Análisis de Ventas y Segmentación de Clientes en un E-commerce

Autor: Italo Buonanno

Mentora: Alana Olivieri

Objetivo General:

Analizar el comportamiento de los usuarios en un entorno de e-commerce para segmentarlos en distintas categorías relevantes para un estudio de mercado.

Objetivos Específicos:

Analizar la cantidad total de productos vendidos.

Examinar las cantidades vendidas por mes.

Analizar el comportamiento de compra de los usuarios.

Agrupar a los clientes según su comportamiento de compra utilizando el modelo Recency/Frequency/Monetary (RFM).

Aplicar un modelo de Random Forest para predecir la categoría de los clientes.

Dataset

UC Irvine Machine Learning Repository

Datasets   Contribute Dataset   About Us   [Search datasets...]   Login

Online Retail
Donated on 11/5/2015

This is a transactional data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

Dataset Characteristics
Multivariate, Sequential, Time-Series

Subject Area
Business

Associated Tasks
Classification, Clustering

Feature Type
Integer, Real

 # Instances
541909

 # Features
6

DOWNLOAD (22.6 MB)

IMPORT IN PYTHON

CITE

99 **8 citations**
👁 **364664 views**

Keywords
(sales)

Creators

Daqing Chen
chend@lsbu.ac.uk
School of Engineering, London South Bank University

DOI
10.24432/C5BW33

Dataset Information

Additional Information
This is a transactional data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

Has Missing Values?
No

◈ https://archive.ics.uci.edu/dataset/352/online+retail

Dataset

	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
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850.0	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850.0	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047.0	United Kingdom

Herramientas

Analisis Descriptivo

Productos Mas Vendidos

Cantidad De Ventas Por Mes

Usuarios Que Mas Gastan

Top 10 De Los Usuarios Que Mas Gastan

Total Gastado / Usuarios

0.3M, 0.2M, 0.1M, 0.0M

14646, 18102, 17450, 14911, 12415, 14156, 17511, 16684, 13694, 15311
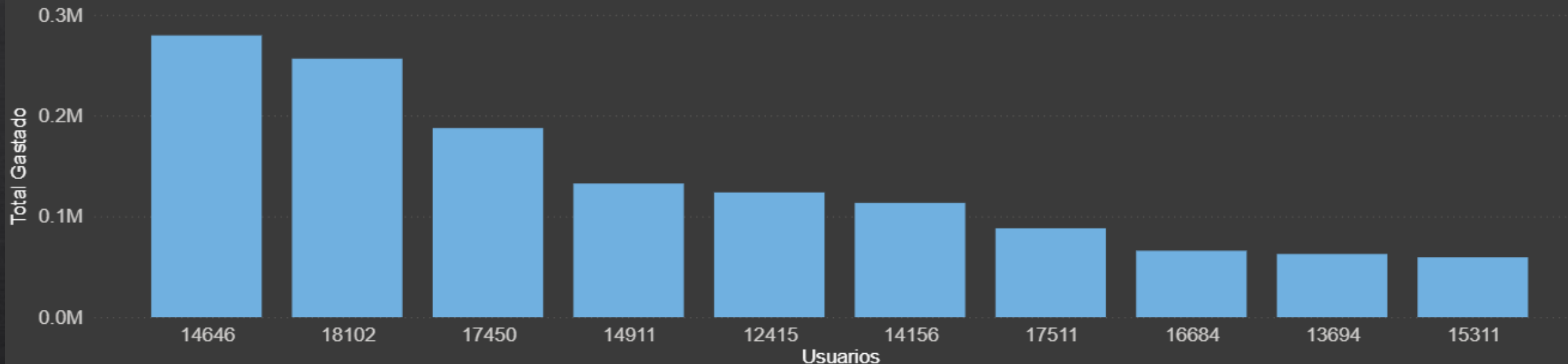
Compras De Los Usuarios

● Ventas Brutas   ● Ventas Netas   ● Devoluciones

Numero De Compras / Usuario

200, 100, 0

14911, 14646, 18102, 17450, 12415, 16297, 14586, 15973, 17044, 18276
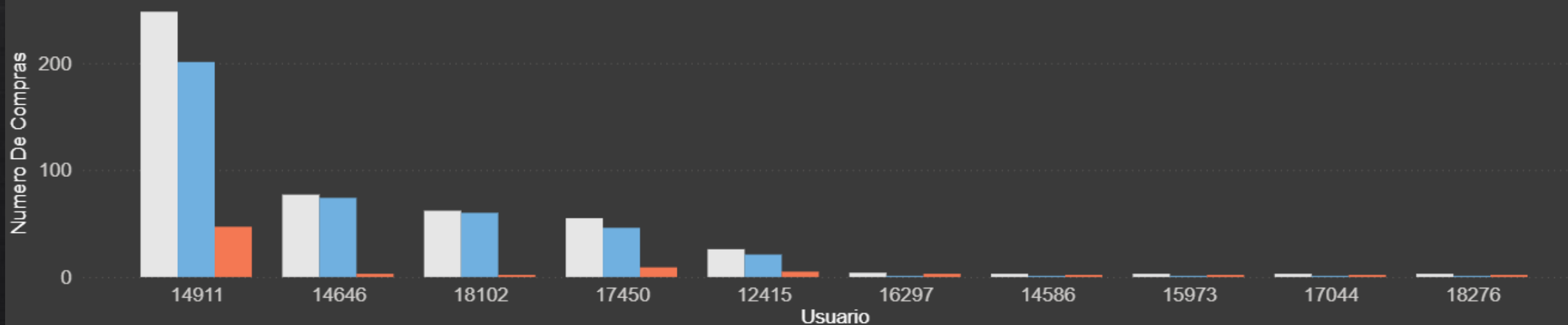
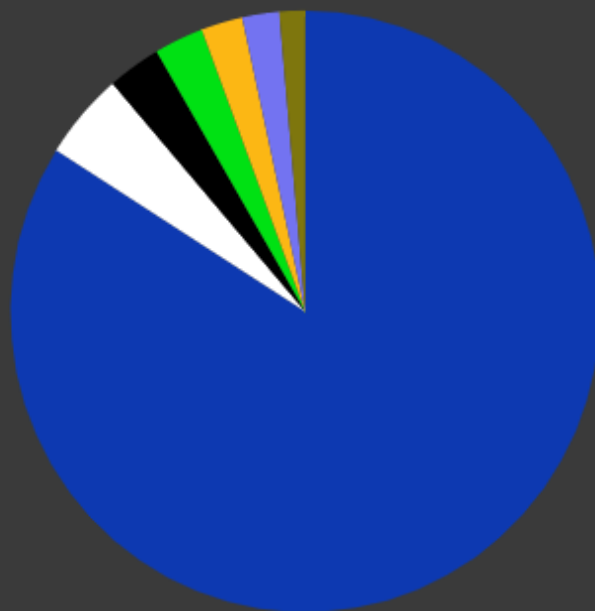# Paises Que Mas Compran

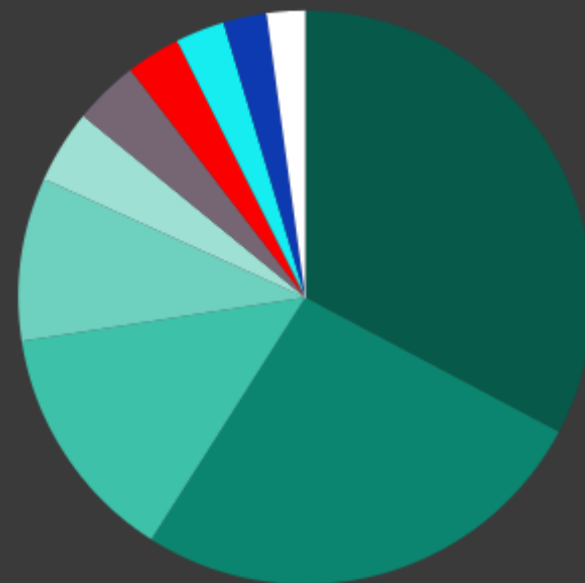Legend: United Kingdom, Otros, Germany, France, Spain, Netherlands, EIRE, Singapore, USA, Hong Kong, Bahrain
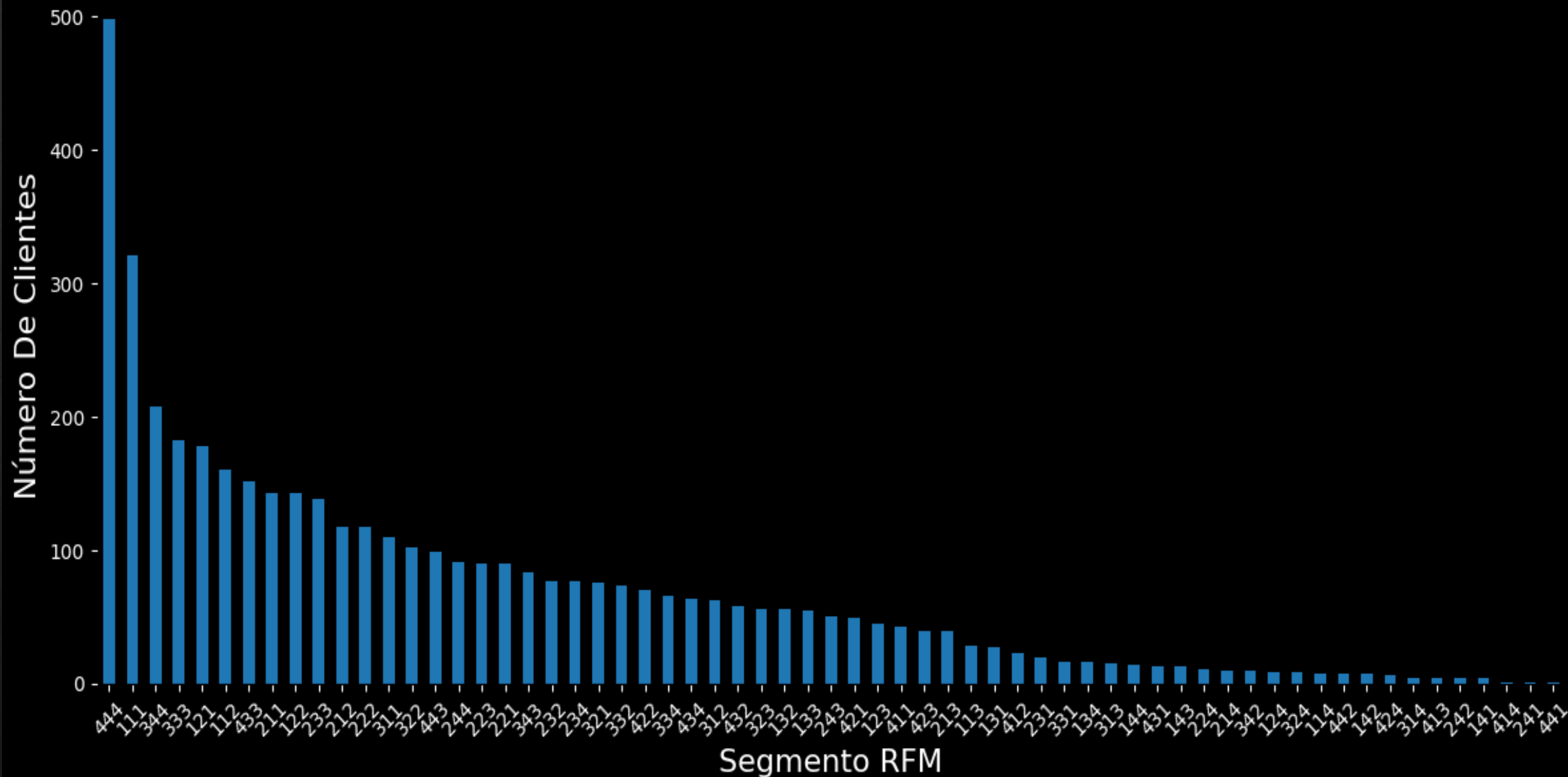
