CodeX

Data-Driven Insights into Consumer Behavior and Brand Perception in the Beverage Industry

Banuvathy R

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. A survey was conducted in those 10 cities and received results from 10k respondents

Objective

The Marketing team is responsible for increasing brand awareness, market share, and product development. a marketing data analyst is tasked to convert these survey results to meaningful insights which the team can use to drive actions.
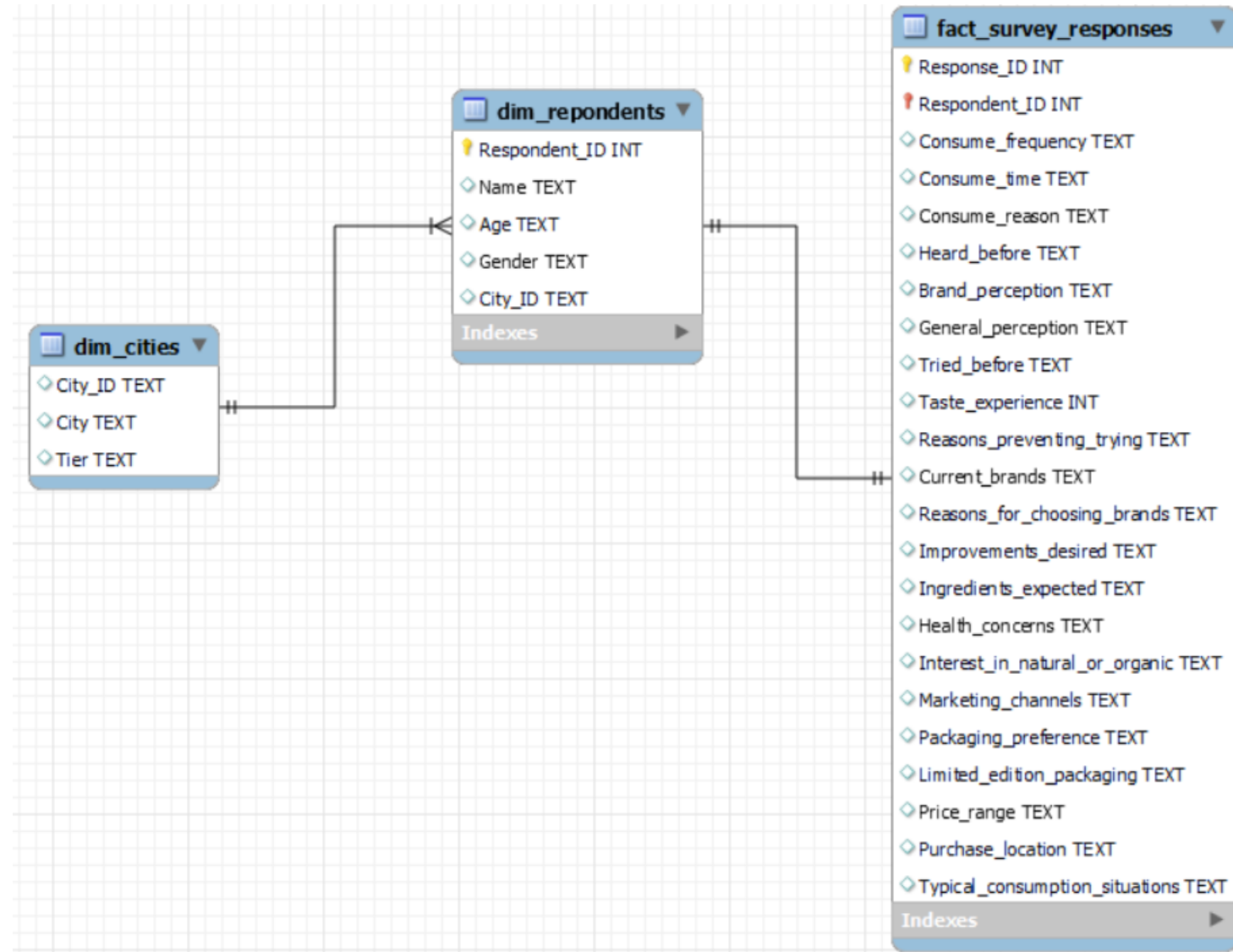
Dataset

The dataset contain 3 tables:

1. dim_respondents

2. dim_cities

3. fact_survey_responses

Data model

dim_cities
- City_ID TEXT
- City TEXT
- Tier TEXT

dim_repondents
- Respondent_ID INT
- Name TEXT
- Age TEXT
- Gender TEXT
- City_ID TEXT
- Indexes

fact_survey_responses
- 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
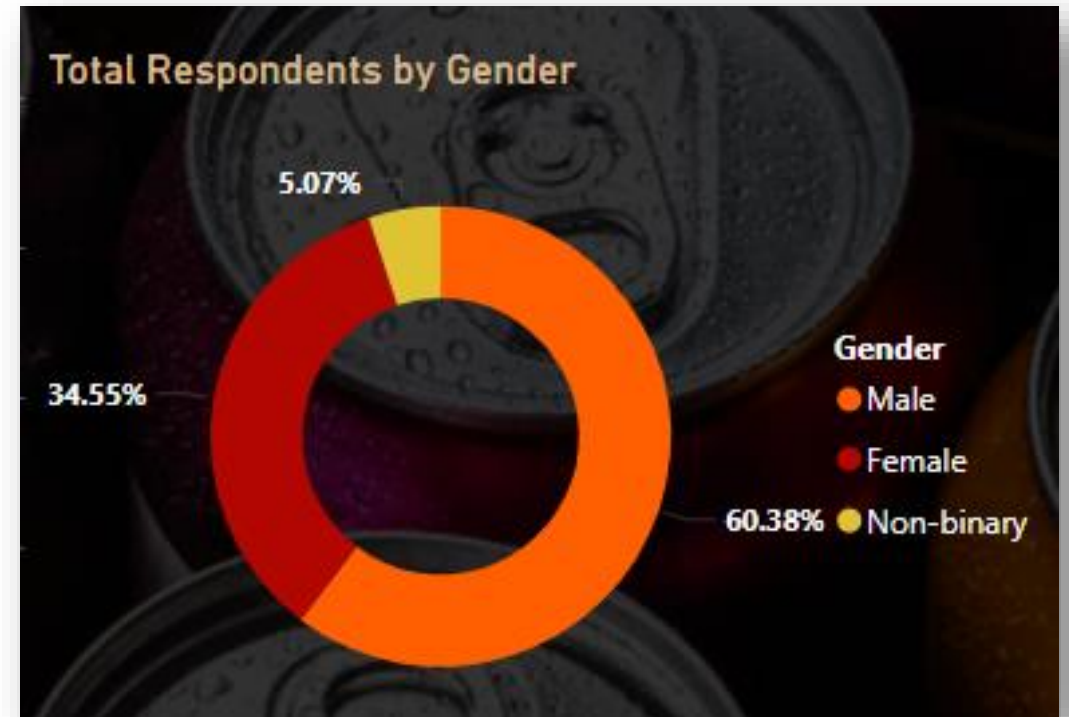- Typical_consumption_situations TEXT
- Indexes

