

COMP4621 – Programming for Data Science

Semester 1, 2024/2025

Project

DUE: Wed. December 04, 2024 at 21:30hrs

Presentations: Wed. December 11, 2024 at 17:30–21:30

Changing viewpoints of the general population about Covid-19 has influenced how businesses decide to make investments and strategize for growth. A new convenient food shop franchise has asked for your help in assessing whether it is an appropriate time and economically appropriate step to begin their expansion overseas. They have been considering different regions of the world but would also be satisfied with country specific guidance. Although they have depended on traditional accounting and economic indicators, they would like to augment their strategic data with additional indicators to include current sentiments surrounding Covid-19 and profit maximization potential and purchase patterns in different regions / countries.

NOTE: Marks will be apportioned based on results and creative ways of presenting results.

Task I – [10 marks]

- a) Retrieve the archive provided and select tweets from three separate days in August 2021 (all separated by at least 6 days) and conduct data cleaning and preparation. [3]
- b) Generate the *sentiment* for each date selected, extract tweets with subjectivity ≥ 0.4 and analyse the dominant sentiment for each day (you are allowed to change the cut-off value for subjectivity, with explanation). What is the overall dominant sentiment? [5]
- c) Conduct ONE additional analysis of the tweets/sentiments not specified in this section. Include a visualization. [2]

TASK II – [35 marks]

Download the dataset specified for your team from VLE and conduct the specified analysis below.

a) Explore, Prepare Data

- i. Explore the data and provide details on all fields retrieved. You should ensure each feature in the dataset (each column) is reviewed and summarized to verify things such as value ranges, missing values etc. Be sure to generate relevant representations/explanations as necessary to demonstrate your review and decision making. [4]
- ii. Fix noise, outliers and any other issues discovered (example: missing values). You must provide discussion / explanation of all activities done and why each decision has been made. [4]

b) Analysis & Recommendations

Write code to conduct analysis that will answer the questions below. You are encouraged to ensure the results are clearly presented. Additionally, your code should be shown along with each question, the result and notes that explain the results.

- i. What is the average spend on vegetables vs. meat in each region? [3]
- ii. Which country has the highest spending on vegetables? [2]
- iii. Which country consumes the most meat? [2]
- iv. What is the average profit from the sale of meat vs. vegetables in each region? [4]
- v. What has been the total revenue from meat vs. vegetables for each year since 2012? [4]
- vi. Plot a graph of the average revenue from meats and vegetables for any 12-month period in the dataset. [4]
- vii. What is the average units sold for each item in the dataset (not just meat and vegetables) by region? Is vegetables or meats units sold above the overall mean for units sold? [4]
- viii. In which season (Spring, Summer, Autumn, Winter) do persons spend the most meat vs. vegetables? [4]

Task III – [5 marks]

Now that you have conducted your analysis, provide recommendations using results as evidence to answer the following:

- a) When is the best time to open the franchise given that their primary products are made from meat and vegetables?
- b) Which country/region provides the best opportunity for profit maximization based on the data analysed?
- c) Are there any limitations, ethical issues to declare in relation to your results or analysis?

Submission

- 1. A single file (PDF generated by markdown) should be submitted before the stated due date.
- 2. You will be required to make a presentation (10 minutes per team), which should summarize, but give sufficient details of your project work. The submitted pdf notebook can be used to do the presentation without prejudice.