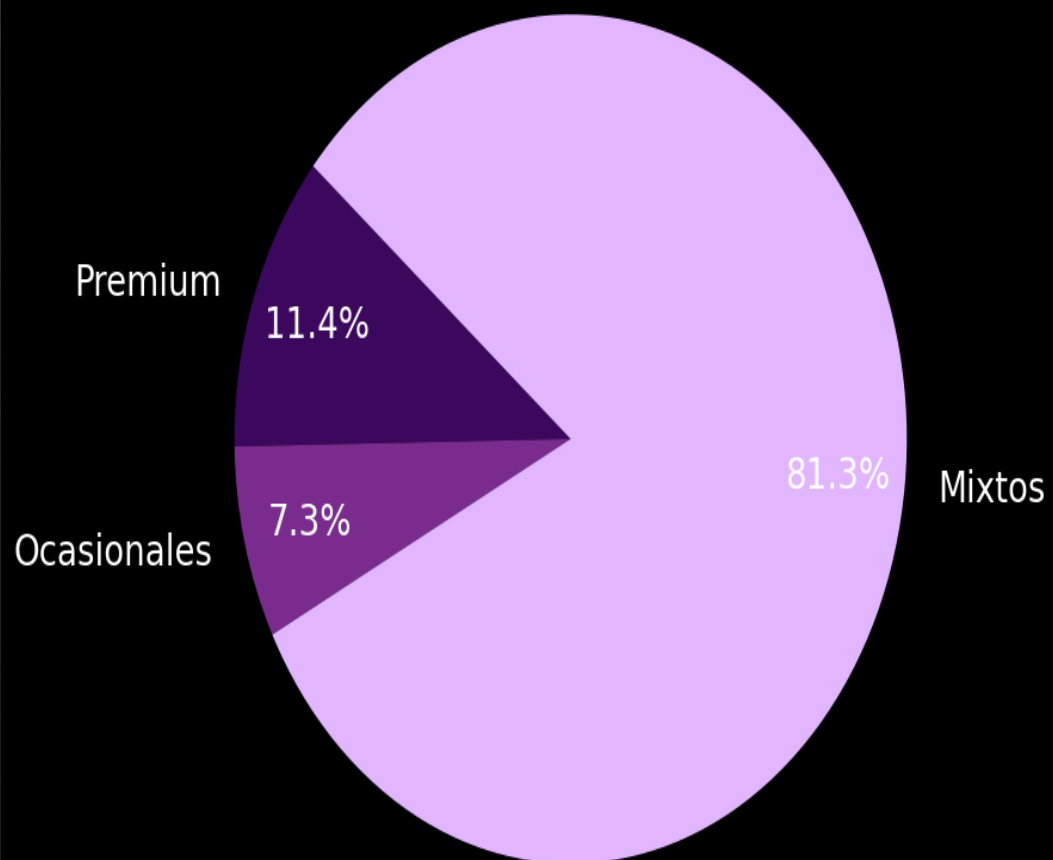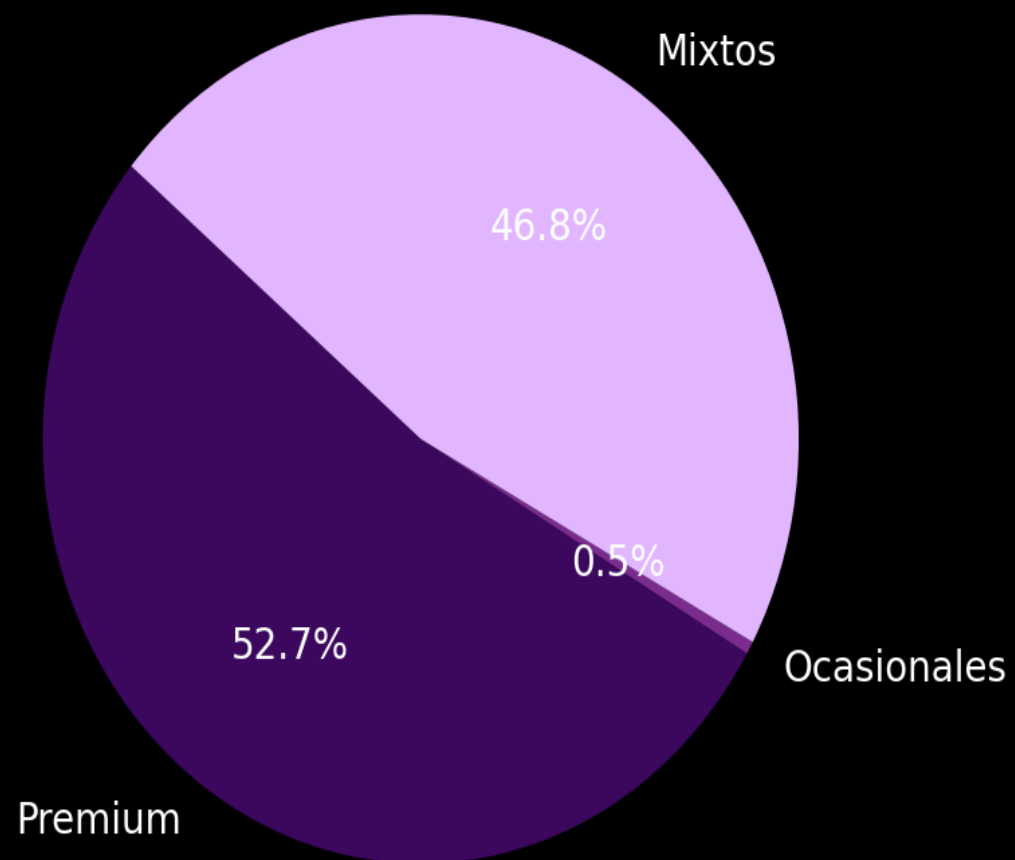Clientes Por Pais
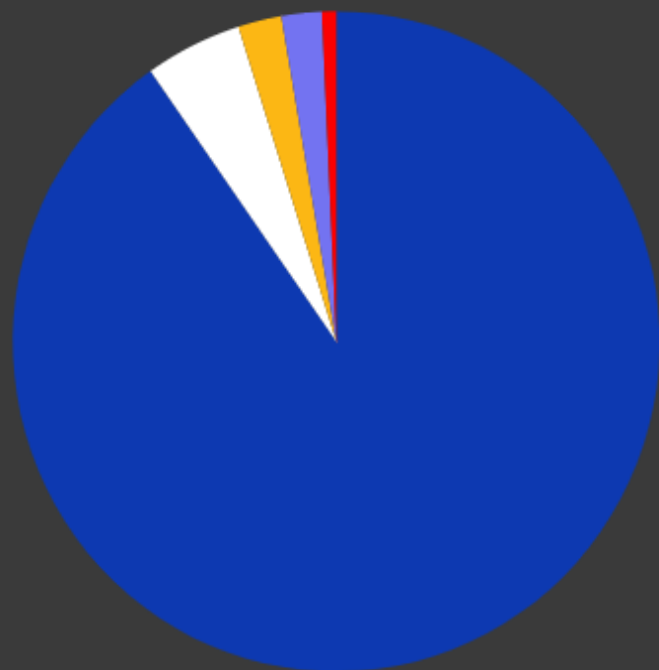
Consumo Neto Por Pais

