A study of E-commerce Sales and Forecast of retail sales in U.K: A time series analysis

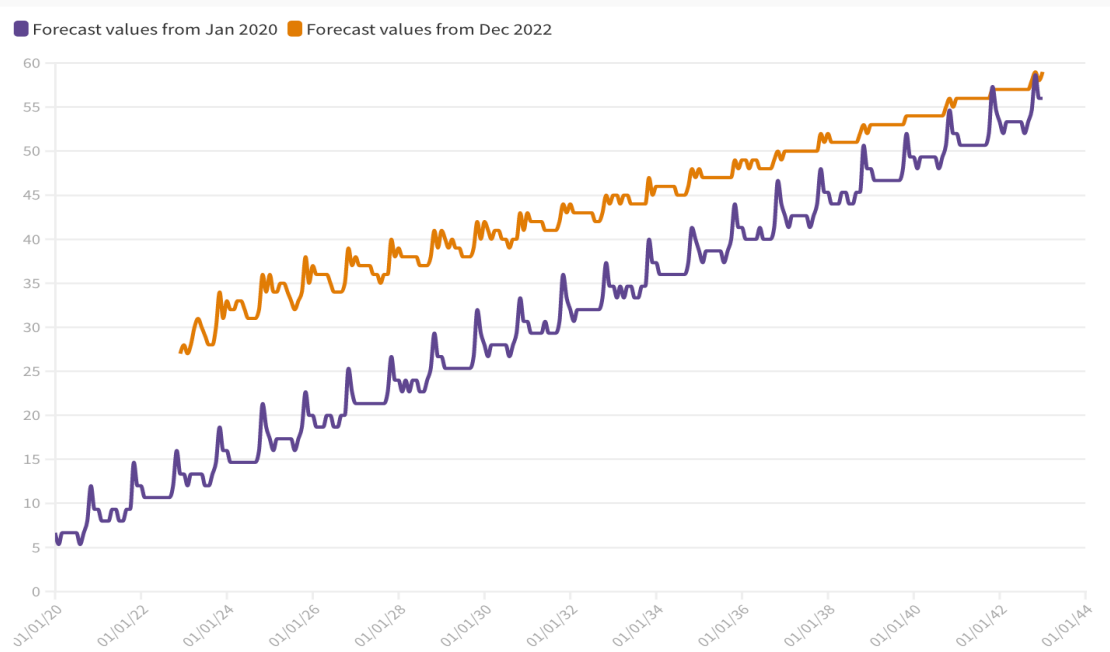
Module Code: BEMM466J
Module Name: Business Project

Table of Contents

Executive Summary

Purpose of this business project:
The main purpose of this report is to analyse the trends in e- commerce sales and in particular the online retail sales in the United Kingdom. This is done by employing time series analysis. The findings of this analytical study will be of use to retailers, policymakers, and other stakeholders in the United Kingdom. However, it is the aim of this study paper to assist all novice entrepreneurs of the United Kingdom who wish to start a business in retail sector. The study includes a detailed analysis of how e- commerce sales progressed over time in the United Kingdom. The study used Exploratory Data Analysis (EDA) and related statistical models to investigate the patterns in e- commerce sales in the United Kingdom. With recourse to a statistical model called Autoregressive Integrated Moving Average (ARIMA), this paper generated a forecast for sales on e- commerce platforms in the retail sector of the UK.

Key findings:
The findings indicate that the proportion of online sales have increased 12.3 per cent in the past 8 years. Further, the retail sales worldwide on e- commerce platforms increased almost 4.2 times since the year 2014. In United Kingdom, e- commerce generates a market of 693 billion pounds in year 2019, with over 50 per cent of this revenue being generated by sale over websites. It is pertinent to note that although large companies earned significantly more revenue compared to Small and Medium sized Enterprises (SMEs) through online sales, these SMEs in the United Kingdom, despite their size limitations, were successful in registering a revenue growth by widening their customer bases through online sales. In the United Kingdom, the wholesale industries and manufacturing industry generated highest revenue from online sales. Specifically, the wholesale sector made the highest sales through online platform in the year 2019 and the construction sector was the industry making least revenue on e- commerce sites. Over a period of last two decades, the UK population between the age group of 25 to 34 have been making the highest percentage of online purchases. However, by the year 2020, buyers in the age bracket of 16 to 24 have also increased significantly in using online markets. Both men and women by the year 2020 are shopping online in equal proportions. A total of 34.3 per cent of UK's population used to shop online 11 or more times. Furthermore, 42 per cent of the UK's population spends at least 100 pounds to 499 pounds in online purchase. The items that were purchased the most online in the year 2022 were clothes, shoes and accessories. South-East of the UK has most number of internet users. As is evident from the correlation map, most e- commerce sales are made in this region. E- commerce market witnessed an unexpected spike in sales during the year 2020 which can be accounted by the onset of corona virus. Due to this boost of online sales the percentage of retail sales shifting to e- commerce increased by almost by an average of 8 per cent in the year 2020 than the previously predicted values.



Recommendations

The recommendation of this study is of particular assistance to novice entrepreneurs who wish to enter the business in the retail sector in the United Kingdom. The forecast model generated predicts that by the year 2040 almost an average of 40 per cent of retail sales in the United Kingdom will be conducted on the internet. From the exploratory analysis of relevant data published on Office of National Statistics (ONS), it can be observed that most of the e-commerce sales revenue will be generated from website sales. Hence, having a website to conduct retail sales will yield higher revenues and reach for a retail business in the United Kingdom. It can also be drawn that it will be profitable if they start a retail e commerce business which appeals to people of the United Kingdom aged below 54 years by specifically targeting buyers of ages 16 to 34 years. Starting an e commerce business which is either located or is available to the population living in South East region of United Kingdom will help boost their probability to increase revenues from online sale of goods. Since around 42 per cent of the United kingdom population is likely to spend between 100 and 499 pounds worth of good and majority of the items bought are clothes, shoes and accessories, it would be ideally profitable to stock on inventories that match this requirement. For an already existing retail business which that does not have e-commerce sales platform, it is highly recommended to establish online shopping portal since the forecast graph shows an increasing trend in the shift of retail sales to e- commerce platforms.

Conclusion:

The results in this study can help new entrepreneur in the United Kingdom who wish to enter the UK's retail sector to gain insights into trends and patterns in online retail sales in the United Kingdom. The study also discusses various time series related to e- commerce sales in the United Kingdom. The same can also be used to observe the positive impact Covid 19 pandemic on e- commerce retail sales in the United Kingdom and to forecast future sales in the United Kingdom. Sales over an e commerce platform will generate high revenues for companies with 1000 or more employees and increase the revenue generated for any company with less than 1000 employees. Due to covid 19 pandemic many people are willing to shift to online shopping. The retail e- commerce sales increased due to pandemic by an average of 7 per cent more than the predicted values in the year 2020. And this trend seems to continue in the foreseeable future with almost 42 % of retail sales in United Kingdom shifting to e- commerce platforms by year 2031.

Business Report

Introduction

In recent years, many activities have moved from offline to online platforms in the wake of invention of the internet and as internet has become more widely used and accessible. Likewise, business is also being conducted online through platforms that create a new kind of virtual market. With the invention of e- commerce and the widespread popularity of online shopping, businesses from around the world have attempted to create online presences to increase their sales and revenue. It is important for business, existing or new, to be able to evaluate these contemporary methods of conducting business and determine if it is suitable for their business. E-commerce sales have also had a significant impact on a country's Gross domestic product, creating a new source of revenue for the country's economy. Therefore, analysis of data related to e-commerce activities in a country is important not only for business owners but also for policy makers. This extended report will analyse the data on e-commerce sales and that in particularly more specific to the united kingdom's e commerce sales. the analysis will be of a how the online market has changed over time, especially with the sudden change in scenarios due to an unforeseen event, the corona virus pandemic. Such analysis of time series data relating to the e- commerce sector in the United Kingdom as well as establishing a trend and shift in trend is very useful for evaluating the success of a business and the economy as a whole. In addition, since the study concentrates on the e-commerce activity within the United Kingdom, the report will also examine general, physical, demographic, and geographic data of the country and analyse it to identify correlations and make a forecast of how the retail e-commerce sales might be impacted in the future.

Background

The United Kingdom has been one of the top countries with highest per cent of its population shopping online. The retail sector in United Kingdom has seen a rapid increase in sales over internet. E- commerce made up 30% of the UK's total retail market in 2021, with annual revenue reaching over $120 billion. The drive for a continuous quest of value for money is developing the internet commerce. People continue to seek online for the greatest offers and rates. This is driving the share of money spent online, which has been showing an increasing trend. Shopping for fashion, health and beauty, home and garden, consumer electronics, and travel services is increasingly taking place on the internet. Click and collect is also becoming increasingly popular. Over 80% of the UK population made at least one online purchase that year. Amazon.co.uk is the leading e-commerce company in the UK, followed by tesco.com and argos.co.uk. (International Trade Administration, 2022) With the continued expansion of social networking and mobile internet connectivity, social media marketing is becoming increasingly popular. Hence, it is important to combine business and data analysis and establish trends to either create a strong market value for businesses or uplift the country's economy.

Research Questions

This analysis report can be used by enterprises trying to enter into the current retail sector in the united kingdom or looking to expand their ecommerce presence. The results also cater to government policymakers who are seeking to understand the general trend in the united kingdom's electronic commerce. This report conducts a comprehensive analysis of the time series of e-commerce in the United Kingdom and more specifically into the retail sector of the United Kingdom. The report seeks to answer questions about the performance of e-commerce sales compared to offline sales globally and then specifically for United Kingdom, and how preferences for online purchases have evolved across

different modes of e-commerce sales in the United Kingdom. The report also examines if a company is required be a certain size to perform well in the e- commerce market. Additionally, the report looks at how different sectors are performing in the e -commerce market in the United Kingdom with a focus on how the retail sector has been doing on the online platforms in the United Kingdom. A further analysis of the detailed review of how retail sector has been performing on the e- commerce front compared to its overall performance over past few years. the report also analyses the change in demographics of united kingdom's customers of e- commerce platforms over time, like what age group of people are making the most online purchase are and which gender shops more online in the United Kingdom The study also aims to determine which area within the United Kingdom has increased internet access hence increased e-commerce sales in a specific geographical area in the United Kingdom. It also analyses the spending habits of united kingdom's online shoppers like how many times the customers are making online purchase and how much they are spending and which items they are spending most of their money on. The report also considers the impact of the COVID-19 pandemic on the UK e-commerce market and how the pandemic changed the trend in e- commerce market. The study also tries to forecast the effects of covid -19 on the retail sector of e-commerce market in the United Kingdom to analysis makes predictions about its future effects on the retail sector of the e-commerce market.

Literature Review

E- commerce refers to a transaction that involves the electronic purchase or sales of goods over the internet or online services. (Wikipedia Contributors, 2019) online shopping is a part of e- commerce where in customers shop on the internet from businesses directly over the internet web.

In the digital age, consumers have access to a broader range of merchandise and services on the internet from domestic and international retailers. The relationship between internet shopping and high street shopping habits is complex and evolving. The perceptions of online shopping have changed over time as a result of factors such as comfort and freedom, price, and access to information about the products available. This has led to a shift in popularity towards online shopping. (Pyne, 2014)

E-commerce makes up a substantial portion of the market in Europe as a part of Gross Domestic Product (GDP), with Britain having the highest rate at 7.9 %. In Europe, the United Kingdom serves as the hub of e- commerce. Eighty-one per cent of all European internet sales are made in the United Kingdom, Germany, and France. This indicator demonstrates consistent increase over time. The global sum of e-commerce sales is anticipated to more than treble as a result of the COVID-19 epidemic. (Fedushko et al., 2021). United kingdom has the highest percentage of population who use e- commerce's platform for purchases, with over $120 billion in e- commerce revenue in the United Kingdom , 82 % of United Kingdom residents participating in at least one online purchase in the year 2021. (International Trade Administration, 2022). E- commerce platforms may gain more insight about customer behaviour by analysing user related product information. Furthermore, big data technology's analytical capabilities will allow companies to optimise customer service, product inventories and customer services and also assist the businesses in crucial management choices. (Liu, 2021).

Developing a strong e- commerce strategy requires integrating business strategy with data analytics. Forecast of global trends and statistics show that conducting of business online is becoming increasingly inevitable. (AbuRaya, 2020). In their relentless pursuit of thriving during the pandemic, brick and mortar stores or owner managed traditional businesses aimed at transitioning to e- commerce. Prior literature confirms that some of the transformational strategies employed by the businesses could not earn from this alternative revenue streams as desired. Two of the most common reasons attributable to this problem were as follows. One, lack of reliable data related to customers. Secondly, merchants failed to attract a wider customer base owing to website limitations. Studies have shown that businesses did not have sophisticated information models that could incorporate the data related to customers like their

spending habits, demographics. Merchants were unable to boost revenue owing to non-availability of refined data validation practices. (Teh et al., 2021).

The study by Kirby-Hawkins et al. (2018) look at the geography of corporate e- commerce sales in the United Kingdom food sector, focusing on the importance of geographic location and the distribution of online sales across different regions. The authors analysed the market share of several e- commerce merchants in different regions of the United Kingdom using data from the Nielsen Retail Index. They discovered that e- commerce sales are largely concentrated in urban areas, with London and the Southeast of England accounting for the lion's share of the market. The authors also discovered that different merchants' market shares varied dramatically across different regions, with some shops having a strong presence in some locations but a lower presence in others. (Kirby-Hawkins et al., 2018)

In this report, the shift of retail sales to e- commerce platform in the United Kingdom will be forecasted and summarised and also an exploratory data analysis of the United Kingdom's e- commerce market will be conducted to decide if it is profitable avenue of conducting business.

An overview of the methodology

The data used for the analysis is a sequence of data points collected over a period of time is using this specific method, forming a time series. In time series analysis, data points are recorded at consistent intervals over a specified period of time, rather than just being recorded arbitrarily (Tableau, n.d.). In order to analyse the time series and predict future values, an Autoregressive Integrated Moving Average model is used.

An Autoregressive Integrated Moving Average model is an autoregressive integrated moving average model that can be used for forecasting and back casting. (Box et al., 2016). Autoregressive Integrated Moving Average models are also known as the 'BoxJenkins' approach due to the work of Box and Jenkins . (Wikipedia Contributors, 2020).A general Autoregressive Integrated Moving Average (Autoregressive Integrated Moving Average) Model is a combination of a three process parts models. The general equation of Autoregressive Integrated Moving Average (Ho & Xie, 1998) can be expressed as

$$\mathbf{y'(t) = c + \phi 1^* y'(t-1) + \cdots + \phi p^* y'(t-p) + \theta 1^* \varepsilon(t-1) + \cdots + \theta q^* \varepsilon(t-q) + \varepsilon t}$$

where, the trend component is denoted by **c**
the order of autoregressive model, (denoted by 'p') equals { $\phi 1, \phi 2, \ldots, \phi p$ }
the order of moving average model, denoted by 'q') equals { $\theta 1, \theta 2, \ldots, \theta q$}
random errors are denoted by { $\varepsilon t, \varepsilon(t-1), \ldots, \varepsilon(t-q)$}

There are 3 parts of Autoregressive Integrated Moving Average model are the Auto regressive part or the AR component, Integrated part (I) and the Moving Average component (MA). (Nau, 2019).
These are defined as below:

AR: The Auto Regression: ('p')
The fundamental premise of the auto regressive model holds that any stationary time series can be derived by taking its past values and errors into account. It is permissible to allow a defined lag of p for all past values along with an error term εt. (Office for National Statistics, 2008). In the above equation these past values are represented by y(t - 1), y(t - 2) and p stands for the order.

I: Integration ('d'):
Every observation must be uniquely recognised in the time series in order to perform a time series analysis. The data that is nonstationary must be converted to stationary. The series must be differentiated

to attain stationarity (unless it was stationary to begin with) and the number of times a series is differentiated to attain stationary time series is called order of differentiation represented by d.

MA: Moving Average: (‘q’)
The primary premise underlying the MA portion (denoted q) of a time series is that the observed value is composed of a random error term and then linearly combination of earlier random error terms up to a given maximum lag . Regression is performed on the time series using residuals from previous observations, that is, error $\varepsilon(t-1)$, error $\varepsilon(t-2)$ and so on.

In the above equation y′t represents the differenced series and the first AR term's coefficient is $\phi\mathbf{1}$, the first MA term's coefficient is $\boldsymbol{\theta}\mathbf{1}$ and so on. The Autoregressive Integrated Moving Average model is denoted as Autoregressive Integrated Moving Average(p,d,q), but this is a general model of Autoregressive Integrated Moving Average model. SARIMA or seasonal Autoregressive Integrated Moving Average is an extended model of Autoregressive Integrated Moving Average model used when dealing with timeseries having seasonality. Autoregressive Integrated Moving Average (p, d, q)(P, D, Q) is the general notation of a seasonal Autoregressive Integrated Moving Average model, where the non-seasonal order of the AR,I,MA TERMS is indicated given by p, d, q respectively, the seasonal order of AR, I ,MA is represented by P, D, Q. the equation below is the general form of SARIMA model (Chatfield, 2004)

$$\boldsymbol{\Theta(L)^{p}\,\theta(L^{S})^{P}\,\Delta^{d}\,\Delta^{D}\,s\,y_{t} = \Phi(L)^{q}\,\phi(L^{s})^{Q}\,\Delta^{d}\,\Delta^{D}\,s\,\epsilon_{t}}$$

Where,
p, q are order of non- seasonal Autoregressive Integrated Moving Average ,
d , D is number of times differencing required and seasonal difference respectively.
Back shift operator is L, that defines seasonal operator of AR and MA.
Δd and ΔDs are seasonal and nonseasonal differencing operators respectively. (Chang & Liao, 2010)

The p value is obtained by analysing the partial autocorrelation plot and q value is determined by analysing the auto correlation plot. Since the order of Autoregressive Integrated Moving Average model is chosen based on human analysis, Akaike’s Information Criterion (AIC) test can be used to determine if the chosen model fits the data and it is explained by the formulae:

$$\mathbf{AIC\ (p) = nln\ (RSS\ /\ n) + 2K}$$

Where in , n is the number of datapoints and RSS is residual sum squares. (Chen et al., 2018). The test works on the criteria that the lower AIC value the better fitting the Autoregressive Integrated Moving Average model.In this study, the time series approach based on Autoregressive Integrated Moving Average models is investigated for repairable system failure analysis and forecasting. Autoregressive Integrated Moving Average model achieves more accurate forecasts as compared to the Duane model. (Ho & Xie, 1998) and Autoregressive Integrated Moving Average model is also a promising alternative to Holt-winters model (Office for National Statistics, 2008). The model uses it past values that are the data’s lags and lags of predicted errors to generate a forecast. With the advent of sophisticated and automated packages in python such as statsmodels.graphics.tsaplots , statsmodels.tsa.arima.model and other statistics tools, the execution of complex formula od Autoregressive Integrated Moving Average are made easy and more accurate.

