Data Description

I have transformed the *all\_commodities.csv*, *all\_stock\_and\_etfs.csv*, *acs\_5yr\_selected\_economic\_characterics\_2010-2022.csv*, *Nutrition\_Physical\_Activity\_and\_Obesity\_Data.csv*. I will describe each transformed output here.

# *all\_commodities.csv*

src: code/data\_cleaning\_code/data\_transformation\_commodity.ipynb

output: code/cleaned\_data/commodity/coffee\_sugar\_prices.csv

The output csv records coffee and sugar prices for each day.

# *all\_stock\_and\_etfs.csv*

src: code/data\_cleaning\_code/data\_transformation\_stock.ipynb

output: code/cleaned\_data/commodity/Nutrition\_Physical\_Activity\_and\_Obesity\_Data.csv

The output csv records each stock’s close\_price, volume, log\_daily\_return and log\_monthly\_return for each day.

# *acs\_5yr\_selected\_economic\_characterics\_2010-2022.csv*

src: code/data\_cleaning\_code/data\_transformation\_survey.ipynb

output: code/cleaned\_data/survey/\*

These data record the percentage distribution of people across different categories over time.

Specifically, for each kind, I calculated the data for each state and the whole US country.

* POVERTY

record how much people are under the poverty level

* INCOME

record income level of different kinds(has been adjusted by yearly inflation)

* EMPLOYMENT

record the employment rate of different people

* INDUSTRY

record the distribution of industry that people work in

* OCCUPATION

record the distribution of occupations

* CLASS OF WORKER

record the distribution between different class of workers

* COMMUTE TO WORK

record the way people commute to work

* HEALTH INSURANCE COVERAGE

record health insurance coverage across different kind of people

# *Nutrition\_Physical\_Activity\_and\_Obesity\_Data.csv*

src: code/data\_cleaning\_code/data\_transformation\_nutrition.ipynb

output: code/cleaned\_data/questions/\*

There are six kinds of Questions people suffer from, which are:

* FruitConsuming

Not eating enough fruit

* VegetableConsuming

Not eating enough vegetable

* Obesity

suffer from Obesity

* SoftDrink

drink too much soft drink

* PhysicalActivity

not having enough physical activity

* TelevisionWatching

watching too much TV

for each csv, the filename corresponds to its location.

Inside the csv, each column indicates a specific type of person. For example, “AGERY,AGEYR1824” means this column records the data for people between 18 and 24 years old.

Detail can be referred to:

AGEYR,AGEYR2534: 25 - 34

AGEYR,AGEYR3544: 35 - 44

AGEYR,AGEYR1824: 18 - 24

AGEYR,AGEYR65PLUS: 65 or older

AGEYR,AGEYR5564: 55 - 64

AGEYR,AGEYR4554: 45 - 54

EDU,EDUCOTEC: Some college or technical school

EDU,EDUHS: Less than high school

EDU,EDUHSGRAD: High school graduate

EDU,EDUCOGRAD: College graduate

GEN,MALE: Male

GEN,FEMALE: Female

GRADE,GRADE12: 12th

GRADE,GRADE10: 10th

GRADE,GRADE09: 9th

GRADE,GRADE11: 11th

INC,INC3550: $35,000 - $49,999

INC,INC5075: $50,000 - $74,999

INC,INC75PLUS: $75,000 or greater

INC,INCLESS15: Less than $15,000

INC,INC1525: $15,000 - $24,999

INC,INCNR: Data not reported

INC,INC2535: $25,000 - $34,999

OVR,OVERALL: Total

RACE,RACENAA: American Indian/Alaska Native

RACE,RACE2PLUS: 2 or more races

RACE,RACEWHT: Non-Hispanic White

RACE,RACEBLK: Non-Hispanic Black

RACE,RACEHIS: Hispanic

RACE,RACEHPI: Hawaiian/Pacific Islander

RACE,RACEASN: Asian

RACE,RACEOTH: Other