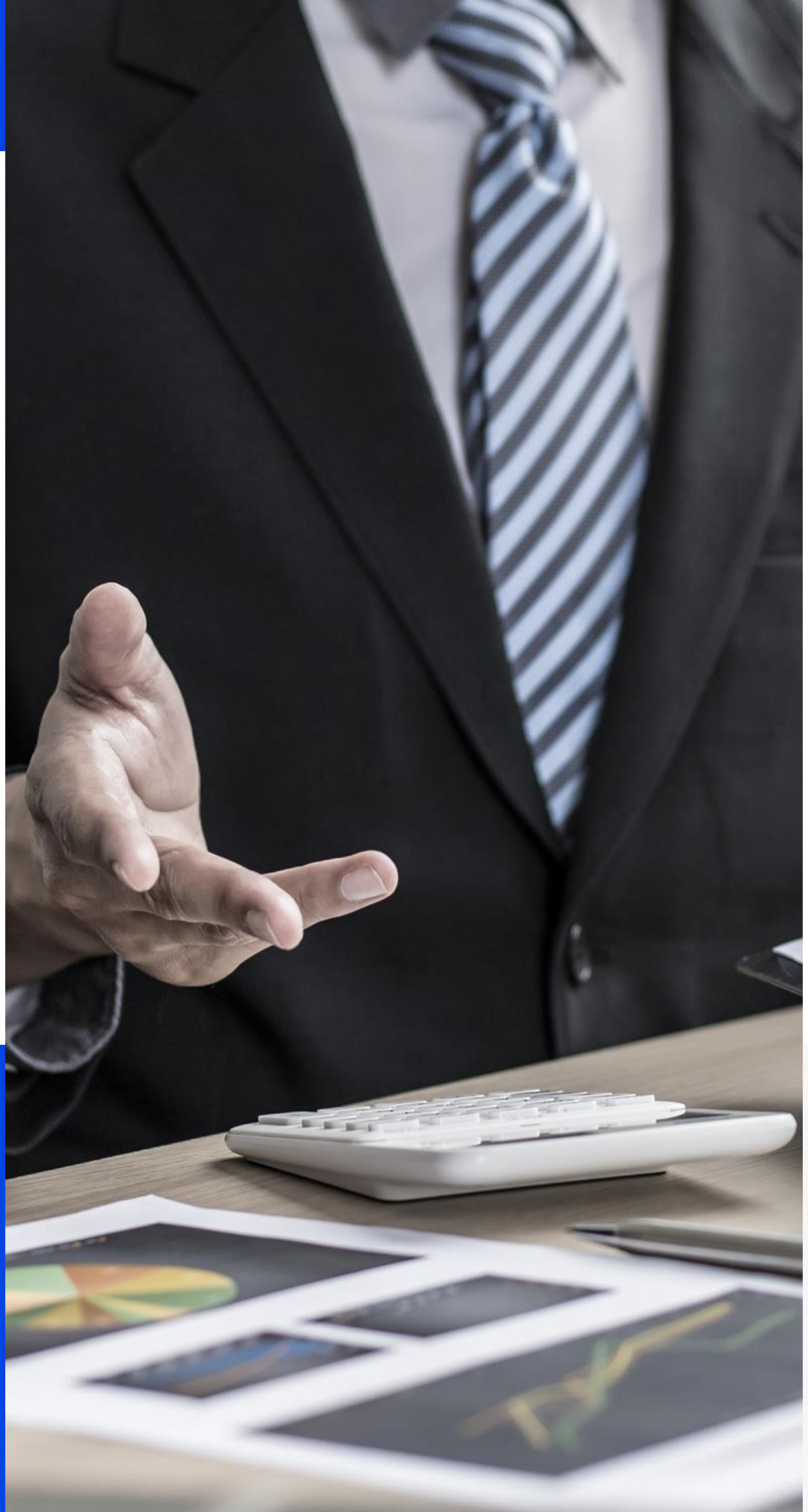
IEEE

CUSTOMER SEGMENTATION

May 2024



CONTENT

OVERVIEW

A company that sells some of the product, and you want to know how well does the selling performance of the product. We have the data that we can analyze, but what kind of analysis that we can do? Well, we can segment customers based on their buying behavior on the market. Keep in mind that the data is really huge, and we can not analyze it using our bare eyes. We will use machine learning algorithms and the power of computing for it.

Total Profit made from

89,526,124$

Number of Customers From 35 Countries

535,243

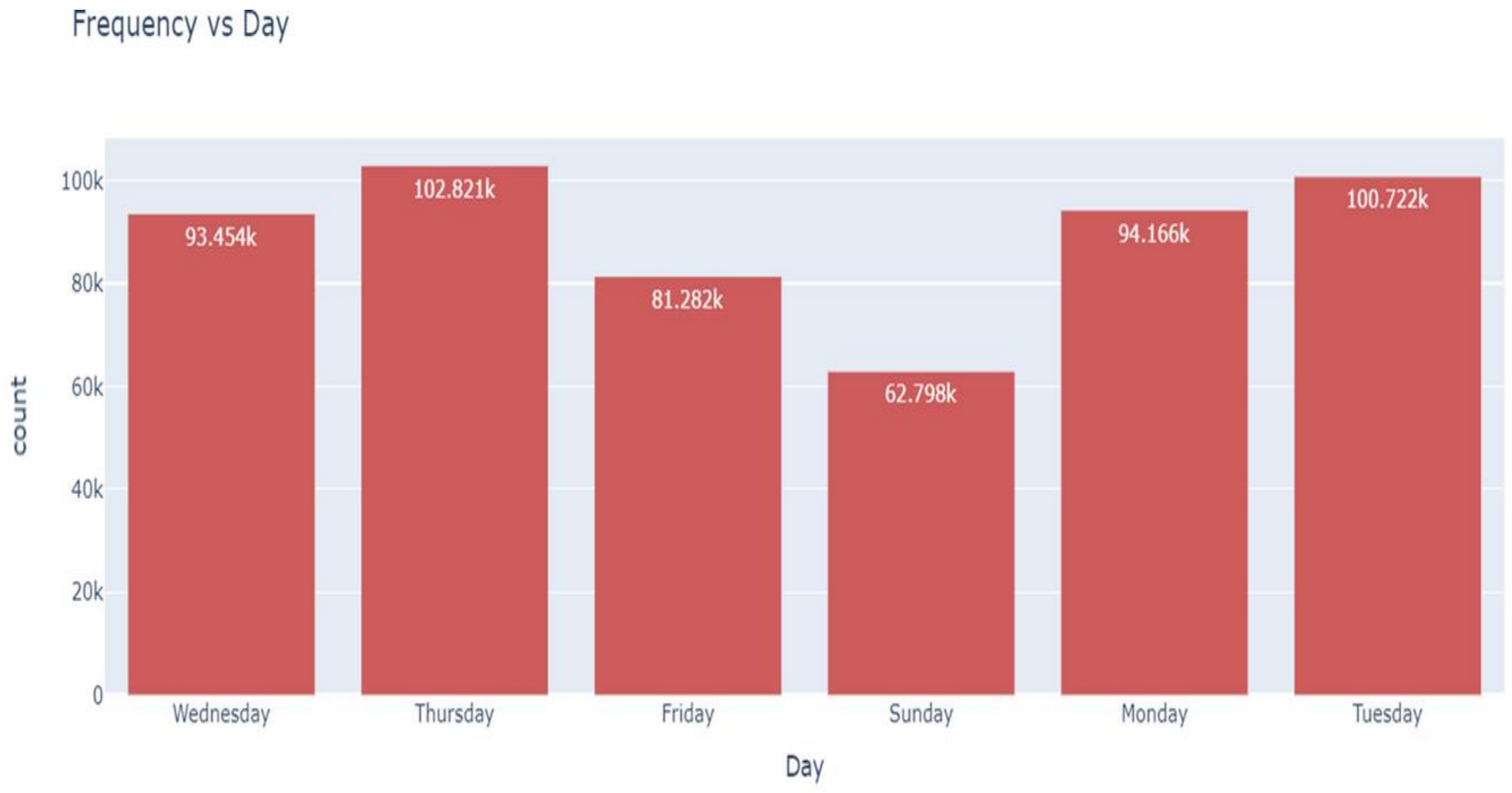
#	Feature	Description
1	InvoiceNo:	Unique identifier for each retail invoice or transaction.
2	StockCode:	Code for the specific product or item being sold.
3	Description	The actual age of the passengers
4	Quantity	Number of units of the product purchased in each transaction.
5	InvoiceDate:	Date and time of each retail transaction.
6	UnitPrice:	Price per unit of the product being sold.
7	CustomerID:	Unique identifier for each customer who made a purchase.
8	Country:	Name of the country where the customer is located.

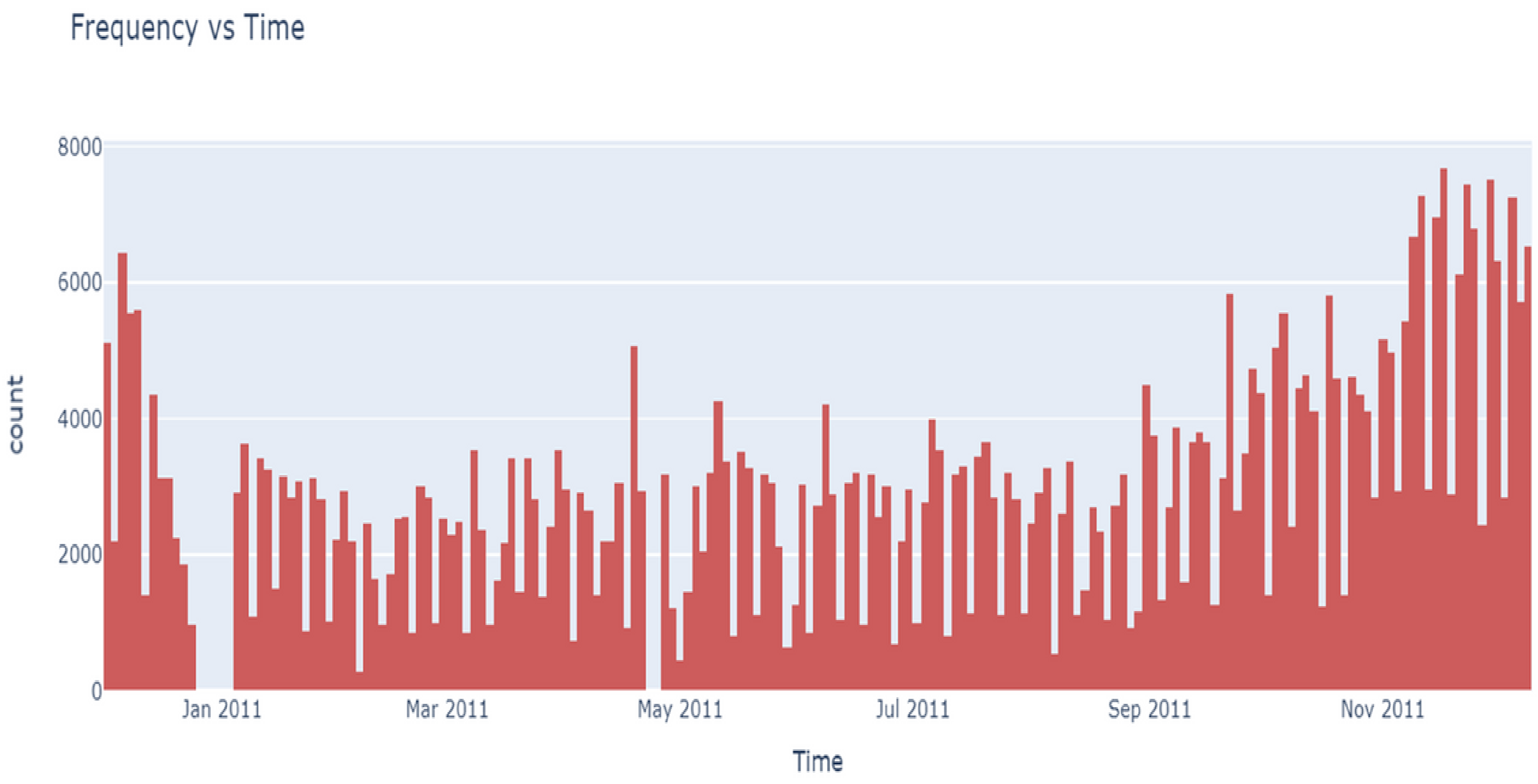
IEEE

EDA

EXPLORATORY DATA ANALYSIS

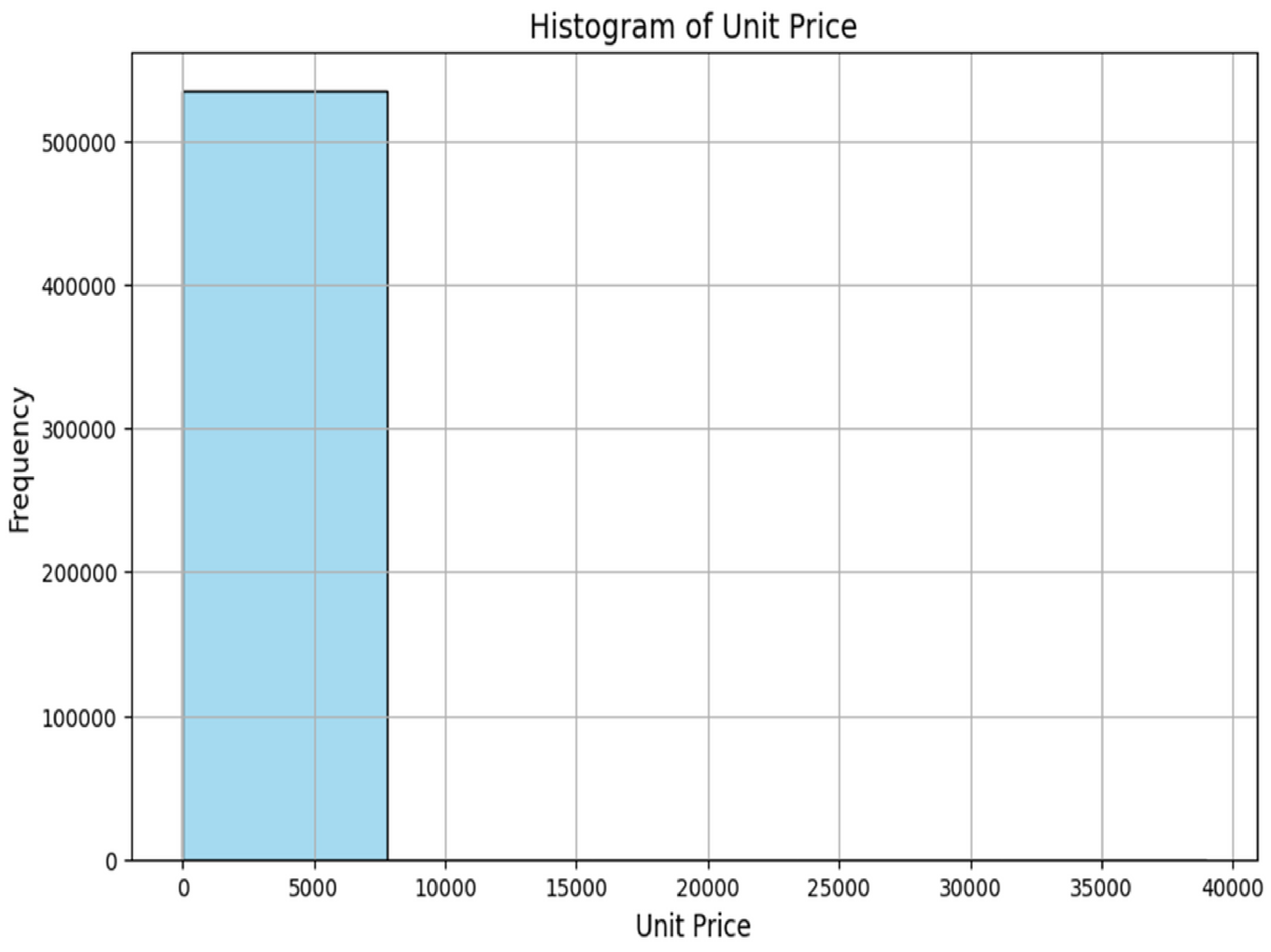
IEEE

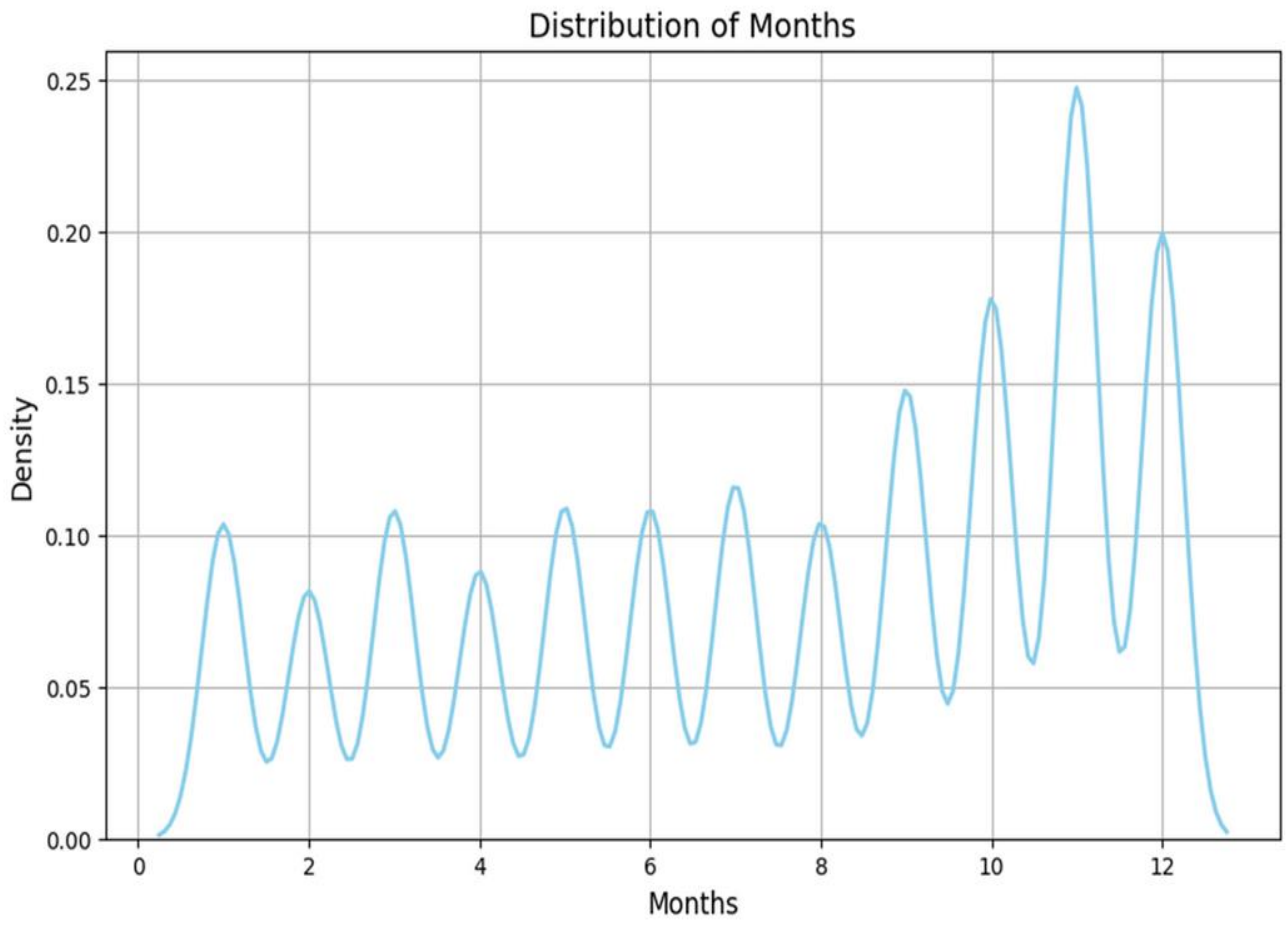




EXPLORATORY DATA ANALYSIS

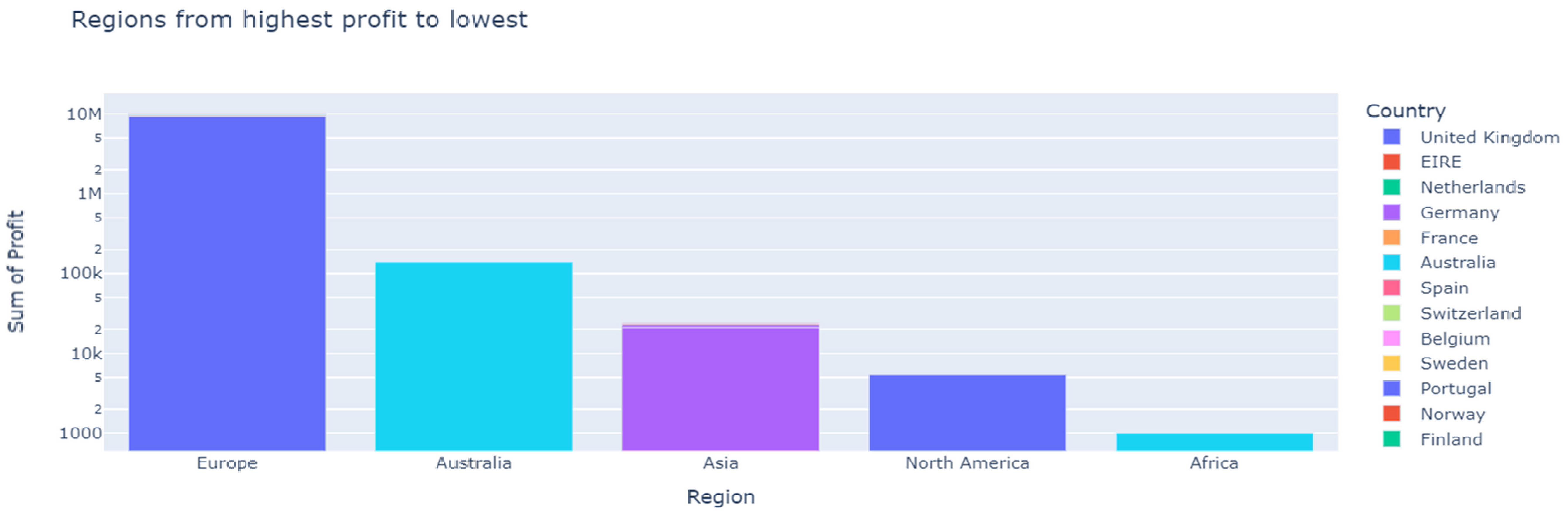
IEEE

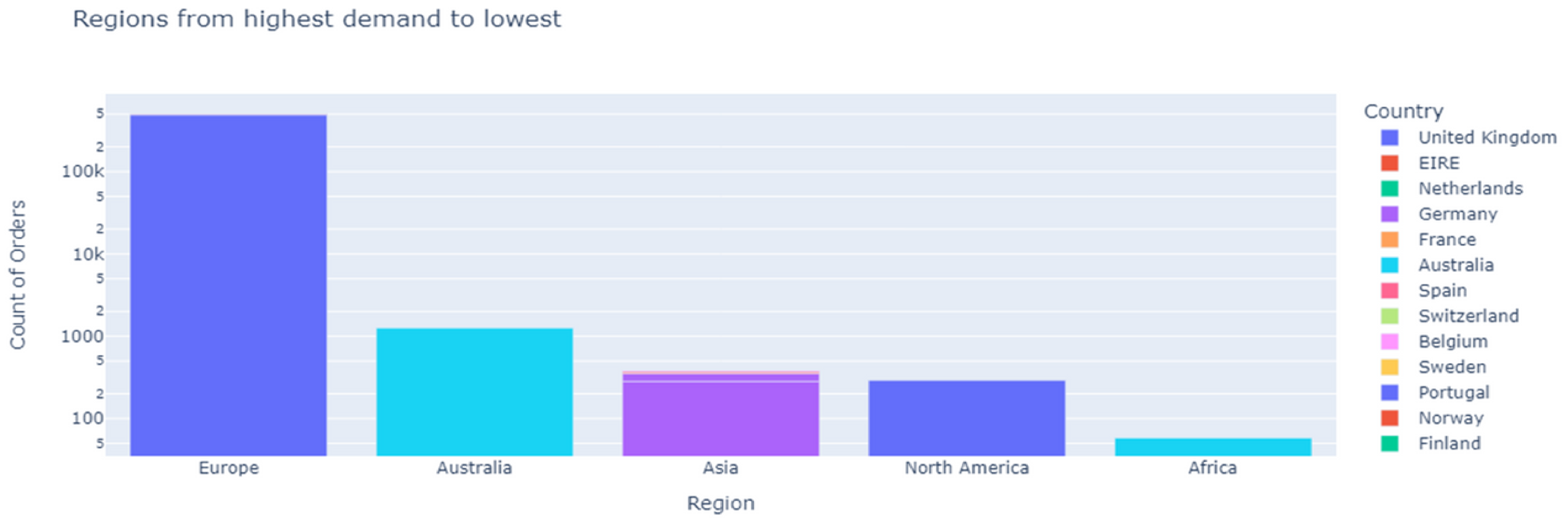




EXPLORATORY DATA ANALYSIS

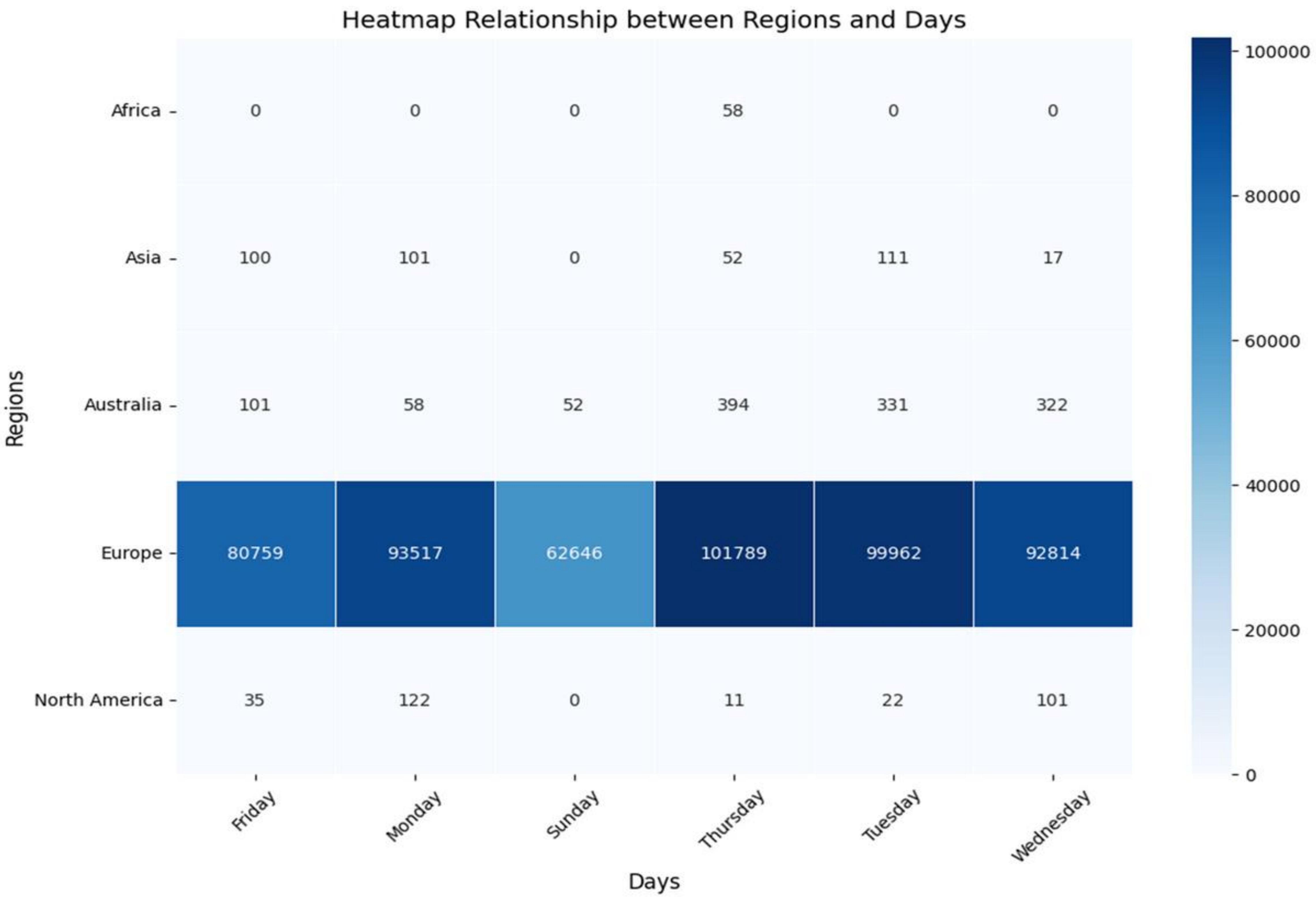
IEEE





EXPLORATORY DATA ANALYSIS

IEEE



EXPLORATORY DATA ANALYSIS

IEEE

- **Regions** Filled with loyal customers:

  Europe..

- **Regions** Filled with potential customers:

  Australia..

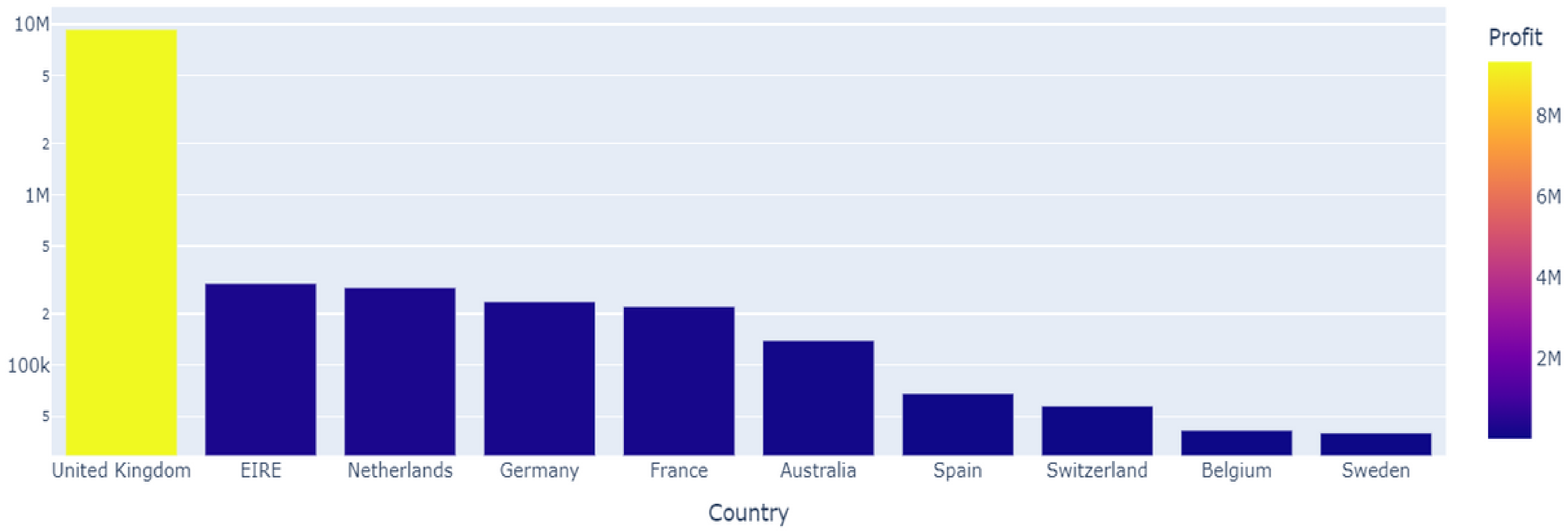
- **Regions** Has most churn customers:

  Africa..

