1. Identify top 10 pet supply products

You are a junior data analyst at a major online pet supplies business. The product development department has asked you to identify the top 10 products with the highest sales in the last month so that they can decide what types of products to develop next. These products come from multiple tables within the company's database: one containing product details and another with sales data.

Reflection

Consider whether you would use a spreadsheet or SQL in the scenario above.

- Which tool would provide the desired output?
- Which tool is appropriate considering the dataset's size and complexity?
- How will you access the data?

Now, write 2-3 sentences (40-60 words) in response to these questions. Identify which tool you would use and then justify your thinking. Enter your response in the text box below.

In this case, SQL is the best tool for the job. In order to assess the top 10 products in the last month, you need to be able to access the data stored in the company database, easily join tables, filter by date, and order by the sales volume. SQL is a great tool in situations in which you need to be able to interact with a database or find yourself working with larger, more complicated datasets that are hard to explore and transform in a spreadsheet.

2. Determine new store location

In this scenario, you work for a small fitness tech company. Your manager has asked you to determine which cities have a large population of customers; this will help them decide where to build the next retail store. You have a .csv file with 300 names, phone numbers, and addresses. The manager also requested a quick, simple visualization, such as a bar chart, so the team can make some quick comparisons.

Reflection

Consider whether you would use a spreadsheet or SQL in the scenario above.

- Which tool would provide the desired output?
- Which tool is appropriate considering the dataset's size and complexity?
- How will you access the data?

Now, write 2-3 sentences (40-60 words) in response to these questions. Identify which tool you would use and then justify your thinking. Enter your response in the text box below.

The most appropriate choice is a spreadsheet, such as Google Sheets or Excel. With a spreadsheet, you can easily import the .csv file, use pivot tables or `COUNTIF` functions to tally customers per city, and visually display the data using a bar chart. In this case,

the dataset you need is already stored in a spreadsheet and is small enough that you don't need to worry about moving it to another tool. Spreadsheets are great for working with smaller datasets directly in one file and generating simple visualizations to communicate data findings quickly with stakeholders.

3. Customer satisfaction survey results

The marketing team of a fashion retailer conducted a survey on customer satisfaction. The results are in a .csv file. The team wants to calculate the average satisfaction score and compare it to the average score from their last survey. They also want to segment responses based on demographic data such as age, gender, and location. This segmentation will enable the marketing team to tailor their marketing strategies to meet the specific needs and preferences of various customer groups.

Reflection

Consider whether you would use a spreadsheet or database in the scenario above.

- Which tool would provide the desired output?
- Which tool is appropriate considering the dataset's size and complexity?
- How will you access the data?

Now, write 2-3 sentences (40-60 words) in response to these questions. Identify which tool you would use and then justify your thinking. Enter your response in the text box below.

The most appropriate choice is a spreadsheet. A spreadsheet is an effective tool for this task because it allows for easy data segmentation, average calculations, and even pie charts to visualize the distribution. The data is already housed in a .csv file, so it can be worked with directly and easily where it is.

4. Calculate course completion rates

The manager of an online education platform wants to know which courses have the lowest completion rates so that they can invest in course enhancements and more targeted student support for those courses. The data is stored in a relational database with separate tables for user registration, course enrollment, and course completion.

Reflection

Consider whether you would use a spreadsheet or database in the scenario above.

- Which tool would provide the desired output?
- Which tool is appropriate considering the dataset's size and complexity?
- How will you access the data?

Now, write 2-3 sentences (40-60 words) in response to these questions. Identify which tool you would use and then justify your thinking. Enter your response in the text box below.

The most appropriate choice is SQL. Using SQL, you can join these tables, calculate the completion rate for each course, and filter to find the courses with the lowest completion rates in the database within which they're already stored. SQL is a great tool for working directly with databases to query data to find the answers you need.