Devoluciones Por Pais

RFM

Top 200 De Cantidad De Clientes Por Segmento De RFM

Distribución De Clientes Por Segmento RFM

Premium 11.4%

Ocasionales 7.3%

Mixtos 81.3%

Distribución Del Gasto Neto Total Por Segmento

Mixtos 46.8%

Ocasionales 0.5%

Premium 52.7%

Machine Learning

	CustomerID	InvoiceDate	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Score	Segmento
0	12346	2011-01-18 10:17:00	326	2	0.00	1	2	1	121	Mixtos
1	12347	2011-12-07 15:52:00	2	7	4310.00	4	4	4	444	Premium
2	12348	2011-09-25 13:13:00	75	4	1797.24	2	3	4	234	Mixtos
3	12349	2011-11-21 09:51:00	19	1	1757.55	3	1	4	314	Mixtos
4	12350	2011-02-02 16:01:00	310	1	334.40	1	1	2	112	Mixtos
...	...	...	...	...	...	...	...	...	...	...
4366	18280	2011-03-07 09:52:00	278	1	180.60	1	2	1	121	Mixtos
4367	18281	2011-06-12 10:53:00	181	1	80.82	1	2	1	121	Mixtos
4368	18282	2011-12-02 11:43:00	8	3	176.60	4	3	1	431	Mixtos
4369	18283	2011-12-06 12:02:00	4	16	2094.88	4	4	4	444	Premium
4370	18287	2011-10-28 09:29:00	43	3	1837.28	3	3	4	334	Mixtos