Data

The data used for this report is majorly obtained from the Office for National Statistics and Statista. The data used to forecast the the retail sales over the internet in the United Kingdom as a percentage of the total retail sales in the United Kingdom, that is to predict how impactful the retail sales are being

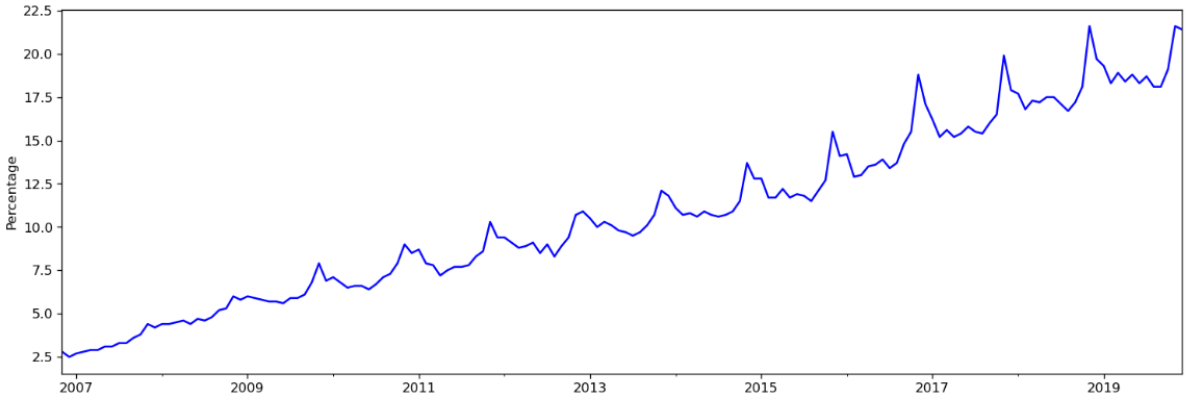
impacted by the e- commerce was taken from form the Office for National Statistics website. from the work sheet INTERNET of the dataset 'Retail Sales Index internet sales'. (Office for National Statistics, 2022). The data about detailed e commerce sales in the United Kingdom has also been obtained from the Office for National Statistics website from the dataset titled e- commerce and ICT activity. (Office for National Statistics, 2021). All the details relating to the e- commerce sales in the united kingdom and further calcification by business and sales using e- commerce within the united kingdom is used to perform an exploratory data analysis. All the data in this data set ranges over the time period from 2007 to 2019 and 2009 to 2019. The data set titled Internet access - households and individuals is also extracted from the office for national statistics. This dataset provides the data about how access of internet trends was in the United Kingdom for the years 2009 to 2020 and how the people of United Kingdom have been using the e commerce applications on the internet. (Office for National Statistics, 2020). The data from Statista is also used in the analysis which is a survey results of people attitude towards online shopping post covid 19 pandemic and this survey was conducted by a company called Appinio. (Tighe, 2021). All the data used in this report is open sourced and publicly available. The data has been extracted and used as smaller files, but no data was manipulated with intention of generating a favourable outcome. The data was handled with adherence to the United Kingdom Government Data Ethics Framework.

Data analysis and Findings

Forecasting the proportion of retail sales on internet in the United Kingdom.

The dataset consists of data of retail sales in the United Kingdom, more specifically the total retail sales information and values of retail sale over internet and Internet retail sales as a percentage of total retail sales for the time periods of November 2006 to November 2022. From this data, the data for time period November 2006 to December 2019 is extracted to simulate a forecasting model to predict how much per cent of retail sales will shift to e commerce platforms.

Internet sales as a percentage of total retail sales (November 2006- December 2019)

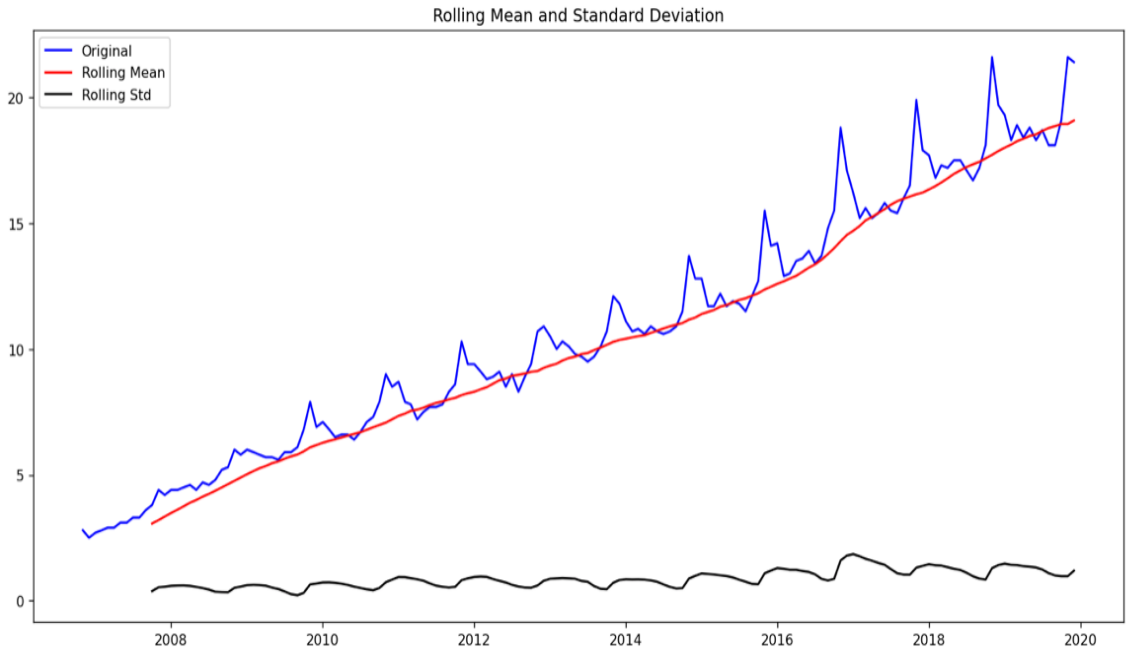


Checking stationary of the time series:

Rolling statistics:
The above line graph shows the retail sales over the internet in the United Kingdom as percentage of total retail sales in the United Kingdom over the time period of 13 years from 2006 to 2019. This percentage of online retail sales has increased from 2.8 % in November 2006 to 21.4 % in December 2019.

Since the data is plotted over continues months, forming a time series, the statistical model of **autoregressive integrated moving average** (**Autoregressive Integrated Moving Average**) is being used to forecast how much per cent of retail sales will be conducted over internet in the United Kingdom by the year 2024.

The plot below shows the rolling mean and rolling standard deviations, which are calculated for an window of 12 (months), to illustrate a trend in the taken data for every year. The increasing in rolling mean and standard deviation show that there is a trend that exists in the given time series.



Augmented Dickey– Fuller Test (or ADF test).:

Now to check if the time series is stationary or non- stationary, we employed another statistical test called Augmented Dickey– Fuller Test (or ADF test). This Augmented Dickey– Fuller Test tests the null hypothesis which in this case checks if the given time series is non-stationary. The p – value, if it should be less than or equal to a significant value (i.e. 0.05) for the given time series, then we can reject the null hypothesis . The results of Augmented Dickey–Fuller Test for the time series which shows the percentage of retail sales generated via the Internet as compared to total retail sales are shown below:

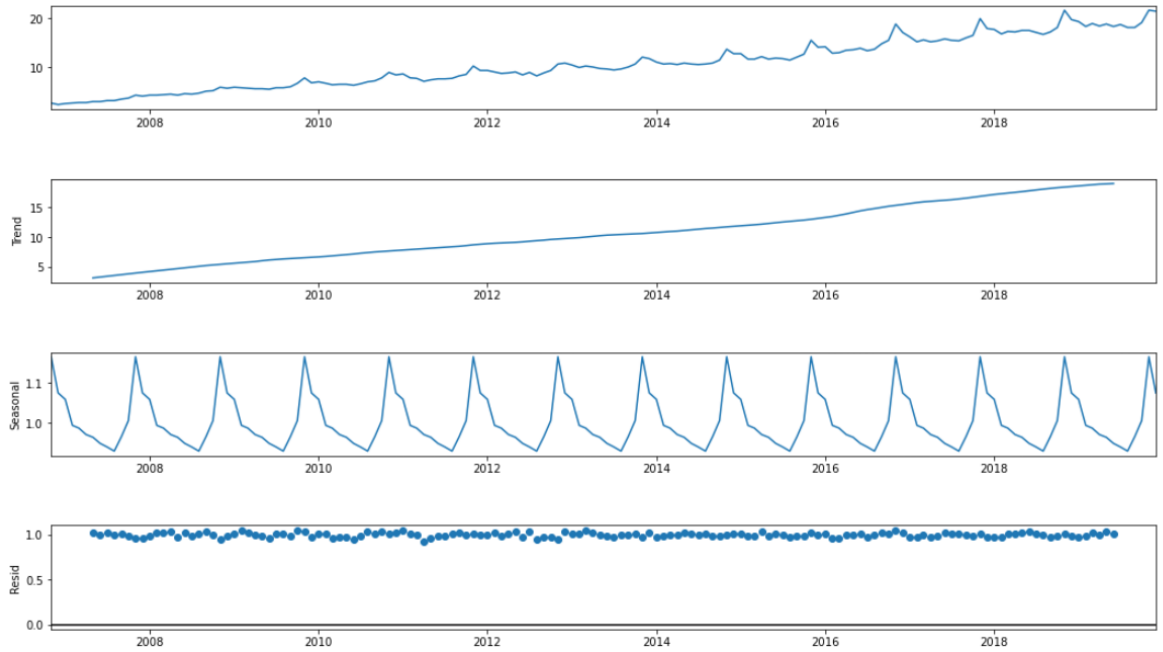
```
      Results of dickey fuller test
Test Statistics                  0.755012
p-value                          0.990874
No. of lags used                14.000000
Number of observations used    143.000000
critical value (1 %)            -3.476927
critical value (5 %)            -2.881973
critical value (10 %)           -2.577665
                dtype: float64
```

Since the p – value is 0.990874 which is greater than the significant value, the test is in favour of null hypothesis and it can be concluded that the data is very possibly non- stationary.

Checking trends and seasonality components of the time series:

To analyse the time series values the trend and seasonality in the time series should be separated to make the time series stationary. We use seasonal _ decompose function from stats model in python to

represent the time series as a product of trend component, linear component, Seasonal component and residuals. (Lewinson, 2022). The below are the decomposed plots of the given time series:

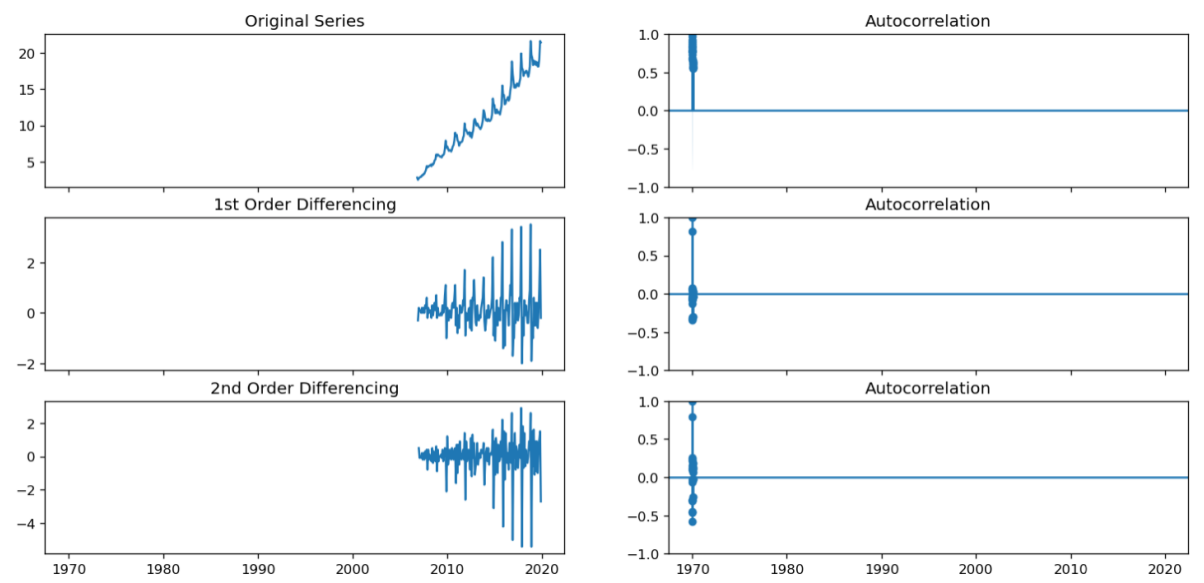


The above decomposed graphs of the time series show that there is an increasing trend that exists in the time series. The seasonality plot also shows uniform and repetitive short term patterns for every 12 months or for every year.

Autoregressive integrated moving average (Autoregressive Integrated Moving Average) model:

An Autoregressive Integrated Moving Average model is based on 3 variables namely p,d,q. 'p' is the order of Auto Regressive(also called AR term). 'q' is the order of moving average(also called MA) term and d is the order of differentiation need to make the time series stationary.

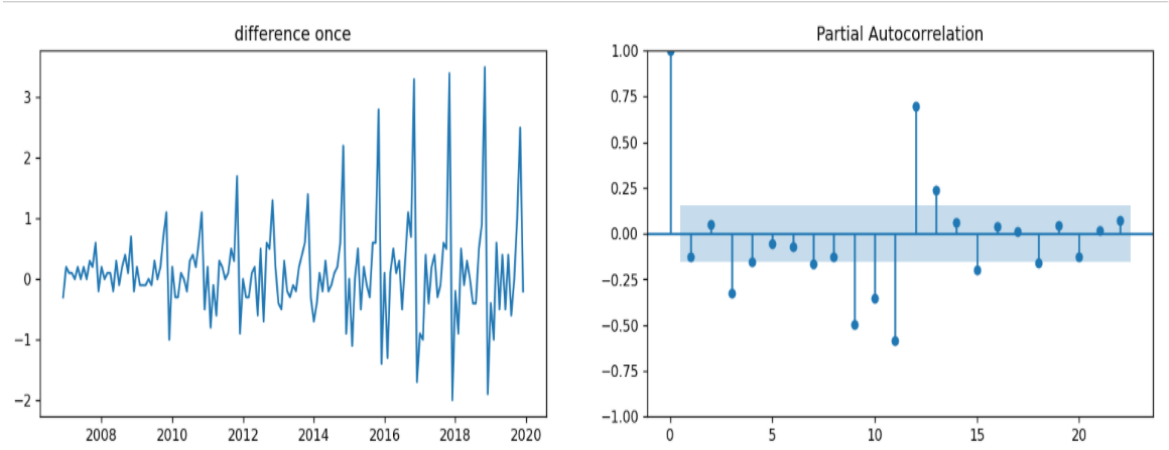
To find the term 'd' , the time series is differentiated using .diff function in python

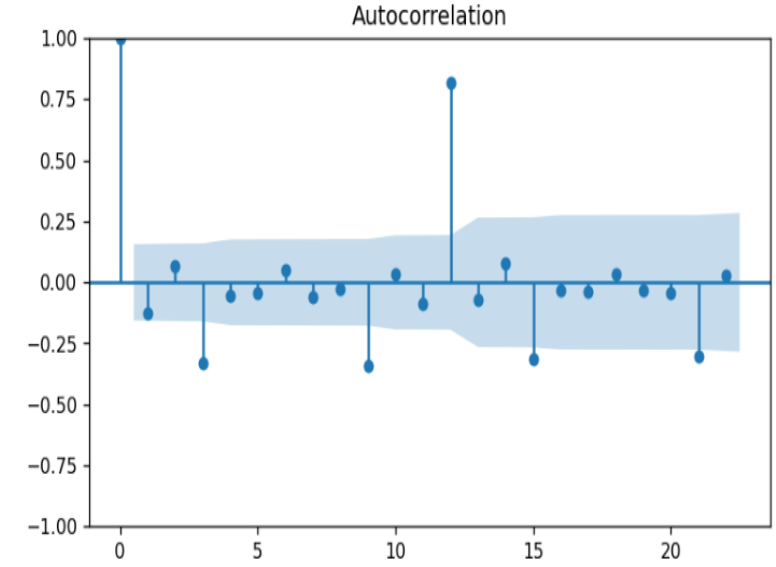


The second order of differentiation of the time series is distributing the mean values in the same pattern as first order of differentiation but only pushing it to higher negatives. Hence, the optimal order of differentiation to make the time series to make it stationary will be **1.**

The p is used to calculate the lag in the time series so it can be used as a predictor and the q term is used to measure the forecast errors in that are needed to be input into the Autoregressive Integrated Moving Average model. The p value is calculated using a partial autocorrelation plot. In python we use the function plot _ pacf from statsmodels.graphics.tsaplots packages to get the p value.

For the given the retail sales over the internet in the United Kingdom as percentage of total retail sales in the United Kingdom time series we can see from the below partial autocorrelation plot, p can be fixed as 0





For the term q the plot _ acf is used from statsmodels. graphics.tsaplots packages of python to generate autocorrelation plot. Stationaries series with autocorrelation are removed by removing an appropriate number of MA terms from them. So, the order of moving average, after inspection of the autocorrelation graph can be taken as 1

Hence for the the Autoregressive Integrated Moving Average model for United Kingdom Internet retail sales as a percentage of overall retail sales in United Kingdom will be (p,d,q)=(0,1,3). Executing Autoregressive Integrated Moving Average model using these values we get the following results.

