5011CEM Big Data Programming Project

Chan Khai Shen

Agenda

Number	Agenda
1	Data Analysis – Objective
2	Data Analysis – Dataset Explanation
3	Data Analysis – Findings and Discussion
4	Conclusion

Data Analysis - Objective

- To study the **eating habit** of people based on data of Tesco Grocery 1.0 by splitting all lower super output areas (LSOAs) into different **clusters based on food purchase** by weight.

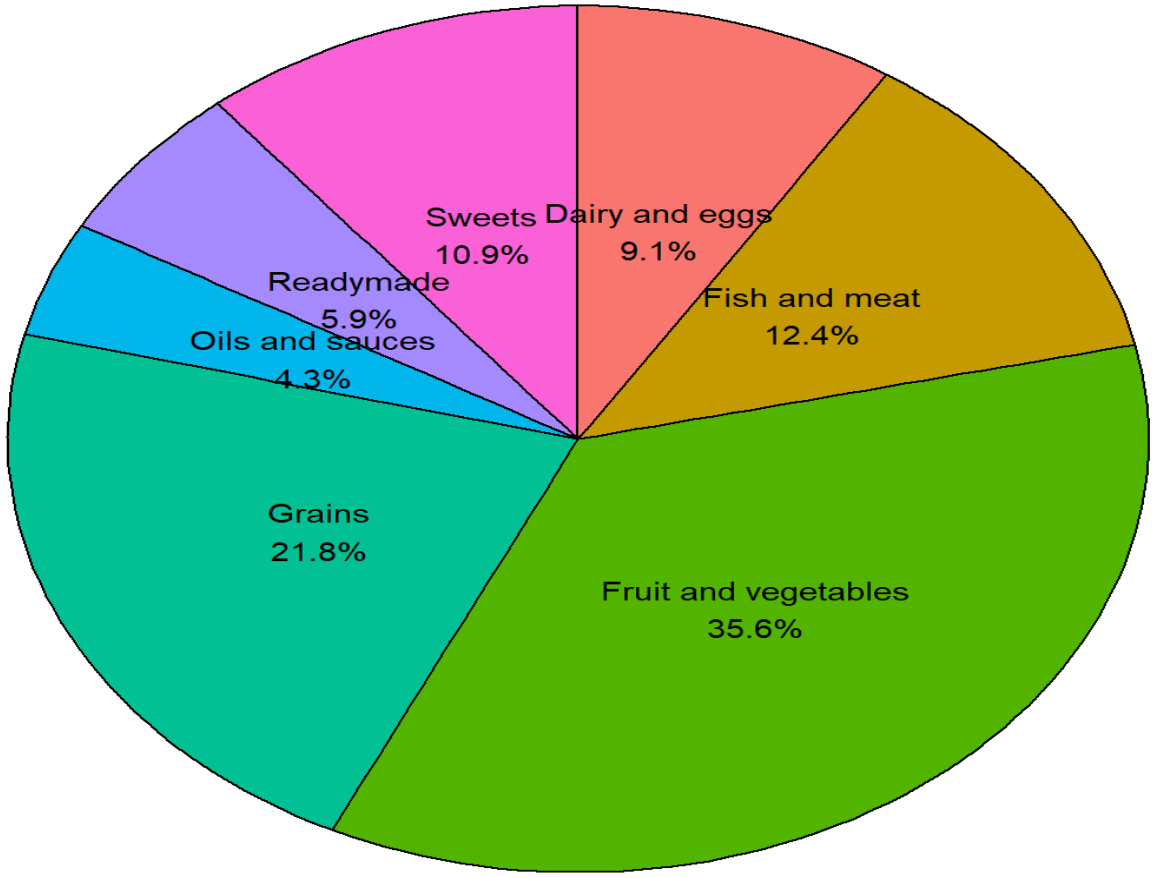
Data Analysis – Dataset Explanation

- Dataset used is f_{category}_weight from Tesco Grocery 1.0
- f_{category}_weight explains the fraction (by weight) of that category of food purchase of Londoners in Tesco outlet in 2015

