**BREAST CANCER**

**NAME: SRAVAN SAI CHINNAMSETTI**

**W9621054**

**SECTION 2 : BUSINESS INTELLIGENCE SOLUTIONS**

1. **Executive Summary**
   1. **Introduction**

An extensive collection of information pertaining to breast cancer patients is provided, encompassing critical particulars such as patient identification, age at the time of diagnosis, the type of breast surgery conducted, the specific categorization of cancer type, detailed classification of cancer subtypes, cellularity, administration of chemotherapy, categorization based on PAM50 and Claudin-low subtypes, cohort specifications, estrogen receptor (ER) status determined through immunohistochemistry (IHC), histologic grade of the neoplasm, HER2 status evaluated using SNP6 analysis, the histologic subtype of the tumor, details about hormone therapy, inferred menopausal state, information on integrative clusters, the laterality of the primary tumor, the number of positive lymph nodes examined, mutation count, Nottingham prognostic index, On cotree code, duration of overall survival, overall survival status, progesterone receptor (PR) status, the application of radiotherapy, duration and status of relapse-free intervals, patient gender, classification based on the 3-gene model, tumor size, tumor stage, and the vital status of the patient. This all-encompassing dataset serves as a valuable tool for investigating trends, associations, and insights into the diagnosis, treatment, and outcomes of breast cancer cases.

1. Top of Form

1.2 **Key findings**

**Survival Analysis:** Assess the overall duration and status of survival to uncover trends and factors impacting patient outcomes. Investigate the correlation between different features (such as tumor stage, subtype, and treatment types) and overall survival.

**Treatment Effectiveness:** Evaluate the efficacy of diverse breast surgery methods, chemotherapy, hormone therapy, and radiotherapy in relation to both overall survival and relapse-free intervals. Explore the influence of the 3-gene classifier subtype on the response to treatment.

**Biomarker Associations:** Examine associations between molecular biomarkers (ER, PR, HER2) and specific subtypes to gain insights into their prevalence and impact on prognosis.

**Clinicopathological Characteristics:** Analyze connections between age at diagnosis, tumor size, histologic grade, and lymph node involvement to detect patterns and trends.

**Genomic Analysis:** Explore the relationship between the mutation count and overall survival, providing insights into the genetic composition of breast cancer within this cohort.

**Cohort-Specific Patterns:** Investigate distinct patterns within specific cohorts, taking into consideration variables like PAM50 and Claudin-low subtypes, integrative clusters, and the Nottingham prognostic index.

**Relapse-Free Periods:** Examine factors influencing intervals without relapse, including the effects of radiotherapy, hormone therapy, and tumor characteristics.

**Demographic Insights:** Assess the impact of demographic factors such as gender, inferred menopausal state, and primary tumor laterality on outcomes.

A pink sign with text

Description automatically generated

Fig 20: Breast cancer home page (navigator page)

**1.3 Scope of work and Power BI questions**

This power BI project will focus on BREAST CANCER analysis.

The questions that this project seeks to answer are as follows:

* **Overview:**

How is the distribution of breast cancer subtypes related to patients' ages at diagnosis, shedding light on age-specific prevalence within the dataset?

* **Breast Cancer Treatment:**

What treatment patterns emerge concerning chemotherapy and hormone therapy usage across different receptor statuses (ER/HER2) and tumor subtypes?

* **Clinical Characteristics:**

Are there discernible relationships between tumor size, histologic grades, and lymph node involvement among distinct breast cancer subtypes?

* **Survival Analysis:**

Can we identify correlations between overall survival durations and specific treatment interventions, receptor statuses, or tumor stages, elucidating potential prognostic factors within the dataset?

**1.3 Recommendations**

Her are the recommendations for this Breast cancer project:

1.Analyze the effectiveness of different breast cancer treatments, such as chemotherapy and hormone therapy, in improving overall survival and reducing relapse.

2.Consider developing personalized treatment plans based on receptor statuses (ER, HER2) and other relevant factors.

3.Identify high-risk patients by evaluating factors such as tumor size, lymph node involvement, and tumor stage.

4. Evaluate the impact of radiotherapy on overall survival and relapse-free intervals.

5. Investigate the impact of hormone therapy on overall survival and relapse-free periods. Identify specific subtypes or patient groups that may benefit the most from hormone therapy interventions.

