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

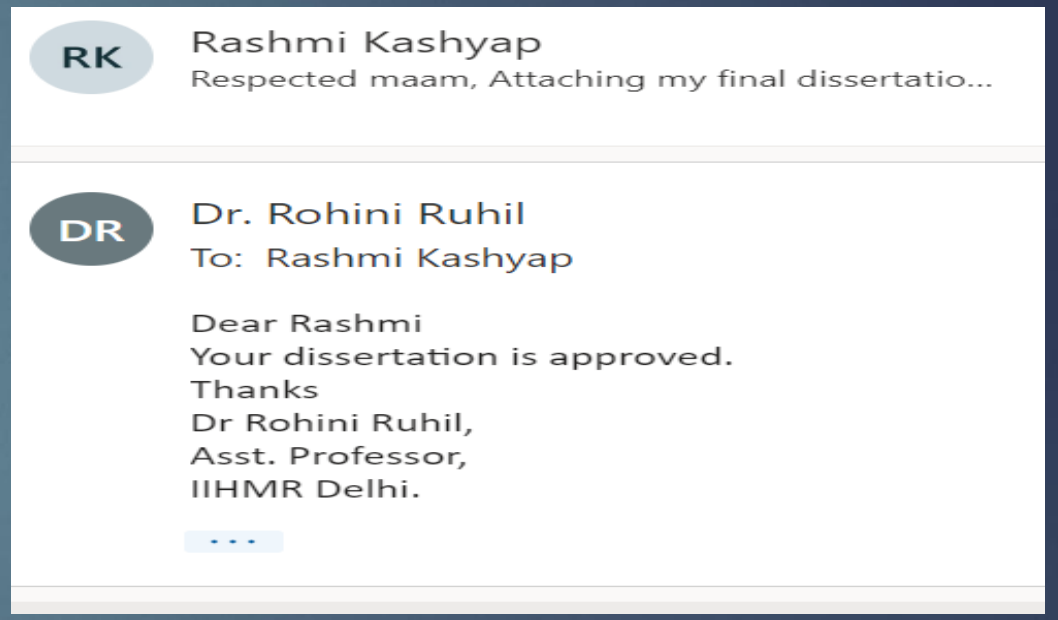
Karkinos Healthcare

RASHMI KASHYAP

DR. ROHINI RUHIL

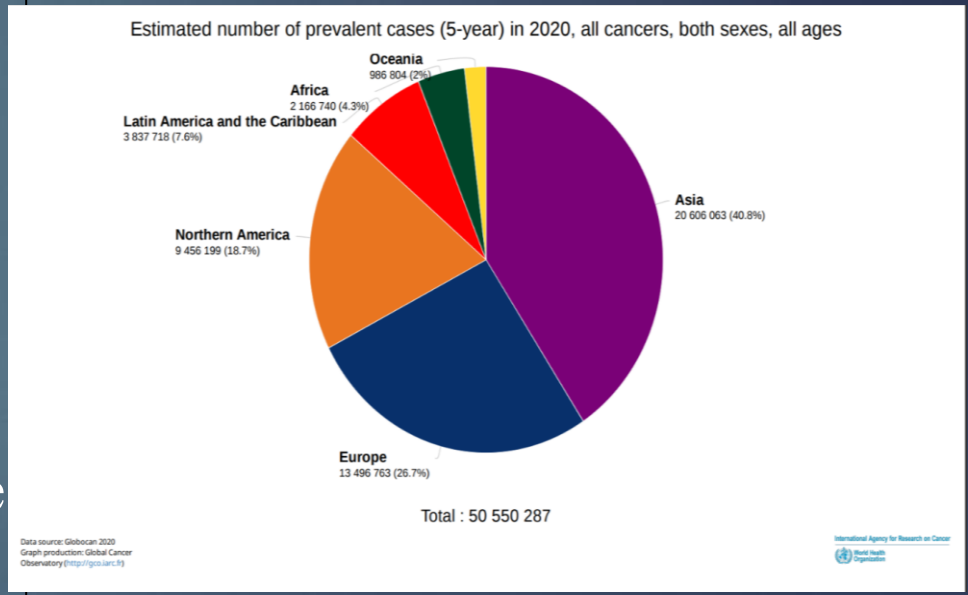
IIHMR DELHI

Screenshot of Approval



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.
- 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.



