Internship Training

at

Karkinos Healthcare

'Feasibility of Utilising Tableau for analysing cancer data : Case Study on BRCA data'

By

Name : Rashmi Kashyap

Enroll No. PG/20/053

Under the guidance of

Dr. Rohini Ruhil

PGDM (Hospital & Health Management)

2020-22

TO WHOMSOEVER IT MAY CONCERN

This is to certify that ___Rashmi kashyap_____ student of PGDM (Hospital & Health Management) from International Institute of Health Management Research, New Delhi has undergone internship training at __karkinos healthcare_____from _07-03-2022____ to ___12-06-2022_____.

The Candidate has successfully carried out the study designated to him during internship training and his/her approach to the study has been sincere, scientific and analytical.

The Internship is in fulfillment of the course requirements.

I wish him all success in all his/her future endeavors.

27/6/2022

Dr. Sumesh Kumar
Associate Dean, Academic and Student Affairs
IIHMR, New Delhi

Mentor
IIHMR, New Delhi

Certificate of Approval

The following dissertation titled "Feasibility of Utilising Tableau for analysing cancer data : Case Study on BRCA data**"** at "karkinos Healthcare" is hereby approved as a certified study in management carried out and presented in a manner satisfactorily to warrant its acceptance as a prerequisite for the award of PGDM (Hospital & Health Management) for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the dissertation only for the purpose it is submitted.

Dissertation Examination Committee for evaluation of dissertation.

Name	Signature
_____	_____
_____	Dr Ankur Kathuria
Dr. Anuj Pandey	_____

Certificate from Dissertation Advisory Committee

This is to certify that Mr./Ms./Dr. Rashmi kashyap, a graduate student of the PGDM (Hospital & Health Management) has worked under our guidance and supervision. He/ She is submitting this dissertation titled "Feasibility of Utilising Tableau for analysing cancer data : Case Study on BRCA data" at "Karkinos Healthcare" in partial fulfillment of the requirements for the award of the PGDM (Hospital & Health Management).

This dissertation has the requisite standard and to the best of our knowledge no part of it has been reproduced from any other dissertation, monograph, report or book.

Institute Mentor Name,
Designation,
Organization

Organization Mentor Name
Designation,
Organization

INTERNATIONAL INSTITUTE OF HEALTH MANAGEMENT RESEARCH, NEW DELHI

CERTIFICATE BY SCHOLAR

This is to certify that the dissertation titled "…Feasibility of Utilising Tableau for analysing cancer data :Case Study on BRCA data …" and submitted by…Rashmi kashyap…. Enrollment No. ……PG/20/053… under the supervision of …Dr. Rohini Ruhil……. for award of PGDM (Hospital & Health Management) of the Institute carried out during the period from …07-03-202….. to …12-06-2022….. embodies my original work and has not formed the basis for the award of any degree, diploma associate ship, fellowship, titles in this or any other Institute or other similar institution of higher learning.

Signature

FEEDBACK FORM

Name of the Student: Rashmi Kashyap

Name of the Organization in Which Dissertation Has Been Completed:

Karkinos Healthcare . Mumbai

"Feasibility of Utilising Tableau for analysing cancer data : Case Study on BRCA data".

Attendance: 99%

Objectives achieved: Completed 3 months internship in the K cloud team . Rashmi was part of a team working on the clinical data and building analytics platform in Karkinos and has completed all essential training like SNOMED , NLP tools , Google analytics for the project works assigned.

Deliverables: 1 Data team project workflow for the pipeline

2 Data analysis summary document

3 Clinical validation test report

4 Mapping between SnoMED and clinical data

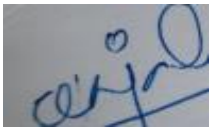
5 Requirement document for the new Clinical pathways

Strengths: Good understanding of the healthcare domain and ready to learn new skills and tools. Quick learner and timely submission of project deliverables.

Suggestions for Improvement: Can have more research mindset and read related new developments in the field . Should develop skills to write and publish articles related to work .

Suggestions for Institute (course curriculum, industry interaction, placement, alumni):

Can have more close industry and academia collaboration .



Signature of the Officer-in-Charge/ Organisation Mentor (Dissertation)

Date: 09-06-2022

Place: Bangalore.

ACKNOWLEDGEMENT

I would like to express my sincere thanks and gratitude to Karkinos Healthcare Private Limited, Bangalore for giving me a wonderful opportunity to work along with and at the same time complete my dissertation project titled 'Feasibility of Utilising Tableau for analysing cancer data : Case Study on BRCA data'.

I sincerely thank Arup Gosh and Dr. Anjali Kulkarni who were kind enough to spare their valuable time and provided the suitable environment and optimum guidance in the interest of my project completion.

Also, I am sincerely grateful to Dr. Rohini Ruhil IIHMR Delhi, for her continuous encouragement in the completion of this project.

Thanks to one and all.

Rashmi Kashyap

PGDM (Hospital & Health Management)

IIHMR, New Delhi

Table Of Figures

Table Of Contents

Chapter 1-INTRODUCTION

The term ‘Cancer’ finds its origin from a Greek word ‘Karkinos’ (meaning crab) which refers to a group of non-communicable diseases where body cells grow uncontrollably, form a lump and spread to other parts of the body. Cellular growth in humans typically consists of multiplying (through a process called cell division) and forming new cells on a regular basis. They die as they age or become damaged, and new ones replace them. The orderly process of cancer is disrupted in diseased states, and cancerous cells grow and multiply when they shouldn't. This results in formation of lumps of cancerous cells/ tissue also called a tumour in different parts of the human body. Tumours can be benign (confined to its original location) in early stages but if not detected and treated in time, the cells begin to migrate to other body parts (by a process known as metastasis) and result in malignant form of cancer. Although several forms of cancer have been detected in various parts of the human body, the most common sites are breasts, colorectum, prostate, lung and thyroid. One in five people worldwide is estimated to contract cancer in their lifetime, and one in eight men and one in eleven women die of it. IARC (International Agency for Research on Cancer) released data for the year 2020 and found that there were 19.3 million cancer cases and 10 million cancer deaths worldwide. According to these estimates, more than 50 million people are currently dealing with the consequences of a cancer diagnosis within the last five years. (figure 1).

Figure 1. Estimated number of prevalent cases of cancer worldwide (5-year) in 2020 .

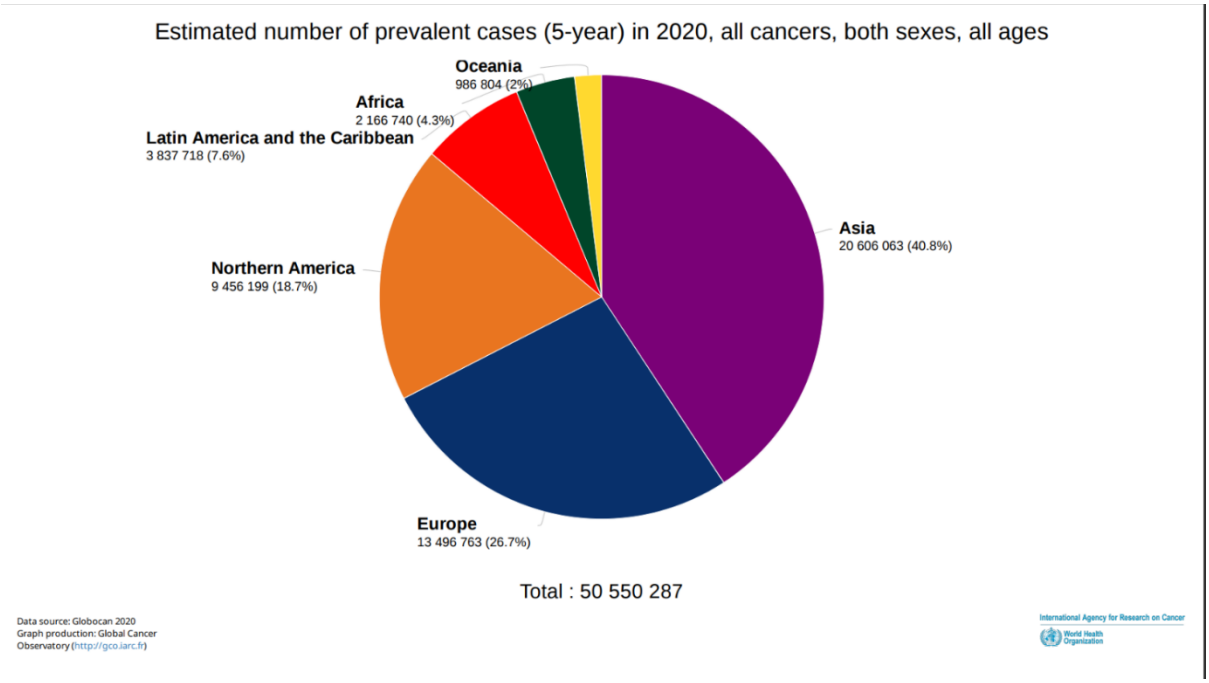


Figure 2. Estimated number of prevalent cases of cancer worldwide (5-year) in 2020 for both sexes and all age groups.

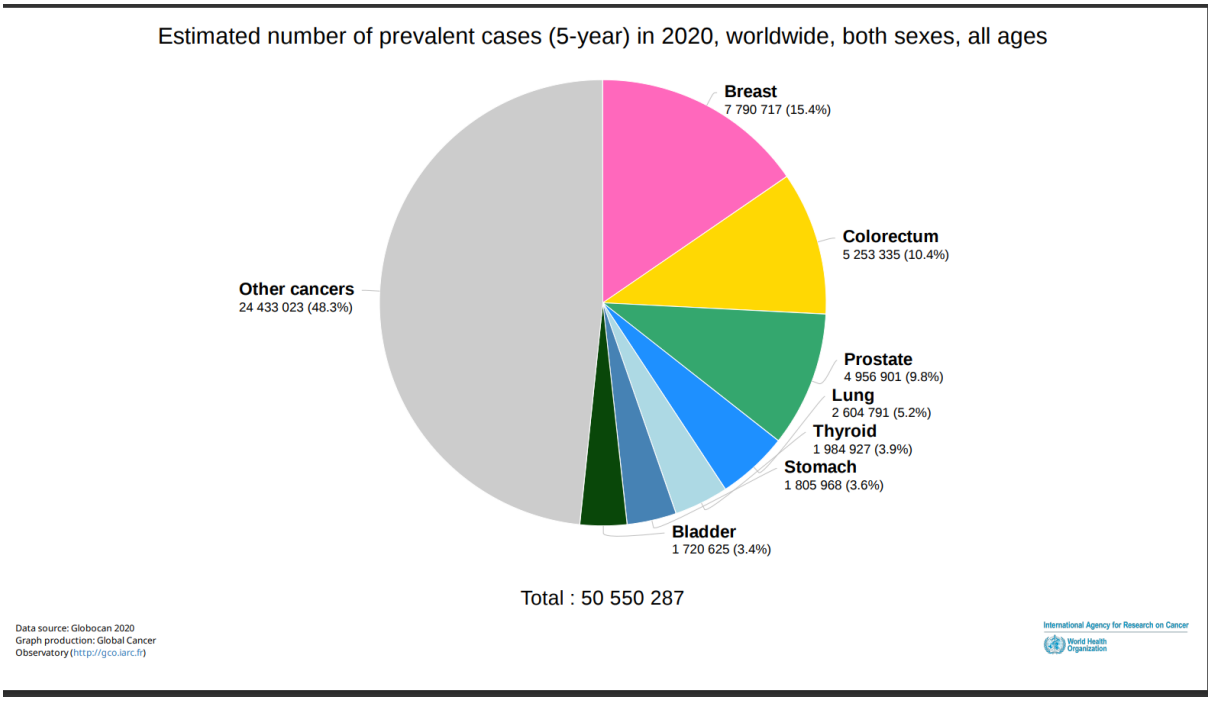
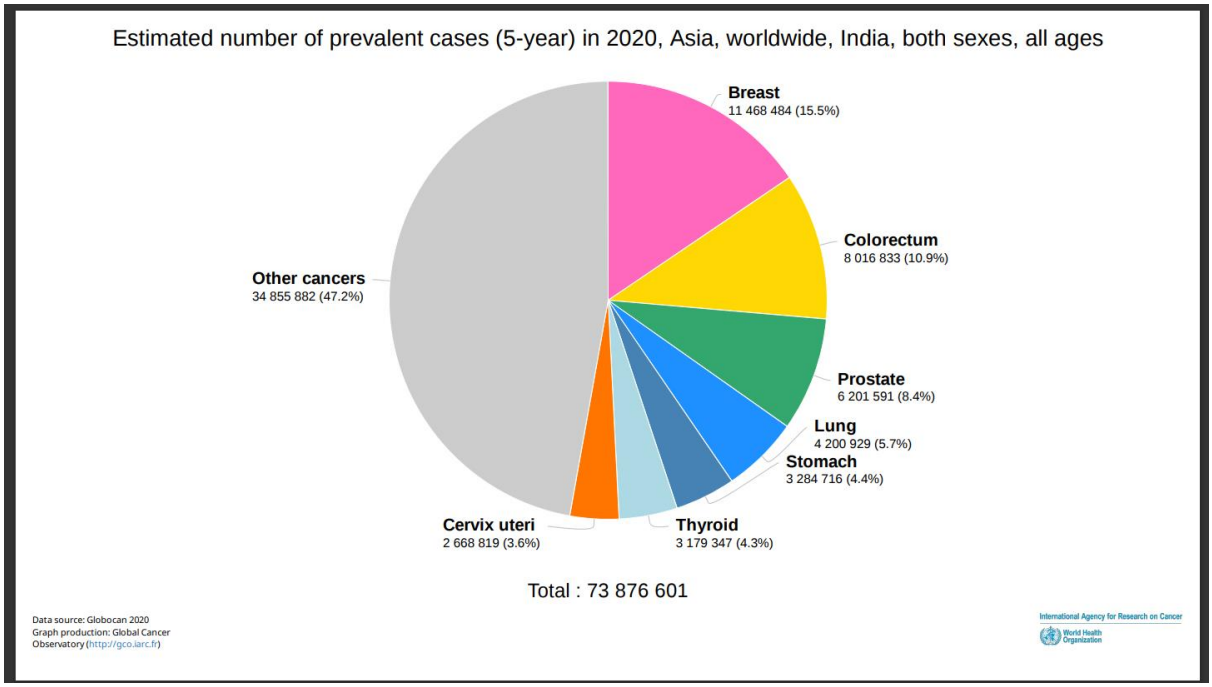


Figure 3. Estimated number of prevalent cases of cancer in India (5-year) in 2020 for both sexes and all age groups.

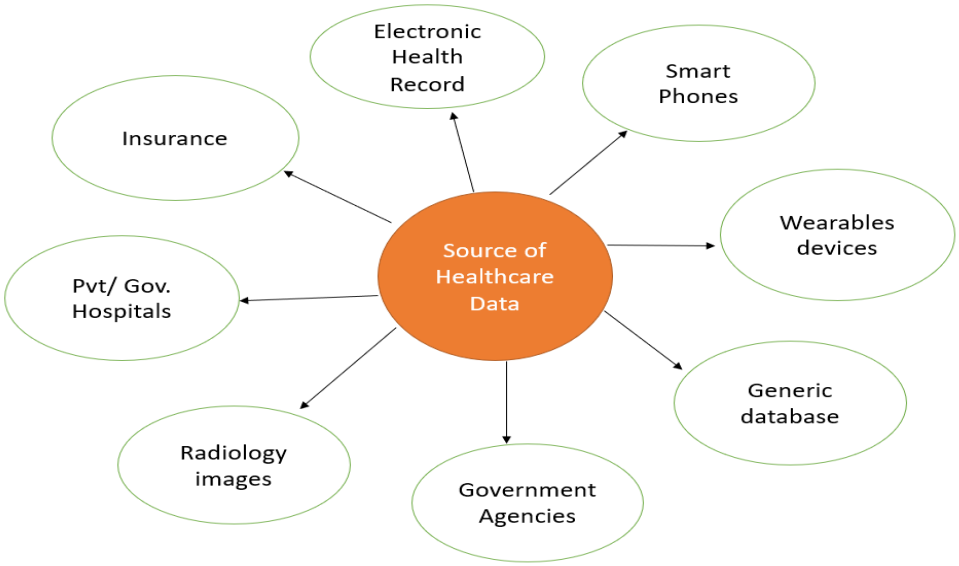


Cancer is leading cause of death, because people are not able to detect cancer at early stage of their life.

Healthcare data

Diverse forms of healthcare data sources include clinical text, biomedical images, EHRs, genomic data, biomedical signals, sensing data, and social media. Clinical text mining transform data from clinical notes that are organized in an unstructured format to useful information.

Figure 4: Different source of healthcare data.



Healthcare generated more than 25,000 petabytes of data (Source: Healthstatics). Huge amount of data is generated in healthcare organisation because of this they should shift their operation to data- driven mindset. Data management is a priority necessity of all stakeholders in the health care industry. This illustrates how incredible rate of expansion is in the digital domain. Here we can take the example of google, it retain a range of information based on our preferences, such as browing history, shopping list, advertisement preferences etc. It store and analyses over 30 petabytes(PB) of users data. When we deal with such a huge amount of data it is called as Big Data. With the growing use of electronic health records (EHRs) and patient monitoring devices, vast volumes of data and physiological data are constantly flowing and piling up, demanding data mining and analysis. According to studies, the volume of data and information in health care is expected to grow over time as technology is introduced to help healthcare professionals perform better by utilising meaningful and relevant data.

Data Analytics

It is an advance analytical technique that analyses data sets that include structured and unstructured. It helps in making better and faster decision. Advance analytical tools are used such as, text analytics, machine learning, data mining statistics, and Natural language processing for big data analytics.

There are many source of healthcare data, sources of data like; transactional data, machine data, biometric data, human generated data, social media data.

But why is data analytics in healthcare becoming so popular? Because of digitalization and advanced technologies, huge amount of heterogenous data from different sources like hospitals, insurers, pharmaceuticals, researchers and government agencies have become accessible. The authors however, state that this data is siloed. Insight generated from integrating such different types of data would facilitate the design of programmes that would result in improved patient outcomes, and possibly reduce incidences of chronic diseases (Task Force 7 Health subgroup [TF7.SG3], 2016

Despite its growing adoption, big data analytics is prone to several challenges. Luna, Mayan, García, Almerares, and Househ, (2014) warn that big data analytics is prone to at least three core challenges. The authors point out that the first challenge has to do with the structure and accessibility of raw data – most raw data captured by organisations is usually scattered in several silos which are sometimes hard to consolidate and integrate. Further, the authors show that most organisations lack clear business cases to guide in the process of harnessing raw data using big data analytics. Belle et al., (2015) wade in to support the authors third challenge by showing that organisations experience a lack of robust coordination among big data analytics teams when attempting to manipulate and interpret raw data in their custody.

In their study to explore the opportunities and challenges that may emerge when applying big data analytics in the health sector, Kruse et al., (2016) found that organisations are likely to experience difficulties in the areas of "… data structure, security, data standardization, storage and transfers, and managerial skills such as data governance" (p.38). At the same time however, the authors found that opportunities manifest in the form of "... quality improvement, population management and health, early detection of disease, data quality, structure, and accessibility, improved decision making, and cost reduction" (p. 38). These findings echo those of TF7.SG3 (2016) to the effect that health sector organisations must be willing to overcome challenges to optimise the benefits and opportunities that come with big data analytics. It is therefore not surprising that the adoption of big data approaches in the healthcare sector is still low (TF7.SG3, 2016).

Data Analytics set up For healthcare data

Traditional health analytics is similar to conceptual frame work. In healthcare system the difference is that how we are using the data. Because the volume of data in healthcare is unpredictable, its physical infrastructure is based on a distributed computing architecture, in which data is typically kept in multiple locations and linked via networks different analytic tools for huge data are used.

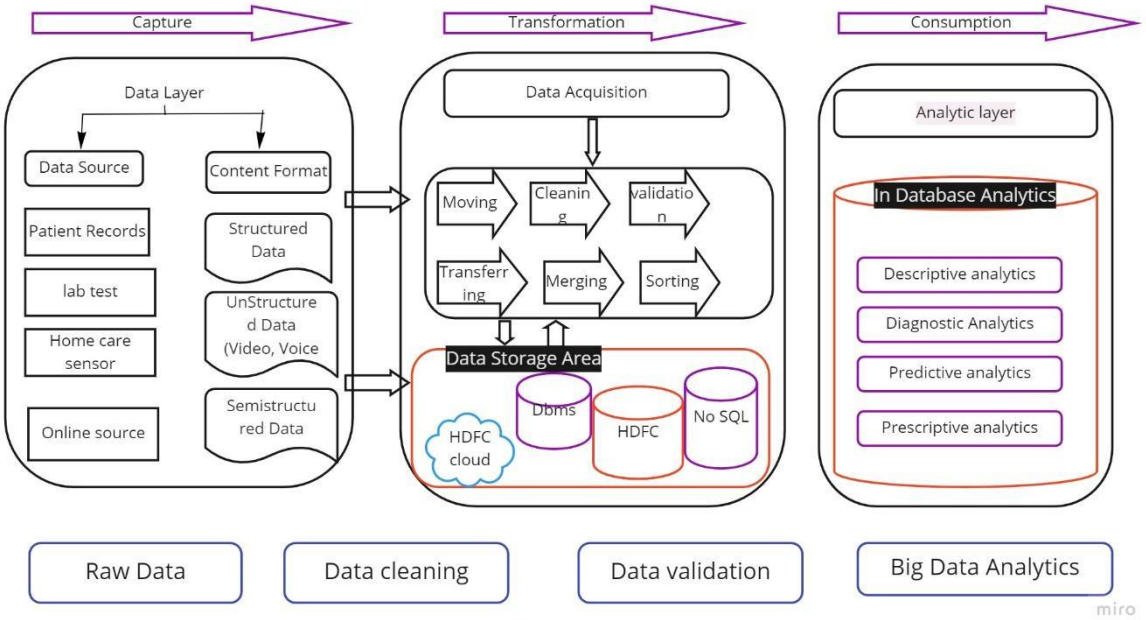
Data goes through different phases in data analysis process, after receiving the data from different sources.

Phases of data

1. **Raw data-** First Data is received.
2. **Prepossessing-** After Data received cleaning/ merging is done.

3. **Make it in structure format** - If it is unstructured data then make it in structure format. This is where we use different Analytical tools.

Figure 5: Phases of data.



As we can see in this figure the complexity begins with the data itself. Healthcare data can be generated from internal (e.g., electronic health records, clinical decision support systems, etc.) and external sources (government sources, laboratories, pharmacies, insurance companies & HMOs, etc.), and it can be in multiple formats (flat files, .csv, relational tables, etc.). and residing at multiple locations (geographic as well as in different healthcare providers' sites).

Organisation like karkinos providing end to end technology driven oncology focused healthcare platform. Dealing with huge amount of oncology data. They deal with unstructured data to get more insight from it. The aim is to detect cancer at early stage

by training patient data. Here dealing with such a huge amount of data is called as Big Data.

Objective Of Study

1. To understand the applicability of tableau to analyze cancer patient data.
2. To understand the status of cancer patient mortality rate based on Age and GENDER.

Chapter 2. Review of Literature

1. Inseok Ko and Hyejung Chang, 2017, in their study mentioned that Big data analysis is receiving increasing attention in many industries, including healthcare. Visualization plays an important role not only in intuitively showing the results of data analysis but also in the whole process of collecting, cleaning, analyzing, and sharing data. In the paper they presents a procedure for the interactive visualization and analysis of healthcare data using Tableau as a business intelligence tool.[1]
2. Shenhui Jiang , Shiaofen Fang, et.al., 2016. In their study talks about health data visualization system which emphasizes integration of geospatial and temporal information in healthcare data. Their focus is on two new visualization methods that they developed for public health data: Spatial Textures, and Spiral Theme Plot. Spatial Texture approach is effective because geospatial visualization intrinsically provides additional screen space (surface areas) that can be taken advantages of to encode additional data and attributes.[2]
3. K. A. Narayan, M. Siva Durga and Prasad Nayak 2021, talks about Each visualization technique has its own advantage. Pivot table and Chart with The Drill Down will make the end-user can visualize the data based on the filters. For effective and prompt use of data visualization it is also important that the information generated is disseminated to the stakeholders. Although different software and platforms came into existence for creating types of visualizations of data. It is not about pretty pictures or statistical graphics. One perfect visualization configuration may not answer a question. The process requires trying out different visualization techniques.[3]
4. Gotz D, Wang F, And Perer A. 2014, developed methods that provides an interactive visual environment for the exploration and analysis of temporal medical event data.

Their method is specifically focused on analyzing collections of temporal event sequences, mining for patterns of events that are strongly associated with outcome, and visualizing the detected patterns.[4]

5. Islam Ms, Hasan M M, et.al,.2018 identified three types of analytics descriptive, predictive and prescriptive. In their study they concluded that descriptive analytics is commonly used in healthcare 48%. Predictive analytics is used 43% among clinical decision support system and in risk and morbidity prediction for heart attack, and other disease.[5]
6. Ying Yang And Tao Chen, 2019, in their paper first the workflow of DBN, a deep learning algorithm and summarizes computational algorithm.

   The classification function is translated into an assembler using an instruction set-based assembly language, and the program is evaluated for performance. visualization and display of medical data is realized, which has certain promotion effect on the research and development of medical big data visualization analysis.[6]
7. Jolly Anjali and Tripathi 2019, Mentioned that With the increased demand for big data, healthcare has come to be a statistics wealthy area. In healthcare, information is specifically acquired from diagnostic and treatment tactics. Also, predictive modelling allows in figuring out the future techniques and control plans precisely.[7]
8. Ratra Ritu and Gulia Preeti. (2019) Mentioned that a large volume of data is generated in different forms i.e. text, audio, video, image etc. It is very difficult to handle this large amount of data i.e. big data and to search out associations, patterns and to investigate the massive information sets from this. This paper gives the idea about the analysis of big data. A number of tools and techniques are discussed to analyze big data.[8]

9. Kumar sunil and singh Maninder. (2019) Provides an in-depth description and a brief overview of big data in general and in healthcare system, which plays a significant role in healthcare informatics and greatly influences the healthcare system and the big data four Vs in healthcare. They also proposed the use of a conceptual architecture for solving healthcare problems in big data using Hadoop-based terminologies, which involves the utilization of the big data, generated by different levels of medical data and the development of methods for analyzing this data and to obtain answers to medical questions.[9]
10. Liu Ning e.al.,(2018) Mentioned that male breast cancer patients have had worse survival outcomes compared with those of female patients. The 5-year survival rate for male patients was lower than that for female patients (82.8% vs. 88.5%). After controlling for other factors, the risk of death in men was 43% greater than that in women during the follow-up period.[10]
11. Maria Teresa and jane A 2017. In their study provides quantitative and qualitative evidence that many SGM breast cancer patients face a dearth of appropriate social supports, both from breast cancer survivor organizations and from within the medical system.[11]
12. Amanze, Chibuike Bethran et.al., 2022, In this paper, they demonstrate data analytics and data visualization that could be used for health data. The visual reports generated by the python program include frequency of the cases in each field in the dataset. Descriptive and linear regression analyses were used in analysing the data collected with the aid of Statistical Package for Social Sciences (SPSS) and Excel Worksheet. The result generated shows the descriptive statistics of the dataset and also predicted the effect of the independent variables on the dependent variables.[12]

Chapter 3. Methodology

Study Design : Descriptive study

Study Data : Secondary Data Analysis

Data Source : Dataset was taken from Cancer imaging Archive. TGCA BRCA which is a platform that holds a huge archive of medical data and images . Dataset comprises of 334 breast cancer patients

Data Analysis tools : Tableau was used to generate visualizations on cancer data

Google Spreadsheet was used to do Data profiling i.e understanding the structure of data

Duration of study : 7 march 2022 to 7 june 2022

Ethical clearance : Since it is secondary data no ethical clearance needed.

Operational definition : Tableau links users to many data sources and gives them the ability to make charts, maps, dashboards, and stories using a straightforward drag-and-drop interface. Tableau is still not commonly utilised in the healthcare sector, despite the attention it has recently gotten. To promote its wider usage, this article introduces Tableau and describes how to utilise it for the interactive display and analysis of healthcare data.

Process :

1.Installing Tableau

There are three variations of Tableau Desktop available. The Personal and Professional editions can be purchased for a cost or with a free 14-day trial, however the Public version is free.

2. Connecting for breast cancer dataset

This dataset is Excel formatted data, so Excel is selected to import the data. The attached datasets appear in the upper left corner, and the sheets are listed at the bottom of the datasets.

3.Generate Insights

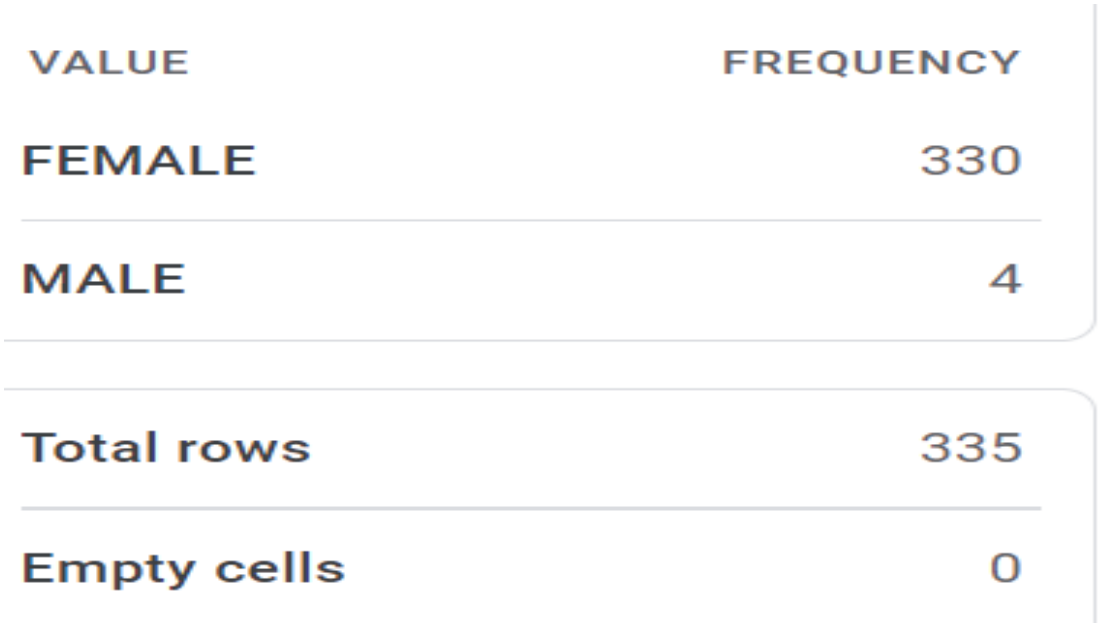
Select the variables for visualisations using drag and drop method.

Chapter 4. Result

1) Data Profiling Output

- There are total number of 334 patient data included in the dataset, out of which 330 are females and four male patients are there. In this analysis value talks about the two categories Female and male. And frequency includes number.(refer figure 6)

Figure 6: Number of male and female.

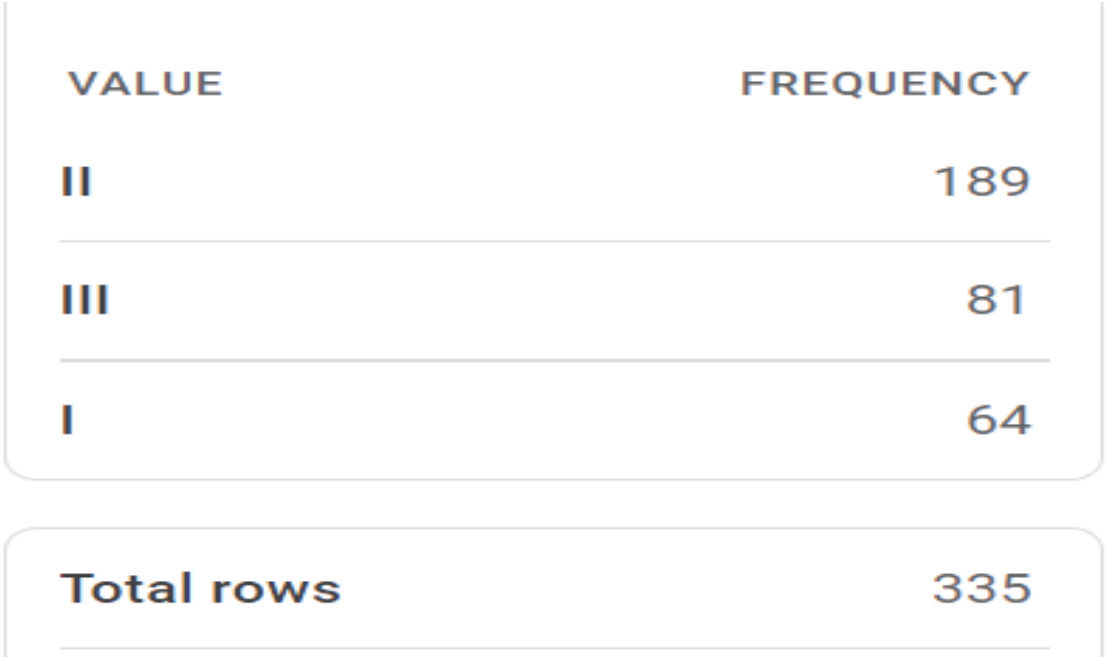


VALUE	FREQUENCY
FEMALE	330
MALE	4

Total rows	335
Empty cells	0

- The figure below talks about the Three Tumour stage categories 1, 2 and 3. Frequency depicts the total number of people at each stage. (refer figure 7)

Figure 7. Number of Tumour staging.

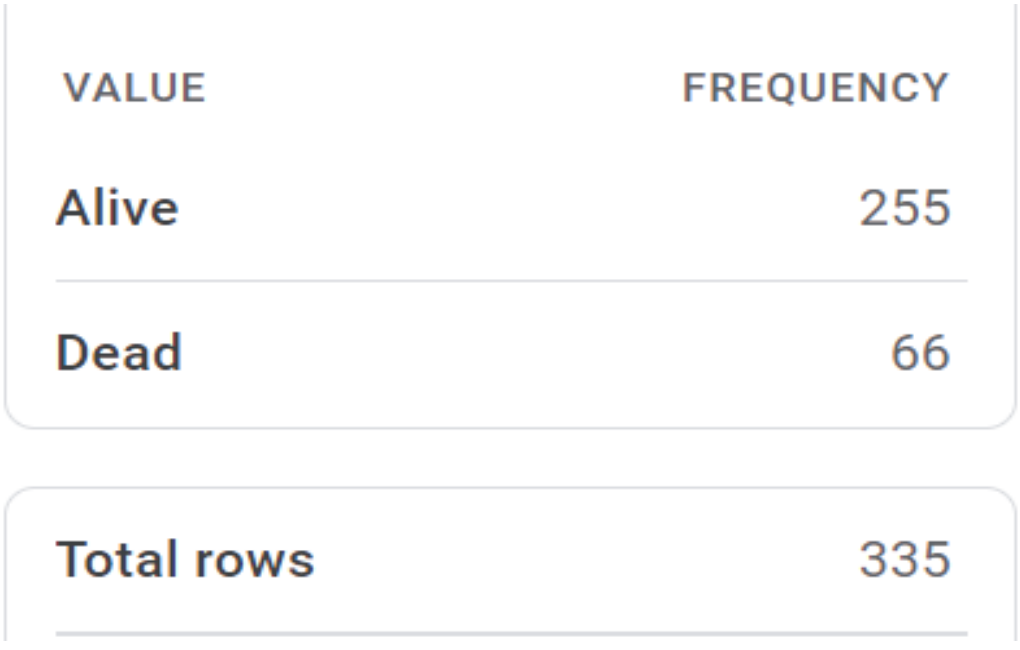


VALUE	FREQUENCY
II	189
III	81
I	64

Total rows	335

- The distribution of patient status was also studied (refer figure 8)

Figure 8. Count of patient status

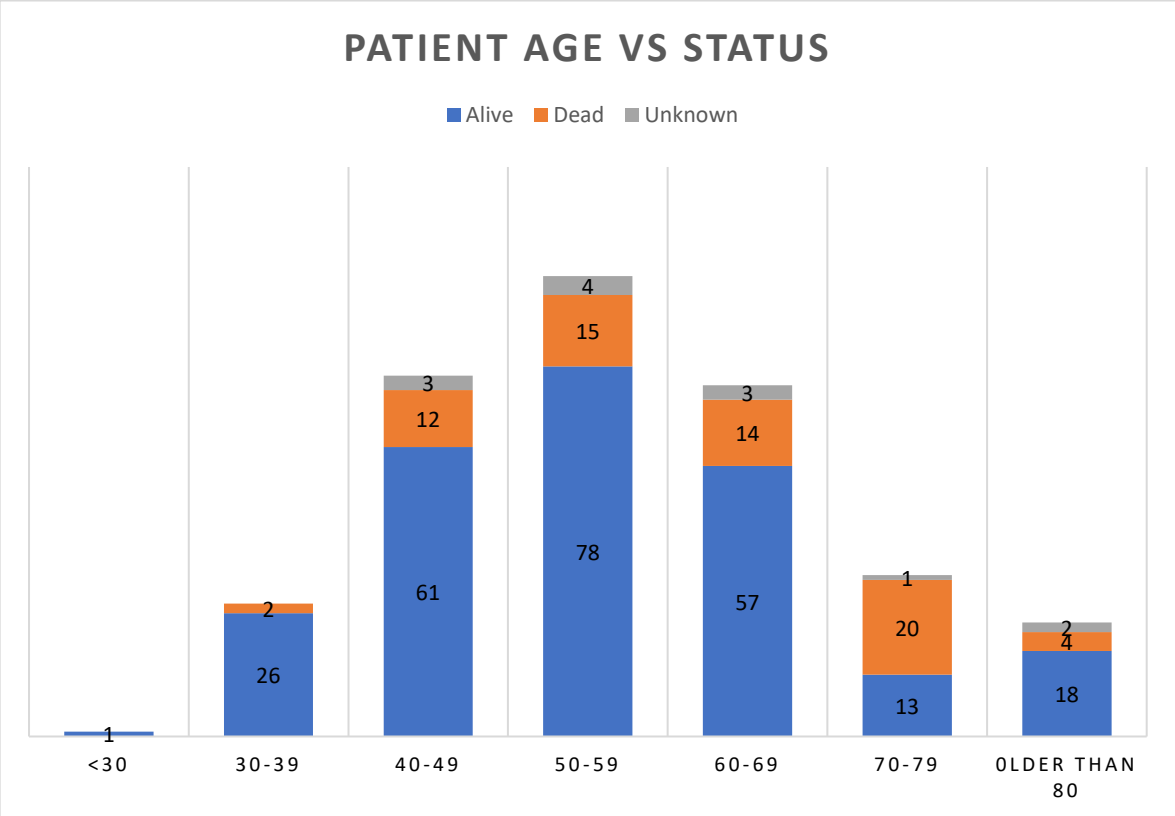


VALUE	FREQUENCY
Alive	255
Dead	66

Total rows	335

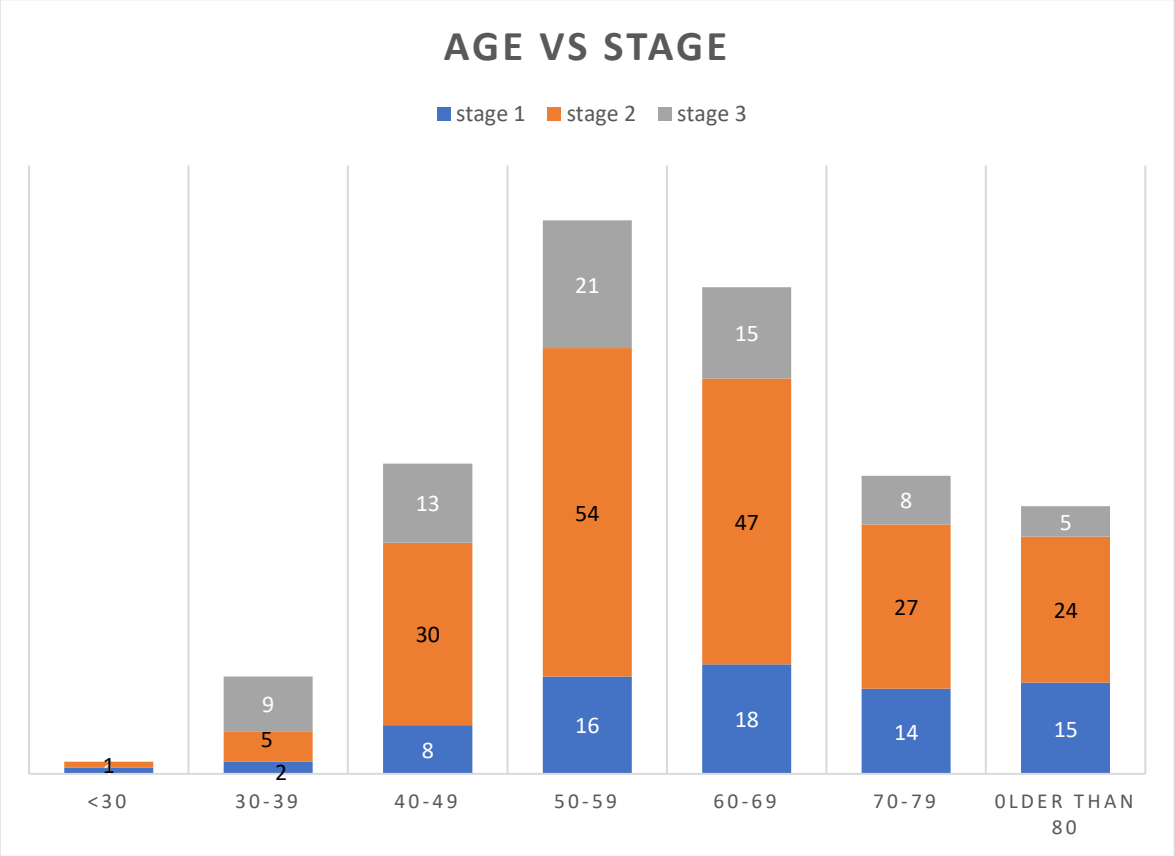
It displays the patient's status and age. This shows that there are 8% between the ages of 30 and 39, 0.5%of them are dead and 7% of them are alive. 22% patients, between the age of 40 and 49, Out of these, 18% are alive, 3.5% have passed away, and 0.89% have undetermined conditions. 29% patient between the age of 50 and 59, Out of these, 23% are alive, 4.49% have passed away, 1.19% have undetermined conditions. 22% patients between the age of 60 and 69 out of these 17% are alive, 4% are dead and 0.89% patients have undetermined condition. 10% patient between the age group of 70-79, out of these 3% is alive, 6% are dead and 0.2% patient's condition is undetermined. 7% Patient who is older than 80 years are out of these 5% patients are alive, 1% patient are dead. (refer figure 9)

Figure 9. Count of patient Age vs status.



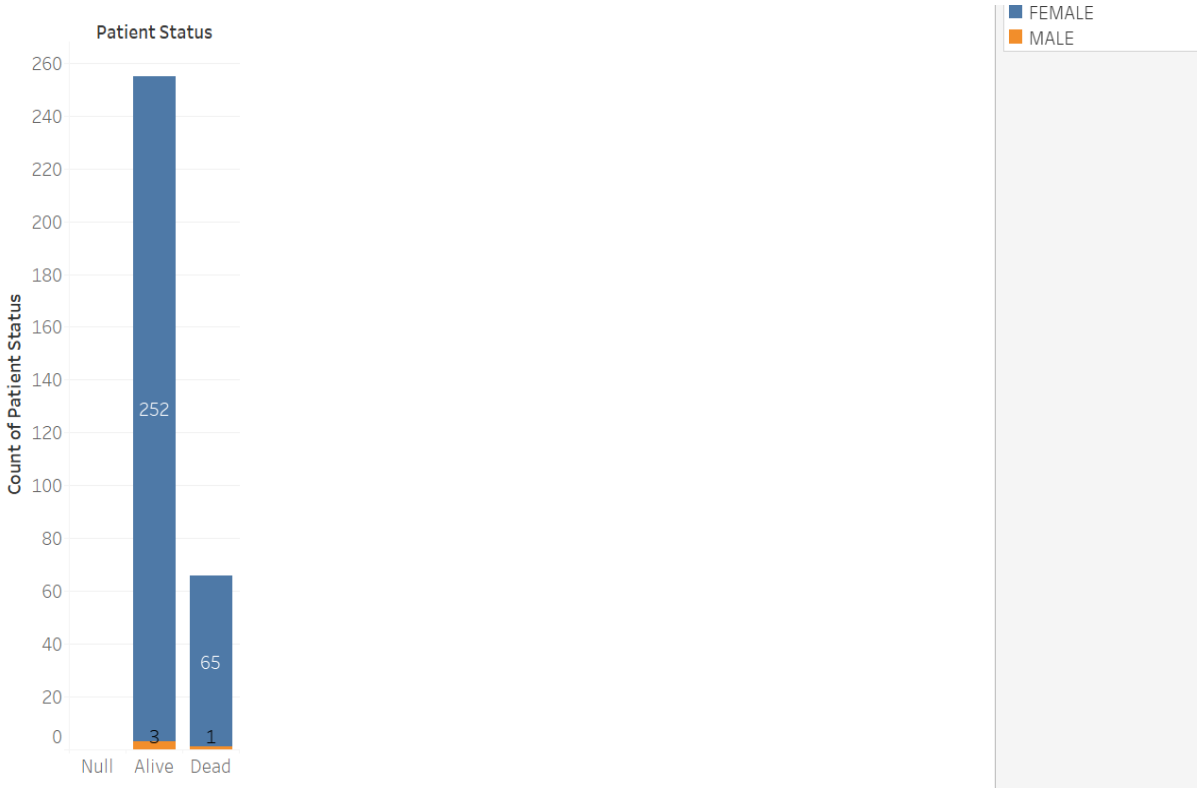
It shows the patient's age and the stage of their tumour. Here, we can observe that the age group with the highest percentage of persons Tumour stage is between 50 and 59, at 27%, and between 60 and 69, it is 24%.(refer figure 10)

Figure 10. Count of patient Age vs stage.



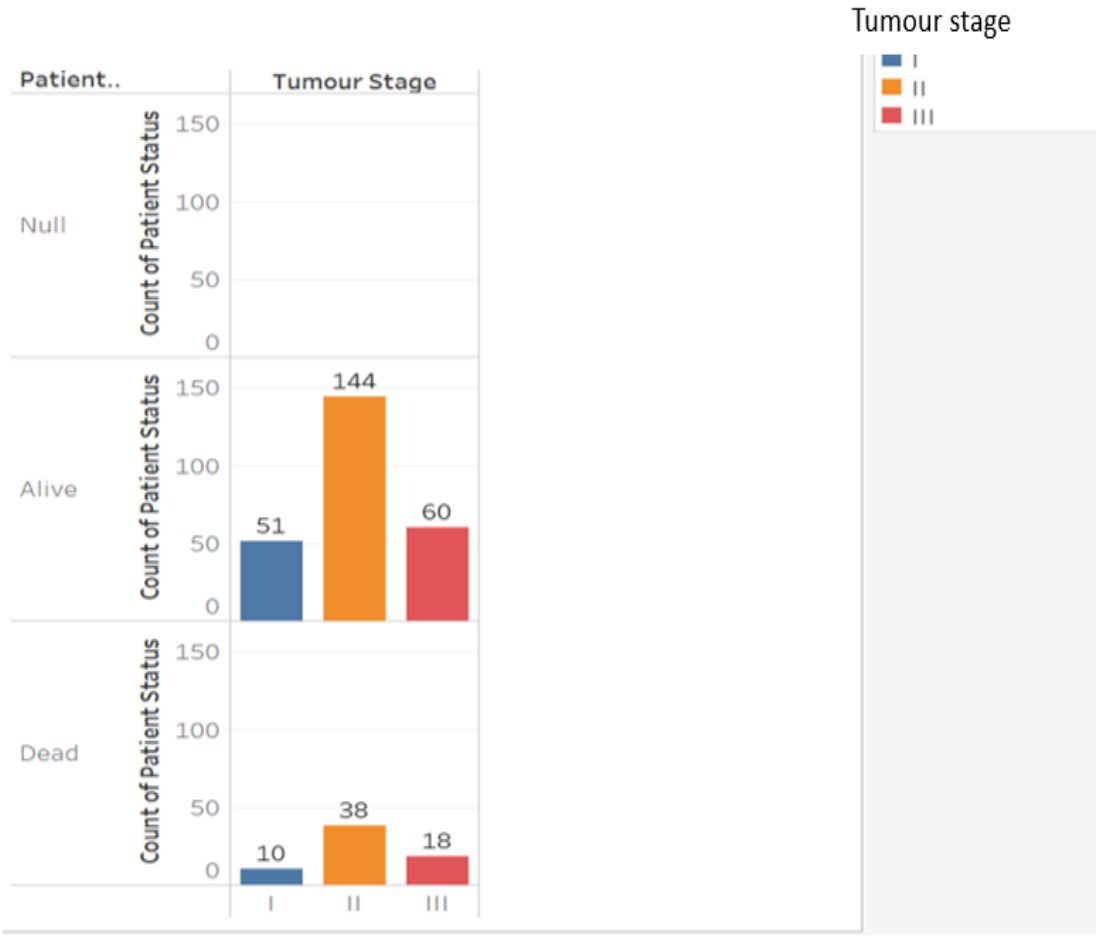
This analysis shows that 95% of women included in this study, 79%women are alive, 20% of women are dead, and 1% of females condition is not known.1% male included in this study. out of which 75% men are alive and 25% men are dead.(refer figure 11)

Figure 11. Count Gender vs patient status.



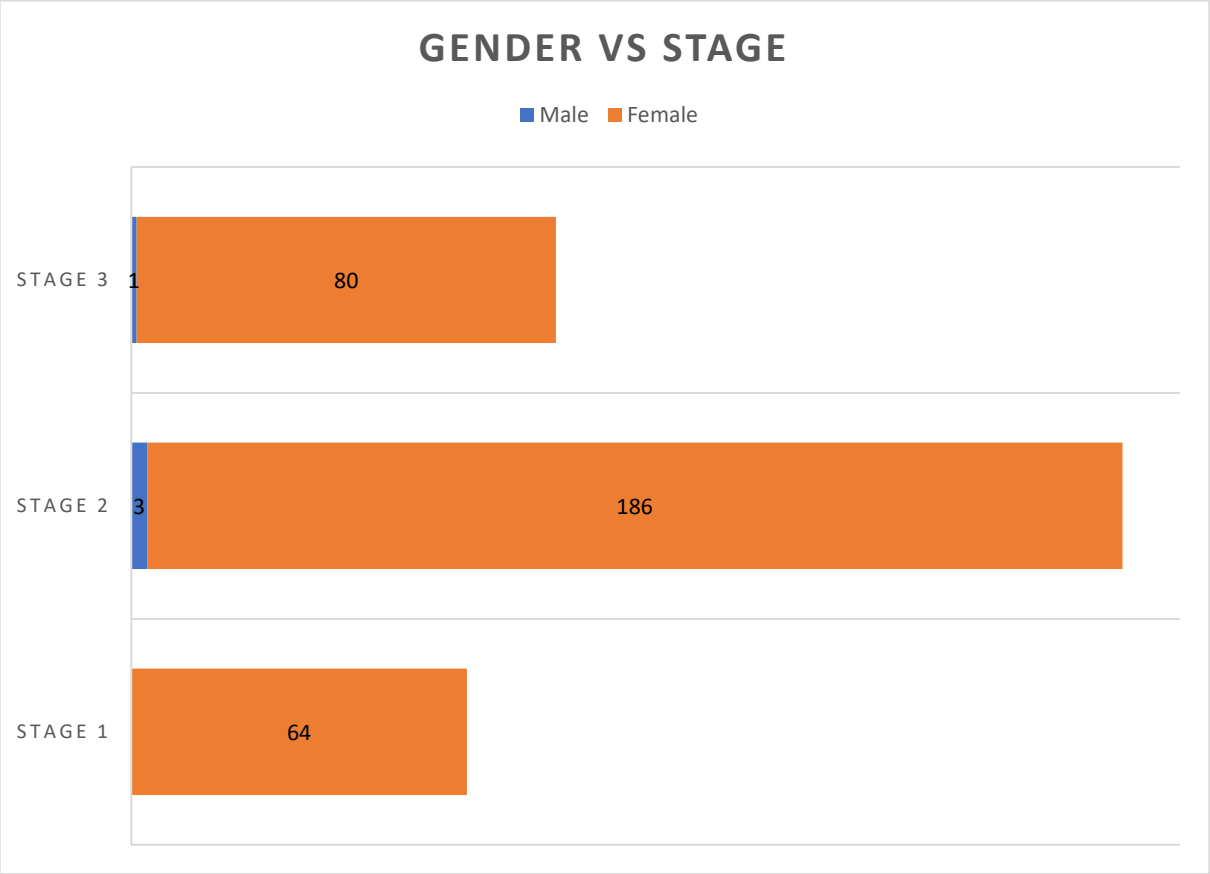
Here it depicts the Patient status vs tumour size, it shows that 15%, 43%, and 17% patient who are at there Tumour stage 1, 2, 3 are alive. And 2%, 11%, 5% Patient who are at there respective tumour stage of 1, 2, 3 are dead. And 7% patient's data is not known.(refer figure 12)

