FORECASTING PERSONAL CONSUMPTION EXPENDITURE FOR OCTOBER 2022

1. INTRODUCTION:

Personal consumption expenditure is a term used to measure the consumer spending habits on goods and services in the U.S economy. The consumer category includes households and non-profit institutions. (Pettinger, n.d.). On the other hand, overspending might result in inflation. So, understanding the customer spending trend might serve as a valuable tool for managing one's investments. It also acts as an indicator of long-term trends of inflation. This report aims to forecast the personal consumption expenditures for October of 2022.

2. DATASET:

The dataset provided is a JSON file. It consists of consumption data from the year 1959 to 2021. There are a total of 756 observations in this file. The file structure is as follows:

- Date
- Pce

```
'data.frame':   756 obs. of  2 variables:
 $ DATE: chr  "01/01/1959" "01/02/1959" "01/03/1959" "01/04/1959" ...
 $ PCE : num  306 310 313 312 316 ...
```

Fig 1. Structure of the file

2.1 TIME SERIES OBJECT CONVERSION

Firstly, the data is converted into time series for further analysis. Time series refers to a series of observations recorded over time. Past behaviour is used to predict future variables. Time series are used for forecasting when we have historical and quantifiable data. There are three types of time series models.

- Explanatory
- Pure time series
- Mixed model

Pure time series is chosen over the others since it is proven to give more accurate results.

2.2 HANDLING MISSING VALUES

Upon careful analysis, 37 NA data were found in the dataset. Looking at the data, it is evident that the data is Missing Completely at Random (MCAR). The two methods available for handling missing data are removing and imputing data.

```
> sum(is.na(pcedata$PCE))
[1] 37
```

Fig 2. Missing data

Since the missing data is significantly less, using imputation methods over deletion is advisable. This method helps in developing reasonable guesses for the missing data. It helps us forecast accurate results without any loss of data. The various imputation methods available in *the imputeTS* package are:

- na_interpolation

- na_ma
- na_kalman
- na_kalman – Arima model

A dummy time series with randomly introduced NA was created to find the best imputation method out of the above ones. All four methods were performed on dummy time series. The method with the most accurate results was selected with the help of the table below.

	head(pcemissingTS, 10)	pcefew	na_interpolation(pcefew)	na_ma(pcefew)	na_kalman(pcefew)	na_kalman(pcefew, model = "auto.arima")
1	306.1	306.1	306.100	306.1000	306.1000	306.100
2	309.6	309.6	309.600	309.6000	309.6000	309.600
3	312.7	312.7	312.700	312.7000	312.7000	312.700
4	312.2	NA	313.975	312.2353	313.7683	313.975
5	316.1	NA	315.250	315.1357	314.8884	315.250
6	318.2	NA	316.525	318.0353	316.1893	316.525
7	317.8	317.8	317.800	317.8000	317.8000	317.800
8	320.2	320.2	320.200	320.2000	320.2000	320.200
9	324.2	324.2	324.200	324.2000	324.2000	324.200
10	NA	NA	324.200	322.1429	327.0313	324.200

Fig 3. Analyzing imputation methods

The moving average method *na_ma* has the values with the highest accuracy. It replaces the missing values with moving mean values. Hence, this method was selected amongst all.

2.3 DECOMPOSING THE TIME SERIES

Decomposing the time series is a critical factor in helping understand the time series better. The time series has four components.

- Trend
- Seasonal
- Cyclical
- Randomness

Understanding these four components helps in selecting the best forecasting method. Trends can be understood by smoothing techniques. PCE is decomposed as below:

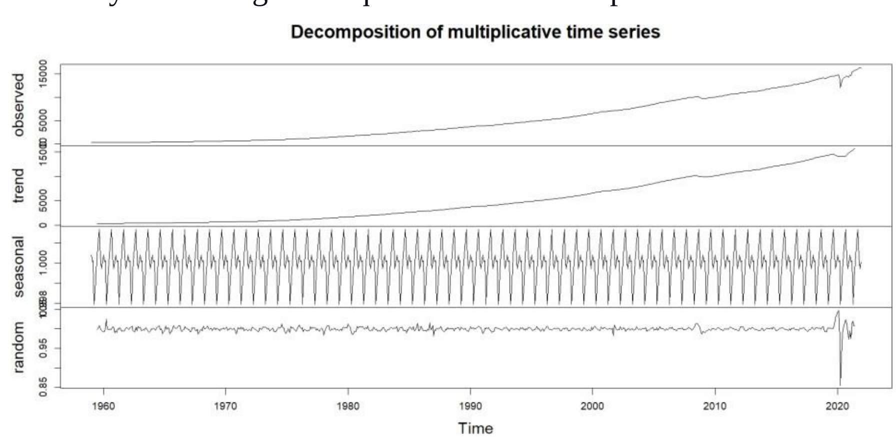


Fig 4. Time series decomposition

The plot clearly explains all four components. The trend component has an upward trend which increases over time. Seasonality is also observed in the above plot. This plot confirms that the time series has both trend and seasonality.

2.4 SELECTING FORECASTING MODEL

Before proceeding to forecast, the data is divided into training and testing data based on an 80/20 ratio.