Demographic Insights

Problem Statement 1: Identify which demographic groups exhibit the highest preference for energy drinks based on age and gender

```sql
SELECT
    Gender,
    COUNT(*) as cnt
FROM
    dim_repondents
GROUP BY
    Gender
ORDER BY
    cnt DESC;
```

Gender	cnt
Male	6038
Female	3455
Non-binary	507

Total Respondents by Gender

5.07%

34.55%

60.38%

Gender
- Male
- Female
- Non-binary

- Male prefer to drink more compared to Female and non binary

Demographic Insights

Problem Statement 1: Identify which demographic groups exhibit the highest preference for energy drinks based on age and gender

```sql
SELECT
    Age,
    COUNT(*) as cnt
FROM
    dim_repondents
GROUP BY
    Age
ORDER BY
    cnt DESC;
```

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

Total Respondents by Age
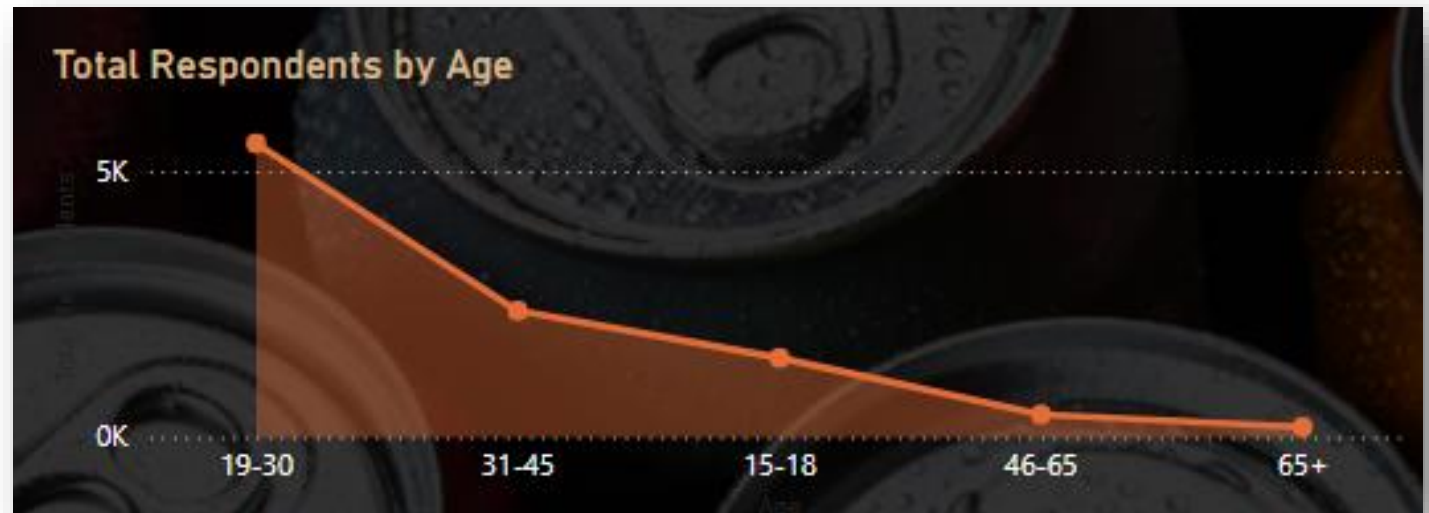
- Age group 19-30 prefer to drink more followed by the age group 31-45

Problem Statement 2
Determine which type of marketing channel is most effective in reaching the youth demographic (ages 15-30).

```sql
SELECT
    fsr.Marketing_Channels,
    COUNT(*) as cnt
FROM
    fact_survey_responses fsr
INNER JOIN
    dim_repondents dr ON fsr.Respondent_ID = dr.Respondent_ID
WHERE
    Age in ('15-18', '19-30')
GROUP BY
    fsr.Marketing_Channels
ORDER BY
    cnt DESC;
```

