

Project : E-commerce Data analysis

TABLE OF CONTENT

Introduction Data Cleaning

Data Wangling

Data Exploratory

INTRODUCTION

This is a transactional dataset which contain all the transaction 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. Which contain informations such as Invoice number, StockCode,Product Descripton, InvoiceDate, Unitprice, CustomerID, Country.

Questions to be answered With this Dataset

DERIVE THE PRODUCT-SALES BY MULTIPLYING THE UNITPRICE AND THE QUANTITY PURCHASED.

AVERAGE UNITPRICE OF PRODUCTS.

TOP 10 COUNTRIES WITH HIGHEST PRODUCT SALES.

WHAT ARE THE 10 MOST POPULAR PRODUCT DESCRIPTIONS.

WHAT ARE THE 10 MOST POPULAR POPULAR STOCKCODE.

WHAT IS THE EARLIEST DATE OF THE INVOICE DATE AND DAY OF THE WEEK

WHAT IS THE EARLIEST DATE OF THE INVOICE DATE AND DAY OF THE WEEK.

WHAT IS THE NUMBERS OF DAYS BETWEEN THE FIRST DAY OF THE DATASET AND THE LAST DATE

```
In [56]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [57]: commerce = pd.read_csv('C:/Users/ADMIN/Desktop/MY DATABASE/E-commerce.csv')
commerce
```

	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	844068 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	84029G RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
...
541904	581587	22613	PACK OF 20 SPACEBOY NAPKINS	12	2011-12-09 12:50:00	0.85	12680.0	France
541905	581587	22899	CHILDRENS APRON DOLLY GIRL	6	2011-12-09 12:50:00	2.10	12680.0	France
541906	581587	23254	CHILDRENS CUTLERY DOLLY GIRL	4	2011-12-09 12:50:00	4.15	12680.0	France
541907	581587	23255	CHILDRENS CUTLERY CIRCUS PARADE	4	2011-12-09 12:50:00	4.15	12680.0	France
541908	581587	22138	BAKING SET 9 PIECE RETROSPOT	3	2011-12-09 12:50:00	4.95	12680.0	France
541909 rows x 8 columns								

```
In [58]: commerce.columns
Out[58]: Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate', 'UnitPrice', 'CustomerID', 'Country'],
dtype='object')
```

```
In [59]: commerce.isna().sum()
Out[59]: InvoiceNo      0
StockCode      0
Description    1454
Quantity       0
InvoiceDate    0
UnitPrice      0
CustomerID    135888
Country        0
dtype: int64
```

```
In [60]: commerce.duplicated().sum()
Out[60]: 5268
```

```
In [61]: commerce.nunique()
Out[61]: InvoiceNo      25990
StockCode      4878
Description     4223
Quantity       722
InvoiceDate    23260
UnitPrice      1638
CustomerID     4272
Country        38
dtype: int64
```

Errors detected in the Dataset

The Dataset contains Missing Values

The Dataset contains duplicate Row

UnitPrice Column needed for the analysis has rows equal to zero, which the Product has no Price.