- Training data – 80 percent
- Testing data – 20 percent

2.4.1 SIMPLE FORECASTING METHOD

There are four types of simple forecasting methods -

- Average – Forecasts values with the help of an average of historical data
- Naïve - Forecasts based on previous observation value
- Seasonal – Forecasts based on last observation's value of same year or the same month
- Drift – Forecasts based on the average change in historical data.

A forecast has been made for all the methods. The best method is selected with the help of accuracy measures and plots.

```
> accuracy(pceAvg, testData)
                       ME     RMSE      MAE        MPE      MAPE     MASE      ACF1 Theil's U
Training set 1.361438e-13 2954.368 2504.979 -196.14897 230.32633 12.53604 0.9951516        NA
Test set     6.804076e+03 6806.977 6804.076   66.94053  66.94053 34.05066 0.8437127   181.733
> accuracy(pceNaive, testData )
                    ME      RMSE       MAE       MPE      MAPE      MASE      ACF1 Theil's U
Training set  15.70447  28.82402  19.20968 0.5704799 0.6659049 0.0961339 0.1469924        NA
Test set     370.84464 420.72206 370.84464 3.6123951 3.6123951 1.8558735 0.8437127  11.33158
> accuracy(pceDrift, testData) #finalizing this !!
                        ME      RMSE       MAE        MPE     MAPE       MASE      ACF1
Training set -6.370651e-15  24.17011  16.32616 -0.8081765 1.123155 0.08170346 0.1469924
Test set      2.059477e+02 232.80881 205.94771  2.0065288 2.006529 1.03065499 0.8351165
             Theil's U
Training set        NA
Test set      6.267148
> accuracy(pceSNaive, testData)
                   ME     RMSE      MAE      MPE     MAPE     MASE      ACF1 Theil's U
Training set 194.9086 249.0486 199.8222 6.707418 6.757564 1.000000 0.9704526        NA
Test set     181.5175 345.9912 296.9035 1.739083 2.905634 1.485839 0.7634971  9.120272
```

Fig 5. Accuracy measures

The model with less RMSE, less MASE, and a MAPE value lesser than one is considered the best model. This resulted in selecting *the pceDrift* method as the best simple forecasting model.

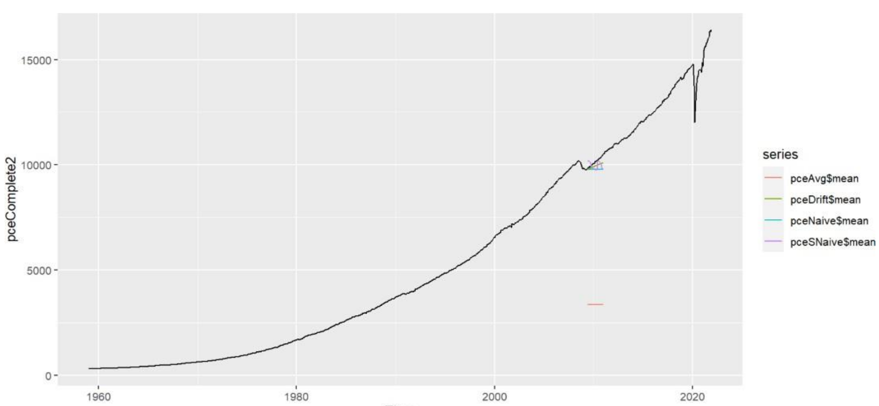


Fig 6. Simple forecast plot

2.4.2 EXPONENTIAL SMOOTHING MODEL

As per *Fig.4*, the PCE time series has both trend and seasonality. Holt-Winters and ets methods were used, and the best one is selected based on accuracy measures as explained above.

```
> accuracy(forecast(fit),testData)
                      ME      RMSE       MAE        MPE      MAPE       MASE      ACF1
Training set  -0.2073152  21.73532  11.64092 0.03548565 0.3959803 0.05825641 0.1393695
Test set     537.6525899 613.90669 537.65259 5.18371774 5.1837177 2.69065550 0.8760151
             Theil's U
Training set        NA
Test set      15.76082
> accuracy(fchw)
                     ME     RMSE     MAE        MPE      MAPE       MASE      ACF1
Training set -0.5237421 25.30779 14.8901 0.02033865 0.7253601 0.07451677 0.2190521
```

Fig 7. Exponential smoothing model

The method with the highest accuracy occurred to be *the ets model.*

2.4.3 ARIMA MODEL

Auto-Regressive Integrated moving average is a combination of Autoregression where values are predicted based on past values and the MA model, which used past prediction errors to model the relationship

```
> accuracy(forecast(pceArima),testData)
                     ME      RMSE       MAE        MPE      MAPE       MASE         ACF1
Training set  -0.218879  21.28734  11.87414 0.01557004 0.4019744 0.05942353 -0.009503355
Test set     611.637756 697.74373 611.63776 5.89737896 5.8973790 3.06091057  0.876699958
             Theil's U
```

Fig 8. ARIMA model

Out of the above three methods, the accurate method is decided by plotting them together in one graph. It helps to understand which model predicted better results. The plot below explains all three models.

