



# BREAST CANCER (METABRIC) DATASET EXPLORATORY VISUALIZATIONS

By

Dr. Serifat Folorunsho (PhD)

There are 2,509 unique breast cancer patients in METABRIC dataset as mentioned above. Patients had undergone to two different surgeries; Mastectomy (removal of all breast tissue from a breast) or Breast-conserving surgery (removal of a part of the breast that has cancer). Those patients are diagnosed at ages between 21.9 -96.3 and their mean diagnosis age is 60.4. Also, there are 2,506 breast cancer and 3 breast sarcoma patients in the dataset since breast sarcomas are a very rare form of breast cancer, that cover fewer than 1% of all breast cancers. The most common histological subtype of the breast cancer is invasive ductal carcinoma (IDC) with 1865 occurrences. IDC is the most common form of breast cancer, representing 80% of all breast cancer diagnoses. Those indicators show that this dataset reflects real world scenarios very accurately.

Data Source :

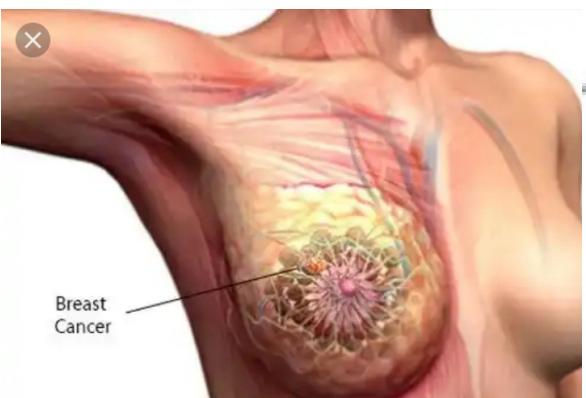
<https://www.kaggle.com/datasets/gunesevitan/breast-cancer-metabric>

Below is the link to the data manipulation code on github:

**Population(Female)**

**2509**

Count of Sex



Average Age of Patients

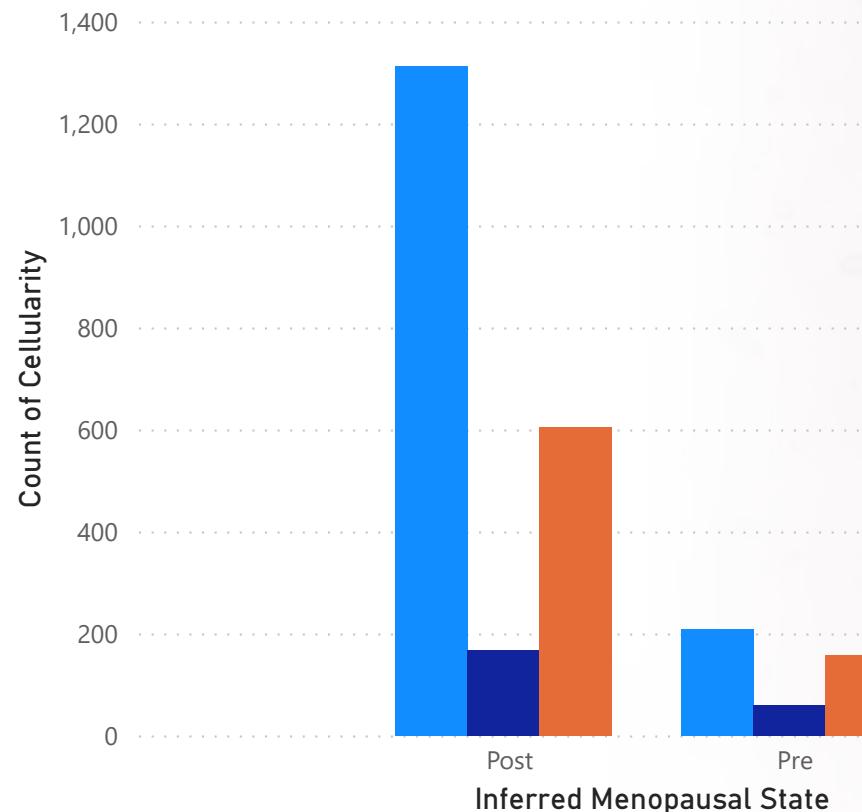
60.42

Age at Diagnosis

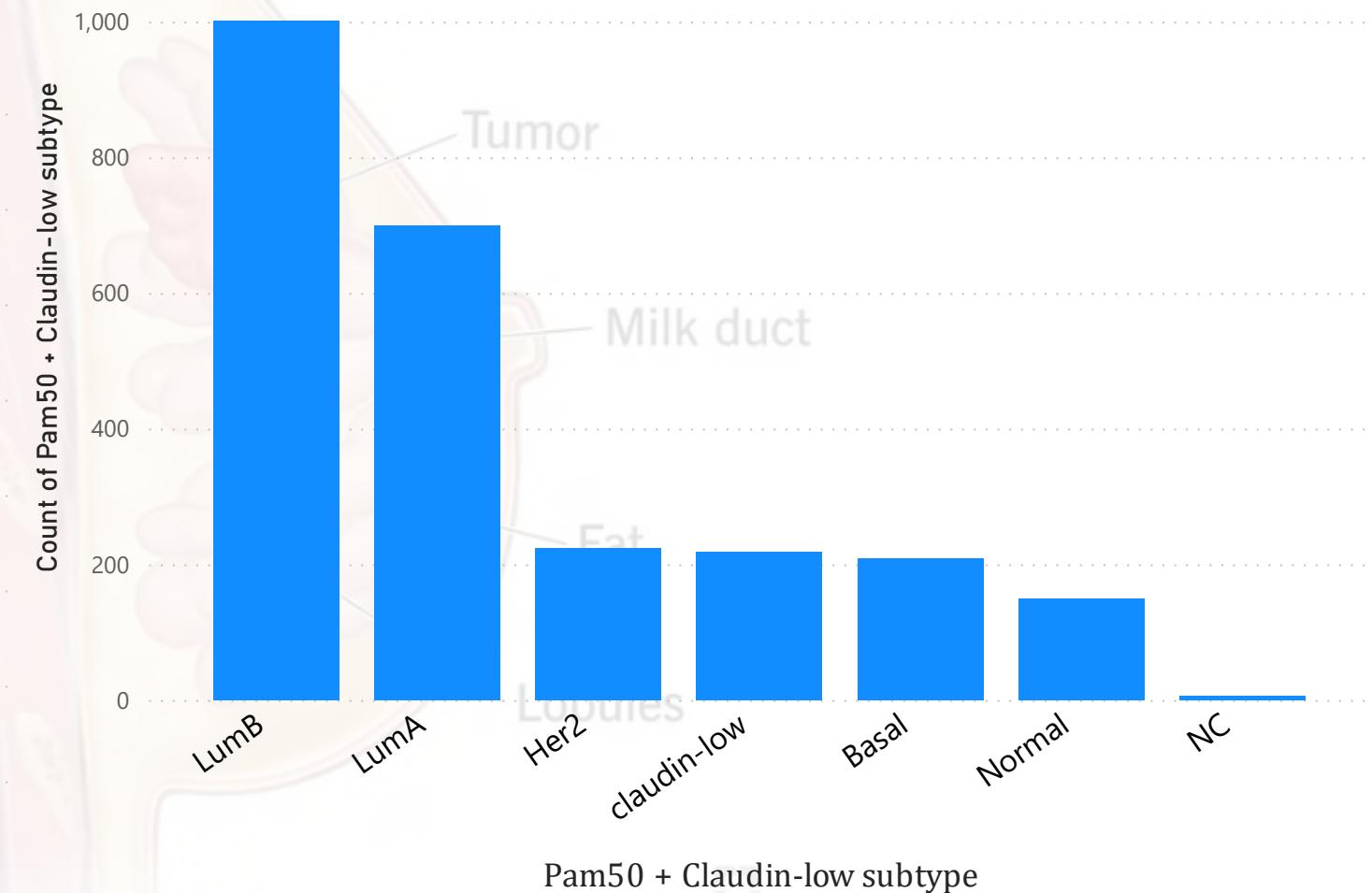
# General Demographic Informations and Clinical Status of Patients

