```
!pip install ipython-sql

# Step 2: Load the SQL magic extension
%load_ext sql
```

```
Requirement already satisfied: ipython-sql in c:\users\mreth\
anaconda3\lib\site-packages (0.5.0)
Requirement already satisfied: ipython in c:\users\mreth\anaconda3\
lib\site-packages (from ipython-sql) (7.31.1)
Requirement already satisfied: ipython-genutils in c:\users\mreth\
anaconda3\lib\site-packages (from ipython-sql) (0.2.0)
Requirement already satisfied: sqlalchemy>=2.0 in c:\users\mreth\
anaconda3\lib\site-packages (from ipython-sql) (2.0.35)
Requirement already satisfied: prettytable in c:\users\mreth\
anaconda3\lib\site-packages (from ipython-sql) (3.11.0)
Requirement already satisfied: six in c:\users\mreth\anaconda3\lib\
site-packages (from ipython-sql) (1.16.0)
Requirement already satisfied: sqlparse in c:\users\mreth\anaconda3\
lib\site-packages (from ipython-sql) (0.5.1)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\mreth\
anaconda3\lib\site-packages (from sqlalchemy>=2.0->ipython-sql)
(1.1.1)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\
mreth\anaconda3\lib\site-packages (from sqlalchemy>=2.0->ipython-sql)
(4.12.2)
Requirement already satisfied: pickleshare in c:\users\mreth\
anaconda3\lib\site-packages (from ipython->ipython-sql) (0.7.5)
Requirement already satisfied: jedi>=0.16 in c:\users\mreth\anaconda3\
lib\site-packages (from ipython->ipython-sql) (0.18.1)
Requirement already satisfied: prompt-toolkit!=3.0.0,!
=3.0.1,<3.1.0,>=2.0.0 in c:\users\mreth\anaconda3\lib\site-packages
(from ipython->ipython-sql) (3.0.20)
Requirement already satisfied: decorator in c:\users\mreth\anaconda3\
lib\site-packages (from ipython->ipython-sql) (5.1.1)
Requirement already satisfied: colorama in c:\users\mreth\anaconda3\
lib\site-packages (from ipython->ipython-sql) (0.4.5)
Requirement already satisfied: matplotlib-inline in c:\users\mreth\
anaconda3\lib\site-packages (from ipython->ipython-sql) (0.1.6)
Requirement already satisfied: pygments in c:\users\mreth\anaconda3\
lib\site-packages (from ipython->ipython-sql) (2.11.2)
Requirement already satisfied: setuptools>=18.5 in c:\users\mreth\
anaconda3\lib\site-packages (from ipython->ipython-sql) (63.4.1)
Requirement already satisfied: traitlets>=4.2 in c:\users\mreth\
anaconda3\lib\site-packages (from ipython->ipython-sql) (5.1.1)
Requirement already satisfied: backcall in c:\users\mreth\anaconda3\
lib\site-packages (from ipython->ipython-sql) (0.2.0)
Requirement already satisfied: wcwidth in c:\users\mreth\anaconda3\
lib\site-packages (from prettytable->ipython-sql) (0.2.5)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\mreth\
```

```
anaconda3\lib\site-packages (from jedi>=0.16->ipython->ipython-sql)
(0.8.3)
```

```
import sqlite3
import pandas as pd

# Create the SQLite connection
conn = sqlite3.connect(':memory:')  # In-memory database

# Load the CSV files into DataFrames
path = "C:\\Users\\mreth\\Downloads\\C6 Input Files\\Dataset\\"
dim_cities = pd.read_csv(path + "dim_cities.csv")
dim_respondents = pd.read_csv(path + "dim_respondents.csv")
fact_survey_responses = pd.read_csv(path +