Shapiro-Wilk

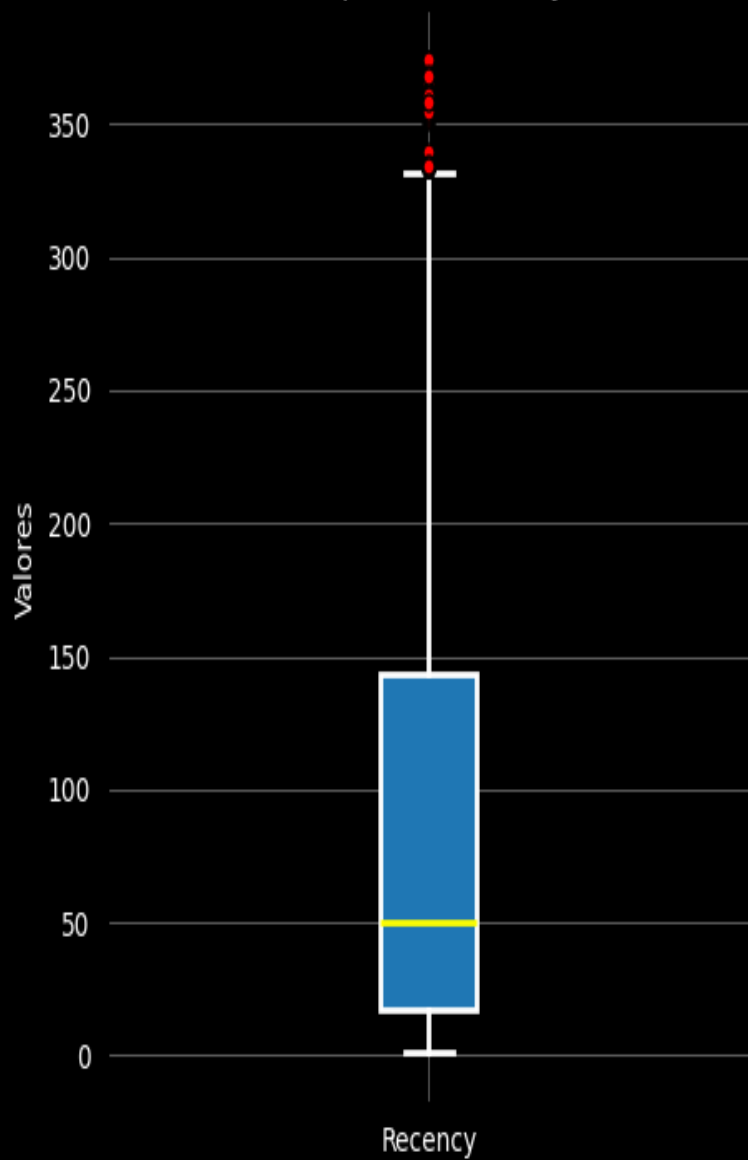Recency: Estadístico=0.8106, p-valor=0.0000 -> No es normal

