TASK 1: DATA AGGREGATION AND ANALYSIS

1.1 HERE IS THE BACKGROUND INFORMATION OF THE TASK

CBA is embarking on a long-term vision with trusted partner InsightSpark, a data science specialist.

This vision is focused around using the huge volumes of CBA's transactional data, open-source data, and advanced data science to build a platform that will provide insights for businesses, governments, and investors in Australia.

The data engineering team is responsible for building data engineering pipelines, running analysis on big data, and deploying algorithms into production at scale. This project's success lies heavily on the data engineering team.

1.2 HERE IS THE TASK

The first task is to use the provided "CSV Data Set" (You can find it in the Resources) of transactions at a supermarket and analyse the data to answer some questions. We suggest you complete this task in Excel or an open-source alternative like LibreOffice Calc.

Task details and the answers you need to find are listed below.

Hello,

I have provided you with a data set named "supermarket_transactions.csv." This data set contains three years of transactional data we collected from supermarkets across Australia.

InsightSpark wants to access our data to perform some analysis across Australian supermarkets. Before this can happen, I need you tom perform some analysis on the provided CSV data set (a sample of the database) to answer the following questions:

- Across locations, how many apples were purchased in cash?
- How much total cash was spent on these apples?
- Across all payment methods, how much money was spent at the Bakershire store location by non-member customers?

You should conduct your analysis in Excel or a spreadsheet program of your choice. In your file, document the formulas used to generate your values.

1.3 QUESTION

1. Across locations, how many apples were purchased in cash?

To answer this, question, we filtered the data sheet to include only rows where the “product_name” is “apple” and where the “payment_method” is “cash”. Then, we summed the “quantity’ column to get the answer of 117 apples.

2. How much total cash was spent on these apples?

We left the datasheet filtered to include only rows where the “product_name” is “apple” and where the “payment_method” is “cash”. Then we summed the “total_amount” column to get an answer $537.03.

3. Across all payment methods, how much money was spent at the Bakershire store location by non-member customers?

For this question, we cleared all the previous filter criteria. Then, we filtered the data sheet to include only rows where the “customer_type” was non-member, and the “store” was Bakershire. Then, we summed the total_amount” column to get and answer of $2,857.51