

Celebal Summer Internship 2025



ABHISHEK MOHAN
IIT MANDI
CT CSI DS 2176

Customer Lifetime Value Prediction

Goal :

Predicting the future value of a customer to a business over the entire duration of their relationship, incorporating factors such as past purchase history, frequency of purchases, and customer demographics.

Dataset :

<https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset>

Benefits of CLTV Prediction

Better Customer Segmentation:

- Identify high-, medium-, and low-value customers to tailor marketing strategies accordingly.

Improved Revenue Forecasting:

- Accurately estimate future revenue streams based on predicted customer value.

Increased ROI on Marketing Spend:

- Focus promotional budgets on high-value customers who generate the most profit.

Enhanced Customer Retention:

- Take proactive actions to retain valuable customers and reduce churn.

Data-Driven Decision Making:

- Empower sales, marketing, and strategy teams to make informed decisions backed by predictive analytics.

About the Dataset

- This online Retail data set contains all transactions occurring for a UK-based, registered, non-store online retail business between 01/12/2009 and 09/12/2011.
- **Key Attributes:** InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country.
- **Shape:** 541910 Rows and 8 Columns.
- **Missing Values:** 1454 in Description and 135080 in CustomerID column

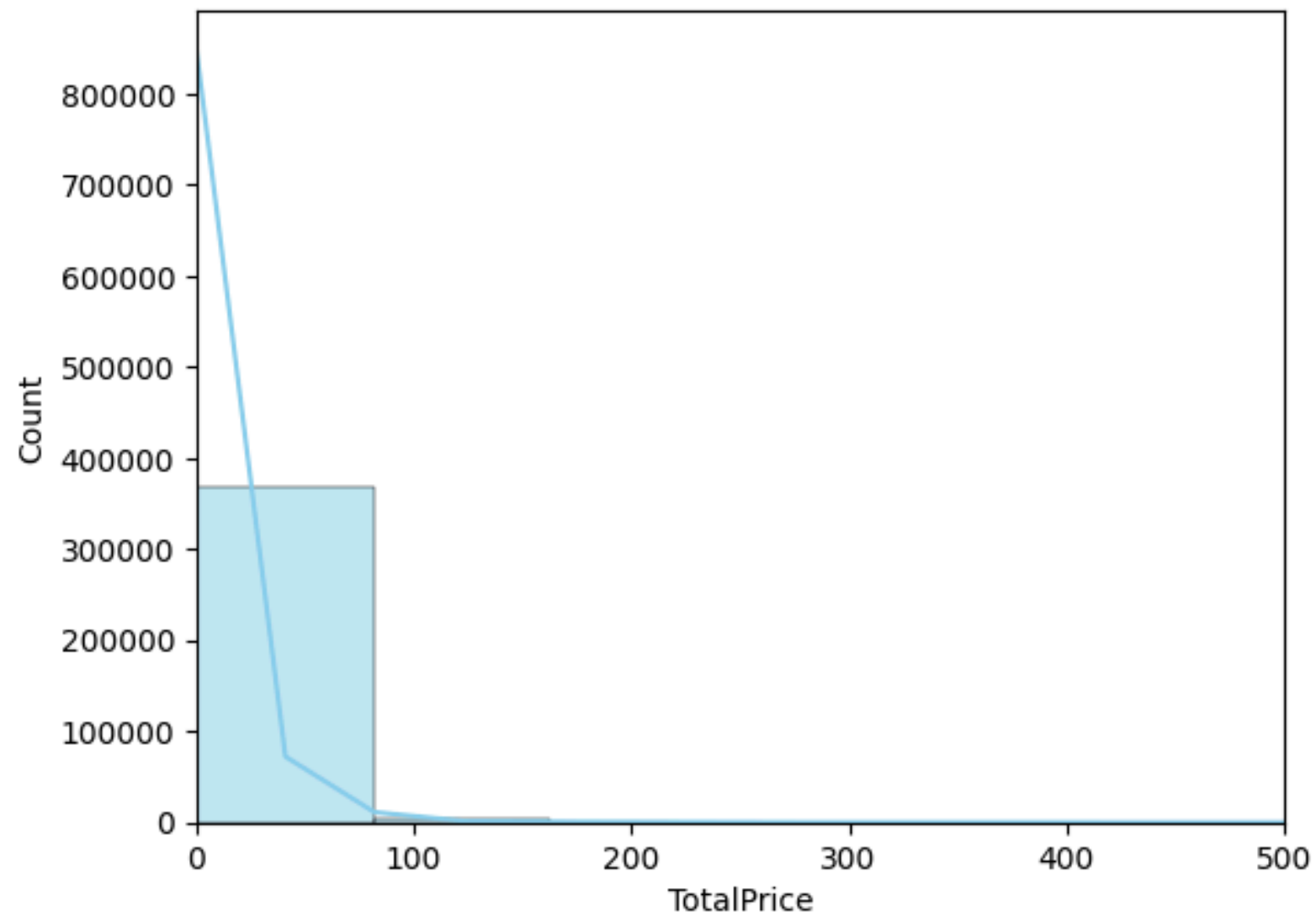
Data Preprocessing

- Removed missing values from the dataset.
- Removed Cancelled transactions (Invoice starting with 'C').
- Removed negative quantities or prices.
- Removed records without Customer ID.
- Removed Outliers from the dataset.
- Created new column, $\text{TotalPrice} = \text{Quantity} \times \text{UnitPrice}$.

