Tesco Grocery 1.0 Dataset vs. Educational Attainment

For Data Experts in Education and Health

Introduction: Tesco Grocery 1.0 Dataset

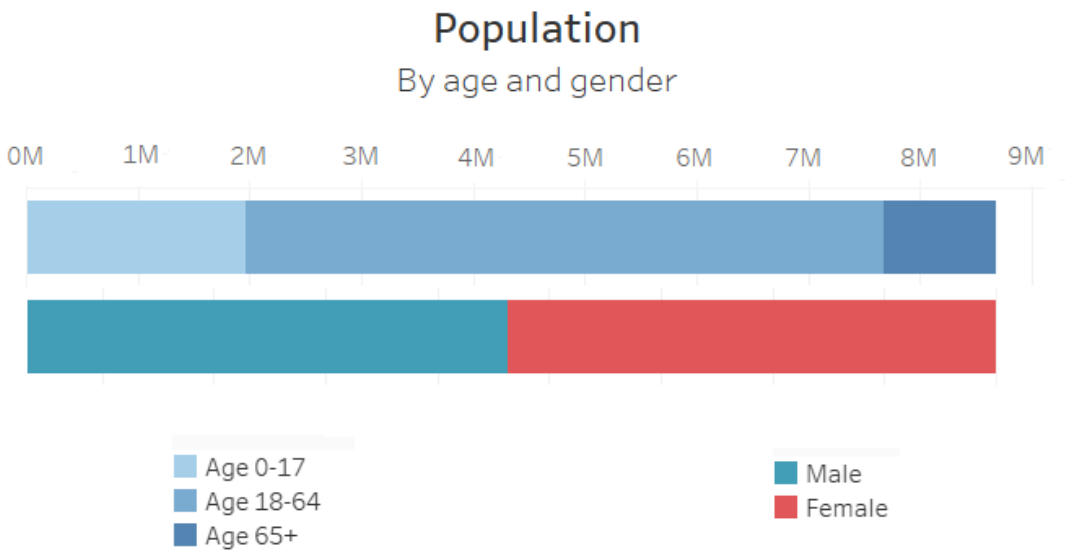
- The Tesco Grocery 1.0 dataset records 420M food items purchased by 1.6M Clubcard owners who shopped at the 411 Tesco stores in Greater London over 2015.
- The dataset contains four geographic aggregation: LSOA, MSOA, Ward, Borough.
- In this analysis, we work at borough level, covering 33 local authorities in London.

