**BANA 8083 - 006 MS Capstone  
Project Proposal**: **Amazon Sales Analytics Hub - Exploring Product Data**

**Problem Statement:**

Retailers on Amazon face challenges in understanding customer preferences, optimizing pricing strategies, and evaluating product performance across diverse categories. Current methods often lack the capability to provide comprehensive, real-time insights into sales data, making it difficult to make informed decisions. An interactive and user-friendly platform is needed to analyze and visualize sales data, helping retailers address these issues effectively.

**Objectives:**

The primary objectives of this project are to compare the performance of various product categories, examine pricing strategies and discount patterns, analyze the distribution of product ratings, and correlate the number of ratings with discounted prices. Additionally, the project aims to enable dynamic data exploration through an interactive platform, allowing users to gain actionable insights and make informed decisions.

**Technologies:**

Python with Pandas for data analysis, Scikit-learn for EDA, Matplotlib/Seaborn and Tableau for visualization, Streamlit for interactive application development, and AWS for cloud hosting.

**User Capabilities:**

Upload sales data, get data statistics, compare category performance, identify trends, visualize data interactively, perform correlation analyses, conduct custom analyses, filter data, set alerts for key metrics, and download reports and export visualizations.

**Datasets:**

The project will utilize the [Amazon Products Sales Dataset](https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset) from Kaggle, which contains comprehensive sales data for various product categories on Amazon. This dataset includes information such as product names, categories, ratings, number of ratings, prices, and discounts.

**Scope:** Analyze various product categories, provide real-time insights and trends, and develop an interactive platform for data exploration.

**Limitations:** Dataset may not cover all product variations and seasonal changes, real-time updates depend on data source refresh rates, and potential biases in user ratings and reviews.