Data Analysis – Dataset Explanation

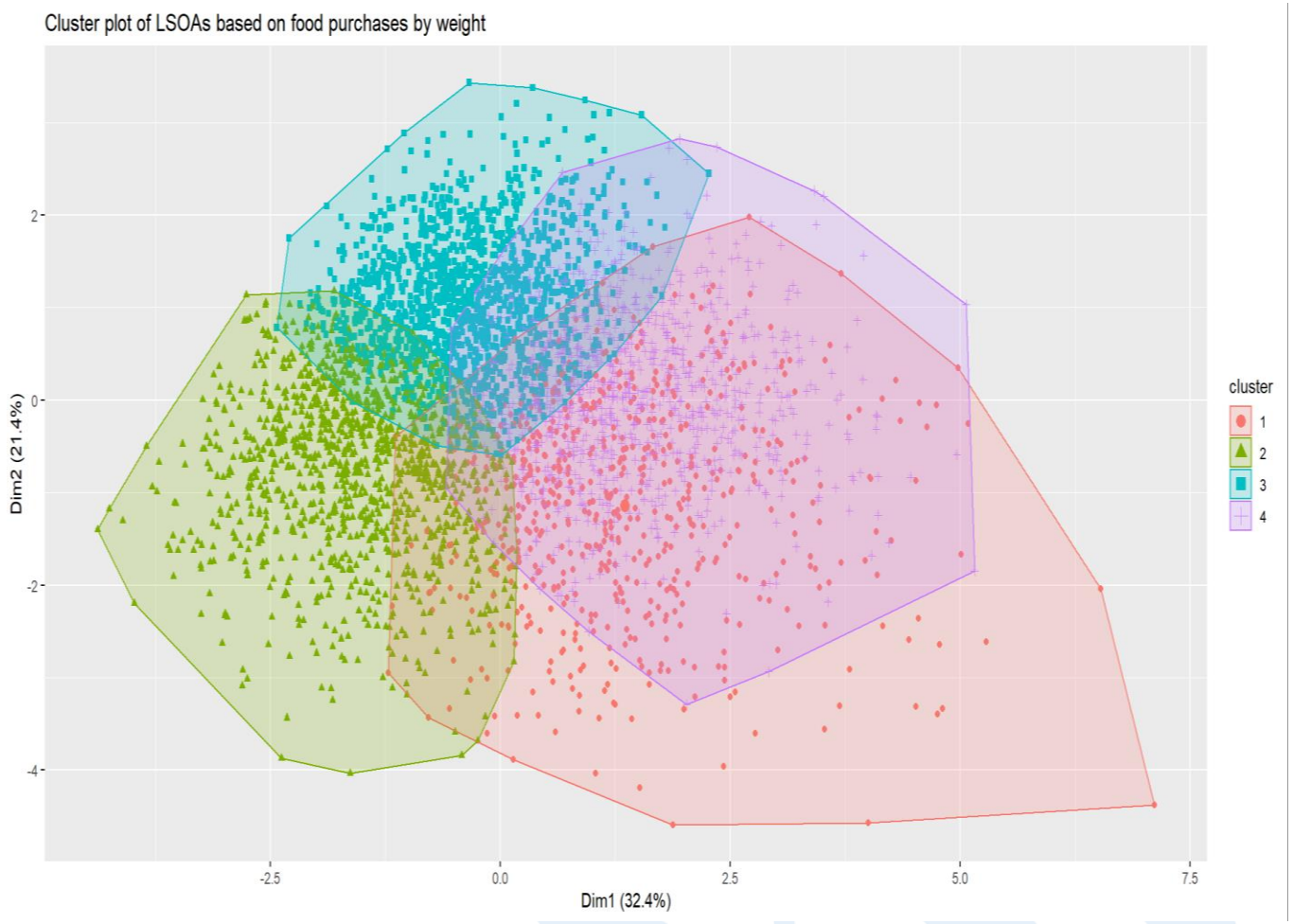
The most purchased category is fruit and vegetable (35.6%), followed by grains (21.8%), fish and meat (12.4%), sweets (10.9%), dairy and eggs (9.1%), readymade (5.9%), and the least purchased is oils and sauces (4.3%).

