18	495
	Outdoor billboards	15-18	117
	Other	15-18	94
	Print media	15-18	75
	Online ads	19-30	2666
	TV commercials	19-30	1290
	Other	19-30	608
	Outdoor billboards	19-30	585
	Print media	19-30	371
	TV commercials	31-45	737
	Online ads	31-45	490
	Outdoor billboards	31-45	431
	Other	31-45	408
	Print media	31-45	310
	TV commercials	46-65	117
	Online ads	46-65	109

2. How effective are different marketing strategies and channels in reaching our customers?

Each channel analysis with Age Group,



```
1 • WITH Marketing_Age AS (
2     SELECT
3         Marketing_channels,
4         Age,
5         COUNT(*) AS reach_count
6     FROM dim_repondents
7     JOIN fact_survey_responses USING (Respondent_ID)
8     GROUP BY 1, 2
9     ORDER BY 1, 3 DESC
10 )
11 SELECT
12     *,
13     reach_count * 100 / (SUM(reach_count) OVER(PARTITION BY Marketing_channels)) AS reach_pct
14 FROM Marketing_Age
15 ORDER BY Marketing_channels DESC;
```

Result Grid Filter Rows: Export: Wrap Cell Content: I A				
	Marketing_channels	Age	reach_count	reach_pct
▶	TV commercials	19-30	1290	47.9911
	TV commercials	31-45	737	27.4182
	TV commercials	15-18	495	18.4152
	TV commercials	46-65	117	4.3527
	TV commercials	65+	49	1.8229
	Print media	19-30	371	44.1141
	Print media	31-45	310	36.8609
	Print media	15-18	75	8.9180
	Print media	46-65	57	6.7776
	Print media	65+	28	3.3294
	Outdoor billboards	19-30	585	47.7162
	Outdoor billboards	31-45	431	35.1550
	Outdoor billboards	15-18	117	9.5432
	Outdoor billboards	46-65	65	5.3018
	Outdoor billboards	65+	28	2.2838
	Other	19-30	608	49.6327
	Other	31-45	408	33.3061
	Other	15-18	94	7.6735
	Other	46-65	78	6.3673
	Other	65+	37	3.0204
	Online ads	19-30	2666	66.3184

All channels analyzed with Age Group,

```

1  WITH cte4 AS (
2      SELECT
3          Marketing_channels,
4          Age,
5          COUNT(*) AS reach_count
6      FROM dim_repondents
7      JOIN fact_survey_responses USING (Respondent_ID)
8      GROUP BY 1, 2
9      ORDER BY 2, 3 DESC
10 )
11 cte5 AS (
12     SELECT
13         *,
14         SUM(reach_count) OVER(PARTITION BY Age ORDER BY reach_count) AS cumulative_reach_count
15     FROM cte4
16 )
17 SELECT
18     *,
19     ROUND((cumulative_reach_count * 100 / (SELECT COUNT(*) FROM fact_survey_responses)), 1) AS percent
20 FROM cte5
21 WHERE cumulative_reach_count IN (1488, 5520, 2376, 426, 190);

```

Result Grid Filter Rows: Export: Wrap Cell Content: IA				
	Marketing_channels	Age	reach_count	reach_pct
▶	TV commercials	19-30	1290	47.9911
	TV commercials	31-45	737	27.4182
	TV commercials	15-18	495	18.4152
	TV commercials	46-65	117	4.3527
	TV commercials	65+	49	1.8229
	Print media	19-30	371	44.1141
	Print media	31-45	310	36.8609
	Print media	15-18	75	8.9180
	Print media	46-65	57	6.7776
	Print media	65+	28	3.3294
	Outdoor billboards	19-30	585	47.7162
	Outdoor billboards	31-45	431	35.1550
	Outdoor billboards	15-18	117	9.5432
	Outdoor billboards	46-65	65	5.3018
	Outdoor billboards	65+	28	2.2838
	Other	19-30	608	49.6327
	Other	31-45	408	33.3061
	Other	15-18	94	7.6735
	Other	46-65	78	6.3673
	Other	65+	37	3.0204
	Online ads	19-30	2666	66.3184

