Citadel Datathon 2024

**Goal:** Analyze the provided datasets, potentially in combination with supplementary datasets, in order to better understand the footprint of processed food in the United States

**Task:** Pose your own question and answer it

**Data Categories**:

1. Food Production
2. Health
3. Economic Data
   * financial market
   * price of products

## Analysis Date Range

1999-2024

## Report Outline

### 1. Abstract [Scott]

The US economy has been flourishing since the 2008 financial crisis. Without a shadow of a doubt, prosperity is always one of the goals that every government fights for. Nevertheless, some may have noticed affluence is also associated with some illnesses which are also known as diseases of affluence. Office workers may order fast food so as to strive for more time to work. Some may just want to be lazy.

.

Origin in the US, fast food may be a friendly choice to many citizens. Just go to a fast food store within a walkable distance, delicious burgers, french fries and coke will be prepared in a minute. One ignores the issue of health, processed meat may be very tempting. With this culture, related corporations, like fast food chains and processed meat firms, are having strong financial health.

Unfortunately, without paying much attention on their diet habit and healthy lifestyles, obesity is getting prevalent. As everyone know, having more meat and less vegetables will lead to obesity which can be linked to many illnesses including diabetes.[1] Thus, it is of utmost importance to fight against obesity. Otherwise, the cost to the medical system and even to the society will be overwhelming.

To further investigate the relationship between and explore the effect of increasing processed meat production, we conduct a research to find out how they affect our society. Is it benefiting us or just bring incredible money to the companies’ owner?

[1] Klein S, Gastaldelli A, Yki-Järvinen H, Scherer PE. Why does obesity cause diabetes? Cell Metab. 2022 Jan 4;34(1):11-20. doi: 10.1016/j.cmet.2021.12.012. PMID: 34986330; PMCID: PMC8740746

### 2. Introduction [Mulan]

1. background
2. motivation

### 3. Non-Technical Summary(main questions and key findings)

### 4. Overall WorkFlow [Lihua draw a draft]

draw one overall graph (multi step EDA) to represent our workflow

### 5. Data Exploration and Preprocessing [Lihua]

(some features we propose and we use)

1. Meat Production Data
2. Financial Data
3. Health and Other Data

### 6. Research Questions

#### RQ1: how red meat production is influenced and influenced in the financial market?

meat production negatively related to Trades & Services stocks, “WEN”; positively related to Manufacturing and Technology stocks. (Not very strong with overall financial market)

Manufacturing and Technology Stocks => Meat production grow => meat price decrease => why? higher competition / less profit / cook themselves [Mulan]