**Findings based on analysis and Evaluation**

To analyse the questions mentioned above, various graphs are plotted and each graphs screen shorts and those explanation are given below.

**1.How is the distribution of breast cancer subtypes related to patients' ages at diagnosis, shedding light on age-specific prevalence within the dataset?**

A screenshot of a pink screen

Description automatically generated

Figure 21: card visuals for the No. of patients, Avg Age at diagnosis, Positive ER, Survival (yrs)

**Number of Patients:**

The metric "Number of Patients" signifies the total count of individuals within the dataset who have received a diagnosis of breast cancer. This metric is foundational in comprehending the dataset's scale and the population being considered for further analysis.

**Age at Diagnosis:**

The metric "Age at Diagnosis" represents the average or median age at which breast cancer is identified in the specified dataset. This critical information provides insights into the distribution of patient ages, aiding in the recognition of age-related trends and potential correlations with other clinical characteristics.

**Positive ER (Estrogen Receptor):**

The metric "Positive ER" emphasizes the tally or proportion of patients exhibiting a positive estrogen receptor status. Given the influential role of estrogen receptor status in breast cancer treatment decisions, particularly in guiding hormone therapy choices, understanding the prevalence of positive ER cases is essential for tailoring effective treatment strategies.

A screenshot of a computer

Description automatically generated

Figure 22: slicers for the cancer type detailed and Type of breast surgery

The Types of Cancer in Detail slicer offers an interactive means to delve into the dataset's intricate details concerning specific breast cancer types. This dynamic slicer enables users to selectively filter data, homing in on distinct cancer subtypes for thorough examination. By choosing a particular cancer subtype through the slicer, users can narrow down the dataset to scrutinize cases associated with that specific detailed cancer type. This functionality facilitates focused inquiries into the attributes, treatments, and outcomes linked to each subtype.

For instance, this slicer allows users to investigate distinct molecular subtypes like HER2-positive or triple-negative breast cancer, exploring how these subtypes correlate with patient demographics, treatment responses, and survival outcomes. The detailed cancer slicer elevates the granularity of data exploration, offering a more nuanced perspective on the diverse array of breast cancer types prevalent in the dataset.

The Types of Breast Surgery slicer empowers users to navigate the dataset based on the specific surgical interventions undertaken by breast cancer patients. This slicer enables users to dynamically filter data according to the types of surgical procedures conducted, such as lumpectomy, mastectomy, or other surgical interventions. By selecting a specific type of breast surgery through the slicer, users can concentrate on cases associated with that surgical approach.

This slicer facilitates in-depth examinations into the relationships between the chosen type of breast surgery and various factors like tumor size, lymph node involvement, or overall survival outcomes. It presents a user-friendly interface for exploring patterns and trends related to surgical interventions, allowing healthcare professionals, researchers, or analysts to extract insights into the surgical landscape within the breast cancer dataset.

In conclusion, both slicers stand as invaluable instruments for interactive data exploration, enabling users to customize their analysis based on specific cancer subtypes and surgical interventions. These slicers enhance adaptability and depth of comprehension when navigating the intricate facets of breast cancer data.

A blue and red pie chart

Description automatically generated

Figure 23: Surgery type by total patients

A pie chart depicting the distribution of patient IDs based on the types of surgery serves as a concise illustration of the diverse surgical interventions undergone by breast cancer patients in the dataset. This visual representation facilitates a swift comprehension of the relative prevalence of various surgical procedures within the cohort. Patient IDs function as unique identifiers for individual cases, while the types of surgery categorize cases based on the specific procedures conducted.

In this graphical representation, each segment of the pie chart corresponds to a distinct type of surgery, and the size of each segment correlates with the number of patients who underwent that specific surgical procedure. The types of surgery encompass a range of procedures, including lumpectomy, mastectomy, and other surgical approaches integral to breast cancer management.

The pie chart provides a succinct overview, enabling easy identification of the most common surgical procedures and detection of notable variations in their distribution. This visual tool proves valuable for healthcare professionals, researchers, and clinicians seeking a rapid and intuitive comprehension of the surgical landscape within the breast cancer patient population.

It is crucial to acknowledge that while pie charts effectively convey categorical data and relative proportions, thoughtful consideration is necessary to ensure clarity and prevent visual clutter, especially when dealing with numerous distinct categories. Additionally, for a more comprehensive analysis of specific surgical trends or correlations within the dataset, supplementary information or alternative visualizations may be warranted.

