**BLACK FRIDAY SALES PREDICTION:**

**Objective:**  
This project focuses on predicting the purchase amount customers are likely to spend on Black Friday. By leveraging demographic and product data, the model aims to forecast spending patterns, enabling retailers to optimize inventory and personalize marketing efforts.

**Problem Statement :**

The goal is to create a predictive model that estimates a customer’s purchase amount based on demographic and product attributes. This will assist retailers in identifying high-value customers and tailoring offers to maximize profit.

**Dataset Description**

The dataset includes anonymized information on customer demographics and purchase details, consisting of:

* **Demographics**: Age, Gender, City Category, and Stay Duration in the Current City.
* **Product Categories**: Features related to product types purchased.
* **Purchase Amount**: Target variable indicating customer spending.

Each feature provides valuable insights into factors influencing customer purchases.

**Exploratory Data Analysis (EDA)**

Key insights derived from the EDA:

1. **Gender Distribution**: Males make about 75% of purchases and exhibit a higher average spending amount than females.
   * **Average Purchase Amount by Gender**: Male customers - $9437, Female customers - $8734.
2. **Age Insights**: The 26-35 age group is the most active in terms of purchase count, while the 51-55 age group shows slightly higher average spending.
   * **Highest Spending Age Group**: Age 51-55, with an average purchase amount slightly higher than other groups.
3. **City Trends**: City Category B has the highest transaction count, but City Category C has the highest average spending per purchase.
4. **Stay Duration**: Customers residing in a city for 0–1 year make the most purchases, potentially due to new residents setting up homes.
5. **Product Category Insights**: Product categories 1, 5, and 8 are most frequently purchased, although categories with fewer transactions often have higher average purchase amounts, hinting at higher-priced items.

**Data Preprocessing**

The preprocessing steps involved:

1. **Encoding Categorical Variables**: Applied label encoding to Gender, Age, and City Category.
2. **Dropping Irrelevant Columns**: User\_ID and Product\_ID were dropped as they were not relevant for predictions.

These steps transformed the dataset into a suitable format for modeling.

**Modeling Phase**

The dataset was split into an 80:20 training-to-testing ratio, and various models were trained to predict purchase amounts:

1. **Linear Regression**: This baseline model produced an **RMSE of 4626** and an **R² score of 0.15**, indicating limited effectiveness.
2. **Decision Tree Regressor**: Showed better performance, with an **RMSE of 3362** and an **R² score of 0.55**, capturing non-linear relationships in the data.
3. **Random Forest Regressor**: Improved accuracy with an **RMSE of 3051** and an **R² score of 0.63**, reducing overfitting by averaging multiple decision trees.
4. **XGBoost Regressor**: The best-performing model, achieving an **RMSE of 2904** and an **R² score of 0.66**.

**Hyperparameter Tuning**

Grid search was used for parameter optimization:

* **Random Forest Tuning Results**: Best parameters included a maximum depth of 15, a minimum samples split of 10, and 300 estimators, resulting in an **RMSE of 2915**.
* **XGBoost Tuning Results**: Optimal parameters were a max depth of 10, learning rate of 0.1, and minimum child weight of 10, with a resulting **RMSE of 2846** and **R² score of 0.68**.

**Evaluation Metric**

The **Root Mean Square Error (RMSE)** was chosen to evaluate model performance:

* **Best Model (XGBoost) RMSE**: 2846, showing the model's strong predictive accuracy.

**Conclusion**

The XG Boost model provided the most accurate predictions of customer purchases, with an RMSE of 2846 and an R² of 0.68. This model can enable retailers to optimize their marketing and inventory strategies by forecasting purchase amounts based on customer demographics and product preferences, improving profitability and customer targeting.