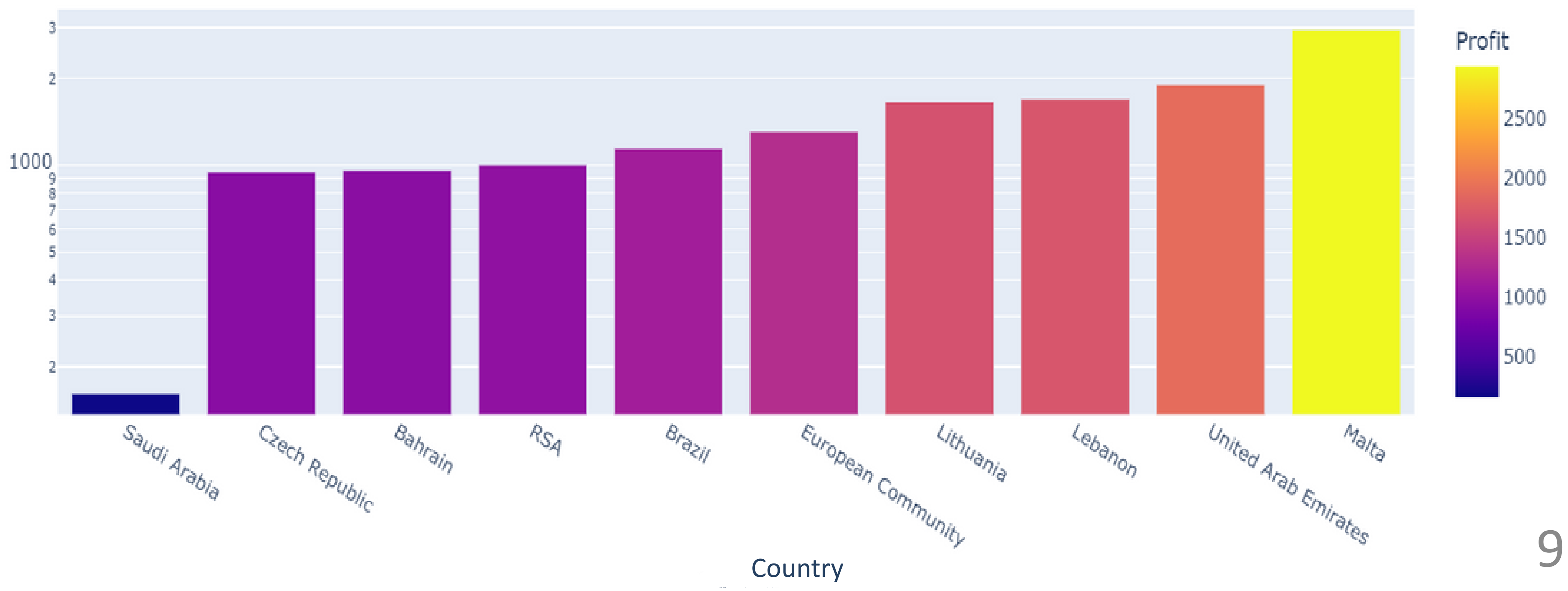
EXPLORATORY DATA ANALYSIS

IEEE

Top 10 Profitable Countries

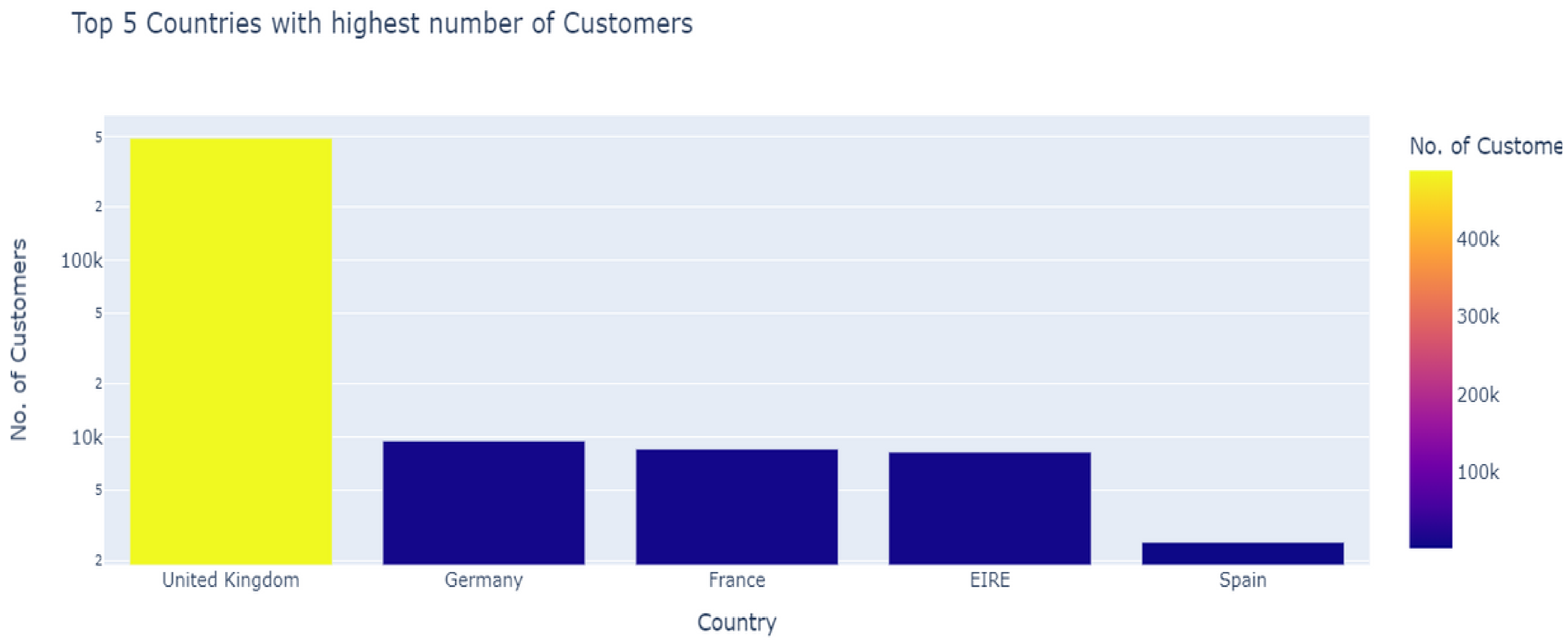


Least 10 Profitable Countries



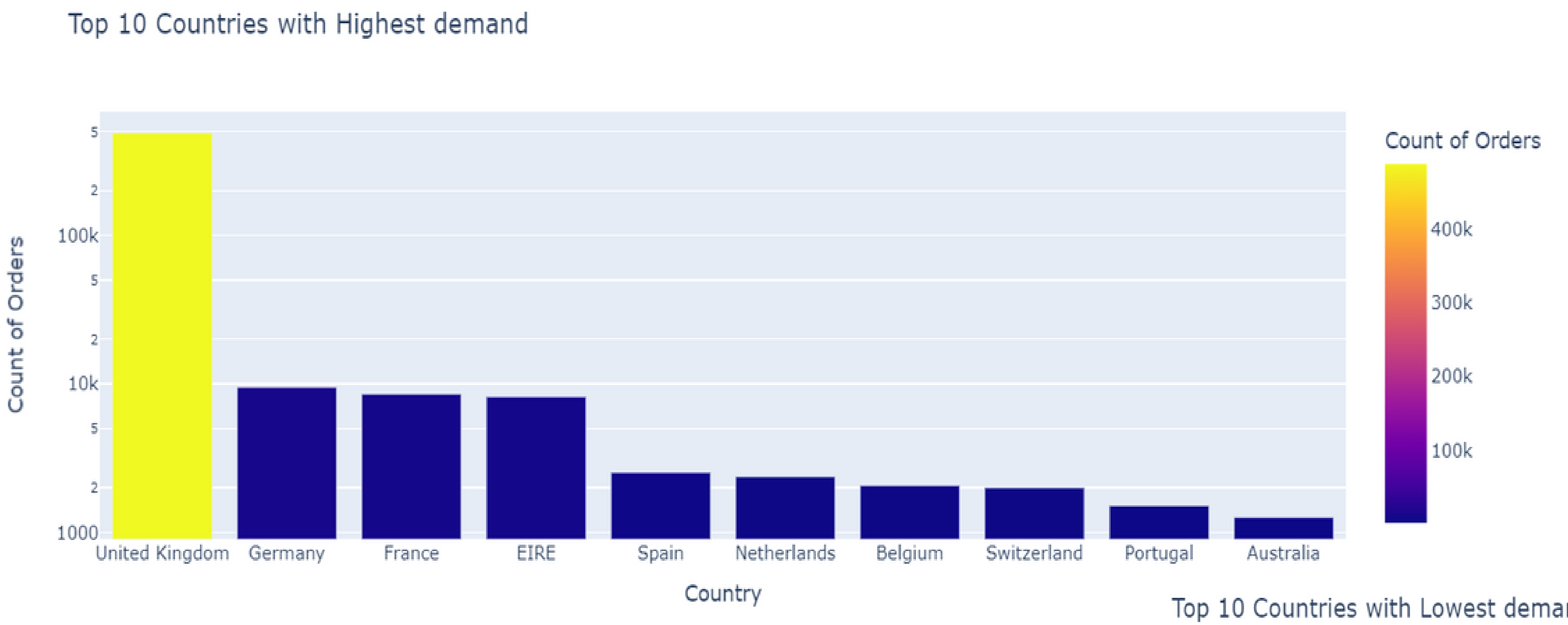
EXPLORATORY DATA ANALYSIS

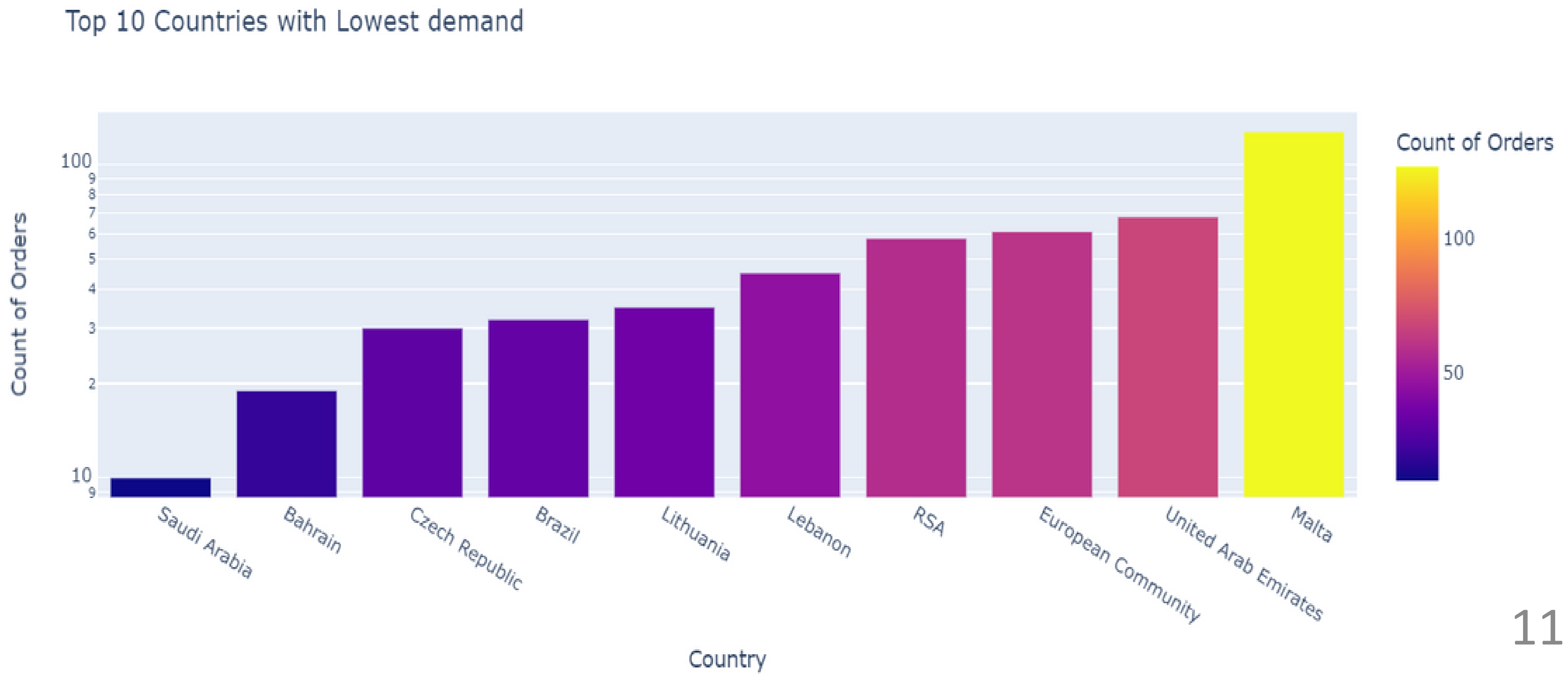
IEEE



EXPLORATORY DATA ANALYSIS

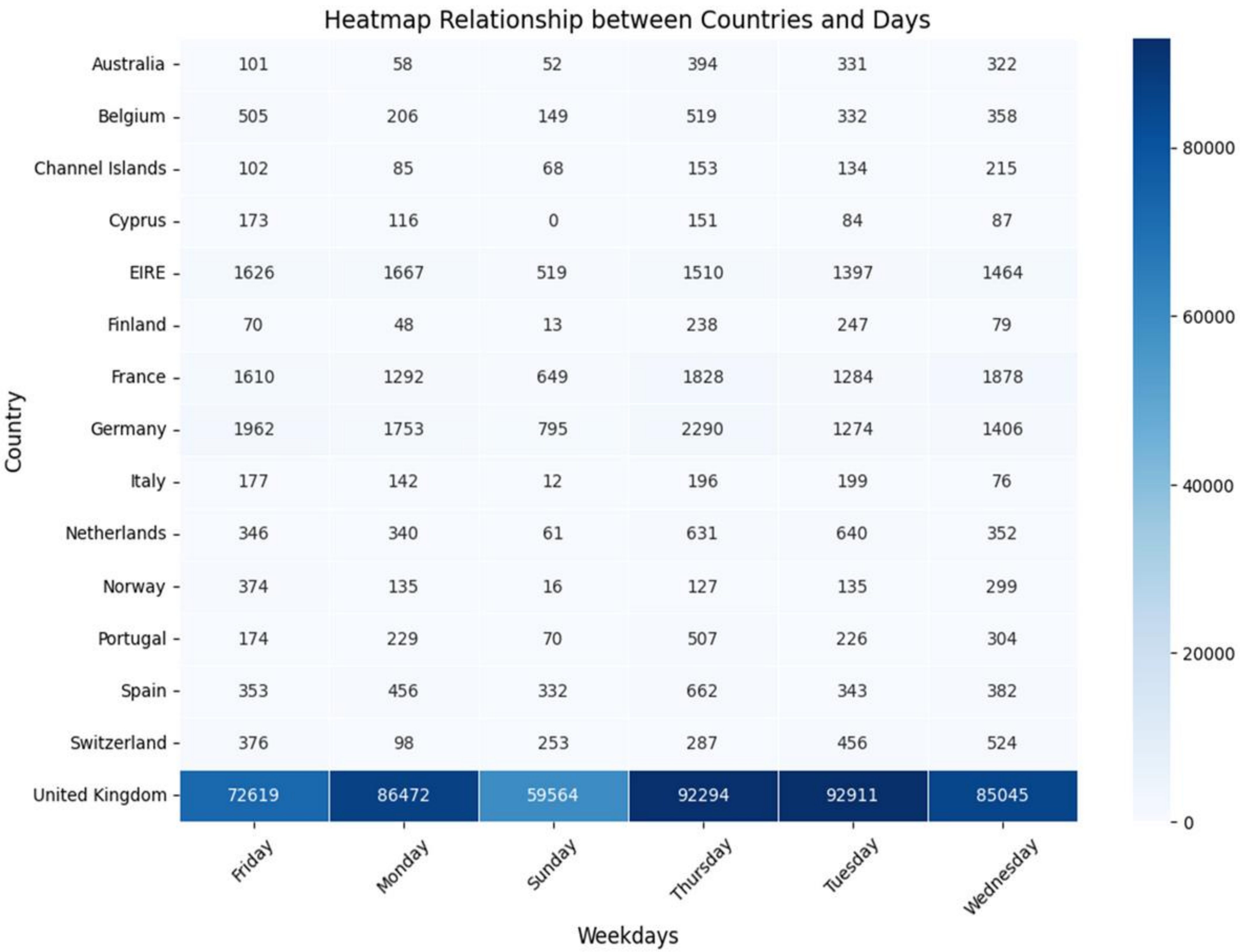
IEEE





EXPLORATORY DATA ANALYSIS

IEEE



EXPLORATORY DATA ANALYSIS

IEEE

- Countries Filled with loyal customers:

UK..

- Countries Filled with potential customers:

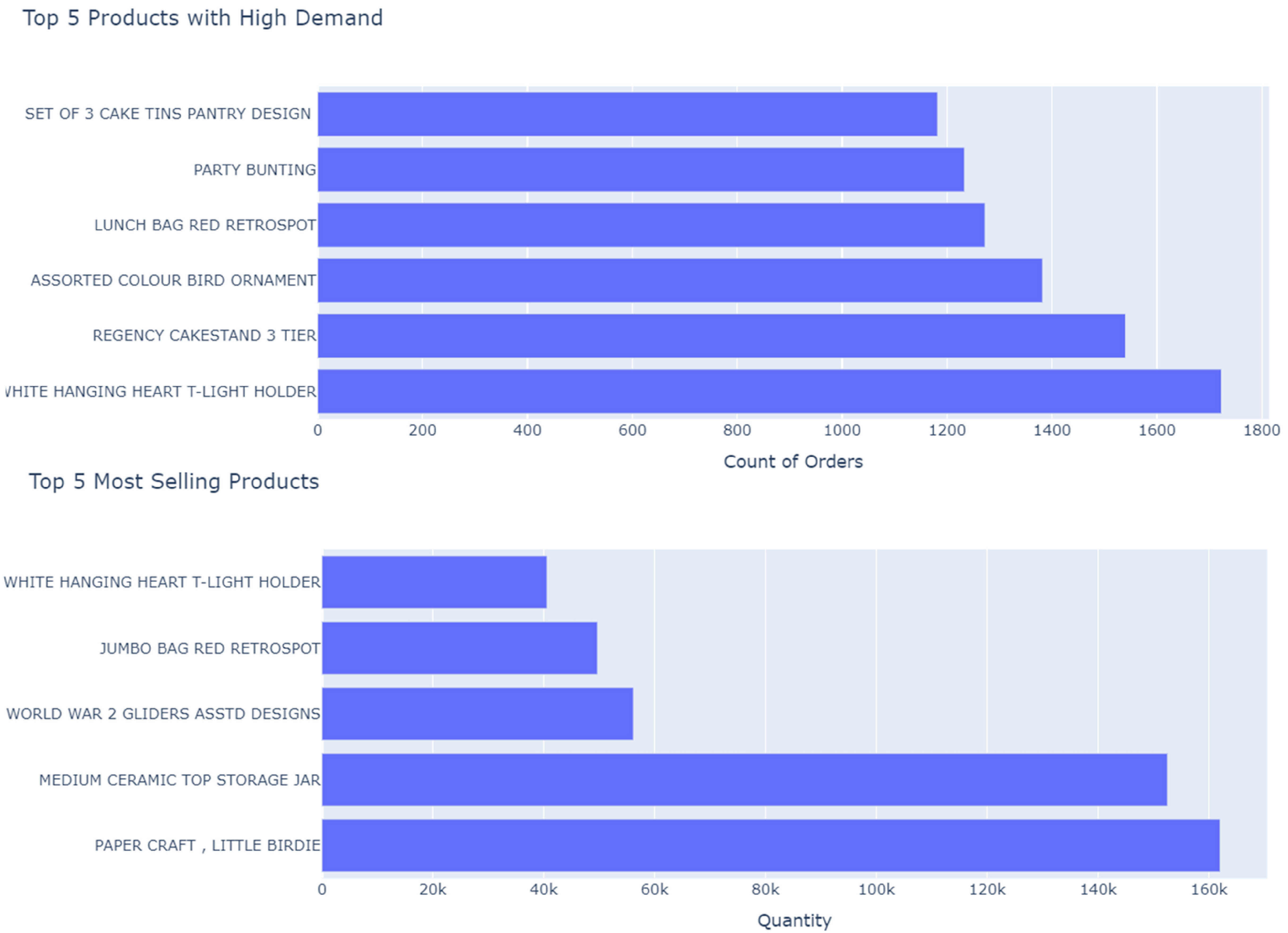
Germany..

