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)
4. 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

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.

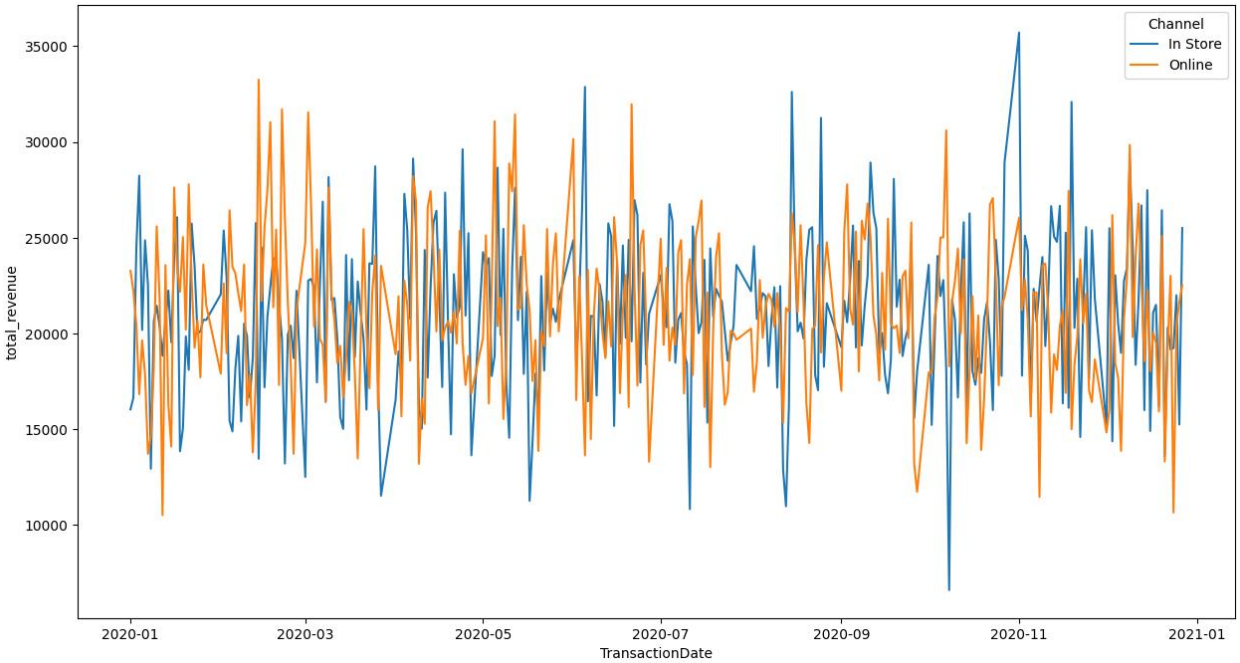
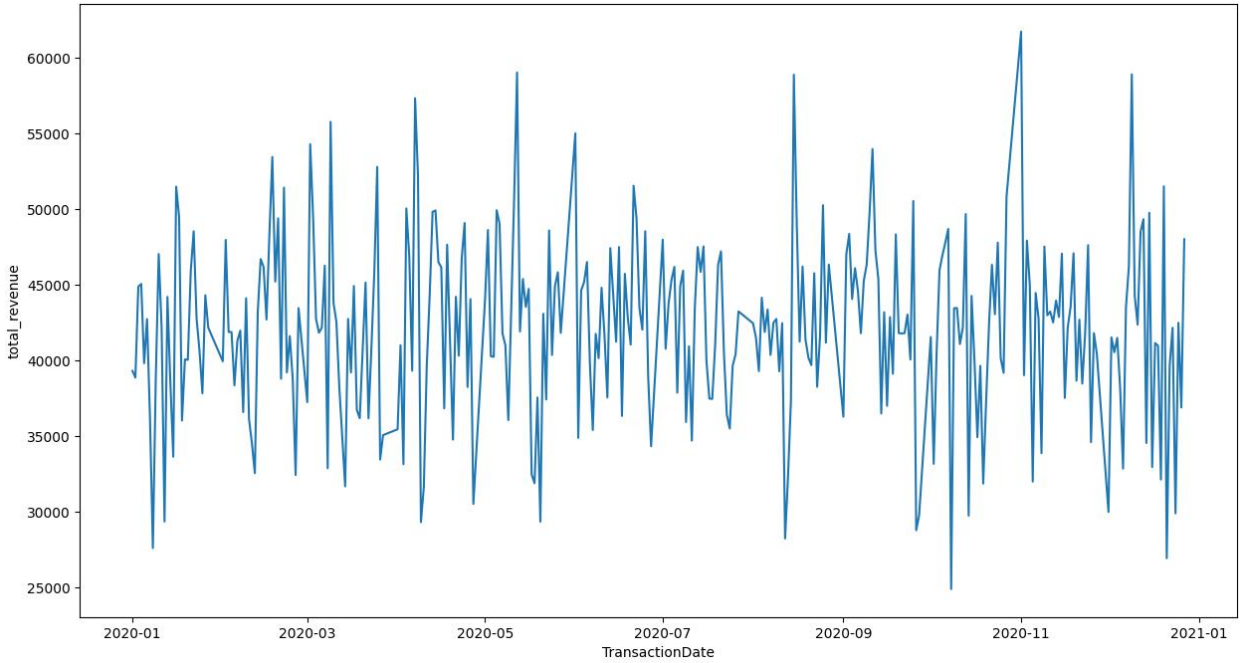
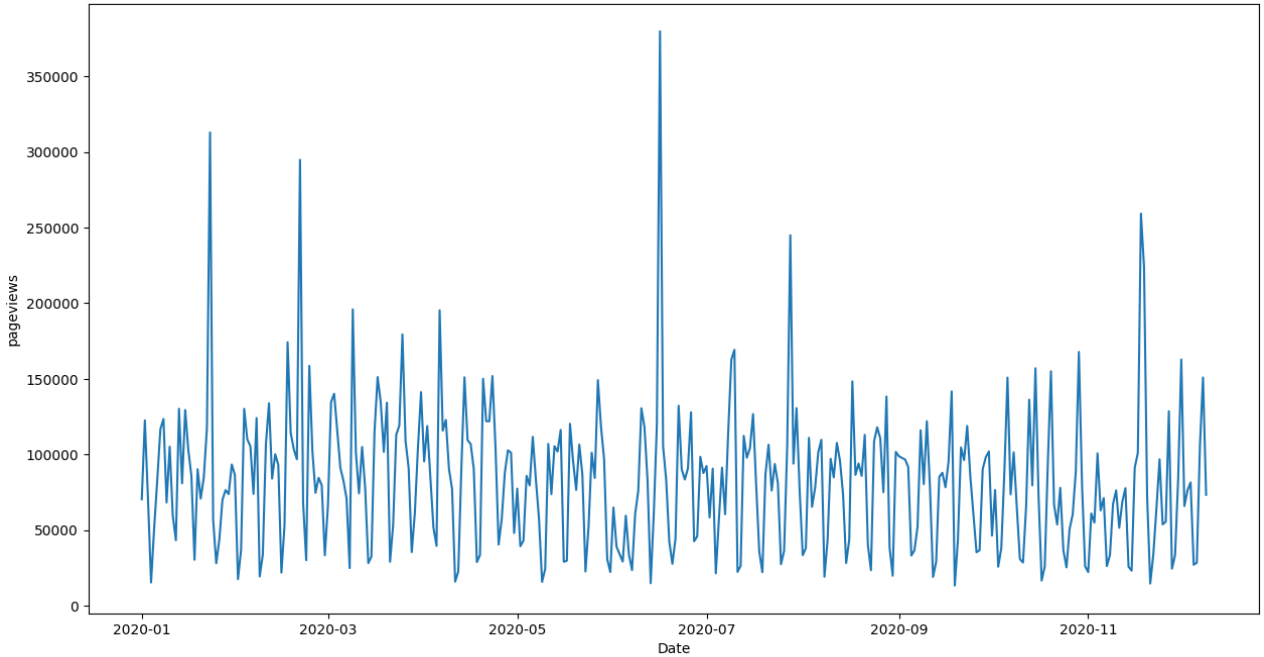


Figure 1 : Total Revenue from 2021-01 and 2020-12 of both Channel

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

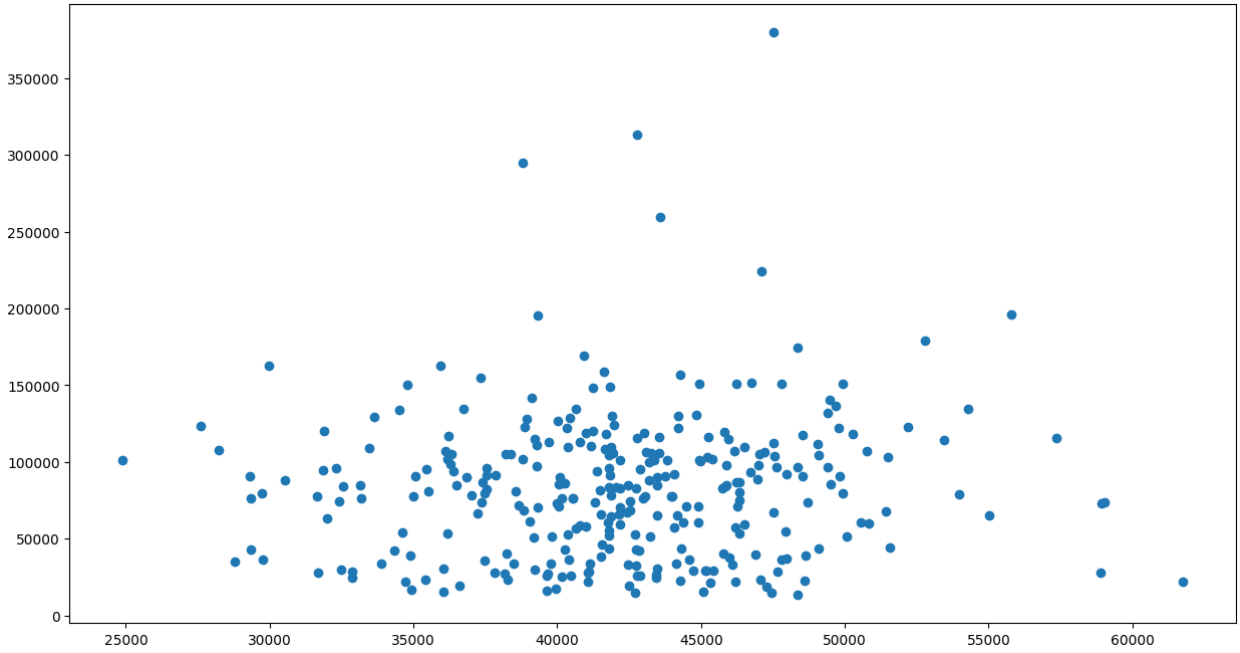


Total Revenue in 2020 by Date



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

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
4298	Il^(xPdB:S`#irqz	hypsrview0.321288570724117	9072	587.0	Cash Cow	Medium
225	Ac8IJsKH,4xtY.Tk	audE-info0.884915261087885	7665	1205.6	Cash Cow	Hight
8760	Relp\+KJ?D,cWw0P	reeelease0.452821711209563	7119	281.8	Cash Cow	Low
11173	YMbpE$ev3qMx-h*E	ymc_orson0.319251813809483	6810	223.2	Cash Cow	Low
9000	Sa?9zXUH5iJbuE'S	samLement0.78565851365938	6762	1590.0	Cash Cow	Most Popular

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

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
9000	Sa?9zXUH5iJbuE'S	samLement0.78565851365938	6762	1590.0	Cash Cow	Most Popular
8666	Relp\+KJ?D,cWw0P	bbcF-date0.977769785530841	6455	2049.6	Cash Cow	Most Popular
4816	Ka>n{\}.q.~P&*Ao	takgeveal0.557457939403364	6454	8596.2	Cash Cow	Most Popular
8407	PoAqnL=>P9Qb"ZUa	poryn-9920.309751616966229	6315	10795.6	Cash Cow	Most Popular
6948	NiPea$n|Eu@<@>'L	g-dX-look0.897569041695992	6145	2588.0	Cash Cow	Most Popular

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

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
8760	Relp\+KJ?D,cWw0P	reeelease0.452821711209563	7119	281.8	Cash Cow	Low
11173	YMbpE$ev3qMx-h*E	ymc_orson0.319251813809483	6810	223.2	Cash Cow	Low
5972	Mi+QB'FqF;cGT)Y'	theftream0.650769488105747	6575	350.2	Cash Cow	Low
3211	Exa~e4sZi*|fpD<^	ausDition0.316805159126848	5336	161.4	Cash Cow	Low
4963	Ki2W*dzo0^'?Jr-"	run_video0.00393012917991042	5322	290.6	Cash Cow	Low

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

	Brand	Product	total_revenue	pageviews	Ranking	Traffic_Ranking
9778	TN~2zg4|A-y2vih~	snofailer0.725276401832327	350	1782.6	Low	Most Popular
11566	ad(;%f6iD'}9EHD[	phak-look0.477724180992789	349	3972.8	Low	Most Popular
6265	MouDd/fn,XEARhBp	newD-20200.0182418730571077	349	2024.6	Low	Most Popular
5627	Ma/J0*ar?Rce}C(@	macNideos0.505902918232793	349	1661.4	Low	Most Popular
10922	WHh2aol.(?D1=4y%	whoBables0.00927985090738437	348	77737.0	Low	Most Popular

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.