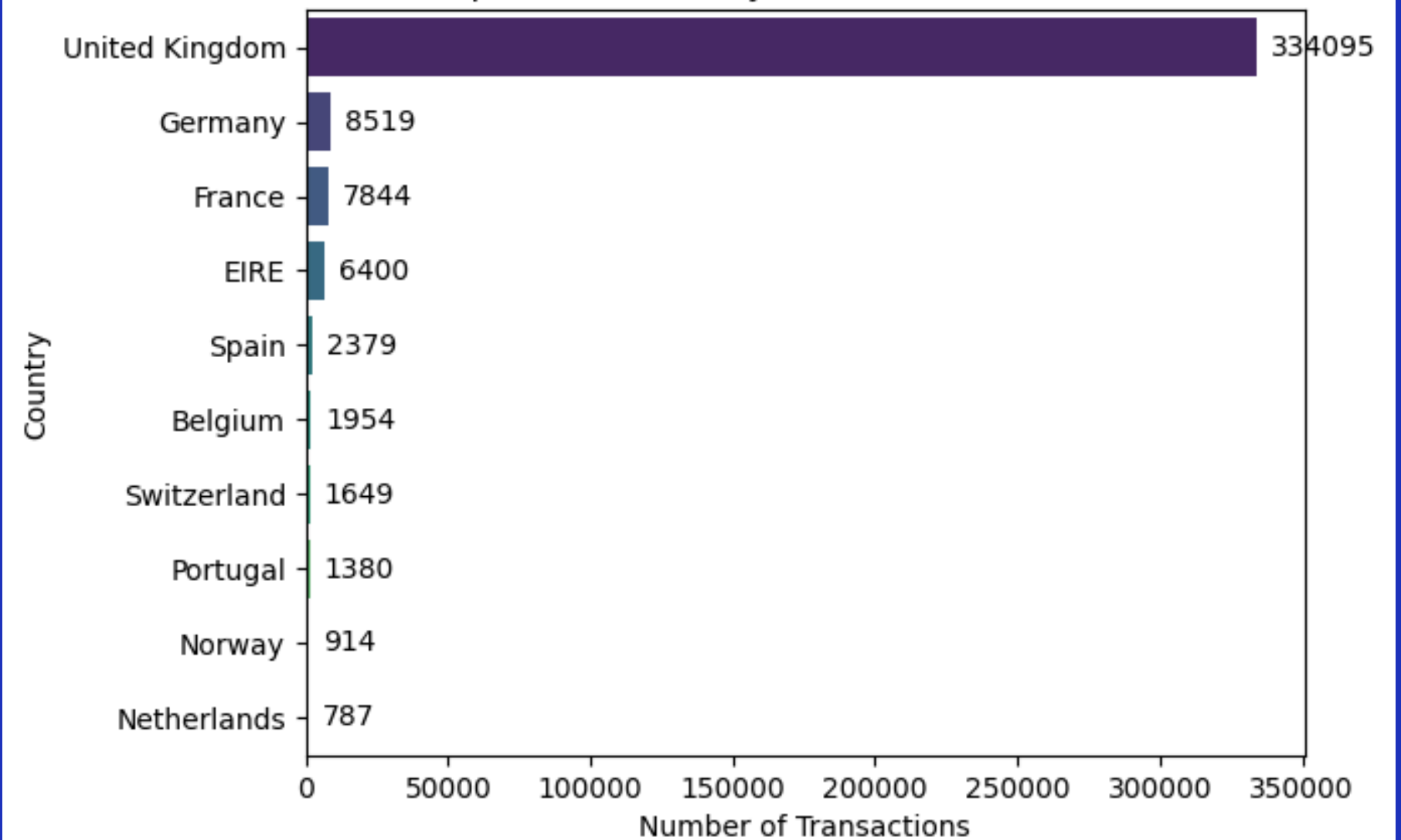


Data Visualization