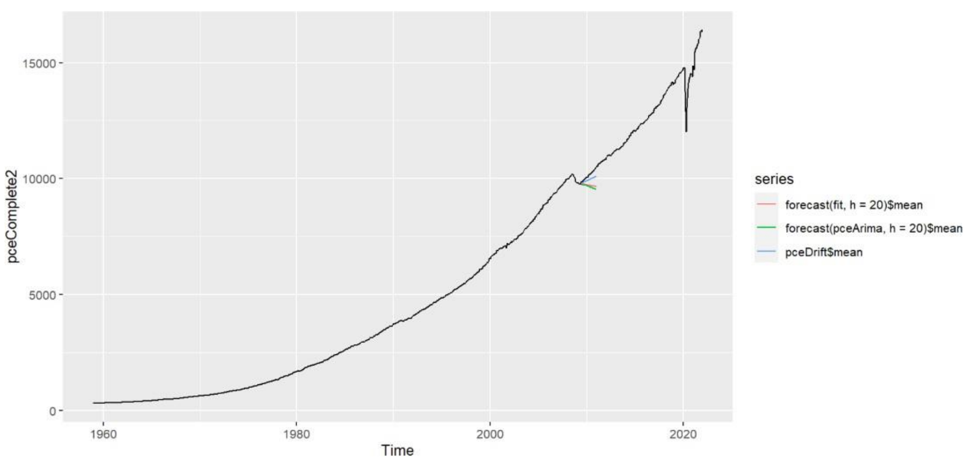


Fig 9. Selecting forecasting model

By comparing the accuracy of models, the Drift method looks only a few points away from being perfect. Therefore, *the Drift method is* the best method to forecast the given time series.

2.5 RESULT

FORECASTING PCE FOR OCTOBER 2022:

The drift method is used to forecast the next ten months of PCE. The forecasted value is plotted alongside the original time series to understand the trend better.

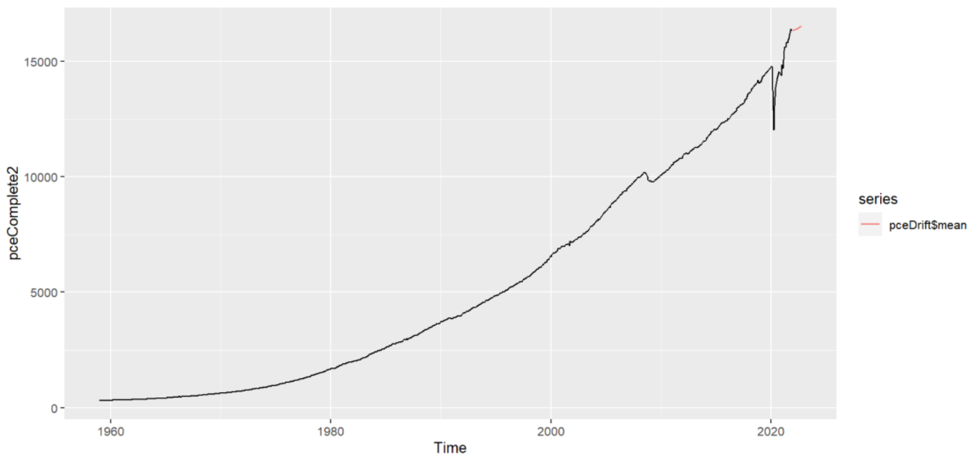


Fig 10. October 2022 forecasting

The above plot clearly states that the spending for the year 2022 is increasing steadily after a dip in 2020. The black line refers to the original data, whereas the orange line refers to the forecasted value. Below are the predicted values for the year 2022 until October.

Forecasts:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2022	16327.49	16199.98	16455.00	16132.48	16522.50
Feb 2022	16348.68	16168.24	16529.13	16072.71	16624.66
Mar 2022	16369.88	16148.73	16591.03	16031.66	16708.09
Apr 2022	16391.07	16135.54	16646.60	16000.27	16781.87
May 2022	16412.26	16126.38	16698.14	15975.05	16849.47
Jun 2022	16433.45	16120.08	16746.82	15954.20	16912.71
Jul 2022	16454.65	16115.95	16793.35	15936.65	16972.64
Aug 2022	16475.84	16113.51	16838.16	15921.71	17029.97
Sep 2022	16497.03	16112.48	16881.59	15908.90	17085.16
Oct 2022	16518.22	16112.60	16923.85	15897.88	17138.57

Fig 11. Forecasted values

From the above *Fig11* we can conclude that by October 2022 the PCE value is expected to be around 16518.22.

2.6. ONE STEP AHEAD ROLLING FORECASTING

This method estimates the model on a single set training dataset. IT computes one-step forecasts on the test data. It is achieved by applying the fitted model to the entire dataset and extracting the one-step forecast, called fitted values. It is used to compare time series models.