		Ranking
Ranking	**Traffic_Ranking**	
Cash Cow	Hight	745
	Low	775
	Medium	775
	Most Popular	738
High	Hight	727
	Low	756
	Medium	762
	Most Popular	797
Low	Hight	784
	Low	747
	Medium	723
	Most Popular	731
Medium	Hight	772
	Low	746
	Medium	774
	Most Popular	715

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)**

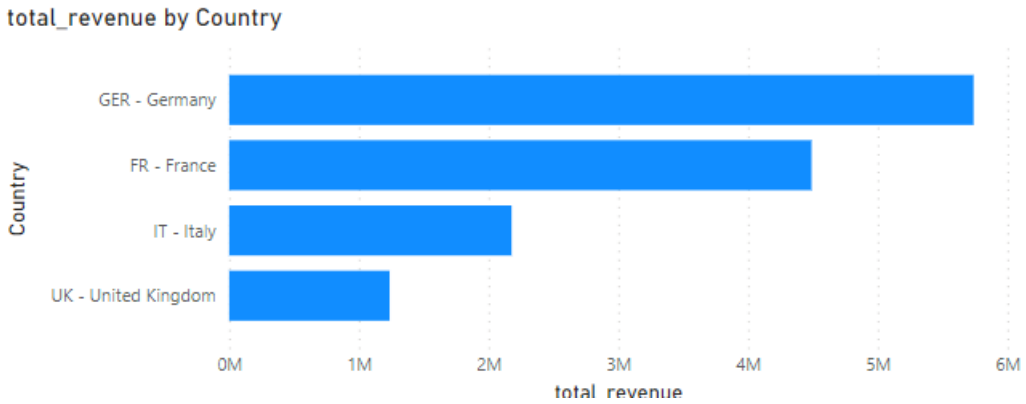
4. Extras

4.1 German customers are the most revenue group

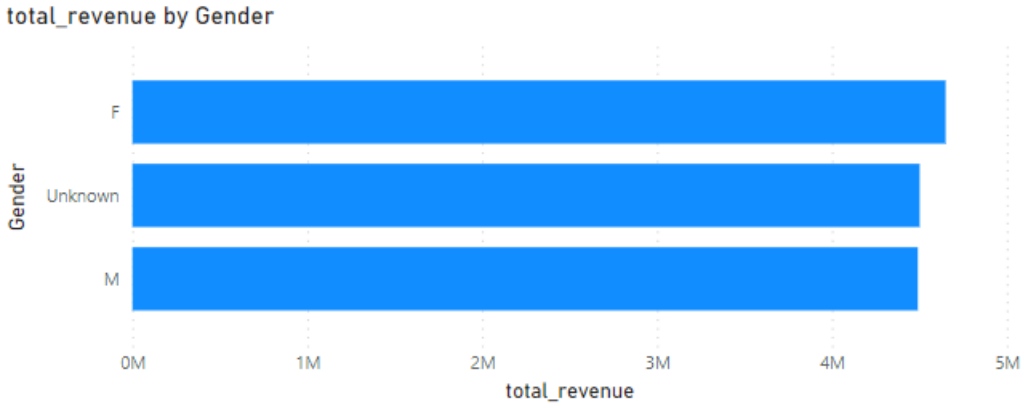
Based on the data given, customer demographic factors such as age , gender are very evenly and equally
distributed, as a result, it is ambiguous to segment customers in terms of age and gender. However, it can be seen that more than half of the customers are German , the second largest group is
French,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

```
GER - Germany           321
FR - France             256
IT - Italy              128
UK - United Kingdom      65
Name: Country, dtype: int64
```