Total Respondents by Marketing_channels

Marketing_Channels	cnt
Online ads	3373
TV commercials	1785
Other	702
Outdoor billboards	702
Print media	446
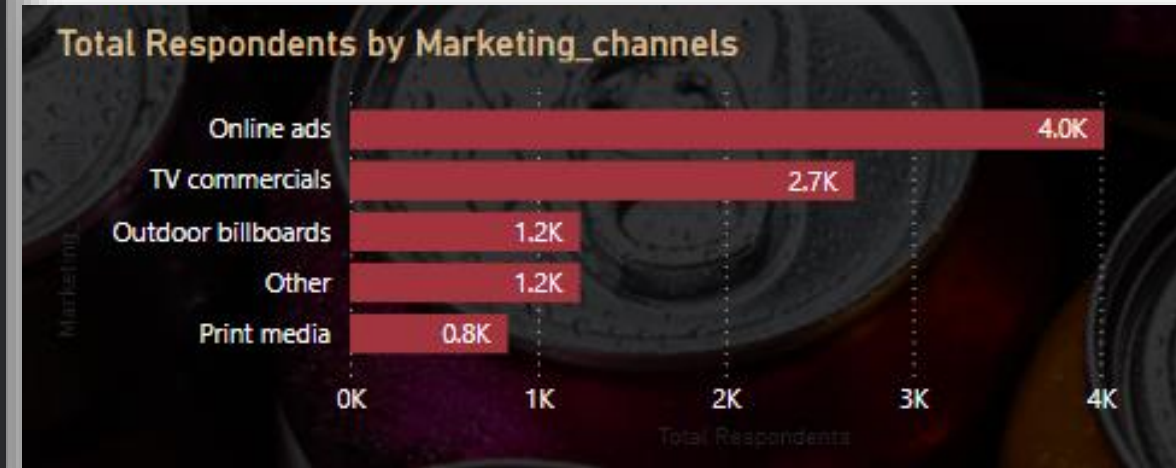
- Online ads reaches the most Youth (15-30)
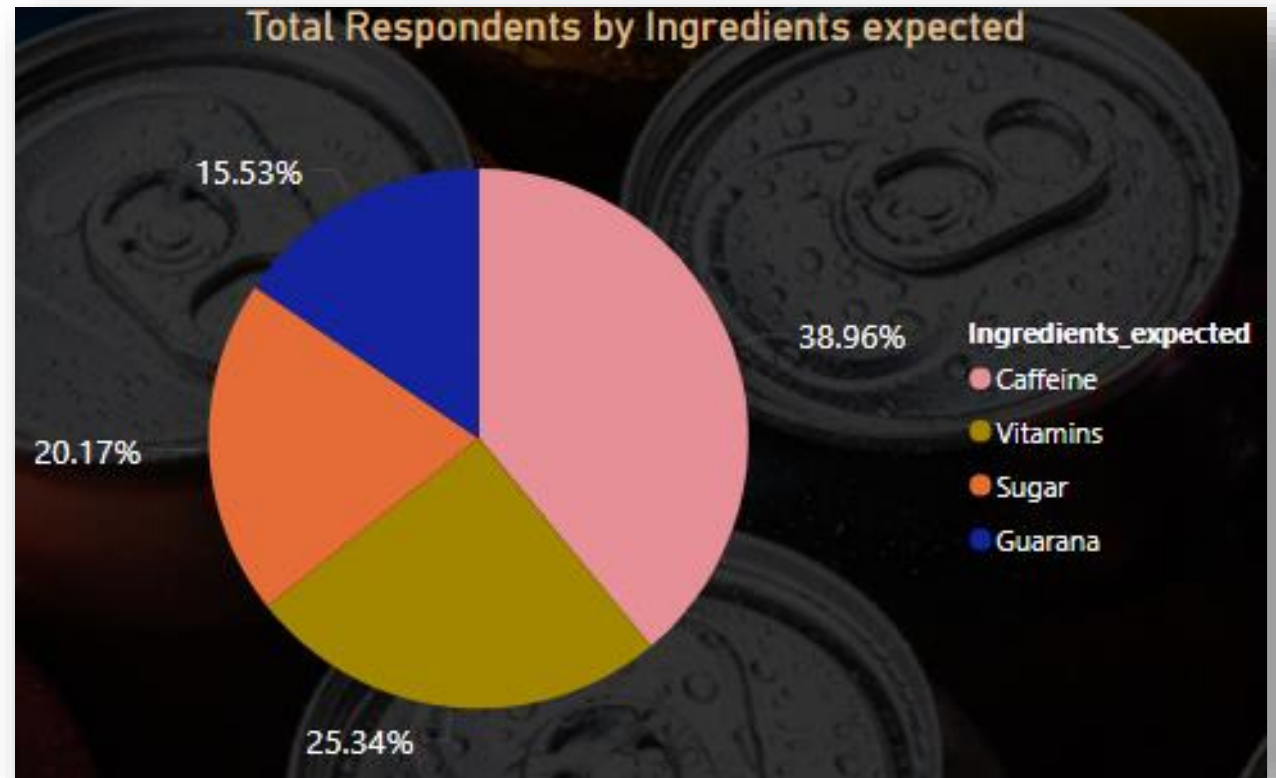
Consumer Preferences

Problem Statement 3
Analyze survey data to determine packaging and ingredients preferences for energy drinks among respondents.

```sql
SELECT
    Ingredients_expected,
    count(*) as cnt
FROM
    fact_survey_responses
GROUP BY
    Ingredients_expected
ORDER BY
    cnt DESC;
```

Ingredients_expected	cnt
Caffeine	3896
Vitamins	2534
Sugar	2017
Guarana	1553

Total Respondents by Ingredients expected

15.53%
38.96%
20.17%
25.34%

Ingredients_expected
- Caffeine
- Vitamins
- Sugar
- Guarana

- Most preferred ingredient of energy drink is Caffeine

Consumer Preferences

Problem Statement 3

Analyze survey data to determine packaging and ingredients preferences for energy drinks among respondents.

```sql
SELECT
    Packaging_preference,
    count(*) as cnt
FROM
    fact_survey_responses
GROUP BY
    Packaging_preference
ORDER BY
    cnt DESC;
```