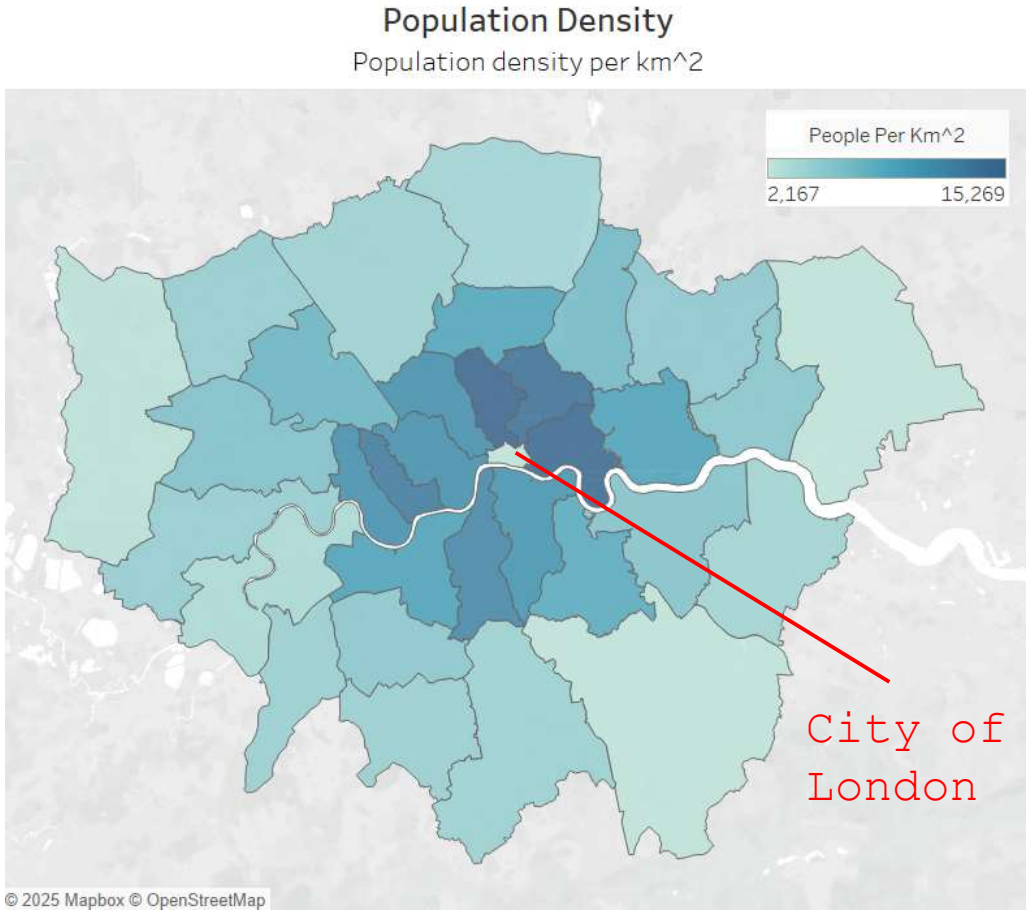
TESCO

figshare

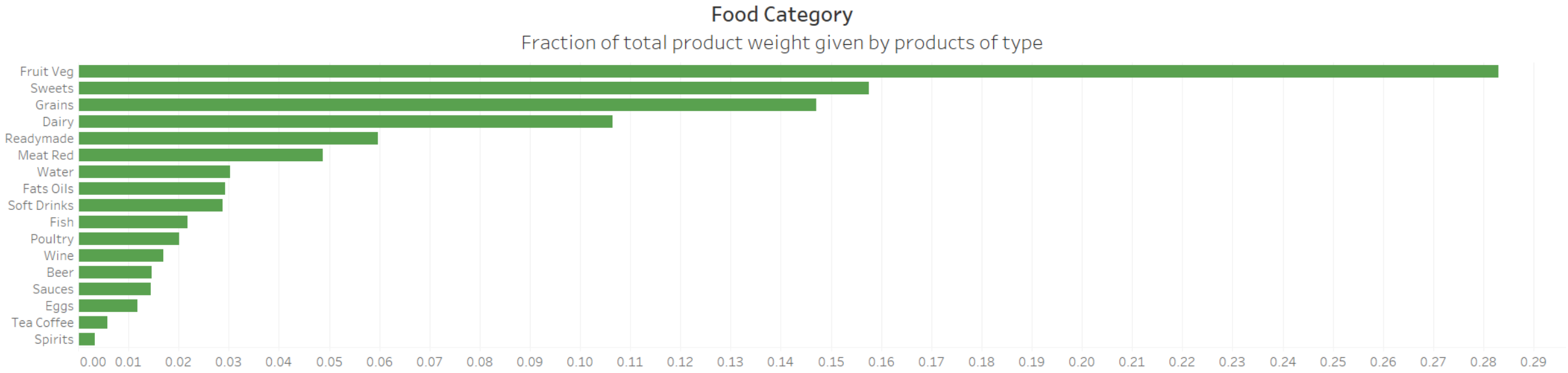
Tesco Grocery 1.0

Population


Population
By age and gender
0M
1M
2M
3M
4M
5M
6M
7M
8M
9M
Age 0-17
Age 18-64
Age 65+
Male
Female


Population Density
Population density per km^2
People Per Km^2
2,167
15,269
City of London
© 2025 Mapbox © OpenStreetMap

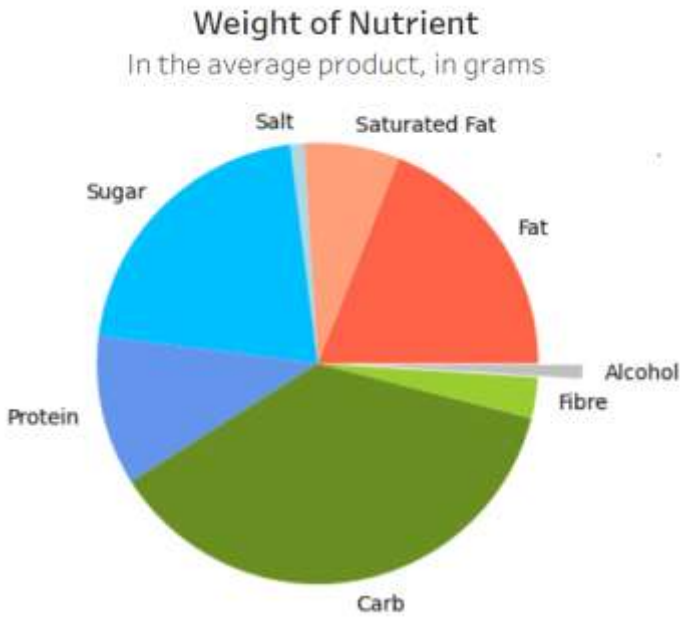
Food Category

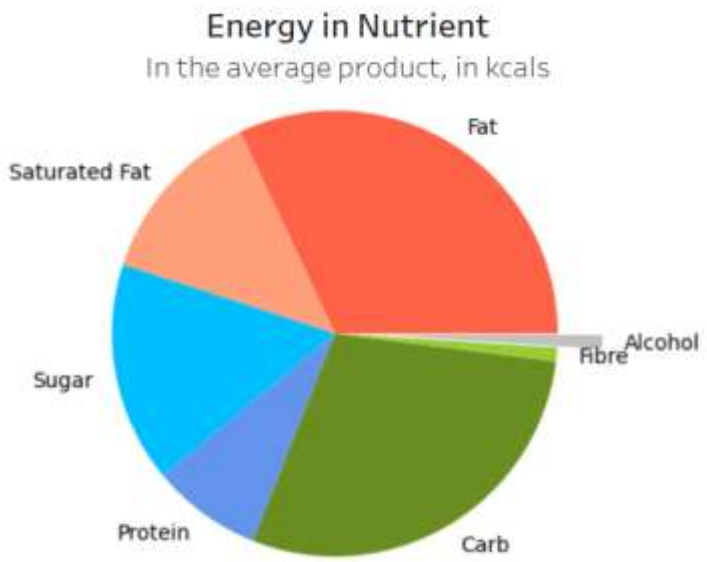


Diversity of food product categories

normalised to [0,1]

In Category	0.79211
In Weights of Category	0.80493





Statistics in Average Product

Volume(liters)	109.8
Weight(grams)	373.9
Energy(kcals)	178.5
Energy Density(kcals/grams)	0.5
Diversity of Nutrients(Energy)	1.7
Diversity of Nutrients(Weight)	1.6

Nutrient

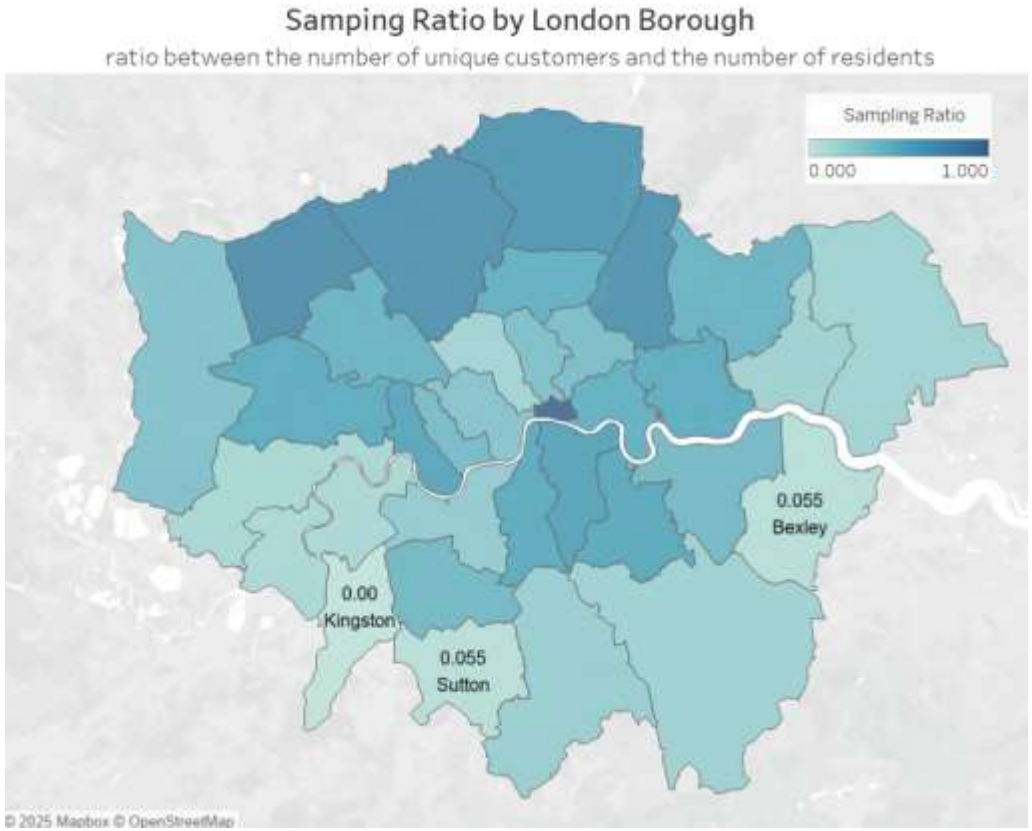
Limitation and assumption

Assumptions:

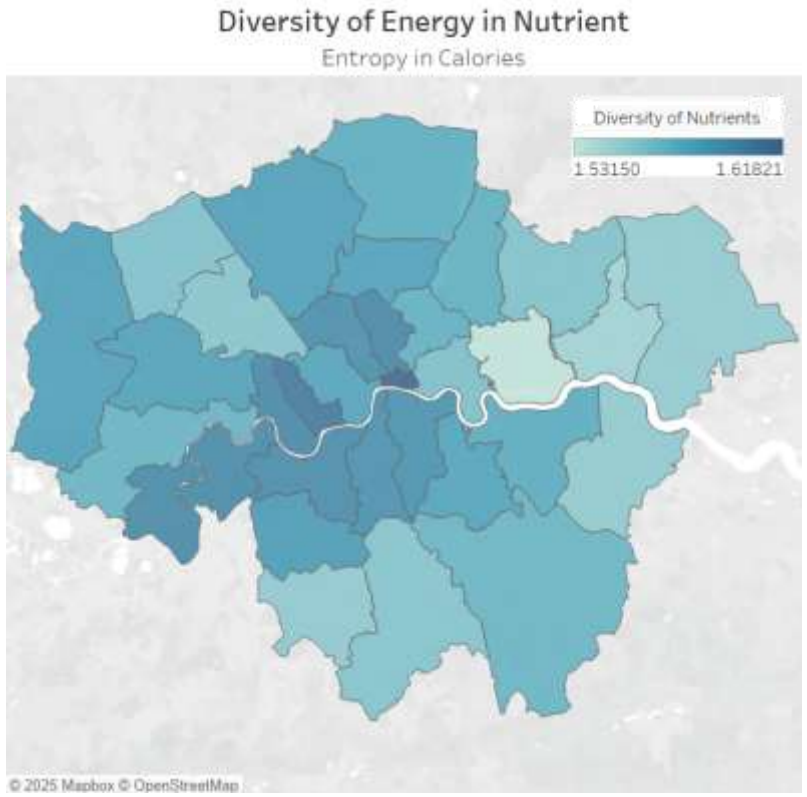
- Sample size - assume 1000 is good enough.

