Week 7: Deliverables

Group Name: Fight on Healthy diet
Name: Sijing Liu
Email: sijingli@usc.edu
Batch code: LISUM13: 30
Country: U.S.
College: University of Southern California
Specialization: Data Science

Problem description:

Does Healthy Diet Help Prevent COVID-19?

On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic. Since then, the worldwide recorded death rate as a result of the illness has surpassed five million (Roberts, 2021). Research shows that the epidemic growth rate for disease spread depends on many factors, including biological, demographic, and social factors. However, dietary risks during the pandemic are void of investigation, given the acknowledged impact of food on health outcomes.

According to the World Health Organization (WHO), eating a healthy diet is very important during the COVID-19 pandemic (WHO, 2021). Now more than ever, we need to prioritize what we put into our bodies to reduce the susceptibility to and long-term implications from the illness. The relationship between dietary habits and diseases has been extensively investigated. However, most of the associations focus on chronic non-communicable diseases (Afshin et al., 2019). Therefore, through this project, I aim to fill this void to make clear the relationship between dietary habits with communicable disease, like COVID-19.

Overall, this project will look into a dataset that measures the nutrition of several food groups, a variety of eating styles, obesity and undernourishment rates, and data on COVID-19 cases from 170 countries. I hope to conduct exploratory data analysis (using descriptive statistics), machine learning (mainly association analysis and prediction), and data visualization to learn more about how diet ultimately influences the contraction and survivability rates of COVID-19. My main objective is to answer the following questions: **Are countries with healthier eating habits less impacted by COVID-19? Does a healthy diet ultimately help prevent COVID-19?**

Business understanding:

The proposal above is mainly focusing on the perspective of academic understanding. When it comes to the perspective of business understanding, the

evidence of data relevance can highlight the importance of healthy diet and thus offer possible marketing strategies for specific food companies.

Project lifecycle along with deadline:

Main part	Description	Deadline
Data Collecting & Understanding	- to get needed raw dataset & supplement more interested data based on the problem framed - by searching official websites/ platforms (Kaggle)	Week 7 & 8

Data Preprocessing	- to preliminary organize my own datasets (reduce redundant data & update data & handle missing data)
	   - by conducting general exploratory data analysis & data visualization using Excel/Tableau
	- to further categorize food data
	   - by checking relevant published literature databases to get domain knowledge & using Excel to deal with personalized categories (healthy and unhealthy food)
		Week 9

EDA	- to explore possible relationships between different food and obesity/undernourished rate & check whether countries with healthier eating habits can be less impacted by COVID-19	Week 10 & Week 11
Data Mining (Modeling)	- by conducting clustering & classification & regression analysis using Python/R, mainly constructing and fine-tuning models based on K-Means algorithm & K-Nearest Neighbors (KNN) algorithm	Week 11 & Week 12
Data reporting	- to clearly conclude and show the findings & review the limitations - by organizing the data visualization and key results of models & presenting them in the final project report	Week 13

Data Intake report:
Name: COVID-19 Healthy Diet Dataset
Report date: 19-Oct-2022
Internship Batch: LISUM13: 30
Version: 1.0
Data intake by: Sijing Liu
Data storage location: https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?resource=download

Tabular data details: Fat_Supply_Quantity_Data

Total number of observations	170
Total number of files	1
Total number of features	32
Base format of the file	.csv
Size of the data	42+ KB

Tabular data details: Food_Supply_Quantity_kg_Data

Total number of observations	170
Total number of files	1
Total number of features	32
Base format of the file	.csv
Size of the data	44+ KB

Tabular data details: Food_Supply_kcal_Data

Total number of observations	170
Total number of files	1
Total number of features	32
Base format of the file	.csv
Size of the data	43+ KB

Tabular data details: Protein_Supply_Quantity_Data

Total number of observations	170
Total number of files	1
Total number of features	32
Base format of the file	.csv
Size of the data	43+ KB

Tabular data details: Supply_Food_Data_Descriptions

Total number of observations	23

Total number of files	1
Total number of features	2
Base format of the file	.csv
Size of the data	3+ KB

Github Repo link:

https://github.com/Sijing98/Internship22Fall/tree/main/Project%20-%20Fight%20on%20Healthy%20diet