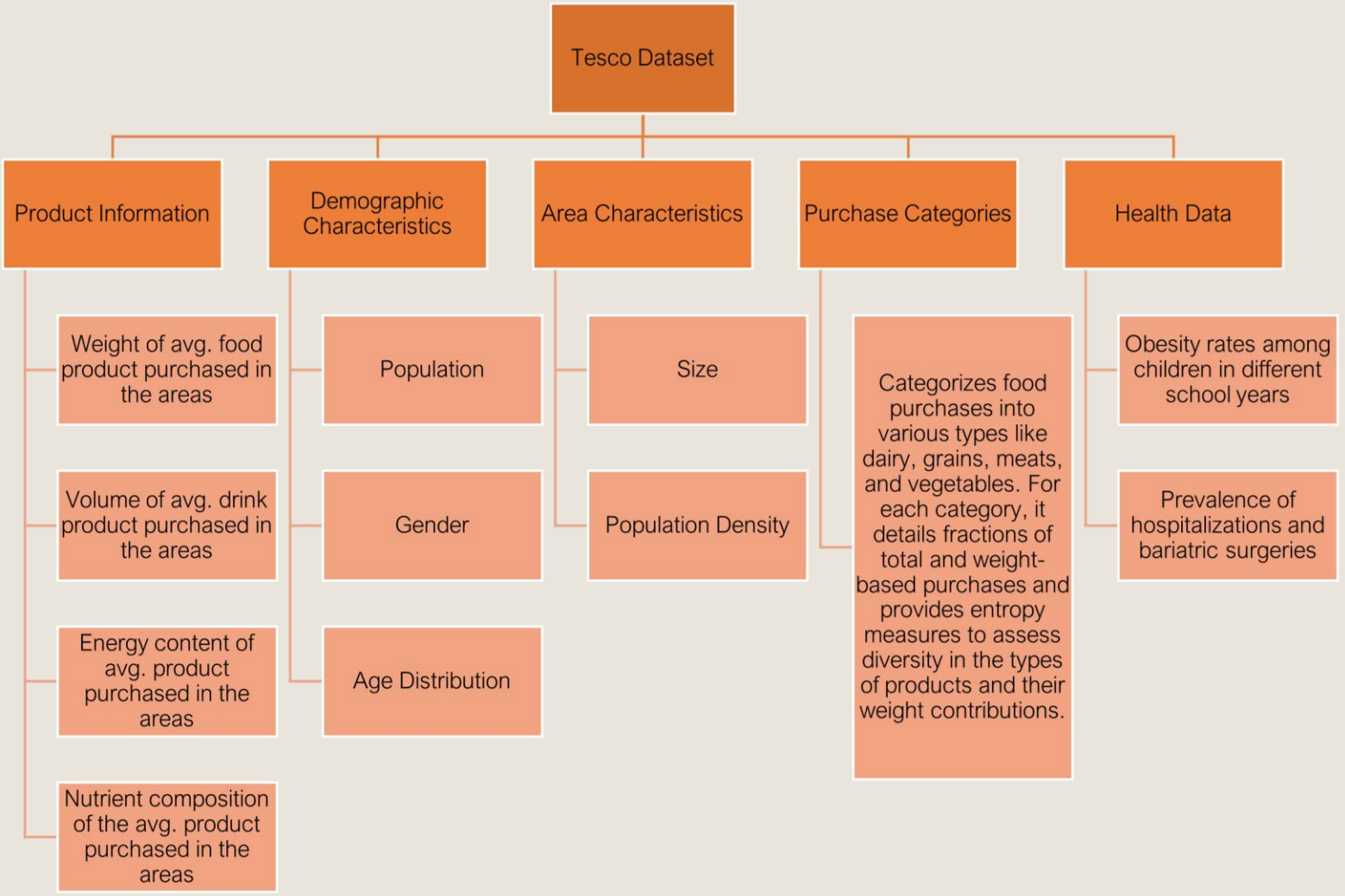
Tesco Grocery 1.0 dataset vs Income dataset

EXPLORING PURCHASING PATTERNS

Dataset Introduction – Tesco Grocery

- The Tesco Grocery 1.0 dataset contains multiple datasets recording 420 million food items purchases by the Tesco clubcard holders who have shopped over the course of the entire year of 2015.
- The datasets are aggregated at different levels of census areas (LSOA, MSOA, Borough and Ward) in accordance with the ONS requirements to preserve anonymity.
- The information for each area includes the number of transactions as well as the nutritional qualities of the typical food item purchased across several food categories such as sweets, fruits and vegetables, eggs, and fish.