Distribution of Cellularity by Inferred Menopausal State

Cellularity ● High ● Low ● Moderate



Distribution of Pam50 + Claudin-low subtype

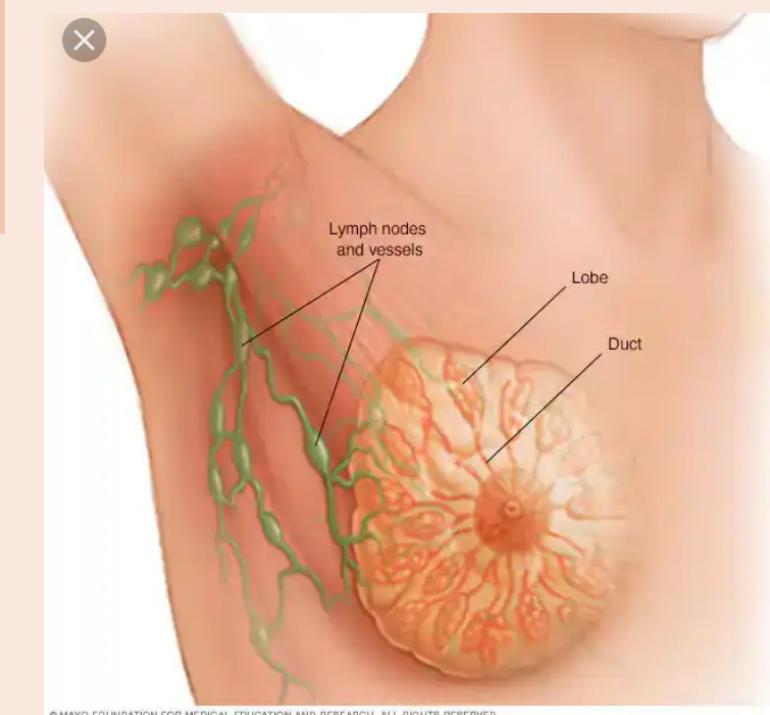
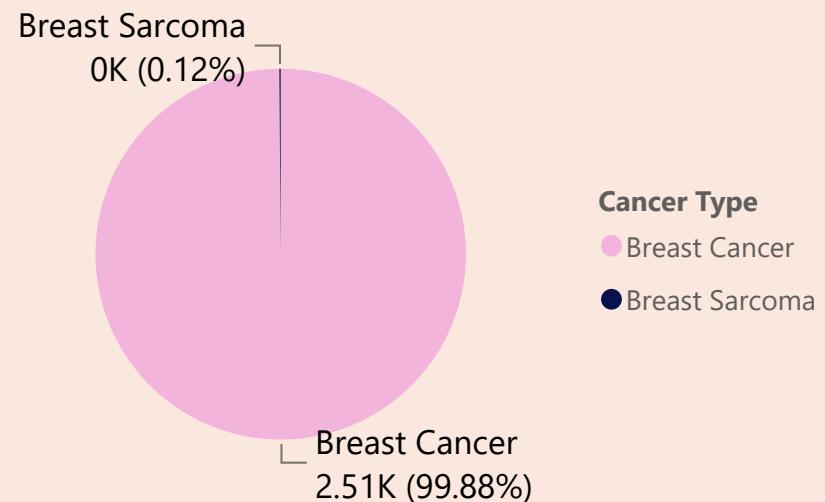