time lag analysis

Causal analysis instead of just correlation

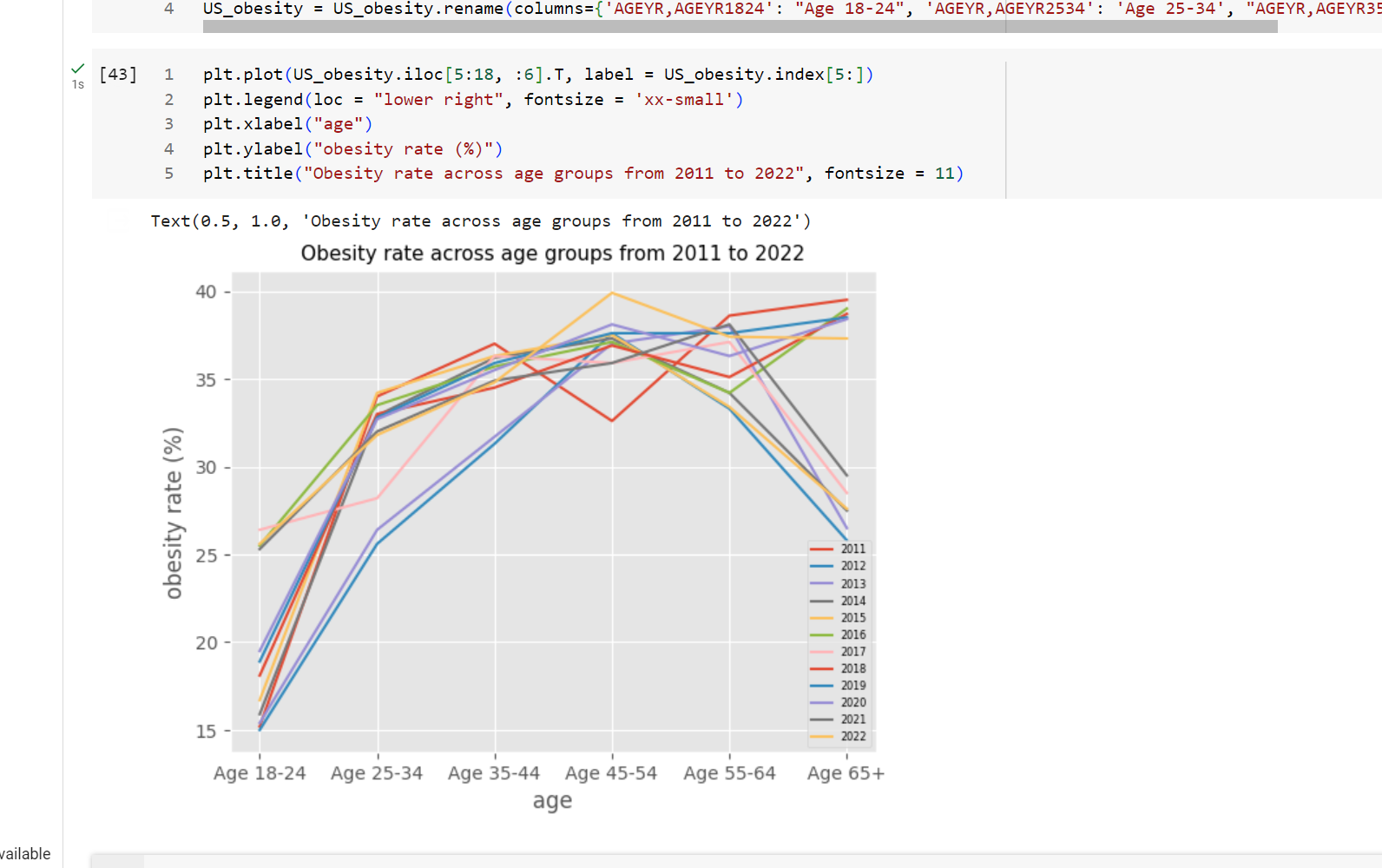
#### RQ2: Does the development of meat production equally impact everyone in the society?

1. Obesity [Scott]

conclusion: overall and for specific kinds of people

evidence:

Obesity rates are always the lowest for the youngest age group. Then for any other age group in all year, the rates never fall below 25%. The figures are very worrying, especially for the middle age group. In some years, the obesity rate drop for age 65+ group while it just maintains at a high level of 30%+ for the middle-age and elderly. This is obviously the result of the change in lifestyle after they graduate.



