A

Report On

Social Media Analysis for Business

Student Name: -

Student Id: -

Table of Contents

[**Abstract:** 5](#_Toc40341242)

[**Introduction** 6](#_Toc40341243)

[**Design and Discussion of Data Sources: -** 7](#_Toc40341244)

[Wikipedia Web Scraping: 7](#_Toc40341245)

[Twitter Sentiment Analysis: - 8](#_Toc40341246)

[Naïve Bayes Classification: - 13](#_Toc40341247)

[(SRS)System Requirement Specification: 14](#_Toc40341248)

[**Results: -** 15](#_Toc40341249)

[**Business Story Telling: -** 17](#_Toc40341250)

[**Conclusion: -** 17](#_Toc40341251)

[**Limitations: -** 18](#_Toc40341252)

[**Implication: -** 18](#_Toc40341253)

[**Recommendation: -** 18](#_Toc40341254)

[**References: -** 18](#_Toc40341255)

**List of Figure**

[**Figure 1** 8](#_Toc39910301)

[**Figure 2** 9](#_Toc39910302)

[**Figure 3** 9](#_Toc39910303)

[**Figure 4** 10](#_Toc39910304)

[**Figure 5** 10](#_Toc39910305)

[**Figure 6** 11](#_Toc39910306)

[**Figure 7** 13](#_Toc39910307)

[**Figure 8** 14](#_Toc39910308)

[**Figure 9** 15](#_Toc39910309)

[**Figure 10** 15](#_Toc39910310)

[**Figure 11** 16](#_Toc39910311)

Biography

# **Abstract:**

As a results of net technology, web users have an oversized quantity of information on the online. These users not solely use the offered resources, however additionally offer feedback. Therefore, extra helpful info is generated. Social networks generate massive amounts of information. a lot of and a lot of individuals area unit expressing their opinions on social networks like Facebook, Twitter and YouTube. The microblogging web site could be a made supply of information for belief mining and sentiment analysis. Sentiment analysis could be a tongue process and data extraction task that identifies the user's perspective and is employed to extract opinions regarding events, products, and folks and to grasp current trends or newest technologies. We have continuously influenced by people's concepts. The concepts and opinions of others have continuously influenced ours. The explosive growth of net a pair of.0 has fueled podcasting, blogging, tagging, RSS feeds, social bookmarking, and social media. As a result, individuals area unit a lot of curious about mining these Brobdingnagian knowledge resources and seeking input. Sentiment analysis or opinion mining could be a method of scheming the opinion, sentiment, and judgment of the text. This report describes numerous challenges and applications of sentiment analysis. Let's take a better check up on the various approaches to playacting emotional and opinion calculations. It describes numerous observance or data-driven approaches to Storm Troops, as well as Naïve Byes, most Entropy, SVM, Voted Perceptron’s, and their strengths and weaknesses. we are going to additionally see a replacement dimension of emotional analysis through psychological science, primarily through the works of Janice Weave. There you'll learn to grasp judgment, narrative views and also the structure of discourse. you'll additionally study specific topics in sentiment analysis and modern add those areas. This project addresses the difficulty of sentiment analysis on Twitter. That is, it classifies tweets in line with the feeling expressed within the tweet: positive, negative, or neutral. Twitter is an internet microblogging and social media platform that permits users to write down short standing updates of up to one hundred forty characters. this can be a apace increasing service with over two hundred million registered users , of that one hundred million area unit active users, 1/2 whom log into Twitter daily and regarding 205 million daily. It generates virtually ten million tweets . the target of this project is to develop a practical classifier for correct and automatic classification of sentiments from unknown tweet streams.

# **Introduction**

We selected to use Twitter as a result of we have a tendency to believe it's a higher approximation of belief, in contrast to ancient web articles and net blogs. the rationale is that Twitter has rather more relevant knowledge than ancient blogging sites. Also, responses on Twitter area unit quicker and additional. Analysis of public sentiment is incredibly vital in macro-level socio-economic phenomena, like prognostication stock exchange rates for a selected company. this will be done by analyzing the overall belief of the corporate over time and mistreatment economic tools to search out a correlation between belief and also the market price of the corporate. Businesses can even estimate however responsive their merchandise area unit to the market, that area unitas of the market are responsive, and that area unit negative (on Twitter. as a result of you'll be able to transfer a sequence of geotagged tweets, firms will use this info to investigate the explanations behind of your geographically differentiated responses and to develop acceptable solutions, like making the correct market section. it's a sort of language process. it's belief a couple of specific product. feelings is additionally known as opinion mining. examine opinion a couple of specific product created during a web log post, comment or tweet. you'll be able to track a particular topic and plenty of firms use it to trace or monitor their product or standing, for instance if somebody is assaultive your complete on social media, sentiment analysis posts you'll be able to produce alerts for posts that have a really opinion score negative, and also the basic tasks of opinion analysis area unit expressed during a document, sentence, or entity characteristic. Divide the polarity of a selected text into a document, sentence, whether or not the opinion is positive, negative, or neutral. The classification of feelings involves searching for emotional states like "angry", "sad", "happy", etc. this is often important to grasp your target market. It primarily measures ROI, that helps produce new audiences and additionally affects current audiences, that additionally describes the audience's approach to the organization. If you recognize the content your audience listens to and brings you new audiences. Is it a media site? Knowing the correct influencer is incredibly vital. continuously air the lookout for those who area unit presently mistreatment your product or service. Take a glance, this can assist you gain a correct understanding of your client. hassle making compelling and effective content. Without it, making effective content that engages customers is incredibly troublesome. Adapt it consequently. Get concepts for social media campaigns here. Analysis of social networks doesn't reveal the present state of the corporate. If you have got reached the correct audience. It does not matter if you are late currently. we are going to still execute our strategy supported the proper social media platform and target market. facilitate improve your social media plans. Social media activity says tons regarding whether or not you are mistreatment the correct social media. Provides results and improves ROI. there's no additional concern regarding whether or not you'll be able to hear your message. it's terribly helpful to enhance. Social media analysis needs knowledge from social media viewers and also the tools to investigate it. it'll additionally inform you of the strengths and weaknesses of you and your competitors. try and straighten your strengths and cut back your weaknesses. we have a tendency to additionally make the most of the weaknesses of our competitors. keep earlier than your competition. whether or not you're late or not. Keep performing on your strategy consequently.

# **Design and Discussion of Data Sources: -**

## Wikipedia Web Scraping:

Web scratching, web accumulate, or web information handling is data scratching wont to separate data from sites. web scratching code will get to the globe Wide web legitimately abuse the machine-decipherable content exchange convention or by means of an online program. Scratching web clients are frequently done physically by code clients, anyway as a rule alludes to programmed forms authorized abuse bots or web crawlers. this is frequently a kind of rehashing inside which bound data is gathered and gotten from the net, once in a while in an exceptionally focal local information or PC program, for later recovery or investigation. web scratching of sites includes watching out and extricating pages. Recover is that the exchange of a page (what the program will once it shows the page). Along these lines, web travel is that the primary component of web scratching, that looks for pages for any procedure. When recuperated, the extraction is performed. you'll have the option to investigate page content, search, group, duplicate data into a PC program, and then some. web scrubbers now and then remove one thing from a page and use it somewhere else for a particular reason. Models grasp finding a notoriety and flagging, or an association and its PC address and rehashing it to a posting (get contact data). web scratching is utilized for contact scratching, just as web classification, web mining and information handling, on-line esteem alteration recognition and worth examination, item audit scratching (to see the opposition ), property postings, earth science data grouping. it's utilized as a component of uses utilized for recognition, site correction identification, explore, on-line nearness and name interest, web mashups, and web data mix. sites square measure made abuse content based markup dialects ​​(HTML and XHTML) and rarely contain an outsized amount of accommodating data in content arrangement. Be that as it may, most sites square measure intended for human completion clients and don't encourage programmed use. Therefore, specific instruments and code are created to encourage scratching sites. a substitution kind of web scratching tunes in for data sources from web servers. for example, JSON is regularly utilized as a vehicle stockpiling system between a purchaser and an online server.

## Twitter Sentiment Analysis: -

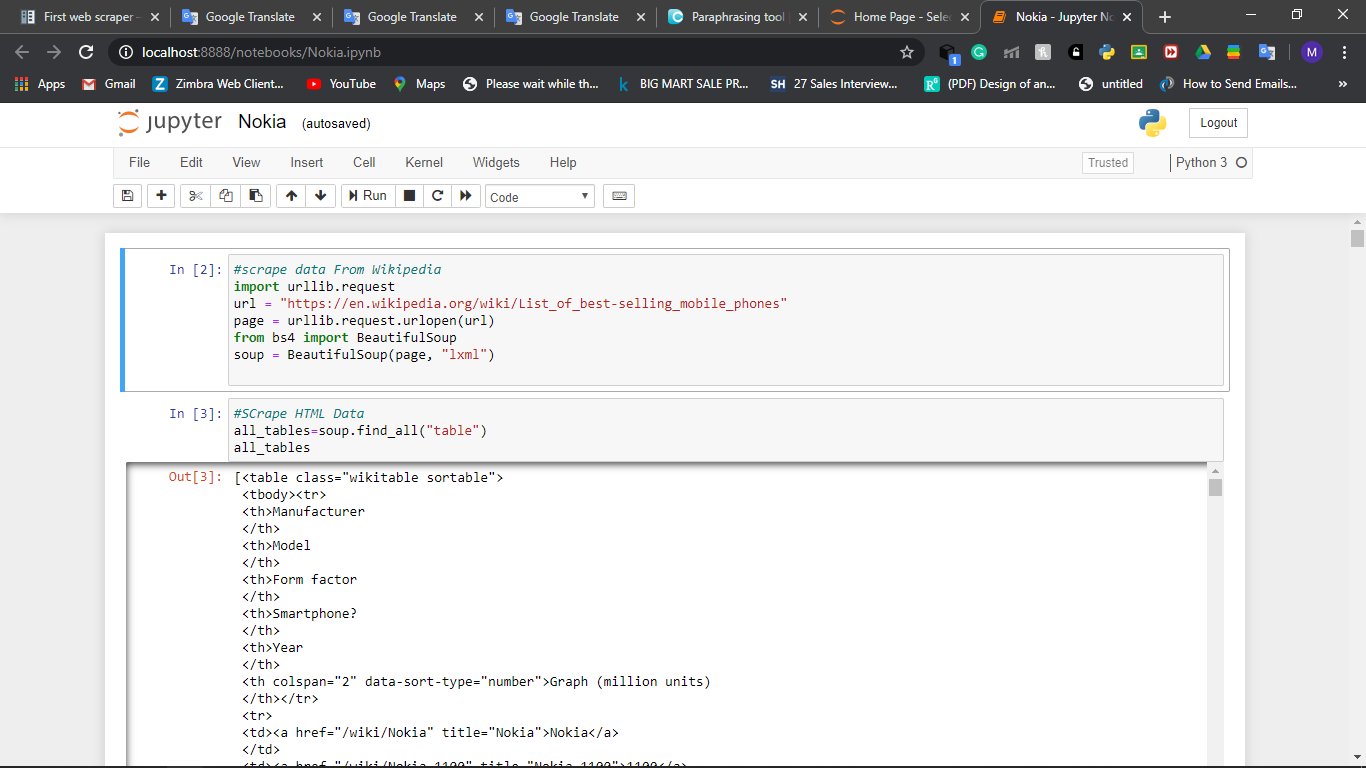
**Introduction to Problem: -**

**Technologies and Platform:-** There are various third party tools which is use to scrape the data and visualizations, Python modules and Library are used in this project like Beautiful Soup, requests, NLTK, re, Sklearn, Wordcloud, Numpy, Pandas, Textblob, Matplotlib.

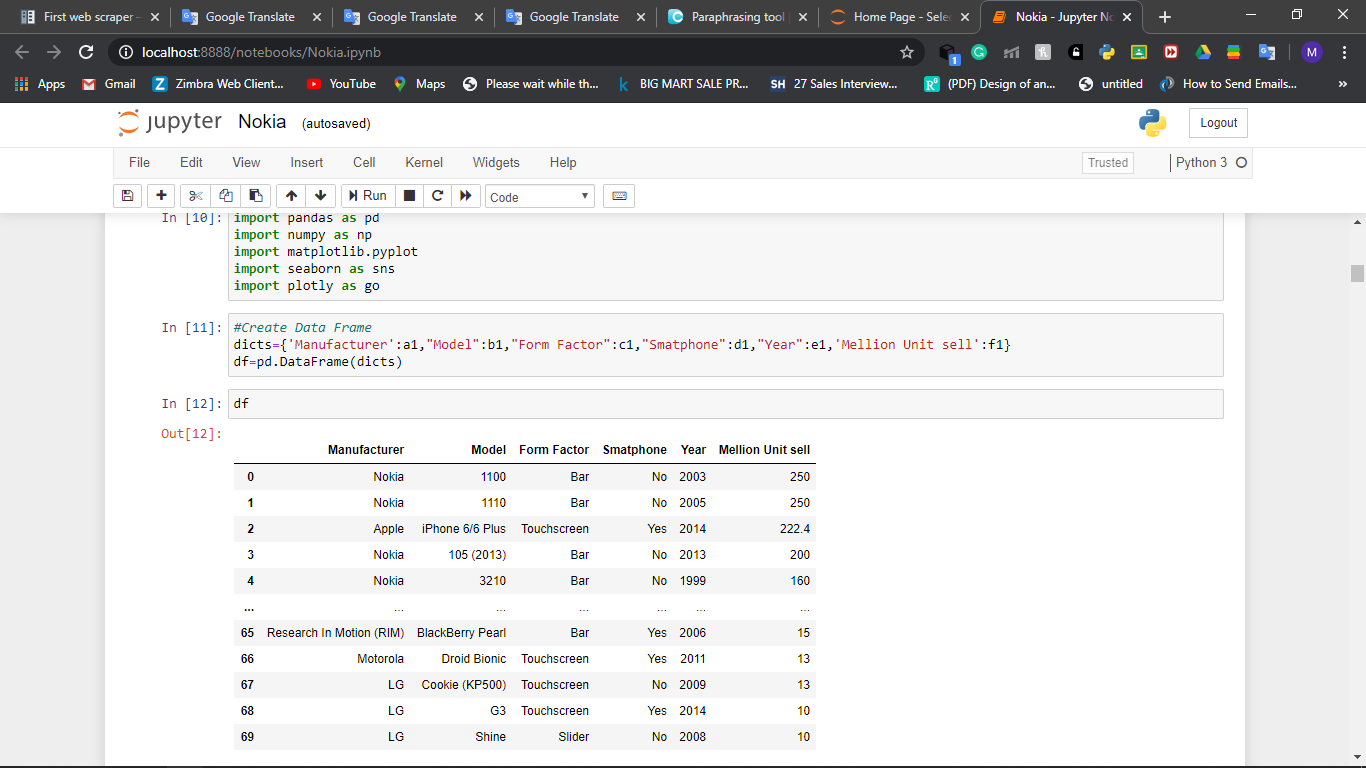
**BeautifulSoup: -** Web scraping is a web knowledge assortment methodology. you'll be able to scratch the net by repeatedly pasting the lyrics on your favourite song. However, the term "web scraping" is sometimes mentioned during a method that involves automation. Some websites aren't getting hot once motor vehicle scrapers collect insights, others do not care. If you are scratching the proper relevant full-featured tutorial pages, you may seldom use retarders. yet, before beginning an outsized project, Nursing Associates United Nations agency have set to review themselves to form certain they are doing not violate the Terms of Service are honest. to seek out out the burden on the legal aspects of net scraping, hunt the instrument on elegant net scraping knowledge. stunning Soup provides some easy shapes and representations for navigating, searching, and modifying the anatomy tree. you are doing not ought to validate the cipher unless the document doesn't specify the associated cipher and Exquisite Soup cannot mechanically acknowledge the cipher. Therefore, you simply got to force it to specify the primary coding. sensible soup is one among the foremost common Python parsers like lxml and html5lib, and it will perform utterly completely different parsing and modification speeds for larger flexibility. let's examine what we are able to do with it.