The Quantity Column needed for the analysis is less than 0

The InvoiceDate was not in datetime

DATA CLEANING

```
In [62]: commerce.dropna(inplace = True)
```

```
In [63]: commerce.drop_duplicates(inplace = True)
```

```
In [64]: commerce.shape
Out[64]: (401604, 8)
```

```
In [65]: commerce[commerce['UnitPrice'] == 0]
Out[65]:
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
	5302	537197	22841 ROUND CAKE TIN VINTAGE GREEN	1	2010-12-05 14:02:00	0.0	12647.0	Germany
	33576	539263	22580 ADVENT CALENDAR GINGHAM SACK	4	2010-12-16 14:36:00	0.0	16560.0	United Kingdom
	40099	539722	22423 REGENCY CAKESTAND 3 TIER	10	2010-12-21 13:45:00	0.0	14911.0	EIRE
	47068	540372	22090 PAPER BUNTING RETROSPOT	24	2011-01-06 16:41:00	0.0	13081.0	United Kingdom
	47070	540372	22553 PLASTERS IN TIN SKULLS	24	2011-01-06 16:41:00	0.0	13081.0	United Kingdom
	56674	541109	22168 ORGANISER WOOD ANTIQUE WHITE	1	2011-01-13 15:10:00	0.0	15107.0	United Kingdom
	86789	543599	845358 FAIRY CAKES NOTEBOOK A6 SIZE	16	2011-02-10 13:08:00	0.0	17560.0	United Kingdom
	130188	547417	22052 CERAMIC BOWL WITH LOVE HEART DESIGN	36	2011-03-23 10:25:00	0.0	13239.0	United Kingdom
	138453	548318	22055 MINI CAKE STAND HANGING STRAWBERRY	5	2011-03-30 12:45:00	0.0	13113.0	United Kingdom
	145208	548971	22162 HEART GARLAND RUSTIC PADDED	2	2011-04-04 14:42:00	0.0	14410.0	United Kingdom
	157042	550188	22636 CHILDS BREAKFAST SET CIRCUS PARADE	1	2011-04-14 18:57:00	0.0	12457.0	Switzerland
	187613	553000	47566 PARTY BUNTING	4	2011-05-12 15:21:00	0.0	17667.0	United Kingdom
	198383	554037	22619 SET OF 6 SOLDIER SKITTLES	80	2011-05-20 14:13:00	0.0	12415.0	Australia
	278324	561284	22167 OVAL WALL MIRROR DIAMANTE	1	2011-07-26 12:24:00	0.0	16818.0	United Kingdom
	282912	561609	22960 JAM MAKING SET WITH JARS	11	2011-07-26 17:09:00	0.0	12507.0	Spain
	285657	561916	M Manual	1	2011-08-01 11:44:00	0.0	15581.0	United Kingdom
	298054	562973	23157 SET OF 6 NATIVITY MAGNETS	240	2011-08-11 11:42:00	0.0	14911.0	EIRE
	314745	564651	23270 SET OF 2 CERAMIC PAINTED HEARTS	96	2011-08-26 14:19:00	0.0	14646.0	Netherlands
	314746	564651	23268 SET OF 2 CERAMIC CHRISTMAS REINDEER	192	2011-08-26 14:19:00	0.0	14646.0	Netherlands
	314747	564651	22955 36 FOIL STAR CAKE CASES	144	2011-08-26 14:19:00	0.0	14646.0	Netherlands
	314748	564651	21786 POLKADOT RAIN HAT	144	2011-08-26 14:19:00	0.0	14646.0	Netherlands
	358655	568158	PADS TO MATCH ALL CUSHIONS	1	2011-09-25 12:22:00	0.0	16133.0	United Kingdom
	361825	568384	M Manual	1	2011-09-27 09:46:00	0.0	12749.0	United Kingdom
	378913	569716	22778 GLASS CLOCHE SMALL	2	2011-10-06 08:17:00	0.0	15804.0	United Kingdom
	395529	571035	M Manual	1	2011-10-13 12:50:00	0.0	12446.0	RSA
	428404	572893	21208 PASTEL COLOUR HONEYCOMB FAN	5	2011-10-26 14:36:00	0.0	18059.0	United Kingdom
	436428	574138	23234 BISCUIT TIN VINTAGE CHRISTMAS	216	2011-11-03 11:26:00	0.0	12415.0	Australia
	436597	574175	22065 CHRISTMAS PUDDING TRINKET POT	12	2011-11-03 11:47:00	0.0	14110.0	United Kingdom
	436961	574252	M Manual	1	2011-11-03 13:24:00	0.0	12437.0	France
	439361	574469	22385 JUMBO BAG SPACEBOY DESIGN	12	2011-11-04 11:55:00	0.0	12431.0	Australia
	446125	574879	22625 RED KITCHEN SCALES	2	2011-11-07 13:22:00	0.0	13014.0	United Kingdom
	446793	574920	22899 CHILDRENS APRON DOLLY GIRL	1	2011-11-07 16:34:00	0.0	13985.0	United Kingdom
	446794	574920	23480 MINI LIGHTS WOODLAND MUSHROOMS	1	2011-11-07 16:34:00	0.0	13985.0	United Kingdom
	454463	575579	22437 SET OF 9 BLACK SKULL BALLOONS	20	2011-11-10 11:49:00	0.0	13081.0	United Kingdom
	454464	575579	22089 PAPER BUNTING VINTAGE PAISLEY	24	2011-11-10 11:49:00	0.0	13081.0	United Kingdom
	479079	577129	22464 HANGING METAL HEART LANTERN	4	2011-11-17 19:52:00	0.0	15602.0	United Kingdom
	479546	577168	M Manual	1	2011-11-18 10:42:00	0.0	12603.0	Germany
	480649	577314	23407 SET OF 2 TRAYS HOME SWEET HOME	2	2011-11-18 13:23:00	0.0	12444.0	Norway
	485985	577696	M Manual	1	2011-11-21 11:57:00	0.0	16406.0	United Kingdom
	502122	578841	84826 ASSTD DESIGN 3D PAPER STICKERS	12540	2011-11-25 15:57:00	0.0	12256.0	United Kingdom

Drop Rows where Unit price of Products is not stated