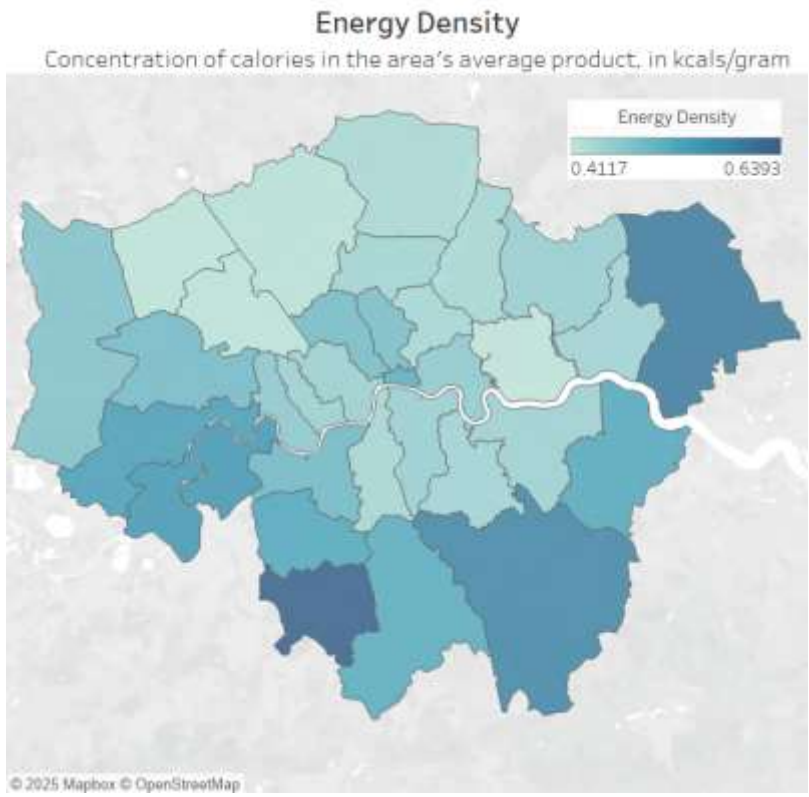
Limitations:

- The sample is not random.
- Fewer Tesco stores in Southern London.
- The data is averaged for each product.