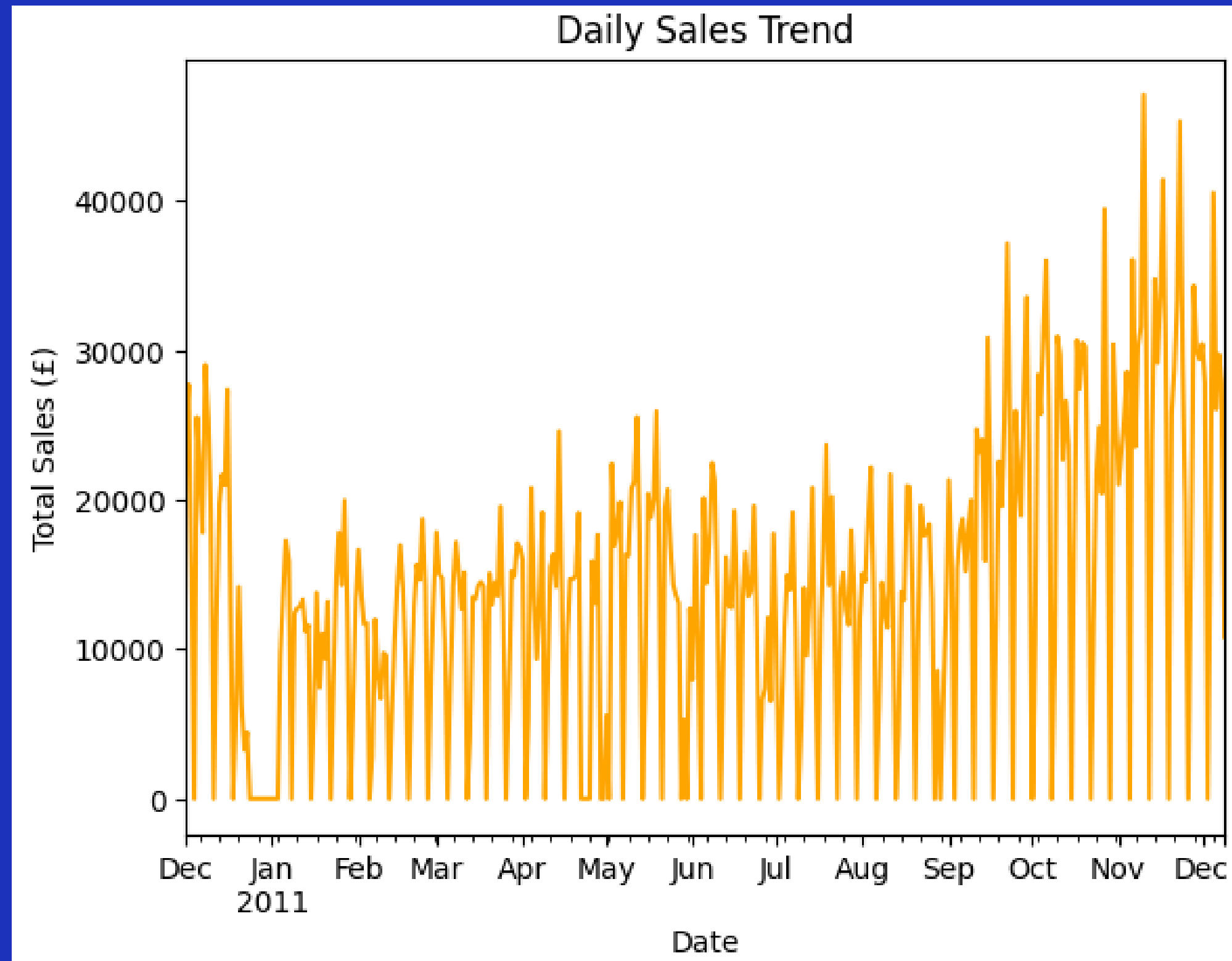
Distribution of Total Transaction Value (TotalPrice)



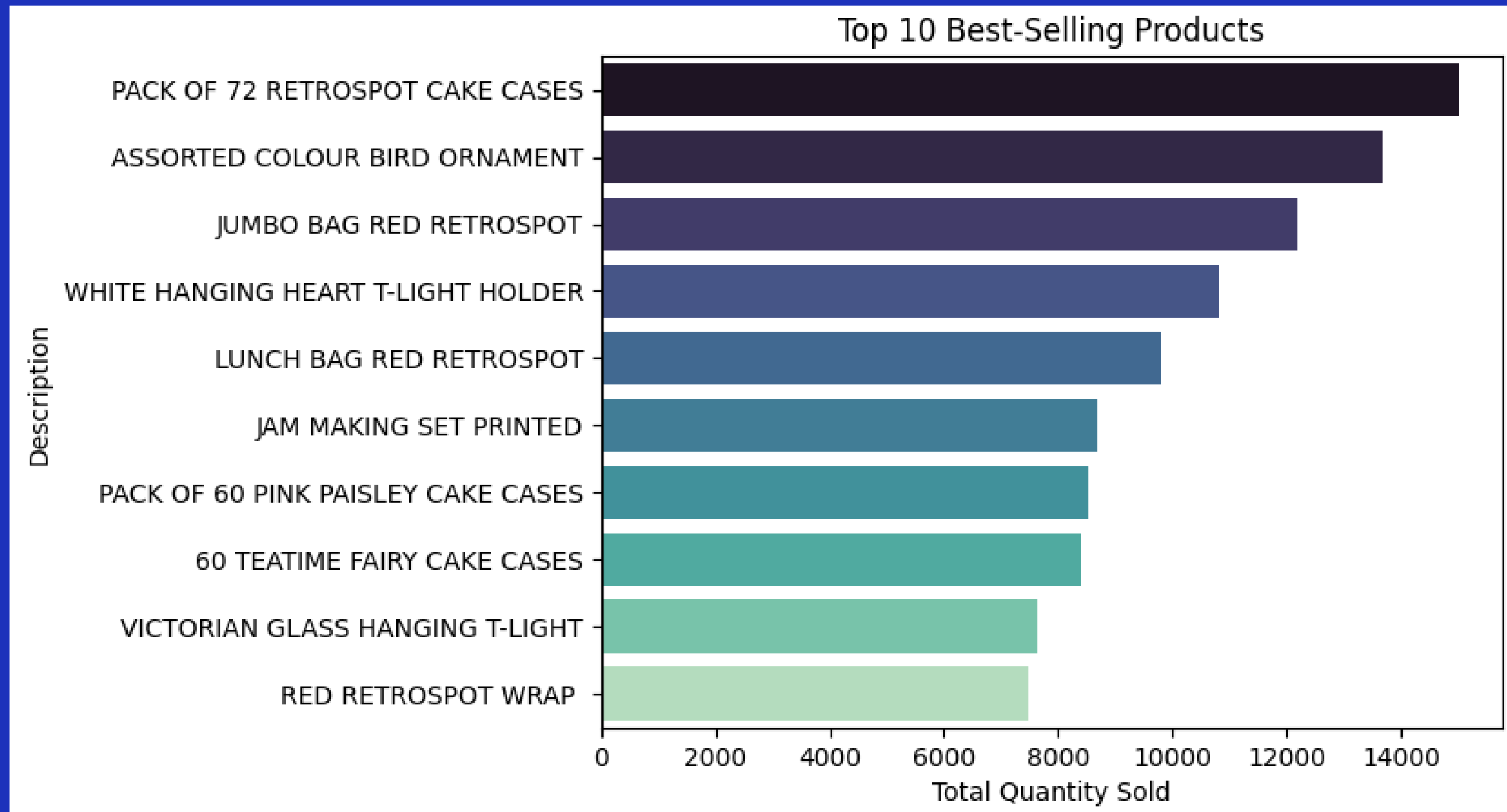
Top 10 Countries by Number of Transactions



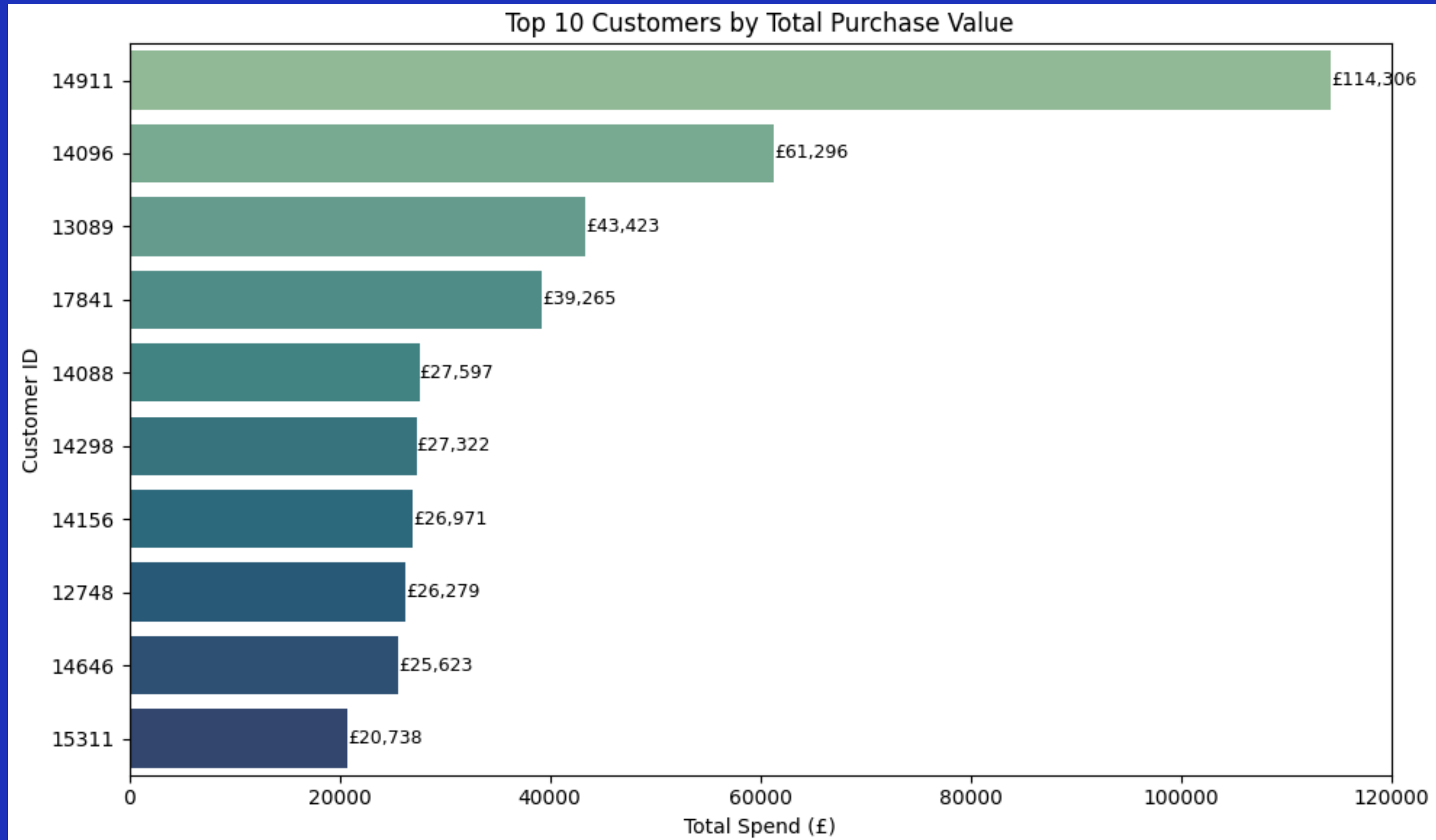
Data Visualization



Data Visualization



Data Visualization



Feature Engineering

Features Extracted:

- **Recency:** Days since last purchase
- **Frequency:** Number of purchases
- **Monetary:** Total amount spent
- **Customer Age:** Days between first and last purchase
- **Average Basket Size :** $(\text{Total Quantity} / \text{Frequency})$
- **Average Unit Price :** Mean unit price paid by the customer.
- **Average Days Between Purchases :-** Mean time gap between consecutive purchases.
- Snapshot taken at the 75% quantile of the Date range.



Target Variable Definition

How CLTV was defined:

Dataset split into:

- Observation Period (75%) – to generate features
- Prediction Period (25%) – to calculate the actual future CLTV

Target = Total purchase value in the prediction window



Model Building

Algorithms Tried:

- Linear Regression
- Random Forest
- XGB Regressor

Preprocessing:

- Log-transformation of target to reduce skew
- Feature scaling using StandardScaler

