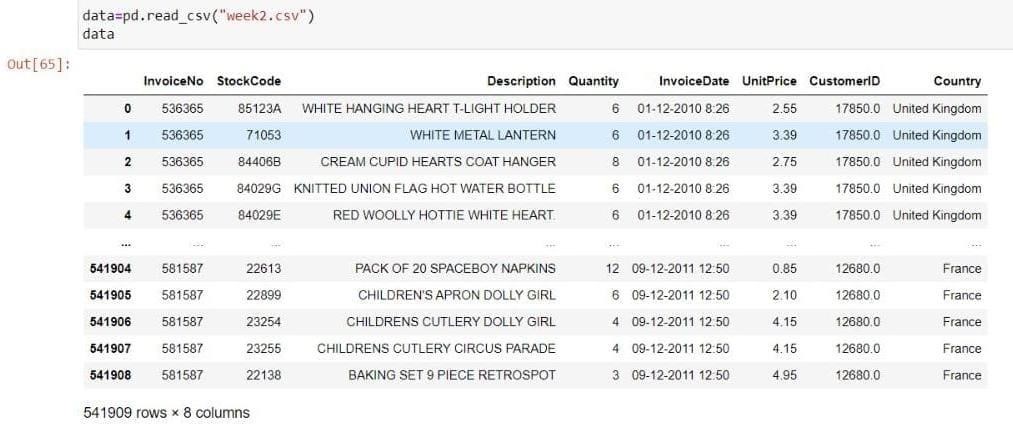
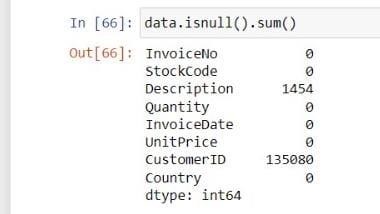
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Data**  **visua**  **lization**      ***Online Retail Data Pre-processing***  ***ASSIGNMENT******2***  *By- Team 3*  *Team Lead-P.Rohith*    **Suggestion and finding**    The analysis of the Preprocessing Online Retail Transaction data using EDA , data analytics software by data visualization methods  A series of sheet are developed to explore the impact of different factors on profitability  1. The data represents a transactional data set with over 500 000 rows 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. we found following things   * 1. By analyzing the data we got to know that United Kingdom has highest sales whereas Saudi Arabia has least sales compared to a total of 38 countries.   2. While considering the invoice date compared to 13 months of data from dec-2010 to dec-2011, November 2011 has highest sales and December 2011 has least sales.   3. By observing description or item name of the product there are total of 3885 items sold in which white hanging heart T-light holder has highest sales   4. By observing the column of Unit Price Manual has the highest unit price of   155507.920 and pads to Manhhall cushions has the least unit price of 0.003.  e. There are a total of 541909 rows of data. After removing the rows containing null values the total rows came out to be 406829.   |  |  |  | | --- | --- | --- | | **Category** | **Description** | **Country** | | **Sub category** | WHITE HANGING HEART T-LIGHT  HOLDER  WHITE METAL LANTERN  CREAM CUPID HEARTS COAT HANGER  KNITTED UNION FLAG HOT WATER  BOTTLE  RED WOOLLY HOTTIE WHITE HEART.  SET 7 BABUSHKA NESTING BOXES  GLASS STAR FROSTED T-LIGHT HOLDER  There are total of 3885 types of items sold | UNITED KINDOM  FRANCE  GERMANY  AUSTREILA  SAUDI ARABIA  NORWAY  There total of 38 countries | | |
| **Data**  **visua**  **lization**       |  |  |  |  | | --- | --- | --- | --- | | Findings from EDA | **JAN –**  white hanging heart T-light holder  **FEB –** set of 3 cake tins pantry  **MAR -**  regency cakestand 3 tries  **APR -**  regency cakestand 3 tries  **MAY -**  party bunting  **JUN -**  party bunting  **JUL -**  party bunting  **AUG -**  Jumbo by reb retospot  **SEP -**  hot water bottle keep clame  **OCT -**  paper chain kit 50’s chirstmas  **NOV –** rabbit night light  **DEC -** rabbit night light | On analyzing the data we found that United Kingdom has 75% of total sales.  Whereas Czech Republic, Saudi Arebia has least sales. | | | **Solution**  **/efficiency** | **1.By observing above data there are exactly 10625 cancellation orders and discount, Rotating Silver Angels has highest cancellation so the board must concentrate on these two products (removing the product or improving the quality check)**  **2.UK has highest sales concentration so board need to concentrate on reducing customer concentration.**  **3.Company has majorly growing upon seasonal products so must improve the sales of seasonal products.**  **4.Company need to take a strong decision on quitting from least sales countries (Czech Republic, Saudi Arebia). As their spending is very much greater than their revenue.** | |  |   ---------------0----------------- |