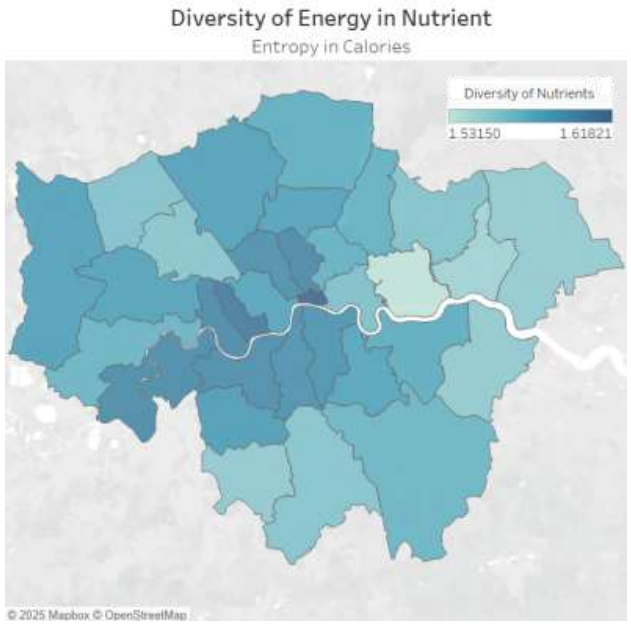


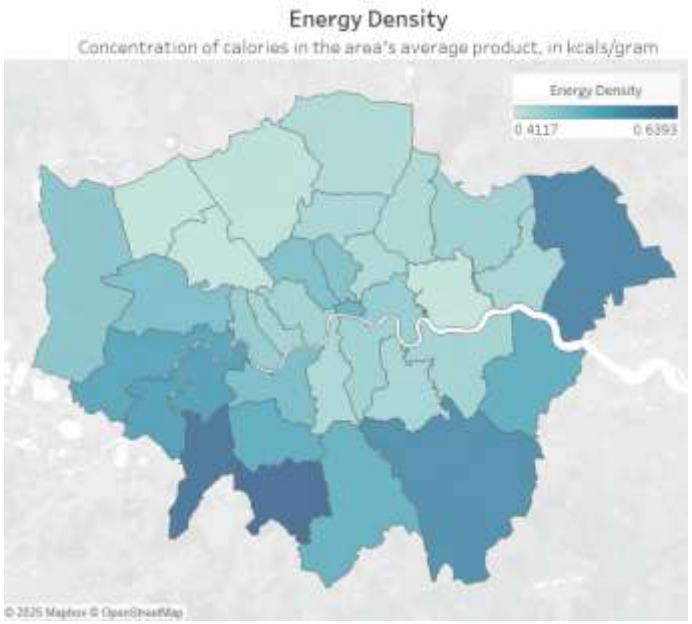
Insight: Diversity of Nutrient vs. Energy Density





Insight: Diversity of Nutrient vs. Energy Density



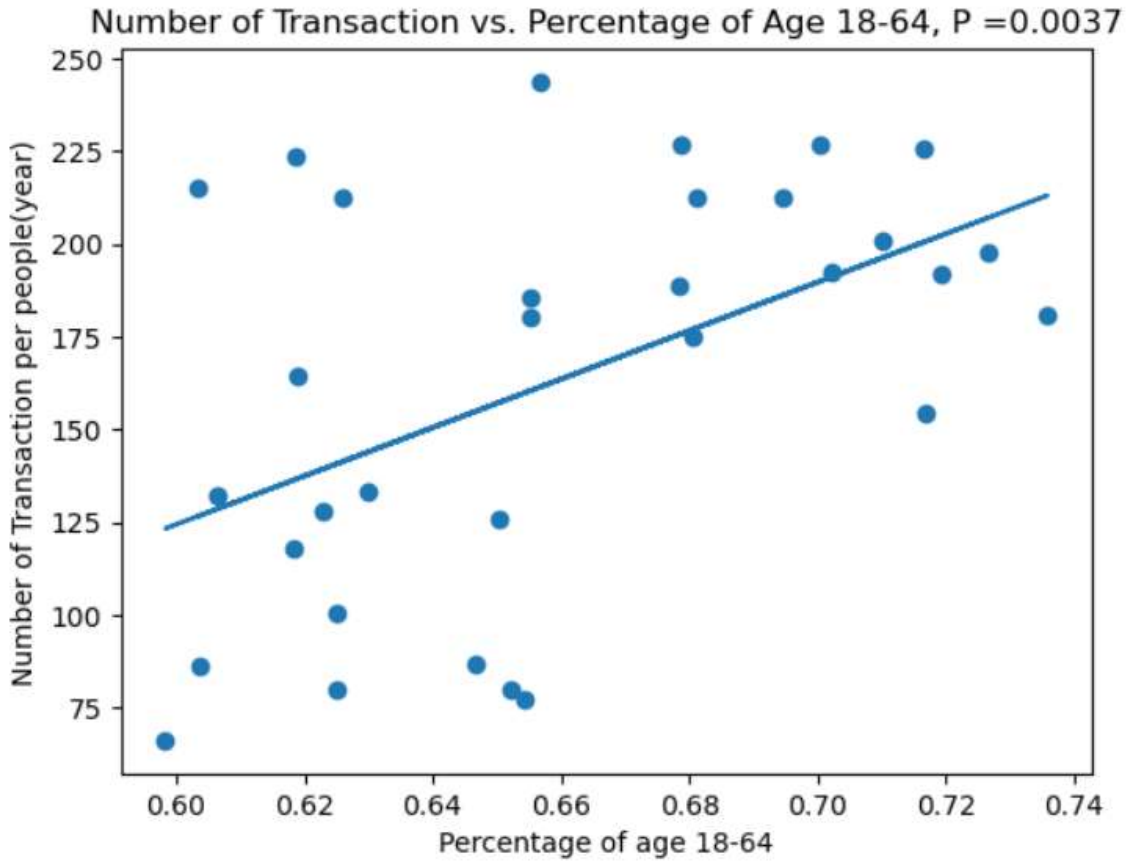


People in Northern London care more about the diversity of nutrients, while in Southern London they prefer to buy food with higher energy density.