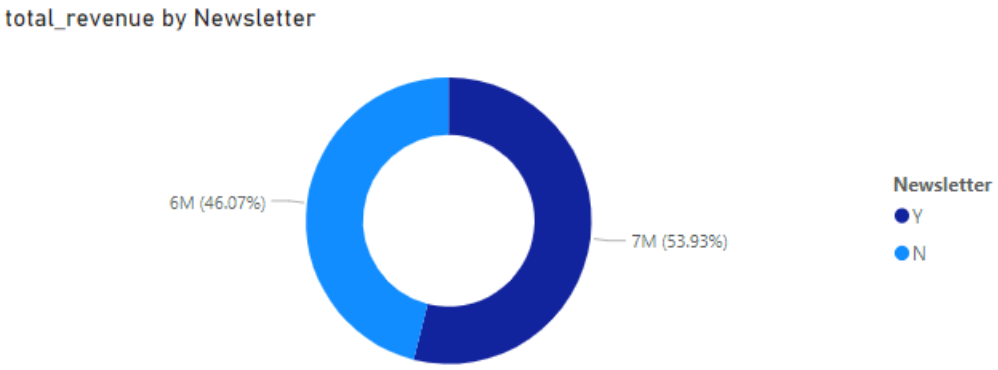
Based on the country ratio in the customers dataset, German customers account for the largest revenue, followed by French



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

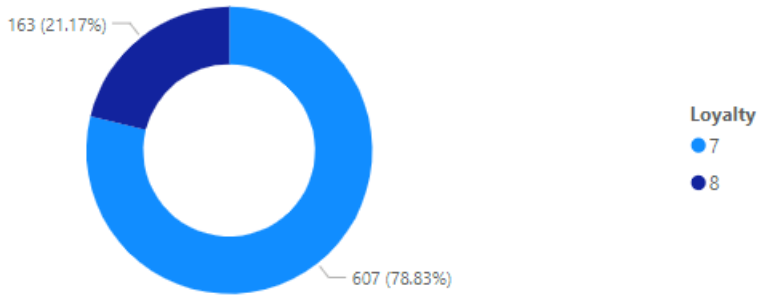


4.3 Newsletter subscribe customers contributes approximately 20% more revenue than who do not



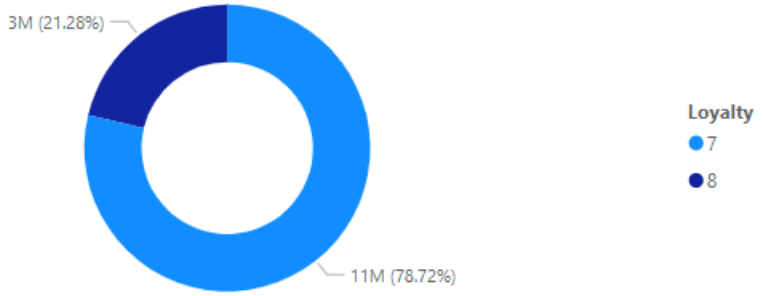
4.4 Loyalty Age does not affect customer spending

CustomerID by Loyalty



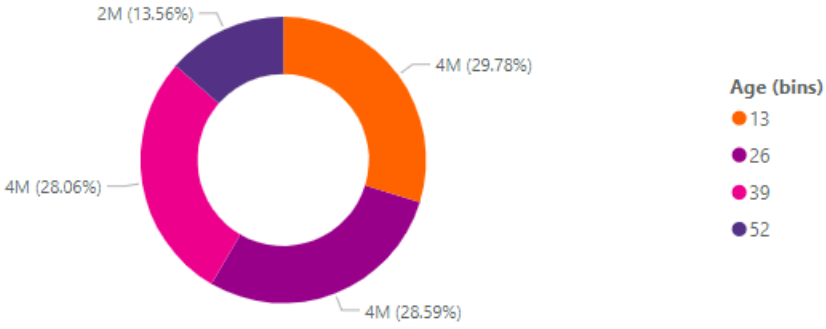
It seems that loyalty does not affect total revenue since the percentage of contribution in total revenue
are also 20 – 80, which means that customers who have 6 years loyalty age don't spend more or less
than those have 5 years.

