**Plan Of Action**

1. **Project Kick-off and Understanding:**
   * Review the project description and provided metadata thoroughly.
   * Understand the business objectives and goals of CodeX in the Indian market.
   * Identify the key questions that need to be answered to achieve these objectives.
2. **Data Exploration and Cleaning:**
   * Examine the three provided CSV files: dim\_respondents, dim\_cities, and fact\_survey\_responses.
   * Check for missing data, duplicates, and outliers.
   * Address any data quality issues through data cleaning and preprocessing.
3. **Data Integration:**
   * Merge relevant columns from dim\_respondents, dim\_cities, and fact\_survey\_responses to create a consolidated dataset.
   * Ensure data consistency and coherence across all merged columns.
4. **Initial Data Analysis:**
   * Perform basic statistics and exploratory data analysis (EDA) to understand the dataset's characteristics.
   * Generate summary statistics, histograms, and correlation matrices.
   * Identify patterns, trends, and initial insights.
5. **Answering Key Research Questions:**
   * Use the survey questions and metadata to define research questions that align with CodeX's objectives.
   * Analyze the data to answer these research questions, including:
     + Demographic insights about respondents (age, gender, city, etc.).
     + Frequency and timing of energy drink consumption.
     + Reasons for consuming energy drinks.
     + Brand perception and preferences.
     + Factors affecting the choice of energy drink brands.
     + Desired improvements in energy drinks.
     + Health concerns and interest in natural/organic ingredients.
6. **Data Visualization:**
   * Create informative and visually appealing charts, graphs, and plots to illustrate key findings.
   * Use tools like Excel or Power BI to generate visualizations.
   * Ensure that visualizations support the analysis and are easy to interpret.
7. **Data Interpretation:**
   * Interpret the insights gained from the data analysis.
   * Identify significant patterns, correlations, and outliers.
   * Formulate conclusions and actionable recommendations based on the analysis.
8. **Insights and Recommendations Presentation:**
   * Prepare a report that includes the following sections:
     + Introduction (briefly explain the project and objectives).
     + Data description and cleaning process.
     + Methodology (how research questions were addressed).
     + Key insights and findings.
     + Recommendations for CodeX:
       - Marketing strategies to increase brand awareness.
       - Product development suggestions.
       - Target audience considerations.
       - Pricing and packaging recommendations.
     + Visualizations to support the analysis.
9. **Final Validation and Review:**
   * Review the report for accuracy, clarity, and completeness.
   * Validate the recommendations against the insights obtained.
   * Ensure that the report addresses all aspects of the project.
10. **Presentation and Delivery:**
    * Present the findings and recommendations to the CodeX marketing team.
    * Prepare a concise and engaging presentation that highlights the most critical insights.
    * Be ready to answer any questions or provide additional details during the presentation.
11. **Documentation and Reporting:**
    * Document the entire analysis process, including code (if applicable) and data transformations.
    * Provide the cleaned and merged dataset for reference.
    * Summarize the project in a clear and concise report for future reference.
12. **Feedback and Iteration:**
    * Gather feedback from the CodeX team on the insights and recommendations.
    * Be open to making adjustments or conducting further analysis if necessary.

**Steps:**

1. Pasted all data into new file, **“Excel Portfolio Project.xlsx”.**
2. Made a new sheet for report **“Report”.**
3. Made a new sheet, **“Combined Data”** to get a dataset for complete observations
4. Checked for duplicates, none found.
5. Sorted **dim\_respondents** and **Fact\_survey** sheets from small to large against Respondent ID (Unique ID)
6. Copied **dim\_respondents** data to **Combined Data**.
7. Turned **dim\_cities** data to table named **“city\_data”**
8. To get **Cities** against city ID, I used VLOOKUP function. The formula was **=VLOOKUP(G2,city\_data,2, FALSE)**
9. To get **Tier** against city ID, I used VLOOKUP function. The formula was **=VLOOKUP(G2,city\_data,3, FALSE)**

|  |  |
| --- | --- |
| **Consume\_frequency** | **Freq** |
| 2-3 times a week | Frequently |
| 2-3 times a month | Monthly |
| Rarely | Rarely |
| Daily | Frequently |
| Once a week | Frequently |

1. Changed the following columns according to specific categories using **VLOOKUP**

|  |  |
| --- | --- |
| **Age** | **Age Group** |
| 15-18 | Teenagers |
| 19-30 | Youngster |
| 31-45 | Adults |
| 46-65 | Elderly |
| 65+ | Senior Citizens |

|  |  |
| --- | --- |
| **Price\_range** | **Price Category** |
| 50-99 | Average |
| 100-150 | High |
| Above 150 | Expensive |
| Below 50 | Cheap |