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

Total Respondents by Packaging preference
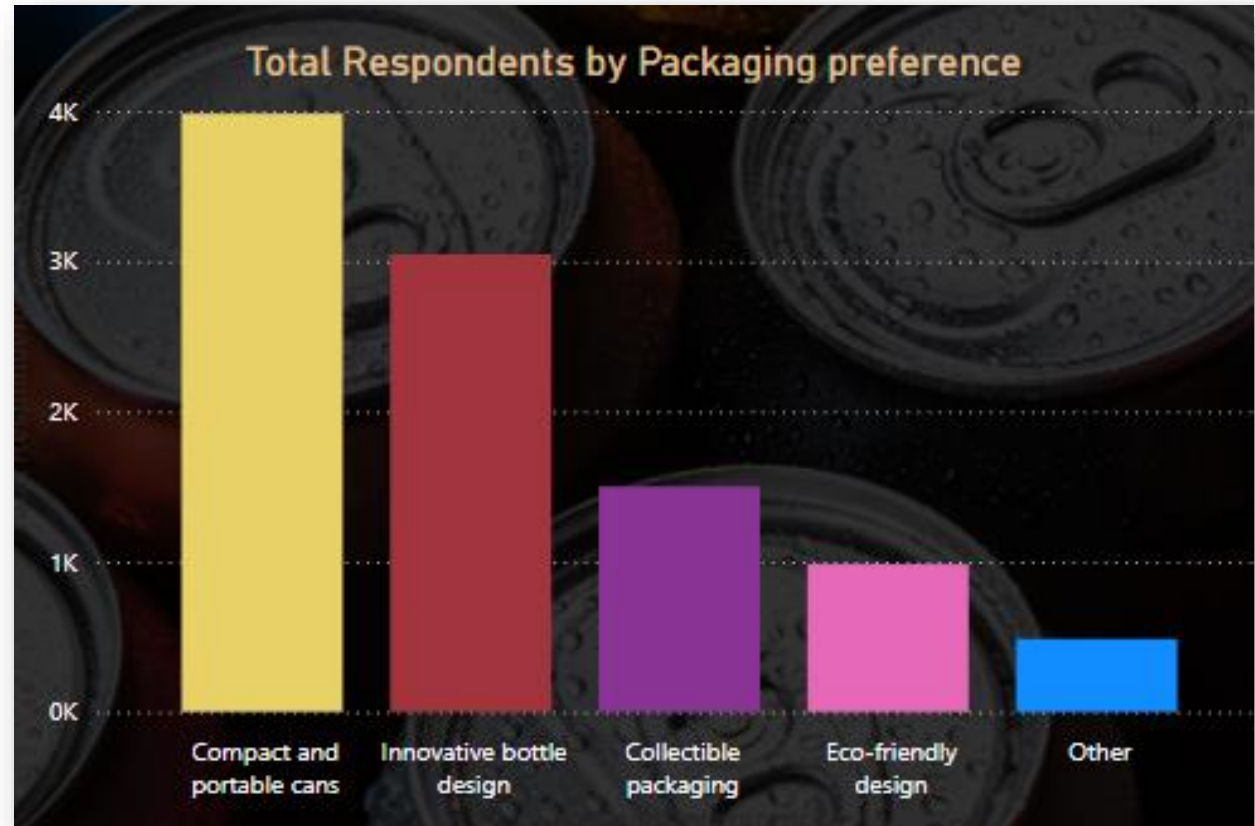
- Most consumers prefer to have package made of Compact and portable cans

```sql
WITH codex_responses AS (
    SELECT r.Age, s.Consume_reason
    FROM dim_repondents r
    JOIN fact_survey_responses s ON r.Respondent_ID = s.Respondent_ID
    WHERE s.Current_brands = 'CodeX'
),
reason_counts AS (
    SELECT Age, Consume_reason, COUNT(*) AS reason_count
    FROM codex_responses
    GROUP BY Age, Consume_reason
),
rank_find AS (
    SELECT Age, Consume_reason, reason_count,
            RANK() OVER (PARTITION BY Age ORDER BY reason_count DESC) AS RNK
FROM reason_counts)
SELECT *
FROM rank_find
WHERE RNK = 1;
```

Age	Consume_reason	reason_count	RNK
15-18	Increased energy and focus	37	1
19-30	Increased energy and focus	192	1
31-45	Increased energy and focus	93	1
46-65	Increased energy and focus	13	1
65+	Increased energy and focus	9	1

Competition Analysis

Problem Statement 4 Current market leaders and the reasons for preferring those brands over CodeX.

```sql
SELECT
    Current_brands,
    count(*) as cnt
FROM
    fact_survey_responses
GROUP BY
    Current_brands
ORDER BY
    cnt DESC;
```

Current_brands	cnt
Cola-Coka	2538
Bepsi	2112
Gangster	1854
Blue Bull	1058
CodeX	980
Sky 9	979
Others	479
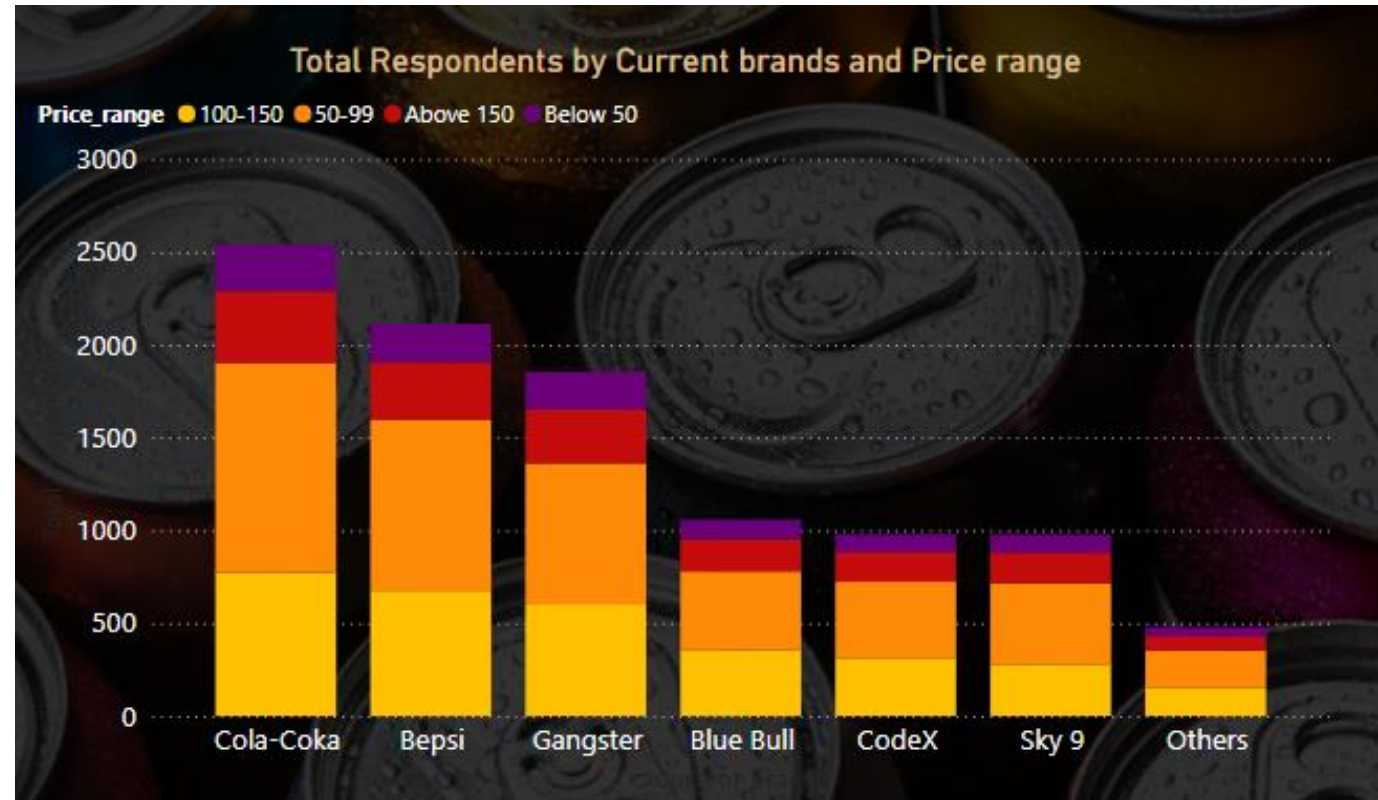
Total Respondents by Current brands and Price range
Price_range ● 100-150 ● 50-99 ● Above 150 ● Below 50