total_revenue by Loyalty



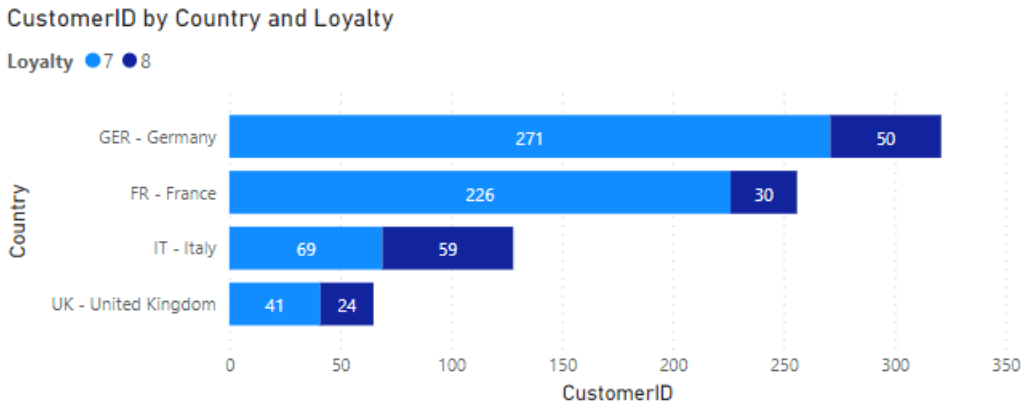
4.5 Age is not a factor when it comes to total revenue contribution

total_revenue by Age (bins)