Brand Penetration

1. What do people think about our brand? (Overall rating)

Result Grid Filter Rows: Export: Wrap Cell Content: IA			
	Brand_perception	vote_count	percent
▶	Neutral	5974	59.7400
	Positive	2257	22.5700
	Negative	1769	17.6900

Doing more analysis on

How many percent of the consumers have tried our brand or not?

```
190 SELECT
191     Tier,
192     Tried_before,
193     count(Tried_before) as Count,
194     round((count(Tried_before) * 100 / (SELECT count(*) FROM fact_survey_responses)),2) as percent,
195     dense_rank() over(partition by Tier order by count(Tried_before) desc) as Tried_Rank
196 FROM dim_cities
197 JOIN dim_repondents
198 USING (City_ID)
199 JOIN fact_survey_responses
200 USING (Respondent_ID)
201 GROUP BY 2,1;
```

Result Grid

Tier	Tried_before	Count	percent	Tried_Rank
Tier 1	Yes	4115	41.15	1
Tier 1	No	3422	34.22	2
Tier 2	No	1697	16.97	1
Tier 2	Yes	766	7.66	2

51.19% haven't tried our brand! Let's see why?

```
1 With Tier_Reasons as (
2     SELECT
3         Tier,
4         Reasons_preventing_trying,
5         count(Reasons_preventing_trying) AS Vote,
6         dense_rank() over(partition by Tier order by count(Reasons_preventing_trying) desc ) AS Vote_Count
7     FROM fact_survey_responses
8     JOIN dim_repondents
9     USING (Respondent_ID)
10    JOIN dim_cities
11    USING (City_ID)
12    WHERE Tried_before="No"
13    GROUP BY 1,2
14 )
15 SELECT * FROM Tier_Reasons
16 WHERE Vote_Count <4;
```

Result Grid

Tier	Reasons_preventing_trying	Vote	Vote_Count
Tier 1	Health concerns	867	1
Tier 1	Not interested in energy drinks	800	2
Tier 1	Not available locally	725	3
Tier 2	Not available locally	440	1
Tier 2	Not interested in energy drinks	379	2
Tier 2	Health concerns	330	3