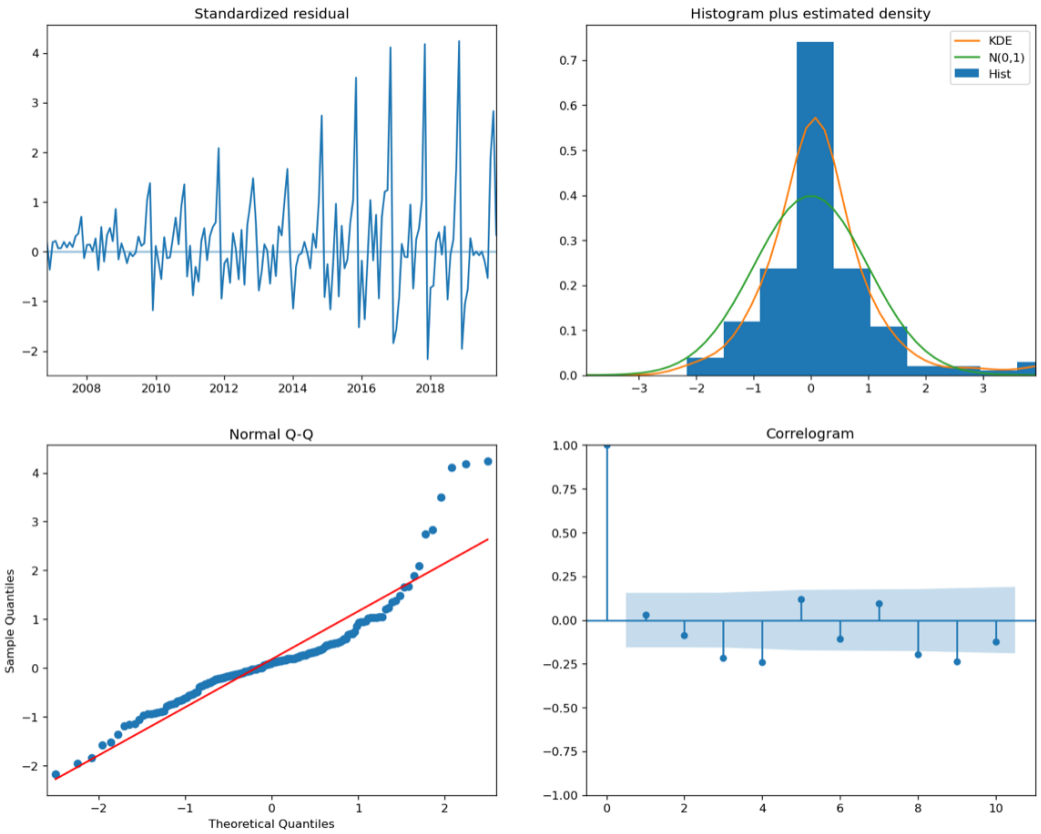
```
                                SARIMAX Results
==============================================================================
Dep. Variable:                  value   No. Observations:                  158
Model:                 Autoregressive Integrated Moving Average(0, 1, 3)   Log Like
lihood                -180.027
Date:                Thu, 05 Jan 2023   AIC                            368.054
Time:                        17:57:41   BIC                            380.279
Sample:                    11-01-2006   HQIC                           373.019
                         - 12-01-2019
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.1433      0.074     -1.926      0.054      -0.289       0.003
ma.L2          0.0647      0.100      0.646      0.518      -0.132       0.261
ma.L3         -0.3428      0.154     -2.223      0.026      -0.645      -0.041
sigma2         0.5786      0.045     12.833      0.000       0.490       0.667
===================================================================================
Ljung-Box (L1) (Q):                   0.24   Jarque-Bera (JB):               413.67
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               7.34   Skew:                             2.09
Prob(H) (two-sided):                  0.00   Kurtosis:                         9.76
===================================================================================
```

The above Autoregressive Integrated Moving Average has an AIC score of 368.053 and the time series is now stationary since all the p values of all the coefficient of AR,MA are given. The Autoregressive Integrated Moving Average model has a skewness of 2.09 and a Kurtosis of 9.76

Running Diagnostics on the Autoregressive Integrated Moving Average model

The plot_diagnostics function is used to check the accuracy of the Autoregressive Integrated Moving Average model.The standardized residual plot show that the mean of the plot is not zero and the Histogram plus estimated density plot data is distributed around zero but the KDE and N(0, 1) are not very similar.

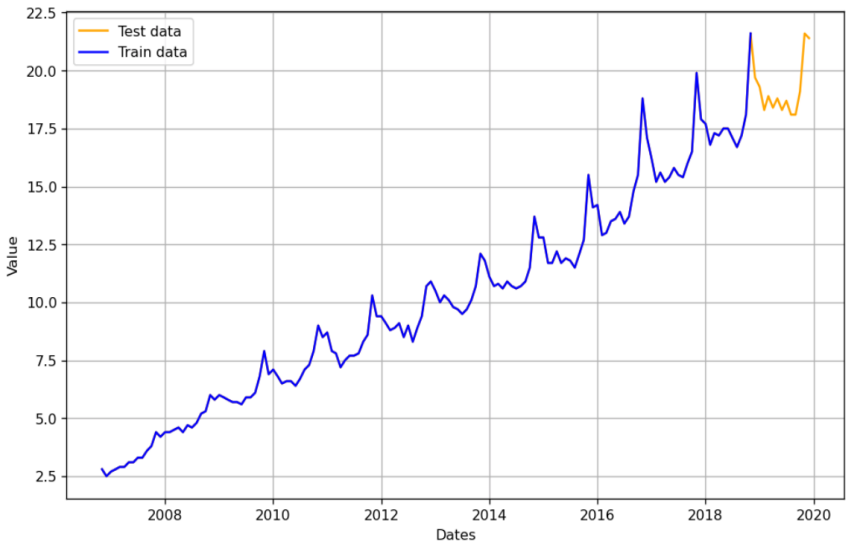
Normal Q - Q graph shows that probability of data point lying on the line is very less and Correlogram (ACF plot) shows that that many data points lie outside the confidence interval suggesting a need to add more predictors to increase the model accuracy.



Splitting timeseries into training and testing data

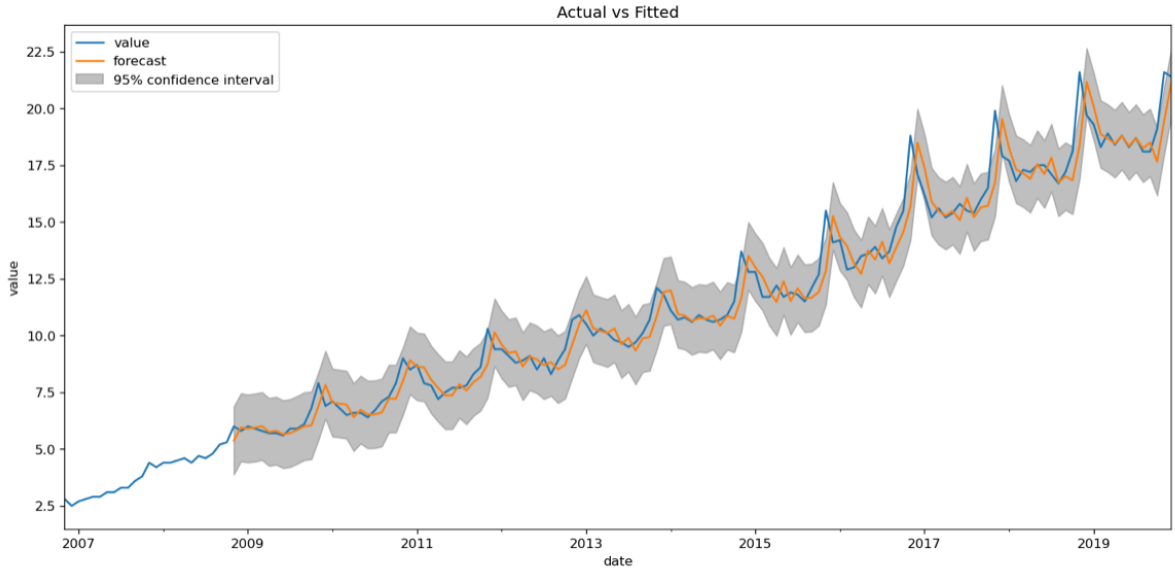
The data in the time series is split into two separate but continuous data fames to train and test the prediction model. The training data frame is made up of 145 values from the November 2006 to November 2019 and the testing data is made up of 13 values from December 2018 to December 2019.

Splitting time series into train and test data



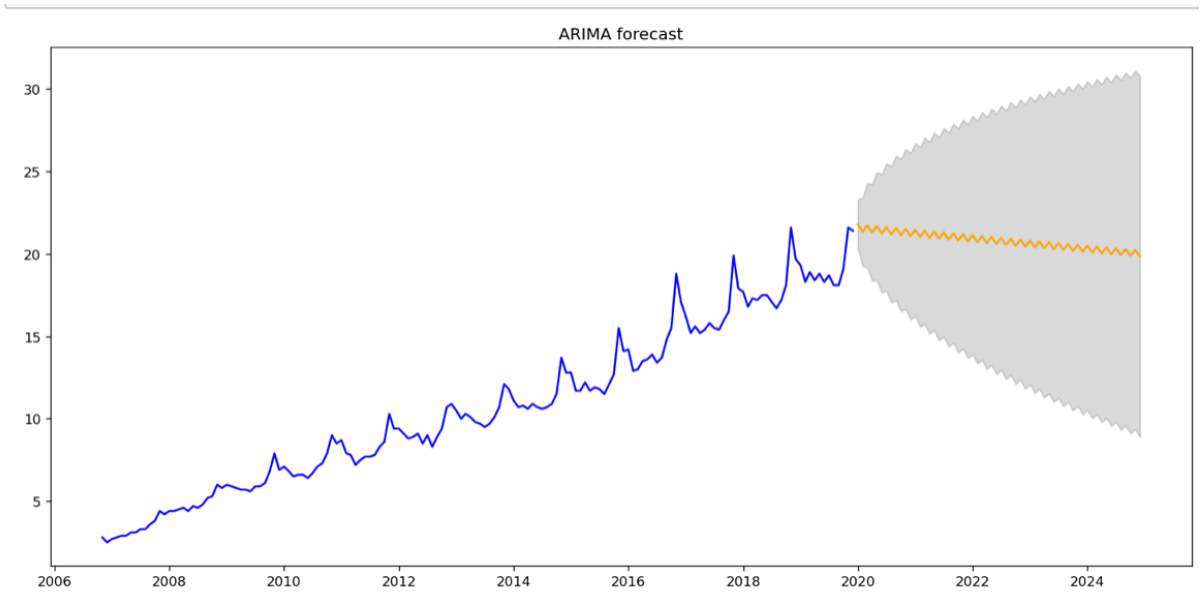
Fitting the Autoregressive Integrated Moving Average forecast graph on to the original graph

The graph illustrates the use of Autoregressive Integrated Moving Average function to train the data to predict the values and with order equalling to (0, 1, 3) and check if the model fits the actual plot. The plot _ prediction function is executed to observe if the model fit the actual data.



Forecast graph

The graph below shows the retail sales over the internet in the United Kingdom as a percentage of the total retail sales in the United Kingdom forecasted over net 60 month time period. This period covers predictions for the years 2020 to the year 2024, both inclusive.



As we can see from the forecast plot above, the graph does not show predicted values in detail and does not meet the objective of the forecast. Addition of the seasonality factor to the timeseries will give more accurate predictions.

Introducing seasonality component with SARIMA model

An seasonal autoregressive integrated moving average model called SARIMA model is executed on the time series. An SARIMA model is an Autoregressive Integrated Moving Average model which also uses the seasonal component. (Graves, 2020). The 'pmdarima' package in python is used to auto generate the most optimal order and seasonal order for the given time series. The summary of results achieved by executing auto_ arima on the data are tabulated below:

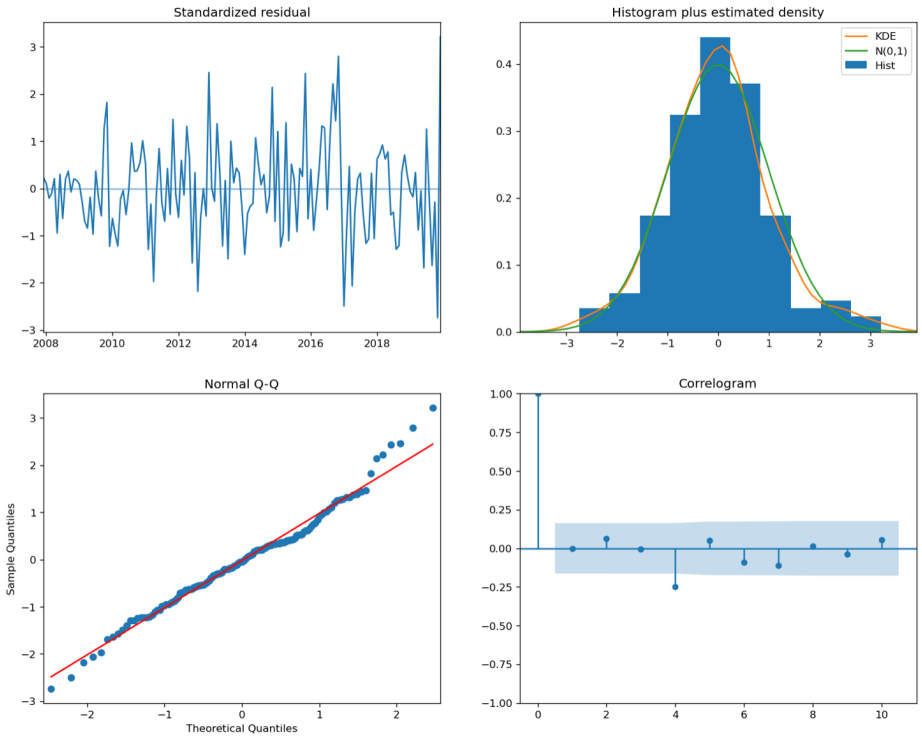
```
                                     SARIMAX Results
==========================================================================================
==
Dep. Variable:                                     y   No. Observations:                  1
58
Model:             SARIMAX(0, 1, 1)x(0, 1, 1, 12)   Log Likelihood                -59.9
35
Date:                            Thu, 05 Jan 2023   AIC                            125.8
69
Time:                                    18:10:32   BIC                            134.7
99
Sample:                                11-01-2006   HQIC                           129.4
98
                                     - 12-01-2019
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5196      0.059     -8.866      0.000      -0.634      -0.405
ma.S.L12      -0.2950      0.081     -3.645      0.000      -0.454      -0.136
sigma2         0.1325      0.014      9.571      0.000       0.105       0.160
==============================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 6.25
Prob(Q):                              0.98   Prob(JB):                         0.04
Heteroskedasticity (H):               2.62   Skew:                             0.30
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.83
==============================================================================
```

The proposed optimal order of the auto SARIMA model is (p, d, q) = (0, 1, 1) and the seasonal order is (0, 1, 1, 12).

Diagnostic test on SARIMA model

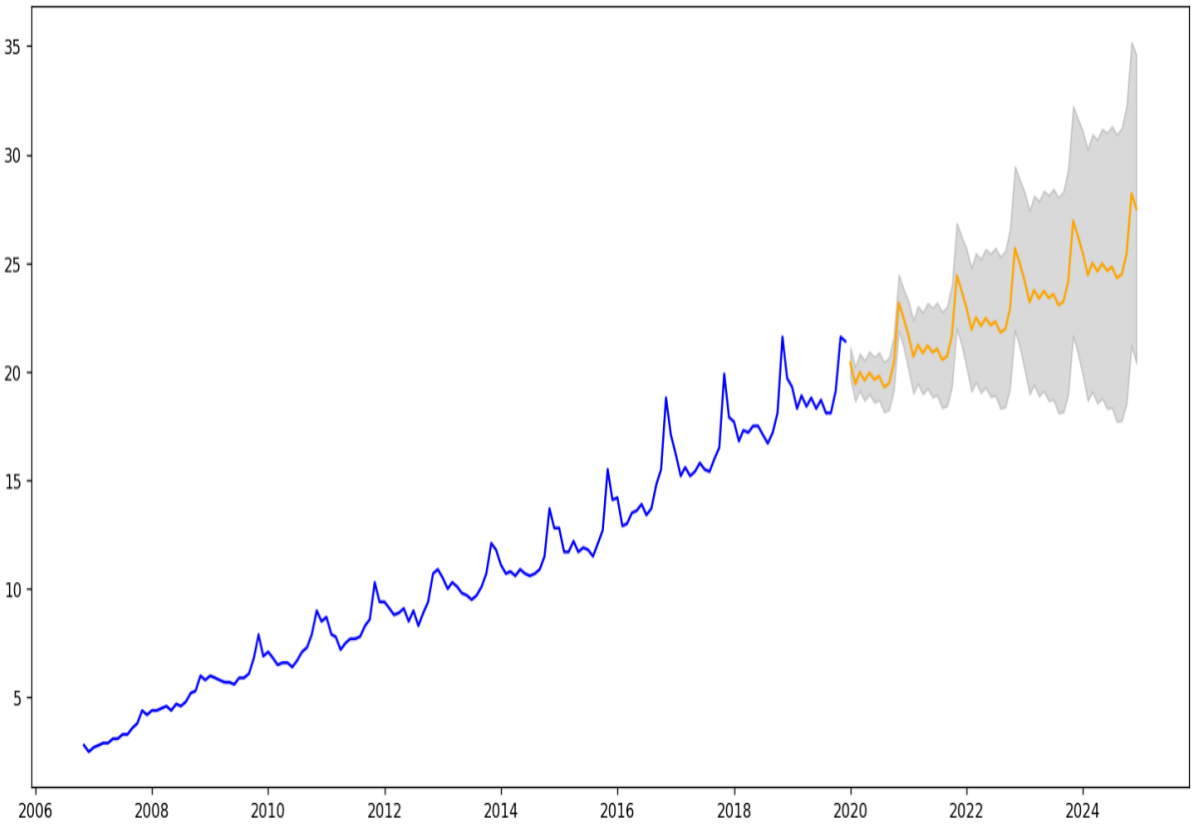
The Standardized residual graph illustrates a much greater consistency in than the Autoregressive Integrated Moving Average model which didn't consider the seasonality of the time series. Standardized residual across the graph is altering between a – 3 and +3 making it more probable to have a mean closer to zero, indicating that the data is now more stationary. The Histogram plus KDE estimate also shows much better results with the KED curve much closer to the normal distribution curve

Normal Q – Q plot presents a much more restricted clustering of data points on the line and the **Correlogram (ACF plot) depicts almost** all values are now within the band of confidence interval. These results prove that the Autoregressive Integrated Moving Average model became much more accurate once seasonality component was added to it.



