**University of Texas at Dallas**

**Report**

**Analytics Practicum – BUAN 6390.501**

**Milestone 1-- Data-Driven Organizations.**

**Professor: Sunela Thomas**

**Tavallaee Melody**

**Harshita Yakkanti**

**Bhavith Akula**

# Index

1. **Project Proposal & Data Preparation** 
   1. **1 Problem Statement**

**○ 2 Objectives**

**○ 3 Methodology**

**○ 4 Data Collection & Preprocessing**

**○ 5 Preliminary Data Analysis**

1. **Analysis & Model Development** 
   1. **1 Exploratory Data Analysis (EDA)**

**○ 2 Customer Segmentation (K-Means Clustering)**

**○ 3 Feature Engineering for Predictive Modeling**

**○ 4 Model Selection: Linear Regression & Ridge**

**Regression**

**○ 5 Model Implementation & Initial Results**

1. **Draft Results & Model Evaluation** 
   1. **1 Performance Evaluation Metrics**

**○ 2 Feature Importance & Error Analysis ○ 3 Model Limitations & Bias Assessment**

**○ 4 Interpretation of Results & Business Insights**

1. **Final Report & Presentation ○ 1 Summary of Findings** 
   1. **2 Business Recommendations**

**○ 3 Future Improvements & Next Steps**

**○ 4 Ethical Considerations & Integrity Statement**

**○ 5 Presentation Overview**

**Customer Segmentation Analysis:**

# Project Proposal

## Problem Statement

Customer segmentation is a critical aspect of retail business strategy. Businesses can optimize marketing efforts, improve customer engagement, and increase sales by identifying distinct customer groups based on purchasing behaviour. The goal of this project is to conduct a **comprehensive customer segmentation analysis** on a real-world retail dataset to identify distinct customer groups and predict future purchasing behavior using multiple analytical methods to grow the business.

## Objectives

* Identify distinct customer segments based on purchasing behavior, demographics, and engagement levels.
* Analyze **customer demographics**, spending patterns, and shopping preferences.
* Provide data-driven recommendations for personalized marketing strategies.
* Enhance customer engagement and retention through targeted promotions.

**Methodology:**

1. **Data Collection:** Utilize Customer Shopping Dataset - Retail Sales Data to analyse behavioural trends.
2. **Data Preprocessing:** Clean and transform the dataset to handle missing values, standardize formats, and remove inconsistencies.
3. **Exploratory Data Analysis (EDA):** Identify key trends and patterns in customer behaviour.
4. **Customer Segmentation Models (Not limited to):**

○ **RFM (Recency, Frequency, Monetary) Analysis**

○ **Clustering Algorithms (K-Means, Hierarchical Clustering)**

1. **Insights & Business Recommendations:** Develop targeted marketing strategies based on segmentation results.

## Expected Outcomes

* Clear **customer segments** that reveal behavioral differences.
* Insights into **high-value customers** and **low-engagement customers**.
* Predictive models that forecast **customer spending**.
* Actionable **marketing strategies** tailored for each customer segment
* The marketing recommendations to drive revenue and customer loyalty.
* Improved business decision-making based on data-driven analysis.

# Data Preparation

## Data Collection & Sources

The dataset consists of 99,457 customer transactions, including attributes like age, gender, purchase history, product categories, quantity, price, and payment methods.

* **Customer Demographics:** Age, Gender, Location
* **Purchase Behavior:** Item Purchased, Category, Purchase Amount, Previous Purchases
* **Engagement Metrics:** Subscription Status, Discount Usage, Frequency of Purchases, Payment Method

**Overall Data Description:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column Name** | **Data Type** | **Description** | **Example**  **Value** | **Unit (if applicable)** |
| **customer\_id** | String (ID) | Unique identifier for each customer | C241288 | N/A |
| **gender** | Categorical  (String) | Gender of the customer (Male/Female) | Female | N/A |
| **age** | Integer | Age of the customer | 28 | Years |
| **category** | Categorical  (String) | Type of product purchased  (Clothing, Shoes,  Books, etc.) | Clothing | N/A |
| **quantity** | Integer | Number of items purchased | 5 | Count |
| **price** | Float | Total price of the purchase | 1500.4 | Currency  (e.g., USD,  INR, etc.) |
| **payment\_method** | Categorical  (String) | Payment mode used (Credit Card, Debit Card, Cash) | Credit  Card | N/A |
| **invoice\_date** | Date  (MM/DD/YYYY) | Date of the transaction | 5/8/2022 | N/A |
| **shopping\_mall** | Categorical  (String) | Name of the mall where the transaction took place | Kanyon | N/A |

## Data Cleaning & Preprocessing

* **Checked for Missing Values:** No missing values were found in the dataset.
* **Checked for Duplicates:** No duplicate records were detected.
* **Standardized Text Data:** Converted categorical values to lowercase for consistency.
* **Ensured Correct Data Types:**

○ Converted purchase amounts and review ratings to numeric types.

○ Standardized categorical features for uniformity.

**Preliminary Data Analysis:**

This preliminary data analysis confirms that the dataset is **clean, structured, and rich in information**. The insights from demographics, spending behavior, and purchase trends serve as a strong foundation for **customer segmentation using clustering techniques** and **predictive modelling using Linear and Ridge Regression and multiple methods to find the best outcome**.

# Next Steps

* Perform advanced clustering techniques and validate customer segments.
* Develop detailed business insights based on segmentation results.
* Recommend personalized marketing strategies and engagement plans.

**Dataset:**

**https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset**