**Data Analytics Report**

Report Summary

The datasets given are the sales and traffic access of customersin 2020 as well as products information

and customer information. This report is made not only to answers the three management questions,

but also to reveal extra insights base on the datasets.

To cut a long story short , based on the analytics, it can be concluded that :

1. The website traffic has little impact on the revenue.
2. Top 5 highest revenue products are not products that have highest pageviews.
3. here are 4 main customer segments :

* Junior customers from age 13-34 who don’t purchase much (191 customers )
* Junior customers from age 13-34 who purchase a lot (177 customers)
* Senior customers from age 34 – 58 who don’t purchase much (208 customers )
* Seniors customers from age 34 – 58 who purchase a lot. (194 customers )

1. Extras :

* German customers are the most revenue generated group.
* Demographic factors such as age , gender and country have little impact on their spending.
* Newsletter subscribe customers contributes approximately 20% more revenue than who do not.
* Italy and the UK customer has the largest percentage of6 year loyalty age in total customers.

Further explainations will be provided in the Findings

Contents

Report Summary 1

2 Findings 3

1. The website traffic has little impact on the revenue 3
2. Top Products that have good total revenue and pageviews 6
3. There are 4 Customer Segments 8
4. Extras 8

4.1 German customers are the most revenue group 8

4.2 Gender is not a factor when it comes to total revenue contribution 9

4.3 Newsletter subscribe customers contributes approximately 20% more revenue than who do not. 9

4.4 Loyalty Age does not affect customer spending 10

4.5 Age is not a factor when it comes to total revenue contribution 10

4.6 Italy and the UK customer has the largest 6 year loyalty age percentage in total customers

11

Appendix 12

Comparision between pageviews, uniquepageviews and users in 2020 12

The revenue between In Store and Online is indifferent 13

Customer Demographic 15

2 Findings

1 The website traffic has little impact on the revenue

Since our pageviews ,uniquepageviews and users access are not significantly identical throughout the period, we can use any of the three indicators to measure traffic. (Appendix 1)

Also , the total revenue between In Store and Online Store are not significantly identical throughout the period , it is unnecessary to separetly analyse them in the total revenue. (Appendix 2 )

Hence , we can use the general total revenue for analysing.

A glance at the plots of total revenue to see the similarity between In Store and Online Channel, further statistical descriptions are provided in the Appendix 2 to see that similarity between In Store and Online Channel.

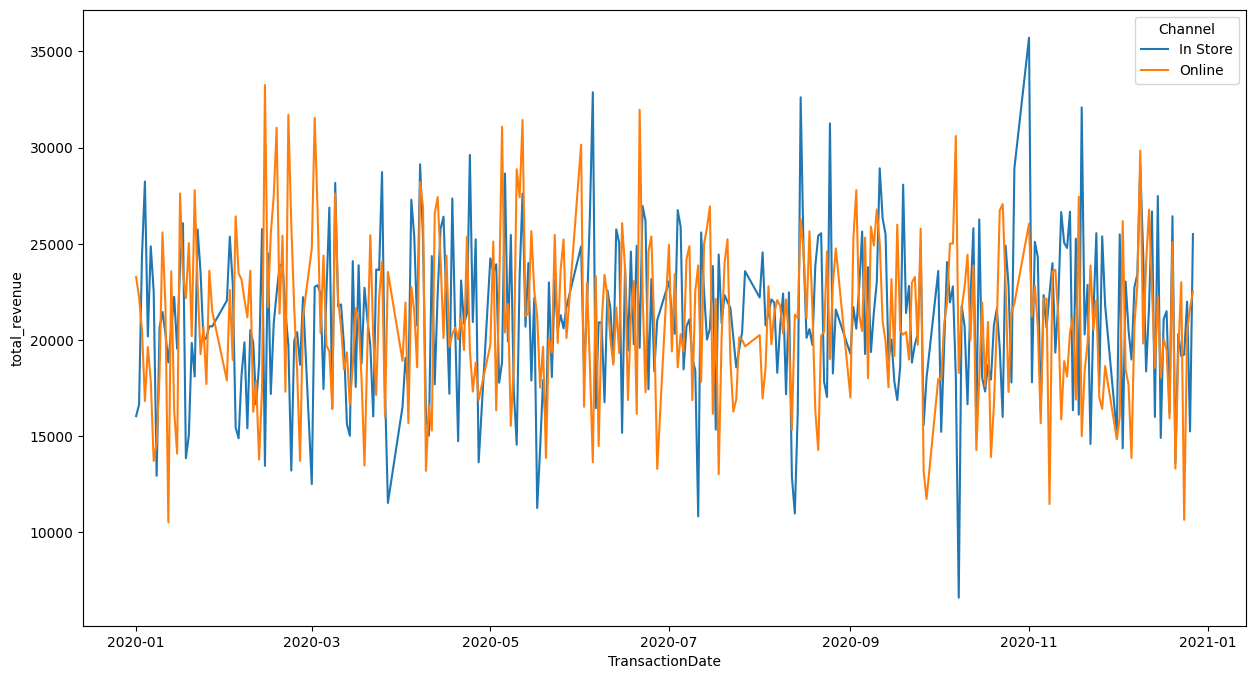
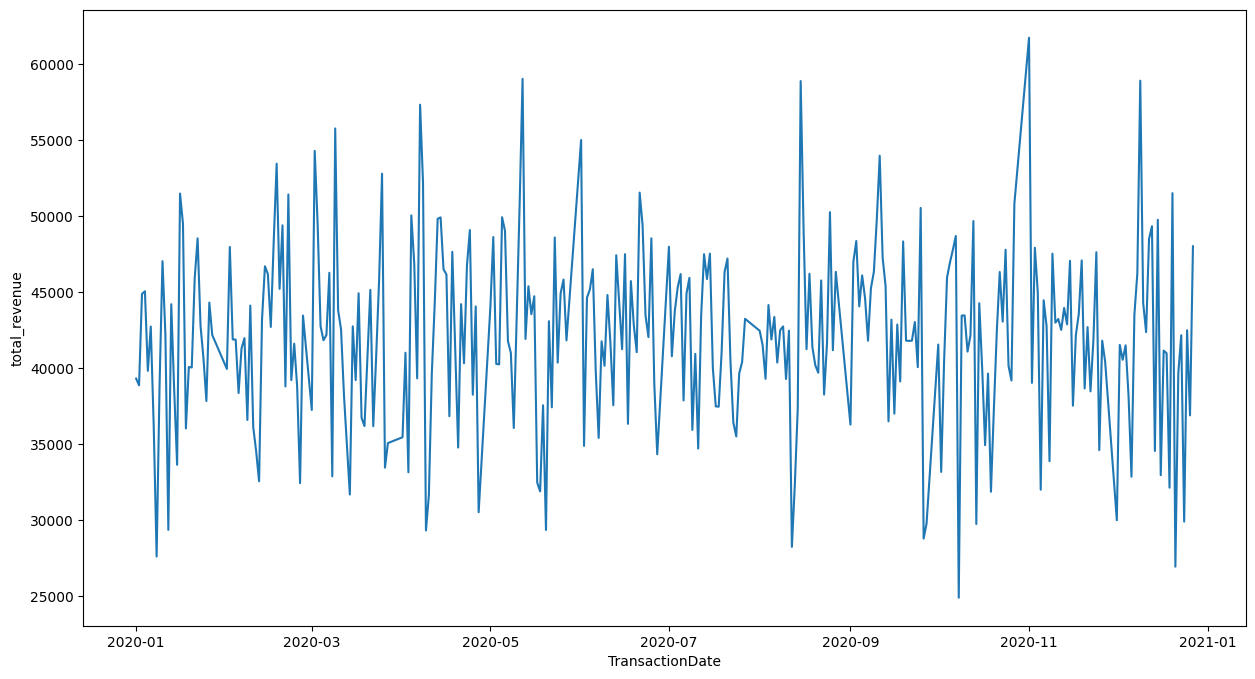
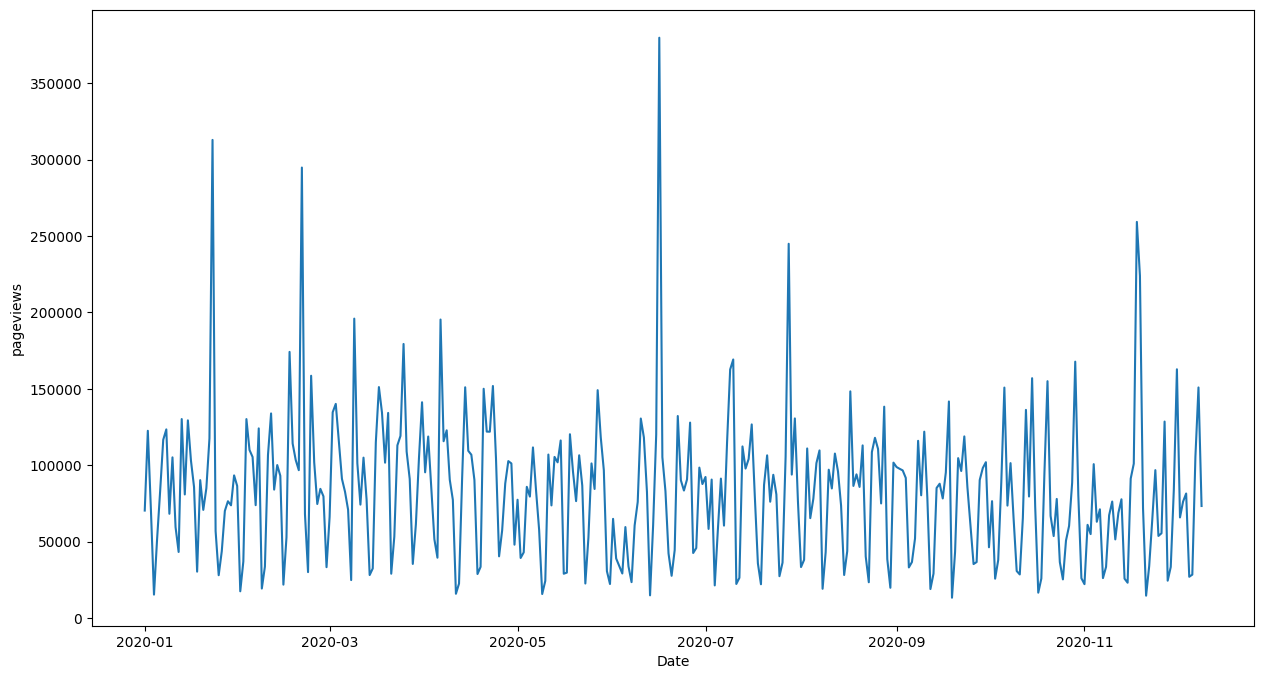


