# **Team Members**

|  |  |  |
| --- | --- | --- |
| **Role** | **ID** | **Name** |
| Evaluation and Interpretation | 22010323 | أمنية محمد عبد الهادي أحمد زايد |
| Exploratory Data Analysis (EDA) | 22011609 | نور الدين سامر عبد الرزاق |
| Applying K-medoids Clustering | 22011606 | محمد عماد مبروك |
| Applying Hierarchical Clustering | 22011589 | محمد مصطفى محمد عبد المجيد |
| Dataset Selection & Data Preprocessing | 22011547 | سارة شريف حسن محمد |

Overview:

This dataset contains data of a company that sells a product online. They want to know how well their product sells. We can analyze this data using data mining techniques to segment the customers based on their buying behavior.

Knowing that this dataset has over 540k rows, we decided -for the sake of simplicity- to work on a small subset of 5k rows from the original dataset.

We’re working on 8 columns in total:

* **InvoiceNo:** Unique identifier for each retail or transaction.
* **StockCode:** Code for the specific product or item being sold.
* **Description:** Name of the product or item being sold.
* **Quantity:** Number of units of the product purchased in each transaction.
* **InvoiceDate:** Date and time of each transaction.
* **CustomerID:** Unique identifier for each customer.
* **Country:** Name of the country where the customer is located.

The objective of this project is to build a clustering model to segment the customers based on their similarities and for that we will apply RFM analysis and cluster the data based on their score.