Frequency: Estadístico=0.3774, p-valor=0.0000 -> No es normal

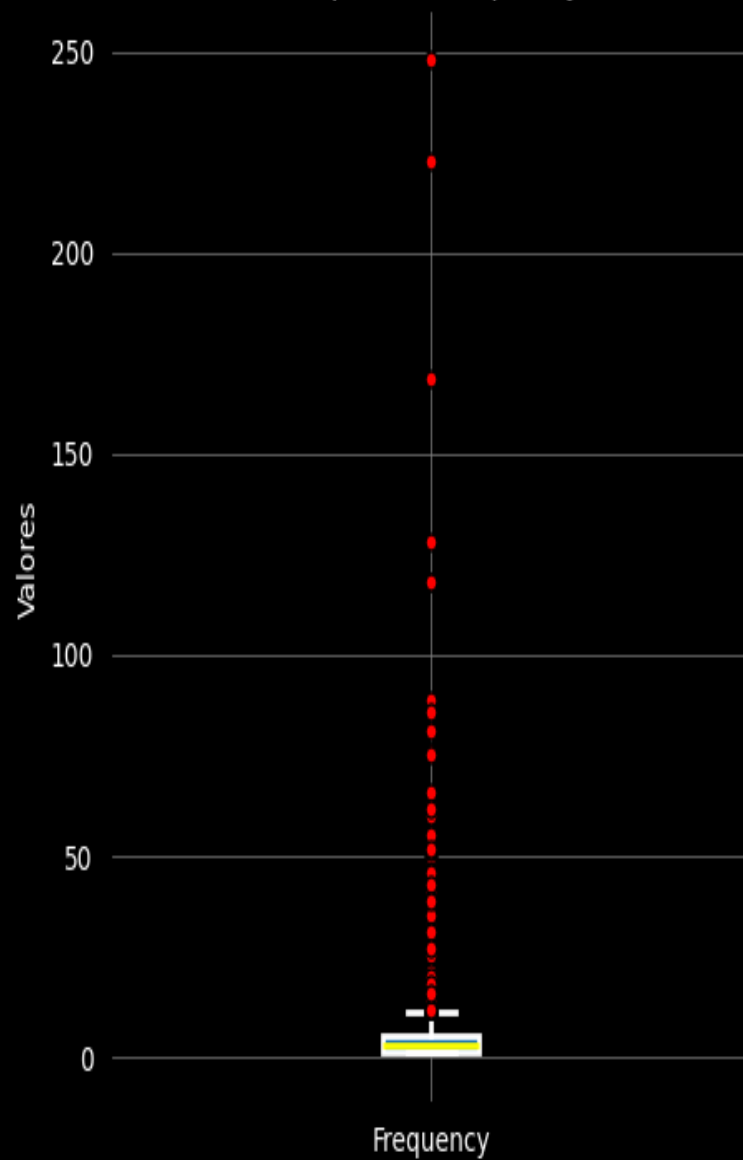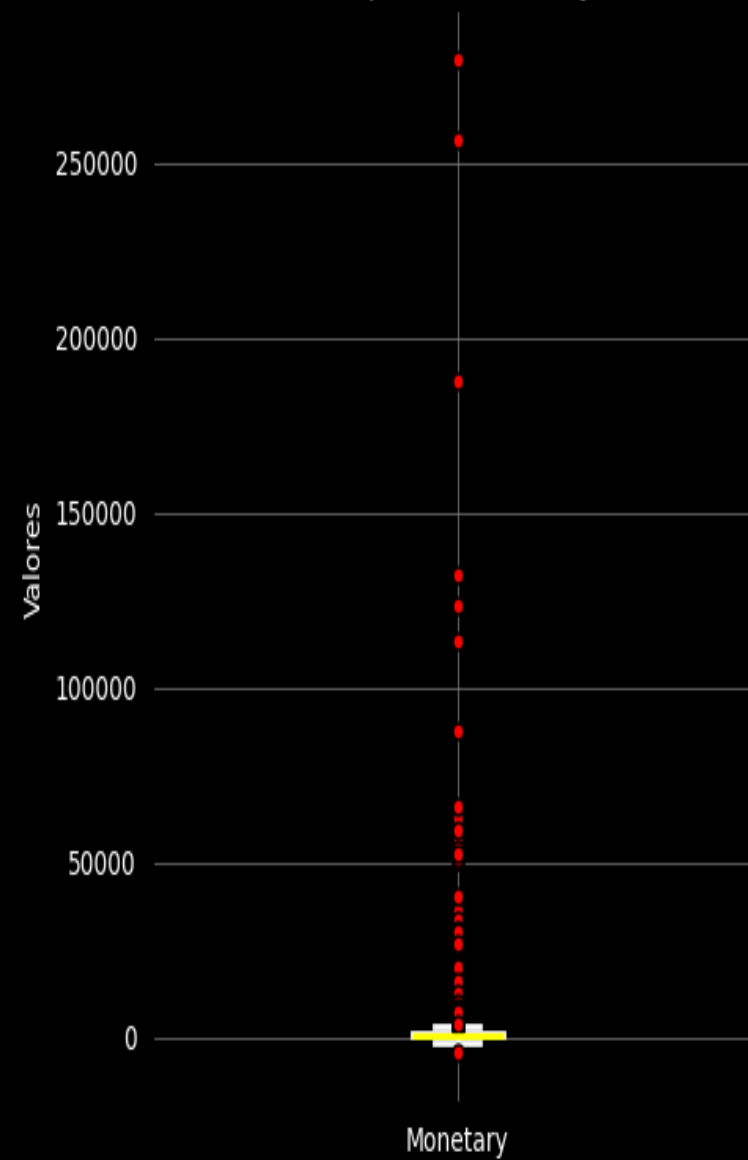Monetary: Estadístico=0.1407, p-valor=0.0000 -> No es normal

Boxplot

Boxplot De Recency

Boxplot De Frequency

Boxplot De Monetary

Random Forest

3.3 Definicion De Las Variables

```python
X = rfm[['Recency','Frequency','Monetary']]
y = rfm['Segmento']
```
[21]   Python

3.4 RobustScaler

```python
scaler = RobustScaler()
X_scaled = scaler.fit_transform(X)
```
[22]   Python

3.5 Segemetacion De Los Datos De Entramiento/Prueva

```python
X_train_scaled, X_test_scaled, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.3, random_state=42)
```
[23]   Python