but still, why does the brand is having more neutral votes from consume

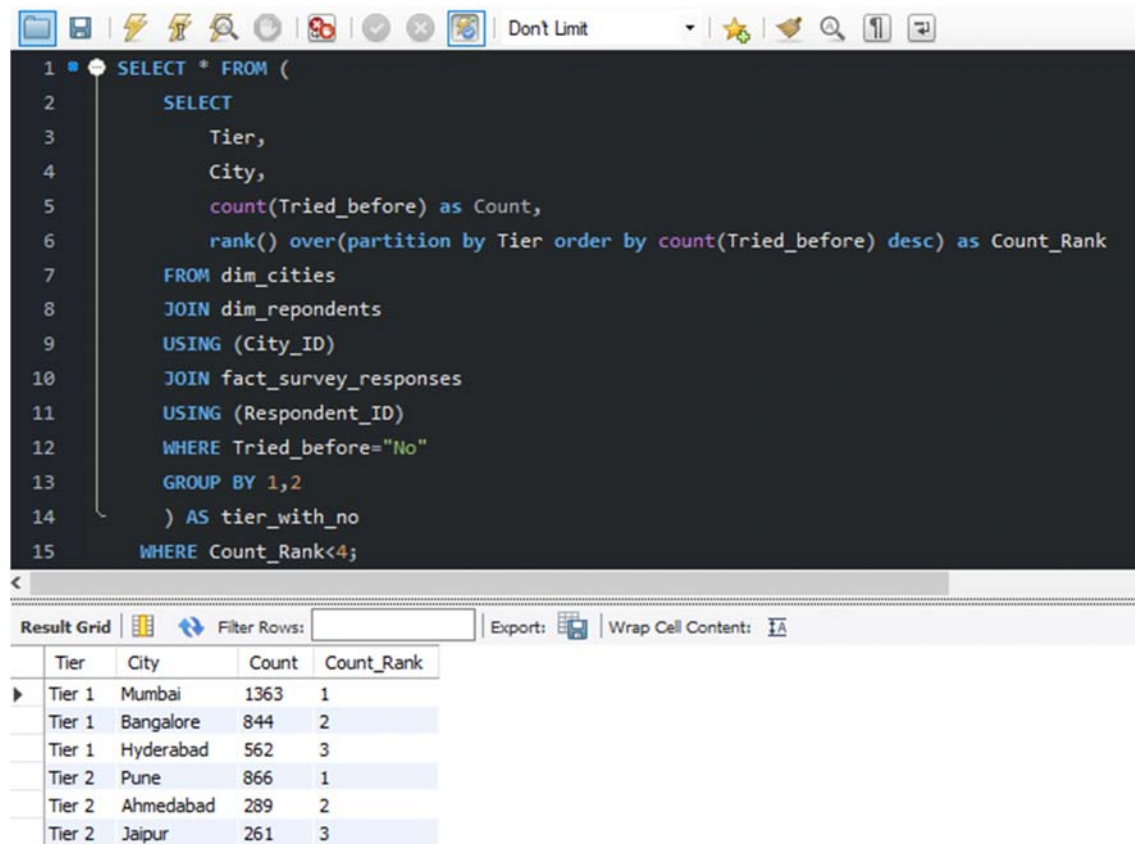
```
1 SELECT * FROM (
2     SELECT
3         Tier,
4         Taste_experience,
5         count(Taste_experience) as Vote,
6         dense_rank() over(partition by Tier order by count(Taste_experience) desc ) as Vote_Rank
7     FROM fact_survey_responses
8     JOIN dim_repondents
9     USING (Respondent_ID)
10    JOIN dim_cities
11    USING (City_ID)
12    WHERE Tried_before="Yes"
13    GROUP BY 1,2
14 ) AS YES_Reasons
15 WHERE Vote_Rank <4;
```

Result Grid

	Tier	Taste_experience	Vote	Vote_Rank
▶	Tier 1	3	1252	1
	Tier 1	4	1029	2
	Tier 1	5	780	3
	Tier 2	3	222	1
	Tier 2	4	190	2
	Tier 2	5	154	3

2. Which cities do we need to focus more on?

Cities with the highest “Not tried before”