Dataset Summary



Dataset Bias and Limitations

- Dataset Representation

  The representativeness of the data may be limited to Clubcard owners and does not represent the overall population.

- Temporal and Geographical Scope

  The dataset collects data of a specific timeframe and geographical areas which might not generalize well to other regions or periods.

- Health Outcomes Correlation

  Health-related outcomes in the dataset cannot be used to represent genuine health issue-causing factors.

- Nutritional Data Limitations

  Nutrient data are averages and may not accurately reflect the diversity of individual products.

Dataset Assumptions

- The data collected by Tesco is done in compliance to privacy laws and regulations.
- The data is collected from its own database and it is as accurate as it can be.

Insight 1

Do demographic factors like gender and age influence food purchasing patterns across different areas?

Purchasing Pattern – Gender Segment

Density Map for Gender Segmentation across London's Boroughs.

- Female – Dominated
- Male – Dominated
- Balanced

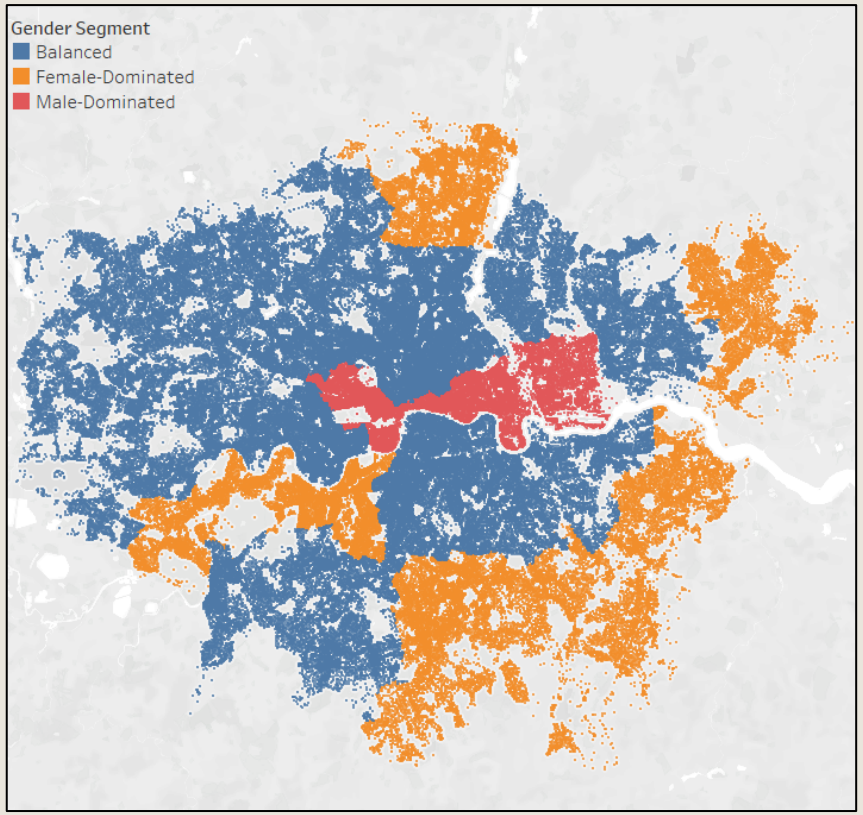


Figure 1: Density Map – Gender Segmentation

Purchasing Pattern – Gender Segment

- Dairy and Eggs: Higher consumption in male-dominated areas compared to female-dominated and balanced areas.
- Grains: Significantly higher in male-dominated areas (22.2%) compared to balanced and female-dominated areas.
- Readymade Foods: Higher consumption in female-dominated areas (6.6%) compared to male-dominated (5.4%) and balanced areas.
- Sweets: Slightly higher consumption in female-dominated areas compared to others.

