**CASE STUDY**

You’re a part of the data analytics team at Adventure Works. The lead data analyst informs you that the company wants an in-depth analysis of sales data from the past year to understand sales trends and customer preferences. This isn't a routine analysis. The company is close to launching a new product line next quarter. Your analysis and insights will guide this massive product launch, including the marketing strategy, sales approach, and even the final tweaks to the product! Each number, trend, and pattern you uncover from the dataset could directly impact hundreds of decisions.

You are assigned some initial tasks related to stakeholder engagement, data collection, and data processing—some early steps in the data analysis process that will contribute to the success of the project.

**Step 1: Identify the stakeholders**

Understanding who the stakeholders are in a given scenario is a critical step in any data analysis project. In this case, you identify stakeholders (individuals or groups) within the company that will be directly affected by, or have a direct interest in, the sales trends and customer preferences that your analysis will reveal. These include the sales, marketing, and product development teams.

A) **Briefly explain why any two of the stakeholder groups at Adventure Works may have an interest in the insights related to the sales performance of various products.**

Several internal stakeholders at Adventure Works may have an interest in the sales performance of various products. These can include:

* Sales team: The sales team is directly responsible for selling products. By understanding the company’s sales performance and sales trends, the sales team can tailor their sales strategies, understand which products are selling well, identify opportunities for improvement, and maximize sales opportunities in preparation for the product launch.
* Marketing team: The marketing team's efforts play a significant role in driving sales. They would be interested in sales data to understand the effectiveness of their marketing strategies and to guide future campaigns that promote sales of the new product line.
* Product development team: The product development team would be interested to check how well their developed products are performing in the market. The sales data of various products can guide them in the development and adjustment of products for launch. For example, they can identify opportunities in the market new products can fill and make data-driven decisions regarding pricing, production volume, and resource allocation for the new line of products.

B) **Which of the stakeholders need to know about customer preferences when it comes to Adventure Works’ products? Why is this information important for them? (Note: Customer preferences refer to customer wants, needs, and expectations regarding products and/or services.)**

Insights into customer preferences are valuable to a number of teams within Adventure Works:

* Marketing team: The marketing team can use this data to target their marketing efforts more effectively and to tailor their messages to better connect with their customers, generating interest and demand for the new product line.
* Sales team: Understanding customer preferences will allow the sales team to better tailor their sales strategies or approaches and to recommend products that better meet customers' needs, ultimately driving more sales.
* Product development team: The product development team can use customer preference information to inform the design and development of new products or to refine existing ones. Insights into customer preferences will help the development team ensure that the products align well with what customers want and enable them to make data-driven improvements or tweaks, increasing the chances of market success.
* Customer service team: Knowing customer preferences can help the customer service team provide more personalized service and support, leading to improved customer satisfaction.

**Step 2: Identify and consider the data sources**

A)**In your own words, what kind of information can you gather from the provided dataset?**

The dataset includes a wealth of information about product details, orders, and customers. Here is a brief description of what can be gathered from each field:

* Product details: The **Product ID**, **Product Category**, **Product Subcategory**, **Product Name**, **Product Description**, **Product Price**, **Product Weight**, and **Product Size** fields provide detailed information about Adventure Works’ products. Understanding the variety and specifications of products can help stakeholders gain insights into the performance of specific products and the reasons behind this performance. For example, fields like **Product ID** make it possible to track the performance of each product individually, and product description, weight, and size fields can provide valuable insights into customer preferences and purchasing behaviors.
* Order details: The **Order ID**, **Order Date**, **Order Status**, **Order Quantity**, and **Order Total** fields provide information about each order, including the total cost, the number of products ordered, and the status of the orders. This data can help stakeholders track sales performance over time and identify any trends or issues (such as bottlenecks in production that delay delivery) and gain insight into the overall order processing workflow.
* Customer details: This dataset includes a unique identifier for each customer in the **Customer ID** field that can be used to track individual customer behavior over time. This can help stakeholders understand customer loyalty and repeat purchase behavior, which can help stakeholders gain insight into customer preferences, tailor marketing strategies, and identify opportunities to improve customer retention. The **Payment Method** field provides information about how customers choose to pay for their orders. It can offer insights into customer payment preferences and potentially inform payment options and strategies.

B)**How can you use the data fields provided to understand sales trends and customer preferences? Briefly discuss your suggested approach for gaining insight into these two aspects separately.**

* Sales trends: You can use the **Order Date**, **Order Quantity**, **Order Total**, and **Product ID** fields to understand sales trends. By grouping and summarizing this data by date and product, you can identify trends in product popularity and sales volume over time.
* Customer preferences: You can infer customer preferences from the **Product ID**, **Product Category**, **Product Subcategory**, **Product Size***,* and **Payment Method** fields. By analyzing which products and product categories sell the most, you can gain insight into what customers prefer. The preferred product size and payment method can also provide insights into customer behavior and preferences.