Figure 12. Count of tumour stage vs patient status



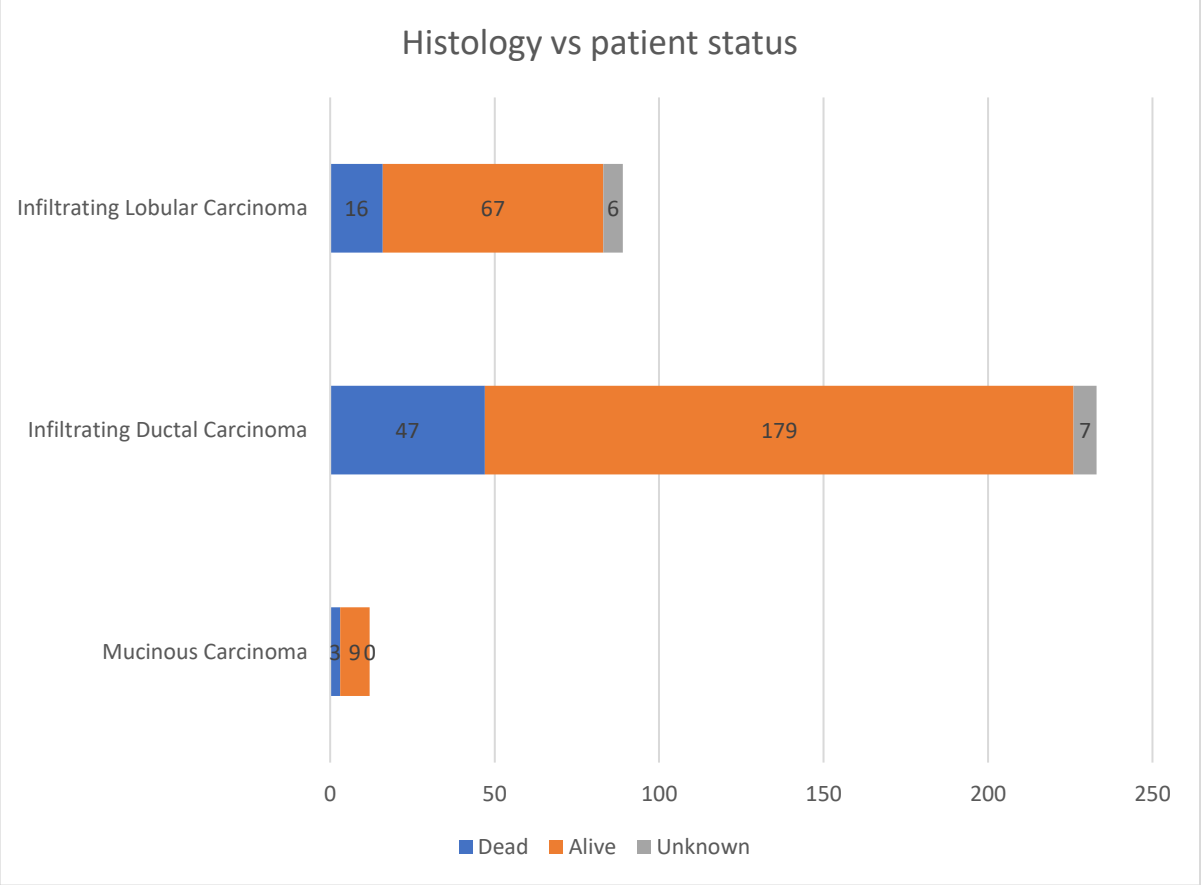
It depicts that 24%of women are at tumour stage1, 56% of female are at stage 2 tumour, and 19% of females are stage 3 of tumour. 25% of male patients are at stage3 of tumour, 75% of are at stage 2 of tumour.(refer figure 13)

Figure 13. Count of Gender vs stage.



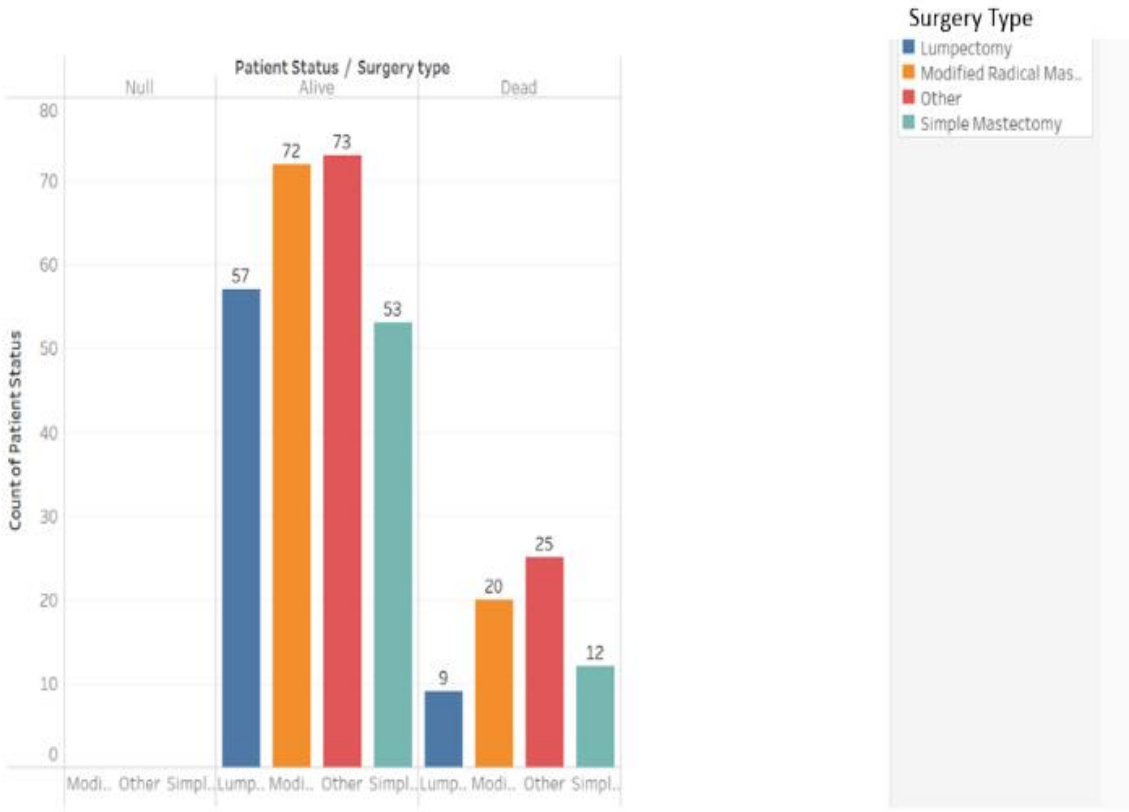
Below analysis depicts histology and status of patients. 26% of patient have Infiltrating lobular carcinoma, out of which 5% are dead, 20% are alive and 2% patient's condition is not known. 70% patients having infiltrating ductal carcinoma, 14% are dead, 53% patients are alive and 2% patient condition is not known. 4% have mucinous carcinoma.(refer figure 14)

Figure 14. Count of Histology vs status.



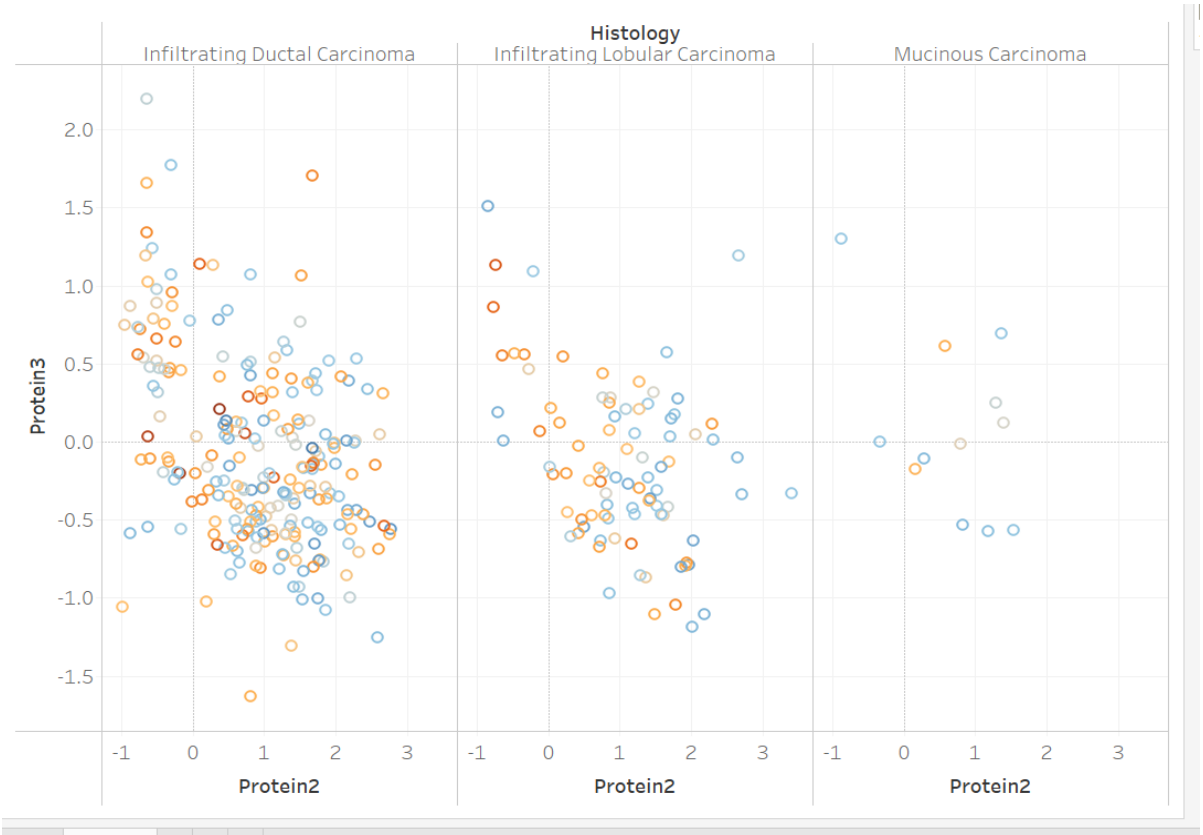
Below analysis shows that there are 96% patient who undergone surgery, 76% of patients are alive and 19% patient are dead.(refer figure 15)

Figure 15. Count of patient status and surgery type.