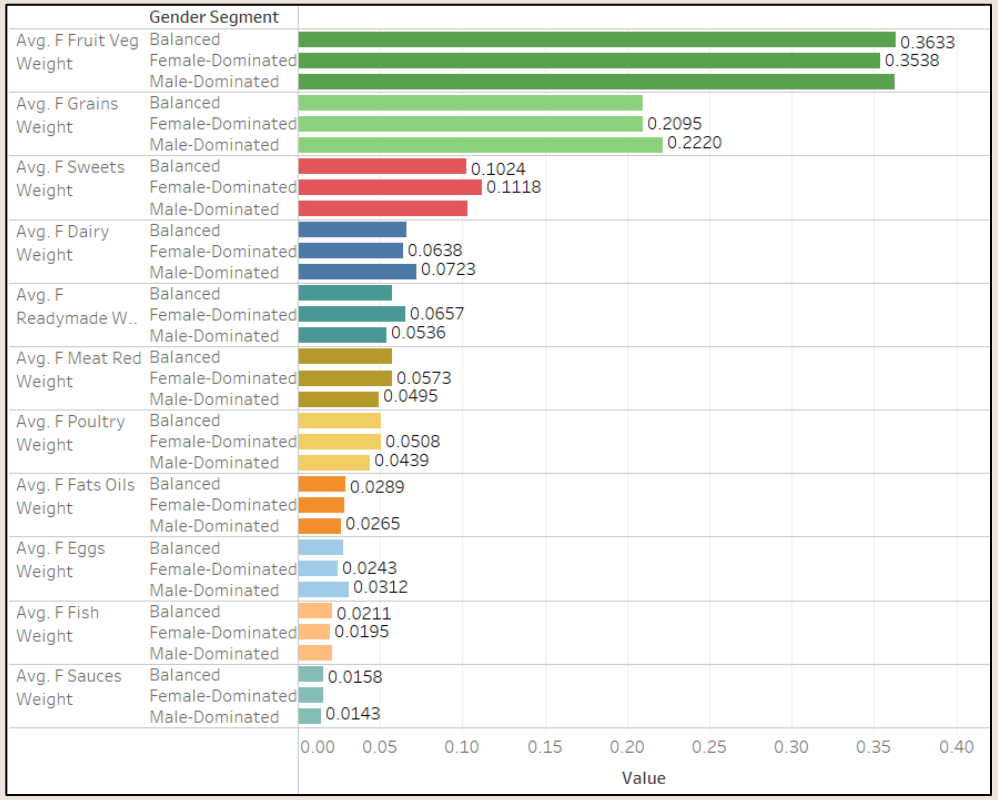


Figure 2: Bar chart for purchasing patterns according to gender Segmentation

Purchasing Pattern – Age Segment

Density Map for Age Segmentation across London's Boroughs.