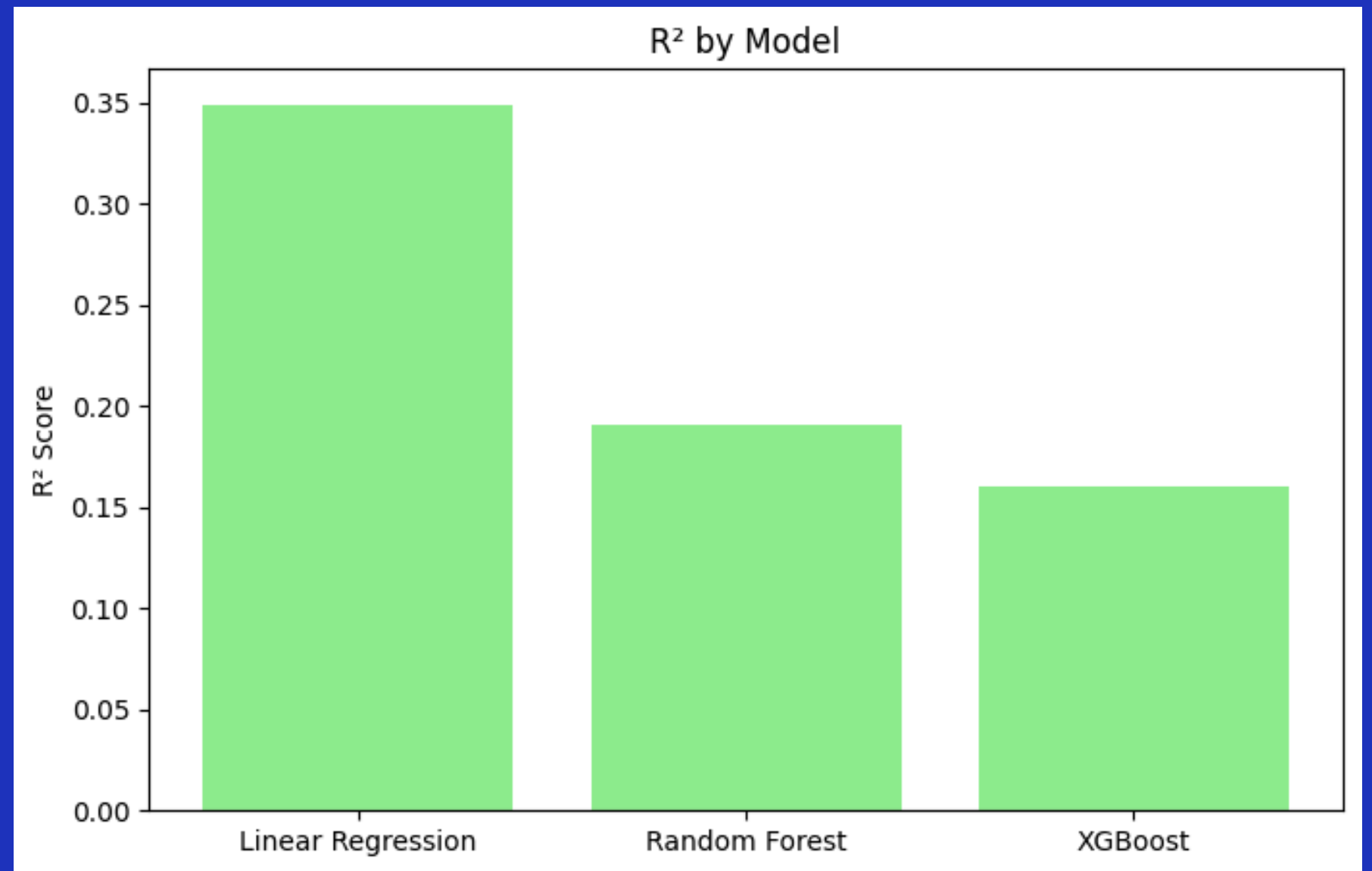