This analysis depicts the type and Histology and patients protein status. Here we can see that patient with Mucinous carcinoma their protein status is low, patient with Infiltrating ductal carcinoma have higher number of protein status.(refer figure 16)

Figure 16. Count of patient histology and protein status.



Chapter 5- Discussion

In my study I analysed the patient age with patient status, can see that people who ages between 30-39 total number of cases are 4.7%. Breast cancer, patient age and tumor stage can help to understand mortality rates among women. Research has established that age effect the survival effect of patient. In my study largest group of patient are from 50-59 year of age group that is 27%. Survival effect among patient above 70 years of age is 11%.

LaufeyTRYGGVAD´ et.al. 2002 in his study said that age should be taken into account when studying about breast cancer. Jasmine Brandt etl 2015 in his study concluded that 80 or more years had a statistically significant higher 10-year mortality rate. When adjusted for potential confounders, including stage at diagnosis, the associations only remained statistically significant for women aged 80 years or more.

In the study Ning liue.al.,2018. Mentioned that breast cancer diagnosis between 2005 to 2010. The five year survival rate for male patient is lower than female patient,(82.8 vs 88.5%).

In my study as 95% of women are included are included in the study, 20% among them are not alive. Out of total male 25% among them are dead. In my study analysis also shows the survival rate among male patient is lower than female patients.

In their study Anita Agrawal et.al., 2021. Mentioned that the mean age of diagnosis for male patient is 68.6 years and male patients are at higher number of stage 3 and 4 tumour stage and death among male patients are higher when compared with female data for age, race, stage and grade.

In my study also it shows that higher number of male patients are at stage 2 and 3 of tumour and death among male patients are higher.

In their study laura c. Hanson et.al., 2022. Said that when compare patient stage and mortality in days 32% of patient at stage 4 died within 60days they said that acute illness hospitalization is a sentinel event in stage 4 cancer sort term mortality is high.

In my study analysis when compared the tumour stage with patient status it depicts that 23% of people who are at their stage are dead.

Additionally. patient's age and the stage of their tumour also have corelation Here, we can observe that the age group with the highest percentage of persons Tumour stage is between 50 and 59, at 27%, and between 60 and 69, it is 24%.

Future scope of Improvement

In my study data includes few indicator column only, we don't have any awareness regarding patient's lifestyle data and smoking habits. Information regarding this can help in better outcome about how patient is affected by their lifestyle. With the advanced technologies and data management practices, Although the healthcare business has made significant progress, data management procedures in the field are still lacking. Only a small amount of data is gathered electronically and organised for study. Healthcare applications necessitate more efficient compilation frameworks as well as data conversion, which includes automated structured and unstructured data conversion. Also different sections of the world have varied levels of data management.

Data is being gathered for record purposes and future reference. It is not enough to just the collect data; what counts is how it is used to produce value for patients and other stakeholders

Chapter 6. Conclusion

The process outlined in this paper may be used to generate an easy-to-use Tableau display for novices to investigate healthcare data. Tableau offers a variety of visualisations, including unique options like radar charts with user purpose and tables, heatmaps, maps, bar charts, tree maps, line charts, and bubble charts. Filters, drop-down lists, the computed field, and other tools can be used by users to construct interactive visualisations that convey the desired format. When we compare the data we can understand the cause of disease and act accordingly. By working on data only we know the pattern among male and female.

These findings confirm that there is need for research for age and gender factor breast cancer patients. During this report we found that majority of patient who suffered from breast cancer belonged to the age group 50-59(27%) , survival rate was higher in females(80%) than males (75%) and majority of male patients had tumour stage 2 (75%).

Chapter 7.Reference

1. Inseok Ko, and Hyejung Chang. (2017). Interactive Visualization of Healthcare Data Using Tableau. PMID: 29181247, 23(4): 349–354.

2. Shenhui Jiang, Shiaofen Fang , Sam Bloomquist , Jeremy Keiper et.al.,(2016). Healthcare Data Visualization: Geospatial and Temporal Integration. Volume 2: IVAPP, pages 214-221, ISBN: 978-989-758-175-5.

3. K. A. Narayan, M. Siva Durga and Prasad Nayak(2021). Need for Interactive Data Visualization in Public Health Practice: Examples from India. International Journal of Preventive Medicine.

4. Gotz D, Wang F, Perer A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. J Biomed Inform. 2014 Apr;48:148–59.

5. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare (Basel) 2018 May 23;6(2):E54. doi: 10.3390/healthcare6020054.

6. Gandomi Amir and Haider Murtaza. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35, 137-144.

7. Luina Pani, Somnath Karmakar, Chinmaya Misra, Satya Ranjan Dash. (2019) Multilevel classification framework of fMRI data: A big data approach, in: Big Data Analytics for Intelligent Healthcare Management, Academic Press, pp. 151–174.

8. Mohammed O Razi and Talab A Samani. (2014). Clinical Data Warehouse Issues and Challenges. International Journal of u-and e-Service, Science and Technology Vol.7, No.5. pp.251-262.

9. Jolly Anjali and Tripathi Pooja. (2019). Role of Big Data Analytics in Healthcare. International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958 (Online), Volume-9 Issue-1

10. Ning lu, Kimberly j. johnson, Cytia x.Ma, (2018). Male breast cancer: An updated surveillance Epidemiology, and End results Data Analysis. E997-E1002 Vol. 18, issue 5.

11. Maria Teresa Brown and Jane A. McElroy (2017). Unmet support needs of sexual and gender minority breast cancer survivors. Supportive care in cancer. 26, 1189-1196.

12. Amanze, Chibuike Bethran and Chikezie Francis Oguji. (2022). Data Analytics and Visualization in the Health Sector. International Journal of Trend in Research and Development, Volume 9(3), ISSN: 2394-9333

13. Ying Yang And Tao Chen (2019) Analysis and Visualization Implementation of Medical Big Data Resource Sharing Mechanism Based on Deep Learning. School of Management, Xinxiang Medical University, Xinxiang 453003, China 2The Third Affiliated Hospital of Xinxiang Medical University, Xinxiang 453003, China

14. Chui KK, Wenger JB, Cohen SA, Naumova EN. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. PLoS One. 2011 Feb 15;6(2):e14683. doi: 10.1371/journal.pone.0014683

15. Dayal Mohit and Singh Nanhay. (2016). Indian Health Care Analysis using Big Data Programming Tool. Procedia Computer Science 89, 521 – 527.

16. Hassan Shafiqul , Dhali Mohsin , Zaman Fazluz , Tanveer Muhammad. (2021). Big data and predictive analytics in healthcare in Bangladesh: regulatory challenges. Heliyon 7 e07179.

17. Ashwin Belle, Raghuram Thiagarajan, S.M. Soroushmehr, Fatemeh Navidi, Daniel A. Beard, Kayvan Najarian, (2015). Big data analytics in healthcare, BioMed Res. Int.

18. Ratra Ritu and Gulia Preeti. (2019). Big Data Tools and Techniques: A Roadmap for Predictive Analytics. International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-2.

19. Kumar sunil and singh Maninder. (2019). Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools. Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools. Big Data Minning and Analytics. ISSN2096- 0654 pp48– 57.

20. Wang Lidong and Alexander Ann. (2019) Big Data Analytics in Healthcare Systems. International Journal of Mathematical, Engineering and Management Sciences, Vol. 4,17–26.

21. Aboudi Naoual EL and Benhlima Laila. (2018) Big Data Management for Healthcare Systems: Architecture, Requirements, and Implementation. Advances in Bioinformatics, Volume 2018, Article ID 4059018.

22. Abbo Ismail Brahim and Dr. R. Suchithra. Big Data Analytics in Healthcare.(2016) International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181.

23. Belle Ashwin, Thiagarajan Raghuram, Soroushmehr S.M. Reza et al. (2015) , Big Data Analytics in Healthcare. BioMed Research International, Article ID 370194

24. Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. Healthcare informatics research, 22(3), 156-163.

25. Batko Kornelia and Ślęzak Andrzej Ślęzak (2022). The use of Big Data Analytics in healthcare. Journel of Big data, 9:3.

26. Sri Venkat Gunturi Subrahmanya1 · Dasharathraj K. Shetty2 · Vathsala Patil3 · B. M. Zeeshan Hameed4 · Rahul Paul5 · Komal Smriti3 · Nithesh Naik6 · Bhaskar K. Somani7. (2021). The role of data science in healthcare advancements: applications, benefits, and future prospects. Irish Journal of Medical Science(1971 -)

27. Mohammed. Far H Behrouz, Naugler Christopher. (2014). Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trendsdz. BioData Mining, 7:22.

28. Pai Vaikunth. (2016). Big Data New Challenges, Tools and Technique. International Journal of Engineering Research and Modern Education (IJERME) ISSN (Online): 2455 – 4200. Volume I, Issue I.