- High Youth
- Low Youth

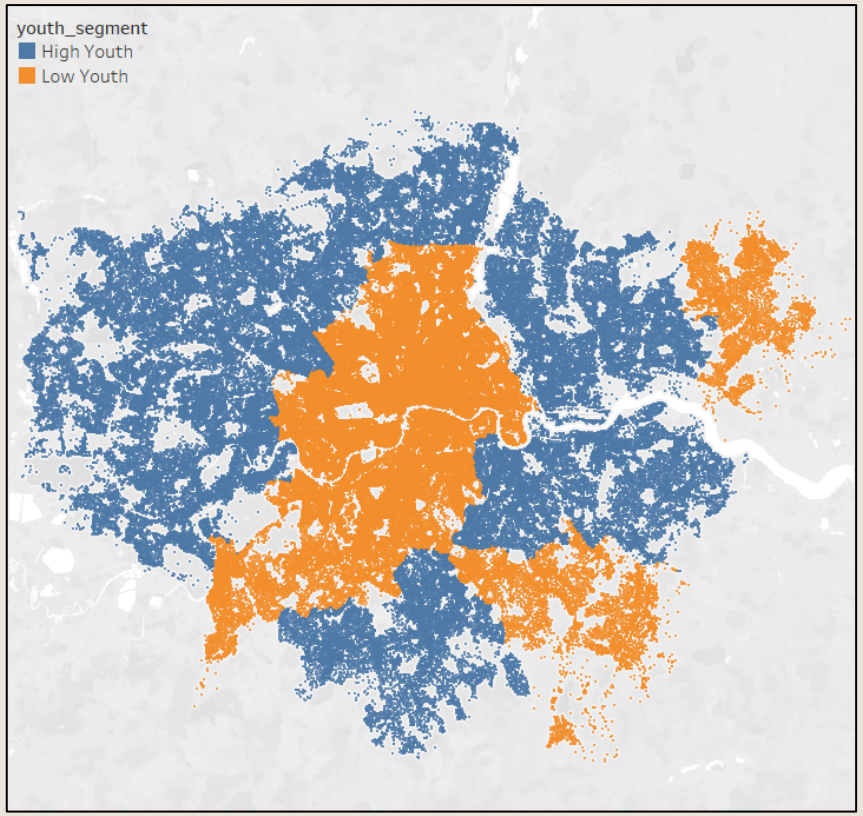


Figure 3: Density Map – Age Segmentation

Purchasing Pattern – Age Segment

- Fruits & Vegetables: Higher consumption in "Low Youth" areas (36.7%) compared to "High Youth" areas (35.5%).
- Grains: Higher consumption in "High Youth" areas (22.1%) versus "Low Youth" areas (20.1%).
- Sweets: Slightly higher consumption in "High Youth" areas (10.7%) compared to "Low Youth" areas (10.2%).
- Readymade Foods: Higher consumption in "Low Youth" areas (6.0%) compared to "High Youth" areas (5.7%).

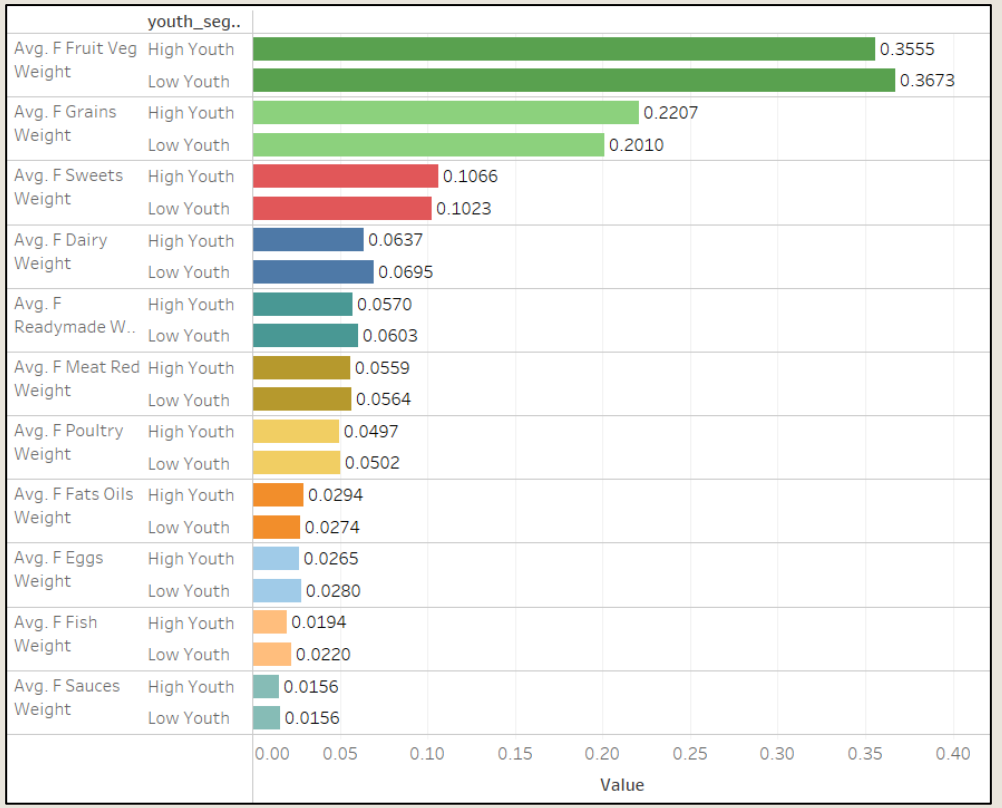


Figure 4: Bar chart for purchasing patterns according to age Segmentation

Insight 2

Are there correlations between food choices/nutrients and obesity-related hospitalizations?

Nutrients and Obesity

- **Fat and Sugar:** Both show negative correlations with obesity-related hospitalizations. This suggests that higher intakes of fat and sugar are associated with lower hospitalization rates in this dataset, which is counterintuitive and may be caused due to the insufficiency and incompleteness of the diet data.
- **Protein:** Displays a positive correlation with hospitalization rates, indicating that higher protein intake correlates with higher obesity-related hospitalizations.
- **Fibre:** Has a negative correlation with hospitalization rates, which aligns with general health advice that higher fiber intake is beneficial and might reduce obesity-related health issues.

