E-commerce Furniture Dataset Project Report

This project analyzes the **E-commerce Furniture Dataset 2024**, containing 2,000 entries of furniture items scraped from AliExpress. The dataset includes product details such as title, original price, discounted price, number of units sold, and additional tags (e.g., free shipping). The main objective is to **predict the number of furniture items sold** based on product attributes and understand the relationship between price, discounts, and sales performance.

Dataset Columns:
- productTitle: Name of the furniture item
- originalPrice: Price before discount
- price: Final selling price
- sold: Number of units sold
- tagText: Shipping/offer tags

Steps Followed in the Project: Data Collection – Dataset loaded from CSV. **Data Preprocessing** – Missing values handled, price columns cleaned, discount percentage calculated, and categorical encoding applied. **Exploratory Data Analysis (EDA)** – Visualized distributions of price and sold, scatterplots, and shipping tag frequencies. **Feature Engineering** – Created new discount percentage feature and label-encoded categorical features. **Model Training** – Trained Linear Regression and Random Forest models to predict sales. **Model Evaluation** – Compared models using Mean Squared Error (MSE) and R^2 Score.

Model Evaluation Results:
- **Linear Regression**: Provides a baseline, works well if relationships are linear.
- **Random Forest Regressor**: More robust, handles non-linear patterns and interactions better.

Based on the evaluation metrics, **Random Forest outperformed Linear Regression** in predicting the number of items sold.

Conclusion:
Discount percentage and shipping tags play a significant role in influencing sales. Random Forest proved to be the better predictive model for this dataset. This project demonstrates how data cleaning, feature engineering, and machine learning can be applied to e-commerce data for actionable business insights.

Acknowledgement:
Dataset sourced from AliExpress via Apify, with proper ethical standards maintained during collection.