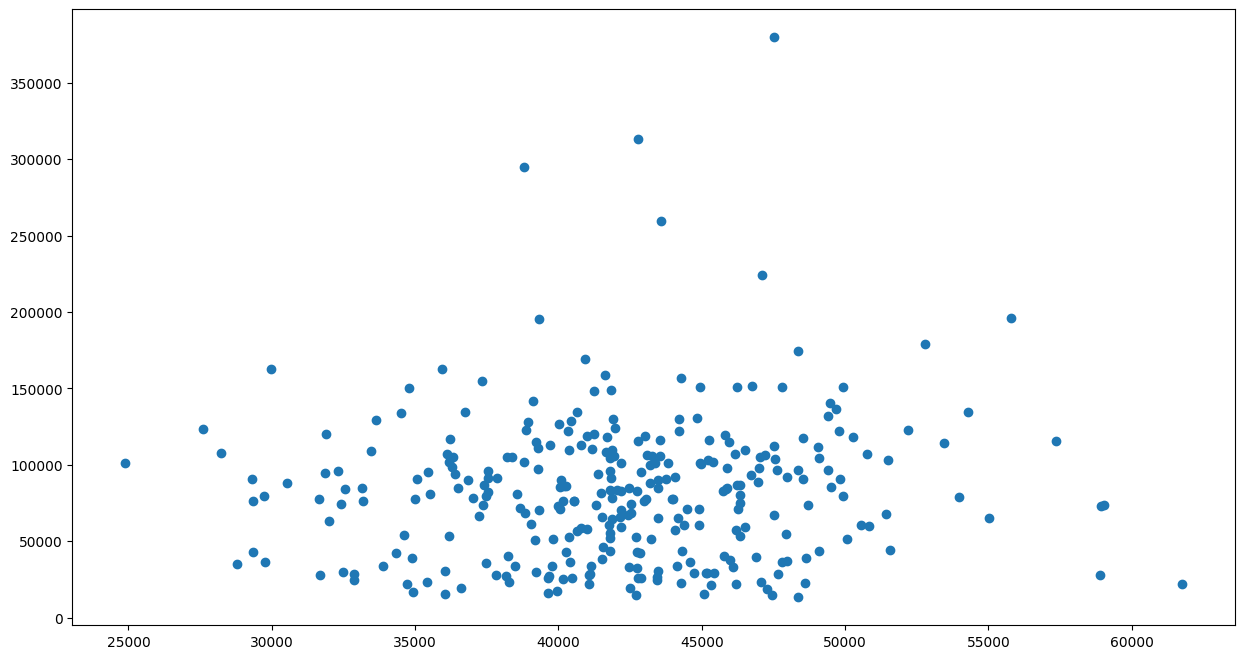
Figure 1 : Total Revenue from 2021-01 and 2020-12 of both Channel

Finally, the comparison between daily Revenue and daily Pageviews in 2020.

Total Revenue in 2020 by Date

Total Pageviews in 2020 by Date

In addition , There are not correlations, or relationships between daily Pageviews and daily revenue when show them in a scatterplot.



