**TEAM-5**

**Supply Stream Insights**

**Leveraging Data Science for Strategic Advantage**

**Topic Proposal:**

The dataset for this project originates from a rich supply chain and sales context, aiming to explore various facets of customer behaviour, product performance, and logistical efficiency. It includes a wealth of data entries and features that span sales information, shipping details, and product categories. The primary research question for this project is: How can data-driven insights enhance operational efficiency and customer satisfaction in supply chain management?

[More about Dataset](https://docs.google.com/document/d/1jhfU20UOilwIJcQ1pM2GxKAdtwqi4aHEzxlYOQRYkx0/edit)

**SMART Questions:**

* What factors most significantly impact delivery times, and how can we mitigate the risk of late deliveries?
* Can customer purchasing patterns be predicted to enhance inventory management and marketing strategies?
* How does the product category relate to sales performance, and can this insight inform inventory stocking?
* What is the relationship between shipping mode choice and customer satisfaction or sales benefit?
* Is there any significant difference in average order profit between regions?

**Modelling Methods:**

* Descriptive Analysis: Summary statistics, correlation analysis, and data visualizations
* Customer Segmentation: K-Means Clustering and Hierarchical Clustering
* Sales and Shipping Forecasting: Linear regression and ARIMA models
* Delivery Risk Prediction: Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM)
* Advanced Predictive Modeling: Gradient Boosting Machines (GBM) and XGBoost
* Optimization: Linear programming for supply chain optimization

**Source:**

The dataset for this project is obtained from Kaggle.  
<https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

**GitHub:**

<https://github.com/ShikhaAtGitHub/data_mining_project>