Breast cancer stage report exploration

Data Analyst: Omosona Alfred

Data Source: kaggle

License: the license permits me to modify and share findings on this dataset

Tableau.Dashboard:
https://public.tableau.com/views/DataExplorationBreastCancerstagereport/Dashboard2?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Exploration questions;

These are the questions I answered with the dataset

1. Exploration of Differentiation and survival month

2. Correlation of Tumor size, survival month and status

3. which of the stages has higher survival rate

4. which combination of the different classification has higher no of dead persons and lowest number of surviving months

Brief explanation of terms

In the TNM system

The **T** refers to the size and extent of the main tumor. The main tumor is usually called the primary tumor.

The **N** refers to the number of nearby lymph nodes that have cancer.

The **M** refers to whether the cancer has metastasized. This means that the cancer has spread from the primary tumor to other parts of the body.

T1, T2, T3, T4: Refers to the size and/or extent of the main tumor. The higher the number after the T, the larger the tumor or the more it has grown into nearby tissues.

N1, N2, N3: Refers to the number and location of lymph nodes that contain cancer. The higher the number after the N, the more lymph nodes that contain cancer.

A stage

In situ—Abnormal cells are present but have not spread to nearby tissue.

Localized—Cancer is limited to the place where it started, with no sign that it has spread.

Regional—Cancer has spread to nearby lymph nodes, tissues, or organs.

Distant—Cancer has spread to distant parts of the body.

6th stage

Stage 0: Stage zero (0) describes disease that is only in the ducts of the breast tissue and has not spread to the surrounding tissue of the breast. It is also called non-invasive or in situ cancer

Stage IA: The tumor is small, invasive, and has not spread to the lymph nodes

Stage IB: Cancer has spread to the lymph nodes and the cancer in the lymph node is larger than 0.2 mm but less than 2 mm in size. There is either no evidence of a tumor in the breast or the tumor in the breast is 20 mm or smaller

Stage IIA: Any 1 of these conditions:

There is no evidence of a tumor in the breast, but the cancer has spread to 1 to 3 axillary lymph nodes. It has not spread to distant parts of the body

The tumor is 20 mm or smaller and has spread to 1 to 3 axillary lymph nodes

The tumor is larger than 20 mm but not larger than 50 mm and has not spread to the axillary lymph nodes

Stage IIB: Either of these conditions:

The tumor is larger than 20 mm but not larger than 50 mm and has spread to 1 to 3 axillary lymph nodes

The tumor is larger than 50 mm but has not spread to the axillary lymph nodes (T3, N0, M0).

Stage IIIA: The tumor of any size has spread to 4 to 9 axillary lymph nodes or to internal mammary lymph nodes. It has not spread to other parts of the body. Stage IIIA may also be a tumor larger than 50 mm that has spread to 1 to 3 axillary lymph nodes.

Stage IIIB: The tumor has spread to the chest wall or caused swelling or ulceration of the breast, or it is diagnosed as inflammatory breast cancer. It may or may not have spread to up to 9 axillary or internal mammary lymph nodes. It has not spread to other parts of the body.

Stage IIIC: A tumor of any size that has spread to 10 or more axillary lymph nodes, the internal mammary lymph nodes, and/or the lymph nodes under the collarbone. It has not spread to other parts of the body.

Stage IV (metastatic): The tumor can be any size and has spread to other organs, such as the bones, lungs, brain, liver, distant lymph nodes, or chest wall (any T, any N, M1). Metastatic cancer found when the cancer is first diagnosed occurs about 6% of the time.

- Overview of the table

Age	Race	Marital_Status	T_Stage	N_Stage	6th_Stage	differentiate	Grade	A_Stage	Tumor_Size	Estrogen_Status	Progesterone_Status	Regional_Node_Examined	Reginol_Node_Positive	Survival_Months	Statu
68	White	Married	T1	N1	IIA	Poorly differentiated	3	Regional	4	Positive	Positive	24	1	60	Alive
50	White	Married	T2	N2	IIIA	Moderately differentiated	2	Regional	35	Positive	Positive	14	5	62	Alive
58	White	Divorced	T3	N3	IIIC	Moderately differentiated	2	Regional	63	Positive	Positive	14	7	75	Alive
58	White	Married	T1	N1	IIA	Poorly differentiated	3	Regional	18	Positive	Positive	2	1	84	Alive
47	White	Married	T2	N1	IIB	Poorly differentiated	3	Regional	41	Positive	Positive	3	1	50	Alive
51	White	Single	T1	N1	IIA	Moderately differentiated	2	Regional	20	Positive	Positive	18	2	89	Alive
51	White	Married	T1	N1	IIA	Well differentiated	1	Regional	8	Positive	Positive	11	1	54	Alive
40	White	Married	T2	N1	IIB	Moderately differentiated	2	Regional	30	Positive	Positive	9	1	14	Dead
40	White	Divorced	T4	N3	IIIC	Poorly differentiated	3	Regional	103	Positive	Positive	20	18	70	Alive
69	White	Married	T4	N3	IIIC	Well differentiated	1	Distant	32	Positive	Positive	21	12	92	Alive