**Case Study: -**In this Project we follow different steps to collect the data From Wikipedia table and Twitter using Scrape , convert the data in a useable form, and make an analytics model for visualization purposes.

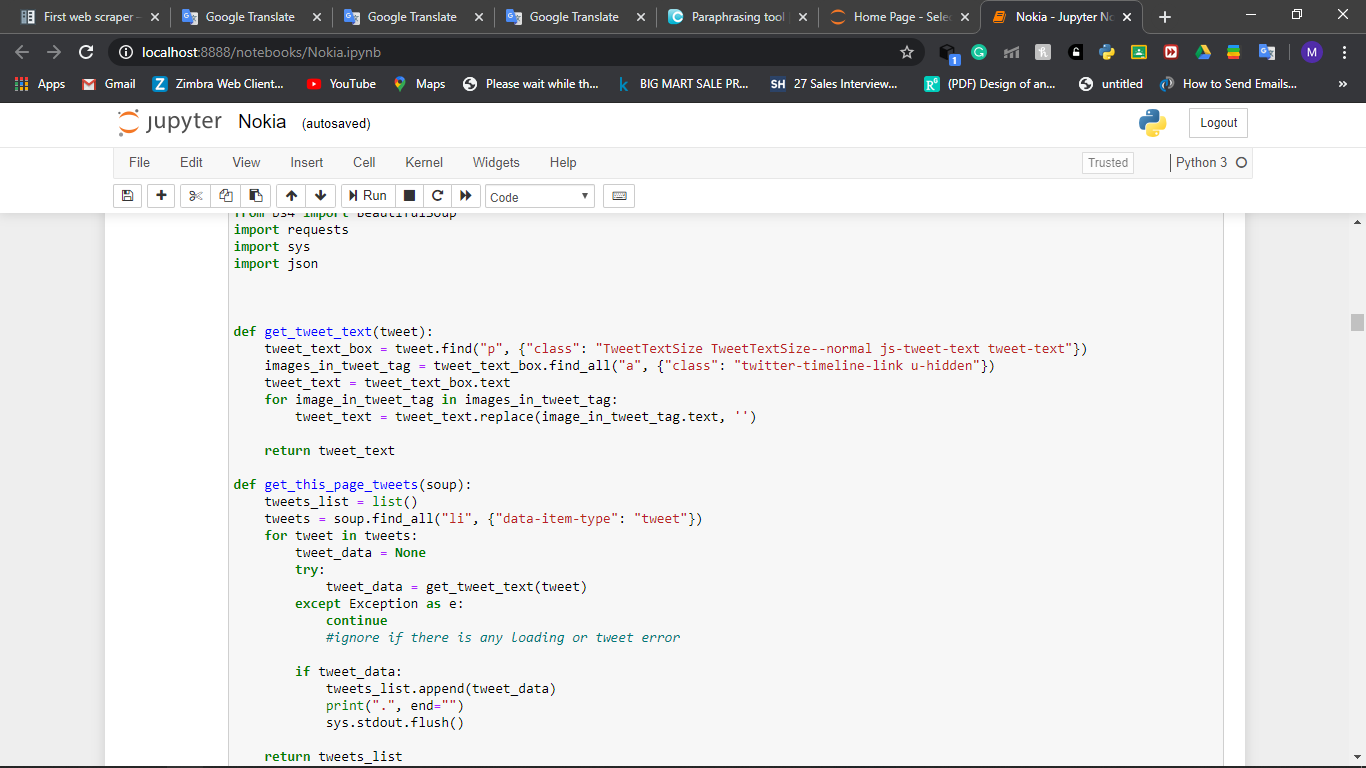
**Data Collection: -** In this project the Model needed the Wikipedia table data that help to compare the different brands of the smartphone data and Reviews of twitter that’s why we tend to collect the information victimisation scrapping ways. The row information is collected through the BeautifulSoup library module. Extracting tweets from nomenclature content: First examine the tweet half on web content. you will see that each one the tweets unit of measurement enclosed in li nomenclature tag. Actual tweet text is at intervals a p tag that's that the descendent of li tag. we are going to initial get all the li tags and then p tags from each li tag. Text contained at intervals the p tag is what we wish.