"fact_survey_responses.csv")

# Create a persistent SQLite database saved on disk
conn = sqlite3.connect('my_database.db')  # This creates a file called
'my_database.db'


## Load CSV data into SQL tables
dim_cities.to_sql('dim_cities', conn, index=False,
if_exists='replace')
dim_respondents.to_sql('dim_respondents', conn, index=False,
if_exists='replace')
fact_survey_responses.to_sql('fact_survey_responses', conn,
index=False, if_exists='replace')

print("Tables loaded successfully into persistent SQLite database.")
```

```
Tables loaded successfully into persistent SQLite database.
```

```
# Load SQL magic and connect it to the same in-memory database
%sql sqlite:///my_database.db
```

```
%%sql
SELECT name FROM sqlite_master WHERE type='table';
```

```
 * sqlite:///my_database.db
Done.
```

```
[('dim_cities',), ('dim_respondents',), ('fact_survey_responses',)]
```

```
%%sql
SELECT f.Response_ID, r.Name, r.Age, r.Gender, c.City,
f.Consume_frequency, f.Consume_time,
```

```
       f.Consume_reason, f.Heard_before, f.Brand_perception,
f.General_perception,
       f.Tried_before, f.Taste_experience, f.Current_brands
FROM fact_survey_responses AS f
JOIN dim_respondents AS r ON f.Respondent_ID = r.Respondent_ID
JOIN dim_cities AS c ON r.City_ID = c.City_ID
LIMIT 10;
```

```
 * sqlite:///my_database.db
Done.
```

```
[(103001, 'Aniruddh Issac', '15-18', 'Female', 'Ahmedabad', '2-3 times
a week', 'To stay awake during work/study', 'Increased energy and
focus', 'Yes', 'Neutral', 'Not sure', 'No', 5, 'Blue Bull'),
 (103002, 'Trisha Rout', '19-30', 'Male', 'Pune', '2-3 times a month',
'Throughout the day', 'To boost performance', 'No', 'Neutral', 'Not
sure', 'No', 5, 'Bepsi'),
 (103003, 'Yuvraj  Virk', '15-18', 'Male', 'Hyderabad', 'Rarely',
'Before exercise', 'Increased energy and focus', 'No', 'Neutral', 'Not
sure', 'No', 2, 'Bepsi'),
 (103004, 'Pranay Chand', '31-45', 'Female', 'Bangalore', '2-3 times a
week', 'To stay awake during work/study', 'To boost performance',
'No', 'Positive', 'Dangerous', 'Yes', 5, 'Bepsi'),
 (103005, 'Mohanlal Joshi', '19-30', 'Female', 'Lucknow', 'Daily', 'To
stay awake during work/study', 'Increased energy and focus', 'Yes',
'Neutral', 'Effective', 'Yes', 5, 'Sky 9'),
 (103006, 'Zeeshan Ratta', '19-30', 'Female', 'Pune', 'Rarely', 'For
mental alertness', 'To combat fatigue', 'Yes', 'Negative', 'Not sure',
'No', 5, 'Cola-Coka'),
 (103007, 'Oorja Anne', '19-30', 'Male', 'Mumbai', '2-3 times a
month', 'To stay awake during work/study', 'Increased energy and
focus', 'No', 'Positive', 'Not sure', 'No', 4, 'Cola-Coka'),
 (103008, 'Rhea Khanna', '19-30', 'Male', 'Hyderabad', 'Rarely',
'Before exercise', 'To combat fatigue', 'No', 'Neutral', 'Healthy',
'Yes', 4, 'Gangster'),
 (103009, 'Zara Joshi', '46-65', 'Male', 'Hyderabad', 'Once a week',
'To stay awake during work/study', 'To enhance sports performance',
'No', 'Neutral', 'Effective', 'Yes', 3, 'Gangster'),
 (103010, 'Sana Dhawan', '19-30', 'Female', 'Hyderabad', 'Once a
week', 'For mental alertness', 'To combat fatigue', 'Yes', 'Neutral',
'Healthy', 'No', 4, 'Sky 9')]
```

Gender Distribution: How gender is distributed among respondents.

```
%%sql
SELECT r.Gender, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
JOIN dim_respondents AS r ON f.Respondent_ID = r.Respondent_ID
```

```
GROUP BY r.Gender
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Male', 6038), ('Female', 3455), ('Non-binary', 507)]
```

Age Group Distribution: How energy drink consumption varies by age group.

```
%%sql
SELECT r.Age, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
JOIN dim_respondents AS r ON f.Respondent_ID = r.Respondent_ID
GROUP BY r.Age
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('19-30', 5520),
 ('31-45', 2376),
 ('15-18', 1488),
 ('46-65', 426),
 ('65+', 190)]
```

City-Wise Responses: Distribution of responses by city.

```
%%sql
SELECT c.City, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
JOIN dim_respondents AS r ON f.Respondent_ID = r.Respondent_ID
JOIN dim_cities AS c ON r.City_ID = c.City_ID
GROUP BY c.City
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Bangalore', 2828),
 ('Hyderabad', 1833),
 ('Mumbai', 1510),
 ('Chennai', 937),
 ('Pune', 906),
 ('Kolkata', 566),
 ('Ahmedabad', 456),
 ('Delhi', 429),
```

```
 ('Jaipur', 360),
 ('Lucknow', 175)]
```

Demographic Insights

1. Gender Distribution

- **Male respondents**: 5038
- **Female respondents**: 3455
- **Non-binary respondents**: 507

Inference: Males form the majority of respondents, followed by females and non-binary individuals. No significant disparity is observed between males and females in terms of representation, but the non-binary group is significantly smaller.

---

2. Age Group Distribution

- **19-30 years**: 5520 respondents
- **31-45 years**: 2376 respondents
- **15-18 years**: 1488 respondents
- **46-65 years**: 426 respondents
- **65+ years**: 190 respondents

Inference: The 19-30 age group is the largest consumer demographic, followed by the 31-45 group. Younger and older age groups represent much smaller proportions of respondents.

---

3. City-Wise Distribution

- **Top cities by number of respondents**:
  - Bangalore: 2828
  - Hyderabad: 1833
  - Mumbai: 1510
  - Chennai: 937
  - Pune: 906
  - Kolkata: 566

Inference: Bangalore, Hyderabad, and Mumbai have the largest number of respondents, making them key markets for energy drinks based on this dataset. Cities like Kolkata, Ahmedabad, and Delhi have smaller respondent pools.

Consumer Preferences:

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

```
%%sql
SELECT f.Ingredients_expected, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Ingredients_expected
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Caffeine', 3896), ('Vitamins', 2534), ('Sugar', 2017), ('Guarana',
1553)]
```

What are the packaging preferences for energy drinks?

```
%%sql
SELECT f.Packaging_preference, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Packaging_preference
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Compact and portable cans', 3984),
 ('Innovative bottle design', 3047),
 ('Collectible packaging', 1501),
 ('Eco-friendly design', 983),
 ('Other', 485)]
```

Caffeine is the most preferred ingredient in energy drinks, followed by vitamins and sugar. Guarana, while less popular than the others, still has a considerable number of respondents who expect it in energy drinks.

The most preferred packaging option is compact and portable cans, followed by innovative bottle designs. Collectible packaging and eco-friendly designs are less favored, but still have a notable group of respondents interested in them.

Competition Analysis:

Who are the current market leaders?

What are the primary reasons consumers prefer those brands over CodeX?

```
%%sql
SELECT f.Current_brands, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Current_brands
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Cola-Coka', 2538),
 ('Bepsi', 2112),
 ('Gangster', 1854),
 ('Blue Bull', 1058),
 ('CodeX', 980),
 ('Sky 9', 979),
 ('Others', 479)]

%%sql
SELECT f.Reasons_for_choosing_brands, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Reasons_for_choosing_brands
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Brand reputation', 2652),
 ('Taste/flavor preference', 2011),
```

```
 ('Availability', 1910),
 ('Effectiveness', 1748),
 ('Other', 1679)]
```

1. Current Market Leaders

Inference:
Cola-Coka and Bepsi dominate the market, with CodeX trailing significantly. Positioned alongside Sky 9, CodeX has much lower market penetration compared to the top three brands. There is a need to enhance brand visibility and consumer engagement to compete effectively in this space.

---

2. Reasons for Choosing Other Brands

Inference:
The primary reasons for choosing other brands are strong brand reputation and better taste/flavor profiles. Availability and product effectiveness also play crucial roles in consumer decisions. For CodeX to improve its market position, the focus should be on building a stronger brand identity and refining the product's taste, while ensuring better availability across key markets.

Marketing Channels and Brand Awareness:

Which marketing channel can be used to reach more customers?