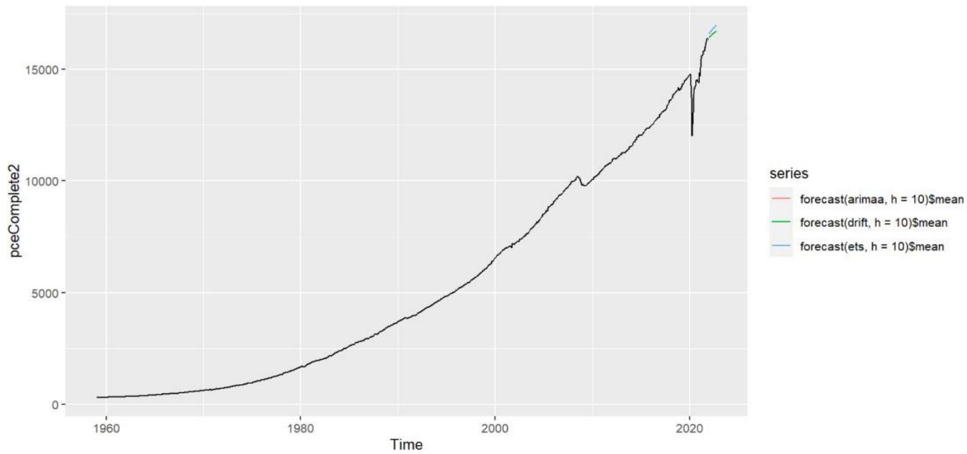


Fig 12. Result

By comparing the models, drift model is the most accurate for this time series.

CONCLUSION:

PCE data from 1959 to 2021 was considered to predict the value for October 2022. The data was cleaned and imputed to proceed with forecasting. Simple forecasting, exponential, and Arima model was performed on the data. The best fit was found with the help of *the accuracy* method. The Drift model proved to be the best fit for the PCE data by forecasting an upward trend for the following year. It is easy to plan the investments and understand the inflation trends with the estimated value.

PERFORMING TOPIC MODELLING ON AMAZON REVIEWS

1.INTRODUCTION:

The world faced the biggest pandemic in the year 2020. The Covid-19 crisis has impacted various industries across the globe. The one industry which boomed amidst all this chaos was the e-commerce industry. Uni-commerce and Kearney studied the e-commerce trends during the pandemic and found a massive rise in customer shopping once the lockdown was announced (Anon., n.d.). Online reviews are a crucial factor for both customers and sellers. It helps customers know the product's authenticity and sellers understand customer satisfaction. We will perform topic modeling on amazon reviews and find the most and most minor factors affecting customer satisfaction.

2. METHODOLOGY:

2.1. DATASET

Amazon reviews of the mobile one Plus have been provided with the following parameters:

- Product
- Product company
- Profile name
- Review title
- Review rating
- Review text
- Helpful count
- Total comments
- Review country
- Review time
- URL
- ID
- Color
- Category
- Sub-category

The above file format is JSON; hence, we use *fromJSON* to read the file. Although we have various parameters listed above, only a few of them will help us achieve our final goal. Variable selection and cleaning are carried out in the data preparation stage. The data preparation stage is the first step before starting any process. It is explained in detail in the below section.

2.2. METHODOLOGY

2.2.1 DATA PREPARATION

For the scope of our analysis, we have selected two parameters as our variables of interest, namely,

- Review text
- Review ratings

These two variables of interest will serve our primary purpose in helping us find out customer satisfaction levels. The next step is to create a sample of this data since our original dataset has more than thirty thousand reviews. The data is sampled into five thousand observations. This sampled data is used throughout the process.

2.2.2 CLASSIFYING REVIEWS

To study customer satisfaction, we need to be able to classify the reviews as positive and negative. So, it is essential to classify the reviews first before proceeding to the next steps. There are two methods to classify the reviews as follows:

- Sentiment analysis
- Review ratings

In a recent study conducted, it is proved that after a positive experience, 85 percent of the customers tend to buy more. The rest, 20 percent, drops the product after just one unpleasant experience. It explains why analyzing review sentiment is vital for understanding customer needs.

2.2.2.1 METHOD 1:

Sentiment analysis is a crucial indicator of understanding customer satisfaction and experience. It is an automated process that analyzes the text and identifies the emotions behind it. This process helps understand whether the review is positive, negative, or neutral with the help of the values mentioned below.

VALUE	RESULT
-1	Extremely negative
0	Neutral
+1	Extremely positive

Table 1. Sentiment analysis scale

This process is carried out by *get_nrc_sentiment* in *the syuzhet* package. It takes the entire sentence and analyzes the emotions behind it. As a result, scores for the below-mentioned emotions are calculated.

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust
- Negative
- Positive

```
e1 <- get_nrc_sentiment("Best phone for gaming under 15k.build quality and battry life ")
e1[,]
anger anticipation disgust fear joy sadness surprise trust negative positive
    0            0       0    0   0       0        0     0        0        1
```

Fig 2. Sentiment analysis example

The following example has taken a customer review from our sample data as input. The results have one in the ‘positive’ field and zero in other fields. The sentiment score is calculated by dividing this value by the number of emotions displayed. In this case, it will be one divided by one. The score for the above sentence is one. As per *Table 1,* this sentence is classified as positive. Similarly, we have classified all the reviews into positive and negative. Below is an example of a negative review with its corresponding score.

	Review	Ratings	Score
2457	Although Built quality and camera is good in this price. But...	2	-0.0434782608695652
143	Battery is not good it takes lot of time to charge and charg...	4	-0.0476190476190476
919	Writing this review after almost a month of use. I was so h...	1	-0.0476190476190476

Fig 3. Negative review score

The reviews with neutral opinions are given a score of zero.