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
1 SELECT * FROM (  
2     SELECT  
3         Tier,  
4         City,  
5         count(Tried_before) as Count,  
6         rank() over(partition by Tier order by count(Tried_before) desc) as Count_Rank  
7     FROM dim_cities  
8     JOIN dim_repondents  
9     USING (City_ID)  
10    JOIN fact_survey_responses  
11    USING (Respondent_ID)  
12    WHERE Tried_before="No"  
13    GROUP BY 1,2  
14 ) AS tier_with_no  
15 WHERE Count_Rank<4;
```

Below the query editor is a "Result Grid" with the following data:

	Tier	City	Count	Count_Rank
▶	Tier 1	Mumbai	1363	1
	Tier 1	Bangalore	844	2
	Tier 1	Hyderabad	562	3
	Tier 2	Pune	866	1
	Tier 2	Ahmedabad	289	2
	Tier 2	Jaipur	261	3

Cities with the Lowest “tried before”

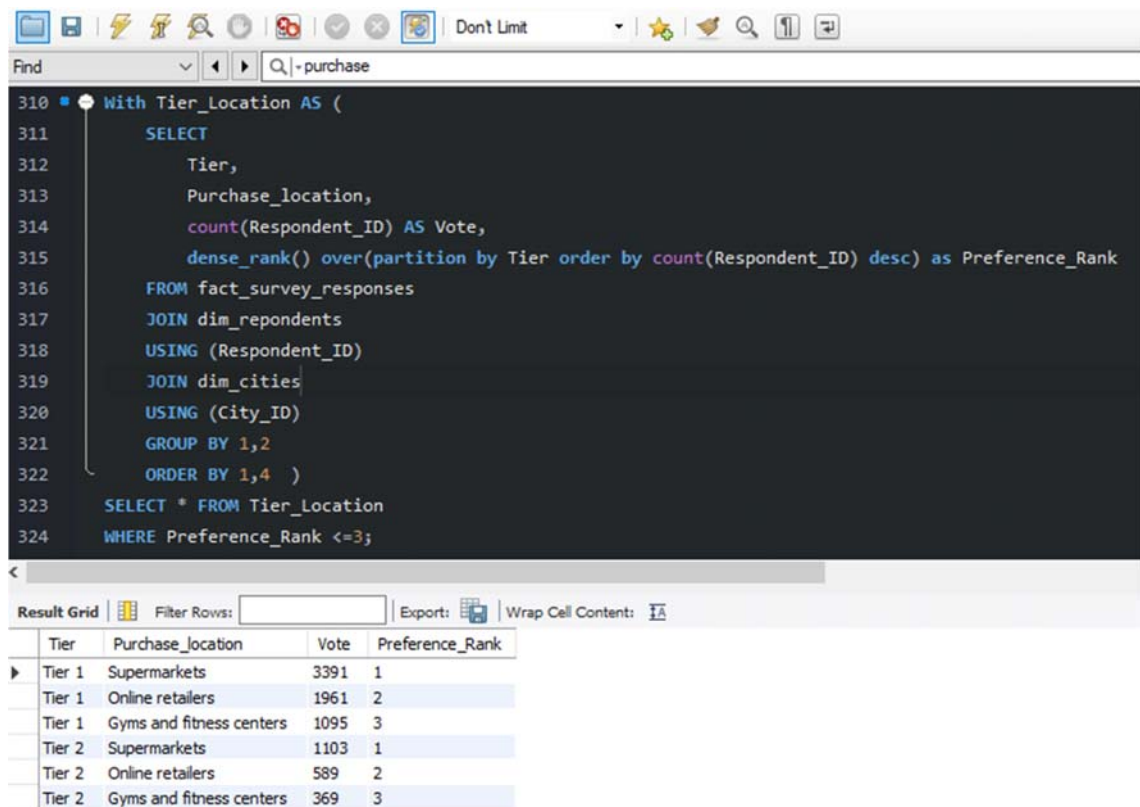
```
1 SELECT * FROM (
2     SELECT
3         Tier,
4         City,
5         count(Tried_before) as Count,
6         rank() over(partition by Tier order by count(Tried_before)) as Count_Rank
7     FROM dim_cities
8     JOIN dim_repondents
9     USING (City_ID)
10    JOIN fact_survey_responses
11    USING (Respondent_ID)
12    WHERE Tried_before="Yes"
13    GROUP BY 1,2
14 ) AS tier_with_yes
15 WHERE Count_Rank<4;
```

Result Grid

	Tier	City	Count	Count_Rank
▶	Tier 1	Delhi	46	1
	Tier 1	Mumbai	147	2
	Tier 1	Chennai	667	3
	Tier 2	Pune	40	1
	Tier 2	Lucknow	58	2
	Tier 2	Jaipur	99	3

Purchase Behavior

1. Where do respondents prefer to purchase energy drinks?

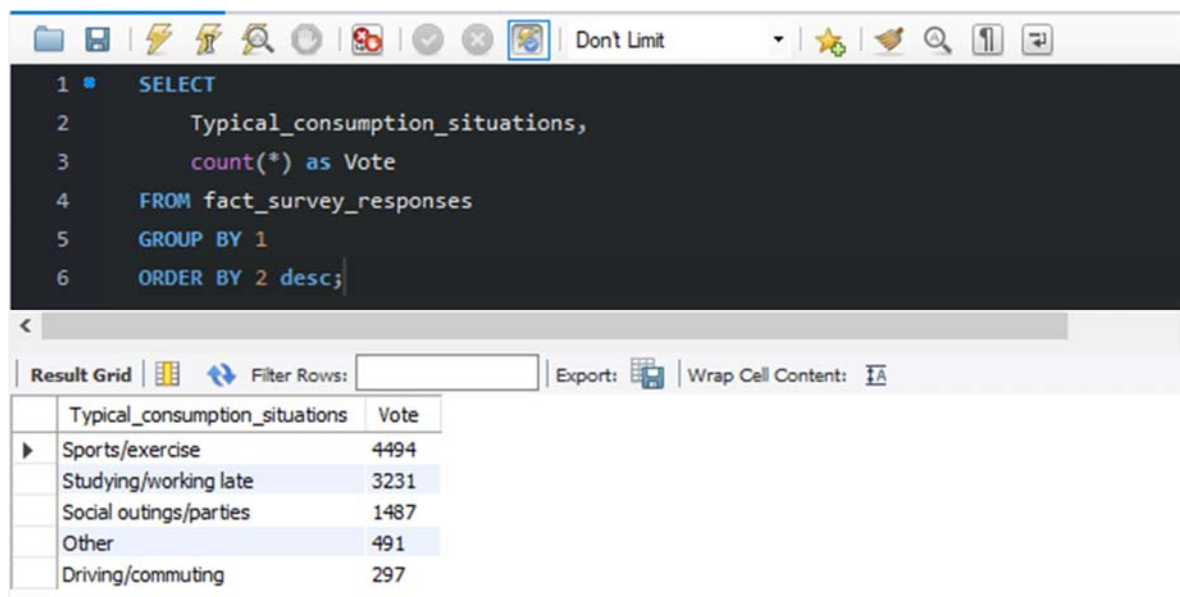


The screenshot shows a SQL query editor with a query that joins fact_survey_responses, dim_repondents, and dim_cities tables. It calculates the count of respondents for each purchase location and ranks them by preference. The results grid shows the top three locations for each tier.