The data represents a transactional data set with over 500 000 rows 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.



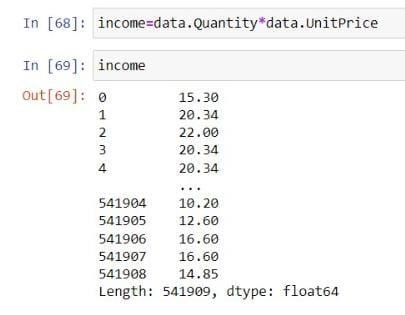
By the above picture we can observe that there are 541909 transactions and there are 8 features given in the dataset They are - 1. InvoiceNo

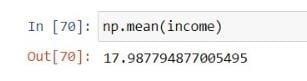
1. StockCode
2. Description
3. Quantity
4. Invoice Date
5. UnitPrice
6. CustomerID
7. Country



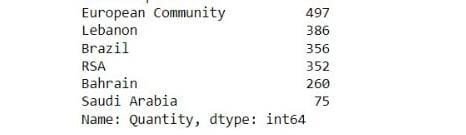
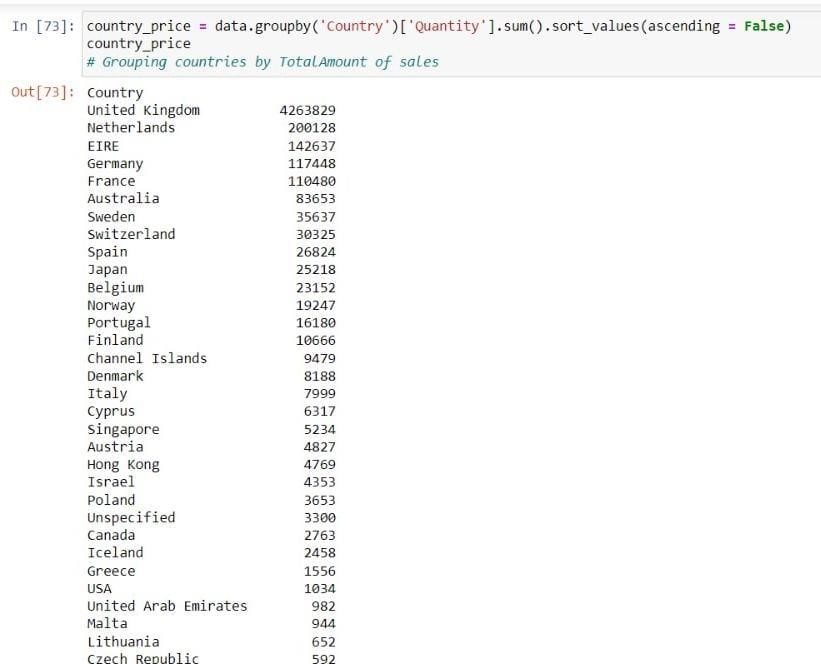
By this picture we observe that there are 1454,135080 null vales in Description ,CustomerID features respectively.

