Executive Summary

Overview
The project revolves around a detailed analysis of the Tesco Grocery 1.0 Dataset in correlation with income data across the Greater London area. The dataset explores consumer purchasing behaviors and their socio-economic implications.

Dataset Description
The Tesco Grocery 1.0 Dataset[1] encapsulates a year's worth of grocery purchases by 1.6 million customers across 411 Tesco stores, aggregated from over 420 million transactions. This dataset is dissected into 52 distinct files offering monthly and annual insights per region, structured around different geographic divisions: Lower Super Output Areas (LSOA), Medium Super Output Areas (MSOA), Wards and Boroughs.

Key Insights:
- The analysis of food consumption across Greater London (MSOA) reveals distinct regional preferences, with central and northern areas predominantly consuming more fresh foods, while southern and outskirt regions favor processed or convenient foods. Monthly fluctuations also affect food consumption patterns; fresh food intake peaks during summer, while processed foods see a slight increase in winter. Overall, fresh food maintains a significant preference, never comprising less than 67% of total food consumption, indicating a general trend towards healthier eating habits. These insights are crucial for crafting targeted public health policies and interventions aimed at promoting healthier food choices.

- The second insight examined the relationship between population & population density and total grocery transactions in 2015. For population, maps[2] and statistical analysis reveal a moderate positive correlation (0.51) between population size and total grocery transactions per borough. While larger populations generally correspond to more transactions, the correlation is not absolute, suggesting additional factors such as economic status, store density, or local shopping behaviors also play significant roles. Also, it shows a weak positive correlation of 0.27 for population density. Outliers identified in the data suggest the presence of other factors that might influence transaction levels, challenging the notion of a straightforward correlation between these variables.

Income Dataset:
The Income dataset[3] provides an extensive history of average earnings in London boroughs from 1999 to 2022, detailing gross weekly earnings and yearly income statistics for regional comparison. Central London boroughs show high income diversity and volatility, reflecting economic growth and varied financial sectors, while outer boroughs like Barnet and Ealing display greater income stability over time.

Two Dataset Comparison:
The analysis consisted of finding correlation of nutrients, food and drinks with the Income. The correlation analysis shows that lower-income groups consume more sugar, salt, fat and carbs, and higher earners drink more alcohol, consistent with literature[4,5,7]; however, there's no clear link or correlation between income and fibre or protein intake, contrary to expectations[4,6]. Higher-income individuals tend to eat more fruits, vegetables, fish, eggs, and dairy, in line with a healthier diet[4,5,6]. Also, they appear to consume fewer staple grains such as bread, pasta, and rice. Meanwhile sauces, fats & oils, and ready-made meals show negative correlation with income however, it is slightly low compared to expected high negative correlation[4,6]. Poultry and red meat intake shows no income-related trend, which diverges from literature[4,6] that suggests a negative correlation.

Conclusions
The analysis showcases the significant influence of socio-economic status on dietary choices, confirming that income affects consumption of various foods and nutrients, as supported by existing research. Unexpectedly, the relationship between income and the consumption of protein, fibre, and red meat is minimal, indicating that factors determining diet quality are more nuanced than income alone. This complexity invites further exploration into the diverse factors that influence dietary habits beyond income levels.

References:

1. Aiello, L.M.,, Quercia, D.,, Schifanella, R., and Del Prete, L., 2020a. Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. *Scientific Data*, 7(1).
2. Authority, G.L., 2014. *Statistical gis boundary files for London*. [online] London Datastore News. Available from: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london [Accessed 10 Apr. 2024].
3. Authority, G.L., 2014. *Earnings by workplace, borough*. [online] London Datastore News. Available from: https://data.london.gov.uk/dataset/earnings-workplace-borough [Accessed 10 Apr. 2024].
4. (2022) *Diet-Related Health Inequalities*. Available at: https://researchbriefings.files.parliament.uk/documents/POST-PN-0686/POST-PN-0686.pdf (Accessed: 12 April 2024).
5. Roberts, C. (2014) *Fruit and vegetable consumption*. Available at: http://healthsurvey.hscic.gov.uk/media/1092/_7-fruit-and-vegetable-consumption_7th-proof.pdf (Accessed: 12 April 2024).
6. De Agostini, P., 2014. *The effect of food prices and household income on the British diet* (No. 2014-10). ISER Working Paper Series.
7. Office for National Statistics, 2017. Adult drinking habits in Great Britain: 2005 to 2016. *OFFICE FOR NATIONAL STATISTICS*