**Figure 1**

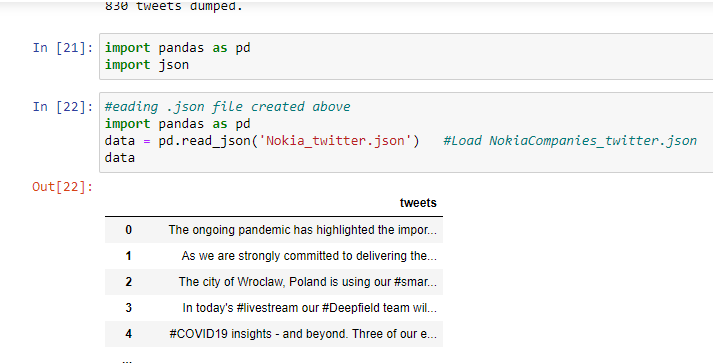


**Figure 2**

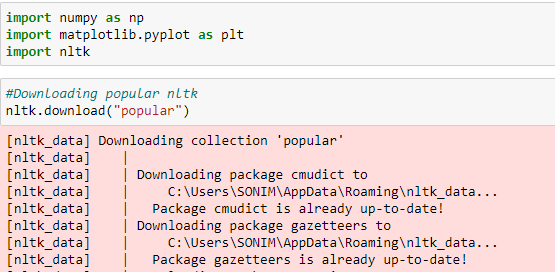


**Figure 3**

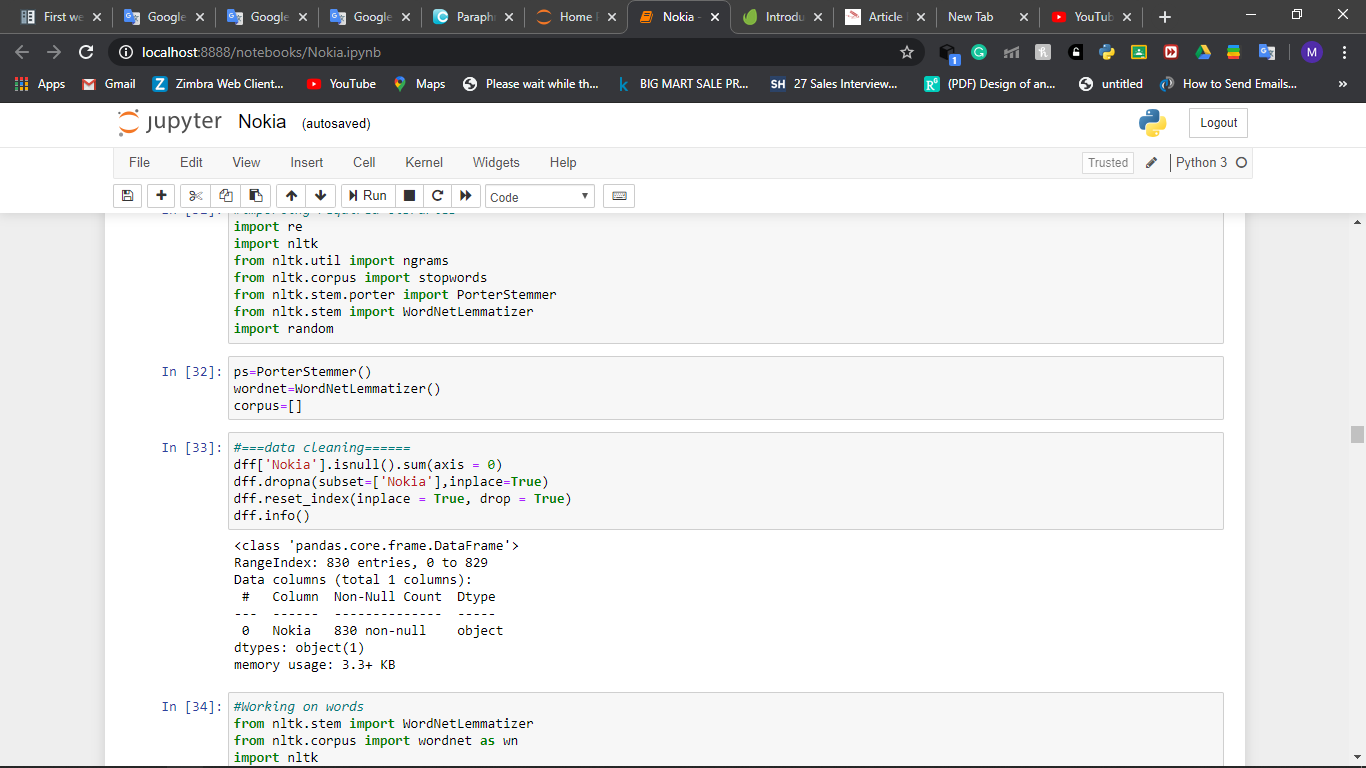
**Create DataFrame:-** A one-dimensional labelled array that may contain any kind of data with axis or index labels. associate degree example of a string object could be a column during a information frame.Using the Pandas class tends to supply a DataFrame, as shown in Figure four.



**Figure 4**

**NLTK:-** Normal Language Toolkit (NLTK) might be a stage acclimated produce content examination programs. The stage was initially propelled by Steven Bird and Edward Lauper in 2001 with the etymology course at the University of Pennsylvania. there's a book packaged with a book alluded to as tongue process with Python. The tongue Toolkit, or a great deal of ordinarily NLTK, might be a lot of libraries and projects for representative and applied arithmetic procedure of land tongue (NLP) composed inside the Python counterfeit language. NLTK is Associate in Nursing truncation for Language Toolkit. This toolbox is one in everything about principal ground-breaking specialized strategy libraries and contains bundles for machines to realize human language associated reactions with right reactions. Tokenization, Derivation, Derivation, Punctuation, Character Count, Word Count square measure the different bundles which will be referenced during this instructional exercise. correspondence methodologies square measure the control or appreciation of content or discourse by a bundle or machine. An associated similarity is that people demonstration, see each other, and react with satisfactory responses. In NLP, this communication, comprehension, and reaction is made by the pc, not the person. NLTK comprises of the first normal calculations like tokenization, part labeling, deduction, assumption investigation, subject division, element acknowledgment, and so on. NLTK enables PCs to break down, preprocess, and see transcripts. This undertaking will in general exchange the well known NLTK exploitation this library. It furthermore figures the exactness of the structure model.

**Figure 5**



**Figure 6**

**Data Pre-processing:**-

In This Data Preparation imports Required libraries for development. information arrangement could be a strategy that changes the gathered column information into graspable AI information. For data preparation, different kind of library is used like NumPy, Pandas, matplotlib. In real-world scrapped data is not complete or shows some Nan Values so This step is Required. During the Data Cleaning process we follow some Following Steps:-

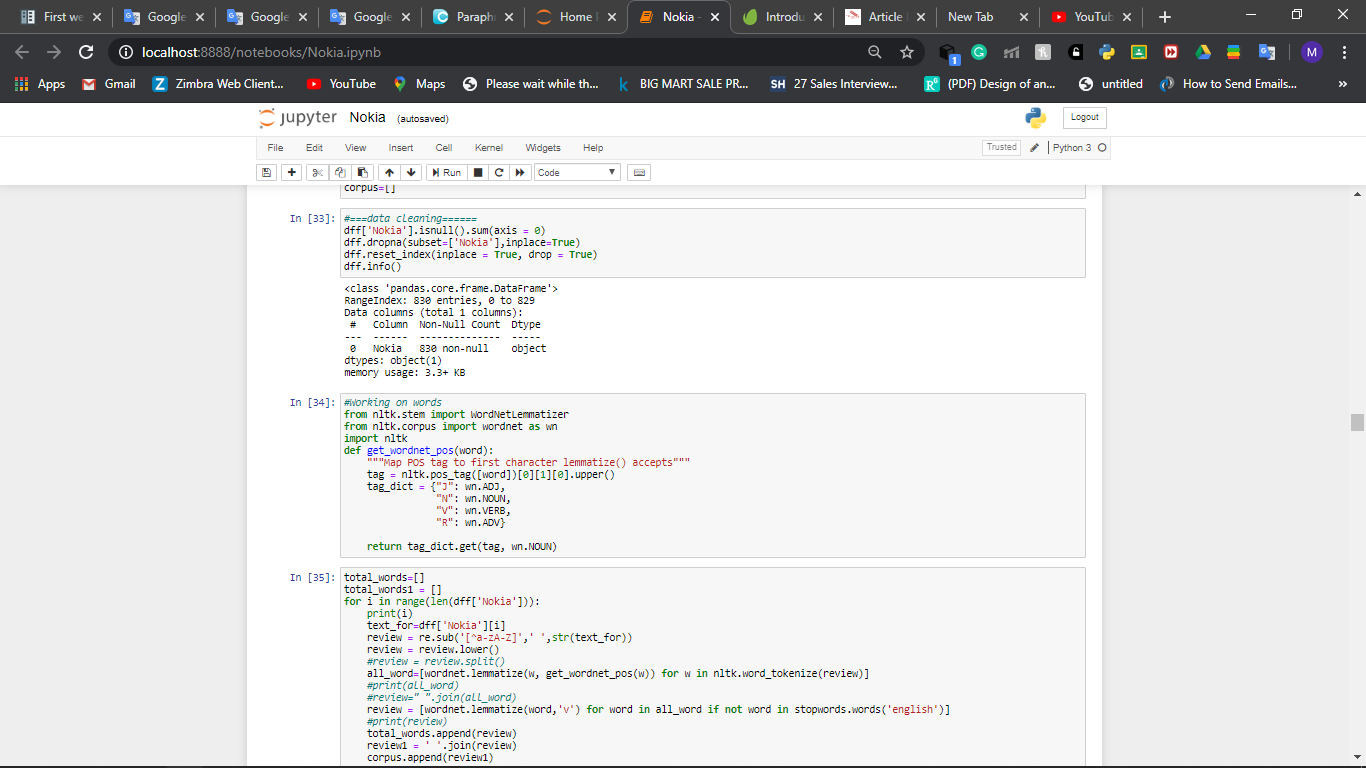
1.Data Cleaning: Data cleaning is a process of detection and correction of corrupted or inaccurate records in a set of records, tables, or information, which is a distinctive portion of incomplete, inaccurate, inaccurate, or irrelevant information. I will. To do. Then replace, change, or remove dirty or chunky knowledge.

