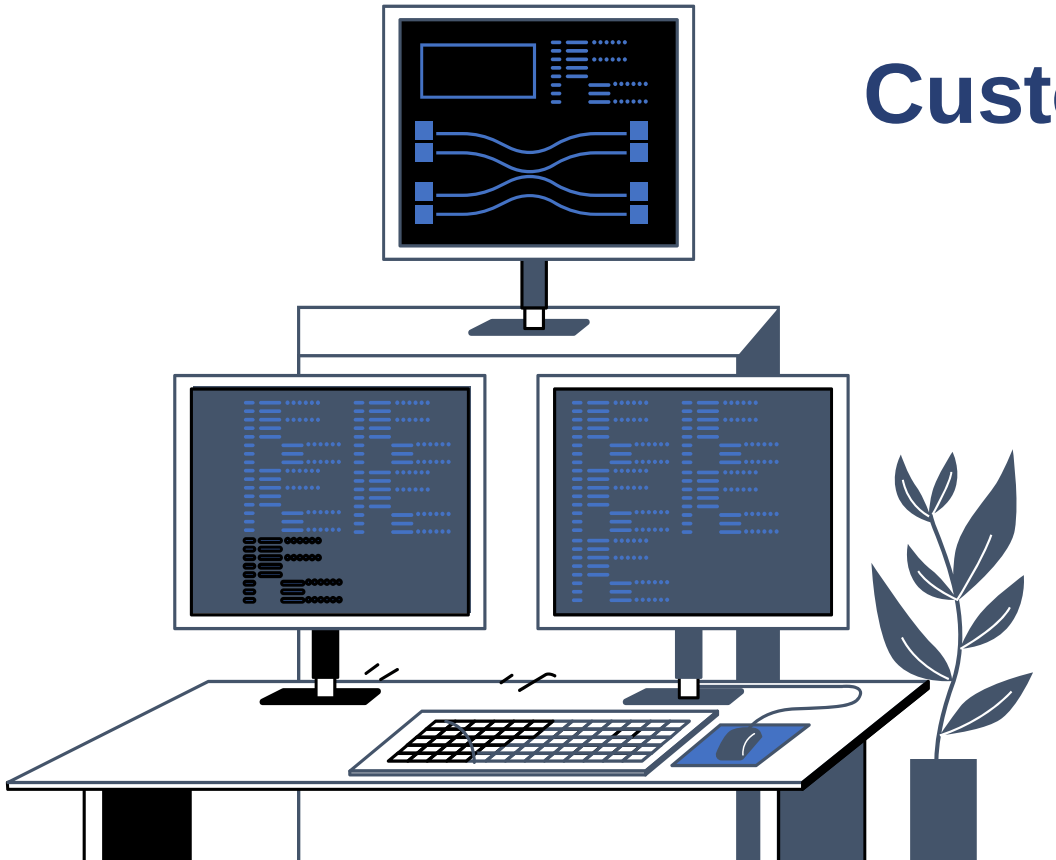
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority



Customer Live Times Value Model

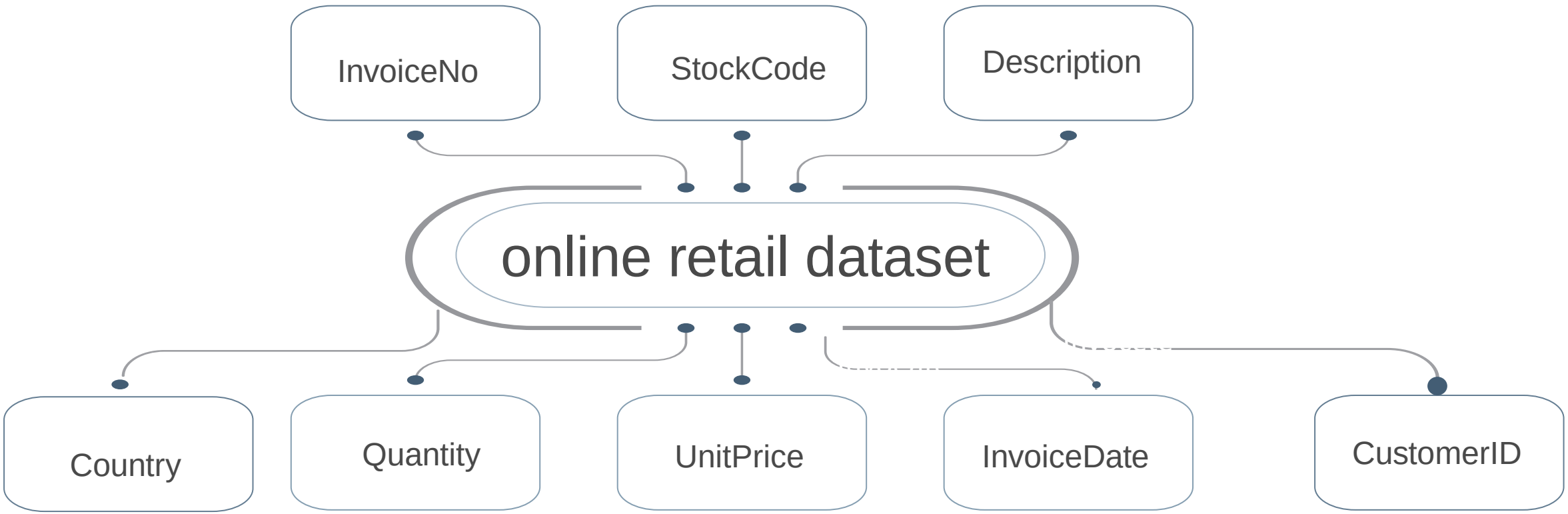
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

What is customer lifetime value (CLV) ?

Data Description

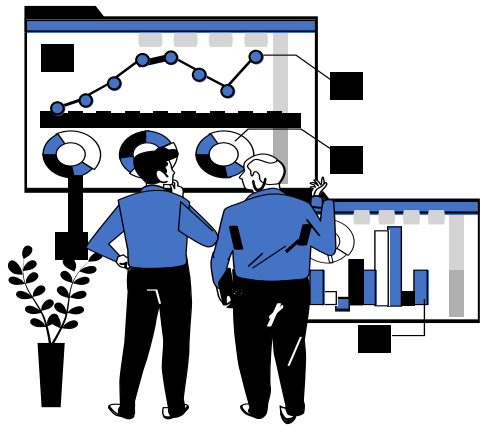
Number of Features: 8

Number of Instance: 541909



SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

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

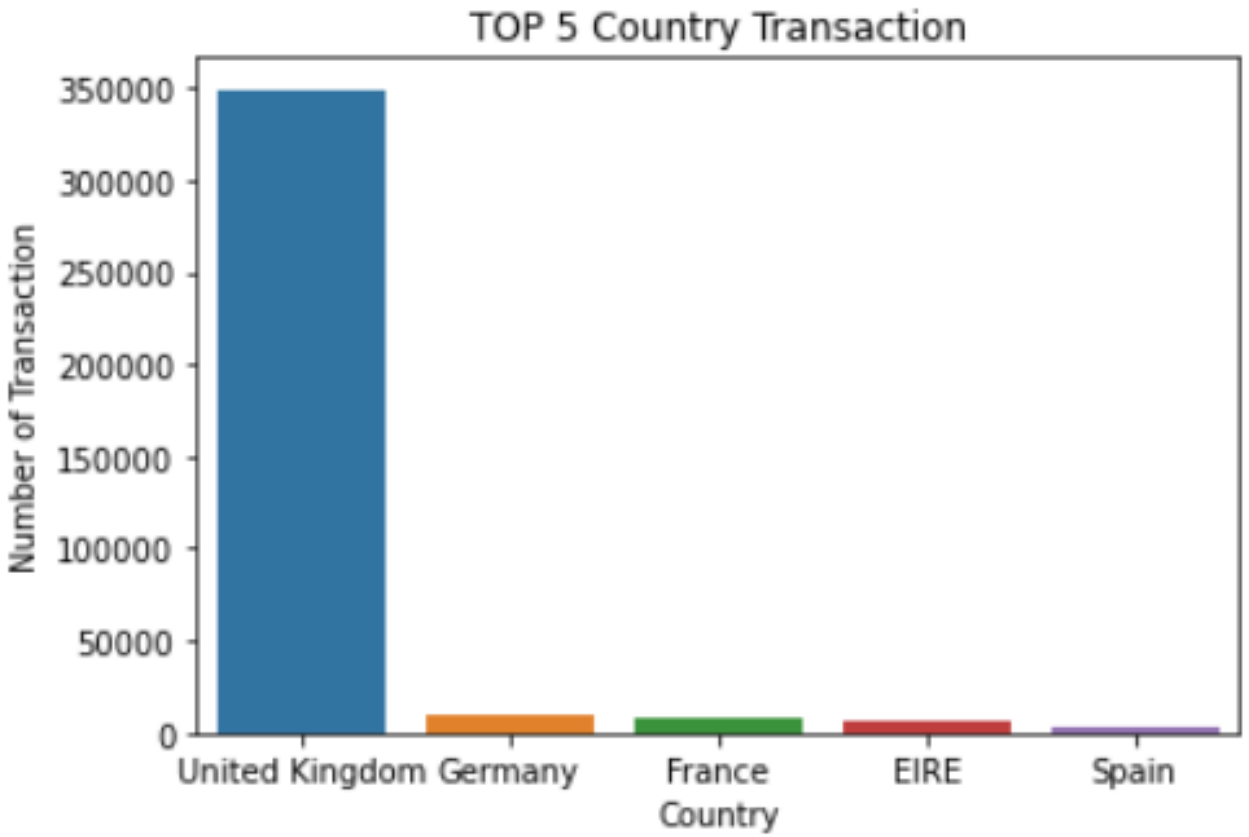


SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

- Top 5 country transaction
- Average Price by Country
- Number of transactions for all countries
- CLV for each customer
- Transactions by Month and Year
- Products purchased more
- Predict Purchases for 1 month for each customer
- Predict CLTV using Linear Regression Model

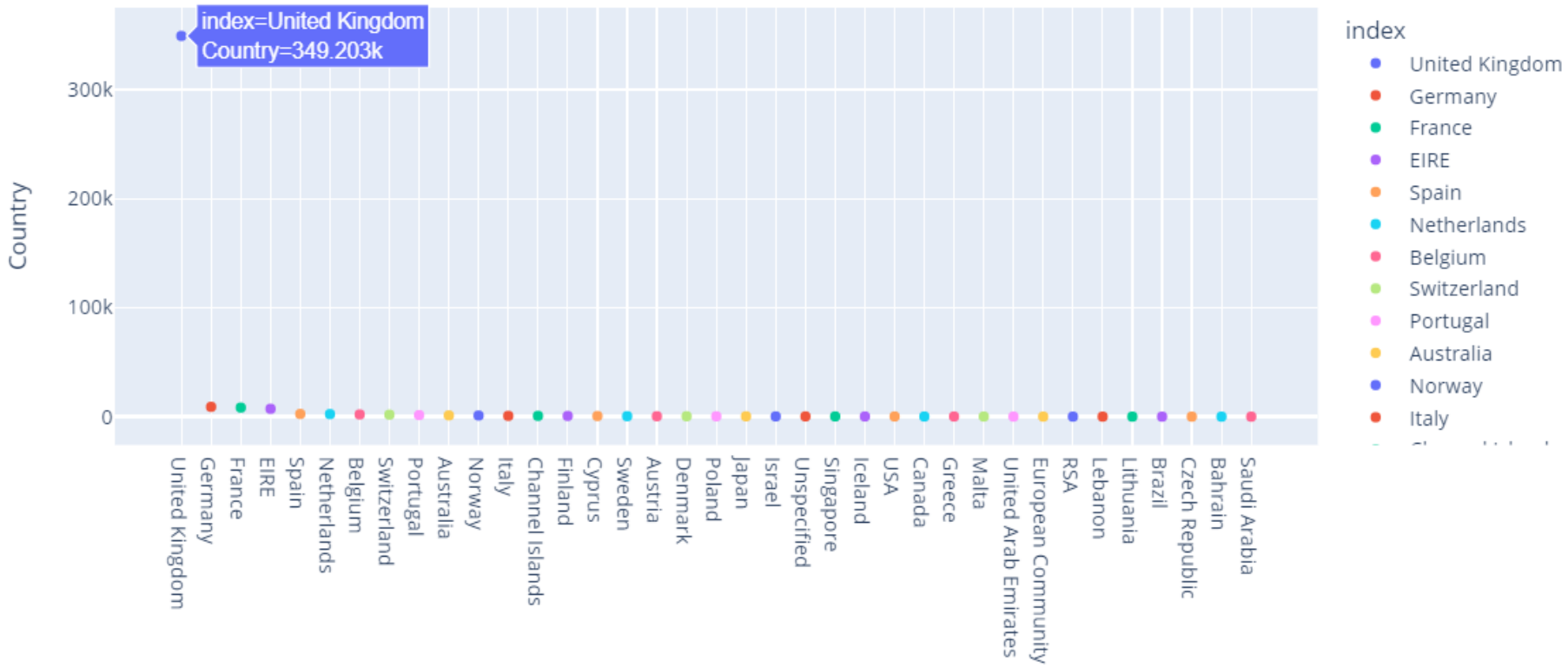
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Top 5 country transaction



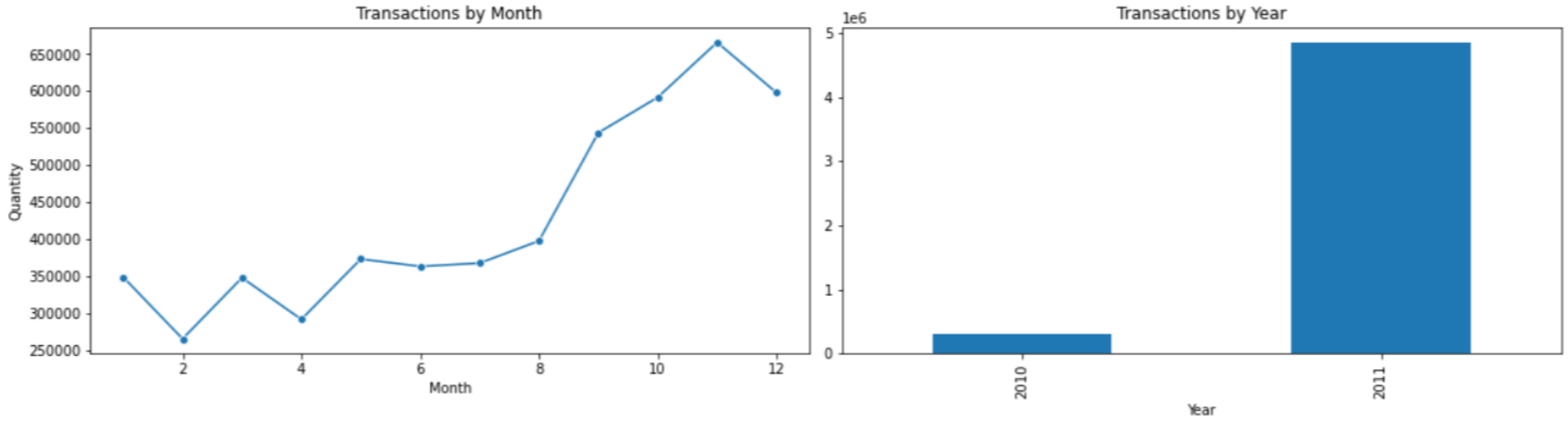
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Number of transactions for all countries