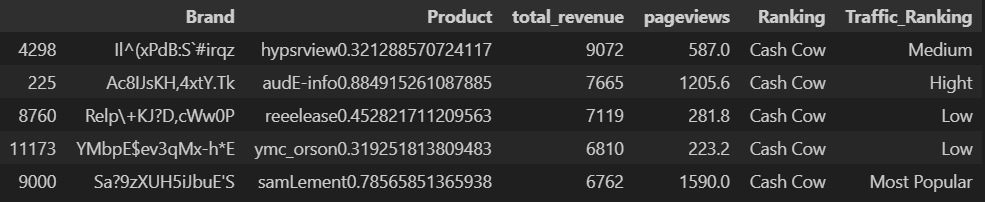
The correlation between the pageviews and total revenue are also measured, which is 0.075, very small and close to 0, which indicates there are no significant correlation between the two variables. As a result, we dont see any patterns between the two variables in the chart.

Notes :

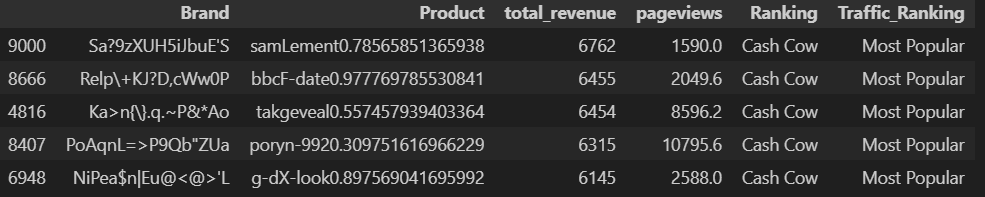
* Correlation > 1 => There are positive correlation between 2 variables
* Correlation < 1 => There are negative correlation between 2 variables
* Correlation = 0 => There are no correlation between 2 variables

2. Top Products that have good total revenue and pageviews

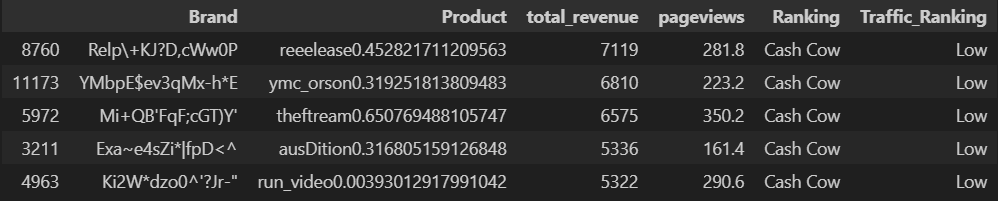
These are the five products that has the highest revenue , but very normal pageviews



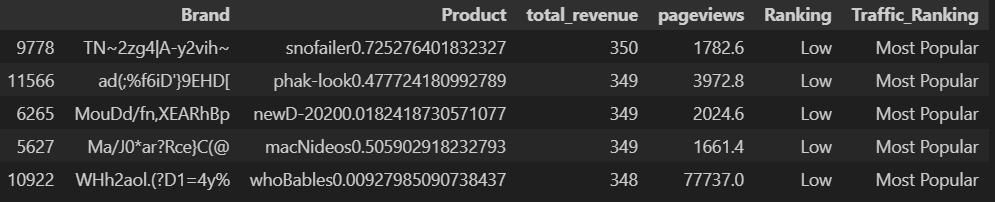
These are the five products that is amongst the highest revenue and highest pageviews



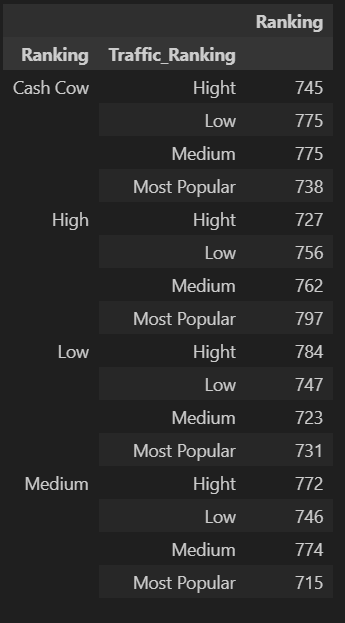
These are the five products that is amongst the highest revenue but lowest pageviews



These are the five products that is amongst the highest pageviews but lowest revenue



Maybe the dataset was evenly distributed at the start , relationship matrix between pageviews and revenue is very evenly distributed , which can be examined in the notebook.



3.There are 4 Customer Segments

Since Demographic Factors such as Age , gender, country do not contribute significantly in total revenue ( see part 4 Extras and Appendix 3). Factors such as Loyalty , Monetary and Youth are used to perform Customer Segmentation.

* Loyalty : Loyalty is the difference between today and date joined of customers. Although this factor has little impact in customer spending, it is included in the algorithm.
* Monetary : Customer’s Monetary is the ranking of their revenue compare to other customers .Monetary varies from 1 to 4 , with 1 is lowest and 4 is highest.
* Youth : Customer’s Youth is their age. Youth varies from 1 to 4 , with 1 is lowest and 4 is highest.

