EXECUTIVE SUMMARY

OVERVIEW:-

The Dataset gives information of 411 Tesco store purchases along with nutritional properties for four regions Borough, Ward, LSOA (Lower Super Output Areas), MSOA (Medium Super Output Areas) in greater London of 420 million Clubcard owners. In addition to the Tesco files other datasets are taken from ONS (office for national statistics) to create insights.

ASSUMPTIONS AND LIMITATIONS:-

- The data provided exits only to Clubcard owners and may not fully represent the food purchasing or dietary habits of the entire population.

- Additionally, the number of stores may differ by region, but population is almost same when data is considered for an entire year.

- Used only 2015 income data for our analysis of last insight and values mentioned in the income dataset are calculated after taking into account that they paid taxes, have saving and spending.

IMPORTANT RESULTS/CONCLUSIONS :-

- I used food categories, nutrient information, and diabetes data from the datasets to examine correlations between them for Ward region. To find the relationship between food categories and diabetes, I used the Spearman correlation and observed a positive correlation between certain foods, such as sweets and soft drinks, which are known to increase the risk of diabetes according to the World Health Organization. I also found that consuming high-protein, fibre foods was associated with a lower risk of developing diabetes. However, it's important to note that this does not mean that one must avoid certain foods altogether, but rather consume them in limit.

- I also examined trends in alcohol consumption, beer and wine purchases during specific months. I found that alcohol consumption is generally high during the Christmas season. Additionally, I observed that people tend to purchase more beer during the summer months compared to other seasons. Moreover, wine consumption in the UK has been increasing over the years, and also second largest importer of wine according to Wine and Spirit Trade Association (WSTA) 2019 report.

- I examined the trends in income for each borough area over the years. I found that the income of individual households has been increasing over the years. Additionally, I explored the correlation between income and food purchases and found that people with higher incomes tend to purchase more healthy food such as dairy, fruits, vegetables, and fish, while those with lower incomes tend to buy more unhealthy foods such as grains and sweets. One of the reasons that fresh foods cost more due to transportation, production and maintenance expenses. As a result, people with higher incomes are more likely to be able to afford these foods. However, it's important to note that people with lower incomes may also consume fresh foods by sacrificing other things.

After analysis the dataset I understand that how data plays an important role for finding trends, correlations and even predictions of certain things and also finding the right dataset for creating insights is more important rather then simply comparing the results without considering all the bias and limitations.

BIBLIOGRAPHY

1. Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London, Aiello, L.M., Quercia, D., Schifanella, R. et al. Sci Data 7, 57 (2020).[accessed on 19 April 23)]; Available online: https://www.nature.com/articles/s41597-020-0397-7

2. Regional gross disposable household income: local authorities by ITL1 region.[(accessed on 19 April 23)]; Available online: https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomelocalauthoritiesbyitl1region

3. Drinking trends in the UK, [(accessed on 19 April 23)]; Available online: https://alcoholchange.org.uk/alcohol-facts/fact-sheets/drinking-trends-in-the-uk

4. The Wine and Spirit Trade Association Our Vision, Your Voice, [(accessed on 19 April 23)]; Available online: https://alcoholchange.org.uk/alcohol-facts/fact-sheets/drinking-trends-in-the-uk

5. pandas 0.16.2 documentation[(accessed on 19 April 23)]; Available online: https://pandas.pydata.org/pandas-docs/version/0.16.2/generated/pandas.core.groupby.DataFrameGroupBy.plot.html

6. Pandas sort dataFrame by month name,Kanoki,[accessed on 19 April 23)]; Available online: https://kanoki.org/2022/06/27/pandas-sort-by-month-name/

7. Style sheet reference for matplotlib[(accessed on 19 April 23)]; Available online: https://matplotlib.org/stable/gallery/style_sheets/style_sheets_reference.html

8. https://www.youtube.com/watch?v=9VK8quGFcSE