The datasets are divided into multiple folders, one of which includes data from 420 million food items purchased by 1.6 million loyalty card owners from Tesco in the greater London area during 2015. This dataset captures the average weight and volume of nutrients like fat, sugar, protein, fiber, saturated fats, carbohydrates, and salt found in food and drink items, segmented by different area sizes—Borough, Ward, MSOA, and LSAO—on a monthly and annual basis. The focus for this case study was on the yearly averages for Boroughs and Wards. Additional datasets include child obesity rates, diabetes estimates, general obesity, and obesity hospitalization rates across London boroughs and wards, with this study specifically using the datasets on diabetes estimates in Wards and obesity rates in Boroughs.

The initial analysis examined the correlation coefficients between obesity and the average nutrient weights of fat, sugar, protein, fiber, saturated fats, carbohydrates, and salt, which were 0.06, 0.40, -0.25, -0.15, -0.01, 0.45, and 0.11, respectively. The correlation coefficient is a statistical measure that quantifies the strength of a linear relationship between two variables, ranging from -1 to +1. A correlation of +1 indicates a perfect positive linear relationship, while -1 indicates a perfect inverse relationship.

The findings indicate that protein and fiber are inversely related to obesity, with protein having a lesser effect. Conversely, the other nutrients—fat, sugar, carbohydrates, and salt—showed positive correlations with obesity, suggesting a direct relationship. However, due to the very low correlation value (-0.01) for saturated fats, it can be assumed to have negligible impact on obesity. This suggests that increased consumption of the other nutrients may be associated with higher obesity rates, whereas saturated fats appear to be less of a concern in this context.

The second analysis, which focused on overweightness, suggested that all nutrients are correlated with being overweight. The correlation values for sugar, carbohydrates, fat, fiber, saturated fats, salt, and protein were 0.44, 0.41, 0.34, 0.33, 0.22, 0.14, and 0.08 respectively, arranged from the most to least impactful. This underscores the significant role that nutrient intake plays in both obesity and overweightness, with varying degrees of influence depending on the nutrient.

The third analysis focused on Type-2 Diabetes, revealing that the correlation coefficients between diabetes and both fiber and protein were -0.31 and -0.52, respectively. This suggests that both nutrients are inversely related to Type-2 Diabetes, with protein having a more pronounced effect. Conversely, carbohydrates, sugar, saturated fat, and fat demonstrated positive correlations with Type-2 Diabetes, suggesting a direct relationship with values of 0.65, 0.47, 0.36, and 0.34 respectively, listed from the most to the least impactful. The correlation with salt was a negligible 0.01, indicating it has a minimal impact on Type-2 Diabetes.

Does the executive summary concisely state the nature of the data being used?

4 Does the executive summary concisely state the two insights produced?

4 Does the executive summary concisely state any observation about the relationship of the two datasets and provide a reasonable conclusion? 4

4 References Have appropriate references been provided in the expected style?

The datasets provided are split into multiple folders. One contains datasets from 420M food items by 1.6M loyalty card owners in the greater London area, that were purchased from Tesco from the year 2015, this was tracked using Tesco’s loyalty cards (Club Cards). This contains the average weight and volume of nutrients found in food and drink items bought per person per trip. It also contains other information such as number of people in a certain age group, the average age and other information.

This case study however has only focused on average weights of fat, sugar, protein, fiber, saturated fats, carbohydrates, and salt. These were split into different averages for different area levels which were Borough, Ward, Middle Layer Super Output Area (MSOA) and Lower Layer Super Output Area (LSAO) which are respective of the size of the area from largest to smallest. These averages were also given in a month-to-month basis and as an average for the whole year. In this case study, the focus was yearly Borough and Ward data. Other datasets provided contained datasets containing values for child obesity in London boroughs and wards, diabetes estimates in wards**,** Obesity in boroughs, and obesity hospitalisation in boroughs. The study focused only on diabetes estimates in Wards and Obesity in Boroughs.