In the given dataset there is no income feature so to get income feature we should multiply Quantity and UnitPrice of each transaction



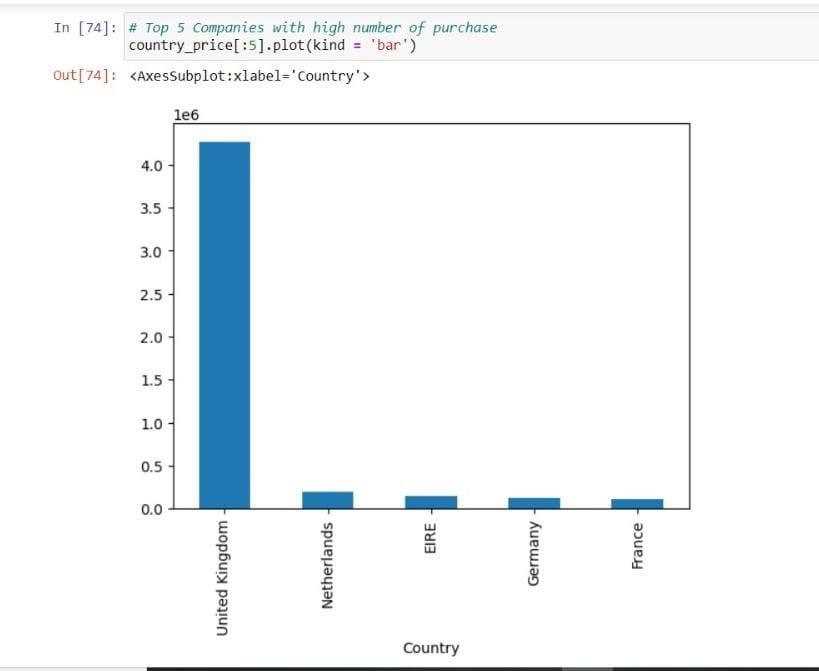


By this picture we can say the average income for each transaction is around 18.



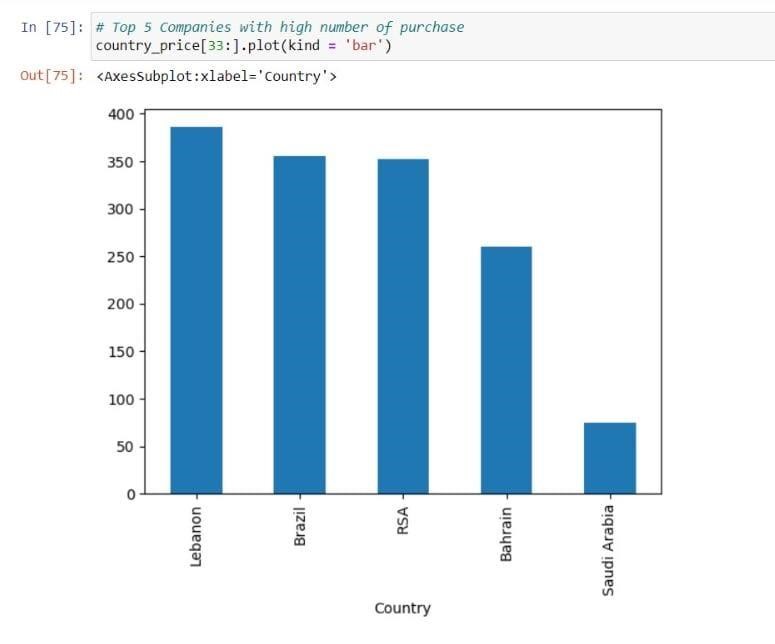
By the above picture we can observe this company is more into online retail across 38 countries shown above

We can also observe the most transactions are done with U.K and least with Saudi Arabia.



