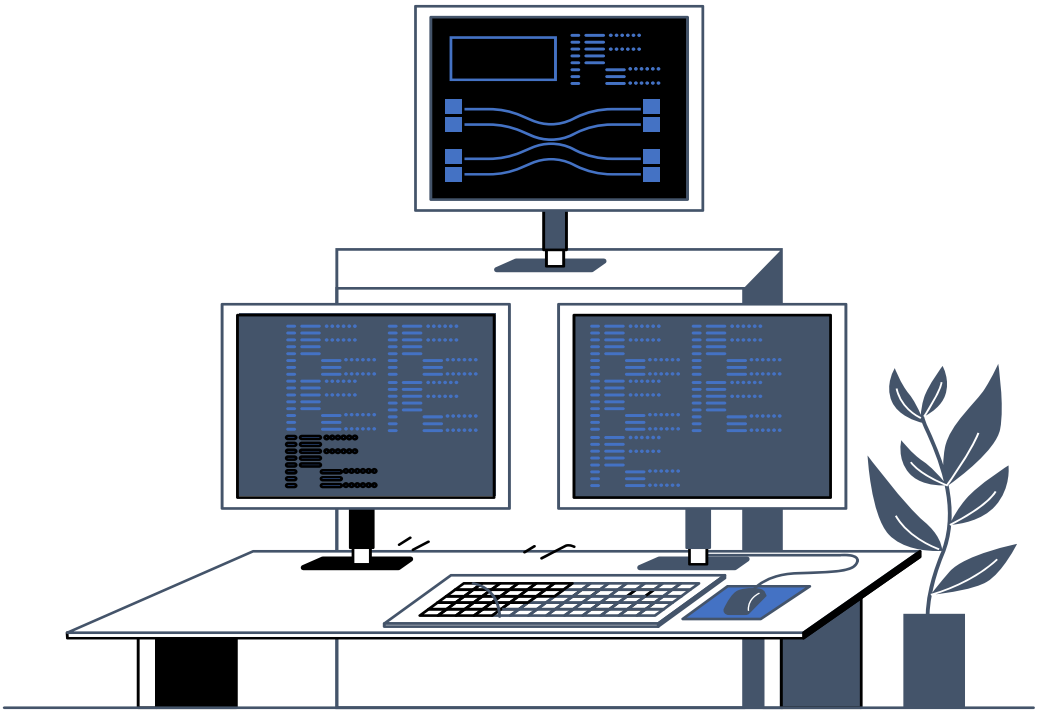
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Customer Lifetime Value Model



Name: Haya Alowias

Email: haya.alowias@gmail.com

Abstract:

Every organization competes to stay on the top in the market, so organizations need to consider all the factors that will result in their long-term success. One of the most crucial factors among all is to provide the best customer experience. Customer Lifetime Value is an important factor that helps in understanding customers.

Customer lifetime value (CLV) represents the total amount of money a customer is expected to spend in your business, or on your products, during their lifetime. This is an important figure to know because it helps you make decisions about how much money to invest in acquiring new customers and retaining existing ones. The cost of building new relationships(customers) is expensive compared to the cost of retaining customers. So, whenever a business focuses on retaining their old customers, they increase Customer Lifetime Value (CLV). That means the higher the CLV, the higher the loyalty of customers towards a particular business.

So, the goal of this project was to use regression models to predict the CLTV model and analysis the data that answer the following questions:

- What is the top 5 country Transaction?
- What is the number of transactions for all countries?
- What is the revenue by country?
- Average Price by Country?
- Transactions by Month?
- Transactions by Year?
- Predicted Purchases for one month for each customer?
- which product has been purchased more?
- What is CLV for each customer?
- Predict CLTV using Linear Regression Model

DATASET:

For the dataset used in this paper, I have chosen an open-source online retail dataset that contains all the transactions occurring between 01/12/2010 and 09/12/2011 for many countries that registered non-store online retail. the companies mainly sell unique all-occasion giftware.

Features	Datatype	Description
InvoiceNo	Nominal	6-digit integral number uniquely assigned to each transaction. If this code starts with the letter 'c', it indicates a cancellation.
StockCode	Nominal	Product (item) code .A 5-digit integral number uniquely assigned to each distinct product.
Description	Nominal	Product (item) name.
Quantity	Numeric	The quantities of each product (item) per transaction
InvoiceDate	Numeric	Invice date and time. The day and time when a transaction was generated.
UnitPrice	Numeric	Product price per unit in sterling (Â£).
CustomerID	Nominal	Customer number. A 5-digit integral number uniquely assigned to each customer.
Country	Nominal	Country name. The name of the country where a customer resides

Steps involved in this project:

Data Importing | Exploratory Data Analysis | Feature Extraction | Predictive Models Building | Model Evaluation.