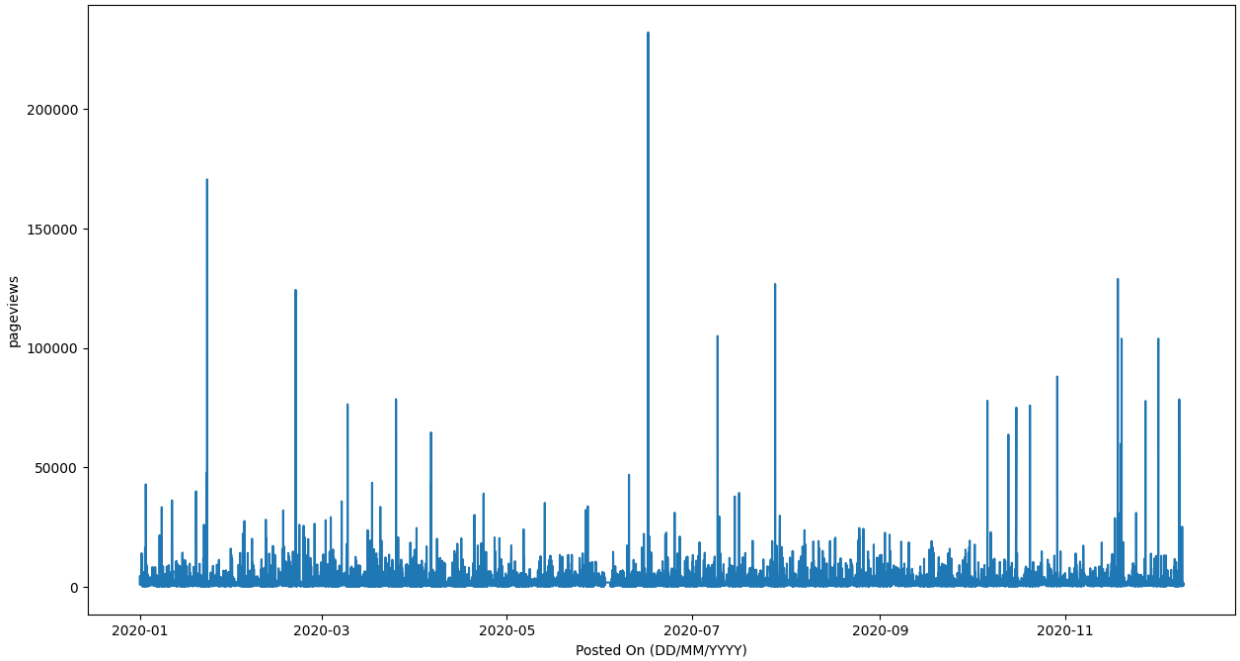


4.6 Italy and the UK customer has the largest 6 year loyalty age percentage in total customers.

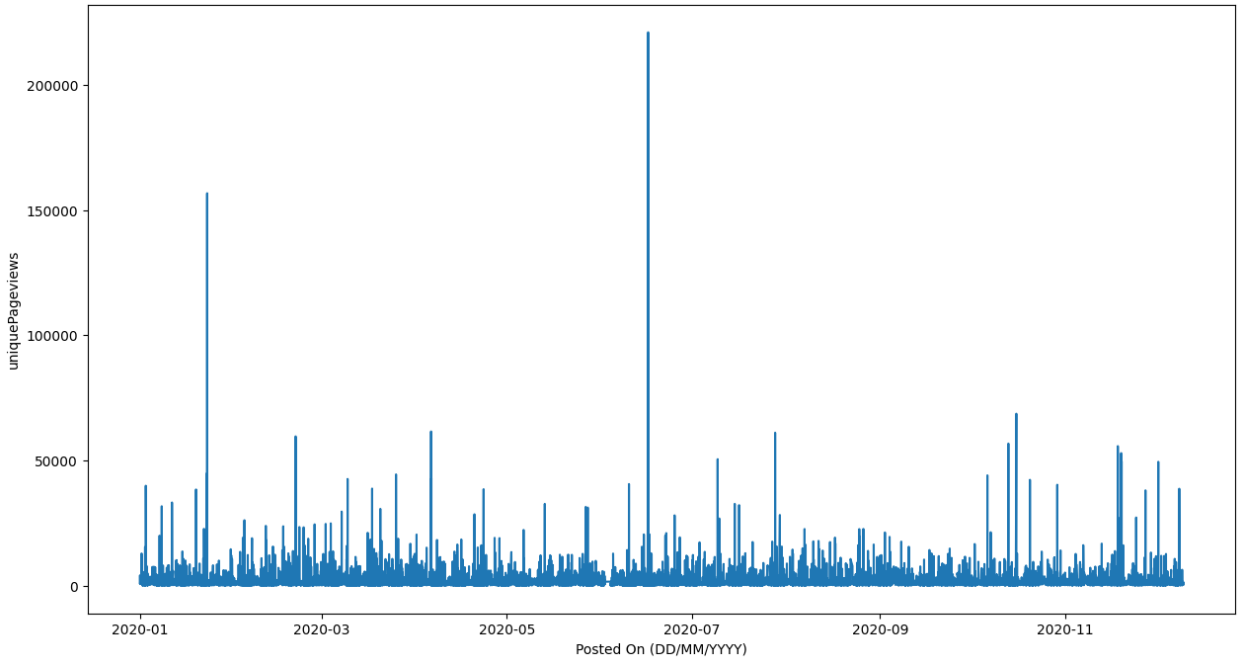
Although Italy and the UK people are