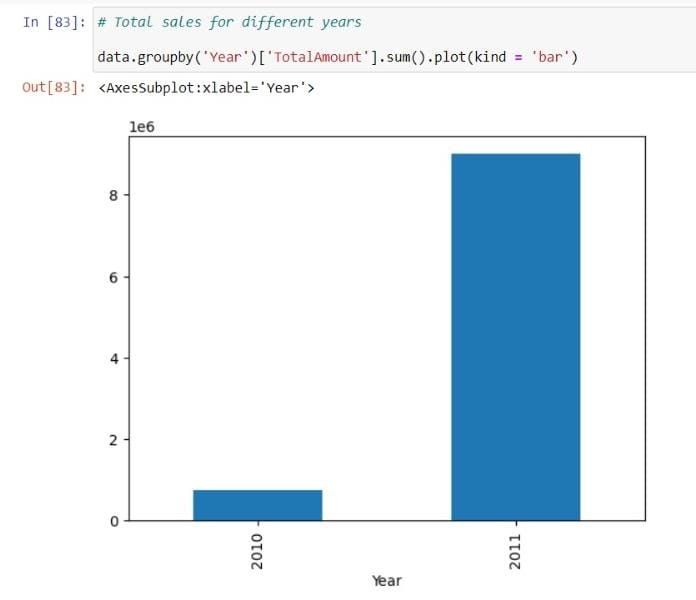
This bar graph shows the top 5 most number of transactions across countries from company.

It concludes that U.K has most transactions with the company.



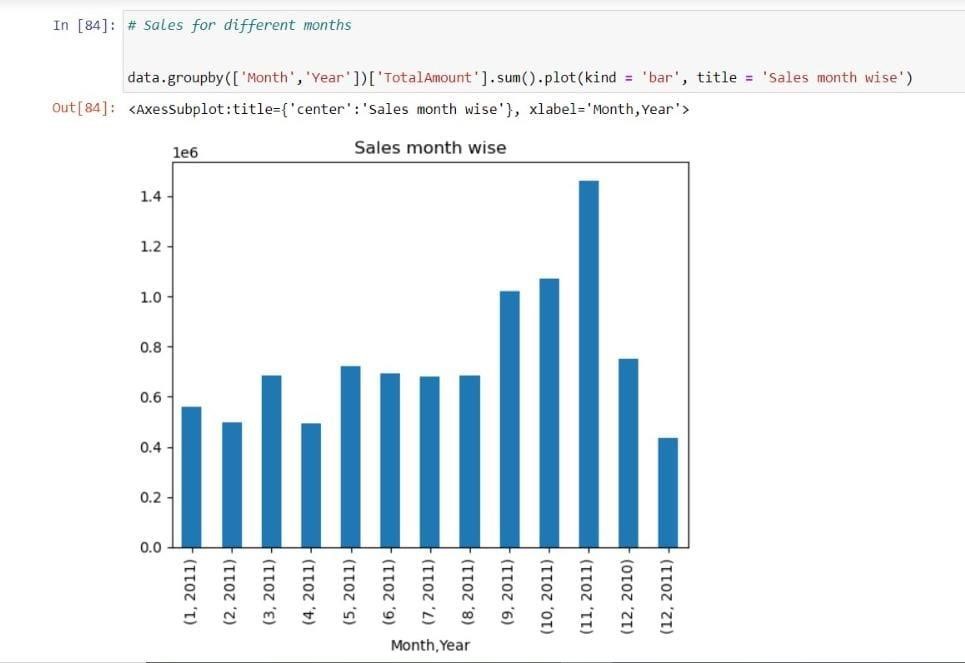
This bar graph shows the top 5 least number of transactions across countries from company.

It concludes that Saudi Arabia has least transactions with the company.



By the above graph says about total sales in different years.

we can say that most number of transactions are done in 2011 year.



In the above bar graph gives sales for different month

it says that the sales by the company are started on December 2010 and data is available upto December 2011.