```
In [66]: NounitPrice = commerce['UnitPrice'] == 0
commerce.drop(index = commerce[NounitPrice].index, inplace = True)
```

```
In [67]: commerce[commerce['UnitPrice'] == 0]
Out[67]: InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
```

Drop Rows where Quantity of Products less than Zero

```
In [68]: commerce[commerce['Quantity'] < 0 ]
Out[68]:
```

154	C536383	35004C	SET OF 3 COLOURED FLYING DUCKS	-1	2010-12-01 09:49:00	4.65	15311.0	United Kingdom
235	C536391	22556	PLASTERS IN TIN CIRCUS PARADE	-12	2010-12-01 10:24:00	1.65	17548.0	United Kingdom
236	C536391	21984	PACK OF 12 PINK PAISLEY TISSUES	-24	2010-12-01 10:24:00	0.29	17548.0	United Kingdom
237	C536391	21983	PACK OF 12 BLUE PAISLEY TISSUES	-24	2010-12-01 10:24:00	0.29	17548.0	United Kingdom
...
540449	C581490	23144	ZINC T-LIGHT HOLDER STARS SMALL	-11	2011-12-09 09:57:00	0.83	14397.0	United Kingdom
541541	C581499	M	Manual	-1	2011-12-09 10:28:00	224.69	15488.0	United Kingdom
541715	C581568	21258	VICTORIAN SEWING BOX LARGE	-5	2011-12-09 11:57:00	10.95	15311.0	United Kingdom
541716	C581569	84978	HANGING HEART JAR T-LIGHT HOLDER	-1	2011-12-09 11:58:00	1.25	17315.0	United Kingdom
541717	C581569	20979	36 PENCILS TUBE RED RETROSPOT	-5	2011-12-09 11:58:00	1.25	17315.0	United Kingdom

8872 rows x 8 columns

```
QuantityLessZero = commerce["Quantity"] < 0
commerce.drop(index = commerce[QuantityLessZero].index, inplace = True)
```

```
In [69]: QuantityLessZero = commerce['Quantity'] < 0
commerce.drop(index = commerce[QuantityLessZero].index, inplace = True)
```

```
In [70]: commerce[commerce['Quantity'] < 0 ]
Out[70]: InvoiceNo StockCode Description Quantity InvoiceDate UnitPrice CustomerID Country
```

DATA WANGLING

```
In [71]: commerce.describe()
Out[71]:
```

	Quantity	UnitPrice	CustomerID
count	392692.000000	392692.000000	392692.000000
mean	13.119702	3.125914	15287.843865
std	180.492632	22.241836	1713.539549
min	1.000000	0.001000	12346.000000
25%	2.000000	1.250000	13955.000000
50%	6.000000	1.950000	15150.000000
75%	12.000000	3.750000	16791.000000
max	80995.000000	8142.750000	18287.000000

DERIVE THE PRODUCT-SALES BY MULTIPLYING THE UNITPRICE AND THE QUANTITY PURCHASED