2.Data Integration: Data integration systems must gradually use machine learning to modify many integration tasks. Examples include knowledge of data schema cataloging and inference.

3. Data Transform: Data is standardized, collected, and generalized.

4.Data Reduction: Data reduction is the conversion of empirical observations or digitized numerical or alphabetical information from empirical observations in a modified, organized, and simplified form. The essential plan is to turn a large amount of knowledge into essential components.

5.Data Discretization: Numerical knowledge discretion is one of the most well-known knowledge preprocessing tasks in information discovery and data processing. -To reduce the data from a large numerical domain to a set of categorical values, the reduction mechanism is considered discrete.



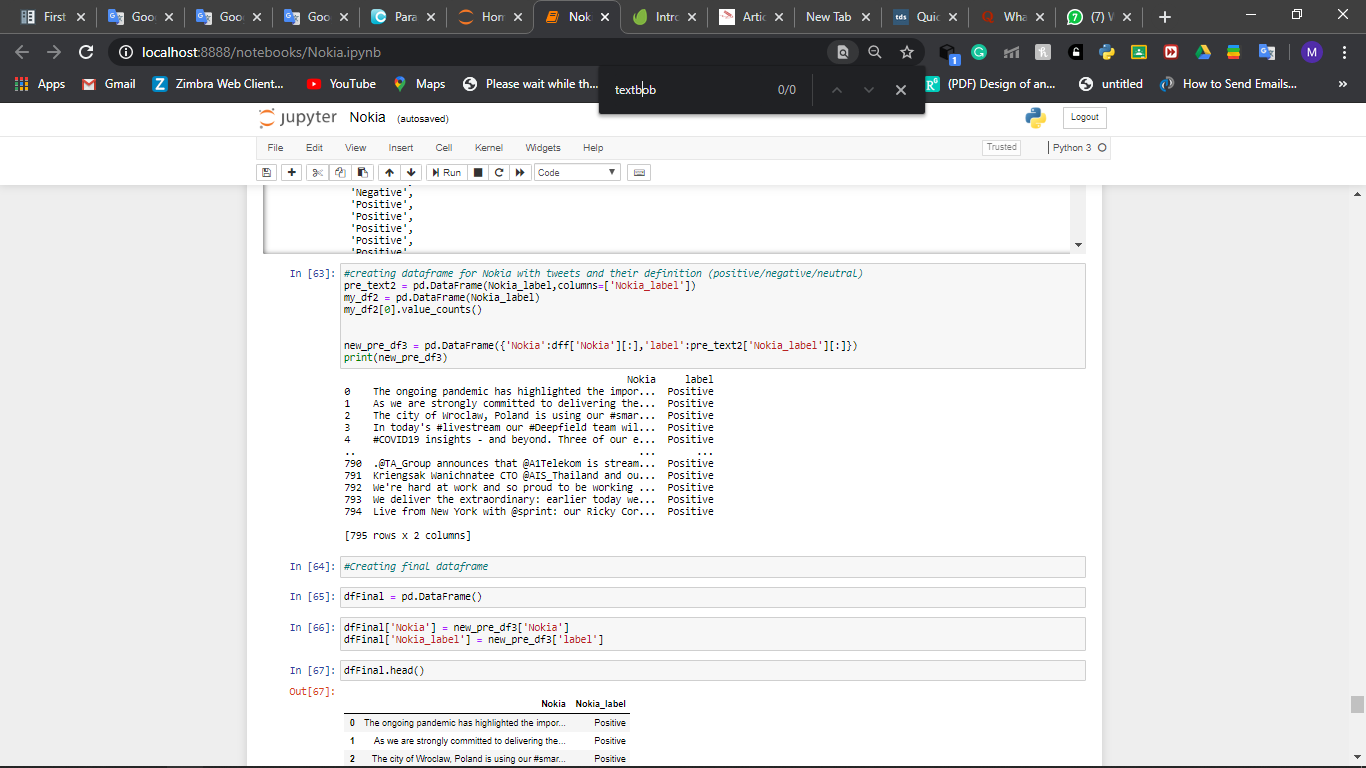
**Figure 7**

**Sentiment Analysis: -**

This is an automatic method that determines if a text expresses a positive, negative, or neutral opinion a few product or topic. By victimization Opinion Analytics, businesses do not have to pay endless hours tagging client knowledge like survey responses, reviews, support tickets, and social media comments. Opinion analysis helps businesses monitor complete name on social media and gain insights from client feedback. Let's take a better check up on what sentiment analysis is, however it's done, and the way it will facilitate your business.

**Finding polarity**: -

The Polarity is a conclusion examination is to recognize the passionate direction (positive, unbiased, negative) of composed or spoken correspondence. various assortments of estimation investigation exemplify fine-grained assumption examination (for instance, horribly positive, positive, unbiased, negative, awfully negative) that will expand precision at the degree of extremity, and notion investigation that is explicit. Emotions in articulation (eg, joy, pity). , Frustration, shocks, and so forth.)

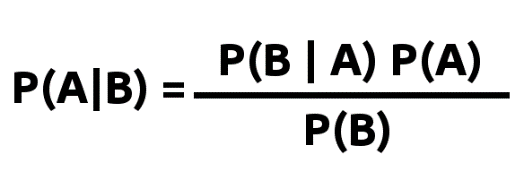


**Figure 8**

* **Working of the Project**
  + Step1 import Libraries and Scrape Wiki table From Wikipedia and Open twitter username and scrape Tweets data using BeautifulSoup scrapping tool.
  + Step2 Create data Frame and adjust the data.
  + Step3 clean the extra values, Spaces, align the make a stable complete format of the data.
  + Step4 Download the NLTK Popular expressions.
  + Step5 Normalize the data in form of list and compare through the NLTK Popular data.
  + Step6 Check the complete data if there is any null value replace Nan value.
  + Step7 Check the positive negative and natural sentiment using Nltk.wordnet class through comparison.
  + Step8 visualize the sentiments with tweets.
  + Step9 Plot graph Sentiment graph using matplotlib and visualize the bar plot.
  + Step10 Plot all the graph visualization using Tableau Third party tools.

## Naïve Bayes Classification: -

This is a portrayal technique that is affected by a hypothesis that suggests self-rule between pointers. Basically, the clear mathematician class anticipates the vicinity of 1 component of 1 classification to be isolated from the nearness of another segment. For instance, in the event that you have red, adjusted and concerning three hairs, a characteristic item are frequently contemplated AN apple of associated grade. It doesn't make a difference if these features territory unit energetic around each other or on the nearness of different features. These attributes unreservedly improve the probability that this common item could be a significant evaluation of apple. this is regularly why it's known as "Creduras". Be that as it may, Thomas Bayes' blameless model is tricky to make and is of express worth for a dreadfully amazing reportage record. Innocent mathematician is known to be a significant outrageously unobtrusive sort of game plan ready. The mathematician theory gives how to work out the probability of P (c), P (x), and furthermore the reaction of P (x | c) to P (c | x). Take the goose underneath the resulting conditions:



**Figure 9**

To group information first, we have to play out the following advances.