Most sold product of each month are

**JAN –**  white hanging heart T-light holder

**FEB –** set of 3 cake tins pantry

**MAR -**  regency cakestand 3 tries

**APR -**  regency cakestand 3 tries

**MAY -**  party bunting

**JUN -**  party bunting

**JUL -**  party bunting

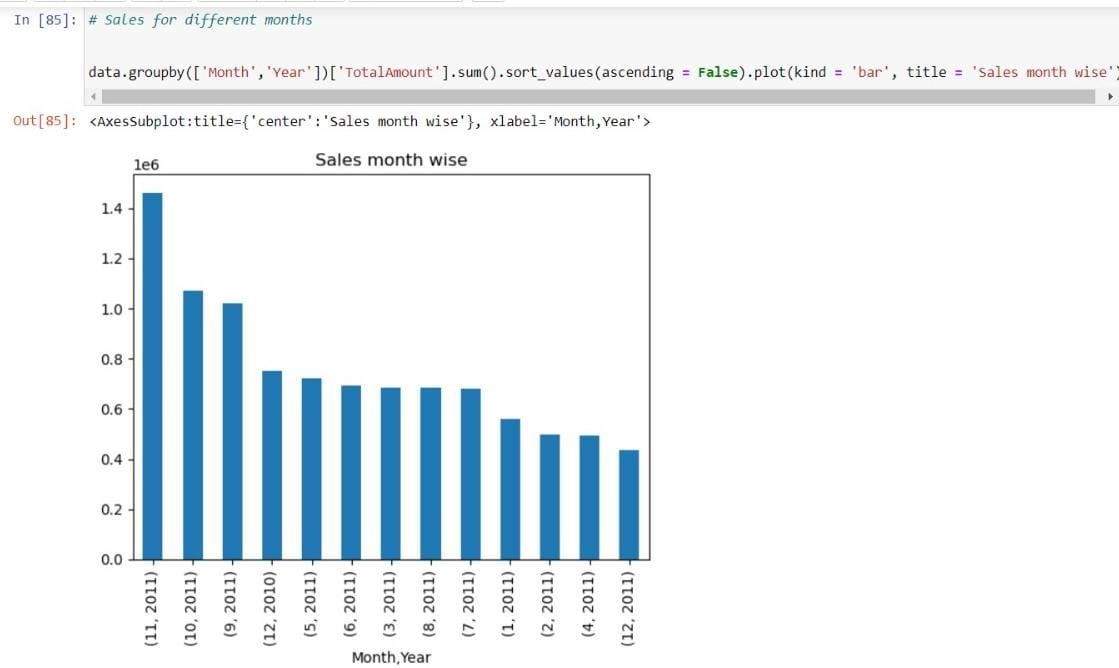
**AUG -**  Jumbo by reb retospot

**SEP -**  hot water bottle keep clame

**OCT -**  paper chain kit 50’s chirstmas

**NOV –** rabbit night light

**DEC -** rabbit night light

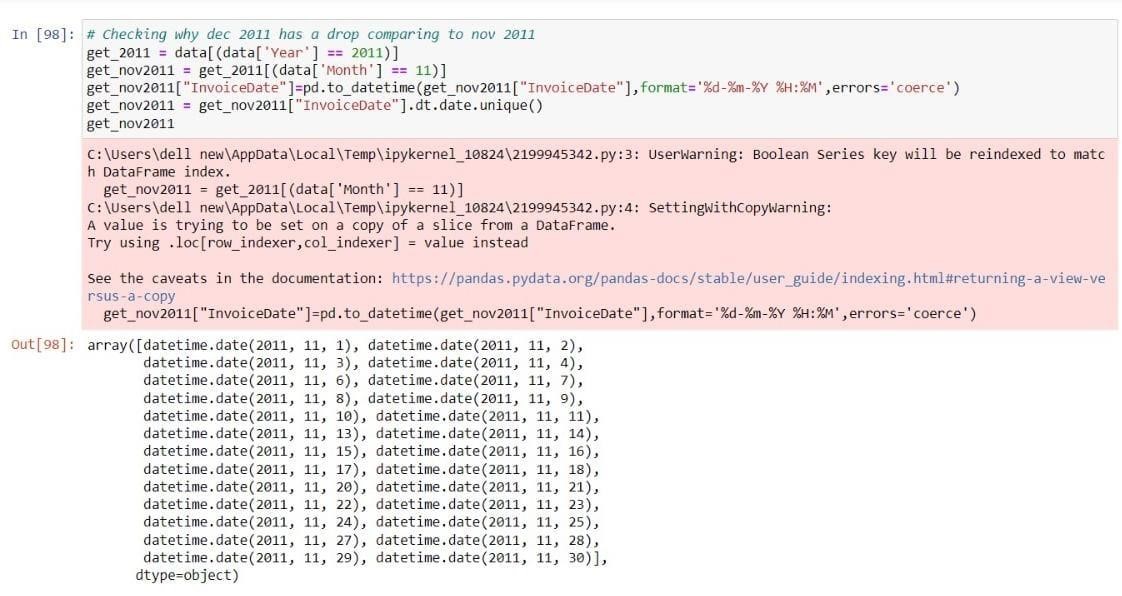


By the graph we can observe that the sales are high in the September 2011 to November 2011 and there is a drop in sales in December 2011.