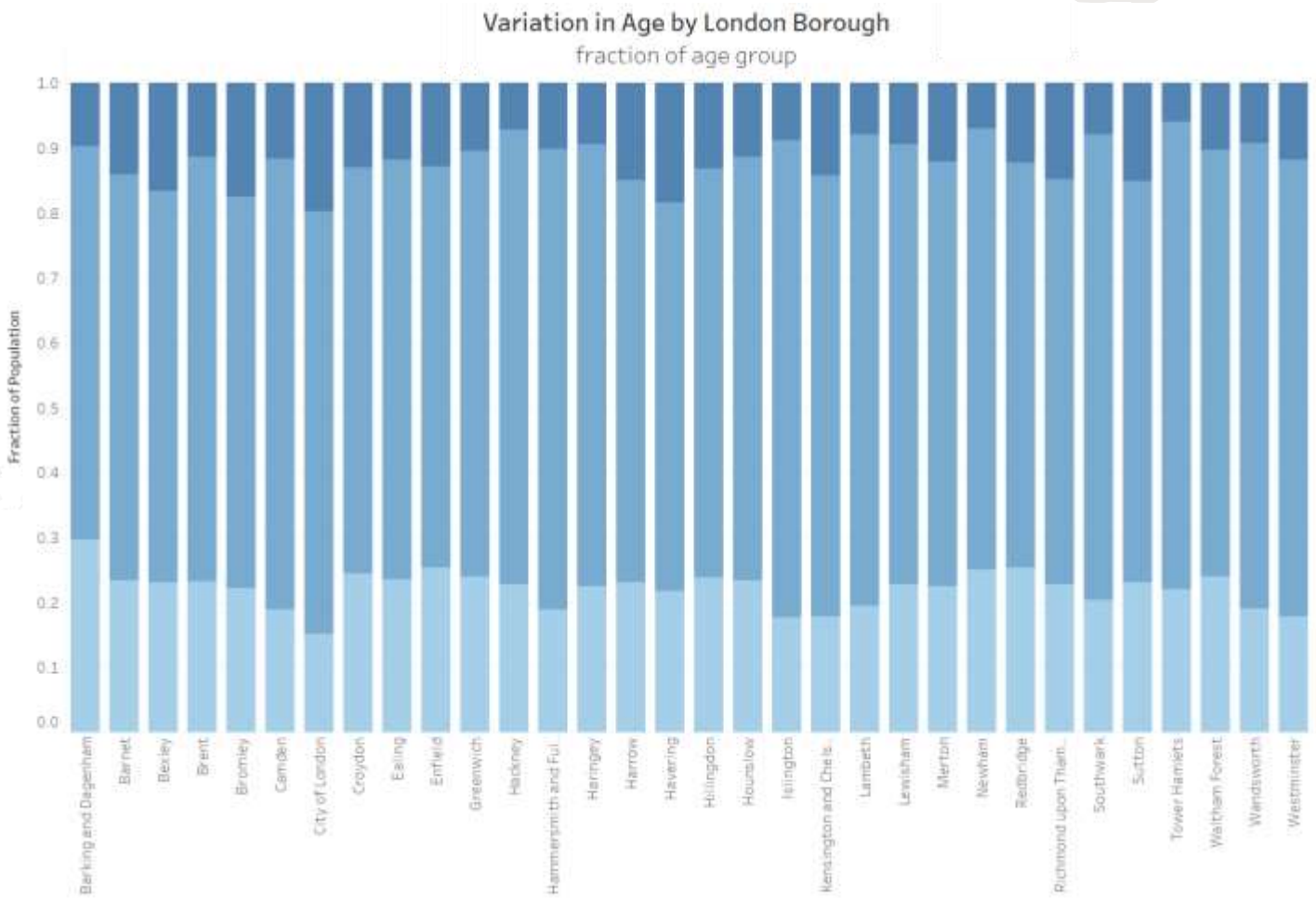
Insight: The Effect of Variance in Age Between Boroughs

The effect

- Intuitively, the main consumption group are people aged 18 - 64.
- Positive correlation between number of transactions and the percentage of age group 18 - 64.
- If the variation is large, it may influence the consumption habits



Insight: The Effect of Variation in Age Between Boroughs



From the graph, the variance in age group is acceptable.

