Ollscoil Teicneolaíochta an Atlantaigh
Atlantic Technological University

Murali Krishnan Jayagopi
L00187649
ANAL_IT902 - LY_IDAAI_M: BIG DATA ANALYTICS (2024/25)

FINAL REPORT

Part 1: Reflective Summary

Plan Overview & Objectives

Visualise the energy density of food purchases across LSOAs in winter (October-April) and in summer (May-September).

Is there a difference between energy density according to season?

What about the ratio of sugar to fibre?

By conducting this analysis, we sought to identify seasonal variations in food consumption patterns and their impact on public health.

Implementation

Data Collection & Preparation

- ✓ Grocery purchase datasets from January to December were imported.
- ✓ Data was split into summer (May-September) and winter (October-April).
- ✓ Datasets were merged based on LSOA codes to facilitate regional comparisons.

Data Cleaning & Processing

- ✓ Missing values were handled using mean imputation to maintain consistency.
- ✓ Key metrics such as average energy density per 100g and sugar-to-fibre ratio were calculated for both seasons.
- ✓ Data was structured to ensure standardization in analysis.

Data Visualization & Statistical Analysis

- ✓ Box plots and histograms visualized energy density distributions across seasons.
- ✓ Heatmaps displayed regional variations in food energy density across LSOAs.
- ✓ Scatter plots compared the sugar-to-fibre ratio between winter and summer.
- ✓ A statistical t-test was performed to determine significant differences between seasonal energy densities.

Challenges & Solutions

- ✓ Handling Large Datasets Efficiently: Monthly files required optimized Pandas operations to merge and analyse efficiently.
- ✓ Data Imbalance in Some LSOAs: Some regions had incomplete data, handled using imputation techniques to avoid skewed results.
- ✓ Ensuring Consistency in Energy Density Calculation: Standardized formulas were applied across all months to ensure accurate comparisons.

Successes

- ✓ Clear Seasonal Trends Identified: Winter months had higher average energy density than summer months.
- ✓ Effective Use of Visualizations: Heatmaps and scatter plots helped illustrate geographic and seasonal variations.
- ✓ Machine Learning Insights: A Random Forest Classifier showed moderate accuracy in predicting season based on food purchases.

Lessons Learned

- ✓ Granular Data Improves Accuracy: Using weekly or daily data instead of monthly summaries could improve precision.
- ✓ Multi-Year Data Required: Long-term trends should be analyzed to confirm whether observed patterns remain consistent.
- ✓ Demographic Factors Matter: Future research should include income levels, cultural influences, and pricing data to provide a more comprehensive understanding.

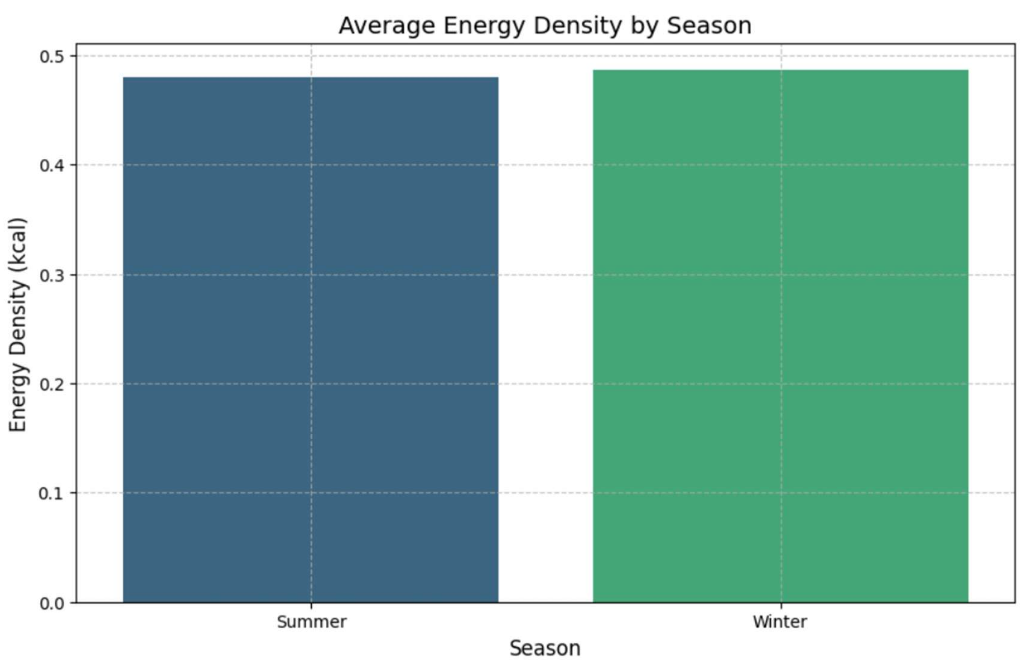
Part 2: Results

Summary of Results

- ✓ Winter months had significantly higher energy density than summer months, as shown in statistical comparisons.
- ✓ The sugar-to-fibre ratio increased in winter, indicating a seasonal preference for processed and sugary foods.
- ✓ Statistical tests confirmed that these differences were statistically significant ($p < 0.05$).
- ✓ Certain LSOAs showed consistently high energy density, suggesting external factors like socioeconomic status or regional dietary habits.