1. Unemployment [Richard]

conclusion: overall and for specific kinds of people

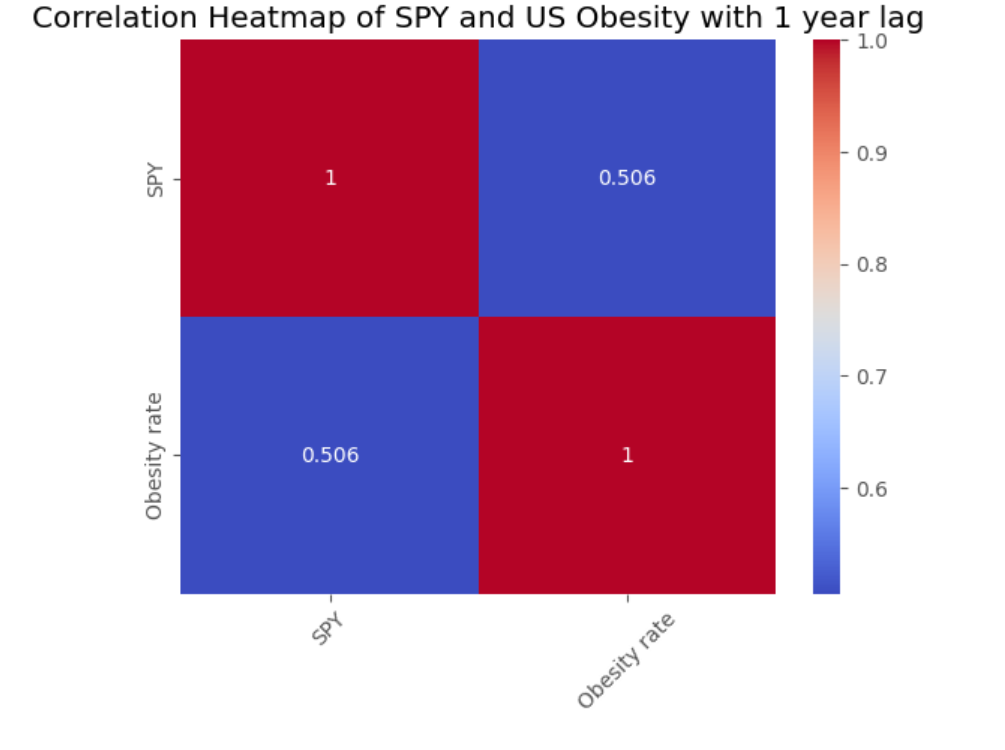
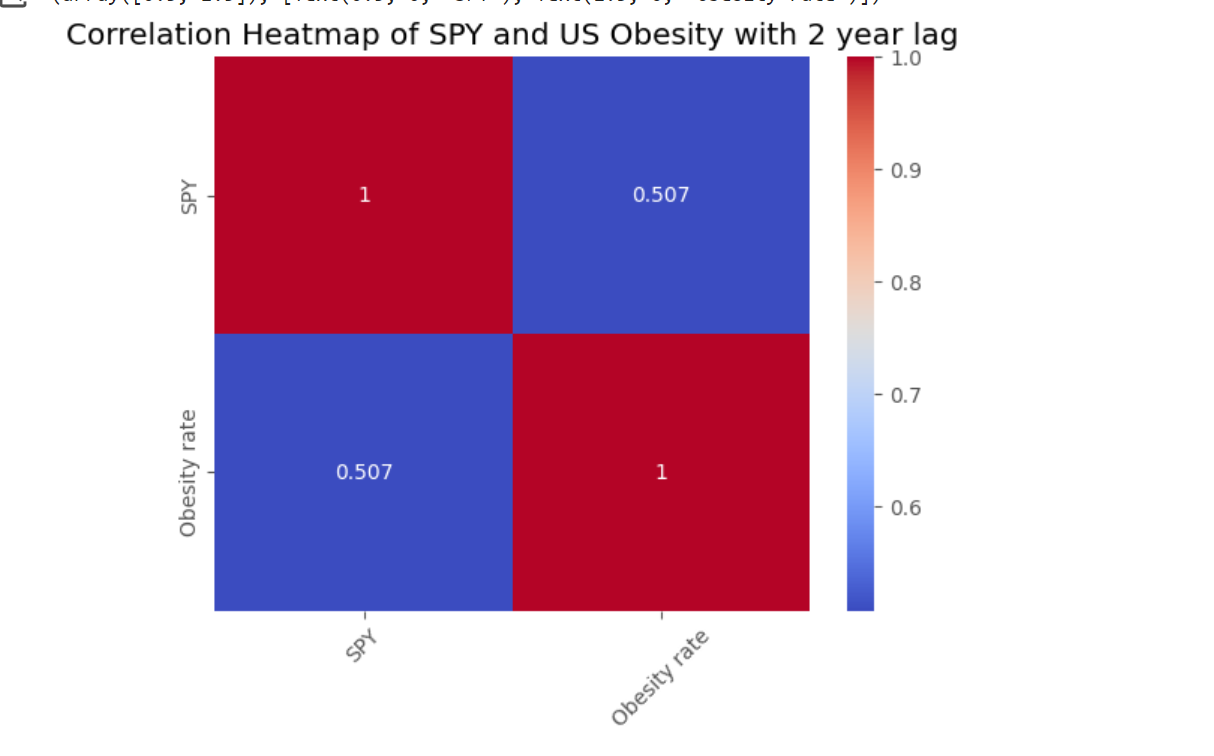
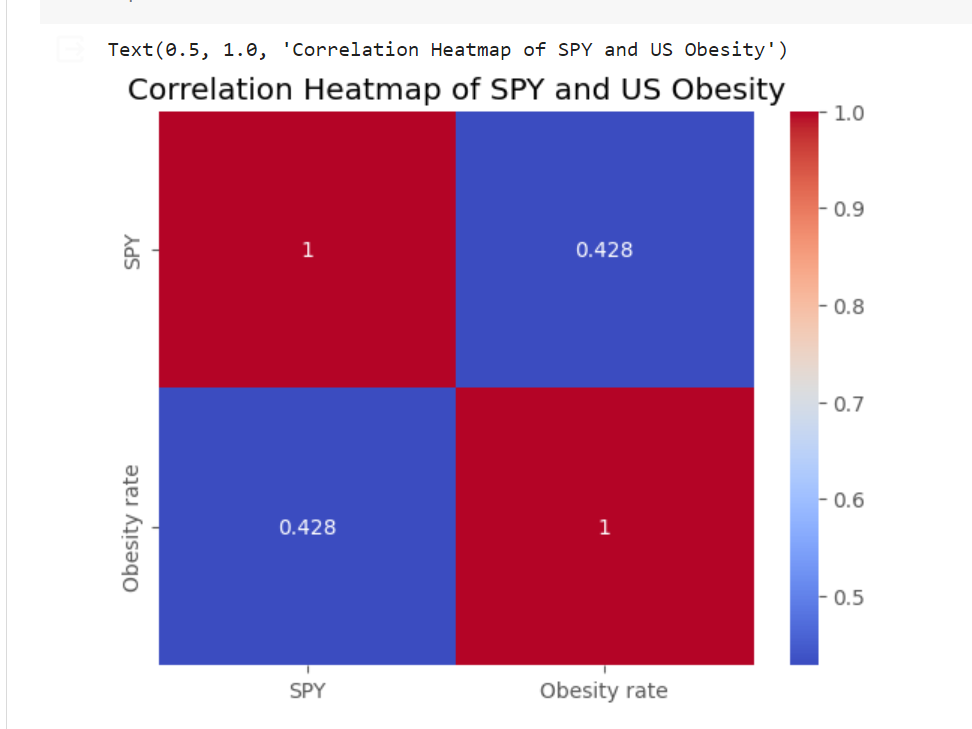
evidence

