**REPORT**

Introduction

In this task we have used e-commerce dataset. [C:\Users\SHREYA\Desktop\Datasets\data.csv](file:///C:\Users\SHREYA\Desktop\Datasets\data.csv)

Tools used: Excel, SQL, Jupyter Notebook

Phase1

Imported this dataset into SQL and completed Phase1 i.e. Data Preparation. We looked into datatypes of all columns, removed nulls, etc. Other than this, also went through return rate by country and marketing channel.

Phase2

In this part we imported cleaned dataset into jupyter notebook and performed EDA part during this phase. Also found out about overall return rate and return rate by country and marketing channel which is same as what we found in SQL during Phase 1. Also top 10 reasons for returning the product.

Phase3

In this phase we are doing Return Prediction using Logistic Regression. We did modelling part and then model evaluation is giving us 98% accuracy which is really good. Our model is doing correct prediction. Then concluded this phase with high-risk products and then exported these products to a csv file.

Phase4

In this part, we will be a building an interactive dashboard in Power BI. In this, we made some kpi cards, some bar charts, new measures, slicers etc.

INSIGHTS:

1. 5 Million products were ordered in total.
2. 485000 products were returned.
3. So, the overall return rate for whole data comes out to be 1.96%.
4. USA is giving us maximum returns followed by Czech Republic and Malta
5. All marketing channels are having nearly equal returning rate but highest of all is Affiliate.