Healthcare data

- Diverse forms of healthcare data sources include clinical text, biomedical images, EHRs, genomic data, biomedical signals, sensing data, and social media.
- 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.
- 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.

Objectives 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.

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

Results

- 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 2)

Figure 2. 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 3)

Figure 3. Number of Tumour staging

VALUE	FREQUENCY
II	189
III	81
I	64
Total rows	335

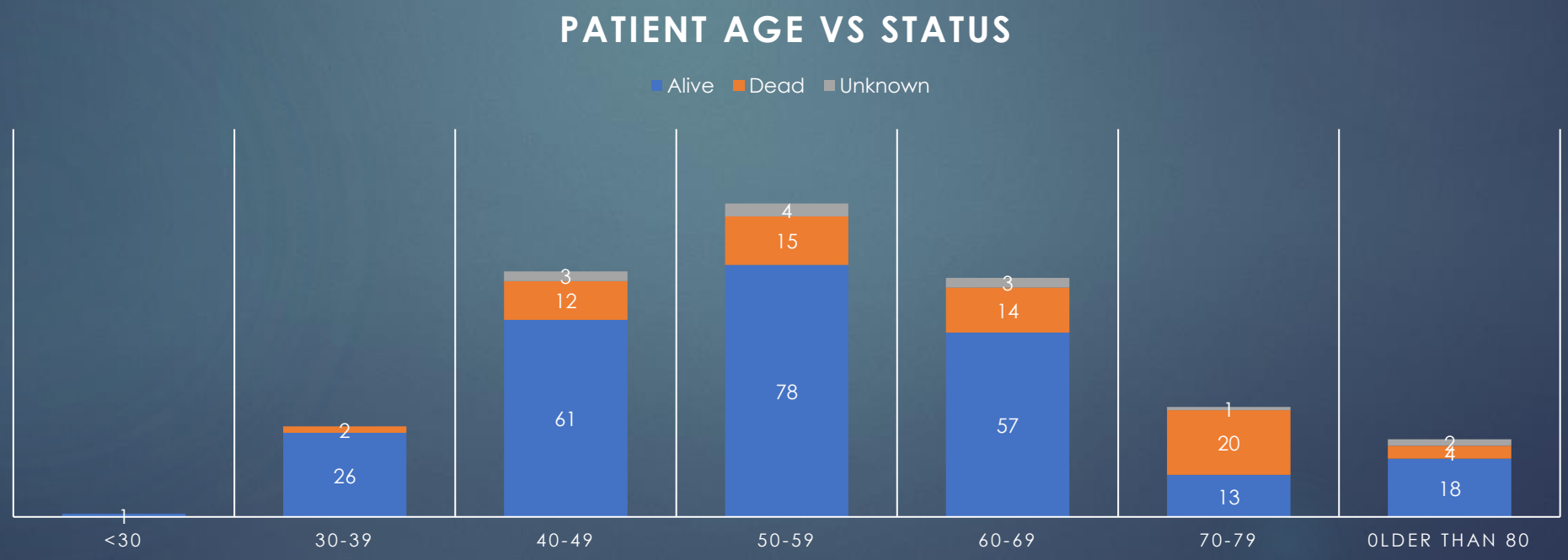
- The distribution of patient status was also studied (refer figure 4)

Figure 4.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 5).

Figure 5. Count of patient Age vs status.