So company should focus onother 8 months to increase there sales.

The starting month i.e December 2011 has higher sales than the ending month i.e December 2011.

By the graph we can observe that the sales are high in the September 2011 to November 2011 and there is a drop in sales in December 2011.



By above picture we can say that on 26 different dates in December 2011 the transactions are done.



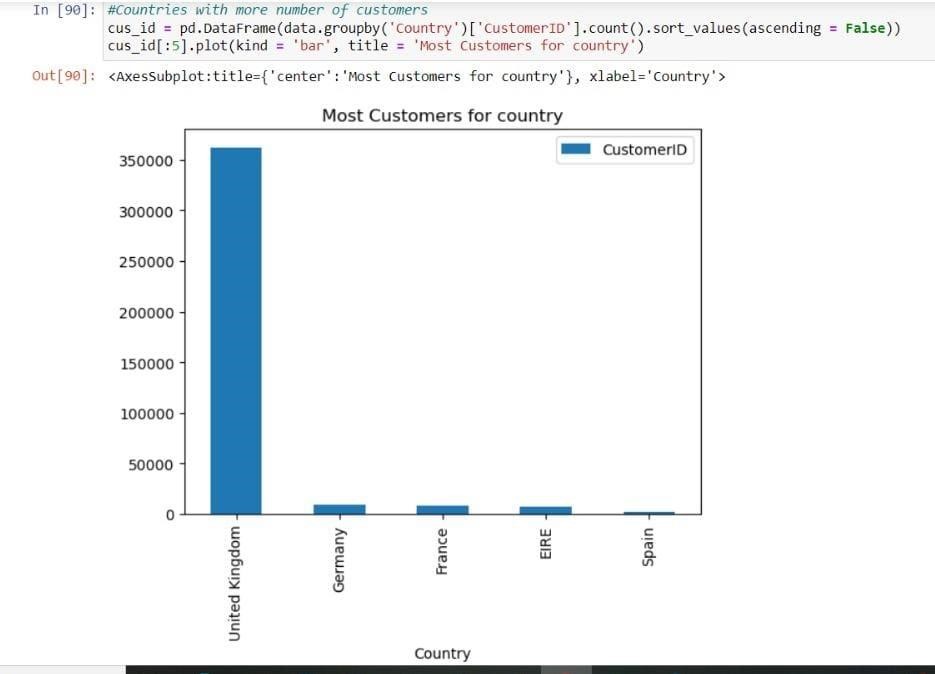
By above picture we can say that on 8 different dates in December 2011 the transactions are done.