the 2 smallest group, the have more than 30% of the total number
customers are 6 years loyalty.



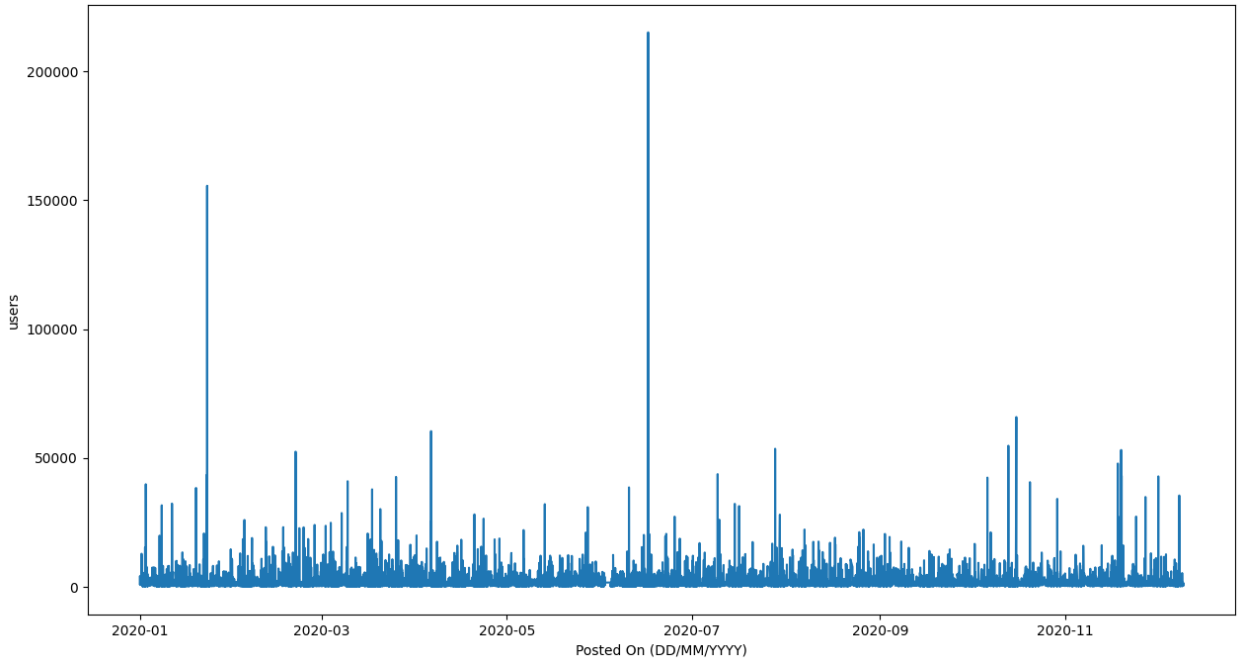
Appendix
Comparision between pageviews, uniquepageviews and users in 2020