- Countries Has most churn customers:

SA

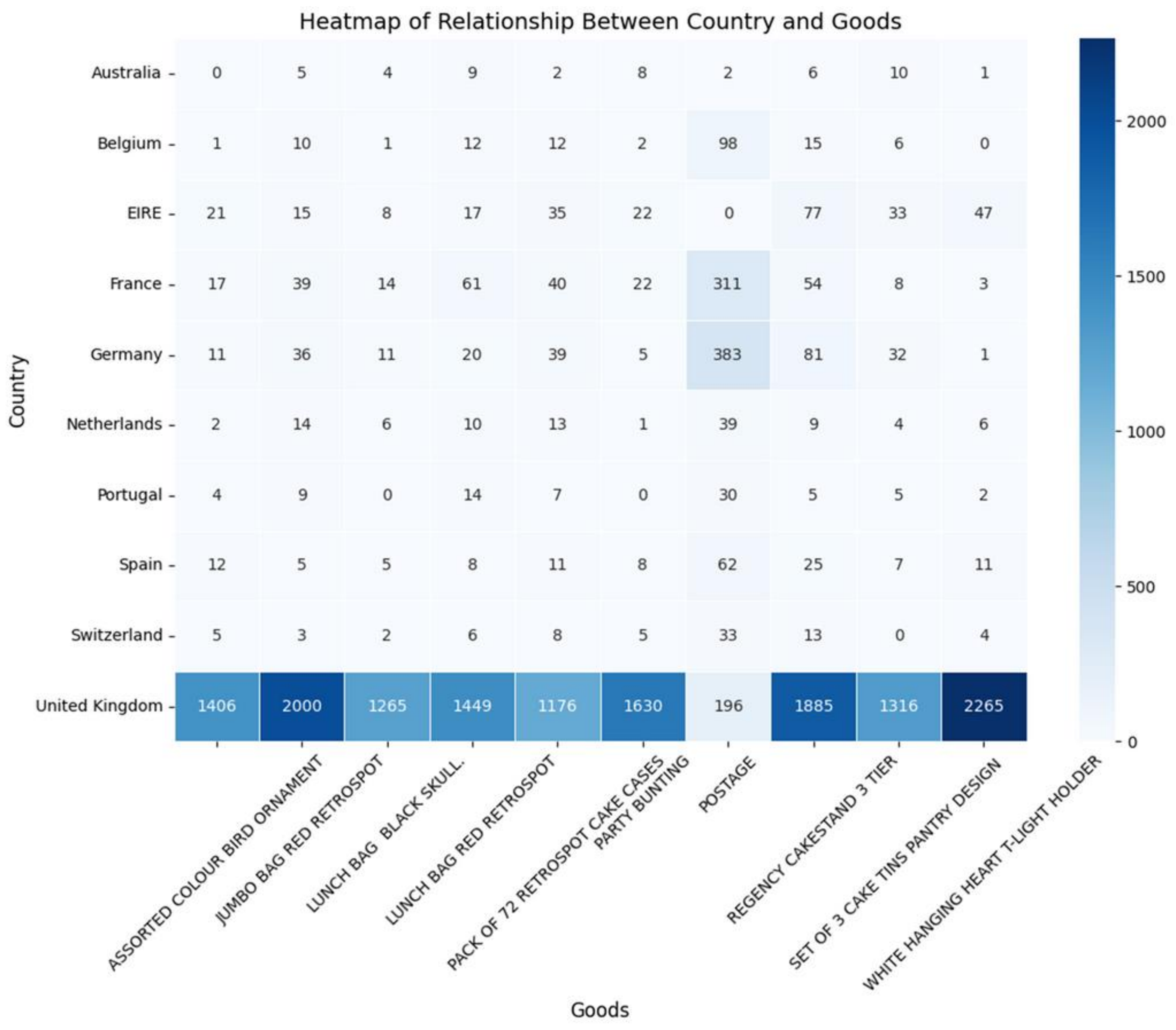
EXPLORATORY DATA ANALYSIS

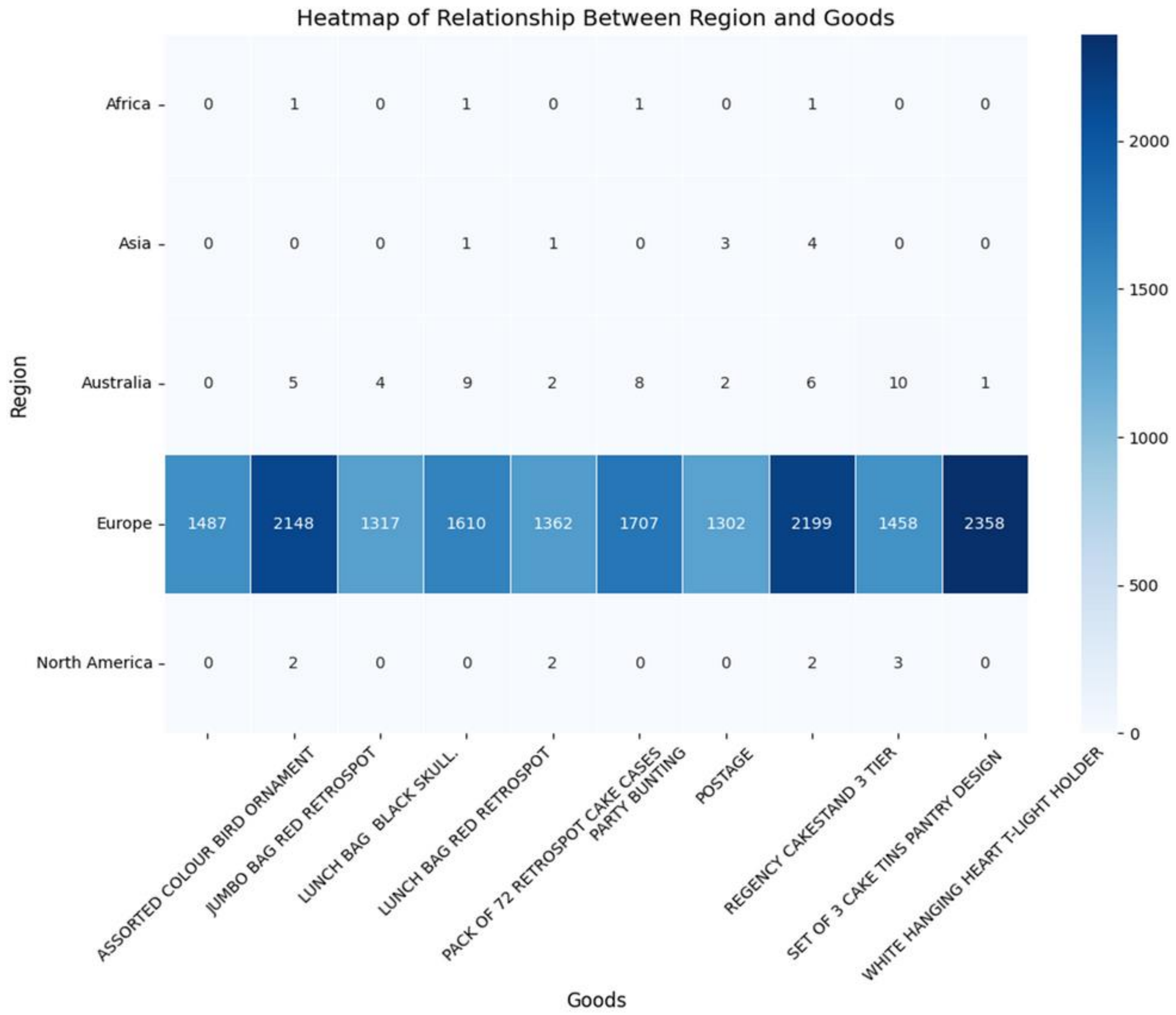
IEEE



EXPLORATORY DATA ANALYSIS

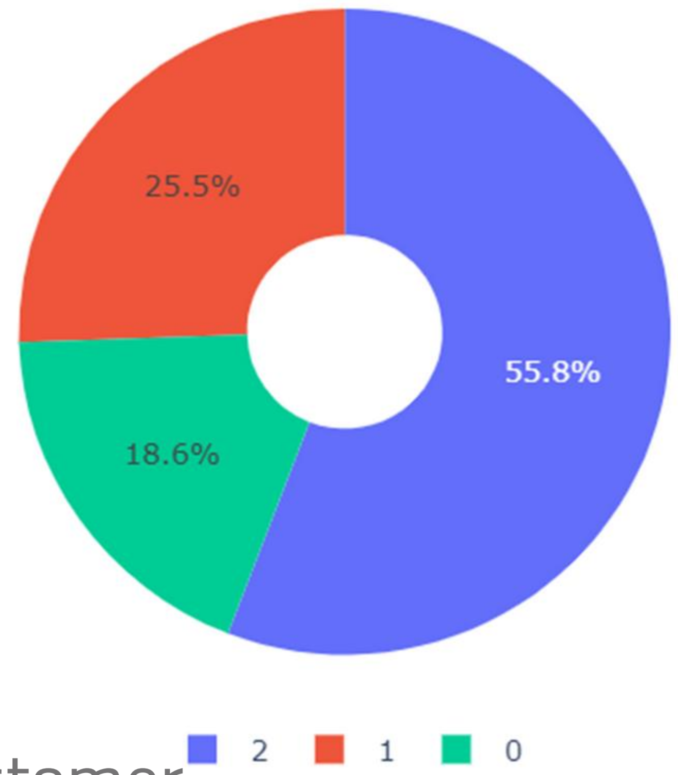
IEEE





EXPLORATORY DATA ANALYSIS

IEEE



Cluster (2) represents Loyal Customers:

Cluster 2 customers shopped recently, with high frequency and monetary, this means that these are the loyal customers that we want to keep.

Cluster (0) represents Potential Customers:

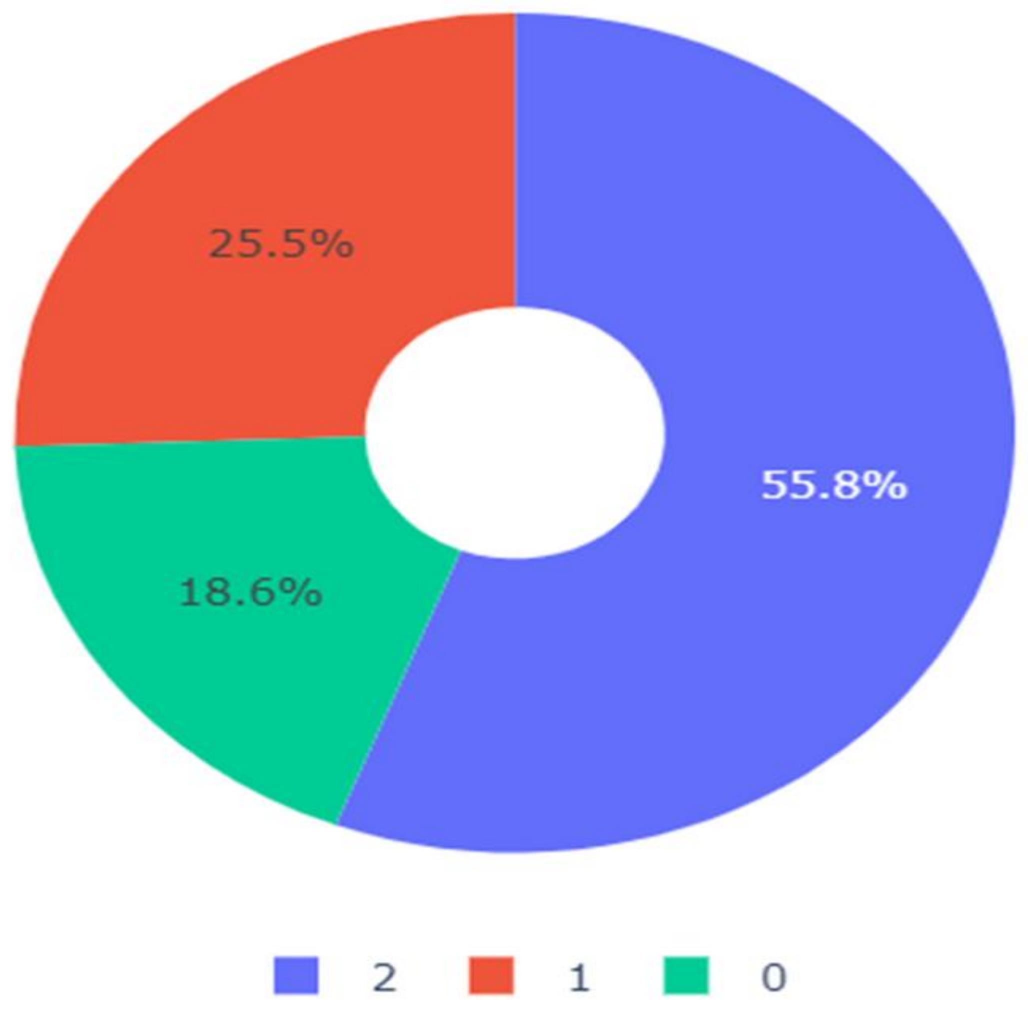
Cluster 0 has landed in the mid range for all three features, these might be our potential customer, that can be encouraged to be more engaged. And maybe there are things we can improve for them.

Cluster (1) represents Churn Customers:

Cluster 1 customers shopped not very recently (note that large recency means more days since their last purchase) and have rather low frequency. Yet, have similar monetary value as cluster 0. These might be some old customers that have quite good purchase ability but became inactive for some reasons, it might be a good idea to figure out why they haven't shopped here for a while.

EXPLORATORY DATA ANALYSIS

IEEE



Good news is the shop has:
mostly loyal customers (56%),
about quarter of customers, those who churned and maybe can be brought back (25%),
And new customers that can be encouraged to buy more (19%)

SOLUTIONS

IEEE

Improve unique selling point (USP)

- Analyze what keeps loyal customers

Increase engagement for potential customers

- Tailor strategies to motivate them.

Re-engage churn customers

- Understand reasons for inactivity.