**股票和失业率的关系，PPI和失业率的关系**

1. general economic impact: S&P 500 positively correlated to red meat production/obesity

S&P 500 has a significant positive correlation with the overall obesity rate. It has a correlation coefficient of 0.428. As we suspect that the effect of the performance of the stock market may have a lagged effect on the obesity, we adjust it with 1 year and 2 year lag. The correlation correlation is indeed even more significant. This result is very reasonable because when people are getting richer, their purchasing power increases. They will then consume more, especially meat. However, they will not be obese immediately. Therefore, it may lag behind for a year.

In this regard, we consider the effect of yearly S&P 500 return versus the change in obesity rate instead of the obesity rate itself because we want to focus on the change in obesity level. We believe a change-to-change comparison can present the effect more accurately.



S&P 500 positively related to Obesity with some time lags.

middle-age: higher Obesity

meat production positively related to question of fruit and vegetable consumption

#### RQ3: What should the government do to help everyone share the benefit of the development of meat production?

### 7. Conclusion and Future Directions [Richard]

### 8. References

## Outline

1. Could red meat production be an indicator for the financial health of processed food companies, as reflected by their stock market performance?

* determine how to measure the meat production: RMI
* seasonal decomposition: explain the chosen of Red Meat
* Measure of the stock Market Change:

monthly

1. Does the development of meat production equally benefit everyone in the society?

* median income level across the year
* health insurance coverage rate
* unemployment rate
* obesity, overweight
* across different income, ages, years, genders, states etc.

1. What should the government do to help everyone share the benefit of the development of meat production?

## Questions

1. About people's health data
   1. formulate the health condition, defining what represent health
   2. how to predict
   3. how to maximize
2. trade-off between health and economy (index represents well-being of citizens) – may be too broad
   1. formulate the index
   2. how to maximize

Gary: Does development of the processed food industry really benefit xx(maybe some specific types) people?

(unemployment rate, health condition => GDP)

Suppose development of meat production can indicate development of financial market, is it a good thing for normal people

## Finalized Question

1. Could red meat production be an indicator for the financial health of processed food companies, as reflected by their stock market performance?

