**Task 6.1 – Sourcing Open Data; Adidas US sales Analysis**

**Data Source:**

External open-source dataset from Kaggle: <https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset>

**Data Collection:**

Unfortunately, no details about the collection of this dataset have been provided.

**Contents:**

An Adidas sales dataset is a collection of data that includes information on the sales of Adidas products. The dataset includes details such as the number of units sold, the total sales revenue, the location of the sales and the type of product sold.

Data Limitation and Ethics:

**Data Cleaning and Consistency Checks:**

Before working with the file, I removed the Adidas logo, the first 4 empty rows and 1 empty column.

Before:

Graphical user interface, application, table, Excel

Description automatically generated

Before working with the file, I removed the Adidas logo; the first 4 empty rows and 1 empty column.

After:

Table

Description automatically generated with medium confidence

* Checked for duplicates
* Checked for missing values
* Checked for mixed-type columns

**Data Profile:**

Dataset contained 9652 rows and 14 columns in total. After cleaning and checking for consistency, the dataset now has 9648 rows and 13 columns.

|  |  |  |
| --- | --- | --- |
| Column name | Column description | Data Type |
| Retailer | Name of the retailer where the sale was made | Qualitative, Nominal |
| Retailer ID | Specific identifying number of the retailer | Quantitative, Discreet |
| Invoice Date | Date of Sale/Invoice | Qualitative, Ordinal |
| Region | Region of the sale | Qualitative, Nominal |
| State | State of the sale | Qualitative, Nominal |
| City | City of the sale | Qualitative, Nominal |
| Product | Type of product | Qualitative, Nominal |
| Price per Unit | Price of the product per unit, in USD | Quantitative, Continuous |
| Units Sold | Number of units sold | Quantitative, Discreet |
| Total Sales | Total number of sales, in USD | Quantitative, Continuous |
| Operating Profit | Total profit for all units sold, in USD | Quantitative, Continuous |
| Operating Margin | Profit for all units sold, in % | Quantitative, Continuous |
| Sales Method | The way the products were sold | Qualitative, Nominal |

Descriptive analytics of Quantitative data:

Table

Description automatically generated

**Data Limitations and Ethics:**

One of the limitations of the dataset is the lack of details on collection methods. It takes away the credibility of the dataset. However, for the purpose of this project it is not a critical factor as we are using the results for practice and educational purposes only.

The dataset does not contain any PII of the customers. Only sales of the products are recorded.

**Questions for explore:**

* Which products generate most revenue, Profit, Profit Margin?
* Which retailers generate most revenue in each region/state?
* What sales method is most preferred?
* Is the sales method correlated to the type of retailer?
* What seasons generate most sales?