# Breast Cancer Informations of Patients and Survival Status

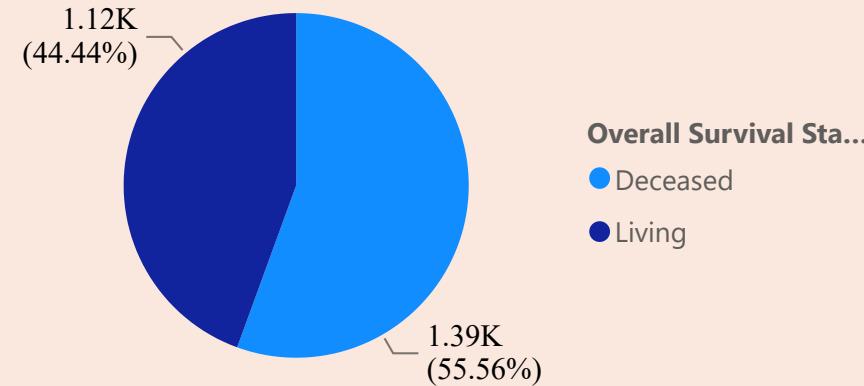
Type of Breast Surgery	Count of Type of Breast Surgery	%GT Count of
Breast Conserving	1319	
Mastectomy	1190	
<b>Total</b>	<b>2509</b>	

< >

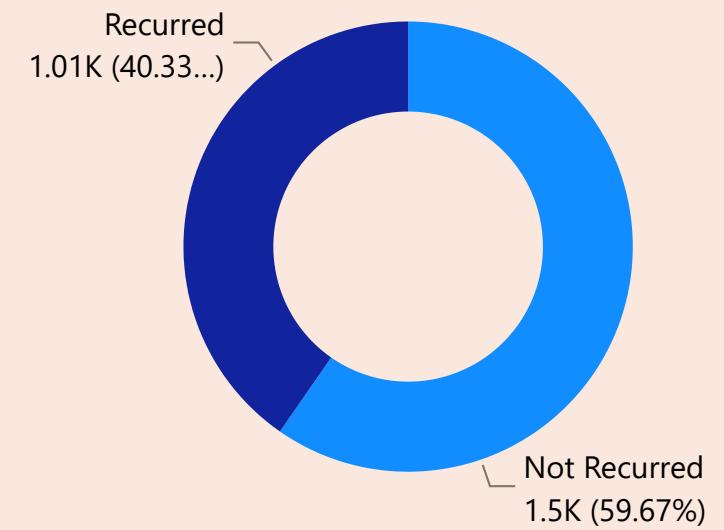
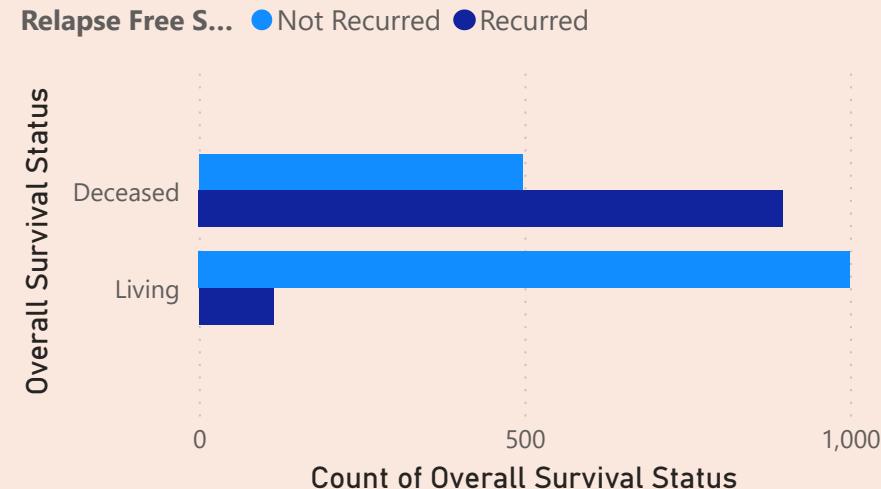
## Count of Cancer Type



