

LODESTONE TAKE HOME PROJECT

SUBMITTED BY

ANKIT JAIN

DATA SET DESCRIPTIONS:

- 1) **SheetA.csv**: This dataset is mainly about the two grill types we have, Charcoal and Propane.

grill Type: the 2 types of grills this data set is about, charcoal and Propane.

pre heat time: amount of time that the particular grill type takes to pre heat.

fuel cost per meal: the cost of fuel for that grill type for each meal.

market share: the market share or more popular grill type in the market.

easiness to use: the ease of use for the particular grill type.

life span year: the life span of the grill type

initial investment: the initial capital that will be required for that grill type

- 2) **SheetB.csv**: this dataset is an extension of sheet A where there are individual items like Ground Beef Patty, Hotdog and Veggie Patty, each made/grilled on a charcoal device and a propane based device.

item material: the individual food item amongst Ground Beef Patty, Hot dog, and Veggie Patty.

grill type: the particular grill type for each food item taken into consideration.

user satisfaction: the user_satisfaction score based on its usage for a particular food item per food item.

fuel cost: the fuel cost that is associated with that grill type for a particular food item

- 3) **SheetC.csv**: this sheet tells us if the user could guess which grill was used for a particular food item for the particular grill type and related score.

sample_item_index: the sample number

item_material: the individual food item

is_frozen: is the food frozen

grill_type: the grill type in associated with the food item

thumbs_up_score: the taste score for that food item for the particular grill

guess_grill_correct: if or not the user guessed it right

TRANSFORMATIONS PERFORMED:

- The data files have been read using Pandas read_csv method.
- The data within the files have been grouped by some column again using Python Pandas using the 'groupby' function and finally, summed some columns using the sum function after the groupby.
- Matplotlib library has been used to perform visualization tasks to produce single bar charts and also multiple bar charts as per requirements.
- Seaborn library has also been used to perform some visualizations in parallel to the matplotlib library.
- In one of the tasks, to create an aggregated dataset, we have merged 3 datafilea as one on some required attributes, grouped on grill_type and item_material and summed on various numerical attributes.
- The function to merge different tables is merge function.
- The task where we had to convert a json file to csv, we had a list of multiple key values and the values itself were key-values in some cases within the json data. We looped to get rid of some attributes that were not required, some keys that were not required and only retrieved the data points that were needed as in SheetC.csv. This was done iteratively using 2 loops.

PROVIDE ANY SCRIPTS/PROGRAM THAT YOU HAVE USED AND TELL US YOU'RE FINDINGS:

The scripts are written in python and have used the Pandas, Matplotlib and Seaborn library to solve the questions. I have attached an additional Jupyter notebook file that contains the code for all questions and also what are the findings for each question and also explaining any assumptions taken.