```
310 With Tier_Location AS (  
311     SELECT  
312         Tier,  
313         Purchase_location,  
314         count(Respondent_ID) AS Vote,  
315         dense_rank() over(partition by Tier order by count(Respondent_ID) desc) as Preference_Rank  
316     FROM fact_survey_responses  
317     JOIN dim_repondents  
318     USING (Respondent_ID)  
319     JOIN dim_cities  
320     USING (City_ID)  
321     GROUP BY 1,2  
322     ORDER BY 1,4 )  
323 SELECT * FROM Tier_Location  
324 WHERE Preference_Rank <=3;
```

Tier	Purchase_location	Vote	Preference_Rank
Tier 1	Supermarkets	3391	1
Tier 1	Online retailers	1961	2
Tier 1	Gyms and fitness centers	1095	3
Tier 2	Supermarkets	1103	1
Tier 2	Online retailers	589	2
Tier 2	Gyms and fitness centers	369	3

2. What are the typical consumption situations for energy drinks among respondents?

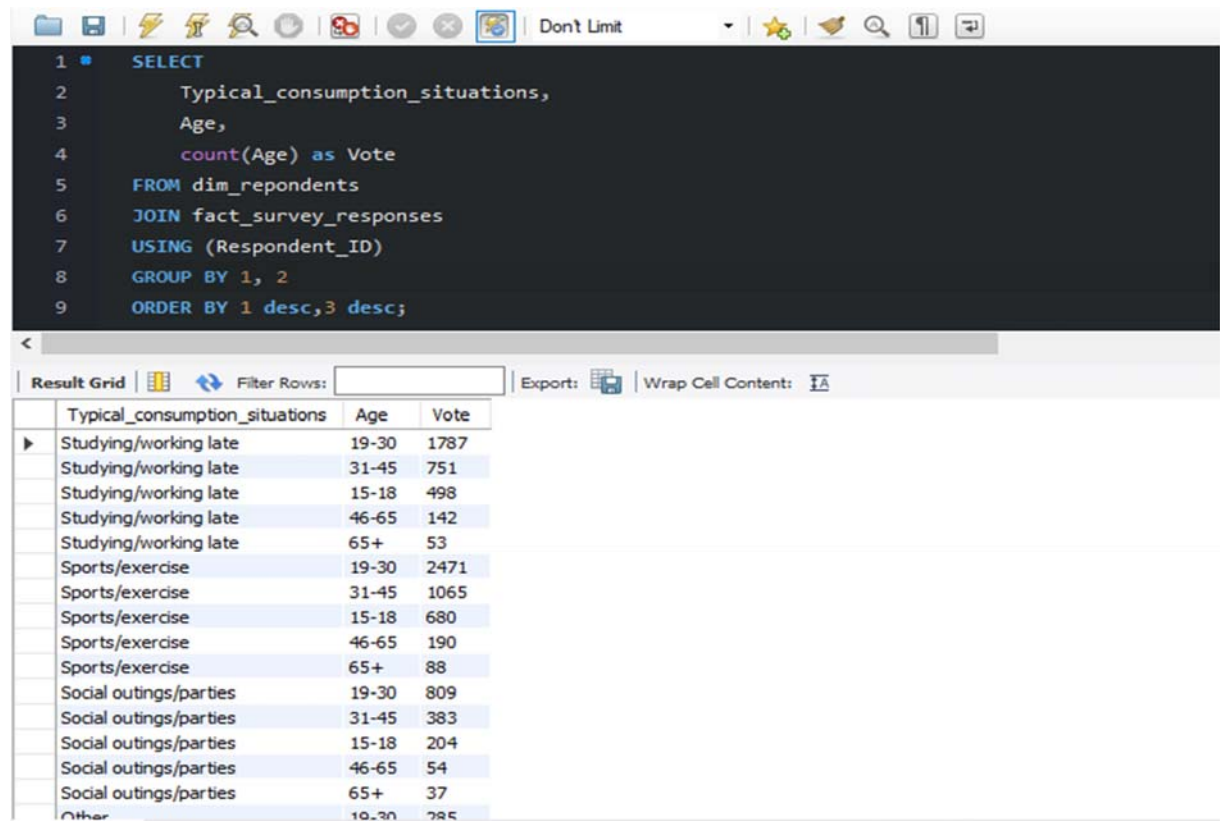


The screenshot shows a SQL query editor with a query that selects typical consumption situations from the fact_survey_responses table, grouped by situation and ordered by vote count. The results grid shows the top five consumption situations.

