

FinalProject_Step1_PuppalaSucharitha

2022-07-31

Identify a topic or a problem that you want to research. Provide an introduction that explains the problem statement or topic you are addressing. Why would someone be interested in this? How is it a data science problem?

1.INTRODUCTION

I have selected the topic of Coffee topic for the research. I am a big lover of Coffee and interested in knowing the consumption of Coffee. While researching about the Coffee topic I found that the Coffee is mostly consumed in USA.About 66% of the Americans drink coffee every day which is more than the tap water.I would like to use the data sciences for analyzing the data sets collected and to know the consumption of Coffee in USA.

2.RESEARCH QUESTIONS

The following are 5 questions that I would like to research:

a.) From the data sets collected what are the data elements available. b.) What is the reason for the high consumption of coffee? c.) Is there any additional data needed for the analysis? d.) Which category of people consume more amount of coffee? f.) Which category of drink has the highest amount of Caffeine?

3.APPROACH

I would like to use the regression models for the analysis.

4.HOW YOUR APPROACH ADDRESSES (FULLY OR PARTIALLY) THE PROBLEM.

Multiple regression analysis will be performed on each category of caffeinated beverages (coffee, tea, energy drinks, soft drinks and others) in order to determine the effectiveness of the reason for the coffee consumption. Multiple regression analysis will be performed simultaneously for all categories.

5.DATA (MINIMUM OF 3 DATASETS - BUT NO REQUIREMENT ON NUMBER OF FIELDS OR ROWS)

A. Dataset 1 - caffeine.csv

<https://www.kaggle.com/datasets/heitornunes/caffeine-content-of-drinks> This dataset contains the data of different drinks and their caffeine content values, calories and type of drink.

B. Dataset 2 - Coffee Chain.csv

<https://data.world/adsafj/coffee-shop/workspace/file?filename=Coffee+Chain.csv> This dataset contains the consumption of coffee and other drinks in different areas in USA

C. Dataset 3 - Coffee_Survey.csv https://data.world/junyu-wei/assignment-1-data-collection-and-curation/workspace/file?filename=Coffee_Survey.csv This data set contains the data of the survey of consumption of coffee, no. of cups etc.

D. dataset 4 - CoffeeConsumption.csv

<https://www.kaggle.com/datasets/nurielreuve/coffee-consumption-by-country-2022> This is the data set of the country wise consumption of Coffee worldwide.

6. REQUIRED PACKAGES

The following are the packages that are required for the analysis of the data

ggm,ggplot2,readr,pastecs,readxl,plyr,dplyr,magrittr,purrr,stringr,QuantPsyc

7. PLOTS AND TABLES NEEDED

The following are the plots that are required for the analysis of the data

Scatter plots, histograms, line graphs

QUESTIONS FOR FUTURE STEPS

How to predict the different samples of the data.