- Brand reputation and taste/ flavour preference was the main reason for choosing other brands over CodeX

Competition Analysis

Problem Statement 4 Current market leaders and the reasons for preferring those brands over CodeX.

```sql
SELECT
    subq.Current_brands,
    subq.Reasons_for_choosing_brands,
    subq.cnt
FROM
    (SELECT
        Current_brands,
        Reasons_for_choosing_brands,
        count(*) as cnt
    FROM
        fact_survey_responses
    WHERE
        Current_brands in ('Cola-Coka', 'Bepsi', 'Gangster')
    GROUP BY
        Current_brands,
        Reasons_for_choosing_brands
    ) as subq
ORDER BY
    Current_brands ASC,
    cnt DESC;
```

Current_brands	Reasons_for_choosing_brands	cnt
Bepsi	Brand reputation	577
Bepsi	Taste/flavor preference	423
Bepsi	Availability	418
Bepsi	Other	355
Bepsi	Effectiveness	339
Cola-Coka	Brand reputation	616
Cola-Coka	Taste/flavor preference	531
Cola-Coka	Availability	510
Cola-Coka	Other	448
Cola-Coka	Effectiveness	433
Gangster	Brand reputation	511
Gangster	Taste/flavor preference	357
Gangster	Availability	339
Gangster	Effectiveness	338
Gangster	Other	309

Current_brands	Total Respondents
⊟ Cola-Coka	2538
Brand reputation	616
Taste/flavor preference	531
Availability	510
Other	448
Effectiveness	433
⊟ Bepsi	2112
Brand reputation	577
Taste/flavor preference	423
Availability	418
Other	355
Effectiveness	339
⊟ Gangster	1854
Brand reputation	511
Taste/flavor preference	357
Availability	339
Effectiveness	338
Other	309
⊞ Blue Bull	1058
⊟ CodeX	980
Brand reputation	259
Availability	195
Taste/flavor preference	182
Effectiveness	176
Other	168

- Brand reputation and taste/ flavour preference was the main reason for choosing other brands over CodeX
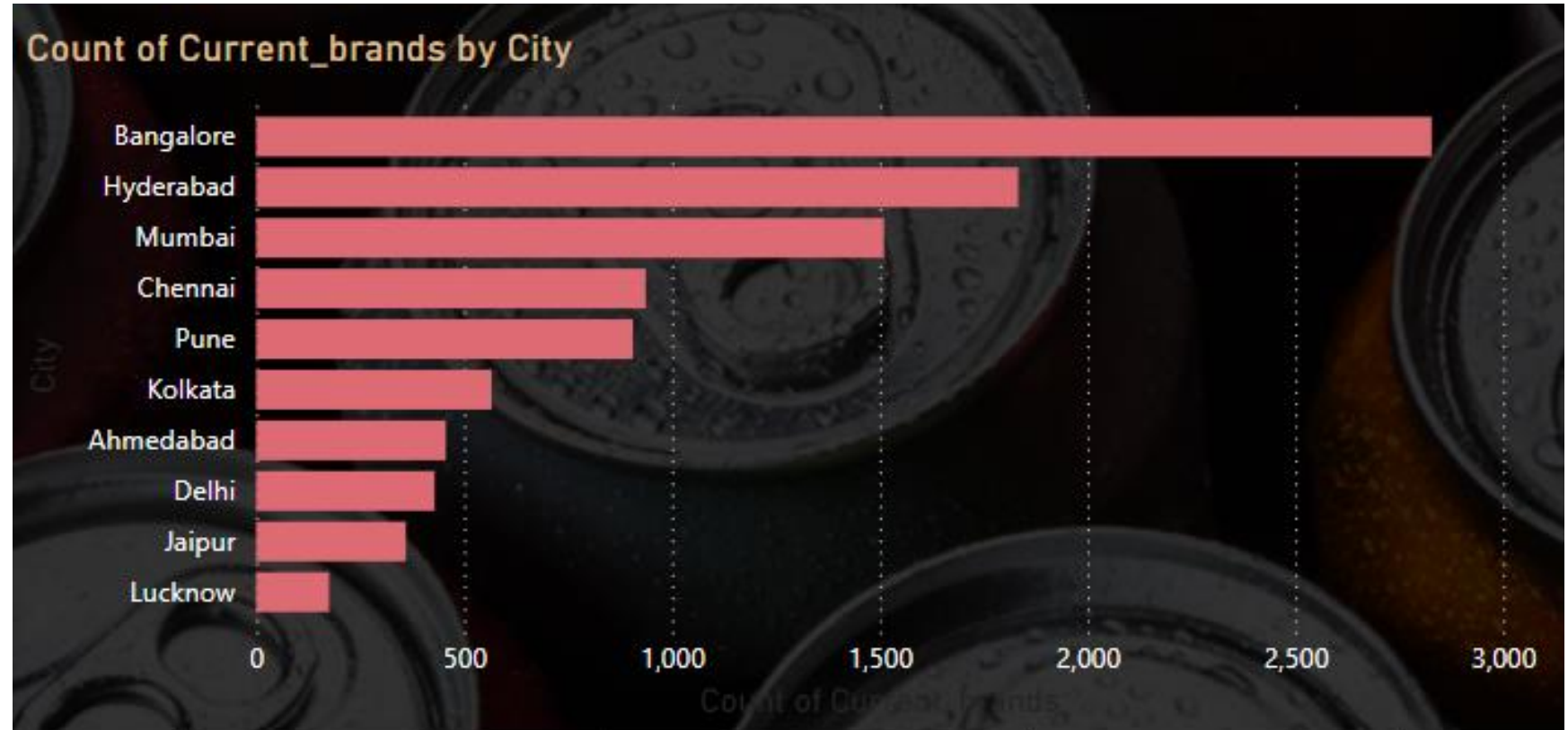
Brand Penetration

```sql
CREATE VIEW
    respondents_city AS
SELECT
    r.Respondent_ID, r.Name, r.Age, r.Gender, c.City_ID, c.City, c.Tier
FROM
    dim_repondents r
LEFT JOIN
    dim_cities c ON r.City_ID = c.City_ID;


SELECT
    rc.City, Count(*) as cnt
FROM
    fact_survey_responses f
LEFT JOIN
    respondents_city rc ON f.Respondent_ID = rc.Respondent_ID
WHERE
    Current_brands = 'Codex'
GROUP BY
    rc.City
ORDER BY
    cnt DESC;
```