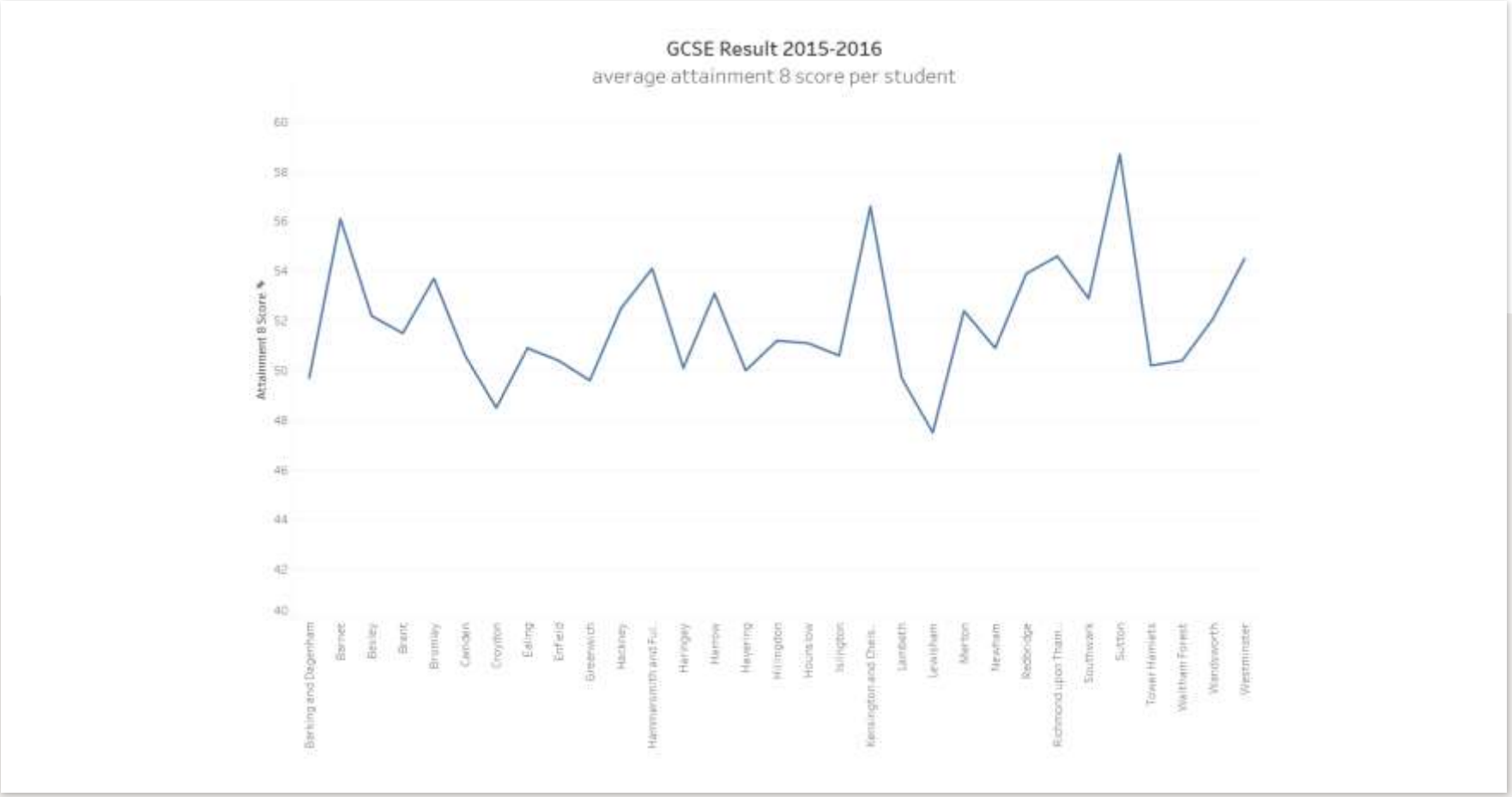
Introduction: Educational Attainment Dataset

Department for Education

GCSE Results by Borough

Department for Education

- Data on GCSE and equivalent entries and achievements at the end of Key Stage 4 by gender, ethnicity, first language, free school meal eligibility, special education needs, disadvantaged status and London boroughs, using the new 1-9 grading system.
- Attainment 8 score measures pupils' average grade across eight subjects.
- In this analysis, we use the average attainment 8 score per pupil for all students by borough,



An insight of Education dataset

Limitation and Assumption

Limitations:

- Mismatch in age:

  Tesco: All age groups.

  Education: Teenagers.

- Mismatch in periods:

  Tesco: 2015.01 – 2015.12

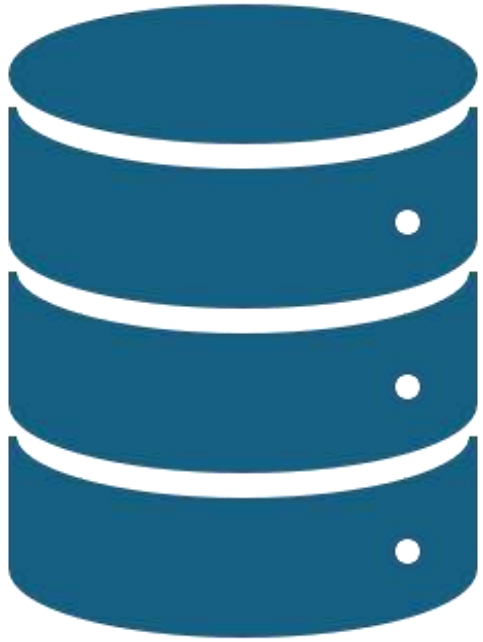
  Education: 2015.09 – 2016.06

- Missing data in city

Assumptions:

Assume the food purchased from Tesco stores have a long-term effect.

The diabetes has effect on the GCSE result in the next year.



Analysis: Data cleaning and handling

1

Concatenate the Tesco 1.0 data from September 2015 to December 2015 and calculate its average.