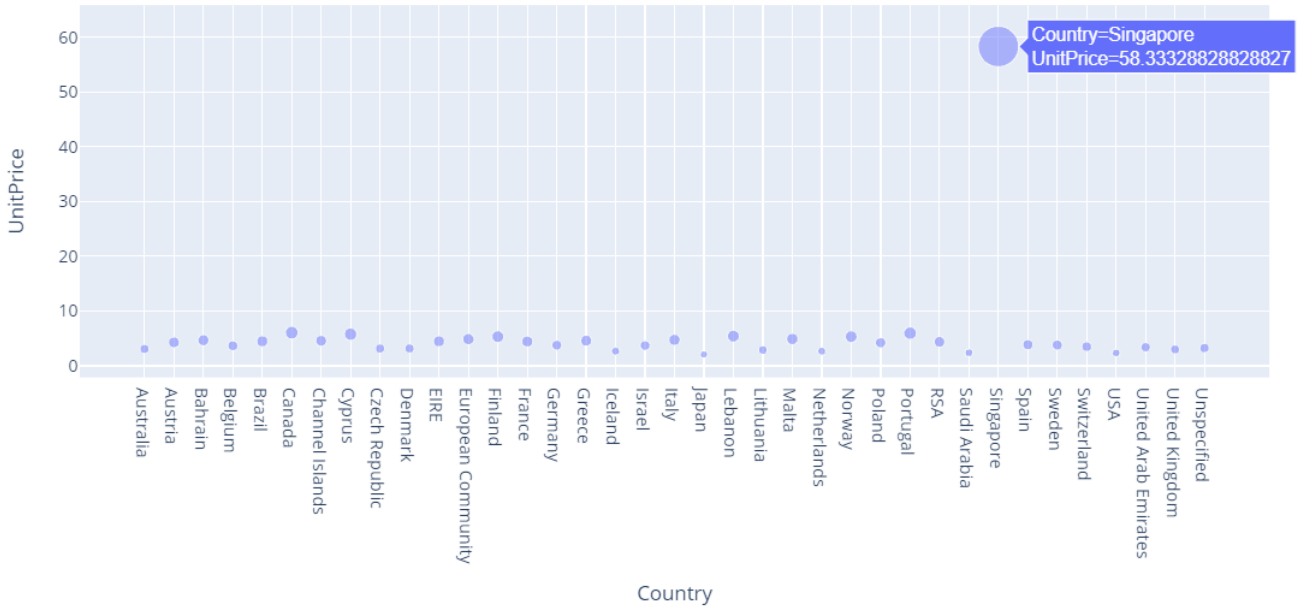
SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