4547	Average result of camera.	4	0
4548	Battery life and charge time are not good.. even realme/oppo midrange p...	3	0
4554	Camera quality is not good, battery also very average	3	0

Fig 4. Neutral review score

The reviews with positive tone of language are given a positive score.

	Review	Ratings	Score
4714	Very very good performance	5	0.2
4719	Product is very awesome good quality, but fingure print sensor is very po...	5	0.2
4723	Everything good	5	0.2
4726	Under 10k is a best best phone good camera and this phone look and d...	4	0.2
4728	Good	4	0.2

Fig 5. Positive review score

2.2.2.2 METHOD 2:

Review ratings given by the customers themselves are taken as a parameter to help understand their viewpoint better. The below scale is used for classifying the reviews as positive and negative.

VALUE	RESULT
<3	Extremely negative
3	Neutral
>3	Extremely positive

Table 2. Review rating scale

2.2.2.3 METHOD 1 + METHOD 2:

The methods mentioned earlier have their limitations and accuracy issues. Sentiment analysis cannot be a hundred percent accurate measure since language can be interpreted by humans and computers differently. Computers will not be able to understand jokes, sarcasm, slang, or irony, which are understood correctly by humans (Anon., n.d.). Hence, we combined both the method's outputs to conclude.

SENTIMENT SCORE	REVIEW RATING	OUTPUT
>0	>3	POSITIVE
<0	<3	NEGATIVE

Table 3. Final classification scale

At the end of classification, we have the below chart showing the positive and negative word distribution. It is clear from the plot below that the reviews are highly positive.

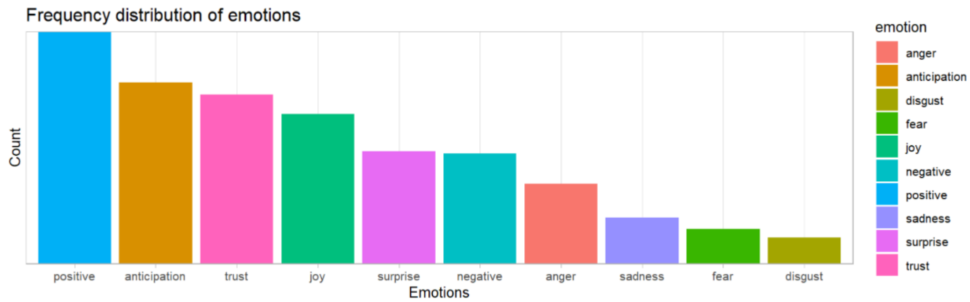


Fig 6. Review distribution

2.2.3 POSITIVE AND NEGATIVE REVIEW ANALYSIS:

2.2.3.1 DOCUMENT TERM MATRIX

The reviews are now classified as positive and negative, respectively. To analyze the reviews, we are converting them into a corpus first. A corpus is a large structure of set of texts in a single or multiple language. The corpus function returns a corpus object with all metadata and contents as shown below.

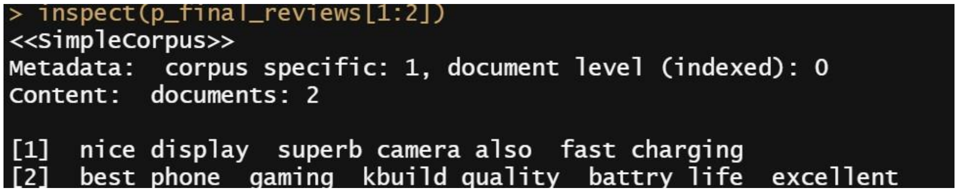
```
> inspect(p_corpus[1:2])
<<SimpleCorpus>>
Metadata:  corpus specific: 1, document level (indexed): 0
Content:  documents: 2

[1] \n  Nice Display and Superb Camera, also very fast charging\n

[2] \n  Best phone for gaming under 15k.build quality and battry life are
 excellent\n
```

Fig 7. Corpus example

This corpus object is used to create a Document term matrix. This matrix contains frequency of terms in the documents. The rows are collections and columns depict each term. We are using TF (Term Frequency) for determining the frequency of terms. The calculation is as below:

$$\textit{Term frequency} = (\textit{Number of times the word occurred in the text}) \\ / (\textit{Total number of words in the text})$$

Data cleaning is done with the help of lemmatization. After considering the contextual meaning, the lemmatization process converts the text to its base form. Whereas stemming cuts of the prefix and suffix without any consideration. Data is further cleaned by removing punctuation, numbers, and common stop words and converted to lower case. Corpus in *Fig.7* looks like the below after data cleaning


```
> inspect(p_final_reviews[1:2])
<<SimpleCorpus>>
Metadata:  corpus specific: 1, document level (indexed): 0
Content:  documents: 2

[1]  nice display  superb camera also  fast charging
[2]  best phone  gaming  kbuild quality  battry life  excellent
```

. Fig 8. Corpus after data cleaning

With the help of the cleaned data, the most used words were in positive and negative reviews were visualized. It was achieved with the help of *worldcloud*. The terms that occurred more than a hundred times are considered and the rest ignored.

