

Project Proposal

Gautham Nagaraj, Romith Bodanda, Lamis

2025-01-20

Introduction: Cancer remains one of the most pressing global health challenges, with early detection and effective treatment being critical to improving patient outcomes and survival rates. This project aims to investigate patterns and trends in cancer patient data to identify key factors influencing survival, tumor progression, and treatment responses. By employing advanced data visualization techniques, we aim to uncover insights that can inform clinical decision-making and contribute to the broader field of medical data analysis. Our study leverages a comprehensive cancer dataset that includes demographic, clinical, and pathological information. This dataset offers an opportunity to explore complex relationships among variables such as tumor characteristics, survival outcomes, and hormone receptor statuses. Through a systematic analysis, we seek to enhance our understanding of cancer progression and provide evidence-based recommendations for improving patient care.

Guiding Questions:

1. What are the demographic factors (e.g., age, race, and marital status) associated with higher survival rates?
 - Age: The correlation between age and survival rates is negatively correlated i.e individuals who are of higher age have a higher mortality rate when compared to younger individuals. This due to the risk of developing diseases due to older age.
 - Race and Ethnicity: Studies consistently show disparities in health outcomes across different racial and ethnic groups. For example, in the United States, Black individuals often experience higher rates of certain diseases (like heart disease and diabetes) and lower life expectancy compared to white individuals. These disparities are often linked to systemic factors like socioeconomic inequalities, access to quality healthcare, and historical and ongoing discrimination. *Reference*
 - Marital Status: It has been studied that married individuals often exhibit better health when compared to those that have been divorced or single. Studies say that this could be due to a better support system and manageable financial stress. *Reference*
2. How do tumor characteristics such as size, grade, and hormone receptor status correlate with survival outcomes?
 - Tumor Size: Smaller tumors - Generally associated with better prognoses. Smaller tumors are often easier to treat completely, reducing the risk of cancer spreading.
 - Tumor Grade: Grade refers to how aggressive the tumor appears under a microscope.
 - Lower grades: Indicate slower-growing tumors with less aggressive features, often linked to better survival.
 - Higher grades: Indicate faster-growing tumors with more aggressive features, often associated with poorer survival.
 - Hormone Receptor Status (Relevant for hormone-dependent cancers like breast):
 - Estrogen Receptor (ER): These tumors rely on hormones for growth. The greater the hormone levels produce the greater the size of the tumor *Reference*

The guiding questions will help provide a blueprint as to how the data will be wrangled/visualized and see if what factors affect the survivalability of the patient the most.

Dataset: This dataset of breast cancer patients was obtained from the 2017 November update of the SEER Program of the NCI, which provides information on population-based cancer statistics. The dataset involved female patients with breast cancer diagnosed in 2006-2010. Patients with unknown tumor size, examined regional LNs, regional positive LNs, and patients whose survival months were less than 1 month were excluded; thus, 4024 patients were ultimately included. The columns included in the dataset are as follows - Age: This data item represents the age of the patient at diagnosis for this cancer. The code represents the patient's actual age in years. Race: is based on the race variables and the American Indian/Native American IHS link variable. Marital status: This data item identifies the patient's marital status at the time of diagnosis for the reportable tumor. Stages: T stage, N stage, 6th stage, these columns provide information on the stage of the breast cancer. GRADE: The grade is a description of how the cancer cells look compared to normal cells. *Reference* A STAGE: Describes whether the tumor has spread to different regions or not. Tumor Size: Describes what size the tumor is at the time of recording. ESTROGEN STATUS: Measures the level of estrogen (a hormone) at the time of recording. PROGESTERONE STATUS: Measures the level of progesterone (a hormone) at the time of recording. REGIONAL NODES EXAMINED: Records the total number of regional lymph nodes that were removed and examined by the pathologist. REGIONAL NODES POSITIVE: Records the exact number of regional lymph nodes examined by the pathologist that were found to contain metastases (cancer spread from another part). SURVIVAL MONTHS: The number of months the patient may survive. Status: Describes whether a patient is alive or dead.