1.Tokenization: This is a technique that divides various registers into smaller pieces called tokens. These tokens can be words, numbers, or accents.

Ex: today is going to rain After performing tokenization the sentence is separated into tokens as follows

"It", "is", "going", "to", "downpour", "today".

2. Stop words: In our programming language, Python uses an instrument called the Characteristic Language Device Kit (NLTK). The NLTK has a summary of stopwords in 16 different dialects.

Ex: I like moving, so I move. After evacuating stop words the sentence will be as per the following Like, dancing, dance.

2. Bag of words idea is applied to these tokens.

3. Finally, the procedure that characterizes the Bayesian naive classifier is applied. It determines the probabilities of all words in the file and provides the result, the positive and negative probabilities of each tweet.

4. Results show the likelihood of each tweet saying whether the tweet is either positive or negative.

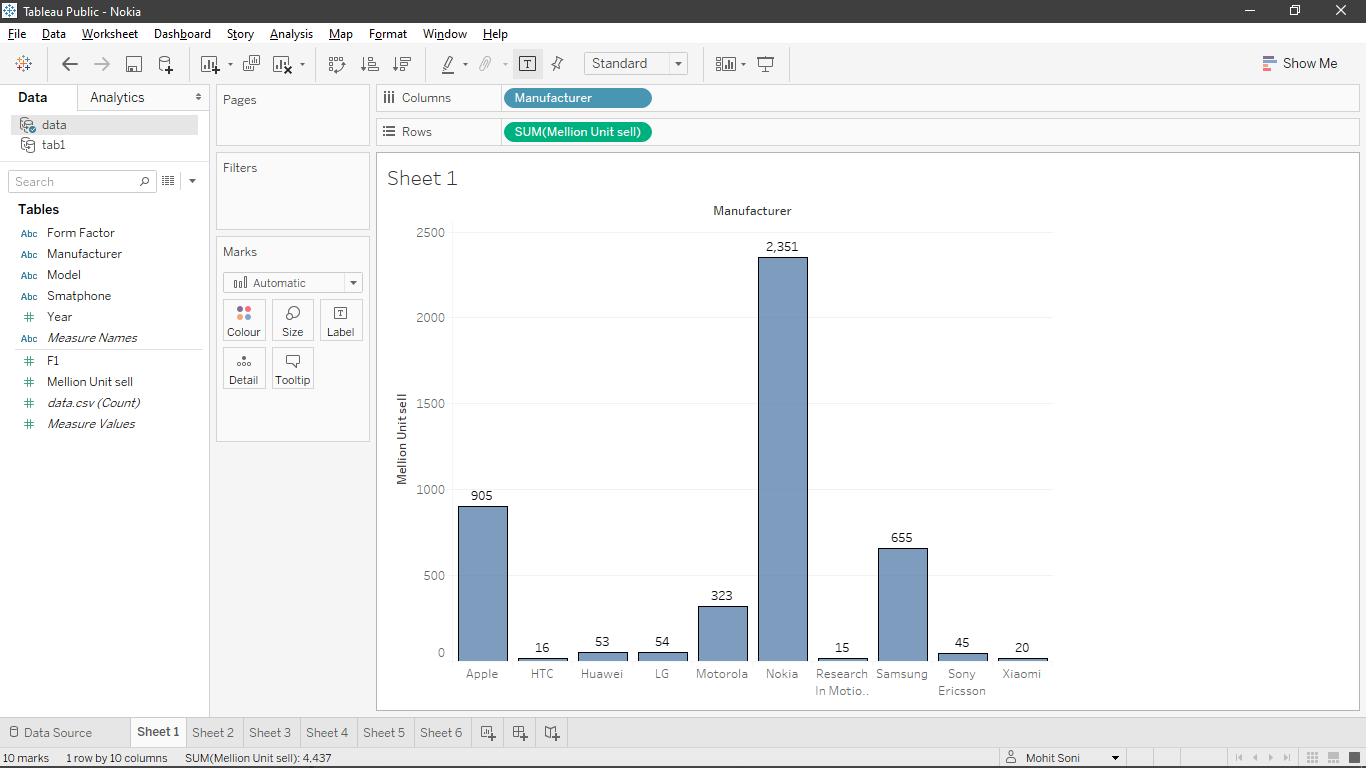
(SRS)System Requirement Specification: For this project A high configuration system with 8 GB ram, Jupyter notebook, and Python 3.8.1 software required,Good Internet and tableau Public Software.

# **Results: -**

After the analysis, several results ar**e** obtained. These ar**e** vital results for locating business ideas and directions. These results ar**e** diagrammatically given within the body section throughout discussions and ar**e** straightforward to know within the sort of bar charts.

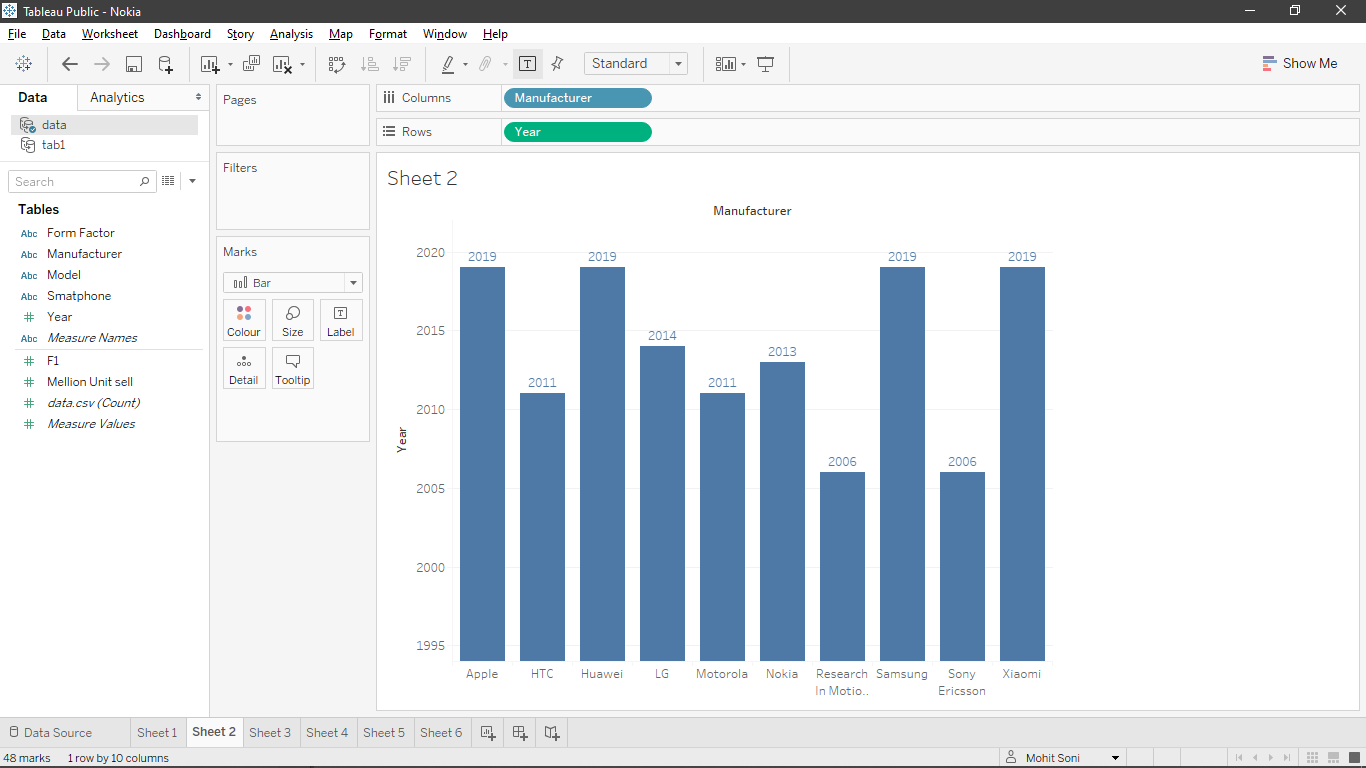
**Visualization of Sentiment Analysis: -**