Fig 9. Positive review term frequency

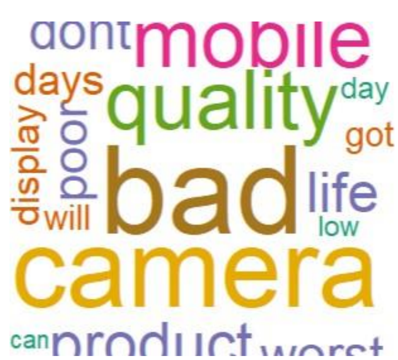


Fig 10. Negative term frequency

The Data term frequency matrix is created for both positive and negative reviews with all the above measures included.

2.2.3.2 TOPIC MODELLING

Topic modelling is a two-step process.

- STEP 1: Generate list of topics covered by documents.
- STEP 2: Later, group the documents based on the topics generated.

Every document has a mixture of numerous topics. This modelling figures out which topic's presence is strongest and finalizes it.

STEP 1:

Latent Dirichlet Allocation is a parameterized model. This model counts words and groups similar word patterns together. It helps us conclude a topic based on how often they appear together. Knowing k number of optimal topics for the documents is one of the critical aspects of the LDA method. It is done with the help of a coherence score. This score is calculated by checking whether the words in the same topic make sense when they are put together. It is calculated both for positive and negative topics. Below are the values achieved.

CATEGORY	K
POSITIVE	9
NEGATIVE	8

Table 4. K value

The accuracy of this value is checked with the help of a coherence matrix. The result is plotted using *ggplot* to help us understand the maximum coherence score. Below is the best topic in positive review calculated through coherence score.

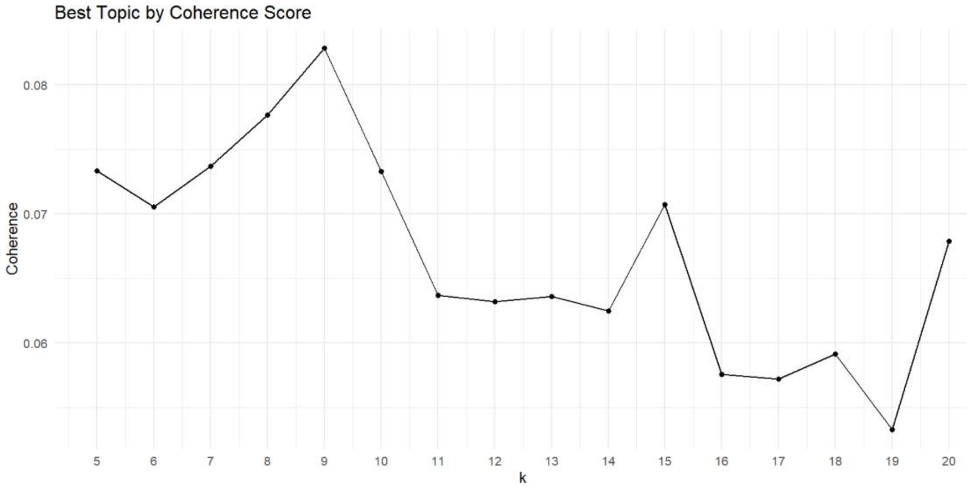


Fig 11. Positive best topic - 9

TOPIC GENERATION:

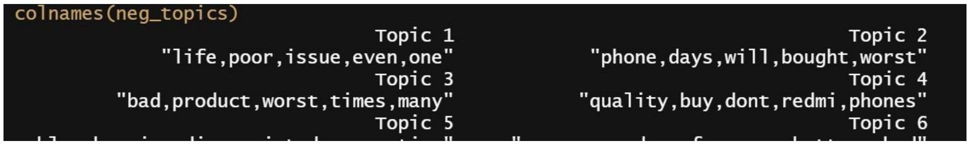
Topics are generated using the topic modeling process. It generates a list of topics covered by documents and the topics at the end of group documents. Distribution over terms for a topic and probability distribution over topics in the documents are created. The topic number generated for positive with the help of a coherence score was 9. So, a list of 9 topics is created as mentioned below

```
> positiveLdaTopics.terms
      Topic 1      Topic 2       Topic 3      Topic 4      Topic 5       Topic 6
 [1,] "price"      "quality"     "phone"      "one"        "battery"     "good"
 [2,] "best"       "camera"      "like"       "oneplus"    "fast"        "money"
 [3,] "phone"      "performance" "everything" "nord"       "camera"      "value"
 [4,] "range"      "super"       "camera"     "display"    "life"        "excellent"
 [5,] "worth"      "best"        "just"       "using"      "fingerprint" "overall"
 [6,] "great"      "superb"      "buy"        "really"     "great"       "performance"
 [7,] "working"    "also"        "use"        "experience" "also"        "really"
 [8,] "overall"    "mobile"      "dont"       "plus"       "reader"      "user"
 [9,] "smartphone" "better"      "looking"    "smooth"     "charging"    "delivered"
[10,] "buy"        "look"        "features"   "screen"     "time"        "best"
      Topic 7  Topic 8   Topic 9
 [1,] "good"   "phone"   "nice"
 [2,] "mobile" "awesome" "good"
```

. Fig 12. Positive topic generation

Similarly, topics were generated for negative reviews as well. Topic 1 in *Fig 12* focuses on a customer who felt this is a great phone in this price range.