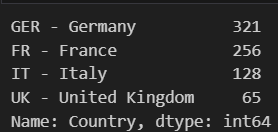
The value from 1 to 4 in each factor is determined using Interquartile Range. After performing analysis using Machine Learning ( KMeans algorithm). It can be seen that there are 4 main customer segments amongst our customers, each group is equally and evenly distributed.

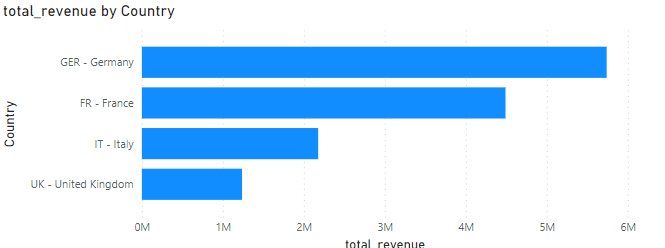
* **Junior customers from age 11-30 who don’t purchase much ( 191 customers )**
* **Junior customers from age 11-30 who purchase a lot (177 customers)**
* **Senior customers from age 30 – 56 who don’t purchase much (208 customers )**
* **Seniors customers from age 30 – 56 who purchase a lot. (194 customers )**

4. Extras

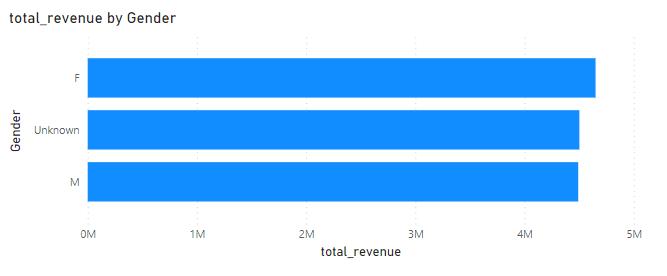
4.1 German customers are the most revenue group

Based on the data given, customer demographic factors such as age , gender are very evenly and equally  
distributed, as a result, it is ambiguous to segment customers in terms of age and gender.  
However, it can be seen that more than half of the customers are German , the second largest group is  
French, followed by Italy. UK people only account for approximately 10% of our customers.

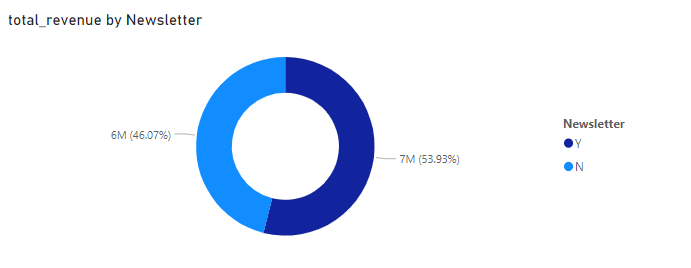


Based on the country ratio in the customers dataset, German customers account for the largest  
revenue, followed by French

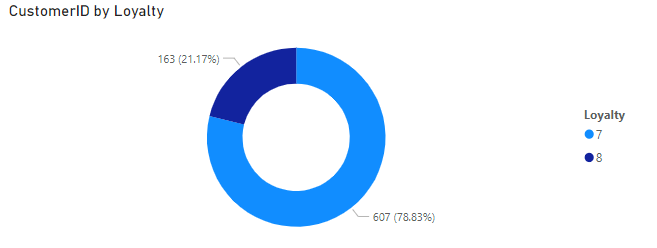
4.2 Gender is not a factor when it comes to total revenue contribution



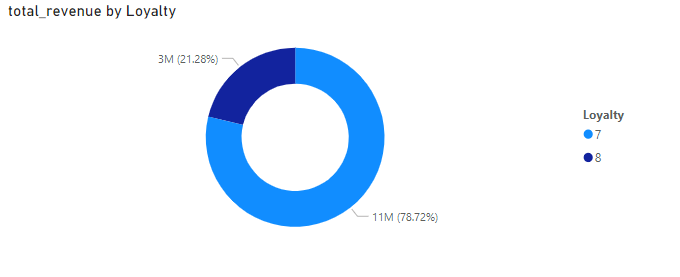
4.3 Newsletter subscribe customers contributes approximately 20% more revenue than who do  
not



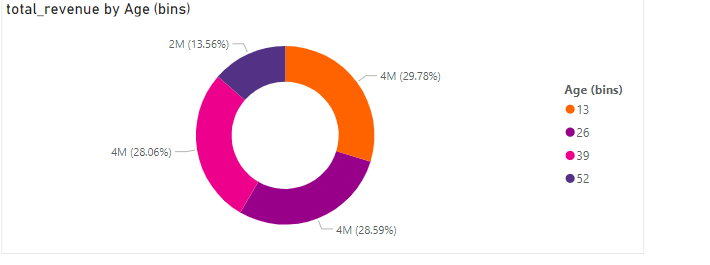
4.4 Loyalty Age does not affect customer spending