1. Hers is the bar plot of different Companies sell unit in a Season.



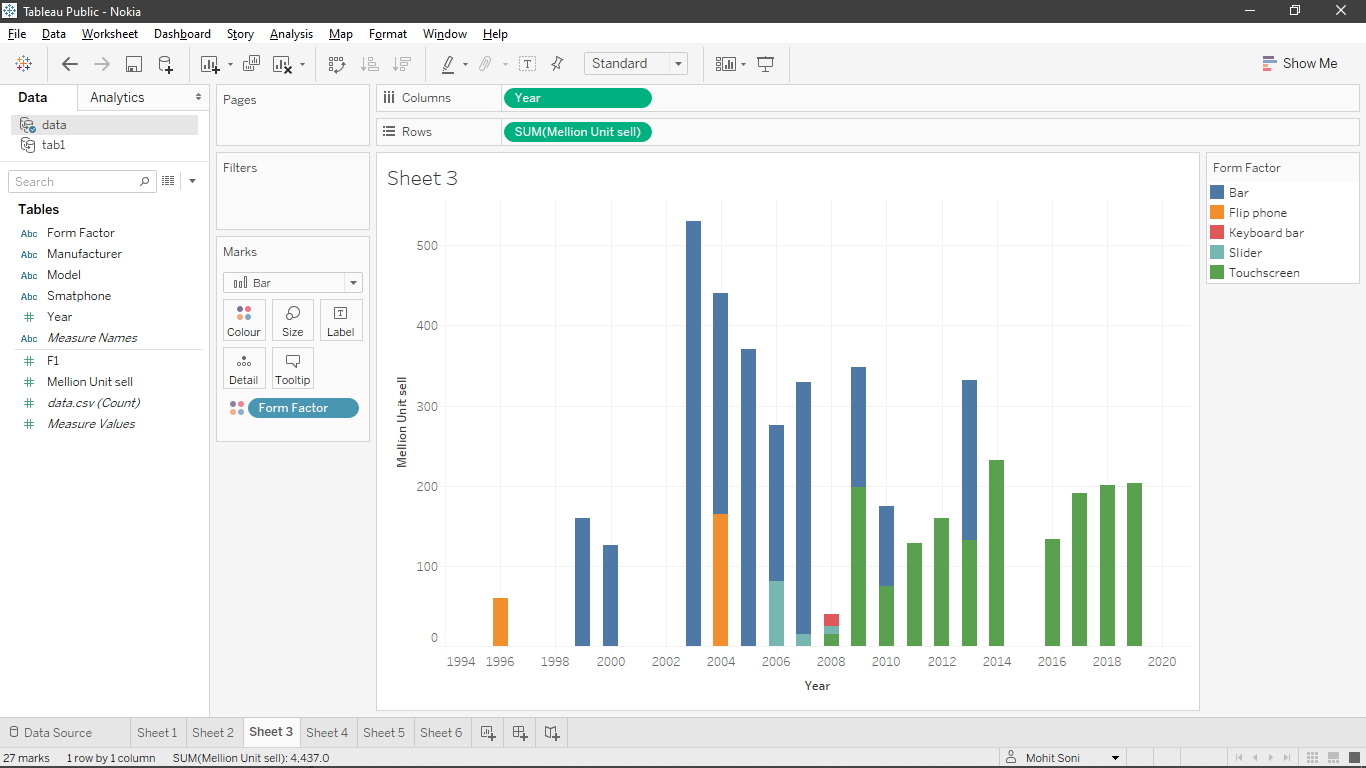
**Figure 10**

1. Here is bar plot of growth of different companies respect to the year.



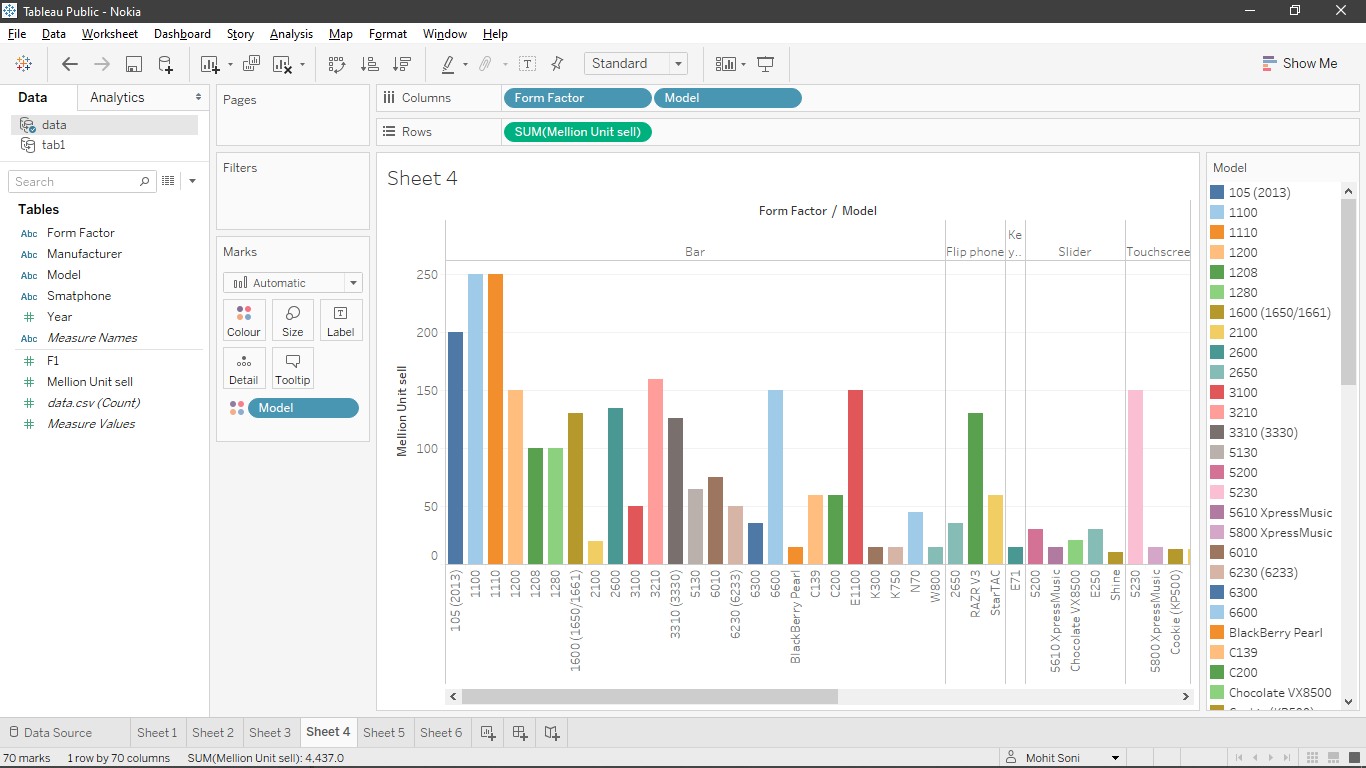
**Figure 11**

1. Here is the Graph of unit sell in million in different years with respect to different form factor.



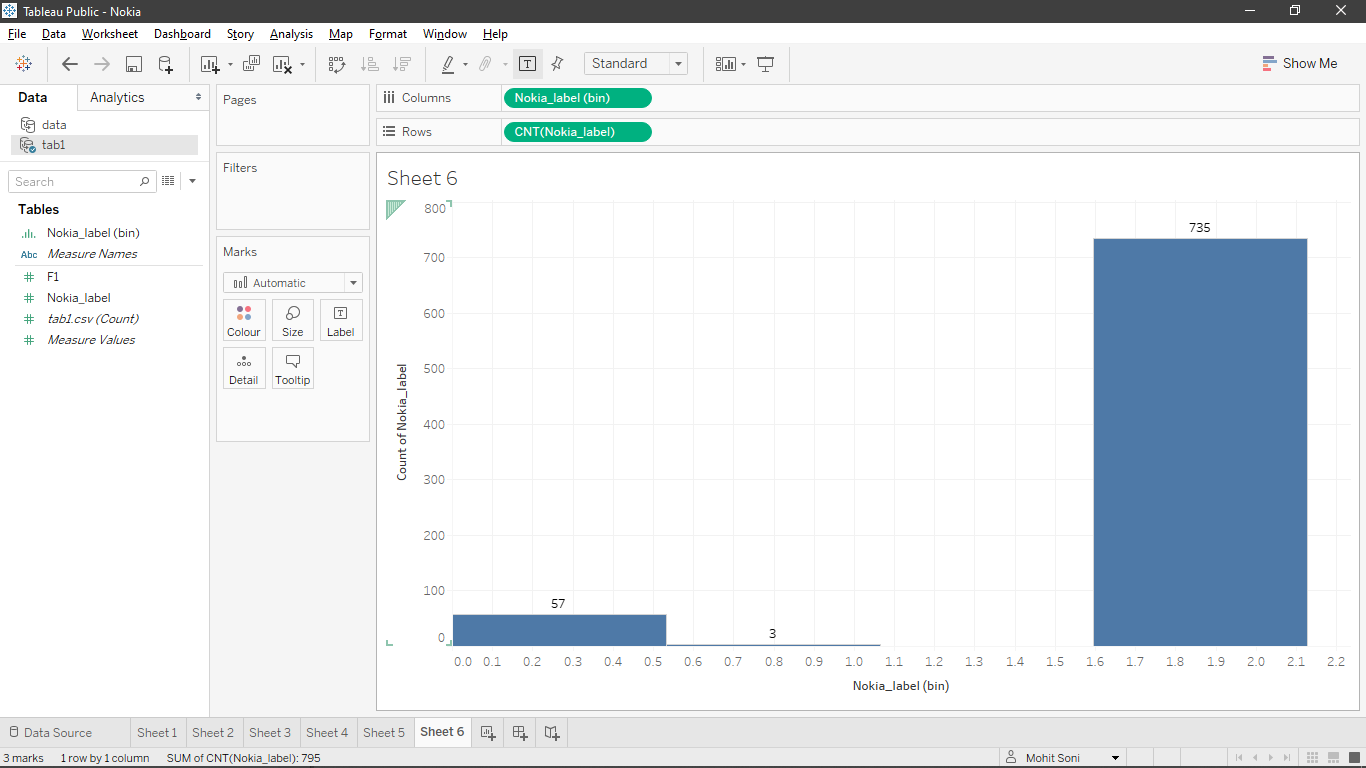
**Figure 12**

1. This bar plot shows the different million unit sell with respect to form factor and Model.



**Figure 13**

1. This bar Plot Shows Nokia Customer Sentiments on twitter.



**Figure 14**

# **Business Story Telling: -**

Companies breathe client satisfaction. The client expertise is safe, negative, or fair. This expertise that may be inferred from the good internet era are the content of social and on-line posts. you'll acknowledge the tones and behavior of this info and classify it by pooled estimates. this may assist you perceive what's happening effectively in relevancy articles, management, client service, and what wants improvement. obtaining a positive response to your article isn't enough altogether cases. Your organization's client support agreement should be satisfactory, in spite of however sensational your government is. Knowing the opinions of rivals opens the door to encourage the animation of the exhibition. Biased analysis in a corporation will facilitate predict client patterns. If you recognize your customers' current patterns, you'll undoubtedly produce techniques to require advantage of them. it'll conjointly ultimately increase its main competitive advantage. A complete isn't characterised by the product it manufactures or the controls it provides. The name and distinction that builds a complete mostly depends on its web-based show, social struggle, content advertising, and client service management. Biased business surveys facilitate live current and potential client impressions of every of those variables.

# **Conclusion: -**

Twitter Sentiment analysis goes under the class of content and sentiment mining. The focus is on investigating the premises of the tweet, processing and preparing the information from the AI model and validating its accuracy, so you can use this model for later use based on the results . This project we Successfully scrape wiki table data and using tableau visualization tool we visualize different graph using different parameter and also perform the twitter sentiment analysis scrape nokia company customer tweets and based on customer tweets we successfully visualize the sentiments.