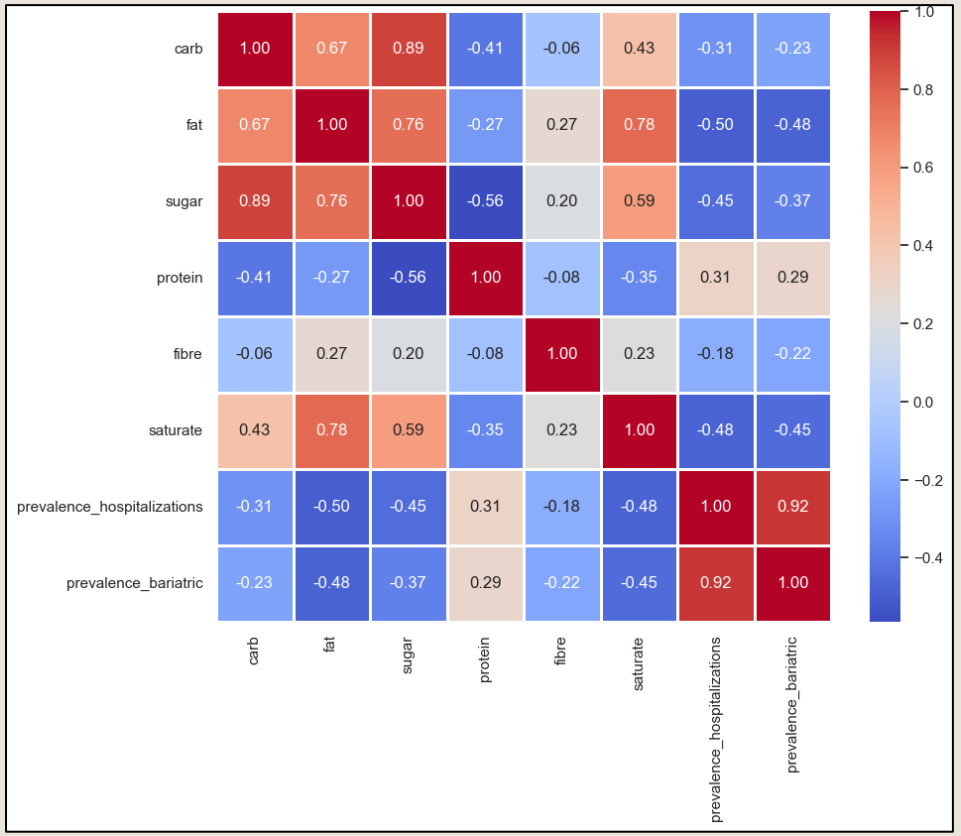


Figure 5: Spearman Correlation Heatmap between nutrients and hospitalizations

Dataset Introduction – Income Dataset

- Gross Disposable Household Income Dataset - The average income of all individuals in the borough for 15 years is sourced from ONS (Office for National Statistics).
- This dataset does not take into account the income of self-employed and pensioners.
- For exploration, the following data is used:
  - total gross disposable household income (in millions) table and
  - total resident population numbers

Dataset Exploration

- GDHI Trends
  - Overall Increase: Most regions, such as Wandsworth, Kensington and Chelsea, and Westminster, reflect general increases in GDHI, ndicating rising income levels.
  - Differential Growth: Regions like Hackney and Tower Hamlets have seen much steeper increases in GDHI.
- Population Trends
  - Population Changes: Regions like Newham and Barking and Dagenham have experienced notable population growth.

