**CodeX Insights: A step-by-step details of the project**

1. **Understanding the dataset:** The dataset provided had 3 files.
   1. **fact\_survey\_responses.xlsx**: The file contained responses of the 1000 respondents
   2. **dim­­\_respondents.xlsx**: It contained the personal details of the respondents like name, age, gender and cityID
   3. **dim\_cities.xlsx**: It contained the information regarding which cityID belonged to which city based on the name of the city
2. **Planning the approach:** The project started with importing all the important Python libraries and reading all the three files and understanding how to merge the datasets and create clean datasets with respect further Exploratory Data Analysis could be performed and the questions could be answered.
3. **Data Cleaning:** Upon inspection a few things were realised:
   1. The city names and city tiers were not mapped on the main dataset **fact\_survey\_responses.xlsx** which contained all the responses. It was solved by creating a ***dictionary*** with cityID as a ***key*** and their respective city names and tiers as ***values*** from the dataset **dim\_cities.xlsx** and mapping this to the main dataset.
   2. The column related to ***consumer frequency*** was not standardized. The responses included rarely, 2-3 times a week, once a week and so on. It was standardized as per monthly consumption. A new column ***monthly consumption*** was created. Rarely was converted to 1, 2-3 times a week was converted 10 and so on a standardized values were given.
   3. Rating column had the values 1-5 which were converted to 1 as Poor and 5 as excellent and so on by creating a dictionary and mapping it to a new column.

Note: Here it was assumed that these things were discussed with the Marketing team and upon discussions these number and standardizations were approved.

1. **Exploratory Data Analysis:** The questions asked in the competition were answered in this stage.
2. **Executive Summary:** An executive summary for the stakeholders to quickly go through providing details of my own recommendations was prepared.
3. **Presentation:** The presentation was created in which stakeholders were informed of the findings of the project in detail.

**Note:** The presentation link is in my LinkedIn profile. It was also a part of the task to present your findings to the team in under 15 minutes.