# **Limitations: -**

* It doesn’t check whether the document contains sentiment or not.
* Another Limitation of sentiment analysis is it take disappointment as a positive comment.
* The speed and Size are also the limitation for sentiment analysis in both Training and Testing.
* Sarcasm, irony, negations, jokes tweets is not possible to predict.

# **Implication: -**

Using Tableau's chart, you'll be able to see that Nokia is that the immeasurable popular firms and their client reviews area unit additional positive than natural or negative. the best average client rating given to the organization is positive. ensuing analysis is negative. Therefore, the rating is neutral. this is often excellent news for Nokia. This analysis shows that client reviews area unit pretty much as good as they must be. For larger accuracy, use retweets and additional to try to to additional correct analysis.

# **Recommendation: -**

For further business growth, I instructed companies to work on this idea. By doing this, the company not only gains a new audience, but the old company also takes over. The company can only drive additional, accurate and productive business creation, so it only complements the target audience further. In line with the current market, each company must follow this approach for the right business.

# **References: -**

1. Narayanan, R. and Choudhary, A., 2019, August. Sentiment analysis of conditional sentences. In Proceedings of the 2019 Conference on Natural Language Processing: Volume 1-Volume 1 (pp. 180-189). Association for Computational Linguistics.
2. Visalli, M., Mahieu, A. and Schlich, P., 2020. Automated sentiment analysis of Free-Comment: An indirect liking Food Quality and Preference, 82, p.103888.
3. Kumar, A. and Garg, KG., 2020. The Multifaceted Concept of Sentiment Analysis. In Cognitive Informatics Computing (pp. 413-421). Springer, Singapore.
4. Elangovan, D. and Subedha, V., FIREFLY WITH MULTILAYER PERCEPTRON BASED ON CLASSIFICATION MODEL FOR SENTIMENT ANALYSIS.
5. Choi, M., Aiello, L.M., Varga, K.Z. D., 2020, April. Ten Social Dimensions of Relationships. In Proceedings of The Web Conference 2020 (pp. 1514-1525).
6. Kumar, T.S. P.M., Manoj, C.K. and Jeyachandran, K., 2020, March. Sentimental Analysis (Opinion Mining) in Social Network by Using Svm Algorithm. In 2020 Fourth International Conference on Computing Methodologies (pp. 859-865). IEEE.
7. Lima, A.Q. and Keegan, B., 2020. Challenges of using machine learning algorithms for cybersecurity: a study of threat-classification models applied to social media Sites for influencers data. In Cyber Influence and Cognitive Threats (pp. 33-52). Academic Press.