Forecasting SARIMA model

Autoregressive Integrated Moving Average forecast with seasonality component



Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors, is used on the testing timeseries data to make more precise prediction. The following results are tabulated by deploying the SARIMAX model:

SARIMAX Results

Dep. Variable:	value	No. Observations:	145
Model:	SARIMAX(0, 1, 1)x(0, 1, 1, 12)	Log Likelihood	-47.366
Date:	Thu, 05 Jan 2023	AIC	100.732
Time:	18:11:58	BIC	109.381
Sample:	11-01-2006 - 11-01-2018	HQIC	104.247
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.4458	0.084	-5.286	0.000	-0.611	-0.280
ma.S.L12	-0.2961	0.083	-3.566	0.000	-0.459	-0.133
sigma2	0.1188	0.014	8.460	0.000	0.091	0.146

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	1.72
Prob(Q):	0.93	Prob(JB):	0.42
Heteroskedasticity (H):	2.29	Skew:	0.19
Prob(H) (two-sided):	0.01	Kurtosis:	3.41

Warnings:

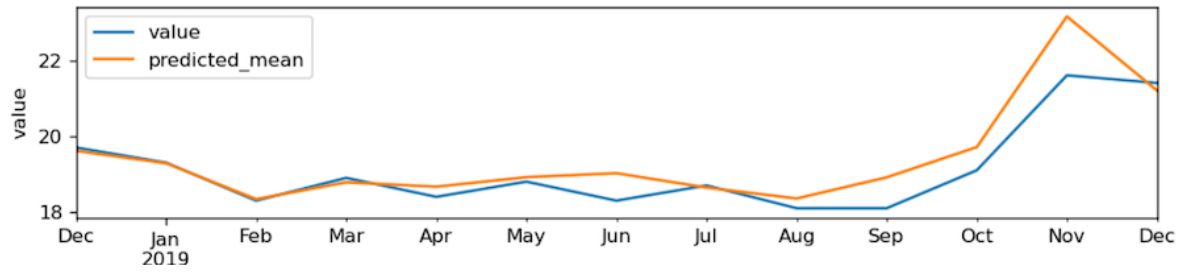
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

The model has an AIC score of 100.732 and all the p values of Autoregressive Integrated Moving Average coefficients indicate the timeseries is stationary with a skewness of 0.19 and kurtosis of 3.41. the following are the generated prediction values and expected values for the test data of the time series.

Predicted vs expected values:

Date	Predicted	Expected
2018-12-01	19.61	19.70
2019-01-01	19.29	19.30
2019-02-01	18.34	18.30
2019-03-01	18.78	18.90
2019-04-01	18.67	18.40
2019-05-01	18.92	18.80
2019-06-01	19.02	18.30
2019-07-01	18.64	18.70
2019-08-01	18.36	18.10
2019-09-01	18.91	18.10
2019-10-01	19.71	19.10
2019-11-01	23.15	21.60
2019-12-01	21.20	21.40

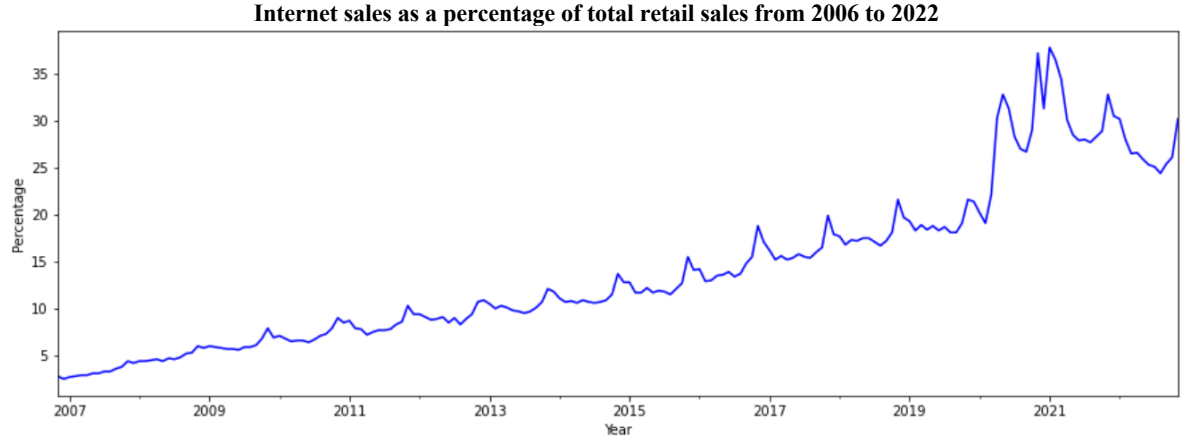
The plot of predicted value and the actual values is illustrated below. It can be seen that the predicted values are following the actual values, which indicated the forecasting model works.



The mean squared error of this model is MES Error: 0.32190004635

Extended Autoregressive Integrated Moving Average model for time series of percentage of retail sales on e- commerce platform considering actual values till November 2022:

The unexpected shift in increase of retails sales to the e commerce platform from the year 2020. Over the period of 16 years from 2006 to 2022, the following line graph illustrates the share of retail sales made over the internet in the United Kingdom as a percentage of total retail sales in the United Kingdom. This percentage of online retail sales has increased from 2.8 % in November 2006 to 30.2 % in November 2022.



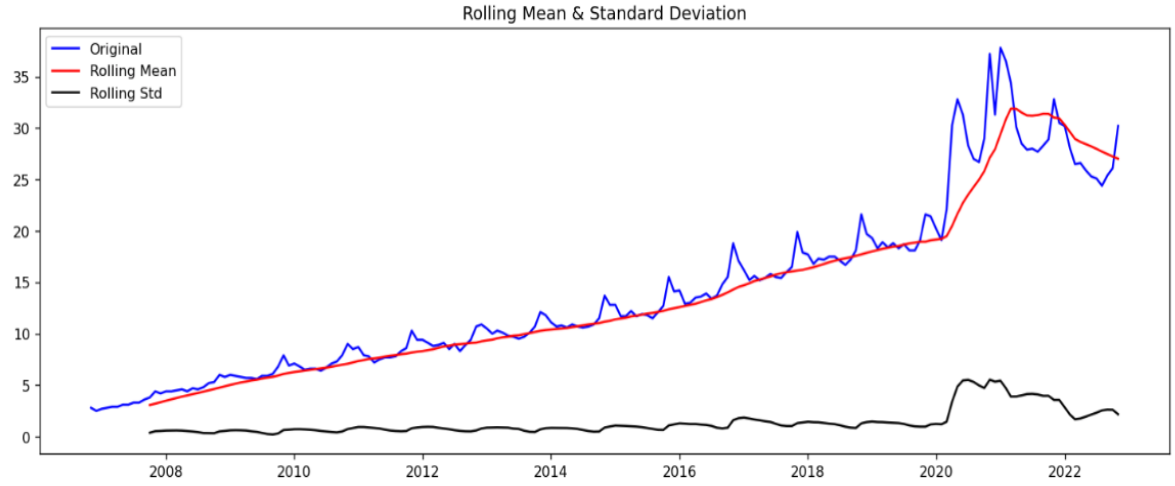
The rolling mean and standard deviation plot of the time series of proportion of sales of retail on the internet in the United Kingdom shows that the time series is not stationary. The results of Augmented Dickey–Fuller Test for the time series extended till year 2022 are shown below

```
       Results of dickey fuller test
Test Statistics                  - 0.707783
 p-value                           0.844723
 No. of lags used                 12.000000
 Number of observations used     180.000000
critical value (1 %)              - 3.467211
critical value (5 %)              - 2.877735
critical value (10 %)             - 2.575403
```

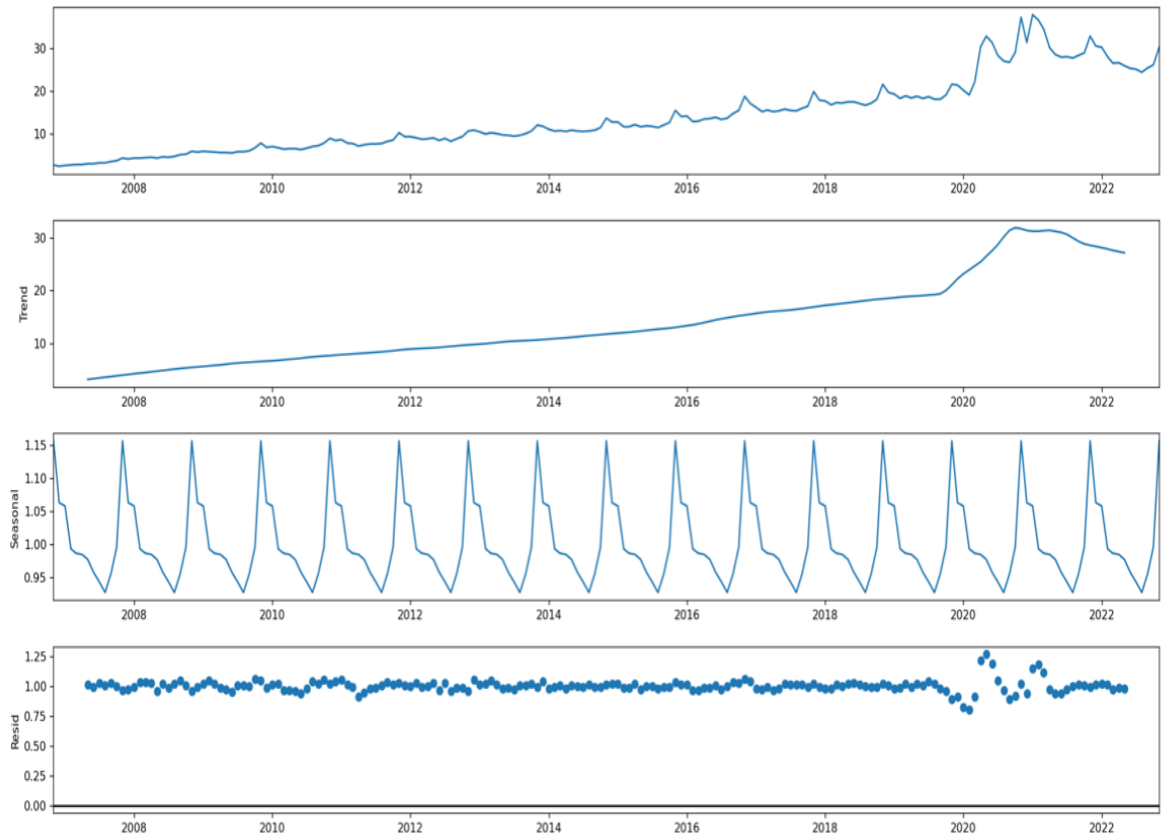
Since the p – value is 0.844723 which is greater than the significant value, the test is in the test is in favour of null hypothesis and it can be concluded that the data is very possibly non- stationary.

Calculating the rolling mean and rolling standard deviation:

The plot increasing rolling mean value and rolling standard deviation value signify there is a trend component in the time series.



The decomposed graph below shows that the seasonality of the extended timeseries is still uniformly varying annually but the trend of the series changed drastically around the year 2020.



The auto_ arima function is employed and seasonality of the timeseries which is very prominent is also considered in generating a SARIMA model. The optimal order of this series is generated as

(p, d, q) = (3, 1, 2) and seasonality order is compiled as (2, 0, 2, 12). The results of the SARIMA model are as follows:

SARIMAX Results

Dep. Variable:	y	No. Observations:	193
Model:	SARIMAX(3, 1, 2)x(2, 0, 2, 12)	Log Likelihood	-301.233
Date:	Thu, 05 Jan 2023	AIC	622.466
Time:	18:33:41	BIC	655.041
Sample:	11-01-2006 - 11-01-2022	HQIC	635.659
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	1.6370	0.083	19.738	0.000	1.474	1.799
ar.L2	-0.9919	0.078	-12.692	0.000	-1.145	-0.839
ar.L3	0.0385	0.045	0.852	0.394	-0.050	0.127
ma.L1	-1.6612	0.083	-19.917	0.000	-1.825	-1.498
ma.L2	0.9181	0.076	12.155	0.000	0.770	1.066
ar.S.L12	1.2925	0.290	4.462	0.000	0.725	1.860
ar.S.L24	-0.3414	0.321	-1.065	0.287	-0.970	0.287
ma.S.L12	-1.3681	0.318	-4.306	0.000	-1.991	-0.745
ma.S.L24	0.6998	0.333	2.103	0.035	0.048	1.352
sigma2	1.1884	0.106	11.209	0.000	0.981	1.396

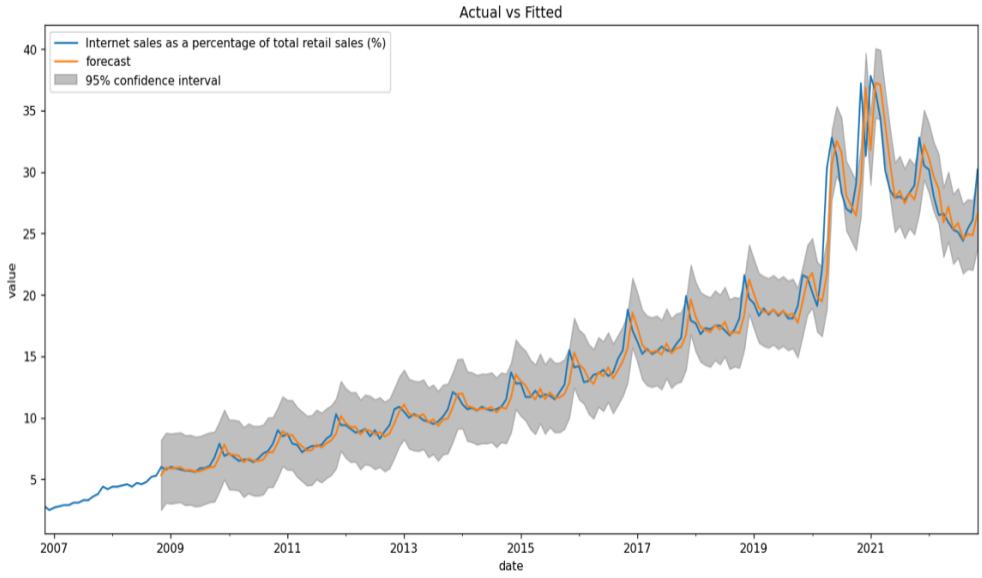
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	3887.49
Prob(Q):	0.97	Prob(JB):	0.00
Heteroskedasticity (H):	33.31	Skew:	2.77
Prob(H) (two-sided):	0.00	Kurtosis:	24.34

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step

The ACI score of this model is 622.466 and the series is skewed at 2.77 and kurtosis is 24.34 .

The time series data till year 2022 has been divided into two separate but continuous data sets again for the purposes of training and testing the prediction model. In the training data frame, 180 values are provided from November 2006 to October 2021, while in the testing data frame, 13 values are provided from November 2021 to November 2022.

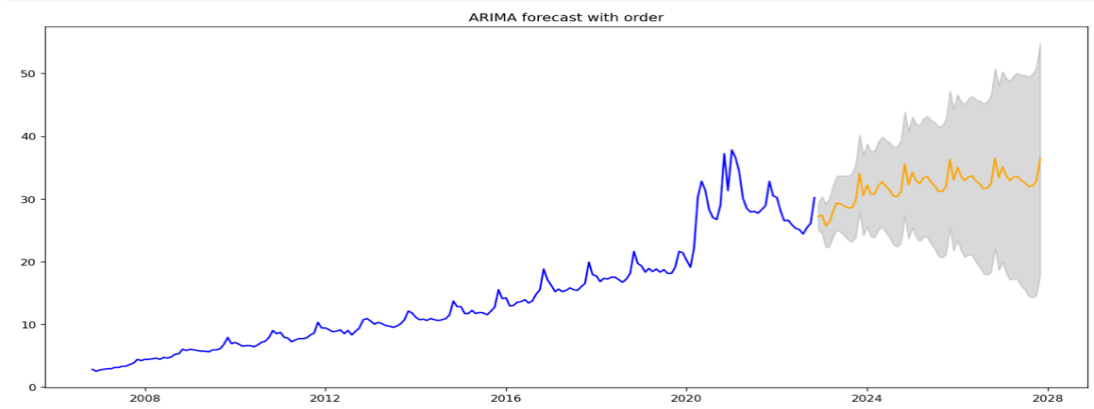
The Autoregressive Integrated Moving Average modelled trained graph is fitted on to the actual graph to check if the model fits the actual plot. And a forecast graph is generated as below.



Predicted vs expected values:

Date	Predicted	Expected
2021-11-01	30.83	32.80
2021-12-01	30.34	30.50
2022-01-01	28.12	30.20
2022-02-01	27.36	28.00
2022-03-01	29.34	26.50
2022-04-01	32.51	26.60
2022-05-01	34.13	25.90
2022-06-01	33.67	25.30
2022-07-01	32.84	25.10
2022-08-01	32.09	24.40
2022-09-01	32.03	25.40
2022-10-01	33.38	26.10
2022-11-01	37.88	30.20

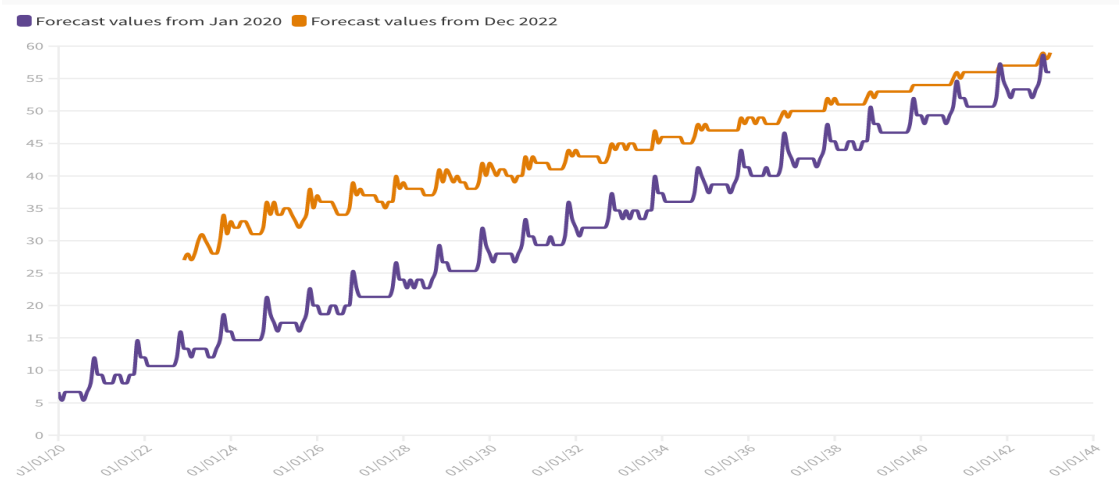
The mean squared error of this model is 28.216801163 and the forecast plot is as shown below. The forecast of values till the year 2028 are shown in graph below using predict function in python.