A blue and red pie chart

Description automatically generated

Figure24:ER status by total number of patients

This pie chart succinctly outlines the composition of breast cancer patients in the dataset based on their Estrogen Receptor (ER) status. The presence or absence of ER in breast cancer is a critical factor influencing diagnostic and treatment decisions. By segmenting patients into distinct categories according to their ER status, the chart provides a swift and visual comprehension of the distribution across these categories.

ER-Positive (ER+):

This portion of the pie chart represents patients with breast cancer whose tumors express the Estrogen Receptor. The specific segment visually highlights the prevalence of ER-positive cases within the dataset.

ER-Negative (ER-):

Conversely, the section designated for ER-negative patients illustrates those with breast cancer whose tumors do not express the Estrogen Receptor. This segment serves to emphasize the proportion of ER-negative cases in the dataset.

Unknown ER Status:

A separate segment accounts for patients with an unknown ER status, where the information is either unavailable or not yet determined. This additional category recognizes the instances where ER status remains unspecified. The simplicity of the pie chart facilitates an immediate visual grasp of the distribution of ER status among breast cancer patients. The size of each segment is directly proportional to the number of patients falling into that specific ER status category. Healthcare professionals and researchers can leverage this visual representation to swiftly comprehend the prevalence of ER-positive and ER-negative cases, contributing valuable insights to treatment planning and ongoing research endeavors.

A graph showing the number of cellulite by breast surgery

Description automatically generated

Figure25: types of cellularity by types of breast surgery

The horizontal axis of the visual likely depicts diverse aspects such as time intervals, categories, or other pertinent variables. In this context, it portrays cellularity types, encompassing classifications like "High," "Moderate," and "Low," serving as the primary categorical variable. The vertical axis measures the quantity of patient IDs, representing the number of individuals associated with each cellularity classification. The shaded regions beneath the lines corresponding to each cellularity type visually convey the patient count, offering an illustrative representation of the dataset. The legend associated with the chart delineates distinct kinds of breast surgery, each identified by a distinct color. This inclusion enables a simultaneous exploration of the interconnected dynamics involving cellularity, patient count, and the prevalence of diverse surgical interventions.

A graph with red lines and black text

Description automatically generated

Figure 26: number of patients by cancer type in detailed

A stacked bar chart effectively depicts the distribution of breast cancer cases across various detailed cancer types, presenting both the total patient count and the number of unique patient IDs associated with each cancer subtype. This visualization offers a comprehensive overview of the dataset, emphasizing the relative prevalence of distinct cancer types and the corresponding number of affected individuals.

On the chart, the horizontal axis categorizes the different detailed cancer types, while the vertical axis represents the patient count. Each bar is partitioned into segments, with each segment representing a specific detailed cancer subtype. The height of each segment corresponds to the number of patients within that particular subtype, contributing to the overall patient count for the given detailed cancer type.

The stacked configuration of the bars enables the differentiation of patient counts within each cancer subtype. Each segment within a bar corresponds to a unique patient ID, collectively contributing to the total count of patients in that detailed cancer type. Consequently, this chart not only portrays the overall distribution of breast cancer subtypes but also offers insights into the individual patient count within each subtype.

To summarize, the stacked bar chart serves as a robust visual aid for understanding the distribution of breast cancer cases across detailed cancer types, providing clarity on both the cumulative patient count and the distinct patient IDs within each specific subtype.

A graph of a patient

Description automatically generated with medium confidence

Figure27: tornado graph for patient’s vital status of breast surgery

A tornado chart is employed here to visually represent the distribution of patients based on their vital status, the count of patient IDs, and the types of breast surgery they underwent. This type of chart is well-suited for displaying the magnitude and directionality of the differences in counts between the various categories.

**Vital Status:** The tornado chart first segregates patients based on their vital status, distinguishing between those who are deceased and those who are alive. The length of each bar reflects the count of patient IDs associated with each vital status category.

**Count of Patient IDs**: The bars extend in opposite directions, with one side representing the count of deceased patients and the other side representing the count of living patients. The length of each bar is **Type of Breast Surgery**: Within each vital status category, the chart further segments the data based on the type of breast surgery performed. The varying lengths of bars for each surgery type provide a clear comparison of the distribution of surgical interventions within both deceased and living patient groups.