Pie chart of food purchase by weight in 2015



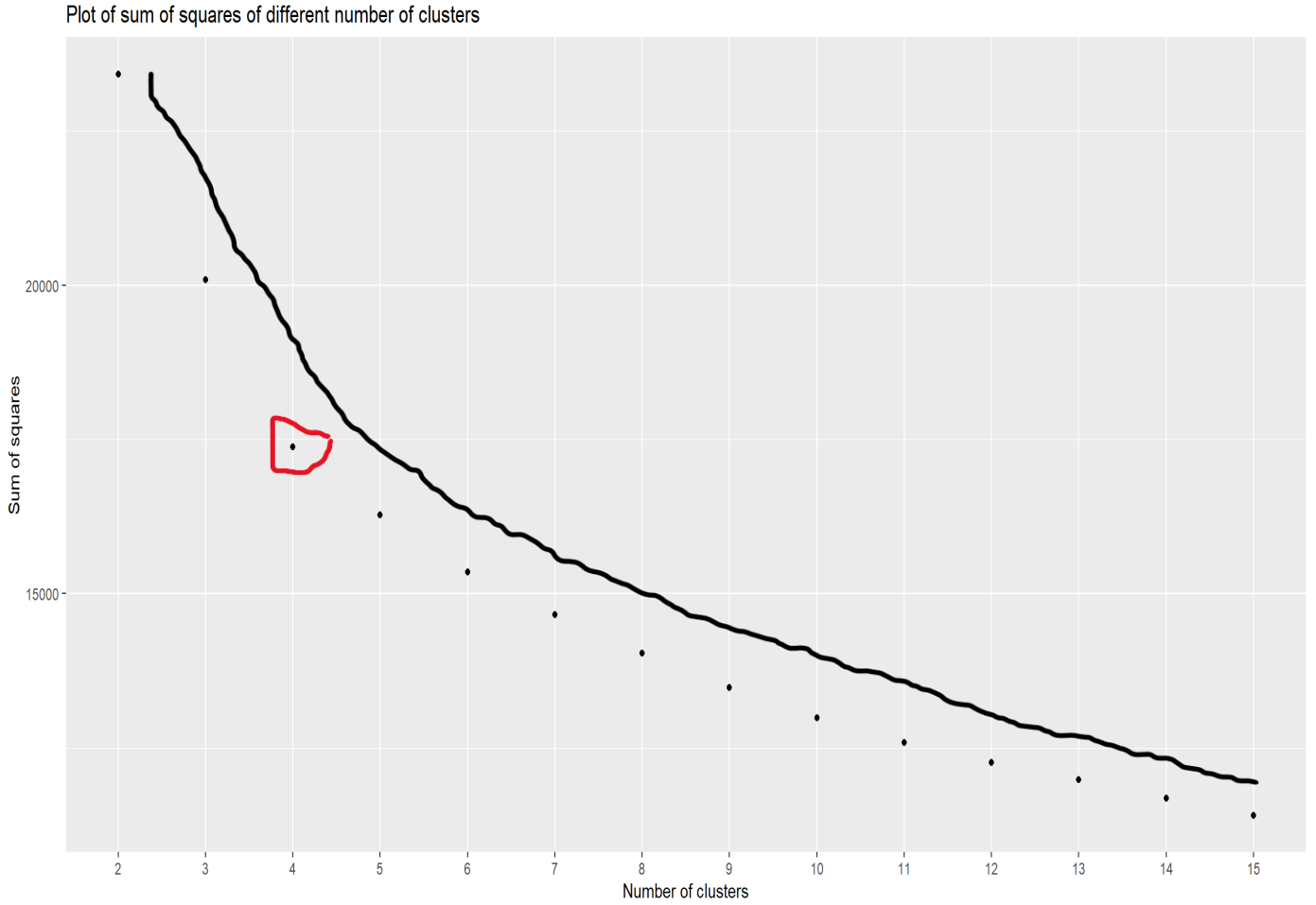
Data Analysis – Findings and Discussion

All areas are grouped into 4 clusters based on the difference in weight of food purchase in 7 categories



Cluster validation

The “elbow point” is at 4, so 4 is a good k value to choose.



Cluster Analysis

Cluster	Description
1	High consumption of readymade and sweets and low consumption of dairy and eggs.
2	High consumption of fruit and vegetable and dairy and eggs and low consumption of grains.
3	High consumption of oils and sauces and fish and meat.

Conclusion

- Managed to achieve objective of data analysis, which is clustering of different areas based on difference in food purchase

Coventry University

END

Ranked 12th UK University
Guardian University Guide 2018

4th for Student Experience and Teaching Quality
The Times and Sunday Times Good University Guide 2017

Queen's Award for Enterprise
International Trade 2015