3.6 Modelo: Random Forest

```python
modelo = RandomForestClassifier(random_state=42, class_weight='balanced')
modelo.fit(X_train_scaled, y_train)
```
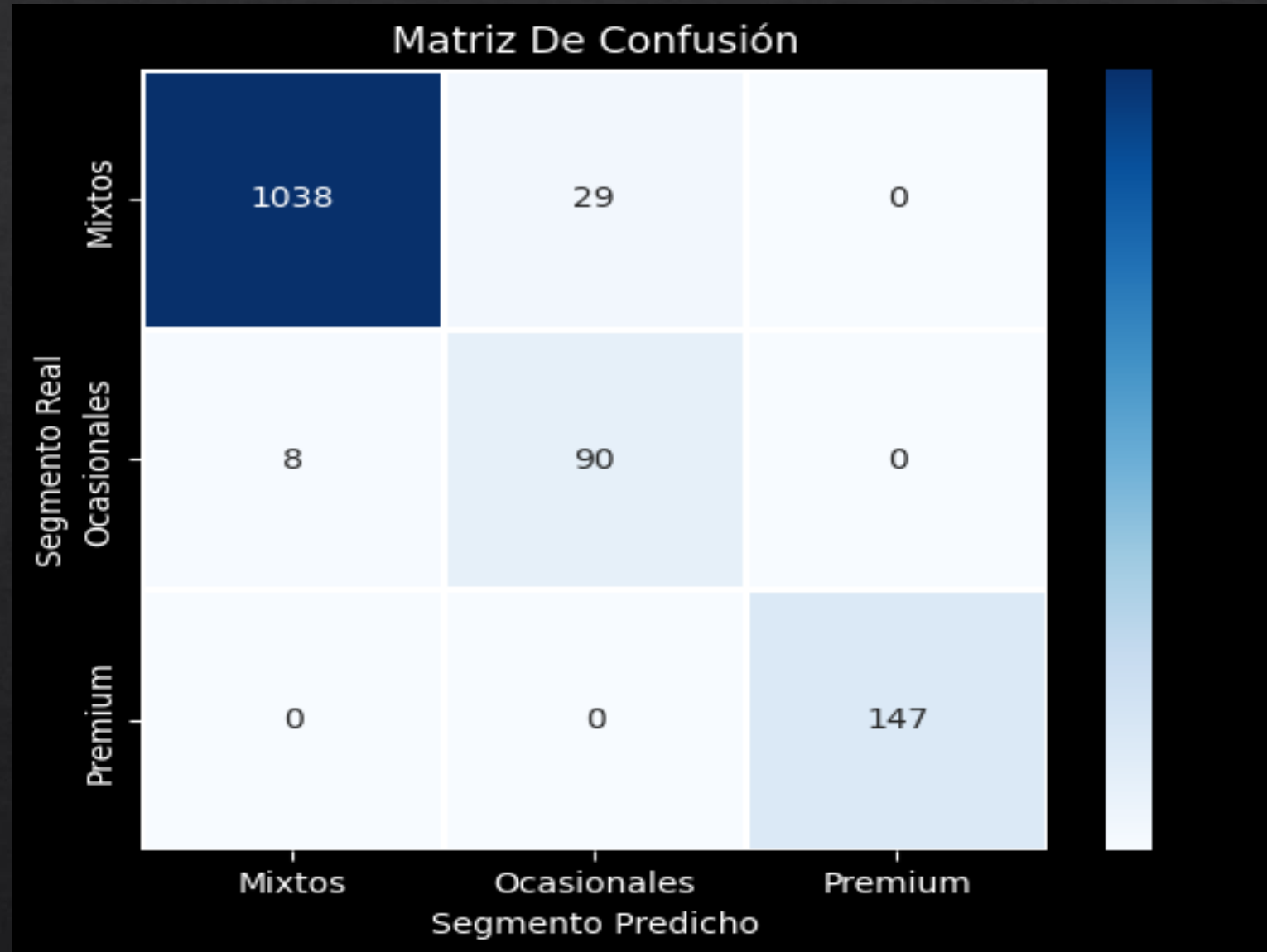[24]   Python

...

```
▼ RandomForestClassifier   ①  ②
▶ Parameters
```

Accuracy = 0.97

Conclusiones

Gracias