

TASK ONE

What you'll learn

- Understand the importance of data engineering and analysis in a long-term data-focused vision.
- Gain experience in using real-world transactional data for analysis.
- Learn to perform data analysis and answer specific questions based on the dataset provided.

What you'll do

- Analyze the "CSV Data Set" of supermarket transactions to address key questions.
- Document the analysis process, including formulas used, using spreadsheet software.
- Submit the analyzed file for review.

Here is the background information on your 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.

Here is your task

Your 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 in an email from the team lead:

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 to 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. When finished, please submit the analysed file for me to review before we share our findings with InsightSpark.

Example Answer

Great work! Take a look at the example answer below to see how a professional would have attempted this task. Think about what you did well and how you can improve.

Example Answer Explanation

The example answer is just one way to approach this problem, not the only solution. For this question, first, we opened the CSV file `supermarket_transactions.csv` in Excel. Then we aggregated and analysed the data using filters and formulas to answer the following questions:

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 an answer of **117 apples**.

Question 2: How much total cash was spent on these apples?

Here, we left the data sheet 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 of **\$537.03**.

Question 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 an answer of **\$2,857.51**.