City	cnt
Bangalore	292
Hyderabad	182
Mumbai	156
Chennai	92
Pune	92
Kolkata	48
Ahmedabad	45
Delhi	40
Jaipur	28
Lucknow	5

Brand Penetration

Count of Current_brands by City

City	Count of Current_brands
Bangalore	~2,800
Hyderabad	~1,800
Mumbai	~1,500
Chennai	~900
Pune	~900
Kolkata	~550
Ahmedabad	~450
Delhi	~400
Jaipur	~350
Lucknow	~200

```sql
SELECT
    c.Tier,
    GROUP_CONCAT(DISTINCT City) as Cities,
    SUM(CASE WHEN s.Heard_before = 'Yes' THEN 1 ELSE 0 END) / COUNT(*) * 100 AS percentage_heard_before
FROM
    dim_repondents r
JOIN fact_survey_responses s ON r.Respondent_ID = s.Respondent_ID
JOIN dim_cities c ON r.City_ID = c.City_ID
GROUP BY c.Tier;
```

Tier	Cities	percentage_heard_before
Tier 1	Bangalore,Chennai,Delhi,Hyderabad,Mumbai	45.4292
Tier 2	Ahmedabad,Jaipur,Kolkata,Lucknow,Pune	41.5347

Total Respondents by Heard_before

No	5.6K
Yes	4.4K

0K — 5K
Total Respondents
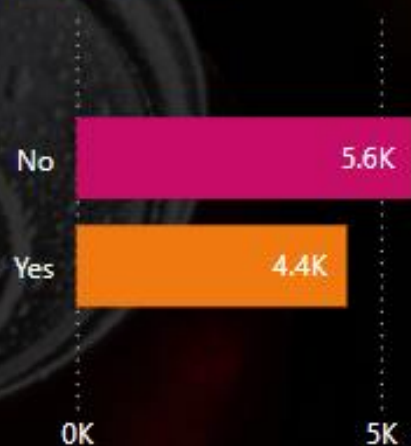
Tier 1

7537	2587	4587
Total Respond...	Total No of Fe...	Total No of Males

Tier 2

2463	868	1451
Total Respond...	Total No of Fe...	Total No of Males

```sql
SELECT
  Price_range,
  count(*) as cnt
FROM
  fact_survey_responses
WHERE
  Current_brands = 'Codex'
GROUP BY
  Price_range
ORDER BY
  cnt DESC;
```

Price_range	cnt
50-99	410
100-150	313
Above 150	159
Below 50	98

Count of Current_brands by Price_range

| 50-99 | 100-150 | Above 150 |
| | | Below 50 |

```
SELECT
    Purchase_location,
    count(*) as cnt
FROM
    fact_survey_responses
GROUP BY
    Purchase_location
ORDER BY
    cnt DESC;
```

Purchase_location	cnt
Supermarkets	4494
Online retailers	2550
Gyms and fitness centers	1464
Local stores	813
Other	679
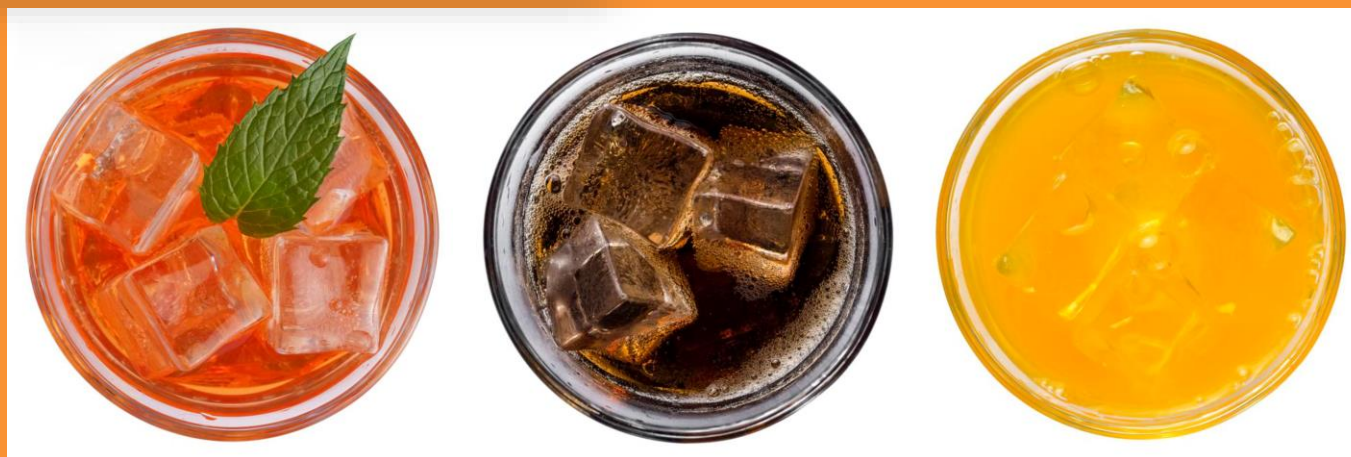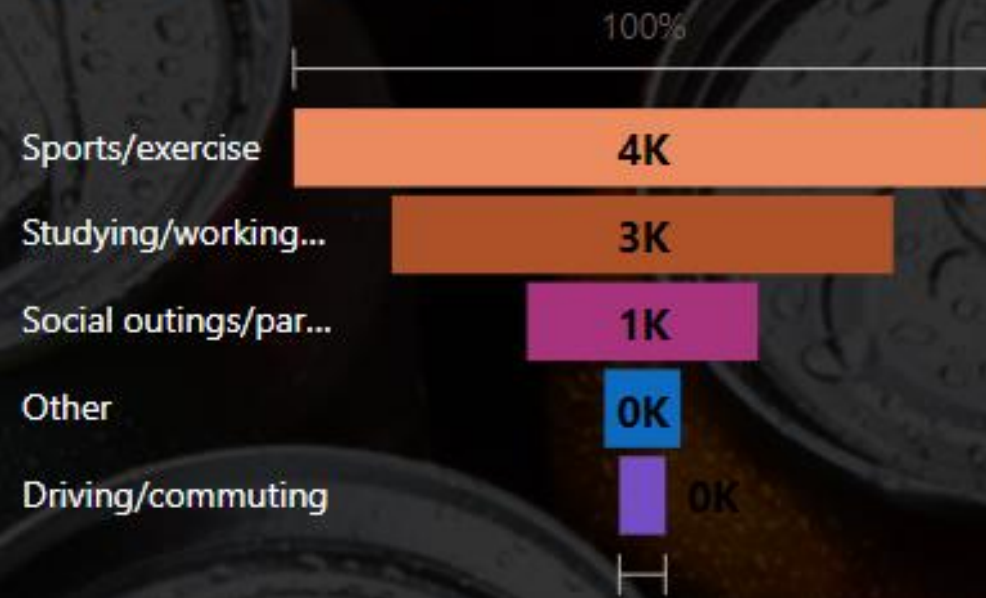
Brand perception

```sql
WITH codex_responses AS (
    SELECT r.Age, r.Gender, s.Brand_perception
    FROM dim_repondents r
    JOIN fact_survey_responses s ON r.Respondent_ID = s.Respondent_ID
    WHERE s.Current_brands = 'CodeX'
)
SELECT
    Age, Gender,
    SUM(CASE Brand_perception
        WHEN 'Positive' THEN 1
        WHEN 'Neutral' THEN 0
        ELSE -1 END) AS sum_brand_perception
FROM
    codex_responses
GROUP BY
    Age, Gender
ORDER BY
    Age ASC,
    CASE WHEN Gender = 'Male' THEN 0
         WHEN Gender = 'Female' THEN 1
         ELSE 2
    END :
```