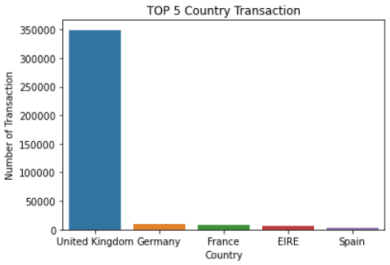
Libraries Used:

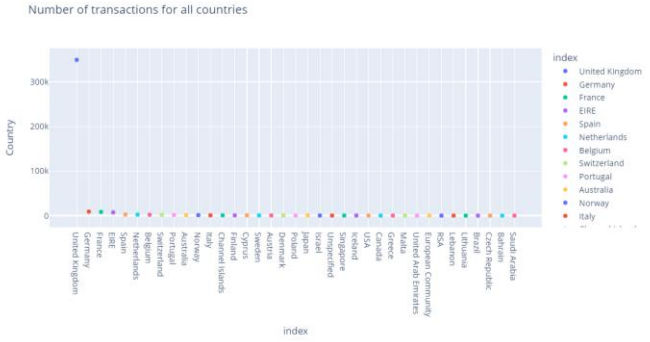
- Scikit Learn
- Lifetimes
- Matplotlib, Seaborn
- Pandas, Numpy

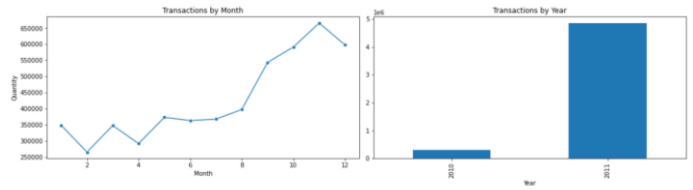
Algorithms Used:

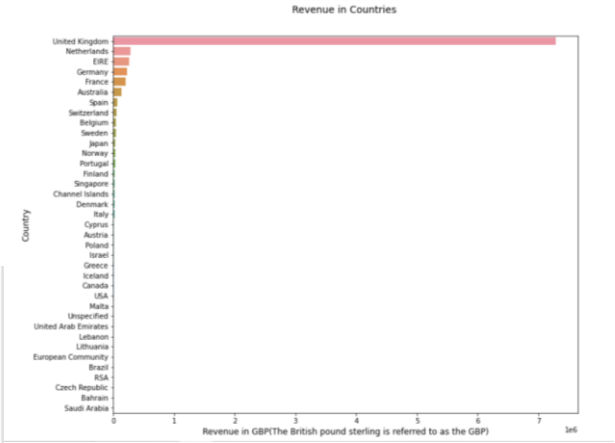
- Linear Regression

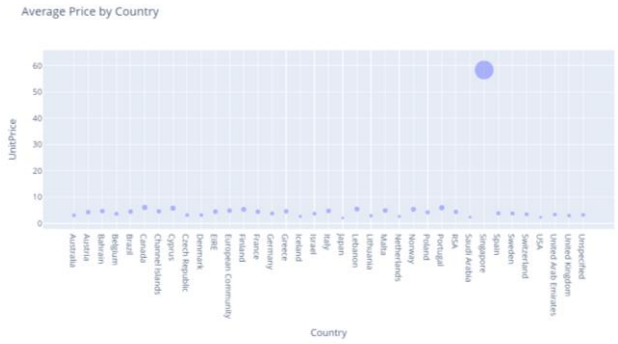
Result:

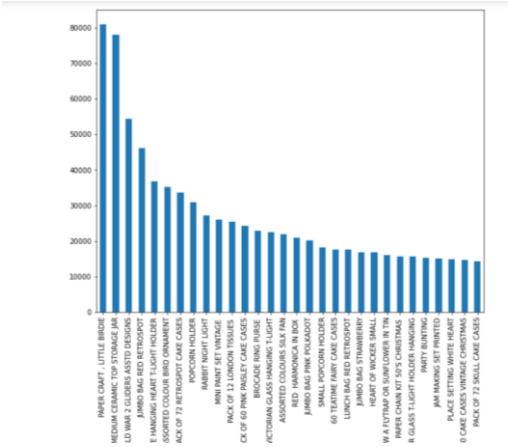


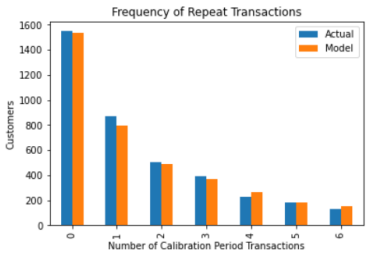












3]:

CustomerID	frequency	recency	T	monetary_value	predicted_purchases	actual	error
12346.0	0.0	0.0	325.0	0.000000	0.063003	0.000000	0.063003
12347.0	6.0	365.0	367.0	599.701667	0.469673	0.493151	0.023478
12348.0	3.0	283.0	358.0	301.480000	0.268852	0.318021	0.049169
12349.0	0.0	0.0	18.0	0.000000	0.285036	0.000000	0.285036
12280.0	0.0	0.0	310.0	0.000000	0.065496	0.000000	0.065496

:

CustomerID	CLV
18280.0	9.019848e+05
18281.0	2.580497e+05
18282.0	7.305777e+05
18283.0	1.604871e+06
18287.0	1.333572e+07

In [43]:
```
# Split the data To understand model performance, dividing the dataset into a training set and a test set
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.8,test_size=0.2)

LR = LinearRegression()


# fit the model to the training data (learn the coefficients)
LR.fit(X_train, y_train)

# make predictions on the testing set
y_pred = LR.predict(X_test)

y_pred_score = LR.score(X_test, y_test)

print("Prediction score is: {}".format(y_pred_score))
```

```
Prediction score is: 0.838339523764265
```