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

```
%%sql
SELECT f.Marketing_channels, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Marketing_channels
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Online ads', 4020),
 ('TV commercials', 2688),
 ('Outdoor billboards', 1226),
```

```
 ('Other', 1225),
 ('Print media', 841)]

%%sql
SELECT f.Marketing_channels, f.Brand_perception, COUNT(*) AS
total_responses
FROM fact_survey_responses AS f
GROUP BY f.Marketing_channels, f.Brand_perception
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Online ads', 'Neutral', 2418),
 ('TV commercials', 'Neutral', 1541),
 ('Online ads', 'Positive', 887),
 ('Other', 'Neutral', 765),
 ('Outdoor billboards', 'Neutral', 743),
 ('Online ads', 'Negative', 715),
 ('TV commercials', 'Positive', 651),
 ('Print media', 'Neutral', 507),
 ('TV commercials', 'Negative', 496),
 ('Outdoor billboards', 'Positive', 281),
 ('Other', 'Positive', 245),
 ('Other', 'Negative', 215),
 ('Outdoor billboards', 'Negative', 202),
 ('Print media', 'Positive', 193),
 ('Print media', 'Negative', 141)]
```

Marketing Channels and Brand Awareness

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

Inference:
Online ads have the highest reach among respondents, with 4020 interactions, followed by **TV commercials** with 2688 responses. These two channels are the most effective at reaching a large audience. Outdoor billboards, other channels, and print media are less effective in terms of overall reach.

---

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

Inference:
Most respondents perceive marketing through **online ads** and **TV commercials** as "neutral," with 2418 and 1541 neutral responses, respectively. However, **online ads** also have a significant number of "positive" perceptions (887), indicating their effectiveness in creating a favorable

brand image. TV commercials are also positively perceived by 651 respondents, but have a higher proportion of negative perceptions (496).

Outdoor billboards and other channels generally receive neutral or mixed feedback, while print media has the least reach and impact, with fewer positive and neutral perceptions.

To improve brand awareness and perception, focusing on **online ads** and **TV commercials** as primary channels seems most effective, while exploring opportunities to strengthen engagement through these platforms can improve brand positivity.

Brand Penetration:

a. What do people think about our brand? (overall rating)

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

```
%%sql
SELECT f.Brand_perception, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Brand_perception
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Neutral', 5974), ('Positive', 2257), ('Negative', 1769)]

%%sql
SELECT c.City, f.Brand_perception, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
JOIN dim_respondents AS r ON f.Respondent_ID = r.Respondent_ID
JOIN dim_cities AS c ON r.City_ID = c.City_ID
GROUP BY c.City, f.Brand_perception
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Bangalore', 'Neutral', 1844),
 ('Hyderabad', 'Neutral', 1191),
 ('Mumbai', 'Neutral', 847),
 ('Chennai', 'Neutral', 615),
 ('Bangalore', 'Positive', 566),
 ('Mumbai', 'Positive', 435),
 ('Pune', 'Neutral', 419),
 ('Bangalore', 'Negative', 418),
 ('Kolkata', 'Neutral', 393),
```

```
('Hyderabad', 'Positive', 376),
('Pune', 'Negative', 282),
('Hyderabad', 'Negative', 266),
('Mumbai', 'Negative', 228),
('Delhi', 'Neutral', 210),
('Ahmedabad', 'Neutral', 208),
('Pune', 'Positive', 205),
('Chennai', 'Positive', 196),
('Jaipur', 'Neutral', 163),
('Ahmedabad', 'Negative', 149),
('Delhi', 'Positive', 145),
('Chennai', 'Negative', 126),
('Jaipur', 'Negative', 106),
('Kolkata', 'Positive', 106),
('Ahmedabad', 'Positive', 99),
('Jaipur', 'Positive', 91),
('Lucknow', 'Neutral', 84),
('Delhi', 'Negative', 74),
('Kolkata', 'Negative', 67),
('Lucknow', 'Negative', 53),
('Lucknow', 'Positive', 38)]
```

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

Inference:
The majority of respondents hold a **neutral perception** of the brand (5974 respondents). While **positive perceptions** (2257 respondents) exceed the **negative perceptions** (1769 respondents), there is room for improvement in shifting more neutral and negative perceptions toward the positive. The high number of neutral perceptions indicates an opportunity to strengthen brand engagement and create a stronger brand identity.