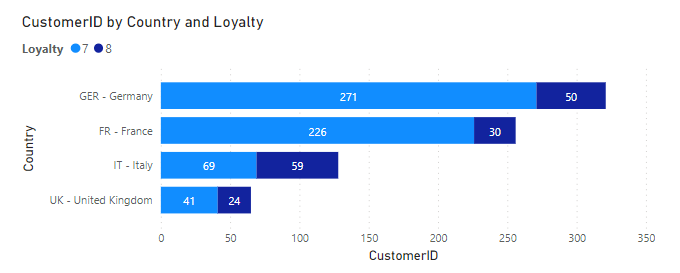
It seems that loyalty does not affect total revenue since the percentage of contribution in total revenue  
are also 20 – 80, which means that customers who have 6 years loyalty age don’t spend more or less  
than those have 5 years.



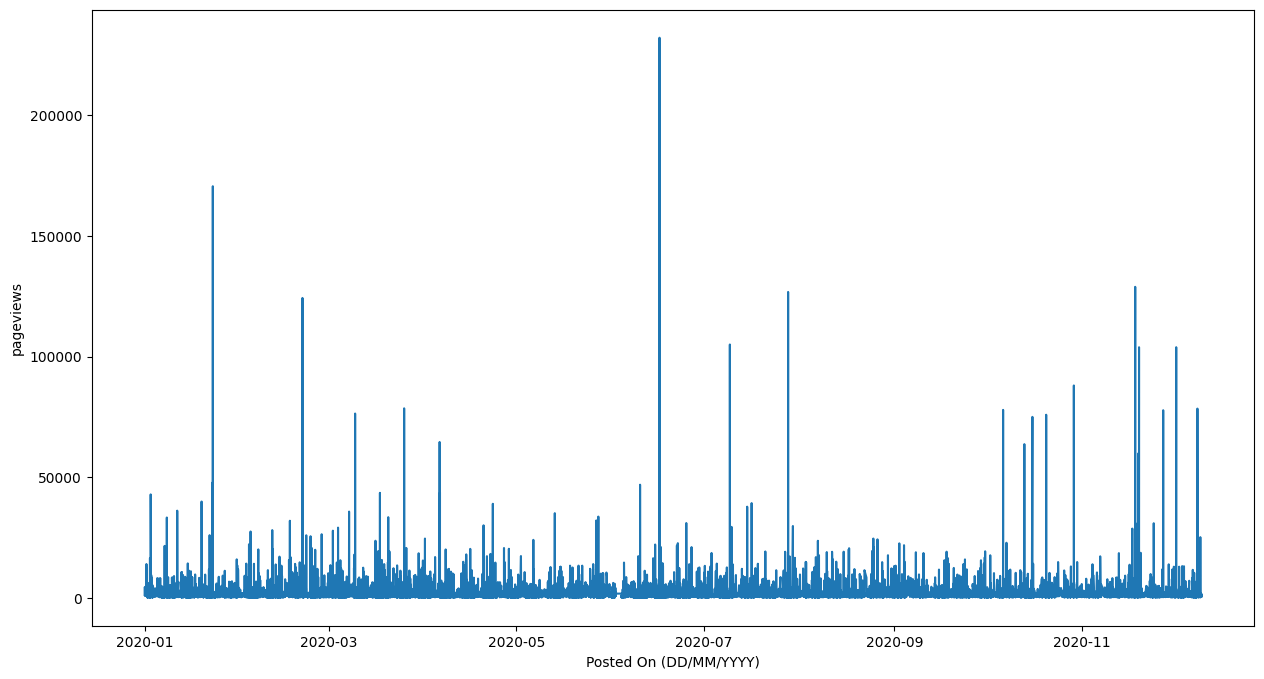
4.5 Age is not a factor when it comes to total revenue contribution