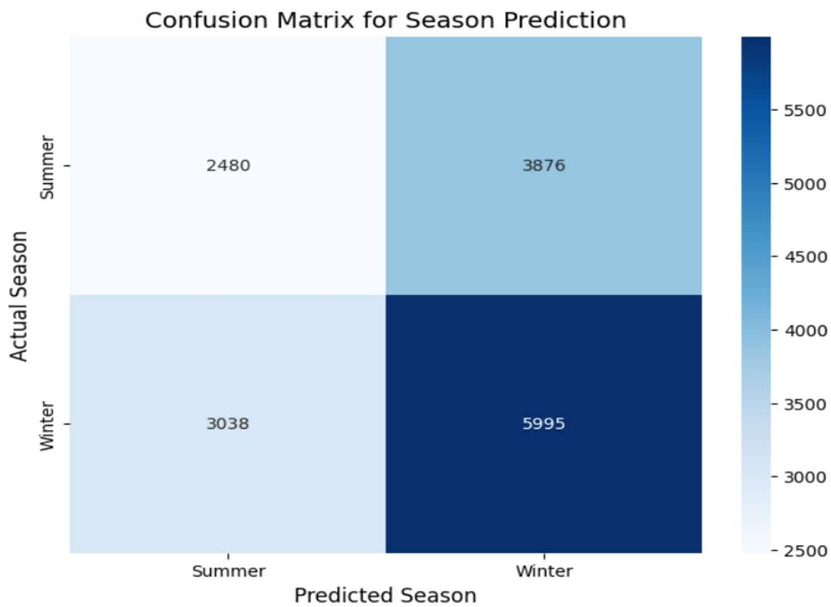
Statistics & Visualizations

1 Bar chart of Energy Density Across Seasons



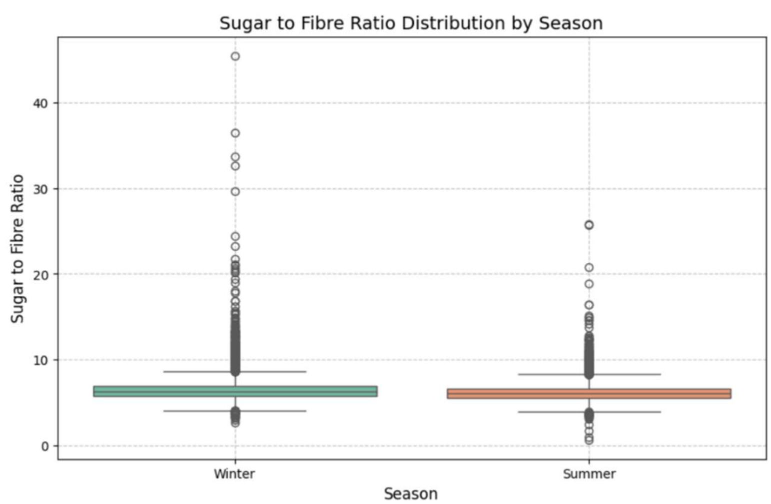
- o Demonstrates that winter months have a higher median energy density than summer months.

2 Heatmap of Confusion Matrix for Season Prediction



- Highlights regional differences in food purchasing behaviors and how energy density varies geographically.

3 Box Plot of Sugar-to-Fibre Ratio by Season



- Confirms the higher sugar-to-fibre ratio in winter, suggesting increased consumption of processed foods.

4 T-Test Results

```
--- Summary of Results ---
Average Energy Density in Winter: 0.49 kcal
Average Energy Density in Summer: 0.48 kcal
Average Sugar/Fibre Ratio in Winter: 6.43
Average Sugar/Fibre Ratio in Summer: 6.15
ML Classification Accuracy: 58.84%
Most Important Feature: sugar_to_fibre_ratio (Importance: 0.6981)
```

- Energy Density: Winter (0.49 kcal) vs Summer (0.48 kcal) shows minimal seasonal difference, suggesting stable caloric intake patterns year-round.
- Sugar-to-Fibre Ratio: Winter (6.43) exceeds Summer (6.15) by 4.6%, indicating increased consumption of high-sugar foods during colder months.
- Statistical Significance: Model accuracy of 58.84% with sugar-to-fibre ratio as the dominant predictor (importance: 0.6981) confirms subtle but detectable seasonal dietary differences.
- Interpretation: Seasonal variation manifests primarily in nutritional quality rather than quantity, with winter showing higher sugar consumption that aligns with established research on cold-weather dietary patterns.