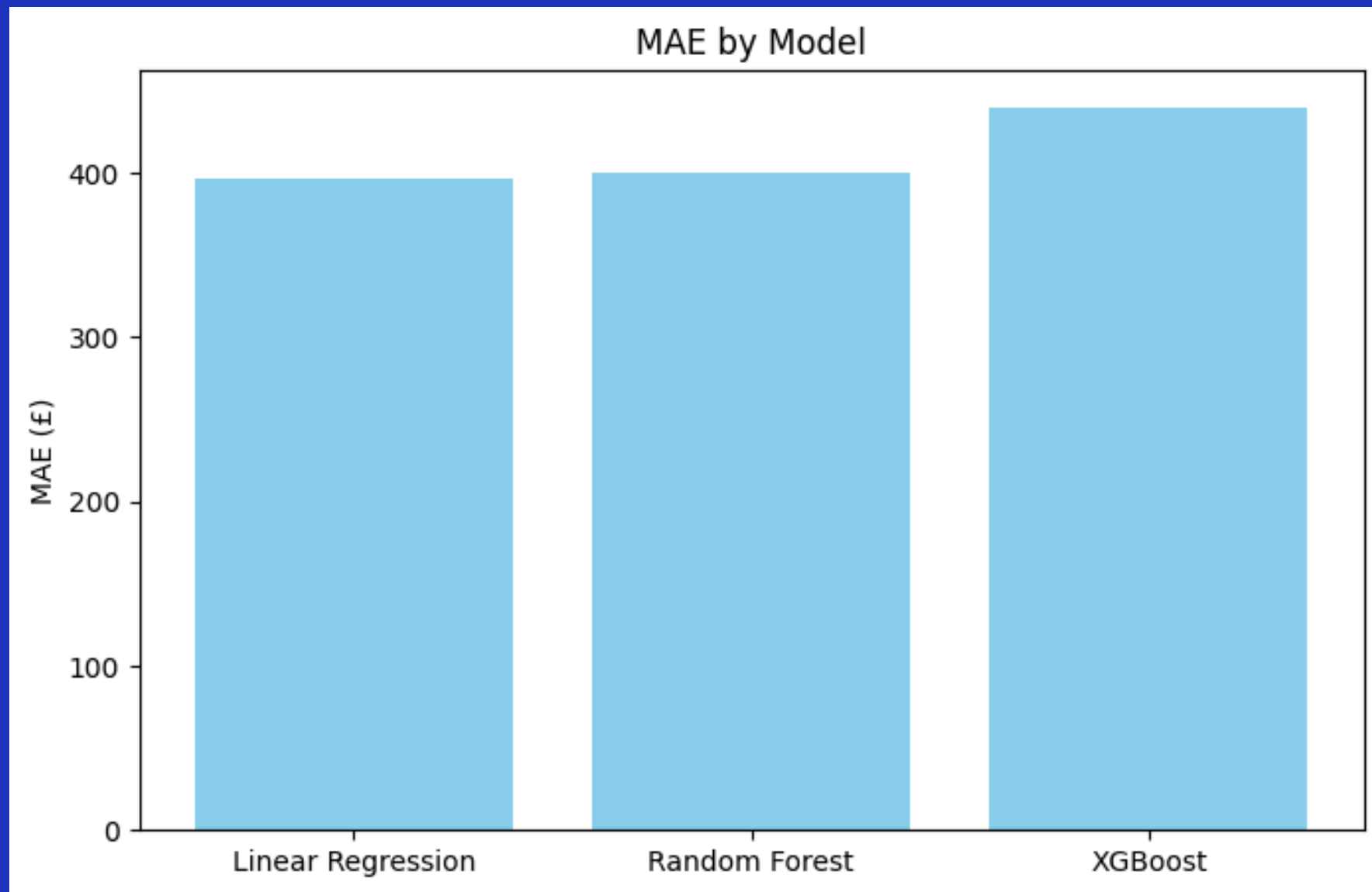