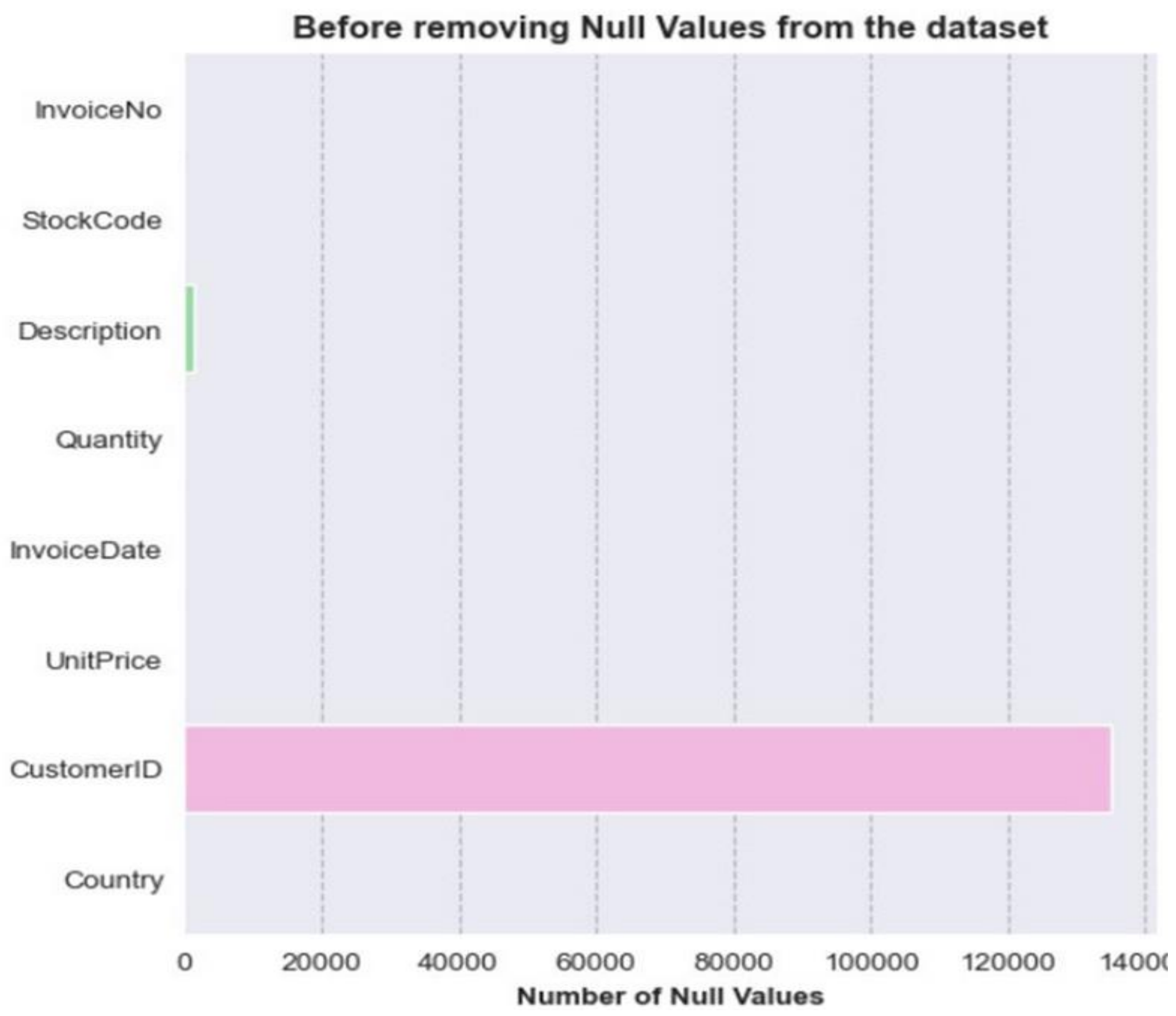
IEEE

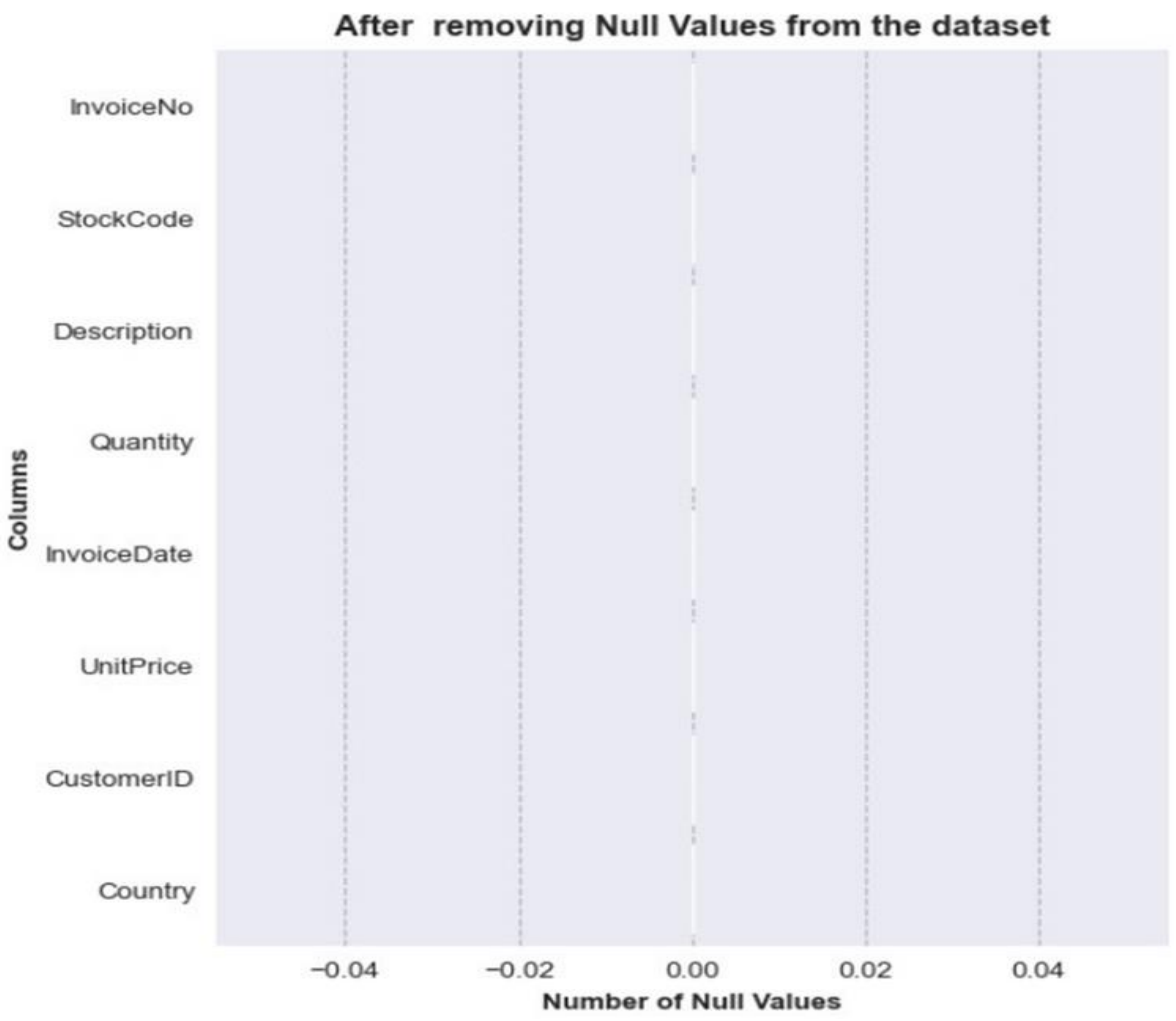
PREPROCESSING

PREPROCESSING

IEEE

Null Values





PREPROCESSING

IEEE

	Total number of null values	Percentage of null values
CustomerID	135080	24.927
Description	1454	0.268
InvoiceNo	0	0.000
StockCode	0	0.000
Quantity	0	0.000
InvoiceDate	0	0.000
UnitPrice	0	0.000
Country	0	0.000

	Total number of null values	Percentage of null values
InvoiceNo	0	0.0
StockCode	0	0.0
Description	0	0.0
Quantity	0	0.0
InvoiceDate	0	0.0
UnitPrice	0	0.0
CustomerID	0	0.0
Country	0	0.0

IEEE

Feature Engineering

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Month of purchase

Day of purchase

Quality X Unitprice

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	InV_Year	InV_Month	InV_Day	WDay	WeekDay	Profit
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	2010	12	1	2	Wednesday	15.30
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	2010	12	1	2	Wednesday	20.34

Year of purchase

Weekday name And number.

IEEE

Feature Engineering

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Number of orders done by the customer

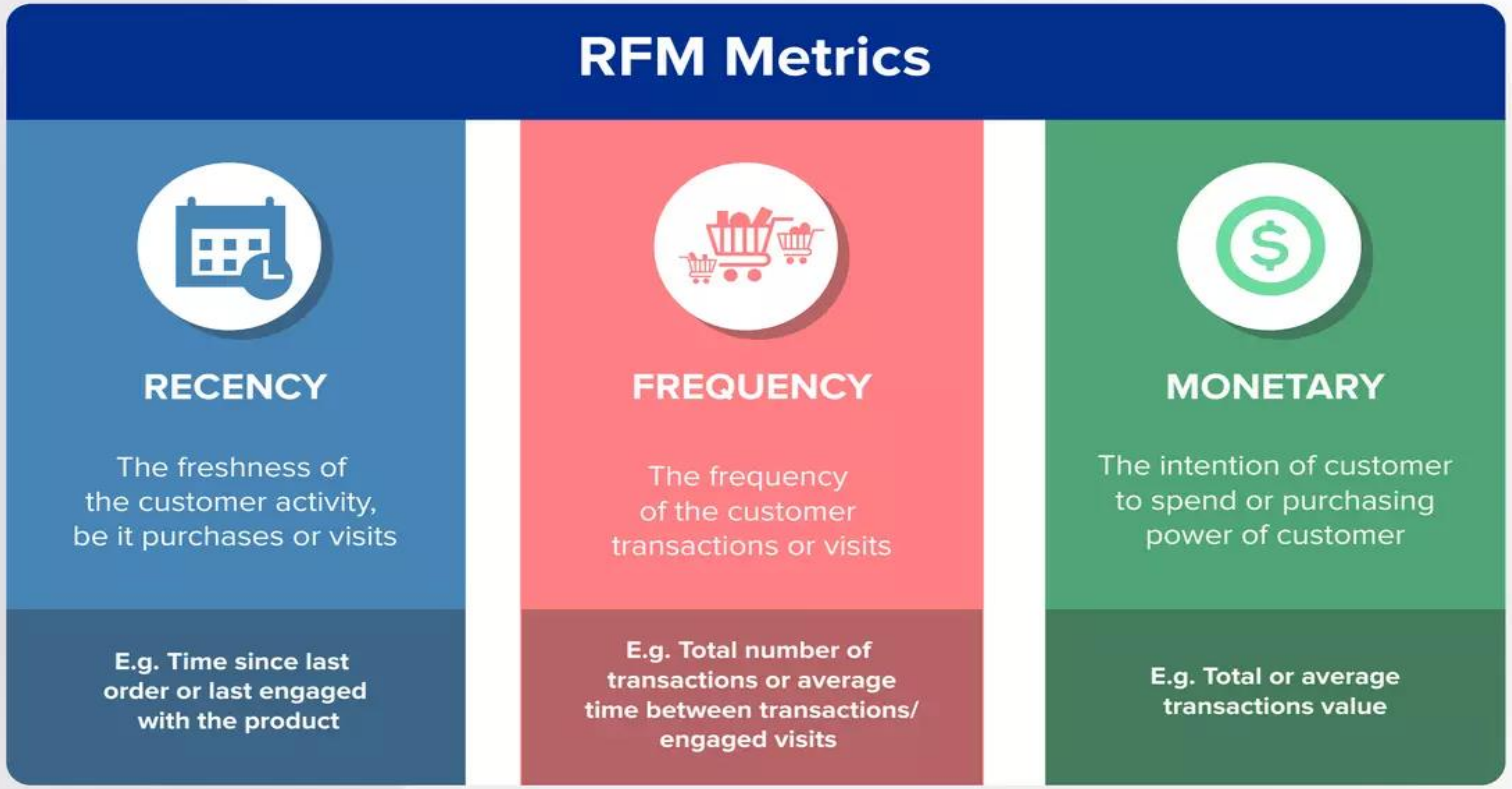
Summation of profit for one customer

CustomerID	Recency	Frequency	Monetary
12346.0	325	2	154367.20
12347.0	1	182	4310.00

Difference between 1st and last purchase day

What is RFM Analysis?

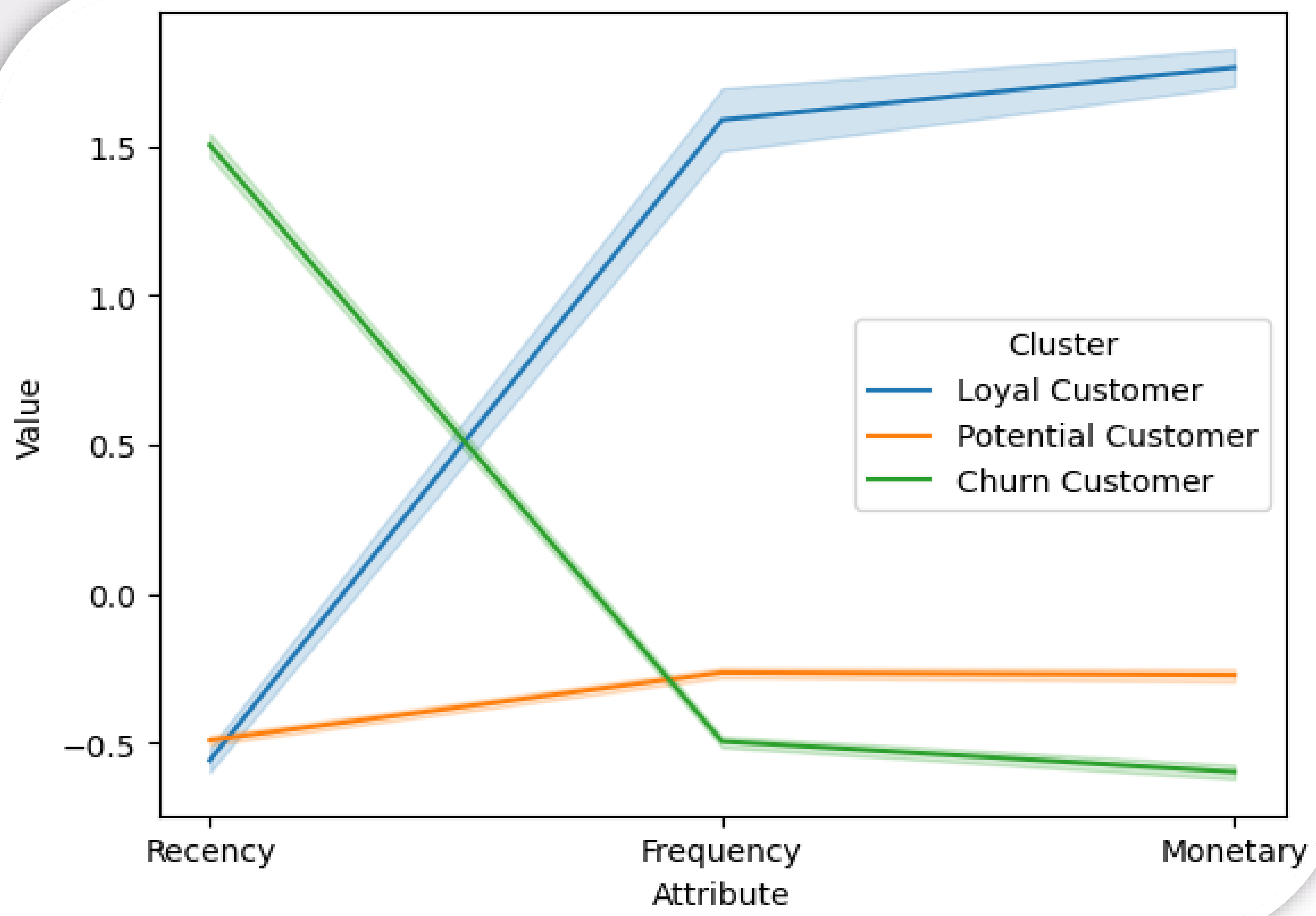
RFM (Recency, Frequency, and Monetary value) Analysis measures how recently, how often, and how much money a customer has given your brand.