```
In [72]: commerce['ProductSales'] = commerce['UnitPrice'] * commerce['Quantity']
commerce['ProductSales']
```

```
Out[72]:
0      15.30
1      26.34
2      22.80
3      29.34
4      26.34
...
```

AVERAGE UNITPRICE OF PRODUCTS

```
In [73]: product_grouped = commerce.groupby(['Description'])['UnitPrice'].mean()
product_grouped.sort_values(ascending = False)[:10]
```

Description	UnitPrice
DOTCOM POSTAGE	744.147500
PICNIC BASKET WICKER 60 PIECES	649.500000
Manual	178.486129
RUSTIC SEVENTEEN DRAWER SIDEBBOARD	158.876823
REGENCY MIRROR WITH SHUTTERS	156.428571
VINTAGE BLUE KITCHEN CABINET	146.750000
VINTAGE RED KITCHEN CABINET	143.423603
CHEST NATURAL WOOD 20 DRAWERS	118.876923
LOVE SEAT ANTIQUE WHITE METAL	114.624398
VINTAGE POST OFFICE CABINET	66.360000
Name: UnitPrice, dtype: float64	

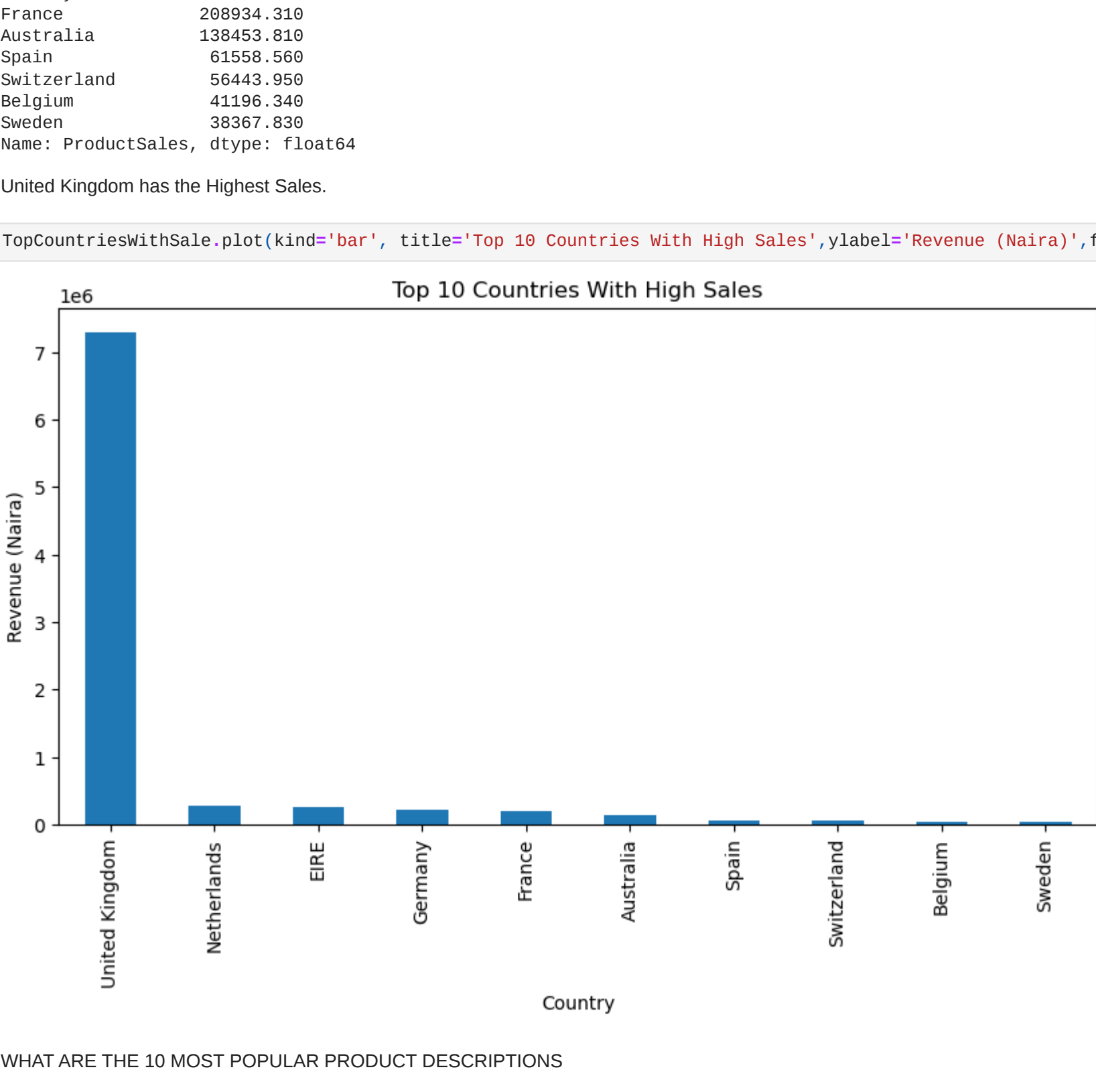
TOP 10 COUNTRIES WITH HIGHEST PRODUCT SALES