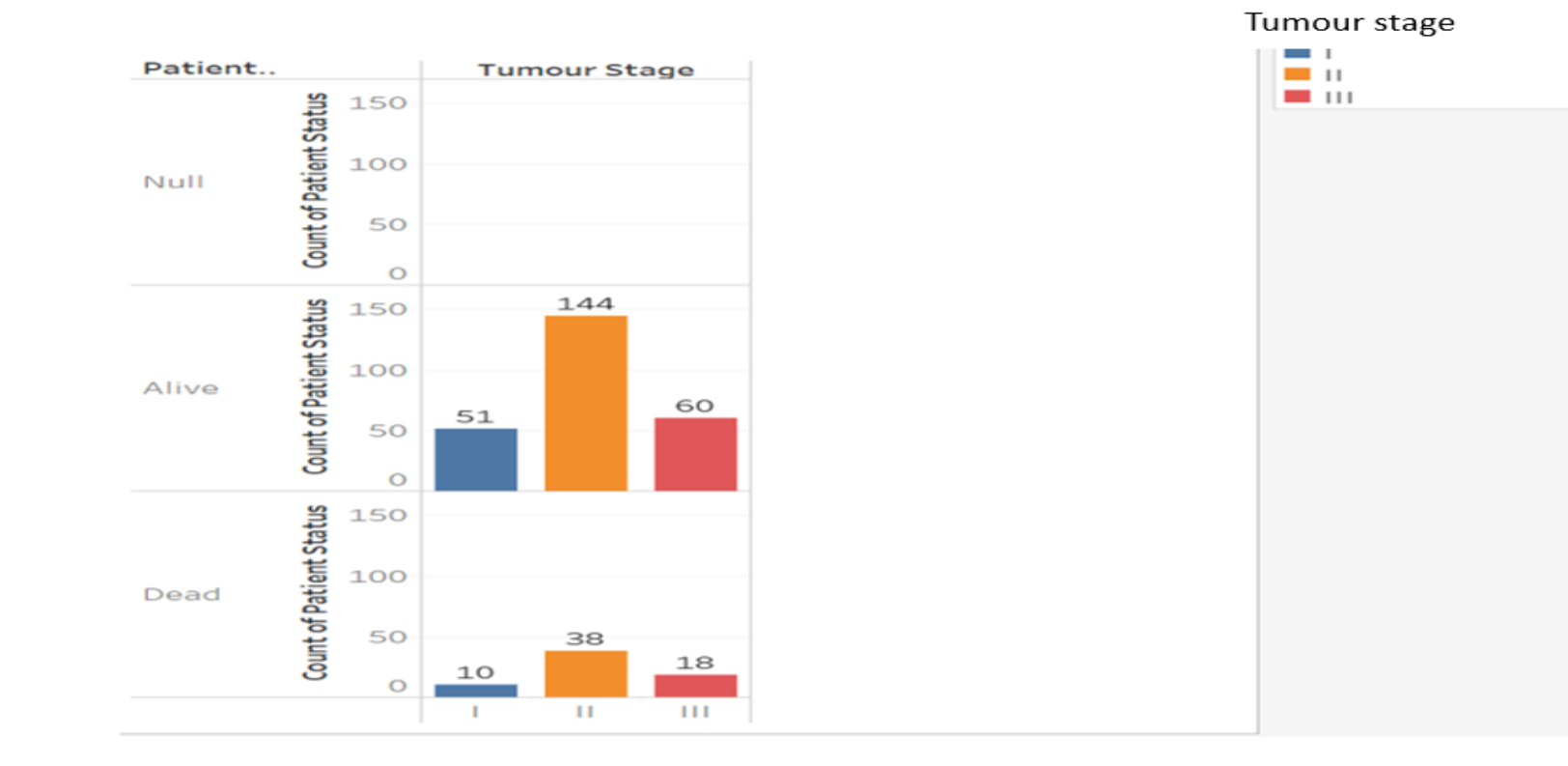


- This analysis shows that 95% of women included in this study, 79%women are alive, 20% of women are dead, and 1% of females condition is not known.1% male included in this study. out of which 75% men are alive and 25% men are dead.(refer figure 6).

Figure 6. Count Gender vs patient status.

