Executive Summary:

The Tesco dataset contains approximately 420 million worth of food and drink purchases by approximately 1.6 million Clubcard owners across all 411 Tesco London stores in 2015. It contains information such as census, demographics, transaction information, nutritional and caloric detail across 17 different food and drink product categories. Limitations of the dataset include the dataset only consists of Clubcard owners and only in London, making it hard to generalise purchasing patterns to the rest of the country. Londoners tend to have higher purchasing power; it is therefore a stretch to generalise the conclusion derived from the dataset for the rest of the UK. Not to mention, the selection of products varies from store to store. As smaller Tesco stores have limited storage and shelf capacity, they cannot display the full range of products offered by larger stores. This too will influence a customer's purchasing pattern. Finally, customers generally do not shop exclusively in a single Tesco store, and they might even make purchases at non-Tesco stores or at different Tesco locations depending on time of the day, price enticement or needs. Thus, we need to take this into account as this factor will reduce the accuracy when attempting to analyse regional purchasing patterns.

Initial analysis reveals there exists correlations between grocery weight and volume with population density and area size, indicating that larger, more populated regions typically have more purchases, although there are exceptions. A small but relatively richer and economically busy region such as the boroughs around center of London such as City of London, Westminster, Kensington and Chelsea will have higher overall grocery weight and volume purchases as well. All these observations are potentially useful information for a marketing campaign or for optimising Tesco's supply chain based on a borough's preference if we can establish a pattern between purchasing preferences and income level.

Before diving deeper into it, we first explored external factors that may be worth considering in conjunction with our goal. Using machine learning, we uncovered two potential factors that might drive purchasing preference. Cluster analysis reveals that there are hidden patterns when it comes to different age groups (*0-17, 18-64, and 65+*) as well as genders (*male and female*). Exploratory data analysis was performed on both factors and appropriate statistical tests were conducted to strengthen the analysis to test for significance.

It reveals that while there are some differences in purchasing preference between male and female, a statistical interrogation has revealed that on average, and at a borough level of granularity, we are 95% confident that the purchasing preference between men and women across all 17 product categories does not differ significantly. However, at a significance level of 0.05, there are significant differences in purchasing preferences across different age groups. For the younger age group, this group shows relatively lower purchasing fractions across most categories, likely reflecting their lesser purchasing power and dependence on adults for buying decisions. Although, there are noticeable fraction in soft drinks and tea & coffee categories compared to other age groups. For the middle age group, they dominate most categories, reflecting their broader economic activity and varied preferences. This group also shows higher fractions in categories like beer, wine, and spirits, which are adult-oriented products. For the older age group, they show interest in ready made category when compared to other age group.

When it comes to income and purchasing preferences, higher-income boroughs tend to spend more on fruit and vegetables, which are often associated with healthier and more expensive food choices. On the other hand, strong negative correlations were observed with purchasing preference for categories such as ready made, soft drink, grains, and sweets, which are often associated with less healthy and more affordable food options. This observation agrees with existing literature such as *Power et. al, 2019* and *French et. al, 2019.*

Notably, beer and wine purchases are positively correlated with income, suggesting higher consumption among wealthier demographics. This observation agrees with an existing research report from NHS Digital (*Niblett, P., 2017*). Similarly, purchase of water shows a negative correlation with income, which aligns with research by *Family et. al (2019)*.

There is no drastic change in median income in borough across the years from 2005 onwards. Therefore, we can expect purchasing patterns of income earners across all spectrums to remain unchanged from 2005 onwards. To take an example, if Westminster (a high-income borough) has a higher purchasing preference for fruit and vegetables in 2015, we expect the same general behaviour to continue in 2024 given that there are no boroughs that dropped from high income to low income or vice versa after 2005. Thus, it is useful to identify the top 5 (*Camden, Richmond-upon-Thames, Kensington & Chelsea, Westminster, City of London*) and bottom 5 boroughs (*Brent, Bexley, Waltham Forest, Newham, Barking & Dagenham*) by median income in 2015 as we expect their purchasing preference to still holds true in 2024.

In conclusion, we predict that the top 5 boroughs by median income in 2015 have a higher preference for healthy food, beer, and wine and the bottom 5 boroughs by median income in 2015 have a higher preference for less healthier food and water in 2024.

Bibliography:

Dataset

I. Aiello, L.M., Quercia, D., Schifanella, R., and Prete, L.D. 2020. *Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London* [Online]. London: Nature. Available from: https://doi.org/10.1038/s41597-020-0397-7 [Accessed 20 March 2024].

II. HM Revenue & Custom, 2024. *Average Income of Tax Payers, Borough* [Online]. London: London Datastore. Available from: https://data.london.gov.uk/dataset/average-income-tax-payers-borough [Accessed 20 March 2024]

III. Greater London Authority (GLA), 2020. *Statistical GIS Boundary Files for London* [Online]. London: London Datastore. Available from: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london [Accessed 20 March 2024]

Report

IV. Niblett, P., 2017. *Statistics on Alcohol (England, 2017)*. Leeds: NHS Digital

Journal

V. Family, L., Zheng, G., Cabezas, M., Cloud, J., Hsu, S., Rubin, E., Smith, L. V., and Kuo, T., 2019. Reasons why low-income people in urban areas do not drink tap water. *The Journal of the American Dental Association* [Online], 150(6) Available from: https://doi.org/10.1016/j.adaj.2018.12.005 [Accessed 10 April 2024]

VI. French, S.A., Tangney, C.C., Crane, M.M., Wang, Y., and Appelhans, B. M., 2019. Nutrition quality of food purchases varies by household income: the SHoPPER study, *BMC Public Health* [Online], *19(231)*. Available from: https://doi.org/10.1186/s12889-019-6546-2 [Accessed 10 April 2024]

Website

VII. Office for National Statistics, 2024. Inflation and price indices [Online]. Newport: Office for National Statistics. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/march2024 [Accessed 10 April 2024]

VIII. The Guardian, 2011. Wages throughout the country: how does your area compare? *The Guardian* [Online], n.d.. Available from: https://www.theguardian.com/news/datablog/2011/nov/24/wages-britain-ashe-mapped [Accessed 10 April 2024]

IX. Daniel, C., 2014. *why poor people buy bottled water* [Online]. Contexts. Available from: https://contexts.org/articles/why-poor-people-buy-bottled-water/ [Accessed 10 April 2024]

X. Power, M., Pybus, K. J., and Pickett, K., 2021. *Families have high awareness of healthy eating but low income means many struggle to access good food* [Online]. York: University of York. Available from: https://www.york.ac.uk/news-and-events/news/2021/research/awareness-healthy-eating-struggle-access-good-food/ [Accessed 10 April 2024]

Disclaimer:

AI tools were used to debug R and Python code.

AI tools were used for grammatical corrections and suggestions.

All images used in the PowerPoint originated from Microsoft 365 Subscribers' stock image library (*None of it were AI generated*).