- Number of Records

count(*)
4005

1. Exploration of Differentiation and survival month

differentiate	survival_months	status
Well differentiated	87	Dead
Moderately differentiated	62	Alive
Poorly differentiated	60	Alive
Well differentiated	54	Alive
Poorly differentiated	44	Dead
Moderately differentiated	14	Dead

	differentiate	survival_rate
▶	Moderately differentiated	had 169693 survival months in total
	Poorly differentiated	had 76381 survival months in total
	Well differentiated	had 39605 survival months in total

2. Correlation of Tumor size, survival month and status

	Correlation between tumor size and survival months
▶	-0.08569832639724019

3.which of the stages has higher survival rate

	6th_stage	survival_months
▶	IIIC	75
	IIIB	63
	IIIA	62
	IIA	60
	IIB	50

	T_Stage	survival_months
▶	T3	75
	T4	70
	T2	62
	T1	60

	N_stage	survival_months
▶	N3	75
	N2	62
	N1	60

4. which combination of the different classification has higher no of dead persons and lowest number of surviving months

n_stage	status	mortality_num
N1	Dead	268
N3	Dead	178
N2	Dead	161

t_stage	status	mortality_num
T2	Dead	299
T1	Dead	156
T3	Dead	113
T4	Dead	39

T_stage	N_Stage	6th_Stage	status	mortality_num
T2	N1	IIB	Dead	133
T1	N1	IIA	Dead	96
T2	N2	IIIA	Dead	87
T2	N3	IIIC	Dead	79
T3	N3	IIIC	Dead	50
T3	N2	IIIA	Dead	32
T3	N1	IIIA	Dead	31
T1	N3	IIIC	Dead	30
T1	N2	IIIA	Dead	30
T4	N3	IIIC	Dead	19
T4	N2	IIIB	Dead	12
T4	N1	IIIB	Dead	8

a_stage	status	mortality_num
Regional	Dead	572
Distant	Dead	35

T_stage	N_Stage	6th_Stage	status	A_Stage	mortality_num
T2	N1	IIB	Dead	Regional	133
T1	N1	IIA	Dead	Regional	96
T2	N2	IIIA	Dead	Regional	87
T2	N3	IIIC	Dead	Regional	67
T3	N3	IIIC	Dead	Regional	44
T3	N2	IIIA	Dead	Regional	32
T3	N1	IIIA	Dead	Regional	31
T1	N2	IIIA	Dead	Regional	30
T1	N3	IIIC	Dead	Regional	28
T2	N3	IIIC	Dead	Distant	12

	N_stage	status	Survival_Months	mortality_num
▶	N1	Dead	14	268
	N3	Dead	42	178
	N2	Dead	87	161

	T_stage	status	Survival_Months	mortality_num
▶	T2	Dead	14	299
	T1	Dead	64	156
	T3	Dead	42	113
	T4	Dead	63	39

	6TH_stage	status	Survival_Months	mortality_num
▶	IIIA	Dead	91	180
	IIIC	Dead	42	178
	IIB	Dead	14	133
	IIA	Dead	64	96
	IIIB	Dead	63	20

	a_stage	status	Survival_Months	mortality_num
▶	Regional	Dead	14	572
	Distant	Dead	53	35

	marital_status	sum_of_survival_months
▶	Married	189391
	Single	43085
	Divorced	34172
	Widowed	16279
	Separated	2752

OBSERVATIONS

- At first glance it's obvious the 6th stage progression correlates with survival month.
- There is no correlation between tumor size and survival month.
- The advanced the stage the less likely the survival. This agrees very well with early detection saving lives from cancer death.
- Race may not be a factor to consider since it's more as though more white people where recorded in this data. However, except more probability study is done
- We may not be able to readily conclude if marital status is truly as impacting on survival. Which of the marital status has higher percentage of survivors per total patients in each class and would the percentage difference be definitive enough to prove the need for a family function around the patient to improve quality of life? it seems so, although there is need to carry out further studies in this regard.