It can be observed that due to the sudden change in trend in the year 2020, the forecasted percentage of retail sales on the e commerce platform in the united kingdom have a significant difference. The table below collaborates this

Year	Forecast of yearly average percentage values based on the data till the year 2019 (monthly averaged per year)	Forecast of yearly average percentage values based on the data till the year 2022 (monthly averaged per year)
2020	20.42	27.94
2021	22.08	30.95
2022	23.67	24.48
2023	25.33	29.50
2024	26.83	32.50
2025	28.42	34.42
2026	30.08	35.75
2027	31.58	36.92
2028	33.25	38.17
2029	34.83	39.42
2030	36.42	40.67
2031	37.92	42.00

In the table above the values in red are the predicted values and the values in black are given values in the database. The column Forecast of yearly average percentage values based on the data till the year 2019 ‘ used the values in the dataset from under the column ‘internet sales as a per cent of the total retail sales (%)’ till the year 2019 and an autoregressive integrated moving average with seasonality component (SARIMA)is fitted on to the values till 2019 to forecast the future values till the year 2031. It is observed that the forecast shows that the retail sales on ecommerce platform as a percentage of total retail sales in the united kingdom will be an average 37.92 % (monthly) by the year 2031. However if the values in the dataset in the under the column ‘internet sales as a per cent of the total retail sales (%)’ are taken from November 2022 then the seasonal autoregressive integrated moving average (SARIMA) forecast fitted on to the data forecasts an average of 42 percent of total retail sales to be will be conducted on the online platforms.

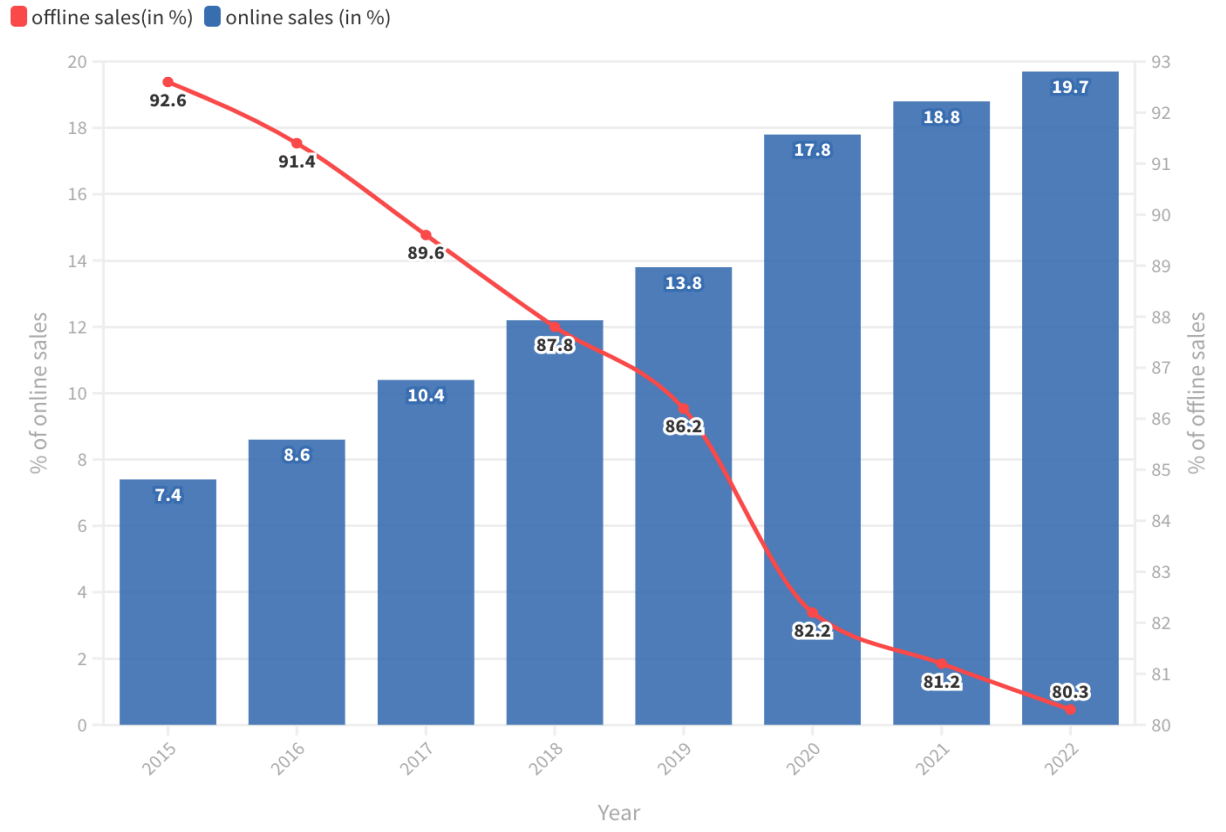


The line graph above show the forecast of monthly predictions of retail sales as an proportion of total retail sales. It can be observed that when the data from the year 2020 to 2022 is considered the prediction model estimates higher forecast values compared to when the data is considered only till 2019. This discrepancy can be explained by the sudden change in trend in the year 2020. The coronavirus pandemic is mostly likely the factor forced the online shopping of retail to change thus leading to the difference in both the models. However, the line graphs also indicates that the gap between the two forecasted values reduces as the values are forecasted further in time.

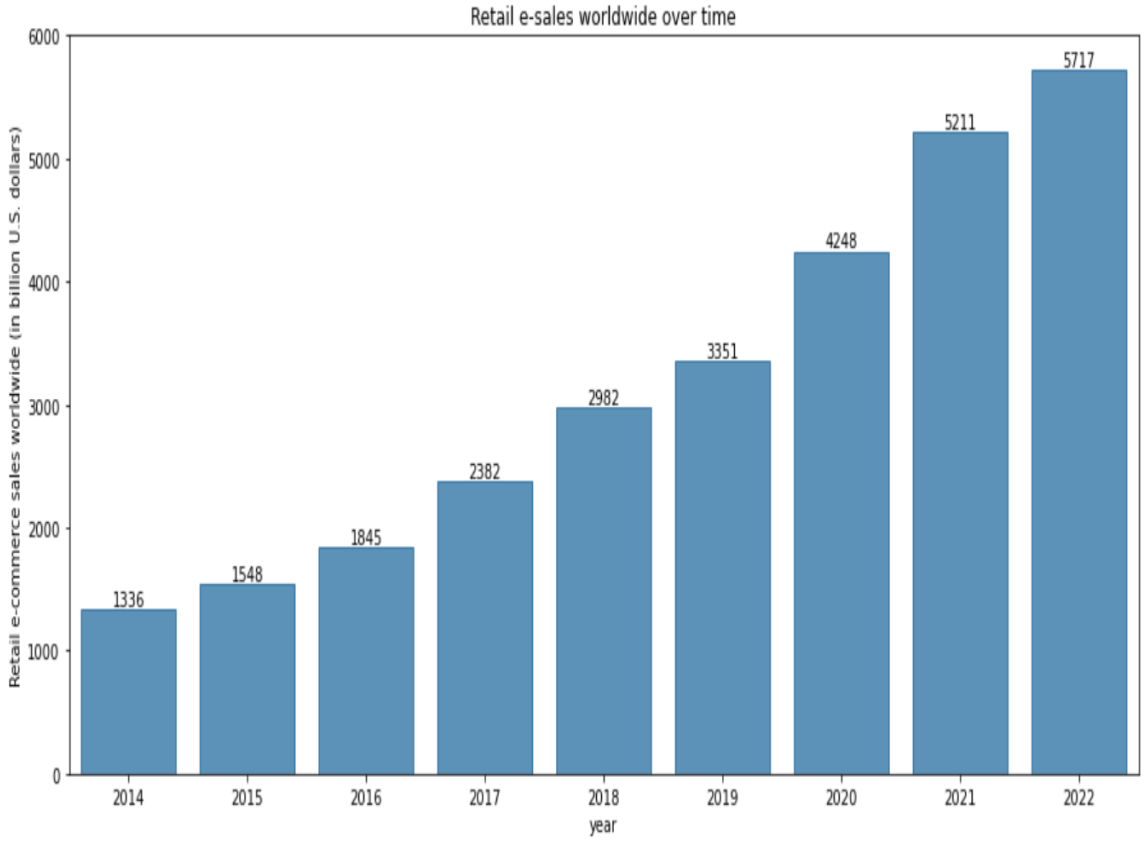
Exploratory Data Analysis (EDA)

As seen in the combined line bar graph below the ratios of online and offline retailing is changing over past few years, with the proportion of online sales increasing and the proportion of the offline sales decreasing, though they are the more conventional and contribute major stakes of overall stakes. The online sales were increasing constantly till 2018 but after 2019 we saw a noticeable increase in online sales.

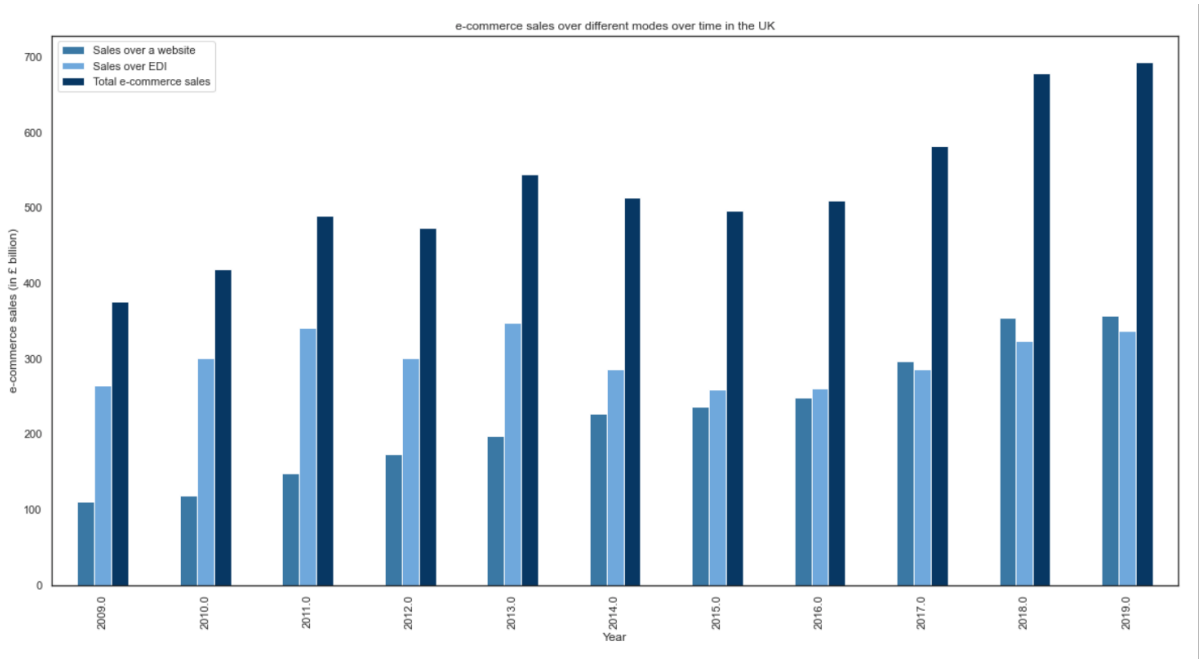
Online sales (in %) vs offline sales (in %) for years 2015 - 2022



The bar graph below portrays worldwide e commerce sales in retail industry in billion U.S dollars, for nearly the past decade. It can be observed that the sales have been constantly increasing from years 2014 to years 2022. As off year 2022 the retail e commerce sales are a market of $ 5717 billion U S dollars which is more than 4 time the value of sales, they were making in the year 2014.



The grouped bar graph below shows e – commerce sales in the United Kingdom over years 2009 to 2019. It also interprets the sales made over a website and sales over Electronic Data Interchange which make up the two streams of e – commerce sales. It can be seen that for the year 2019, the e commerce sales in the United Kingdom are making 693 billion pounds. Of these sales, the sales over website contribute 356.4 billion pounds and the sales over Electronic Data Interchange (EDI) contribute 336.6 billion pounds. The e- commerce sales in The United Kingdom nearly doubled in the past decade, i.e. they increased from 375.1 billion pounds in 2009 to almost 700 billion pounds in 2019. It can also be observed that in year 2009 the sales over Electronic Data Interchange (EDI) , 264.5 billion pounds contributed nearly 70 % of the total e- commerce sale in 2009 (375.1 billion pounds) while sales over websites merely contributed 30 % of the total e commerce sales.(£ 110.6 billions). This ratio gradually changed over years and in year 2017 the sales over website contributed more than the sales over EDI and by year 2019 it can be seen that the sales over website contributed over 50 % of the total sales by contributing 356.4 billion pound to the total 693 billion pounds of e- commerce sales. With the sales over EDI contributing the remaining.

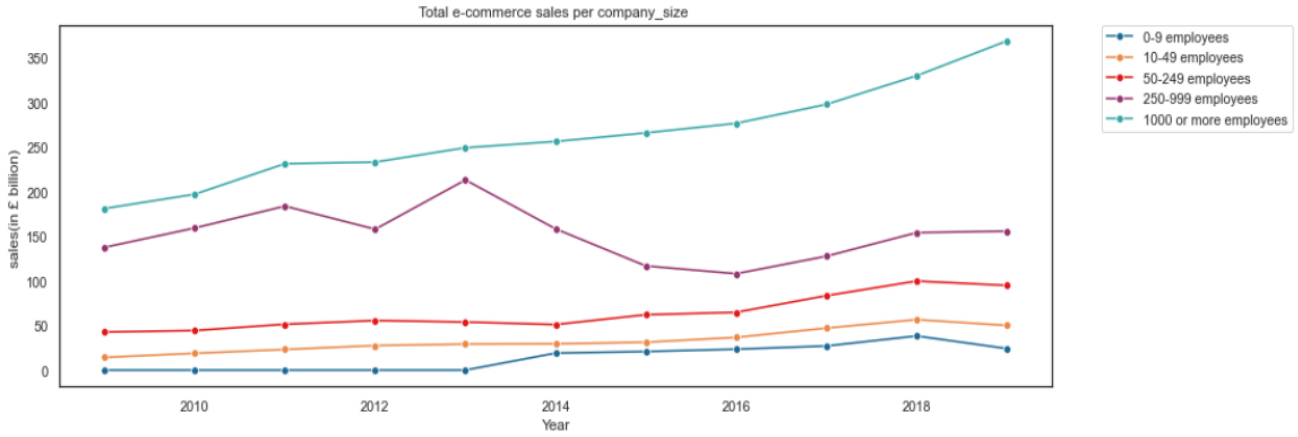


The e - commerce sales by the United Kingdom businesses can be categorized by the size of the business, which can be determined by the number of employees working in each of those company. The business sizes are usually classified as business with 10 to 49 employees, businesses with 50 to 249 employees. And the business having 250 to 999 employees and finally businesses with 1000 or more than thousand employees. However, it is not until in 2014 the survey introduced a concept of micro business hence adding a further classification of business having less than 10 employees.

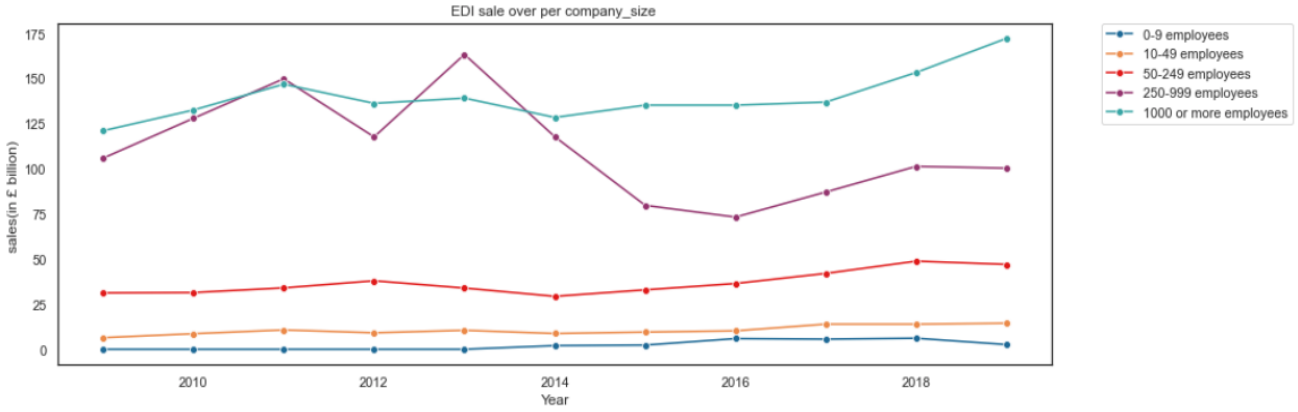
Usually, any business having less than 250 employees is classified as SMEs, i.e., small and medium - sized enterprise and in addition to micro businesses (a businesses with fewer than 10 employees), small and medium enterprises (SME) can be classified into two other types based on their employee counts, the small sized businesses which employees between 10 and 49 employees, and the medium sized businesses with between 50 and 249 employees. Any business is called a large enterprise if it employs at least 250 people. (OECD, 2017)

The below graph depicts the business size wise total e - commerce sales, and business size wise sales over Electronic Data Interchange (EDI) and over website throughout the United Kingdom for years 2009 to 2019.

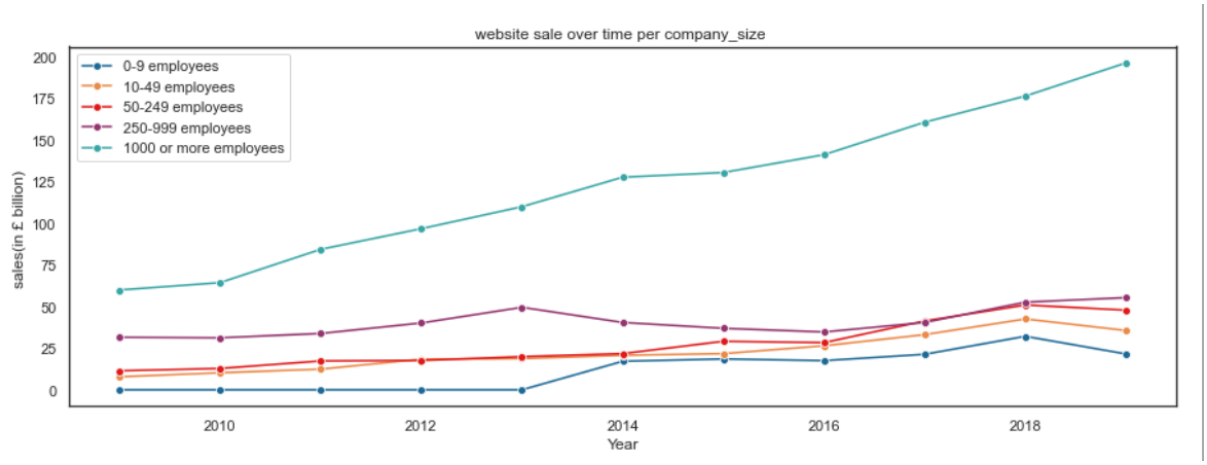
The total e commerce sales per company_size shows that despite the sizes all the business increased their sales over that decade. However, the most significant change can be seen in businesses of size having 1000 or more employees i.e., sales by these businesses increased from £ 180.9 billion pounds in year 2009 to £ 368.5 billion pounds in 2019. Though the rest all business showed increase in sales overall, the business of size 250 to 999 employees, increased their sales from £ 137.4 billion pounds in 2009 to £ 212.9 billion pounds in year 2013 but the sales fell to nearly £ 155.6 billion pounds in year 2019, making it the most nonlinear sized businesses sales wise. The two segments of companies having more than 250 employees have been generating more sales since 2009 compared to the ones employing less than 250 employees.