```
In [74]: TopCountriesWithSale = commerce.groupby(['Country'])['ProductSales'].sum()
TopCountriesWithSale = TopCountriesWithSale.sort_values(ascending = False)[:10]
TopCountriesWithSale
```

Country	ProductSales
United Kingdom	7285024.644
Netherlands	285446.340
EIRE	265262.460
Germany	228678.480
France	208934.310
Australia	138453.810
Spain	61159.560
Switzerland	56443.950
Belgium	41196.340
Sweden	38367.830
Name: ProductSales, dtype: float64	

United Kingdom has the Highest Sales.

```
In [75]: TopCountriesWithSale.plot(kind='bar', title='Top 10 Countries With High Sales', ylabel='Revenue (Naira)', figsize=(10,5));
```



WHAT ARE THE 10 MOST POPULAR PRODUCT DESCRIPTIONS

```
In [76]: PopularDescription = commerce['Description'].value_counts().sort_values(ascending = False)[:10]
PopularDescription
```

Description	count
WHITE HANGING HEART T-LIGHT HOLDER	2618
REGENCY CAKESTAND 3 TIER	1713
JUMBO BAG RED RETROSPOT	1615
ASSORTED COLOUR BIRD ORNAMENT	1395
PARTY BUNTING	1389
LUNCH BAG RED RETROSPOT	1383
SET OF 3 CAKE TINS PANTRY DESIGN	1152
POSTAGE	1099
LUNCH BAG BLACK SKULL.	1078
PACK OF 72 RETROSPOT CAKE CASES	1058
Name: Description, dtype: int64	