IEEE

What is RFM Analysis?

RFM (Recency, Frequency, and Monetary value) Analysis measures how recently, how often, and how much money a customer has given your brand.



How Does RFM Help?

RFM analysis helps marketers answer:

- Who are their best customers?
- Which of their customers could contribute to churn rate?
- Who has the potential to become valuable customers?

IEEE

Feature Engineering

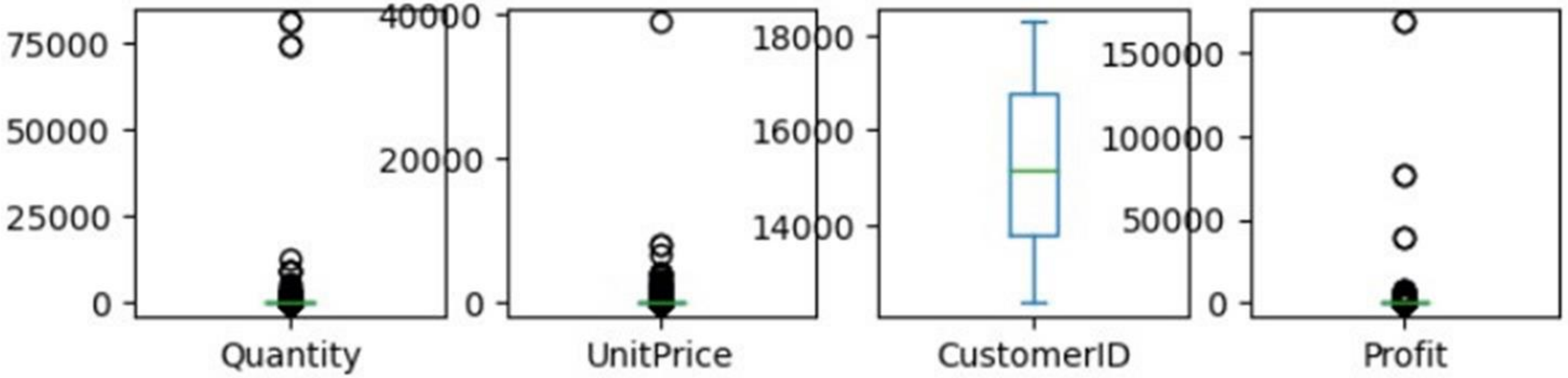
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

CustomerID	Days since last purchase (Recency)	Count of Orders (Frequency)	Sum of Profit (Monetary)	Country
17841.0	1	8459	49248.03	United Kingdom
12748.0	0	5941	55619.79	United Kingdom
14911.0	0	5897	154529.10	EIRE
14096.0	3	5394	63336.12	United Kingdom
14606.0	0	3924	25376.86	United Kingdom

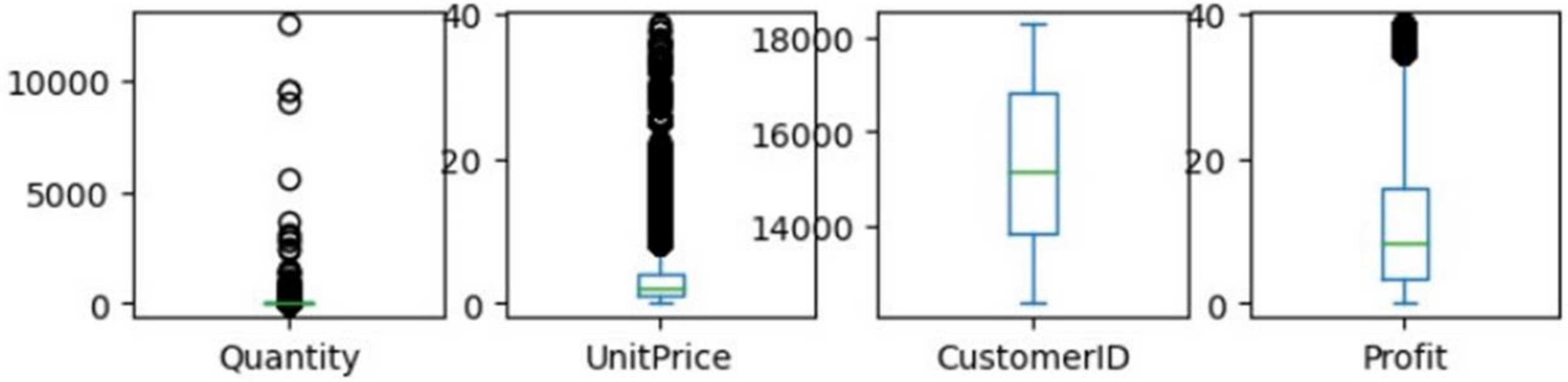
PREPROCESSING

IEEE

Outliers

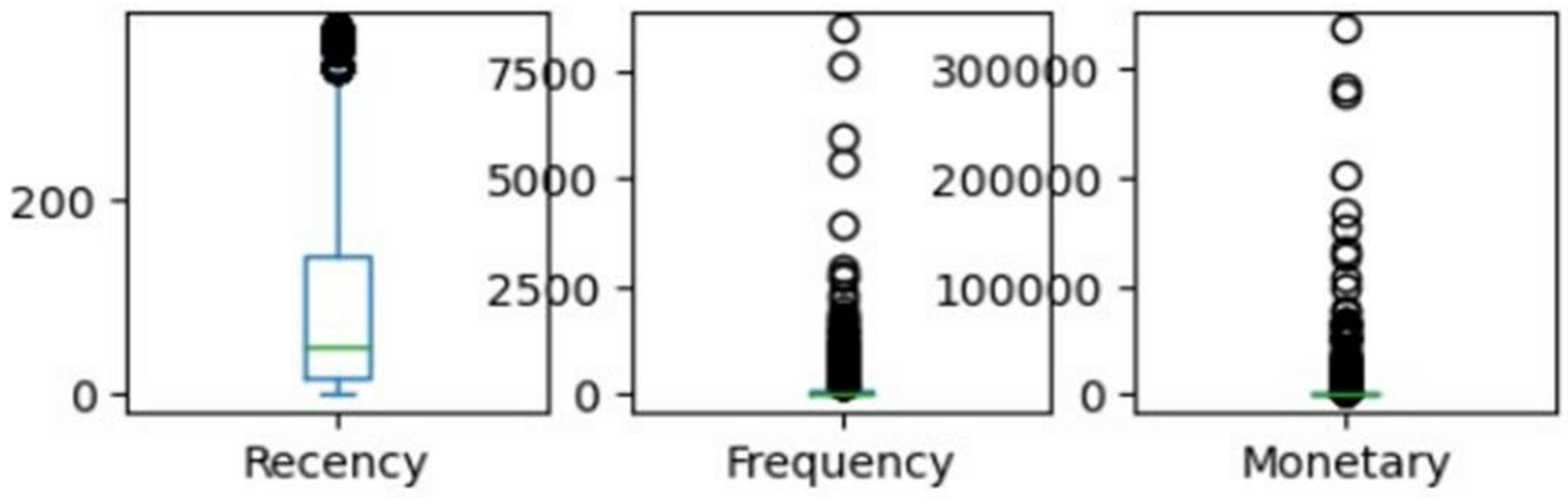


- **Before removing outliers: 535243**
- **After removing outliers: 492206**
- **Outliers: 43037**

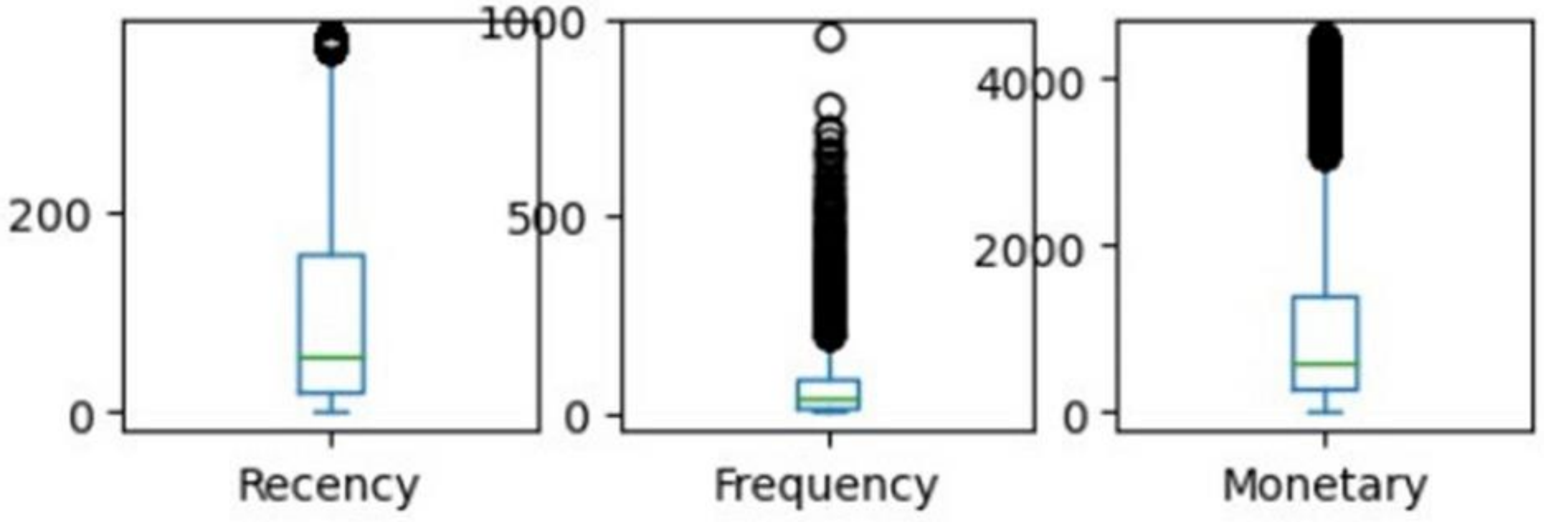


IEEE

Outliers



- Before removing outliers: 4372
- After removing outliers: 3916
- Outliers: 456



IEEE

MODELING

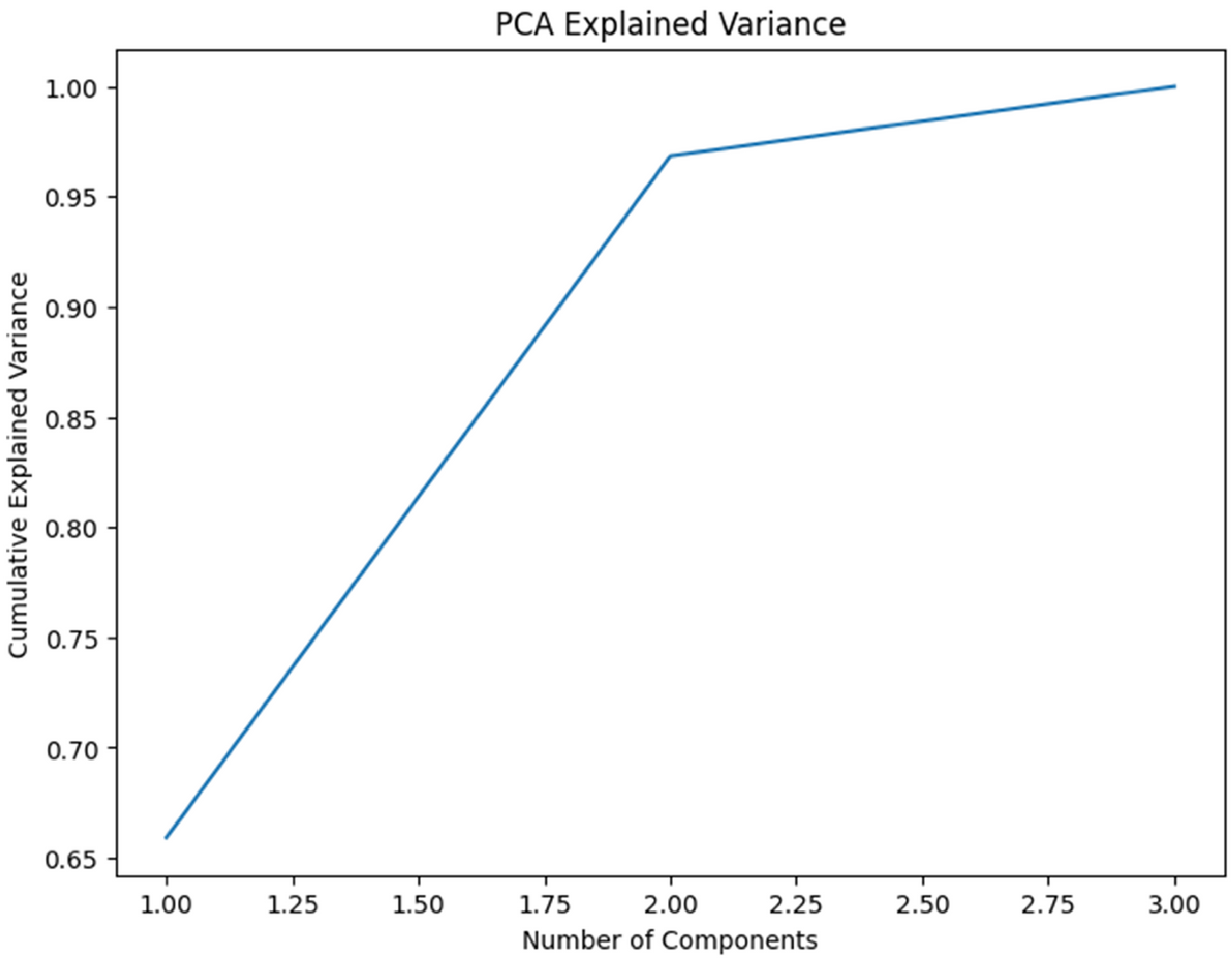
IEEE

Feeding the model!