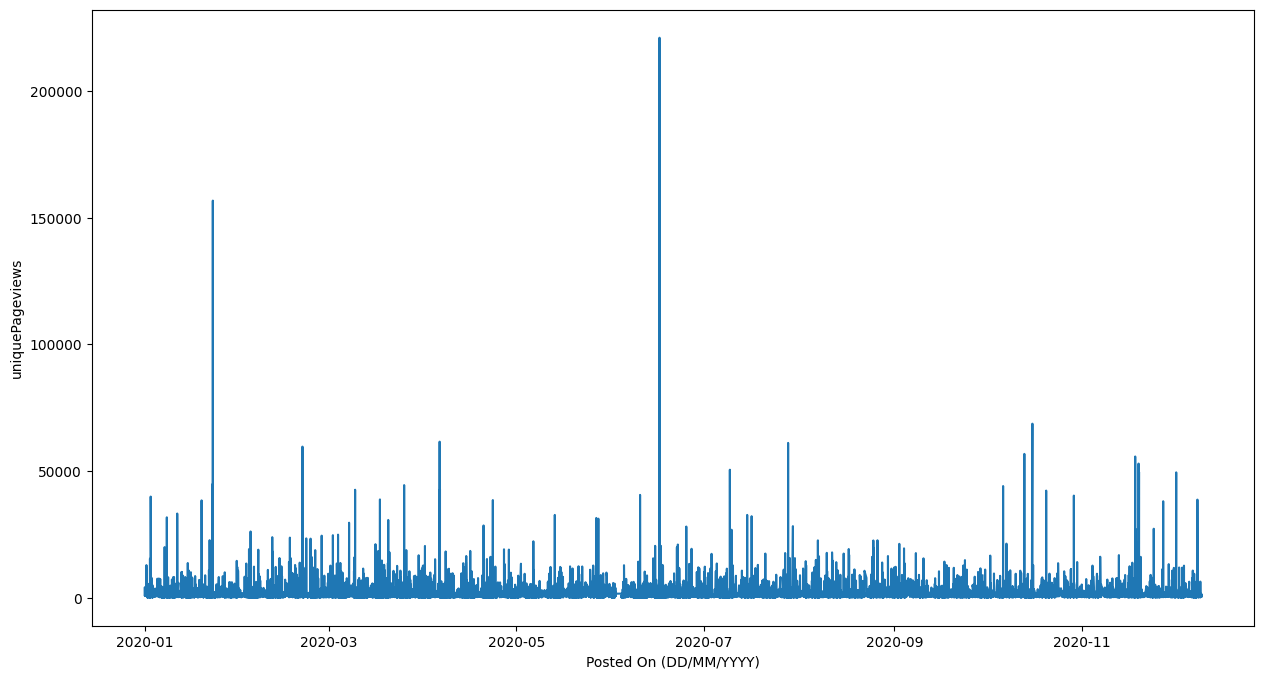
4.6 Italy and the UK customer has the largest 6 year loyalty age percentage in total customers.  
Although Italy and the UK people are the 2 smallest group, the have more than 30% of the total number  
customers are 6 years loyalty.



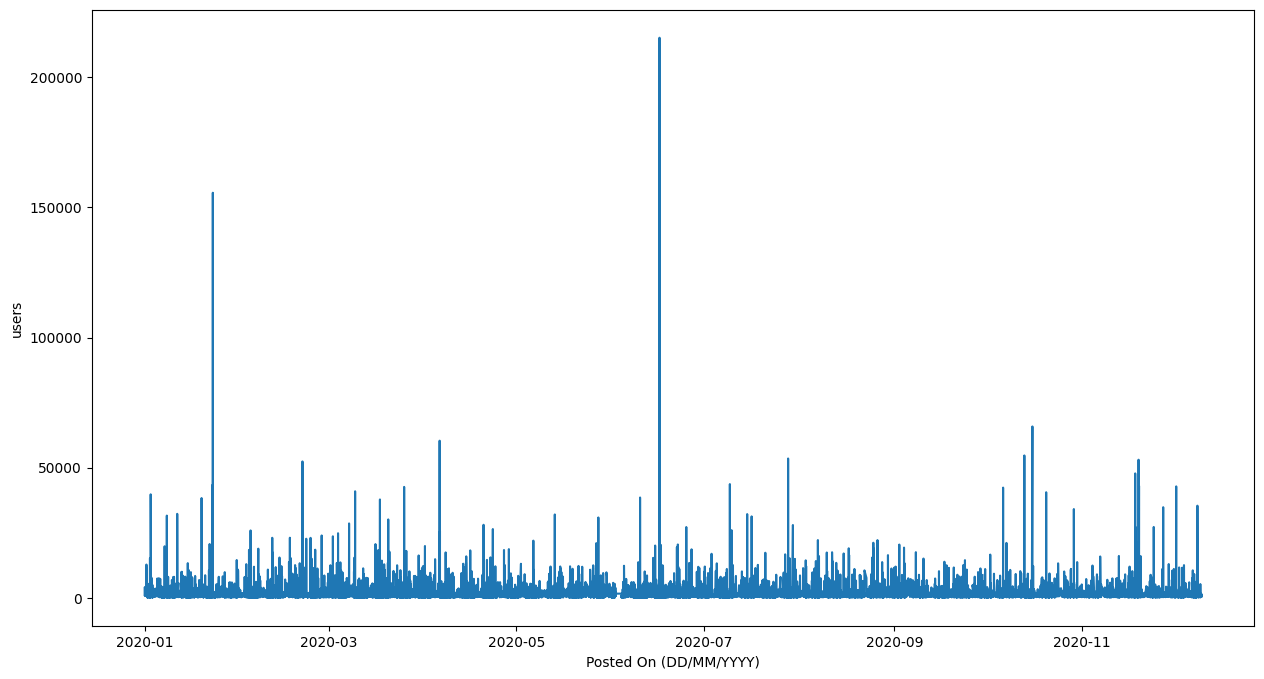
Appendix  
Comparision between pageviews, uniquepageviews and users in 2020