Model Performance

Model	MAE	RMSE	R2 Score
Linear Regression	£396.47	£1909.60	0.3493
Random Forest	£399.55	£2129.90	0.1905
XGB Regressor	£440.16	£2169.29	0.1602



Model Performance





Model Deployment

Key Points:

- **Objective:** Enable real-time CLTV predictions via a simple web interface.
- **Tools:** Streamlit, cloudpickle, Streamlit Cloud.
- **Functionality:** Takes customer features → predicts CLTV → segments customers.
- **Impact:** Helps teams focus efforts on high-value customers, improving ROI.

Application Link: - <https://cltv-predictor.streamlit.app/>

CLTV Predictor App Results

Customer Lifetime Value (CLTV) Predictor

Use this tool to predict the future value of a customer and segment them into tiers.

Recency (days since last purchase)

30 - +

Customer Age in dataset (days)

200 - +

Frequency (number of purchases)

5 - +

Total Quantity Purchased

20 - +

Total Monetary (£)

500 - +

Average Unit Price (£)

25 - +

Average Basket Size

4 - +

Average Days Between Purchases

40 - +

Country

Spain ▾

Predict CLTV

✓ Predicted CLTV: £104.88

🇬🇧 Customer Segment: Low

Customer Lifetime Value (CLTV) Predictor

Use this tool to predict the future value of a customer and segment them into tiers.

Recency (days since last purchase)

15 - +

Customer Age in dataset (days)

300 - +

Frequency (number of purchases)

10 - +

Total Quantity Purchased

20 - +

Total Monetary (£)

50000 - +

Average Unit Price (£)

400 - +

Average Basket Size

10 - +

Average Days Between Purchases

20 - +

Country

United Kingdom ▾

Predict CLTV

✓ Predicted CLTV: £187971249.24

🇬🇧 Customer Segment: High



CLTV Segmentation

Segment	Count	Mean	Sum
Low	403	217.04	87469.24
Mid	403	298.61	120342.55
High	402	403.19	162083.67
Very High	403	1508.91	608092.97

Business Recommendations

Based on CLTV Segments:

-> Very High & High CLTV Customers

- **Retention Focus:** Offer loyalty programs, exclusive deals
- **Upsell:** Personalized product bundles, early access to new products
- **Communication:** VIP-level support, personal recommendations

-> Mid CLTV Customers

- **Nurture Strategy:** Offer time-bound discounts or cart abandonment reminders
- **Incentives:** Encourage higher order frequency or bigger baskets

-> Low CLTV Customers

- **Cost Control:** Limit marketing spend on this group
- **Win-back Campaigns:** Only if the acquisition cost is low
- **Feedback Collection:** Understand why they don't spend more





Thank you!