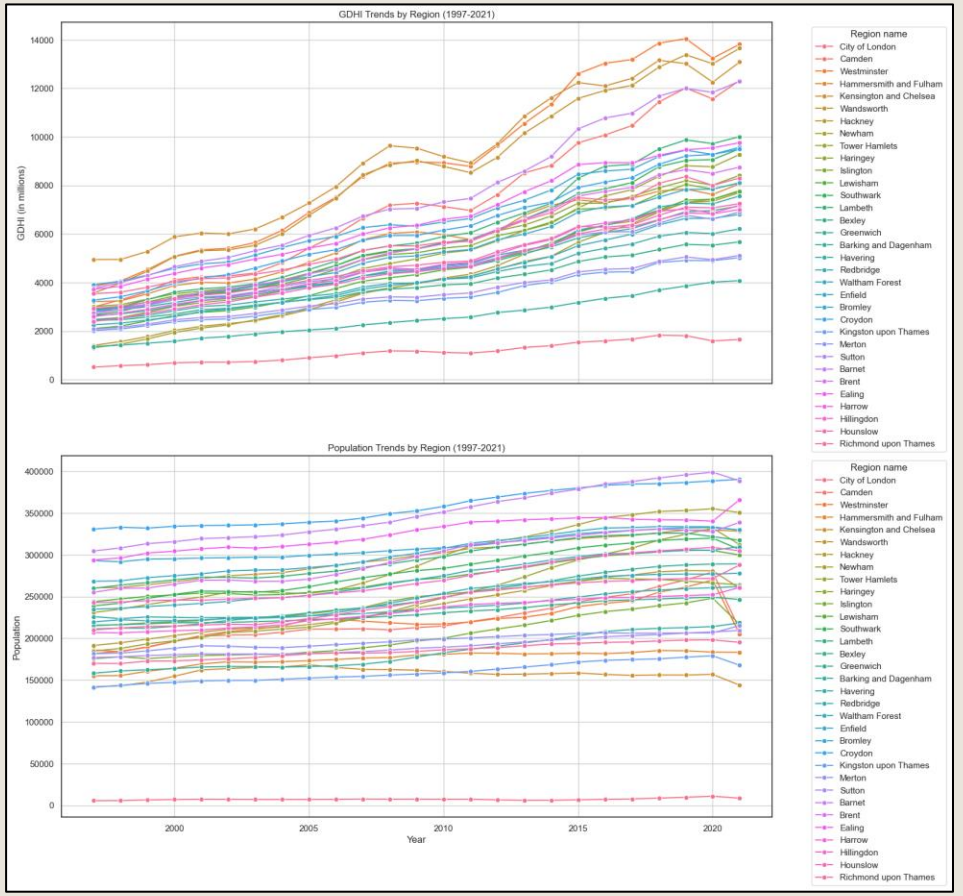


Figure 6: Line Charts to display trends across years for GDHI and Population

Dataset Exploration

The correlation between population growth and GDHI increase is evident in regions like Southwark and Greenwich, where both economic and demographic growth are aligned. This contrasts with areas like Richmond upon Thames, where population has remained relatively stable despite economic growth.

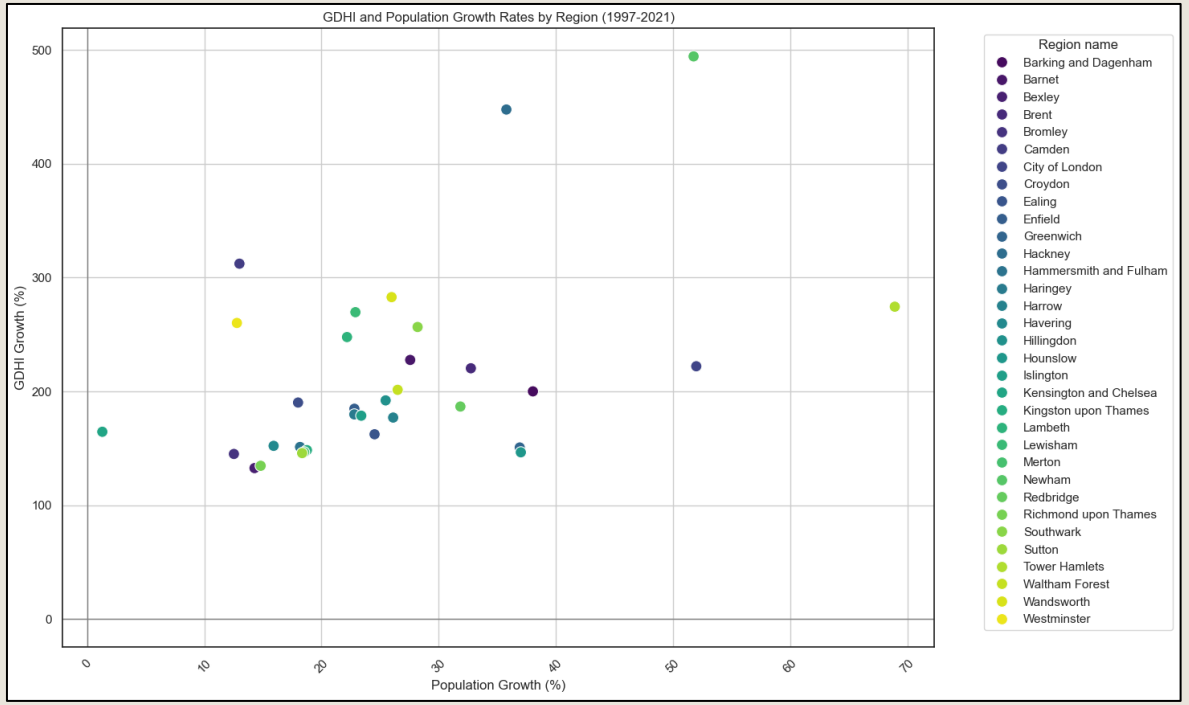


Figure 7: Scatterplot to display the relationship between for GDHI growth and Population growth

Income and Purchasing Patterns

The density map displays the average income in households (in millions) across the boroughs in the year 2015.

- Minimum avg income: 1540
- Maximum avg income: 12612

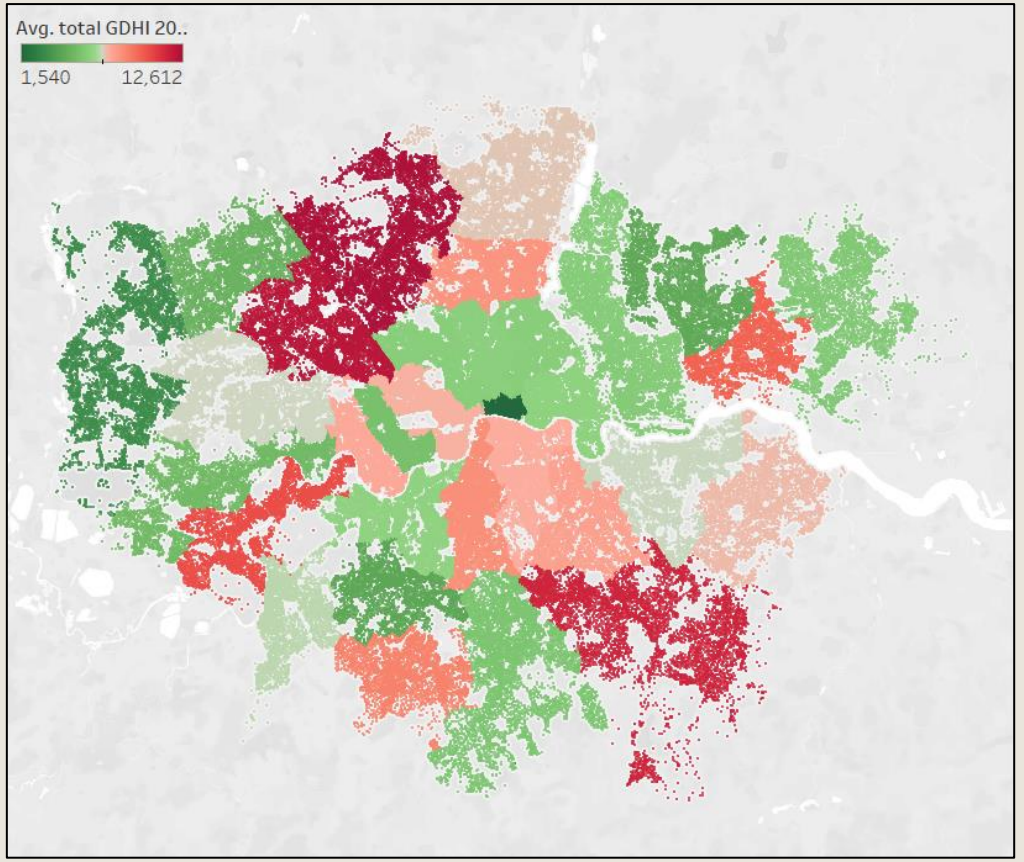


Figure 8: Density Map of Avg Income in household across the boroughs

Income and Purchasing Patterns

Positive Correlations:

- Tea and Coffee: 0.25
- Water and Grains
- Fish and Poultry: 0.19 and 0.15, respectively

Negative Correlations:

- Dairy Products: -0.26
- Wine and Fruit/Vegetables

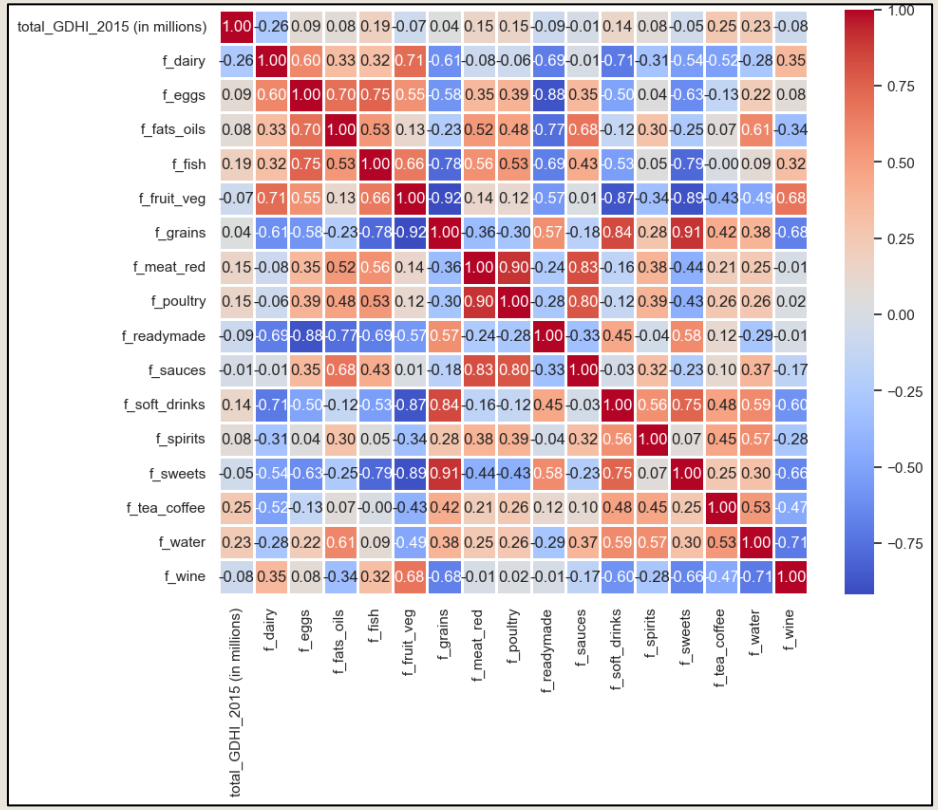


Figure 9: Spearman Correlation Heatmap between food Categories and income

Conclusions

- Male-dominated areas show a preference for more energy-dense food categories like dairy, grains and eggs. Female-dominated areas tend to consume more readymade foods and sweets.
- Areas with a higher number of younger populations consume more grains and sweets, which may reflect lifestyle or energy choices. Areas with a lower proportion of younger residents had a higher intake of fruits and vegetables, and ready-made dishes, showing a preference for convenience and health.
- The data shows counterintuitive and unexpected correlations between dietary intake and obesity-related hospitalizations. Higher intakes of fat and sugar are associated with lower hospitalization rates, which might suggest issues with the data's completeness.
- Conversely, higher protein intake correlates with increased hospitalization rates, possibly reflecting unmeasured dietary or lifestyle factors. On the other hand, a higher fiber intake is associated with lower hospitalization rates, consistent with advice that fiber is beneficial for health.
- Higher income groups often prefer premium or health-oriented products, such as organic grains and bottled water. This preference also extends to high-quality proteins like fish and poultry. Conversely, the negative trends in dairy, wine, and fruit/vegetable consumption might reflect dietary preferences for non-dairy alternatives or other lifestyle choices typical of wealthier demographics.

References

1. Aiello, Luca Maria; Schifanella, Rossano; Quercia, Daniele; Del Prete, Lucia (2020): Tesco Grocery 1.0. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.4769354.v2

2. GDHI_Income: - https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomelocalauthoritiesbyitl1region

3. Bender AE. Food preferences of males and females. Proceedings of the Nutrition Society. 1976;35(2):181-189. https://doi.org/10.1079/PNS19760031

4. Scheelbeek PFD, Cornelsen L, Marteau TM, Jebb SA, Smith RD. Potential impact on prevalence of obesity in the UK of a 20% price increase in high sugar snacks: modelling study. BMJ. 2019 Sep 4;366:l4786. doi: 10.1136/bmj.l4786. PMID: 31484641; PMCID: PMC6724407.

5. RHE Global – Smarter Public Protection: - https://www.linkedin.com/pulse/how-poor-eat-hatchett-rheglobal/