By comparing both November and December in 2011 the number of uniue date transactions are very low in December 2011 . so automatically the sales in December 2011 are decreased.



The above table is dataset without null values.

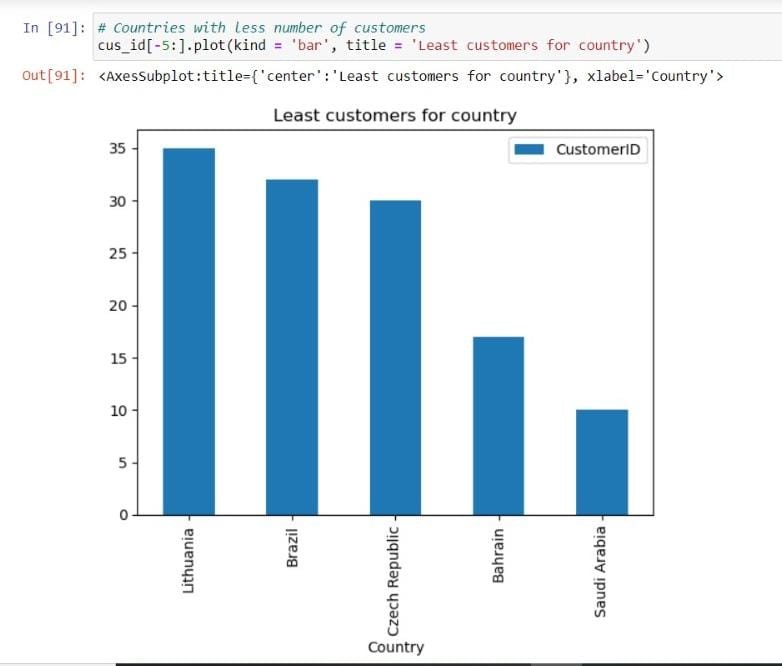
There are total 406829 rows are there in above table. Before removing null values there are 541909 rows so we can conclude there are approximately 14000 null columns.



This bar graph shows the top 5 most number of transactions across countries from company after removing null values.

It concludes that U.K has most transactions with the company.

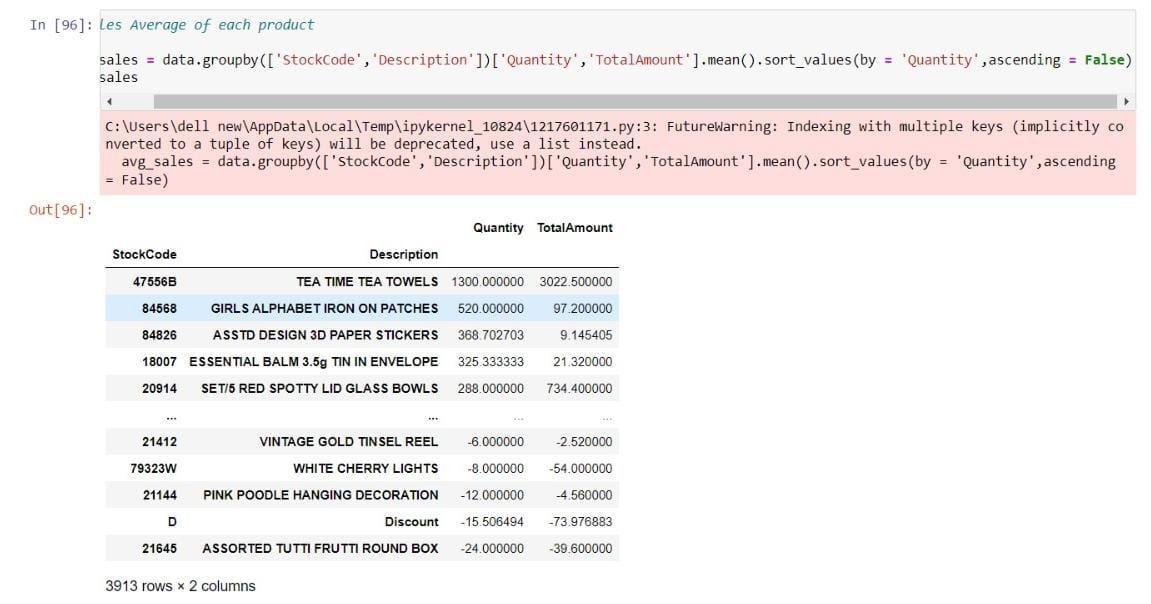
To increase more sales the company should focus on Germany and France.