The next goal is to label these topics generated. Labeling the topics increased the readability and helps easy understanding. The topics are labelled with the most frequently occurred words in that topic. The below snipped shows the title labels of the positive reviews.



```
colnames(neg_topics)
                              Topic 1                                 Topic 2
              "life,poor,issue,even,one"          "phone,days,will,bought,worst"
                              Topic 3                                 Topic 4
           "bad,product,worst,times,many"        "quality,buy,dont,redmi,phones"
                              Topic 5                                 Topic 6
```

Fig 13. Topic Labelling

3TOP FACTORS CAUSING SATISFACTION

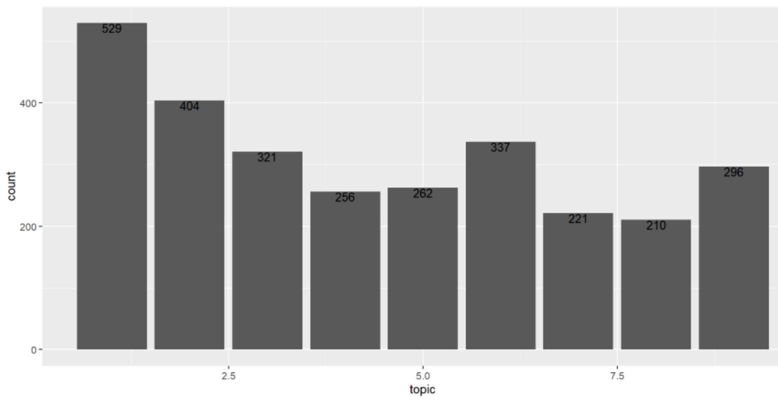


Fig 14. Topic satisfying the customer

From the above chart it is evident that Topics 1,2 and 6 have the highest count. Looking further into these topics will make one understand the customer and his needs better.

```
   Topic 1      Topic 2       Topic 3      Topic 4      Topic 5       Topic 6
] "price"      "quality"     "phone"      "one"        "battery"     "good"
] "best"       "camera"      "like"       "oneplus"    "fast"        "money"
] "phone"      "performance" "everything" "nord"       "camera"      "value"
] "range"      "super"       "camera"     "display"    "life"        "excellent"
] "worth"      "best"        "just"       "using"      "fingerprint" "overall"
] "great"      "superb"      "buy"        "really"     "great"       "performance"
] "working"    "also"        "use"        "experience" "also"        "really"
] "overall"    "mobile"      "dont"       "plus"       "reader"      "user"
] "smartphone" "better"      "looking"    "smooth"     "charging"    "delivered"
] "buy"        "look"        "features"   "screen"     "time"        "best"
   Topic 7   Topic 8     Topic 9
] "good"    "phone"     "nice"
] "mobile"  "awesome"   "good"
] "redmi"   "much"      "product"
] "amazing" "love"      "happy"
] "note"    "device"    "got"
] "budget"  "camera"    "just"
] "pro"     "delivery"  "display"
] "loved"   "amazon"    "thanks"
] "bought"  "satisfied" "oxygen"
```

Fig 15. Top 3 customer satisfying topics

Topic 1: The customer liked that such a great working phone is in this price range. The price factor plays a significant role in customer satisfaction.
Topic 2: The customer loved the camera quality and its sleek design. In addition to price, camera quality and design play a vital role.
Topic 6: The customer loved that it got delivered on time and its exceptional performance.

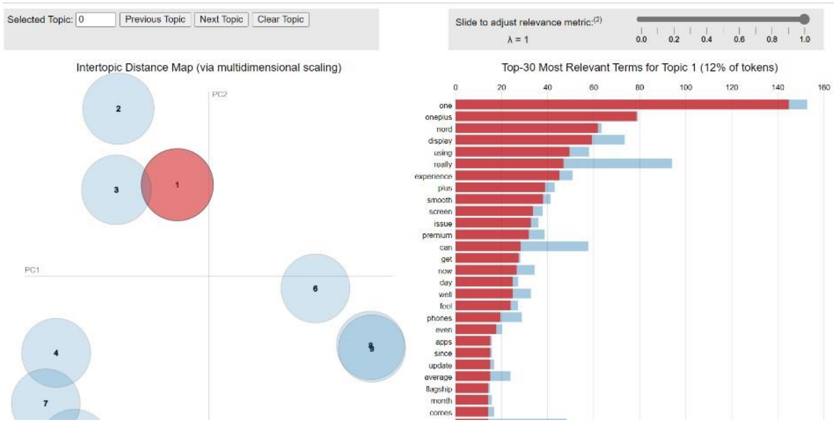


Fig 16. Most frequent terms in TOPIC 1

So, the three top factors would be PRICE, CAMERA, DESIGN.