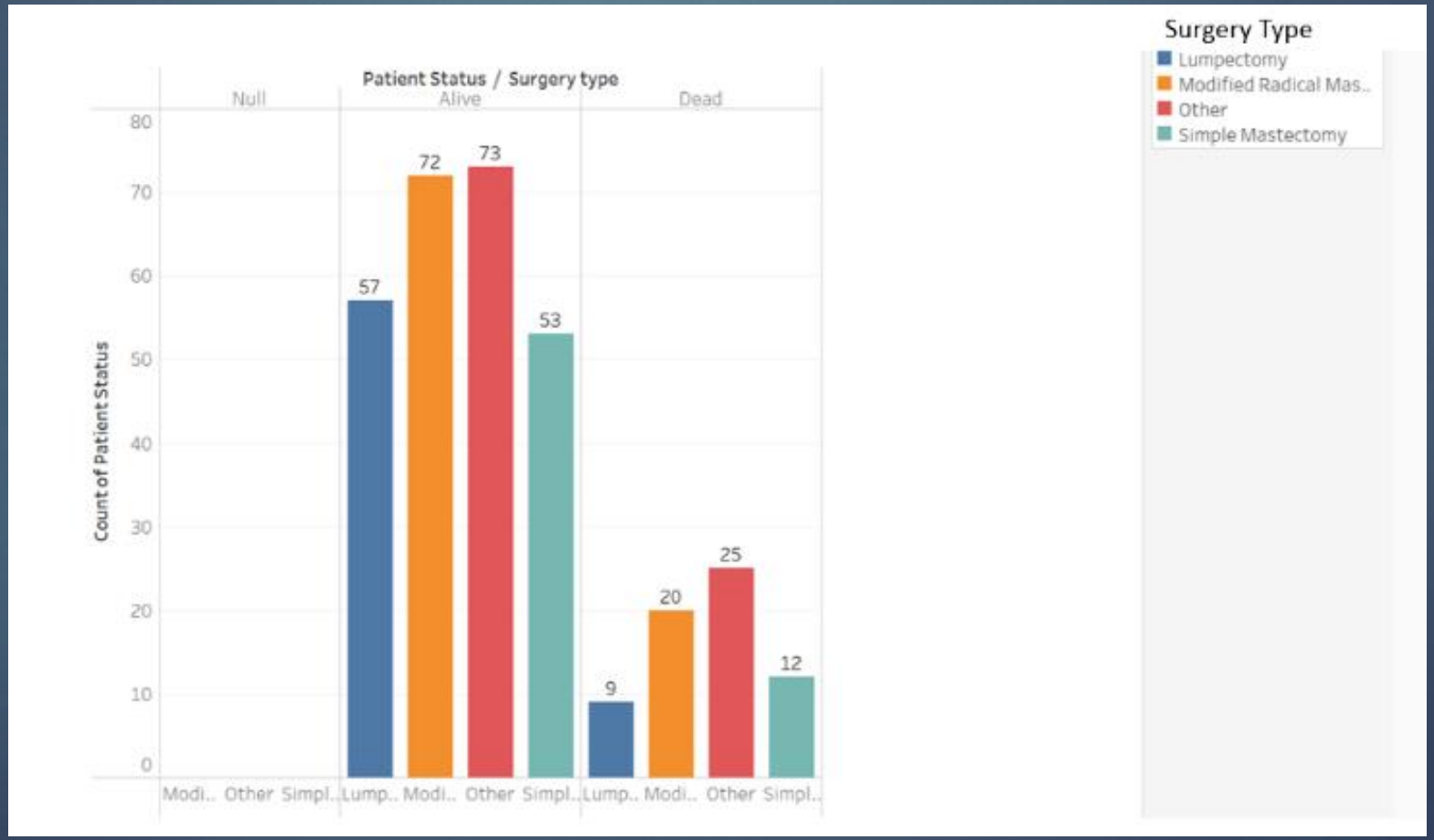
- Here it depicts the Patient status vs tumour size, it shows that 15%, 43%, and 17% patient who are at there Tumour stage 1, 2, 3 are alive. And 2%, 11%, 5% Patient who are at there respective tumour stage of 1, 2, 3 are dead. And 7% patient's data is not known.(refer figure 7)

