Data Challenge Proposal(DC21067)

Overview of Dataset:

The dataset chosen for the Data Challenge is Packaged Meals. The dataset contains the information regarding frozen foods nutrient information in different meal categories. The goal is to analyze this data and provide valuable insights.

Data Analysis:

The team is focussed on analyzing two things. First, the most popular ingredients used in each meal category and overall. The popularity is measured on the basis of percentage of times ingredients are used in each food in the category to the total number of frozen foods in the category. Second, different combinations of ingredients used to prepare the food.

Implications from Data Analysis:

By knowing the most popular ingredients in each frozen food category, prioritization to manufacture the foods with more popular ingredients can be achieved. This also indicates the choice of the consumer. By prioritization one also gets to know the least popular ingredients which can be avoided. By getting popular combinations of the food ingredients, the manufacturer can manufacture new variants of the frozen food to increase the sales and gain profit.

Additional analysis like categorizing each category into vagan, vegetarian and non-vegetarian and performing the same analysis can further help the prioritization process as the manufacturers can target specific groups of consumers knowing the popular ingredients. This information can also help in decreasing the production and manufacturing process in case of unprecedented solutions like Covid-19 by manufacturing only foods with popular ingredients in vegan, vegetarian and non-vegetarian categories and still satisfying the needs of all the people.

Analysis Tools:

- Programming Language: Python version 3.8 is the programming language used for data analysis.

- Data analysis library: Pandas is used to access and process the excel sheet data.

- Visualization library: Matplotlib library is used to provide visualization for the data in the form of bar and pie charts.