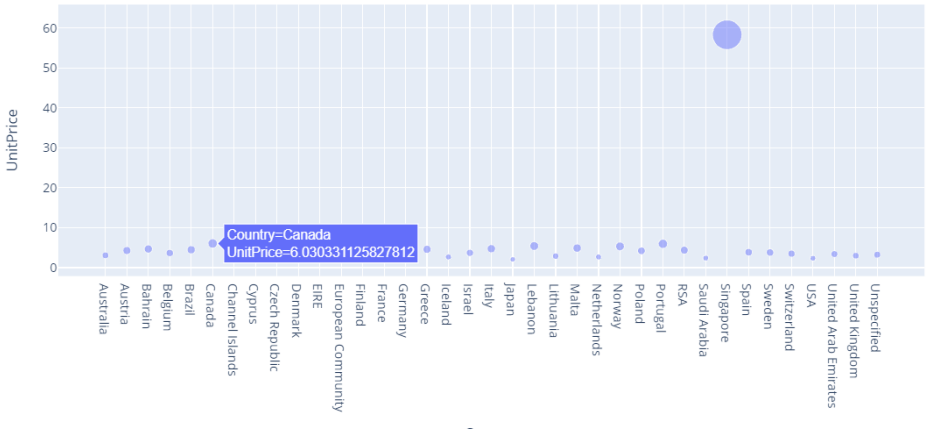
Transactions by Month and year



SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

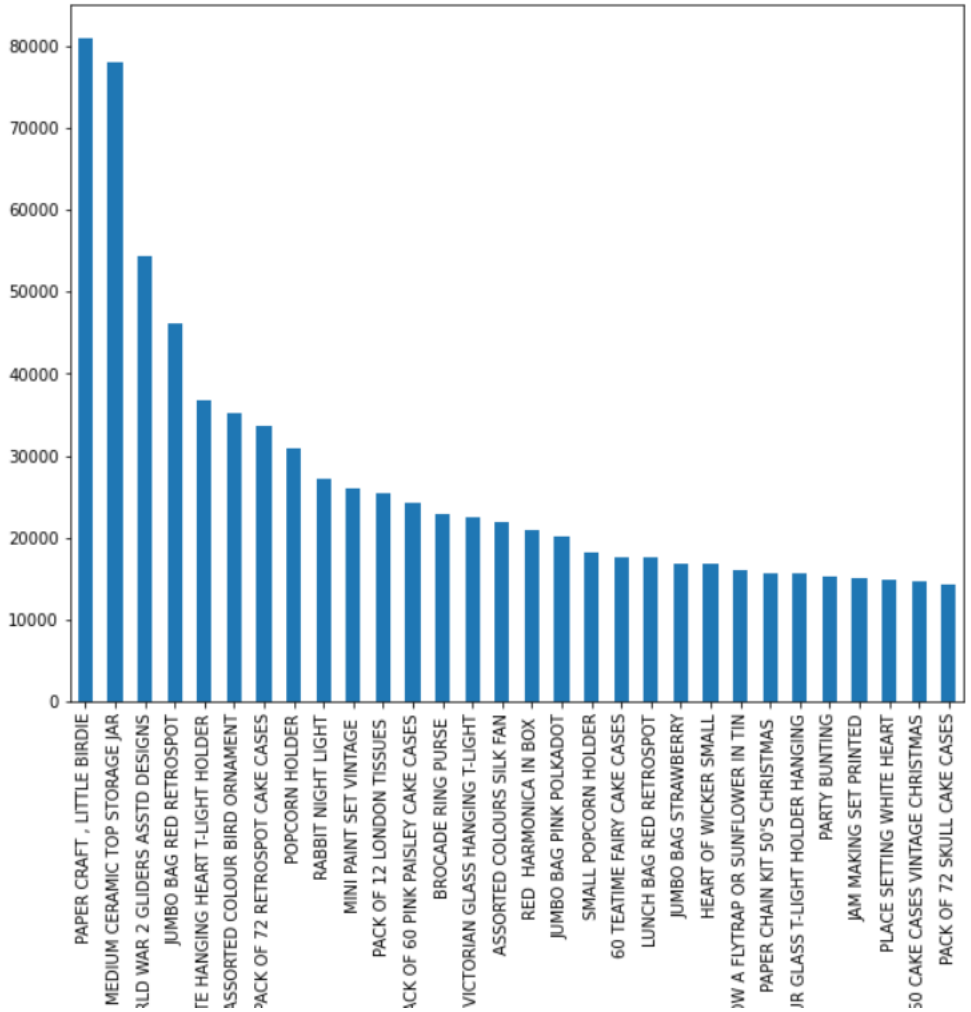
Average price by country





SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Products purchased more



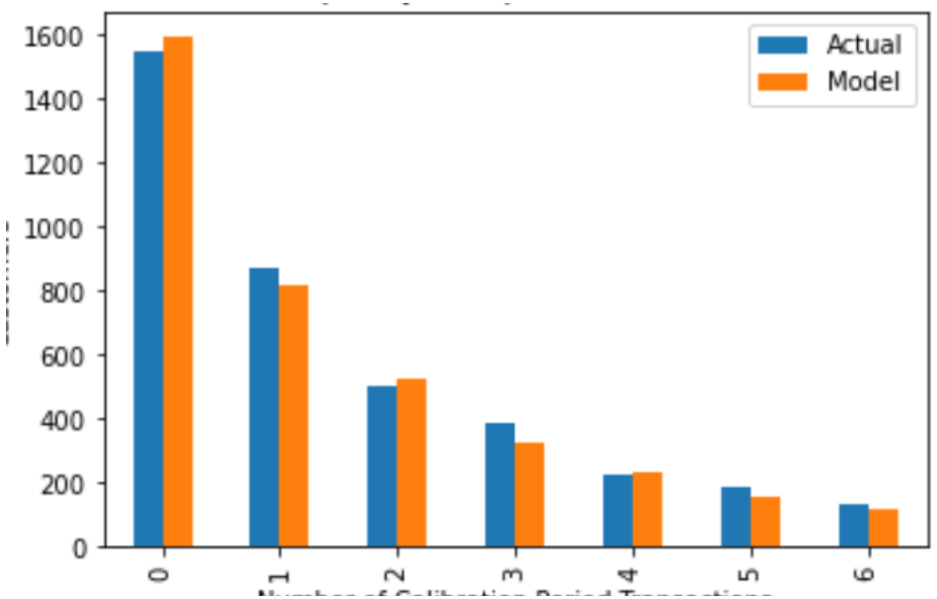
Predict Purchases for 1 month for each customer

```
result = rfm_summary.copy()

#Create instance
pareto_model = lifetimes.ParetoNBDFitter(penalizer_coef = 0.1)
pareto_model.fit(rfm_summary["frequency"],rfm_summary["recency"], rfm_summary["T"])


#Calculate the expected number of repeat purchases up to time t.


#t (array_like) - times to calculate the expectation for.
t = 30
result["predicted_purchases"] = pareto_model.conditional_expected_number_of_purchases_up_to_time(t,result["frequency"], result["
```

```
result["actual"] = (result["frequency"]/result["recency"])*30
result["actual"].fillna(0, inplace = True)
```

```
result["error"] =abs(result["actual"]-result["predicted_purchases"])

result.head()
```