This bar graph shows the top 5 least number of transactions across countries from company after removing null values.

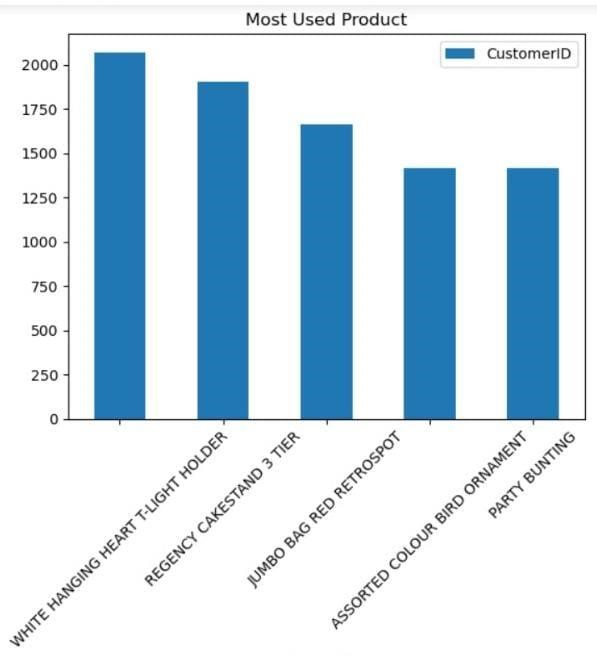
It concludes that Saudi Arabia has most transactions with the company.

The company should focus on the these countries to increase sales.

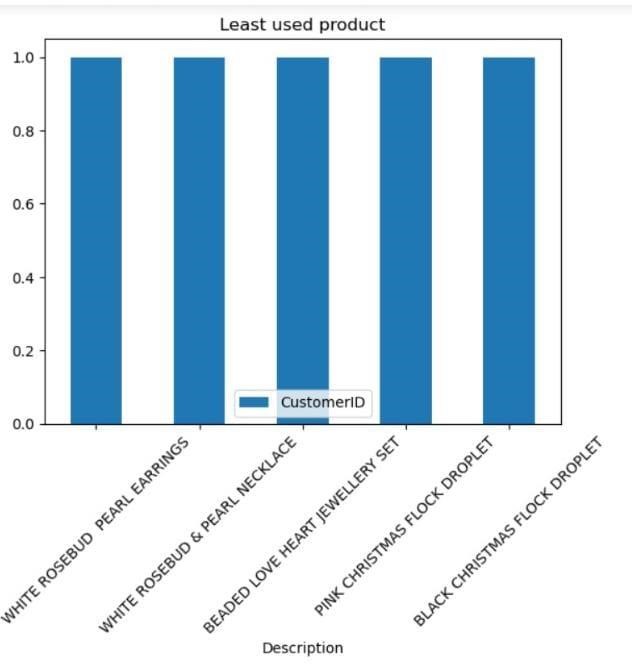


By the above graph we can say which products quantity are more sold and which products quantity are less sold.

So the company can increase the manufacturing the quantity of products which are more sold and company should focus on least quantities sold products.



By above graph we can identify which products are most sold by the company.



By the above graph we can identity which products are least sold products by the company.

The company should focus on these products to gain growth in sales .