| Age	Gender	sum_brand_perception
15-18	Male	-4
15-18	Female	2
15-18	Non-binary	-1
19-30	Male	12
19-30	Female	21
19-30	Non-binary	4
31-45	Male	20
31-45	Female	-4
31-45	Non-binary	0
46-65	Male	0
46-65	Female	-5
65+	Male	0
65+	Female	2
65+	Non-binary	0
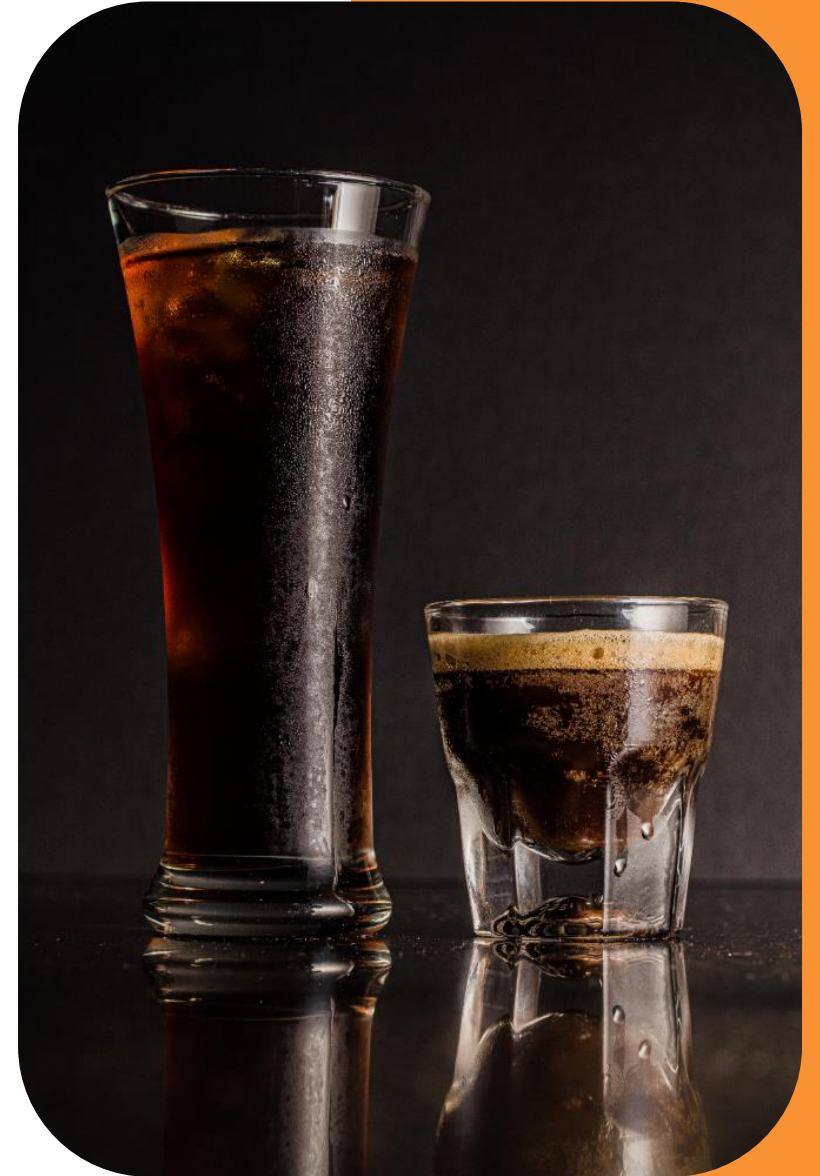
```sql
SELECT
    Reasons_preventing_trying,
    COUNT(*) AS reason_count
FROM
    fact_survey_responses
WHERE
    Current_brands = 'CodeX'
GROUP BY
    Reasons_preventing_trying
ORDER BY
    reason_count DESC;
```

Reasons_preventing_trying	reason_count
Health concerns	233
Not available locally	229
Not interested in energy drinks	210
Unfamiliar with the brand	190
Other	118

Total Respondents by Reasons for choosing brands
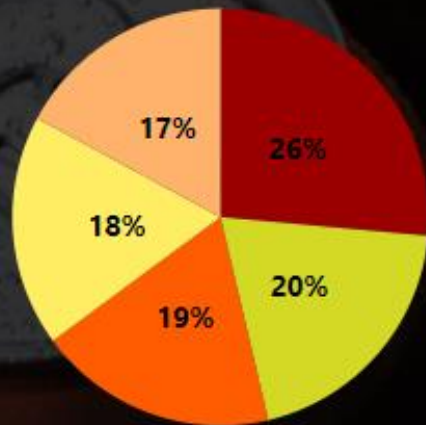
Reasons_for_choosin...
- Brand reputation
- Availability
- Taste/flavor prefere...
- Effectiveness
- Other