WHITE HANGING HEART T-LIGHT HOLDER is the most popular product Description with 2016 Occurrence

WHAT ARE THE 10 MOST POPULAR POPULAR STOCKCODE

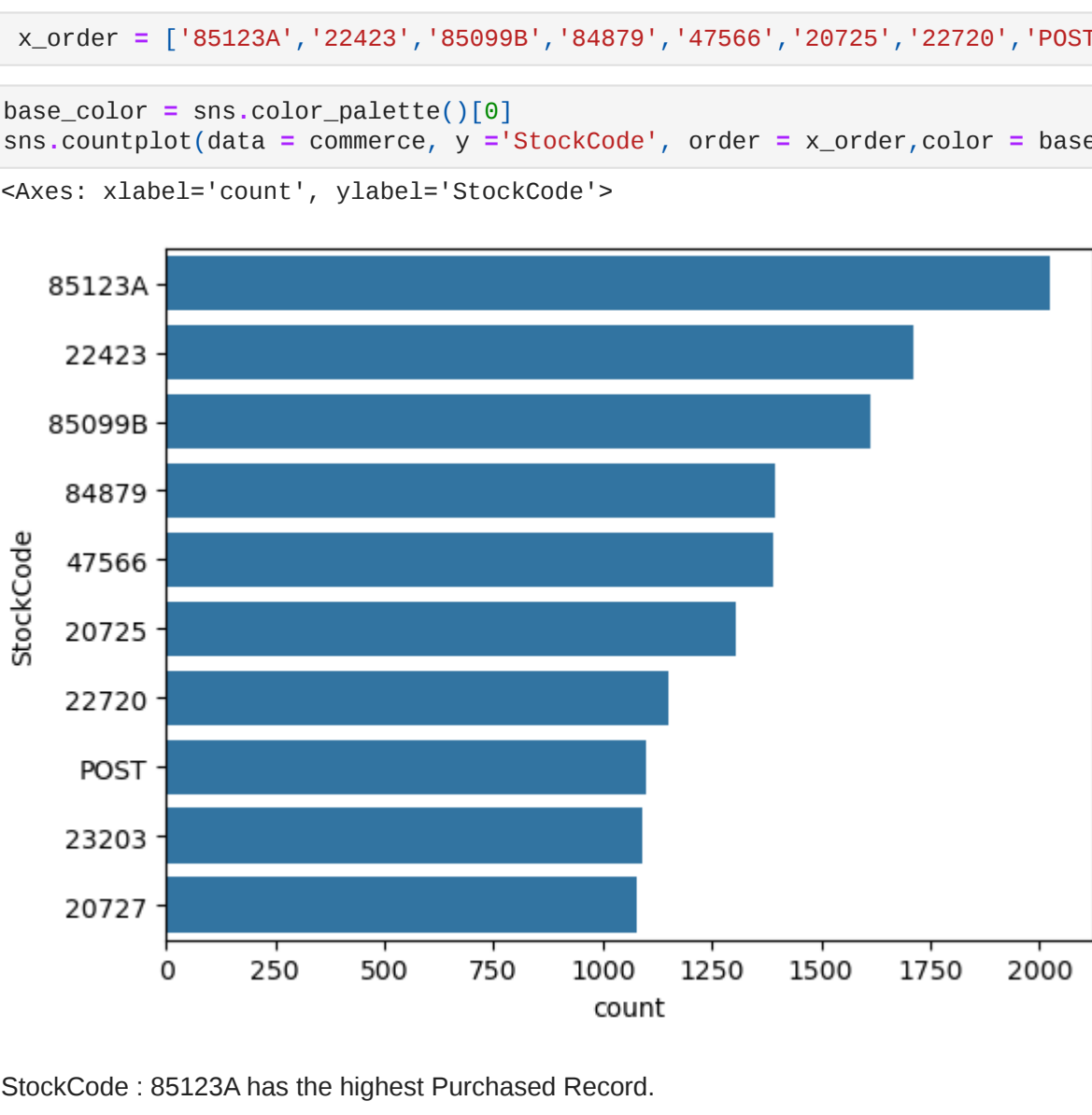
```
In [85]: TopStores = commerce['StockCode'].value_counts().sort_values(ascending = False)[:10]
TopStores
```

StockCode	count
85123A	2623
22423	1713
850998	1615
84879	1395
47566	1389
20725	1384
22728	1152
P08T	1099
23203	1092
20727	1078
Name: StockCode, dtype: int64	

```
In [90]: x_order = ['85123A', '22423', '850998', '84879', '47566', '20725', '22728', 'P08T', '23203', '20727']
```

```
In [95]: base_color = sns.color_palette()[0]
sns.countplot(data = commerce, y = 'StockCode', order = x_order, color = base_color)
```

```
Out[95]: <Axes: xlabel='count', ylabel='StockCode'>
```



StockCode : 85123A has the highest Purchased Record.

```
In [78]: commerce['InvoiceDate'] = pd.to_datetime(commerce['InvoiceDate'])
commerce['DayOfWeek'] = commerce['InvoiceDate'].dt.day_name()
```

Convert to Datetime

```
In [79]: commerce['DayOfWeek']
Out[79]:
0      Wednesday
1      Wednesday
2      Wednesday
3      Wednesday
4      ...
541984    Friday
541985    Friday
541986    Friday
541987    Friday
54198
```