Pageviews in 2020



Uniquepageviews in 2020



Users in 2020

The revenue between In Store and Online is indifferent

	total_orders	total_customers	total_items
count	324.000000	324.000000	324.000000
mean	38.793210	38.793210	38.793210
std	6.310134	6.310134	6.310134
min	22.000000	22.000000	22.000000
25%	34.000000	34.000000	34.000000
50%	39.000000	39.000000	39.000000
75%	43.000000	43.000000	43.000000
max	58.000000	58.000000	58.000000

The statistics of total_orders , total_customers and total_items of Online Channel

	total_orders	total_customers	total_items
count	324.000000	324.000000	324.000000
mean	39.024691	39.024691	39.024691
std	6.549124	6.549124	6.549124
min	19.000000	19.000000	19.000000
25%	35.000000	35.000000	35.000000
50%	39.000000	39.000000	39.000000
75%	43.000000	43.000000	43.000000
max	60.000000	60.000000	60.000000

The statistics of total_orders , total_customers and total_items of In Store Channel

	total_revenue
count	324.000000
mean	21064.543210
std	4136.207078
min	6603.000000
25%	18338.500000
50%	20925.500000
75%	23853.000000
max	35710.000000

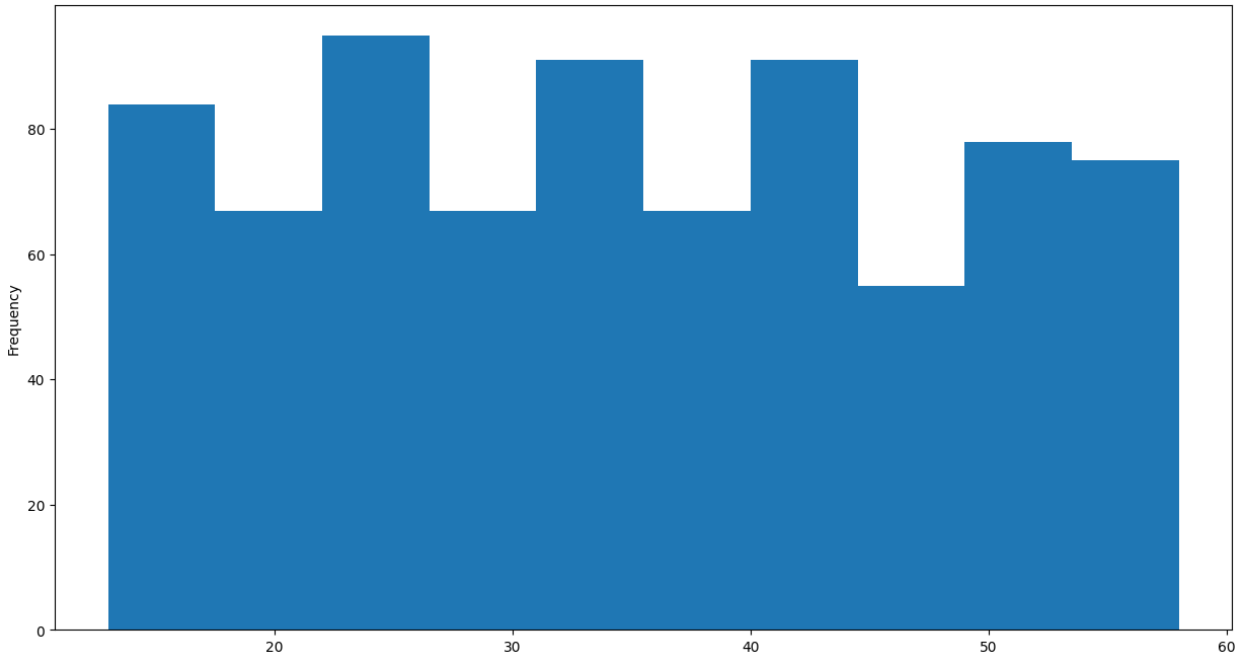
Statistics of Total Revenue of In Store Channel

	total_revenue
count	324.000000
mean	21038.929012
std	4070.871630
min	10515.000000
25%	18441.250000
50%	20903.500000
75%	23599.750000
max	33250.000000

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.