Lihua, Mulan

TSN -> stock.

positive? significance differ

statistical evidence

1. Could processed food companies' **stock** market also serve as predictors for general economic outcomes? Does this correlation vary across states, or among different ethnicity, income level, gender, or age groups? → if so, there may be a loop, food company do well → economic up → consume more → unhealthy → the company do well

Richard & Scott

data transformation 2024.03.22

median household income, every state, every year

median earning for workers, for every state and every year

health insurance coverage rate for every state and every year

problem of by state: each state have diff demographic

Case Example:

in a specific time range

Meat Production up => financial market grow => servers in the restaurant, physical activity down (About Fairness)

Note:

Adjust for overall trend and confounding variables:

* stock value: stock value of food companies/S&P, adjust for inflation and overall trend
* price of sugar fluctuates a lot
* Adjusted Value = Original Value \* (CPI in Base Year / CPI in Current Year)

TSN & Production has a high correlation

## Question 1 thoughts

Measure of Meat production:

Criteria: find the one that can reflect the overall meat supply (quantity of meat available for consumption) that is most relevant to obesity

find correlation between red meat total production quantity, poultry total production quantity and obesity rate -> obesity rate is more closely related to red meat production, and evidence suggests that poultry is healthier -> use red meat production as the measure of meat production

find correlation between production quantity, slaughter weight, and cold storage stocks.

Hypothesis: Cold storage weight is negatively correlated with supply. We want to find an index that reflects the current supply of meat available for consumption.

.

Proposed measure:

1. Red meat production index

RMI = (M\_t - min M)/(max M - min M) \* 100

where M = (dressed red meat weight\_t - red meat cold storage weight\_t + red meat cold storage weight\_t-1)

where t is the date

1. Obesity rate?

OBI = Obesity rate\_t \* 100

where t is the date

Candidate outcome measures as food market indicator:

1. Inflation adjusted stock price at each day
   1. This variable adjusts the stock closing price for inflation using the CPI, which is an accurate measure of stock price in real terms.
   2. food restaurant average stock closing price/CPI \* 100
   3. selected fast food restaurant average stock closing price/CPI \* 100
      1. Chipotle Mexican Grill Inc
      2. McDonald’s Corporation
      3. Yum! Brands Inc
2. proportion food restaurants stock transaction volume at each day
   1. This variable measures the level of trading activity in food restaurant stocks on a daily basis as a proportion of total transaction volume.
3. Inflation adjusted return at each day:
   1. Formula: (1 + Stock Return\_t) / (1 + Inflation) - 1
      1. Stock Return\_t= (closing price\_t - closing price\_{t-1}) / (closing price\_{t-1}) \* 100
   2. Inflation: (CPI level - 100) / 100

Note: can investigate the seasonal trends of poultry production

Assessing the predictive power of RMPI and OBI on return

Thus, we can compare RMPI with OBI and previous data for their prediction abilities.

In the end, we can get a table like

| Outcome measures | RMPI estimate | CI | OBI estimate | CI | R^2 |
| --- | --- | --- | --- | --- | --- |
| Inflation adjusted stock price |  |  |  |  |  |
| food restaurants stock volume |  |  |  |  |  |
| Inflation adjusted return |  |  |  |  |  |

## Question 2

target variable: GDP

1. impact on the stock markets
   1. prediction of the stock price
   2. we want this to be ultimate goal or health + econ?

* generally those stock of companies related to food grow more than SPY but some stock data only available in the recent year. e.g. Coke grow 20+X but SPY and Dow Jones only grow 4X
* They may only be listed recently.

Meeting Time: 8:00 pm, UTC+8

<https://oxfordre.com/economics/display/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-19>

## Tasks:

### 2024.03.19

* Data Cleaning
* ~~Gary: other data~~
* ~~Richard: meat data~~
* Questions Design: Mulan & Scott

### 2024.03.20

* First Question: Mulan & Lihua
* Second Question: Richard & Scott

# Final Tasks:

Content Writing:

* **Abstract + Conclusion**
* Key Findings
  + ~~Mulan for Q1~~
  + Lihua, Scott, Richard for Q2
* 4.2 Feature Engineering Mulan (describe the indexes and the log return of the stock)

1. about the index we propose
2. why choose to use yearly data (figure or data(average value of different months) are both OK) (Lihua)

because the meat production has strong fluctuation across different months, but stocks don’t have. Using yearly data can eliminate this effect

* ~~6.1 (Lihua)~~
* 6.2 (Scott)
* 6.3 (Richard)

Polish:(ref:https://tex.stackexchange.com/questions/85904/showcase-of-beautiful-title-page-done-in-tex)

* ~~Cover~~
* ~~Content Page~~

Figure Style Alignment:

from matplotlib import pyplot as plt

plt.style.use('ggplot')