```
1 SELECT  
2     Typical_consumption_situations,  
3     count(*) as Vote  
4 FROM fact_survey_responses  
5 GROUP BY 1  
6 ORDER BY 2 desc;
```

Typical_consumption_situations	Vote
Sports/exercise	4494
Studying/working late	3231
Social outings/parties	1487
Other	491
Driving/commuting	297

By seeing which Age Group has more preference depending on each consumption situation,



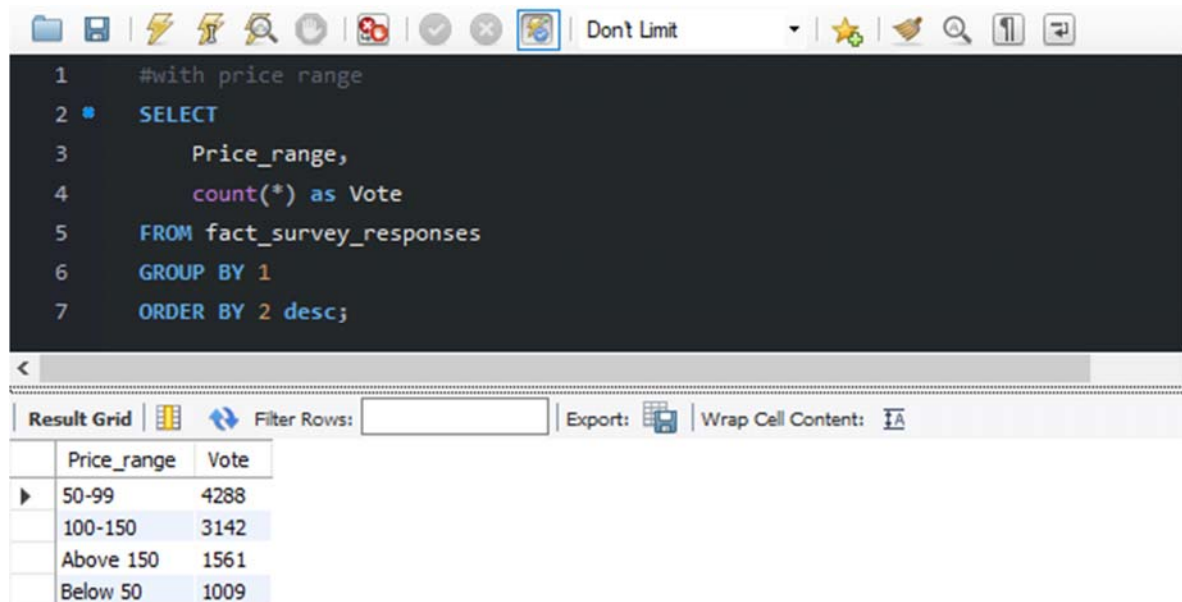
The screenshot shows a SQL query editor window with a dark theme. The query is a SELECT statement that joins 'dim_repondents' and 'fact_survey_responses' on 'Respondent_ID'. It groups the data by 'Typical_consumption_situations' and 'Age', and orders the results by 'Vote' in descending order. Below the query editor is a 'Result Grid' showing the output of the query. The grid has three columns: 'Typical_consumption_situations', 'Age', and 'Vote'. The data is grouped by consumption situation, with each situation having five rows corresponding to the age groups: 19-30, 31-45, 15-18, 46-65, and 65+.