26% 20% 19% 18% 17%

Total Respondents by Reasons preventing trying
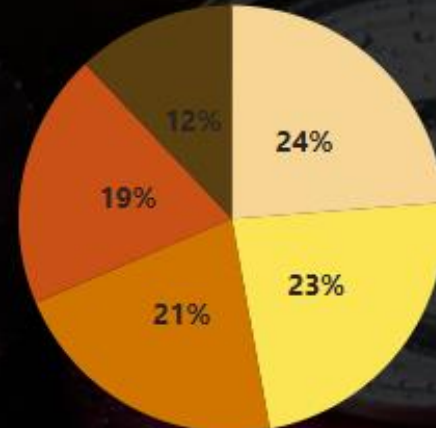
Reasons_preventing_try...
- Health concerns
- Not available locally
- Not interested in ene...
- Unfamiliar with the b...
- Other

24% 23% 21% 19% 12%

Total Respondents by Interest in natural or organic

2K (19.55%)
5K (49.83%)
3K (30.62%)

Interest_in_natural_or_orga...
● Yes
● No
● Not Sure

```sql
SELECT
    Interest_in_natural_or_organic,
    COUNT(*) AS interest_count
FROM
    fact_survey_responses
GROUP BY
    Interest_in_natural_or_organic
ORDER BY
    interest_count DESC;
```

Interest_in_natural_or_organic	interest_count
Yes	4983
No	3062
Not Sure	1955
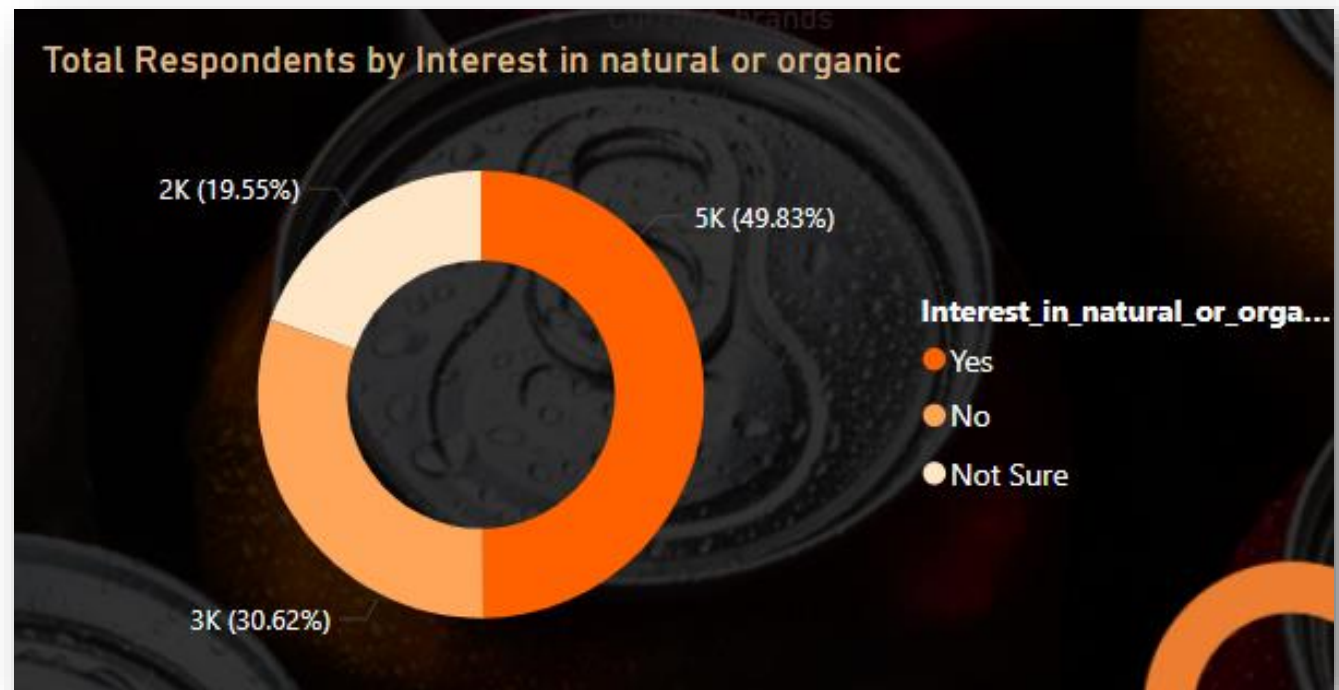
```sql
WITH codex_responses AS (
    SELECT
        s.Current_brands,
        s.Taste_experience
    FROM
        fact_survey_responses s
    WHERE
        s.Taste_experience IS NOT NULL
)
SELECT
    current_brands,
    AVG(Taste_experience) AS avg_taste_experience
FROM
    codex_responses
GROUP BY
    current_brands
ORDER BY
    avg_taste_experience DESC;
```

Current_brands	avg_taste_experience
Others	3.3236
Cola-Coka	3.3018
Blue Bull	3.2977
Sky 9	3.2952
Bepsi	3.2756
CodeX	3.2735
Gangster	3.2395

Recommendation

To boost Codex's sales, the marketing team should focus on the following:

- **Enhance Brand Perception**: Build a distinctive brand identity by highlighting unique benefits and promoting transparency regarding natural ingredients and health benefits.

- **Address Health Concerns**: Reformulate products to reduce sugar and emphasize natural ingredients, appealing to health-conscious consumers with clean labeling.

- **Expand Product Range**: Introduce new flavors and caffeine-free options to cater to diverse consumer preferences and emerging trends.

- **Optimize Pricing and Distribution**: Align pricing with consumer expectations and enhance product availability through expanded distribution channels and promotions.

- **Leverage Effective Marketing Channels**: Invest in targeted digital advertising and create engaging content to educate consumers and reach specific audience segments.

- **Increase Engagement with Key Demographics**: Tailor campaigns for Tier 1 cities and develop educational initiatives for Tier 2 cities to boost awareness and engagement.

- **Enhance Customer Experience**: Implement feedback channels and loyalty programs to improve products and build strong customer relationships.

- **Focus on Key Consumption Situations**: Promote products for sports, exercise, and productivity, partnering with influencers and targeting relevant consumer activities.
- These strategies will help Codex align its marketing efforts with consumer preferences, address key concerns, and strengthen its market presence.

Thank You!