C)**Apart from the provided dataset, what other data sources within Adventure Works might be relevant to your analysis or add value to your insights?**

While the provided dataset is comprehensive, additional data could enhance the analysis:

* Customer details: Additional data about customers such as demographic information (for example, age and location) could provide further insights into customer behavior and preferences.
* Marketing data: Information about marketing campaigns, including dates, marketing channels, costs, and reach, could be used to correlate marketing efforts with sales performance.
* Supply chain data: Information about inventory levels, production costs, and supplier details could be used to understand the cost of goods sold and profit margins. This data could also help stakeholders identify opportunities to manage inventory more effectively.
* Feedback data: Customer reviews or feedback could provide qualitative insights into what customers like or dislike about the products and the ordering process. This could supplement the sales data and provide deeper insight into customer preferences and purchasing behaviors.
* Website analytics: Data from the company's website or e-commerce platform could provide insights into customer behavior online, such as popular search terms or commonly viewed products.

**Step 3: Data import and cleaning**

A)**List and briefly describe a common Excel function that can help you clean each of the following columns:**

* **Product Name**: **TRIM** This function removes leading, trailing, and double spaces in cells. In the *Adventure Works Sales* dataset, it can be used to clean up inconsistencies in text entries and to remove any additional spaces that have occurred during data entry.
* **Product Category**: **PROPER** This function converts the first letter of each word in a text string to uppercase and all other letters to lowercase. In the *Adventure Works Sales* dataset, it can be applied to the **Product Category** column to maintain consistency in data presentation.
* **Product Size**: **UPPER/LOWER** The **UPPER** function converts text to all uppercase and **LOWER** converts text to all lowercase. This could be useful for standardizing entries in **Product Size** columns in the *Adventure Works Sales* dataset.

B) **With reference to the Adventure Works sales dataset, indicate the most appropriate Excel data type for the content of each column.**

|  |  |  |
| --- | --- | --- |
| **Column title** | **Excel data type** | **Explanation** |
| **Product ID** | Number | This column contains unique identification numbers for each product. As IDs are integers (whole numbers), the Excel data type should be set to number, with zero decimal places. |
| **Product Category** | Text | Categories are typically represented by descriptive names (strings of text) making the text data type appropriate here. |
| **Product Subcategory** | Text | Similar to the category, the subcategory is also descriptive and represented as a string, therefore, the text data type is suitable. |
| **Product Name** | Text | Product names are strings of text that describe or label each individual product. The text data type is ideal for this. |
| **Product Description** | Text | Product descriptions provide additional details about each product. They are typically composed of longer strings of text, which makes the text data type the best choice. |
| **Product Price** | Currency | Prices can include decimal values (for example, $19.99). To capture this level of detail, the currency data type is appropriate. |
| **Product Weight** | Number | Because weights often include fractions (for example, 1.25 kg), the number data type, with 2 decimal places, is appropriate. |
| **Product Size** | Text | Sizes can come in various forms such as (S) small, (M) medium, or (L) large. Because they might include text, the text data type is appropriate. |
| **Order ID** | Number | Similar to the **Product ID**, the **Order ID** is a unique identification number for each order, so the number data type with zero decimal places is used. |
| **Customer ID** | Number | **Customer ID** is a unique identifier for each customer, typically represented as a number with zero decimal places. |
| **Order Date** | Date | This field contains the date. Excel’s date data type allows for the correct sorting and time-based analysis of this data. |
| **Order Status** | Text | The **Order Status** column may contain status descriptions such as **Completed** or **Pending**, which are best represented with the text data type. |
| **Order Quantity** | Number | This is the number of products in each order. Because you can't order a fraction of a product, this field is a number data type with zero decimal places. |
| **Order Total** | Currency | This is the total price of each order. This can contain currency symbols, making the currency data type appropriate. |
| **Payment Method** | Text | This column could contain various payment methods like **Credit Card** or **Bank Transfer**. As a result, it should be treated as text. |

C) **Explain how you would exclude the missing entry in the Product Description column in the *Adventure Works Sales* dataset at the source using Excel.**

An appropriate way to exclude the missing entry at the source is to use Excel's filter functionality as follows:

* First, select the **column header** for **Product Description**.
* Next, go to the **Data tab** and select the **Filter button**. This will add drop-down arrows to each column header.
* Select the drop-down arrow on the **Product Description** header.
* In the text box at the bottom of the drop-down menu, type **(Blanks)** and then select **OK**. This will filter your data to only show rows where the product description is blank.
* Select all the rows where the product description is **blank** and delete them, which will exclude the missing entries from your dataset.