3TOP FACTORS AFFECTING SATISFACTION:

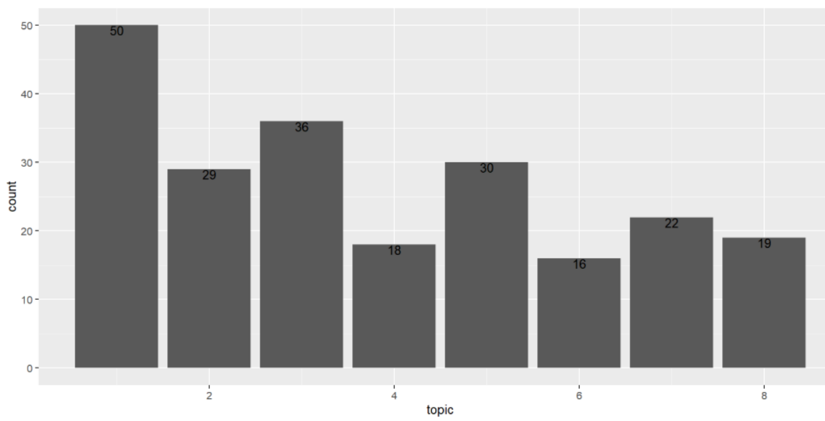


Fig 17. Topics affecting satisfaction

Fig 16 clearly states that topics 1,3, and 5 have the maximum occurrence. To understand what caused dissatisfaction to the customers we need to investigate these topics.

```
negativeLdaTopics.terms
      Topic 1   Topic 2   Topic 3    Topic 4    Topic 5        Topic 6
[1,]  "life"    "phone"   "bad"      "quality"  "problem"      "camera"
[2,]  "poor"    "days"    "product"  "buy"      "hanging"      "good"
[3,]  "issue"   "will"    "worst"    "dont"     "disappointed" "performance"
[4,]  "even"    "bought"  "times"    "redmi"    "money"        "battery"
[5,]  "one"     "worst"   "many"     "phones"   "time"         "bad"
[6,]  "just"    "use"     "phone"    "got"      "video"        "also"
[7,]  "waste"   "much"    "please"   "colour"   "facing"       "cheap"
[8,]  "ram"     "network" "pathetic" "expected" "fast"         "getting"
[9,]  "display" "amazon"  "display"  "hang"     "issues"       "please"
[10,] "fast"    "screen"  "totally"  "mobiles"  "charge"       "voice"
      Topic 7         Topic 8
[1,]  "mobile"        "battery"
[2,]  "quality"       "bad"
[3,]  "dont"          "device"
[4,]  "oneplus"       "charging"
[5,]  "never"         "backup"
[6,]  "automatically" "also"
[7,]  "time"          "low"
```

Fig 18. Top 3 customer dissatisfying topics

Topic 1: Customer is concerned about the ram performance.
Topic 2: The customer is not satisfied with the display.
Topic 3: The customer is facing phone hang and charging issues.

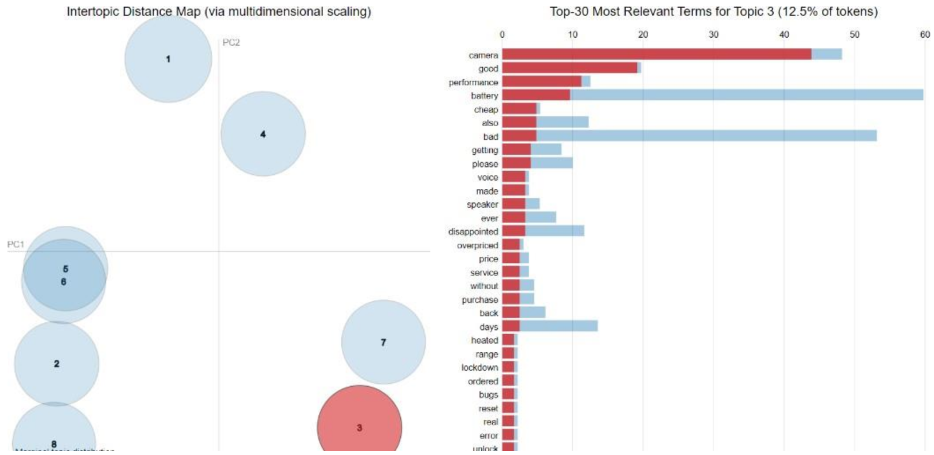


Fig 19. Most frequent terms in TOPIC 3

So, the three top factors would be RAM, DISPLAY, CHARGING.

CONCLUSION:

Understanding customer sentiments play a vital role in product success in the market. To do so, one needs to be able to identify positive and negative reviews. This process was covered in the above part with the help of sentiment analysis. We also found the most common terms that attract the customers and vice versa. This result helps the seller with what to focus on next to increase sales.

References

Anon., n.d. *Forbes.com.* [Online]
Available at: https://www.forbes.com/sites/jiawertz/2018/11/30/why-sentiment-analysis-could-be-your-best-kept-marketing-secret/

Anon., n.d. *UNCTAD.* [Online]
Available at: https://unctad.org/news/covid-19-has-changed-online-shopping-forever-survey-shows

Pettinger, T., n.d. *economicsHelp.* [Online]
Available at: https://www.economicshelp.org/blog/1368/economics/keynesian-stimulus/#:~:text=Keynesian%20fiscal%20stimulus%20is%20a%20decision%20by%20the,sector%20saving%20and%20unused%20resources%20in%20the%20economy.

Topic Modeling with R (slcladal.github.io)

Topic Modeling using R · knowledgeR (rbind.io)

Comparison of different Methods for Univariate Time Series Imputation in R (arxiv.org)

How to Deal with Missing Data (mastersindatascience.org)