Figure 7. Count of tumour stage vs patient status



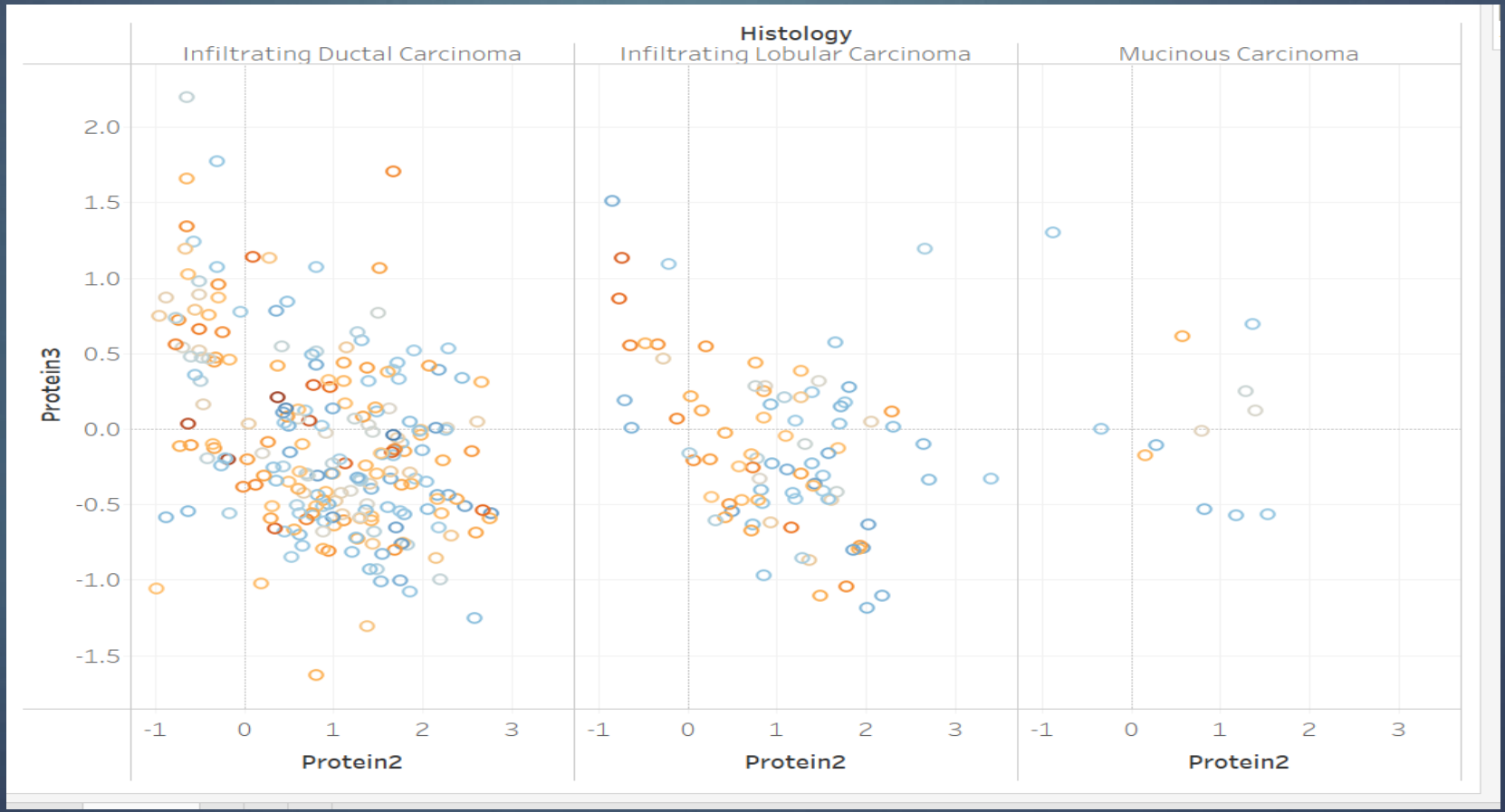
- Below analysis shows that there are 96% patient who undergone surgery, 76% of patients are alive and 19% patient are dead.(refer figure 8)

Figure 8. 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 9)

Figure 9. Count of patient histology and protein status.



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 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.

Limitations of the Study

- Study based on secondary data.
- Few more information could help in better results.

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%).

References

1.GLOBOCAN 2020: New Global Cancer Data | UICC [Internet]. [cited 2022 may 25]. Available from: https://www.uicc.org/news/globocan-2020-newglobal-cancer-data.

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

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

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

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

6. 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.

7.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.

IIHMR DELHI

Thank You

ANY QUESTIONS

Dissertation Experiences

What did you learn (skill/ topic)?

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. Archetype mapping and archetype template formation.

Overall self comments on Dissertation

- I am having great learning experience with karkinos healthcare. Each day learning and exploring new thing.

IIHMR DELHI

Pictorial Journey