CustomerID	frequency	recency	T	monetary_value	predicted_purchases	actual	error
12346.0	0.0	0.0	325.0	0.000000	0.063046	0.000000	0.063046
12347.0	6.0	365.0	367.0	599.701667	0.469673	0.493151	0.023477
12348.0	3.0	283.0	358.0	301.480000	0.268872	0.318021	0.049149
12349.0	0.0	0.0	18.0	0.000000	0.285125	0.000000	0.285125
12350.0	0.0	0.0	310.0	0.000000	0.065540	0.000000	0.065540



Customer Live Times Value

```
# Calculate Purchase Frequency
#Purchase Frequency: Purchase Frequency is the ratio of the total number of orders and the total number of customer. It represent
purchase_frequency = sum(clv_grp['InvoiceNo'])/clv_grp.shape[0]

# Calculate Repeat and Churn Rate
#Repeat Rate: Repeat rate can be defined as the ratio of the number of customers with more than one order to the number of unique
repeat_rate = clv_grp[clv_grp['InvoiceNo'] > 1].shape[0]/clv_grp.shape[0]
#Churn Rate: Percentage of customers who have not ordered again.
churn_rate = 1 - repeat_rate
```

]:
```
# Calculate Profit margin assuming gain of 5%
#Profit margin is the commonly used profitability ratio. It represents how much percentage of total sales has earned as the gain.
clv_grp['profit_margin'] = clv_grp['invoice_value'] * 0.05

# Calculate Customer Lifetime Value
clv_grp['CLV'] = ((clv_grp['AvgOdeVal'] * purchase_frequency)/churn_rate) * clv_grp['profit_margin']

CLV_F=clv_grp['CLV']
CLV_F.to_frame().tail()
#CLV_F.sort_index()
```

]:

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

Predict CLTV using Linear Regression Model

month_yr	CustomerID	Apr-2011	Aug-2011	Dec-2010	Dec-2011	Feb-2011	Jan-2011	Jul-2011	Jun-2011	Mar-2011	May-2011	Nov-2011	Oct-2011	Sep-2011	CLV
4333	18280.0	0.0	0.00	0.0	0.00	0.00	0.00	0.00	0.00	180.6	0.00	0.00	0.00	0.0	361.20
4334	18281.0	0.0	0.00	0.0	0.00	0.00	0.00	0.00	80.82	0.0	0.00	0.00	0.00	0.0	161.64
4335	18282.0	0.0	100.21	0.0	77.84	0.00	0.00	0.00	0.00	0.0	0.00	0.00	0.00	0.0	356.10
4336	18283.0	115.6	0.00	0.0	208.00	100.95	213.75	139.89	296.52	0.0	85.22	637.71	112.99	134.9	3859.86
4337	18287.0	0.0	0.00	0.0	0.00	0.00	0.00	0.00	0.00	0.0	765.28	0.00	1072.00	0.0	3674.56

```
# now i will divide the given columns into two types: 1-target variable 2-feature variables

# 1- latest 6 month as feature variables .

X = sale[['Dec-2011', 'Nov-2011', 'Oct-2011', 'Sep-2011', 'Aug-2011', 'Jul-2011']]

#2--target variable

y = sale[['CLV']]
```

```
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.8,test_size=0.2)

LR = LinearRegression()

LR.fit(X_train, y_train)

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: 0.9597429161719538
```

SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Thanks