```
1 SELECT
2     Typical_consumption_situations,
3     Age,
4     count(Age) as Vote
5 FROM dim_repondents
6 JOIN fact_survey_responses
7 USING (Respondent_ID)
8 GROUP BY 1, 2
9 ORDER BY 1 desc, 3 desc;
```

Typical_consumption_situations	Age	Vote
Studying/working late	19-30	1787
Studying/working late	31-45	751
Studying/working late	15-18	498
Studying/working late	46-65	142
Studying/working late	65+	53
Sports/exercise	19-30	2471
Sports/exercise	31-45	1065
Sports/exercise	15-18	680
Sports/exercise	46-65	190
Sports/exercise	65+	88
Social outings/parties	19-30	809
Social outings/parties	31-45	383
Social outings/parties	15-18	204
Social outings/parties	46-65	54
Social outings/parties	65+	37
Other	19-30	785

3. What factors influence respondents' purchase decisions, such as price range and limited-edition packaging?

With price range,



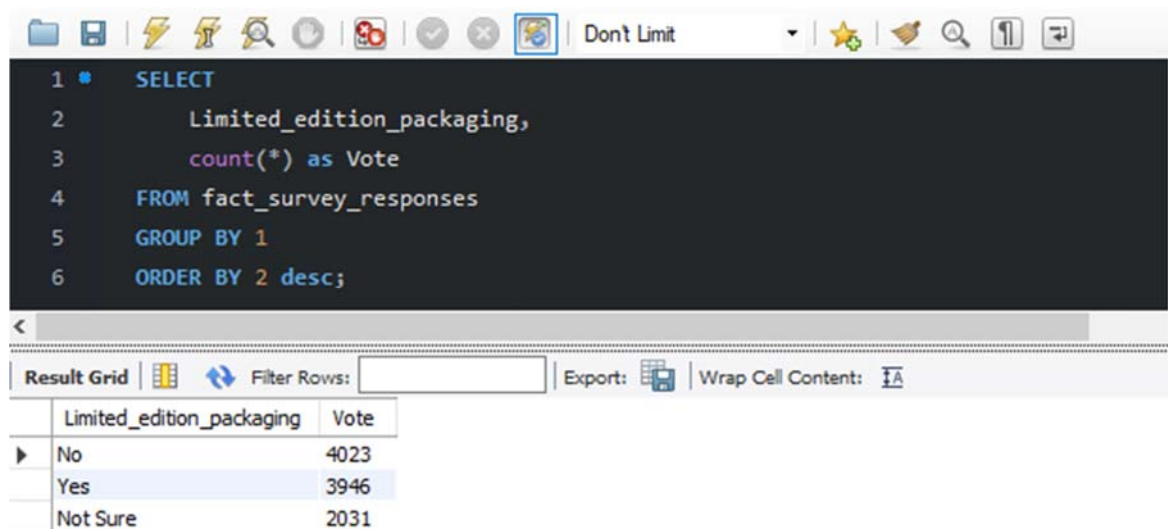
The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
1 #with price range
2 SELECT
3     Price_range,
4     count(*) as Vote
5 FROM fact_survey_responses
6 GROUP BY 1
7 ORDER BY 2 desc;
```