2

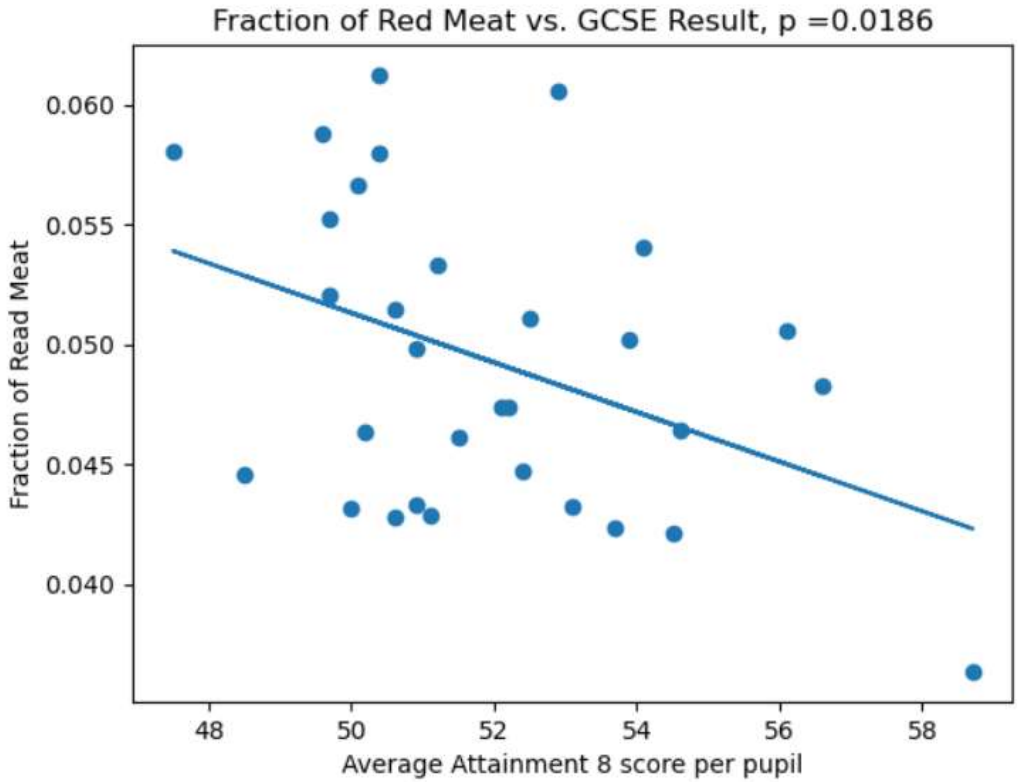
Exclude the data in City of London and Kingston.

3

Link the Tesco 1.0 and GCSE result dataset by London Boroughs.

Analysis: Methodology and Statistics

- For every dietary factor in Tesco 1.0, compute its regression coefficients with GCSE results by boroughs.
- Test the significance of regression at 95% confidence interval.
- The p value is calculated using Wald Test with t-distribution of the test statistic.
- Null hypothesis: No correlation.
- If p value <= 0.05, accept null hypothesis, otherwise reject and accept the alternative



Analysis: Result

Correlation of Dietary Factors and Educational Attainment

Dietary Factor	Correlation	P-value
Fibre	+	0.027
Salt	–	0.004
Red Meat	–	0.025
Poultry	–	0.022
Sauces	–	0.016
Soft Drinks	–	0.045
Diversity in Food Category	–	0.000
Water	–	0.009

Support Literature

- A paper suggested 'young *people with higher education generally have a lower BMI, a higher healthy nutrition index, and healthier habits*'.
- In other way, obesity leads to poor educational attainment.
- The widely known causes of obesity: too much **salt**, **fat**, **sugar**, **energy**, and too little **fibre**.

Analysis: Result

Correlation of Dietary Factors and Educational Attainment

Dietary Factor	Correlation	P-value
Fibre	+	0.027
Salt	–	0.004
Red Meat	–	0.025
Poultry	–	0.022
Sauces	–	0.016
Soft Drinks	–	0.045
Diversity in Food Category	–	0.000
Water	–	0.009

Saturated Fat

Salt

Sugar

Conclusion

Positive Correlation: Fibre
Negative Correlation: Salt, Sugar, Fat

While most of the results are reasonable, there are few things to consider:

- According to the paper, the diversity of nutrients should contribute to the educational behaviors, but the correlation is negative.
- The correlations of fat, fruit & vegetable, oils and so on are not detected.
- Water should not be correlated with educational attainment.

These errors may be caused by the limitations and bias mentioned earlier.

References

- Aiello, L.M., Schifanella, R., Quercia, D., Del Prete, L., 2020, *Tesco Grocery 1.0* [Online]. Figshare. Available from: https://doi.org/10.6084/m9.figshare.c.4769354.v2 [Accessed 23 February 2025].
- Greater London Authority, 2024, *Statistical GIS Boundary Files for London* [Online]. London Datastore. Available from: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london [Accessed 23 February 2025]
- Department for Education, 2025, *GCSE Results by Borough* [Online]. London Datastore. Available from: https://data.london.gov.uk/dataset/gcse-results-by-borough [Accessed 25 February 2016].
- Sandri, E., Pardo J., Cantín Larumbe, E., Cerdá Olmedo, G., Falcó, A., 2024, Analysis of the influence of educational level on the nutritional status and lifestyle habits of the young Spanish population. Front Public Health.12(1341420). Available from: https://doi.org/10.3389/fpubh.2024.1341420 [Accessed 03 March 2016].