	Recency	Frequency	Monetary
CustomerID			
12346.0	325	2	154367.20
12347.0	1	182	4310.00
12348.0	74	205	3150.16
12349.0	18	73	1757.55
12350.0	309	17	334.40

IEEE

PCA dimensionality reduction



MODELING

IEEE

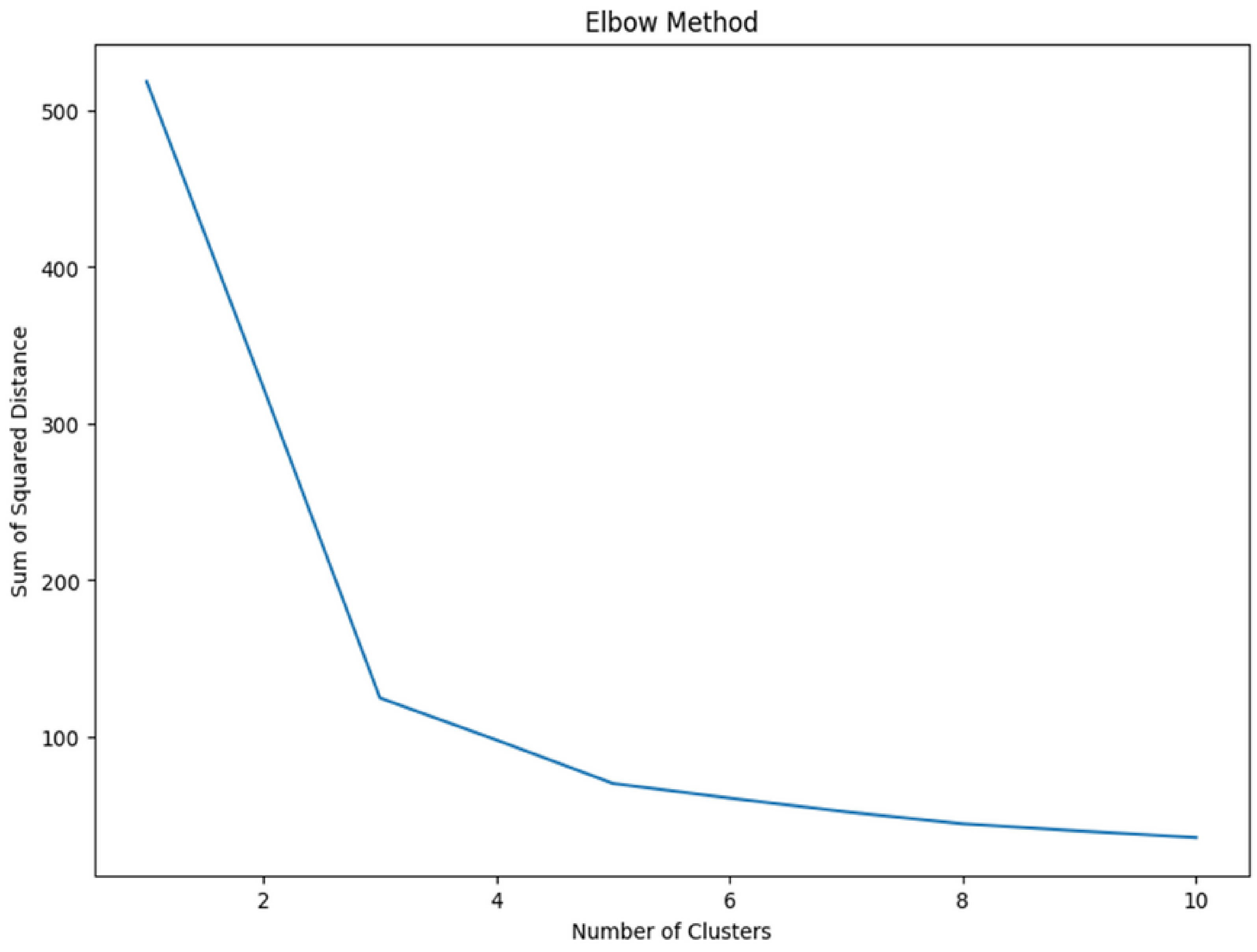
DBSCAN


```
[ ] silhouette = silhouette_score(dfcan,dbscan.labels_)
    print("Silhouette Coefficient:", silhouette)

Silhouette Coefficient: -0.369296214559851
```

- **We Deduce that DBCSAN IS NOT The Best Choice**

KMeans



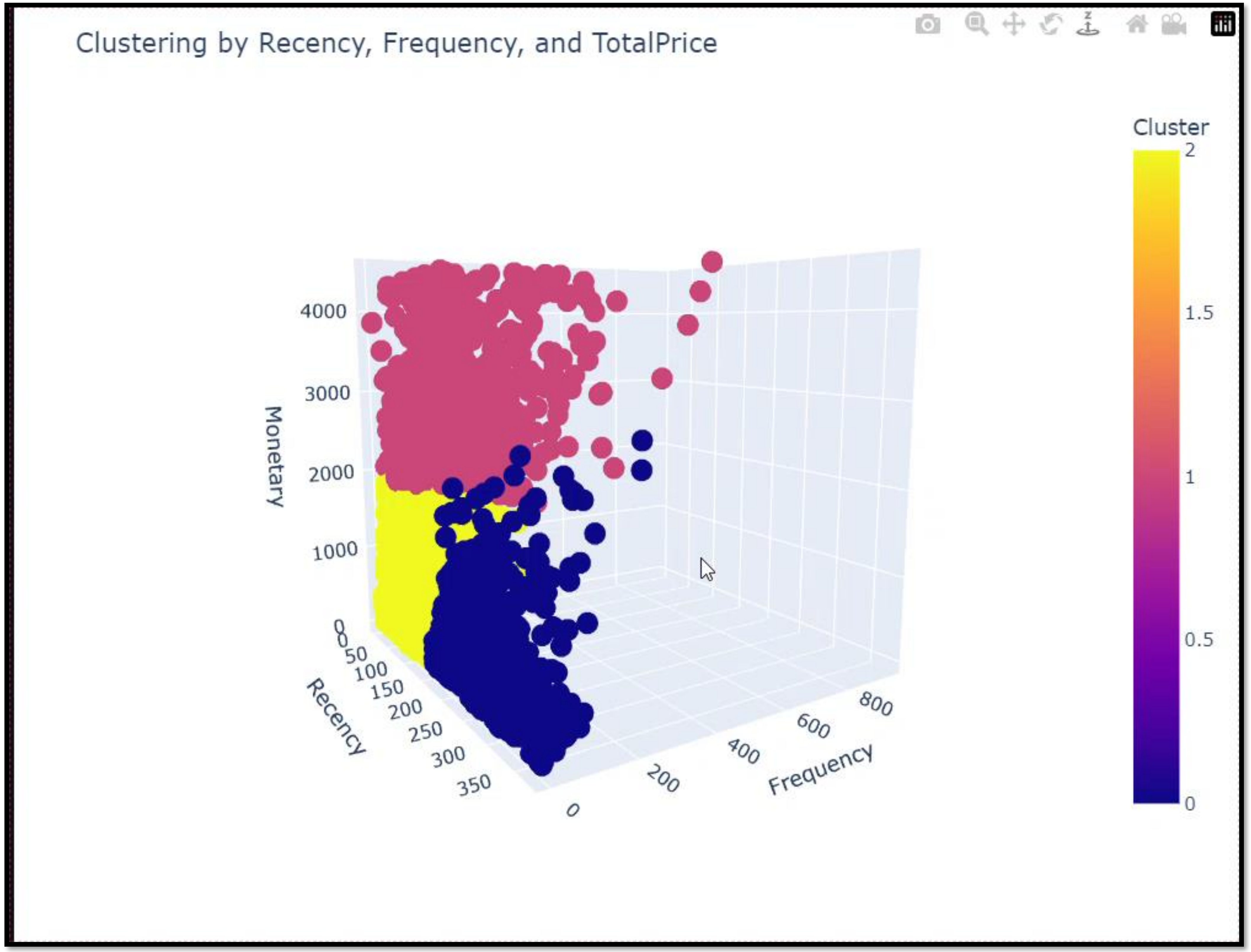
- Elbow Method used to find optimal cluster number; it is 3.

```
Silhouette Coefficient: 0.5483758055140908
```

The Best Model!

MODELING

IEEE

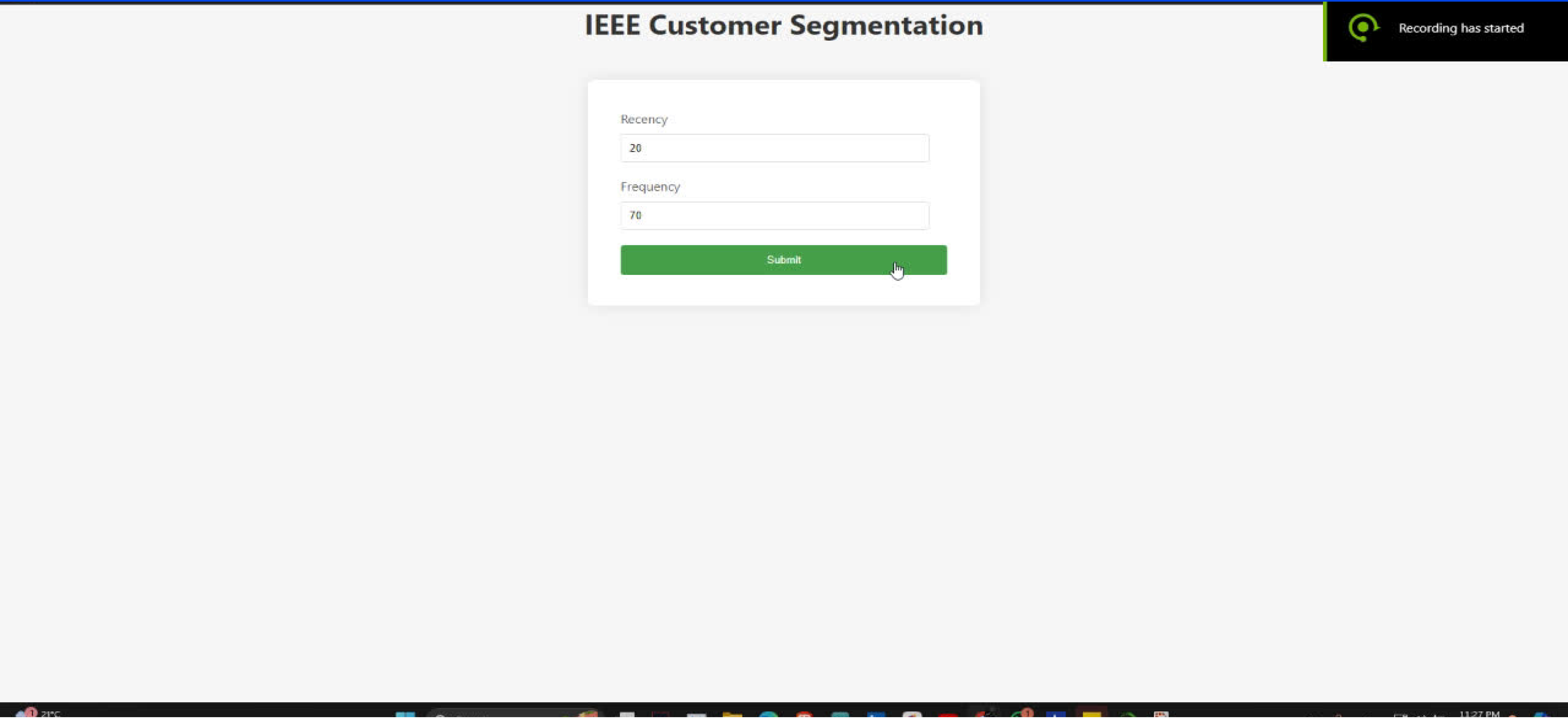


IEEE

DEPLOYMENT

DEPLOYMENT

IEEE



OUR TEAM

Moemn Adel

LeaderTeam

Abdelrahman Alaa

Mohamed Ehab

Ahmed Yusseif

Ahmed Mostafa

Hassan Hossam

Yara Mohy

IEEE

THANK YOU

FOR YOUR ATTENTION

May 2024

[Github]