---

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

Inference:
Bangalore, Hyderabad, and Mumbai are critical markets, with the highest number of respondents showing **neutral perceptions**. Bangalore has a relatively balanced mix of positive and negative feedback, while cities like Pune and Chennai also show considerable negative sentiment, indicating the need for targeted campaigns to improve brand perception.

Cities like **Ahmedabad** and **Kolkata** show a higher proportion of negative feedback relative to their overall response count, suggesting that more attention is needed in these markets to shift consumer sentiment. **Lucknow** also shows weaker brand penetration and should be prioritized for improvement.

Purchase Behavior:

a. Where do respondents prefer to purchase energy drinks?

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

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

```
%%sql
SELECT f.Purchase_location, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Purchase_location
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Supermarkets', 4494),
 ('Online retailers', 2550),
 ('Gyms and fitness centers', 1464),
 ('Local stores', 813),
 ('Other', 679)]
```

```
%%sql
SELECT f.Typical_consumption_situations, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Typical_consumption_situations
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.
```

```
[('Sports/exercise', 4494),
 ('Studying/working late', 3231),
 ('Social outings/parties', 1487),
 ('Other', 491),
 ('Driving/commuting', 297)]
```

```
%%sql
SELECT f.Price_range, f.Limited_edition_packaging, COUNT(*) AS
total_responses
FROM fact_survey_responses AS f
GROUP BY f.Price_range, f.Limited_edition_packaging
ORDER BY total_responses DESC;
```

```
 * sqlite:///my_database.db
Done.

[('50-99', 'No', 1741),
 ('50-99', 'Yes', 1679),
 ('100-150', 'Yes', 1263),
 ('100-150', 'No', 1244),
 ('50-99', 'Not Sure', 868),
 ('Above 150', 'No', 639),
 ('100-150', 'Not Sure', 635),
 ('Above 150', 'Yes', 600),
 ('Below 50', 'Yes', 404),
 ('Below 50', 'No', 399),
 ('Above 150', 'Not Sure', 322),
 ('Below 50', 'Not Sure', 206)]
```

Inference:
The majority of respondents prefer to purchase energy drinks from **supermarkets** (4494 respondents), followed by **online retailers** (2550 respondents). **Gyms and fitness centers** are also notable purchase locations (1464 respondents), while local stores and other venues are less popular. This suggests that focusing distribution and marketing efforts on supermarkets and online channels would yield the best results.

Inference:
Energy drinks are most commonly consumed during **sports or exercise** (4494 respondents), followed by **studying or working late** (3231 respondents). Social outings and commuting situations are less frequent but still significant consumption contexts. Marketing campaigns should focus on these primary situations, emphasizing energy and performance for both physical activity and mental focus.

Inference:
The **50-99 price range** is the most preferred, with a notable interest in **limited edition packaging** within this price range (1679 respondents preferring it). The **100-150 price range** is also popular, with nearly equal preferences for packaging variations. Limited edition packaging appears to have strong appeal across multiple price points, suggesting that offering exclusive packaging options could help drive sales. However, respondents in higher price brackets (Above 150) show less enthusiasm for limited edition packaging, indicating that premium pricing may not benefit as much

Product Development

a. Which area of business should we focus more on our product development? (Branding/taste/availability)

```
%%sql
SELECT f.Reasons_for_choosing_brands, COUNT(*) AS total_responses
FROM fact_survey_responses AS f
GROUP BY f.Reasons_for_choosing_brands
ORDER BY total_responses DESC;

 * sqlite:///my_database.db
Done.

[('Brand reputation', 2652),
 ('Taste/flavor preference', 2011),
 ('Availability', 1910),
 ('Effectiveness', 1748),
 ('Other', 1679)]
```

Which area of business should we focus more on for product development? (Branding, taste, availability)

Inference:
The primary focus for product development should be **brand reputation**, as it is the top reason consumers choose energy drinks (2652 respondents). **Taste/flavor preference** (2011 respondents) and **availability** (1910 respondents) are also critical areas but slightly less influential. To strengthen market positioning, CodeX should prioritize improving brand perception while simultaneously refining the product's taste and expanding availability in key markets.