Below the query editor is a 'Result Grid' showing the results of the query. The grid has two columns: 'Price_range' and 'Vote'. The results are as follows:

Price_range	Vote
50-99	4288
100-150	3142
Above 150	1561
Below 50	1009

With limited edition packaging,



The screenshot shows a SQL query editor with a toolbar at the top. The query is as follows:

```
1 SELECT
2     Limited_edition_packaging,
3     count(*) as Vote
4 FROM fact_survey_responses
5 GROUP BY 1
6 ORDER BY 2 desc;
```

Below the query editor is a 'Result Grid' showing the results of the query. The grid has two columns: 'Limited_edition_packaging' and 'Vote'. The results are as follows:

Limited_edition_packaging	Vote
No	4023
Yes	3946
Not Sure	2031

Covered Topics

The SQL topics covered in this challenge are:

- Joins
- WHERE clause
- GROUP BY clause
- ORDER BY clause
- Sub-queries
- CTEs
- Windows functions
- Aggregate functions

Insights and Recommendation

1. CodeX should immediately improve its taste and flavor to stay on the line with the competitors. Their marketing team should consider the preferred ingredients from the consumers such as caffeine, vitamins, and sugar.
2. Since young people are driving the demand for the energy drinks market in India, CodeX should focus more on them by adapting to consumers' preferred marketing channels. In marketing channels, "Online Ads" is reaching the most to the Age Group 15-18 and 19-30 while it's "TV Commercials" for the Ages 31 and older.
3. CodeX energy drinks should be available more in supermarkets, online retailers, and the sports areas such as gyms and parks. Additionally, 51% of respondents from both Tiers have never tried CodeX's energy drink. Since "Not available locally" is the main reason why Tier-2 consumers have never tried the product, the marketing team should consider their products to be more available in Tier-2 cities.

4. CodeX's marketing team should consider having a young and sport-doing brand ambassador like a cricketer, Virat Kohli, who has one of the top most followers on Instagram. That will help with brand awareness and reputation, inspiration from young people, and many more.
5. CodeX can also make photo booths across different supermarkets in cities, making a contest for online short video content including their brand and rewarding the talented creator with different award categories such as the most liked and shared on social media platforms, creativity, art, and so on.