The graph EDI sales per company_size per time illustrates sales over Electronic Data Interchange with respect to company sizes and how they change over time. It can be observed that the overall sales increased only slightly or dropped below the value in 2009. The most variable plot of sales over EDIs can be observed in sales made by companies of size 250 to 999 employees. We see there is an increase in sale conducted over EDI by this size company in first half of the decade and companies having 249 to 999 employees saw the highest peak in sales in the year 2013 with sales of £ 163.3 billion pounds. By the year 2019 these sales dropped to a £ 100.3 billion pounds. The sales by companies with employees of 1000 or more increased from 121.0 billion pounds in year 2009 to 172.2 pounds in year 2019. Companies having 50 to 249 employees the sales the sales rose to only £ 47 million pounds from 31.2 billion pounds in that decade. The companies with 10 to 49 employees saw an increase of 8 billion pounds in sales since 2009 to 2019 and the micro enterprises only 30 % increase in sale in that decade.



The sales over websites saw a positive increase in sales in the decade 2009 to 2019. The companies having 1000 or more than 1000 employees saw a rapid increase in sales over websites with sales increasing from 110.6 billion pounds in 2009 to 334.9 billion pounds in 2019. The sales over websites in small and medium enterprise also increase over the decade with sales in micro sized companies increasing nearly 20 % and sales in companies sized 250 to 999 employees ranging from 31.5 billion pounds in 2009 to 55.3 billion pounds 2019. Sales by small enterprises increases from 7.8 billion pounds to 35.6 billion pounds. However, the medium enterprises saw an increase from 11.4 billion pound to 47.8 billion pounds.

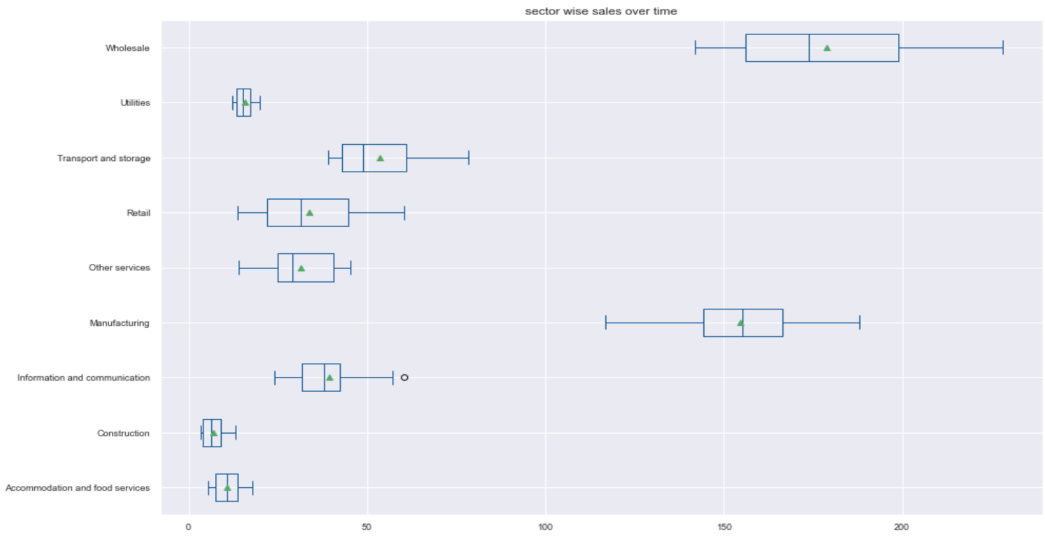


The e commerce sales in the united kingdom can be divided by industry wise sectors. Here the data of sales on internet are divided into nine wide category industries i.e., Accommodation and food services, Construction, Information and communication, Manufacturing, Other services, Retail, Transport and storage, Utilities, Wholesale. The below box plot illustrates the prominence of each industry on e-commerce front. The two major sector of industries making the highest average sales over the decade of 2009 to 2019 are wholesale industry and manufacturing industries, Whereas the Construction industry sectors made the least average sales on e- commerce platforms in that decade.

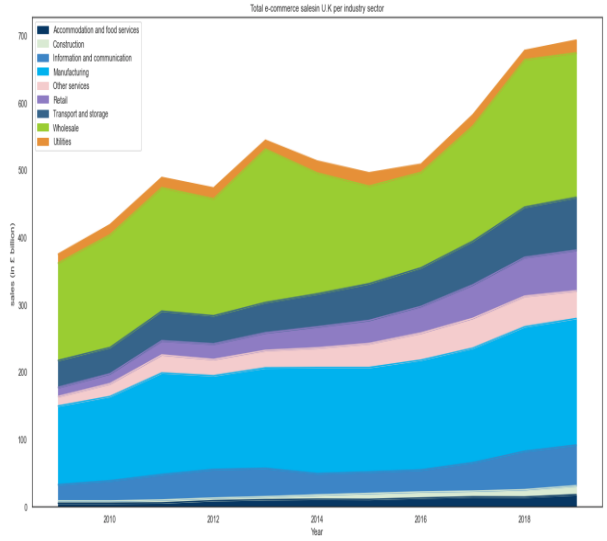
Hence in the decade of years 2009-2019, wholesales industry made an average e- commerce sales of 179 billion pounds in the United Kingdom, while Manufacturing industry made an average of 154.6 billion pounds in that particular decade. Transport and storage industry made an average of 53.4 billion pounds in the same period while Information and communication sector made almost an average of 40 billion pounds. In that decade, retail industry was a market of an average 33.8 billion pounds whereas other services sector accounted for 10.5 billion pounds on average. Utilities industry averaged at 15.6 billion pound in the decade of 2009 to 2019 while Accommodation and food services averaged at 10.7 billion pound during the same period. The least popular industry on e- commerce platforms was the Construction industry sector which made only an average of 6.8 billion pounds in the time period of 2009- 2019.

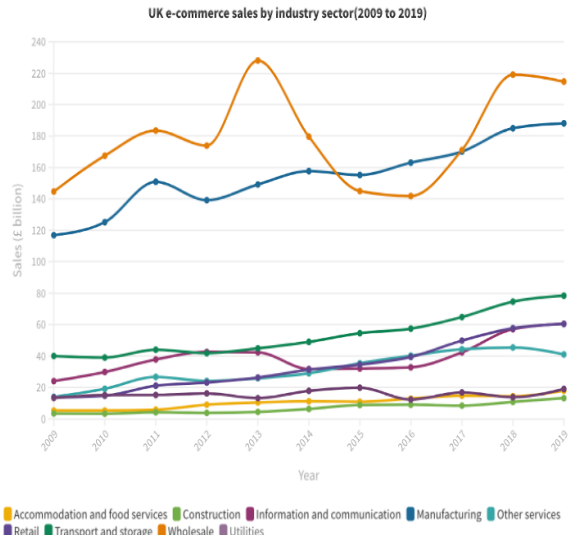
	count	**mean**	**std**	**min**	**25 %**	**50 %**	**75 %**	**max**
Year	11	2014	3.316625	2009	2011.5	2014	2016.5	2019
Accommodation and food services	11	10.70909	4.146916	5.3	7.4	10.9	13.65	17.9
Construction	11	6.872727	3.374638	3.3	4	6.3	8.9	13.2
Information and communication	11	39.35455	11.29171	24	31.85	37.8	42.45	60.5
Manufacturing	11	154.6	22.21382	116.9	144.2	155.2	166.6	188.1
Other services	11	31.36364	10.5572	14	24.95	29	40.6	45.4
Retail	11	33.80909	16.36383	13.6	22.1	31.3	44.6	60.4
Transport and storage	11	53.49091	13.86712	39	42.9	49	61.1	78.4
Wholesale	11	179	30.41533	141.8	156.25	173.9	199.1	228.1
Utilities	11	15.68182	2.477022	12.3	13.55	15.2	17.3	19.8

The below box plot illustrates the e commerce sales by sector wise industries from year 2009 to year 2019 in £ billion. The green triangle on the boxplots indicates the mean values of the different industry's sales in the years 2009 to 2019.

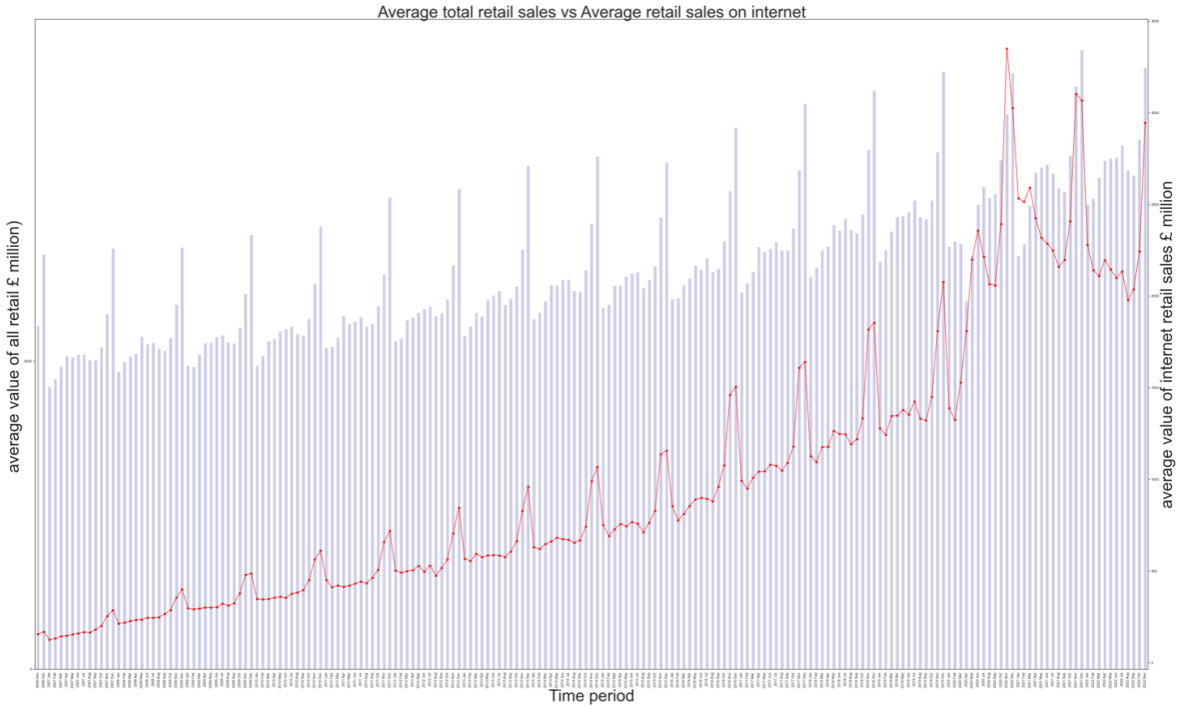


As depicted by the line plots maximum revenue is generated by wholesale industry in the entire decade. E - commerce of Wholesales industry was a 144.7-billion-pound market in year 2009 and these sales increased to 214.7 billion pounds in year 2019 making it the most successful industry on the e commerce platforms. Manufacturing industry followed in the second most successful industry sector with respect to making sales on e- commerce front, by making sales of 116.9 billion pound in 2009, which kept gradually increasing to a 188.1 billion pounds in the year 2019. However, for a brief period of two years i.e. 2015 and year 2016 the manufacturing industry over took the wholesales industry by making more e commerce sales in those years. Of the remaining industries Transport and Storage earned better revenues across the decade with making a 40 billion pounds sales in the year 2009 which gradually increased to 74.8 billion pounds in the year 2019. Information and communication sector made close to 24 billion pounds in the year 2009 which increased to 60.5 billion pounds in the year 2019. The other services category of industry was initial making a sales revenue of 14 billion pounds in the United Kingdom on e- commerce platforms in the year 2009 and these sales increased to 41 billion pounds by the year 2019. The Retail sector in the United Kingdom was a 13.6 billion pound industry in 2009. The e- commerce sales of retail industry eventually registered sales of 60.4 billion pound in the year 2019. The Accommodation and food services industry was generating sales of 5.3 billion pounds on online platform in the United Kingdom. This industry sales increased to an 17.9 billion pound in the year 2019. Utilities was also an smaller industry on the e commerce platforms, it made a total of 13.3 billion pounds worth of sales in year 2009 which only increased by only 5.7 billion pounds by the year 2019. Construction industry related sales on e- commerce platforms were least popular in the United Kingdom. The entire industry generated only an 3.4 billion pounds of revenue in the year 2009 on the internet platforms. However, the sales of the Construction industry on e- commerce site did increase to 13.2 billion pounds by the year 2019. Although the e- commerce sales figures varied drastically industries sector wise in the united kingdom, the area plot clearly illustrates that all industries had a noticeable growth in their respective e- commerce sales in the year 2019 compared to their respective sales figures in the year 2009.

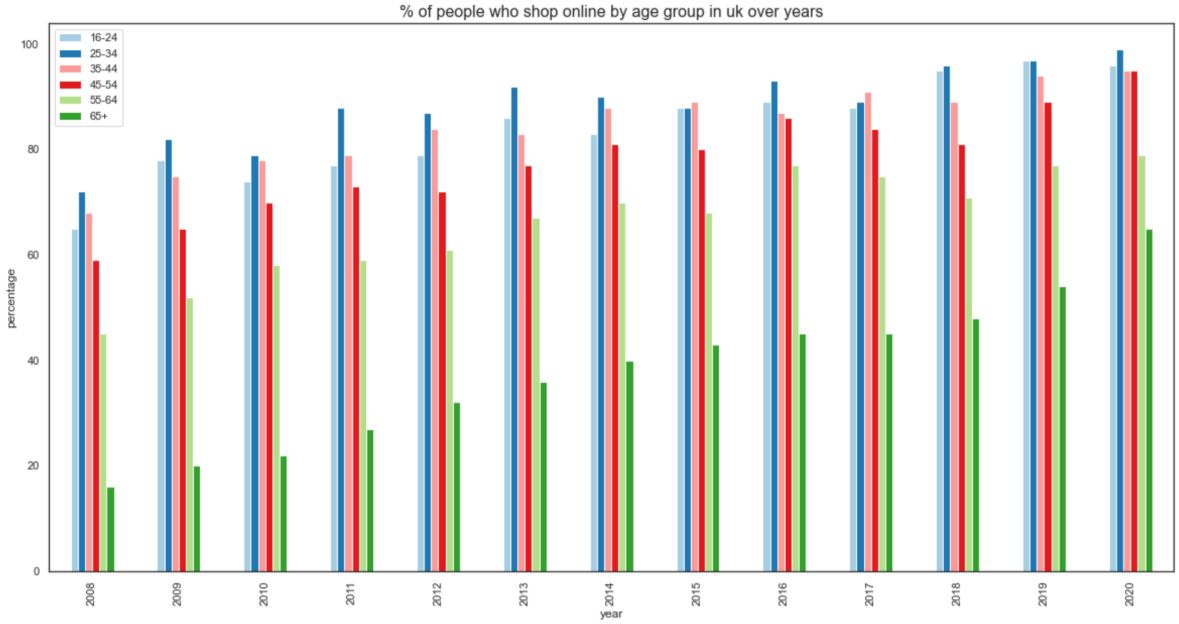




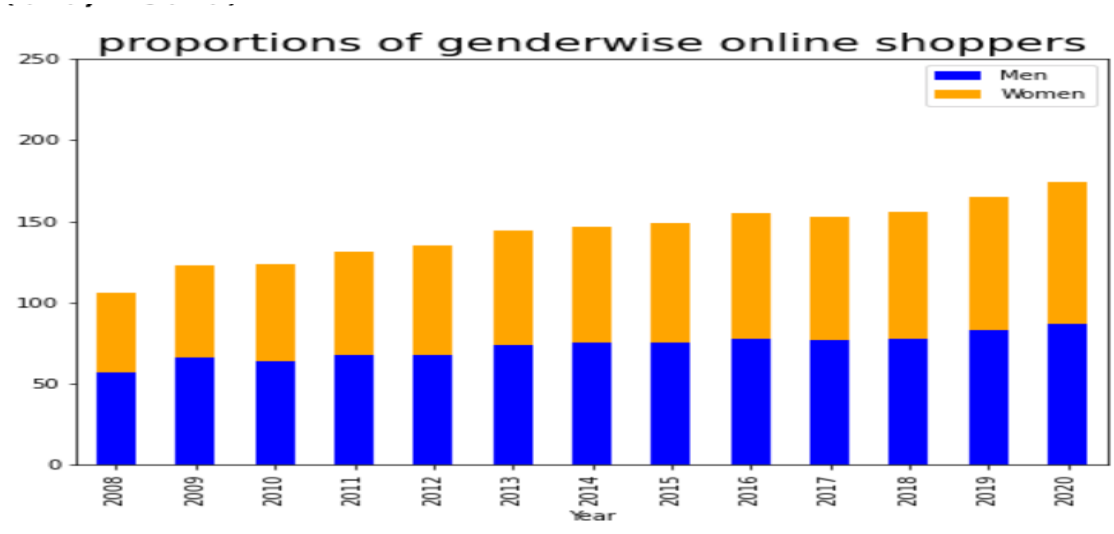
The combine graph below illustrates month wise United Kingdom's Average weekly value for all retailing sales (in £ million) and Average weekly value for internet retail sales (in £ million) for the time period starting from November 2006 to November 2022. For November 2006, the average retail sales value by weekly was 5573.3 million pounds in the United Kingdom and the average retail sales on internet (weekly average) were 155.7 million pounds., that is the sales on the internet were only 2.7 % of the retail sales in November 2006. This ratio was changing gradually with internet retail sales increasing slowly. By November 2020 retail sales on e- commerce platforms were an average value of 3349.2 million pounds when the average of all retail sales that month were 9004 million pounds. Hence by November of year 2020 e- commerce retail sales were nearly 37.2 % of total retail sales.