Pageviews in 2020

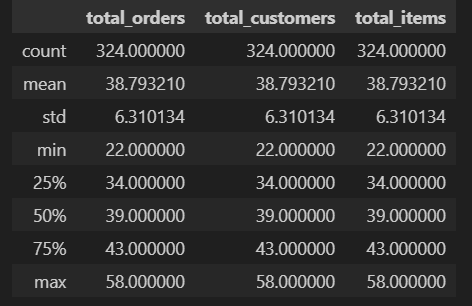


Uniquepageviews in 2020



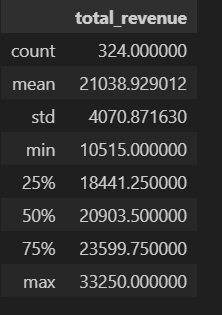
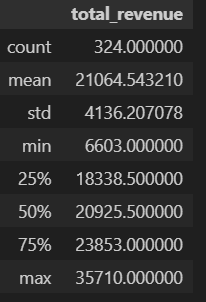
Users in 2020

The revenue between In Store and Online is indifferent



The statistics of total\_orders , total\_customers and total\_items of Online Channel

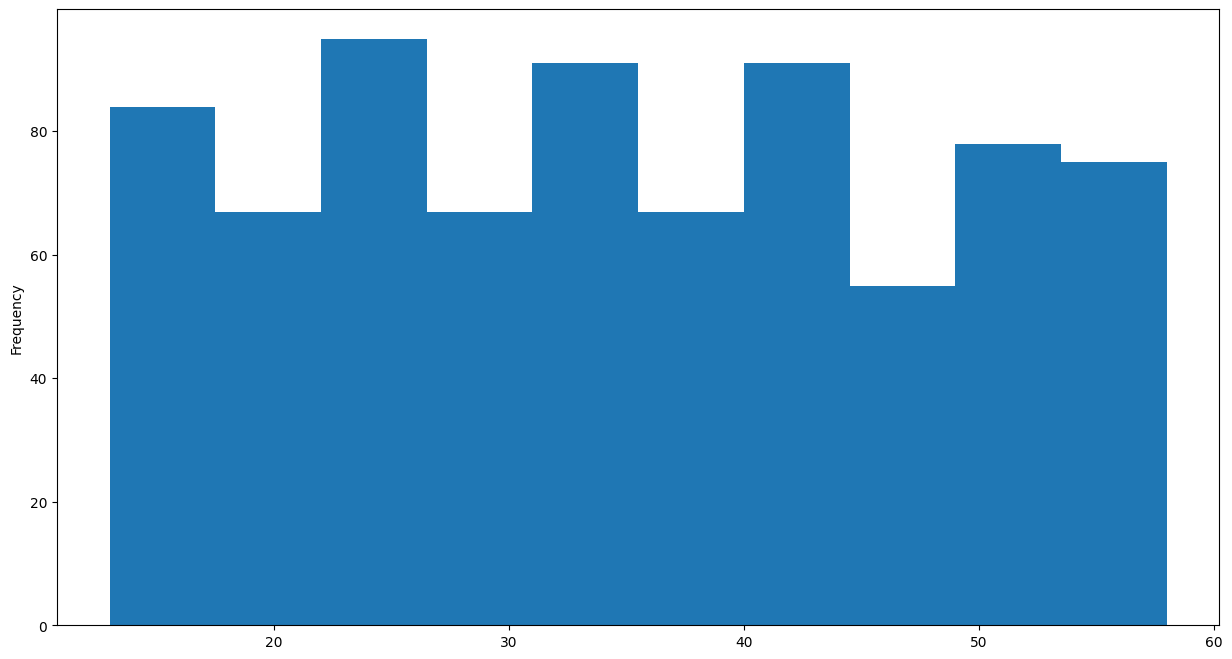


The statistics of total\_orders , total\_customers and total\_items of In Store Channel

Statistics of Total Revenue of In Store Channel Statistics of Total Revenue of Online Channel

The average daily revenue between two channel is also approximately similar (21038 compare  
to 21064 ), the ups and downs in daily revenue between each day is also indifferent ( standard  
deviation of 4070 compare to 4130) in the two channels

Customer Demographic



Age distribution of customers

It can be seen that customers varies from 13 years old to 58 years old, and they are distributed  
evenly into each groups ,there are no dominant age group in our dataset.