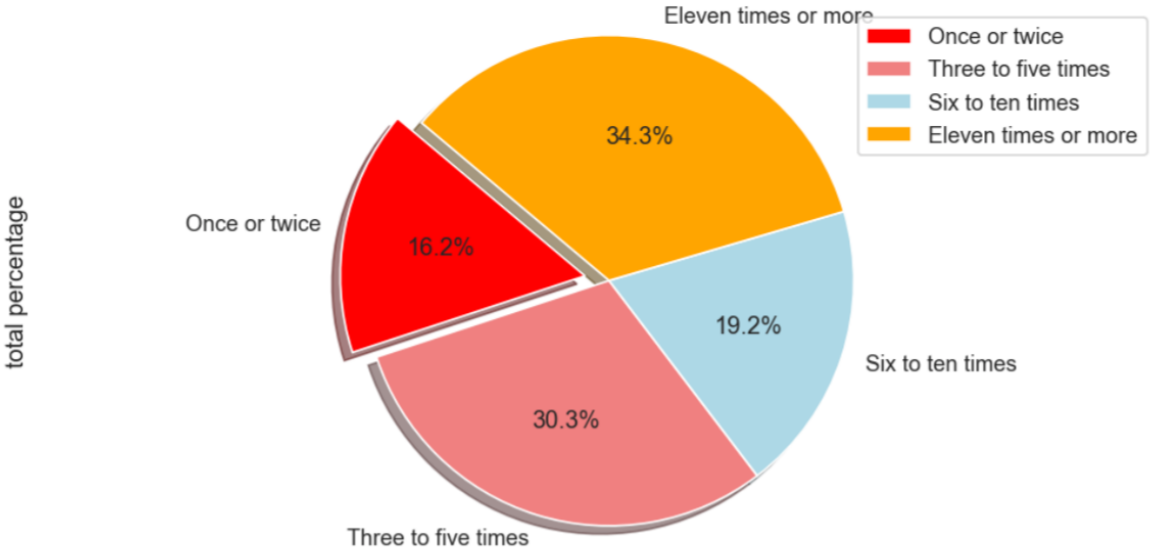
The grouped bar graph below illustrates the percentage of people shopping on the internet in the United Kingdom with further classification of their age group. It is observed that online shoppers belonging to the age of 25 year to 34 years more preferred online shopping compared to other age groups since the years 2008 to 2020. Though as people shopping in all age groups has increased as time progressed from years 2008 to 2020, percentage of people from age group 16 years to 24 years old to shop online has also increase and has become at par with people from age group 25 to 34 years. Percentage of people of age group 55 year to 64 years shopping on the inter has initial increased from year 2008 to 2016, there has be a noticeable dip in this percentage in the year 2015 and there after has followed irregular trend till year 2020.



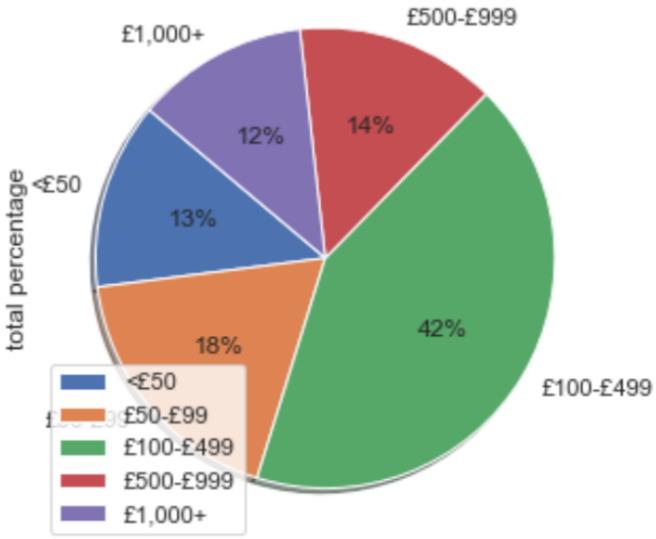
If we observe the proportion of women and men purchasing online in the United Kingdom from year 2008, we notice almost equal per cent of both genders have shopped online. In year 2008 57 % of the males were shopping online while 47 % of women were shopping online. Even if by a minor proportion, it is noticed the percentage of men shopping online has always been higher for a given year, compare to percentage of women shopping online, but by 2020 equal per cent of men and women were shopping online, that is, 87 per cent of both genders were shopping online.

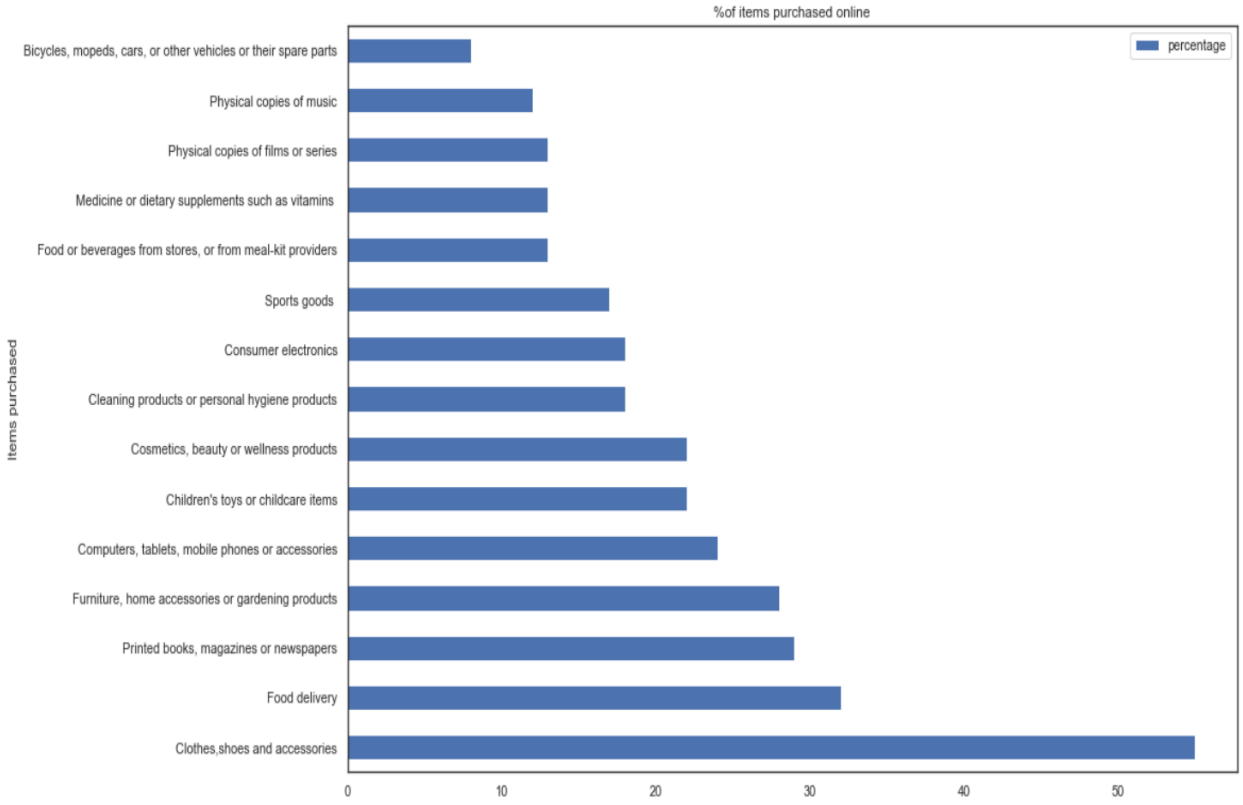


The below pie chart shows how frequent people of the United Kingdom shop online in the year 2020. 34 per cent of the great Britain's population shopped 11 or more times, while 30 % of them shopped 3 to 5 times in the year 2020. 19 % of the population shopped 6 to 10 times and 16 % shopped only once or twice.

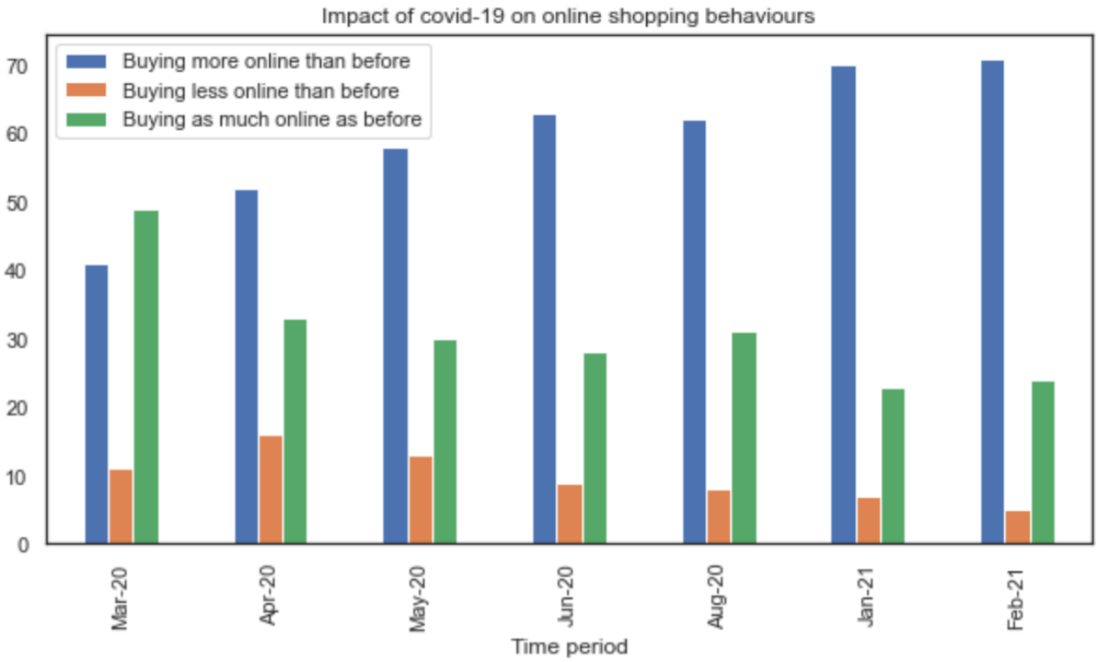


The pie chart shows that for the year 2020, 42 % of Britain's population shopped for goods/services of total value between 100 pounds to 499 pounds on the internet. While 13 s% of the population shopped for less than a total value of less than 50 pounds. 18 % of the population shopped online for goods costing between 50 pounds to 99 pounds. And 14 % of online shopper in United Kingdom bought goods of a value between 500 pounds to 999 pounds and finally only 12 % of the Britans population bought goods valued 1000 or more pounds on the internet in year 2020.

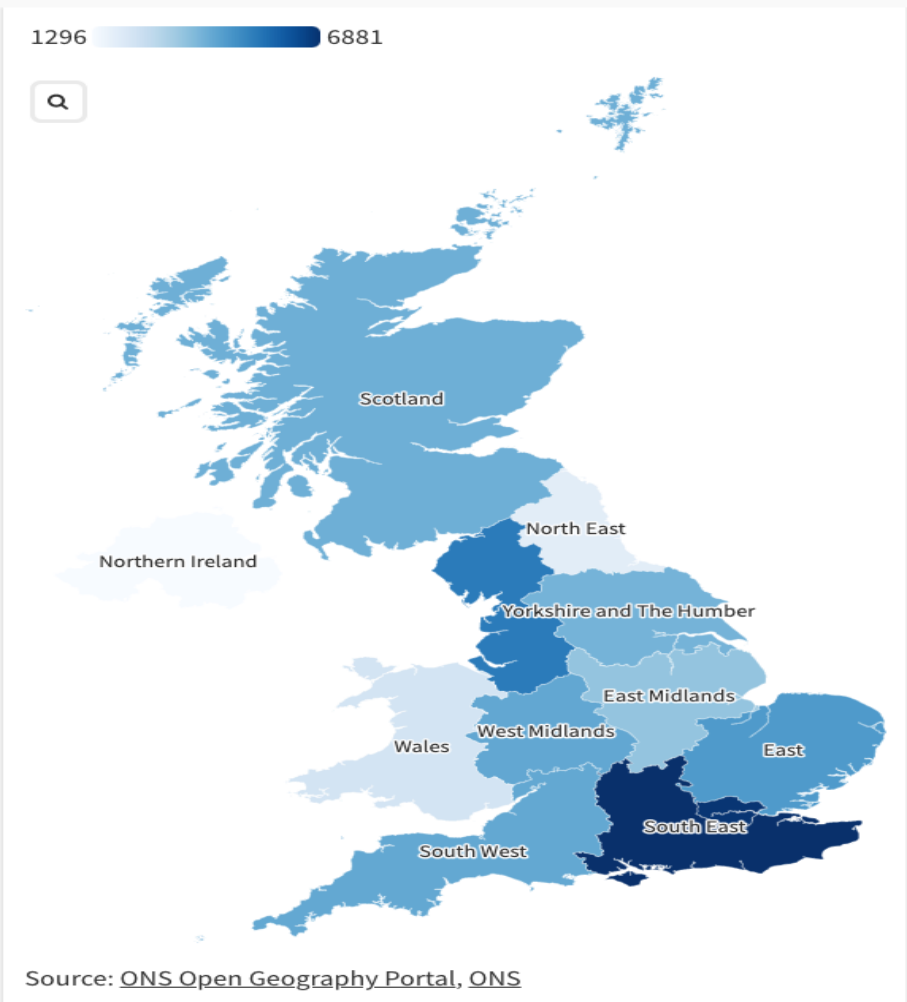




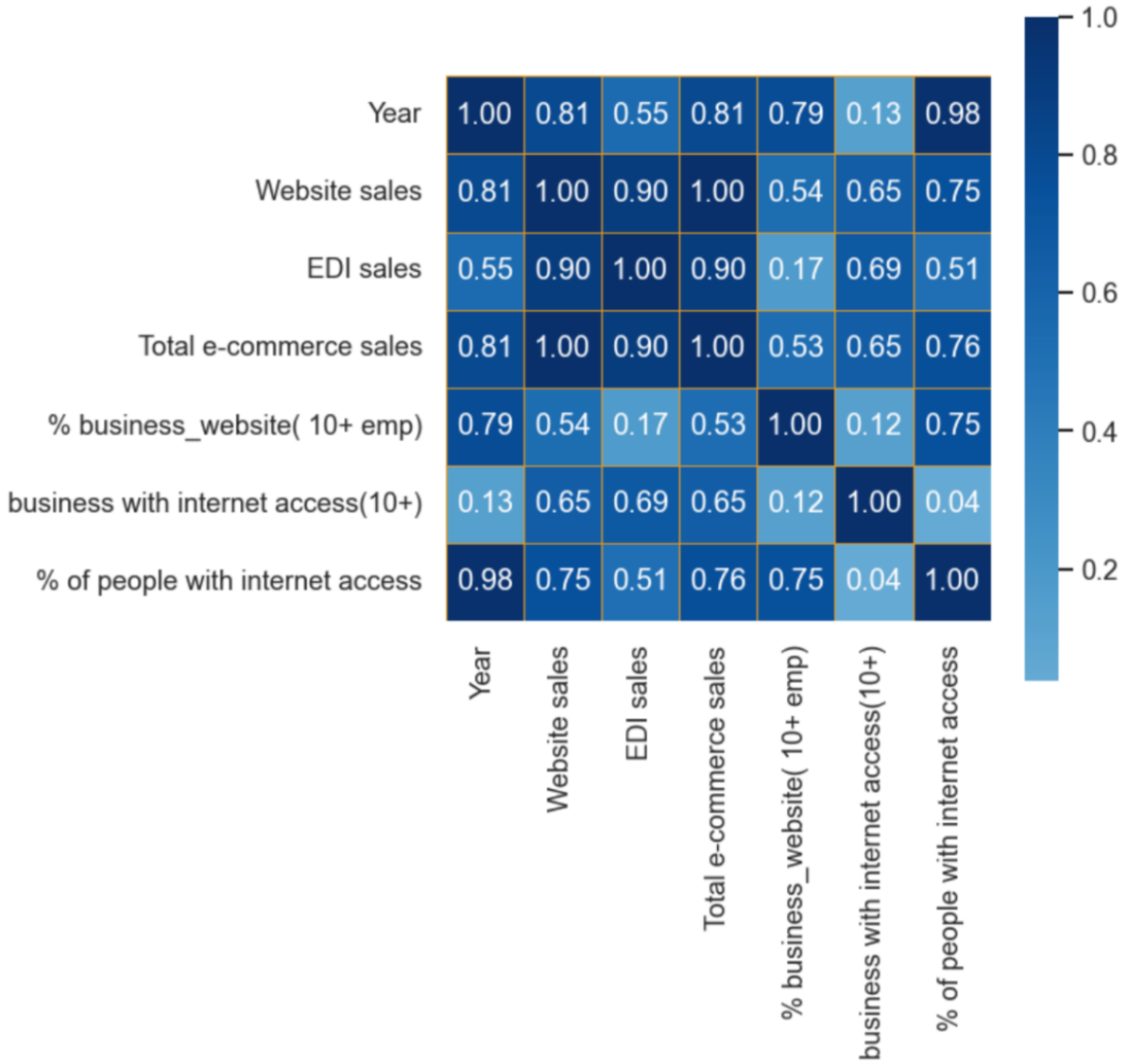
The bar graph show that of all the purchase make on the internet for the year 2022. Maximum type of goods bought in the United Kingdom are clothes, shoes, and accessories. 55 per cent of these goods were bought from the internet in the United Kingdom. And 32 per cent of all food deliveries were placed on the internet in United Kingdom this year. The least type of goods bought on the internet this year was Bicycles, mopeds, cars, or other vehicles or their spare part, with only 8 per cent of their sales taking place online. The same is also portrayed by the sunburst graph as shown above.



The bar graph above shows the results of a survey conducted in the United Kingdom. It clearly shows due to covid 19 pandemic the percentage of people who are buying / prefer buying online is more than before and increased drastically. The percentage of people willing to shop offline has noticeable decreased since April of the year 2020.



The map above shows the density of internet users by geographical location in the United Kingdom, with darker shade of blue indicating higher density of population in that area using internet. In the entire country of United Kingdom there are a little more than 49 million user. Hence the South East region has the highest internet users with almost 6.881 million users located in this area. Norther Ireland has the least internet users with 1.296 million users.



The heatmap above shows the correlation between percentage of people with internet access, percentage of businesses with 10 or more employees, which have internet access and also have their own website, percentage of e-commerce sales. There exist a positive and high correlation of 0.76 between total e commerce sale and percentage of people with internet access. As it can be observed from the graph above there is a positive correlation of 0.75 between percentage of website sales and percentage of people with internet access. However, there exists only a smaller but positive correlation between Electronic Data Interchange sales and percentage of people with internet access that is there is only 0.51 correlation between both the fields. There exists a positive correlation between percentage of companies with internet access and the total e-commerce sales, website sales and sales over electronic data interchange. Percentage of business with websites of size 10 or more employees have high correlation with percentage of people with internet access, that is 0.76 correlation, and they also have positive correlation with total e commerce sales. However, the percent of business with website show higher correlation with sales over website than with sales over Electronic Data Interchange. Total e- commerce

sales are extremely highly correlated (1.00 correlation) with website sales. It can also be concluded that website sale contributes more towards the total e- commerce sales that sales over Electronic Data Interchange. And the business with their own websites has higher e- commerce sales and even higher website sales. Finally, it can be concluded that that the higher the number of people who have access to internet the higher are the e- commerce sales. So, combining the results from the geographic map showing areas having higher density internet users and the the correlation results from heatmap, it can be concluded that areas with high number of internet users have higher e- commerce sales, especially higher sales over websites.

Recommendations and limitations

The retail sale across the world have also increased from $ 1336 billion in year 2014 to $ 5717 billion in year 2022. According to the findings of this study, it is expected that by 2040, almost 40% of retail sales in the UK will be made through e-commerce, and with the majority of e-commerce revenue generated from website sales. This information could be useful for novice entrepreneurs who are interested in starting a retail business in the UK and want to maximize their reach and revenue. It is suggested that targeting a younger audience, specifically those under the age of 54 and particularly those between the ages of 16 and 34, would be profitable for a retail e-commerce business. Additionally, locating the business in or making it available to the population in the South East region of the UK may increase the chances of generating revenue from online sales of goods. The study also notes that a large portion of the UK population is likely to spend between £100 and £499 on goods, particularly clothing, shoes, and accessories, so stocking inventory with these items may be a good idea. For existing retail businesses without an e-commerce platform, the study recommends setting one up, as the trend is towards a shift of retail sales to e-commerce. There is a potential limitation to the study in that it utilises only one data source, the Office for National Statistics, which may not be representative of the entire online retail market in the UK. This could adversely affect the generalizability of the study's findings to other online retail sales data sets or to the broader online retail industry as a whole.

Conclusion

The world- wide trends of e commerce sale show that the percent of offline sales, despite making up the majority of total sales, show a declining pattern, whilst the percentage of online sales have been gradually increasing. While in the United Kingdom the e- commerce sale have been on raise, in particular the sales over website saw a huge increase in the past decade. All sizes of businesses, including micro, small, medium, and large enterprises saw an increase in sales over e commerce platforms in the United Kingdom. However, the businesses with 250 to 999 employees saw noticeable dip in sales over e-commerce in the past decade, particularly in sales conducted over EDI's (electronic data interchange). sales over an ecommerce platform will generate high revenues for companies with 1000 or more employees and increase the revenue generated for any company with less than 1000 employees . The most popular sector making the greatest number of sales in the United Kingdom is the wholesale sector. the sector that generates the least revenue on e- commerce platforms in the United Kingdom, is the construction sector. The most active online shopper in the United Kingdom belong to the age group of 25 years old to 34 years old and South east region of the United Kingdom has the most density of population with internet access. Almost 35 % of people of UK shopped 11 time or more online in the year 2020 and 42 % of Britain's population spent 100 pounds to 499 pounds making purchases on e- commerce platforms. Clothes, shoes, and accessories makeup that 50 % of purchases made from online platforms in the United Kingdom. An auto regressive moving average model was used to forecast the percentage of retail sales to shift to an e-commerce platform. Due to the sudden peak in online retail sales in the mids of the year 2020, the forecasted values of internet retail sales as a percentage of the total retail sales form year 2019 are significantly lower than those values forecasted from the same data including data till November 2022. The results in this study can help novice retailers and policymakers to gain insights into trends and patterns in online retail sales in the United Kingdom and how different time series models can be used observe the positive impact covid had on e- commerce retail sales in the United Kingdom and also to forecast future sales in the United Kingdom.

References

AbuRaya, R. (2020, November 1). *Business Analytics of E-Commerce Policy and Practice: An Ethical Perspective*. IEEE Xplore. https://doi.org/10.1109/DASA51403.2020.9317117

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis : forecasting and control*. John Wiley & Sons, Inc.

Chang, Y.-W., & Liao, M.-Y. (2010). A Seasonal ARIMA Model of Tourism Forecasting: The Case of Taiwan. *Asia Pacific Journal of Tourism Research*, *15*(2), 215–221. https://doi.org/10.1080/10941661003630001

Chatfield, C. (2004). *The analysis of time series : an introduction*. Chapman & Hall/Crc.

Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018). Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. *IOP Conference Series: Materials Science and Engineering*, *394*, 052024. https://doi.org/10.1088/1757-899x/394/5/052024

Fedushko, S., Shevchuk, L., Poritska, A., Kravets, R., & Tymovchak-Maksymets, O. (2021). E-Commerce and E-Health Strategies and Implementation Activities in the United Kingdom: Review Study. *ArXiv:2108.08271 [Cs]*. https://arxiv.org/abs/2108.08271

Graves, A. (2020, July 21). *Time Series Forecasting with a SARIMA Model*. Medium. https://towardsdatascience.com/time-series-forecasting-with-a-sarima-model-db051b7ae459

Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, *35*(1-2), 213–216. https://doi.org/10.1016/s0360-8352(98)00066-7

International Trade Administration. (2022, September 12). *United Kingdom - eCommerce*. Www.trade.gov. https://www.trade.gov/country-commercial-guides/united-kingdom-ecommerce

Kirby-Hawkins, E., Birkin, M., & Clarke, G. (2018). An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Environment and Planning B: Urban Analytics and City Science*, *46*(6), 1148–1164. https://doi.org/10.1177/2399808318755147

Lewinson, E. (2022, March 31). *Time Series DIY: Seasonal Decomposition*. Medium. https://towardsdatascience.com/time-series-diy-seasonal-decomposition-f0b469afed44

Liu, T. (2021, March 1). *Research on the Design of E-commerce Data Analysis Platform*. IEEE Xplore. https://doi.org/10.1109/ECIT52743.2021.00034

Nau, R. (2019). *Introduction to ARIMA models*. Duke.edu. https://people.duke.edu/~rnau/411arim.htm

OECD. (2017). *Entrepreneurship - Enterprises by business size - OECD Data*. TheOECD. https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm

Office for National Statistics. (2008). *From Holt-Winters to ARIMA Modelling: Measuring the Impact on Forecasting Errors for Components of Quarterly Estimates of Public Service Output*. https://www.ons.gov.uk/ons/guide-method/ukcemga/publications-home/publications/archive/from-holt-winters-to-arima-modelling--measuring-the-impact-on-forecasting-errors-for-components-of-quarterly-estimates-of-public-service-output.pdf

Office for National Statistics. (2020, August 7). *Internet access - households and individuals - Office for National Statistics*. Ons.gov.uk. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/ho

meinternetandsocialmediausage/datasets/internetaccesshouseholdsandindividualsreferencetables

Office for National Statistics. (2021, February 5). *E-commerce and ICT activity - Office for National Statistics*. Www.ons.gov.uk. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/datasets/ictactivityofukbusinessesecommerceandictactivity

Office for National Statistics. (2022, December 16). *Retail Sales Index internet sales - Office for National Statistics*. Www.ons.gov.uk. https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/retailsalesindexinternetsales

Office for National statistics. (2019). *Internet users - Office for National Statistics*. Ons.gov.uk. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/datasets/internetusers

Pyne, K. (2014, January 7). *Literature review: The rise of the internet and consumer high street shopping habits – Seven Bridges*. Seven Bridges. http://sevenbridges.ncl-coll.ac.uk/seven-bridges-vol-2-2014/literature-review-the-rise-of-the-internet-and-consumer-high-street-shopping-habits/

Tableau. (n.d.). *Time Series Analysis: Definition, Types, Techniques, and When It's Used*. Tableau. Retrieved January 6, 2023, from https://www.tableau.com/learn/articles/time-series-analysis#:~:text=Time%20series%20analysis%20is%20a

Teh, Z.-H., Lim, C.-Y., & Chen, S.-Y. (2021, September 1). *An Exploratory Review of Malaysian E-Commerce Merchants and Their Readiness in Adopting Business*

Analytics Models for Assessment of Business KPIs. IEEE Xplore. https://doi.org/10.1109/RTSI50628.2021.9597241

Tighe, D. (2021, May 3). *UK: change in online shopping since COVID-19 2020-2021*. Statista; www.statista.com. https://www.statista.com/statistics/1230225/changes-in-online-buying-among-uk-consumers-since-covid-19/

Wikipedia Contributors. (2019, March 7). *E-commerce*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/E-commerce

Wikipedia Contributors. (2020, December 23). *Box–Jenkins method*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Box%E2%80%93Jenkins_method

Appendix:

Python code: ARIMA forecasting and EDA of e-commerce data united kingdom.ipynb

Research Proposal

Introduction

In recent years, retail sales have become increasingly popular for conducting e-commerce transactions. The main aim of my project will be to examine how the trends of online shopping have evolved over time and what lead to this trend to develop. I will also analyse the overall impact such e commerce evolution has on specific individual business entities and on a country's economy. I wish to determine if a certain type of businesses would benefit from multichannel retailing.

Background & Significance to the problem

The global e-Commerce market is experiencing rapid growth and expected to generate $4.891 trillion of which UK contributes 14.5% of the total e-Commerce global retail sales with annual sales of $99 billion. And is one of the top countries to start an e commerce business(Chow, 2022). However, a lot of factors contribute to starting a successful online business in the UK like geography(Kirby-Hawkins et al., 2018), sector of industry , size of business, internet access ,demographics, etc. My analysis is aimed at assessing all these factors to help decide if having a multichannel retailing is profitable to a business. It will establish trends in the online market sales and this analysis can be used by any new business trying to expand into online markets.

Objectives

This project will analyse all relevant data available to determine how e commerce sales are affecting a UK business. The major questions I seek to answer are :
1.What are the trends and purchase behaviour in online retail markets.
2. How has these trends affected businesses and the economy in general
3.Has covid -19 have effected these trends and in what way.
4. Develop a model to prediction model to ensure it generates profits to a business.
5. What kind of business should invest in online retailing to make profitable outcomes.

Methods

To analyse the data a time series analysis will be conducted since I wish to establish a pattern with time series prediction and then correlate and compare the results to reach a conclusion. Python will be the primary language to conduct this analysis and apart from seaborn packages within python, external visualisation tools like Flourish will be used.

An Autoregressive Integrated model will be used to forecast the shift of retail sale to e commerce platform in the united kingdom.

An autoregressive integrated moving average (ARIMA) model is a statistical technique used to predict future values in a time series data set. It combines both autoregressive (AR) and moving average (MA) models, and includes a term to account for the integration of past values. The model is represented by the equation ARIMA(p,d,q), where p is the number of autoregressive terms, d is the number of times the data has been differenced, and q is the number of moving average terms. The AR and MA components of the model identify patterns in the data that are influenced by past observations, while the integration term handles nonstationary in the data, such as trends or seasonal variations. ARIMA models can be utilized to make forecasts by fitting the model to historical data and using it to predict future values. They are commonly used in various fields, such as economics and finance, for forecasting and analysing time series data. (Musarat et al., 2021)

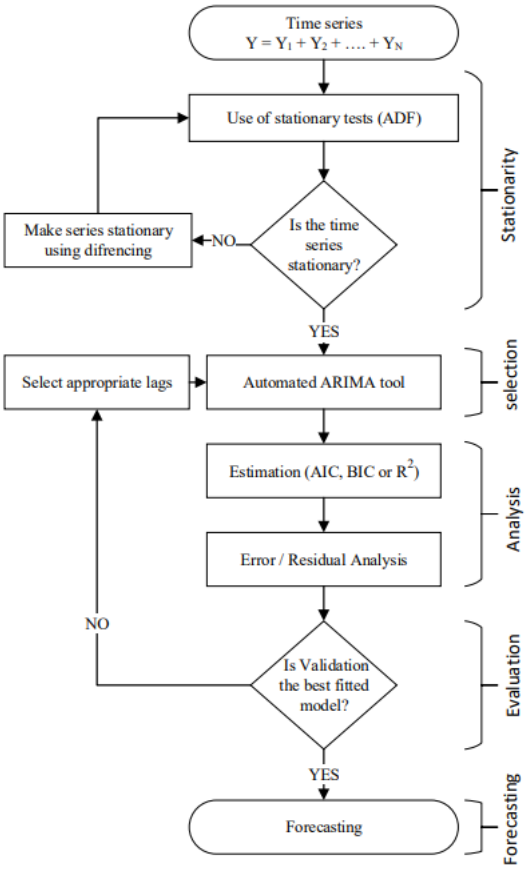


Figure 3. Methodology flowchart.

Literature and Data

There is some literature on how e commerce has changed overall worldwide and impacts covid had on these trends, but no specific literature could be found on how ecommerce progressed in the United Kingdom and if a company based in the United Kingdom will gain profits if it has an online retail avenue.

A paper published in 2011 (White & Ariguzo, 2011) has a similar but more statistical approach to the then e commerce sales in the U.S but the paper was designed to evaluate the overall growth U.S. e-commerce and the impact it had on their economy. I will build on this idea to get a more practical use of similar and wider data to benefit businesses.

ONS and also published an article which solely focuses on of e commerce sales in UK economy (*The Impact of E-Commerce on the UK Economy*, 2015), and a similar article focuses on global economy (Lee et al., 2005) though the data and measures are important for my project, I do not solely wish to focus on how its impacting the economy of the country. Covid -19 also had a huge impact on how tendencies of e commerce changed (Popa et al., 2022) and EY has also investigated the impact of post covid -19 on this market segment sales (Rogers & Cosgrove, 2022).

The data that will be used toward my project will mainly be about detailed ecommerce sales and revenue generated through these online sources over the last decade in UK obtained from Office for National Statistics (ONS) website and detail specific data like the sectors of retail performing well or the size of companies and the revenue they generate through online retail or the geographic location and age of internet users in that location will be used. I will be using ecommerce dB to get United kingdom specific company performance details and Statista to get the consumer survived details. Performing a time series analysis and correlating finding will allow for a relationship to be established from which I can derive the impacts of online retail on the overall profitability of a company and in turn the country's economy.

Ethical awareness

For my project, analysing is being performed on the already existing data available online like data published by Office of national statistics and data collected from published articles and surveys. No first hand surveys or data collection will be conducted.

Hence, I will ensure all terms and conditions are met of using existing data and the data is publicly available and covered by free and fair use agreements. I will ensure that the data is handled in compliance with the UK Government Data Ethics Framework.

Limitations

The major limitation of the study is the period for which the data is available. (White & Ariguzo, 2011). Since a larger number of observations makes time series analysis substantially more accurate, data is expected to be found for only the past decade or 15 years since e-commerce sales and the internet were only popularised fairly recently.

Conclusion

The project will take into consideration multiple aspects and study related data of online sales parameters to perform a time series analysis and correlate results to assesses the trends in online retailing and if it is profitable for businesses to expand sales to online platforms in the United kingdom.

References

Centre for Retail Research. (2018). *Online Retail : UK, Europe & N. America : The Centre For Retail ResearchOnline Retail : UK, Europe & N. America : The Centre For Retail Research*. Retailresearch.org. https://www.retailresearch.org/online-retail.html

Chow, S. (2022, January 6). *10 Best Countries for Opening an E-Commerce Business*. AirCorporate. https://air-corporate.com/blog/ecommerce-country-registration/

Dhali, S., Pati, M., Ghosh, S., & Banerjee, C. (2020, September 1). *An Efficient Predictive Analysis Model of Customer Purchase Behavior using Random Forest and XGBoost Algorithm*. IEEE Xplore. https://doi.org/10.1109/ICCE50343.2020.9290576

E-commerce and ICT activity, UK - Office for National Statistics. (n.d.). Www.ons.gov.uk. https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/ecommerceandictactivity/2019

ecommerceDB.com. (n.d.). Ecommercedb.com. Retrieved December 7, 2022, from https://ecommercedb.com/ranking/stores/gb/fashion?page=1&pagesize=50&specialist=all¤cy=USD

Kirby-Hawkins, E., Birkin, M., & Clarke, G. (2018). An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Environment and Planning B: Urban Analytics and City Science*, *46*(6), 1148–1164. https://doi.org/10.1177/2399808318755147

Lee, S.-Y. T., Gholami, R., & Tong, T. Y. (2005). Time series analysis in the assessment of ICT impact at the aggregate level – lessons and implications for the new economy. *Information & Management*, *42*(7), 1009–1022. https://doi.org/10.1016/j.im.2004.11.005

Musarat, M. A., Alaloul, W. S., Rabbani, M. B. A., Ali, M., Altaf, M., Fediuk, R., Vatin, N., Klyuev, S., Bukhari, H., Sadiq, A., Rafiq, W., & Farooq, W. (2021). Kabul River Flow Prediction Using Automated ARIMA Forecasting: A Machine Learning Approach. *Sustainability*, *13*(19), 10720. https://doi.org/10.3390/su131910720

Popa, A., Spatariu, E., & Gheorghiu, G. (2022). *Changes and Trends in Consumers` Behaviour and Online Purchasing: a Post-COVID-19 Analysis*. https://stec.univ-ovidius.ro/html/anale/RO/2022-2/Section%204/28.pdf

Rogers, K., & Cosgrove, A. (2022). *Future Consumer Index: How to capture the real e-commerce opportunity*. Www.ey.com. https://www.ey.com/en_gl/consumer-products-retail/future-consumer-index-how-to-capture-the-real-e-commerce-opportunity

The impact of e-commerce on the UK economy. (2015). Nationalarchives.gov.uk. https://webarchive.nationalarchives.gov.uk/ukgwa/20160105184546/http://www.ons.gov.uk/ons/rel/rdit2/e-commerce-and-internet-use/analysis-at-uk-level/sty-the-impact-of-e-commerce-on-the-uk-economy-.html

White, D. S., & Ariguzo, G. (2011, October 17). *A Time-Series Analysis of U.S. E-Commerce Sales*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1940960