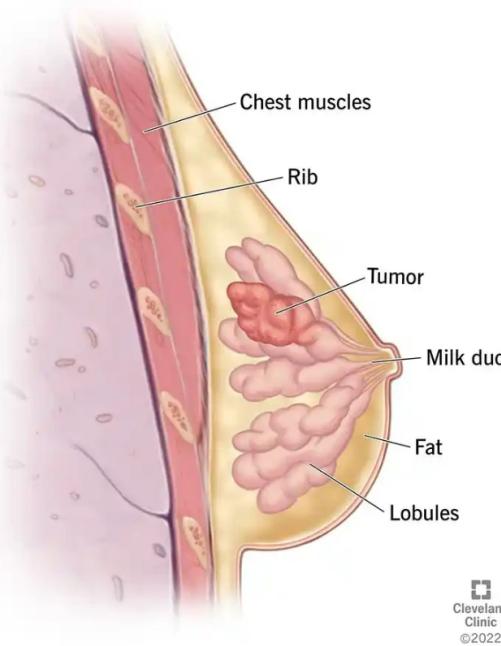
## Count of Overall Survival Status



## Count of Overall Survival Status



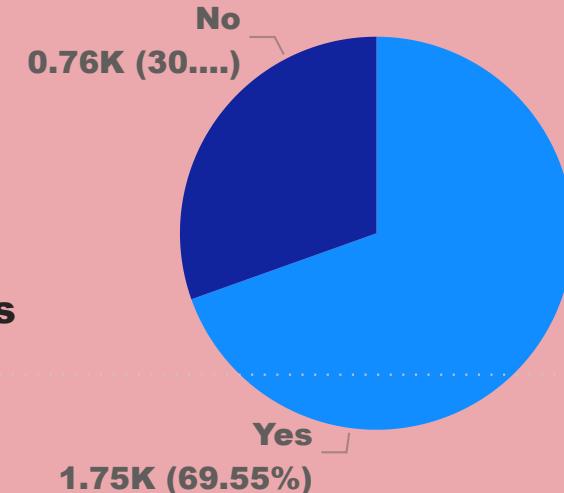
Breast cancer



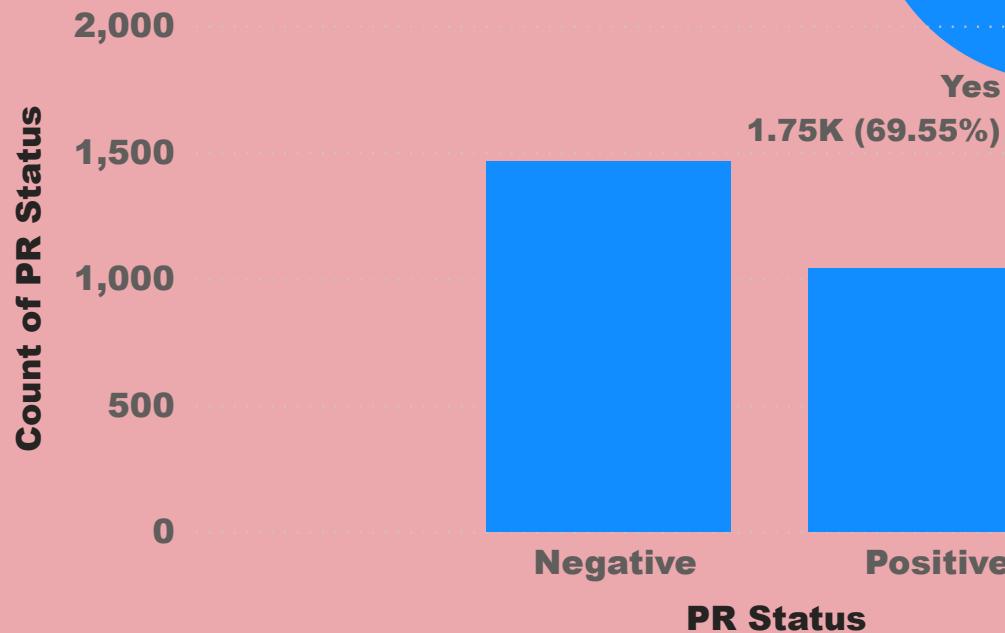
Cleveland Clinic ©2022

# Clinical Informations of Breast Cancer Patients

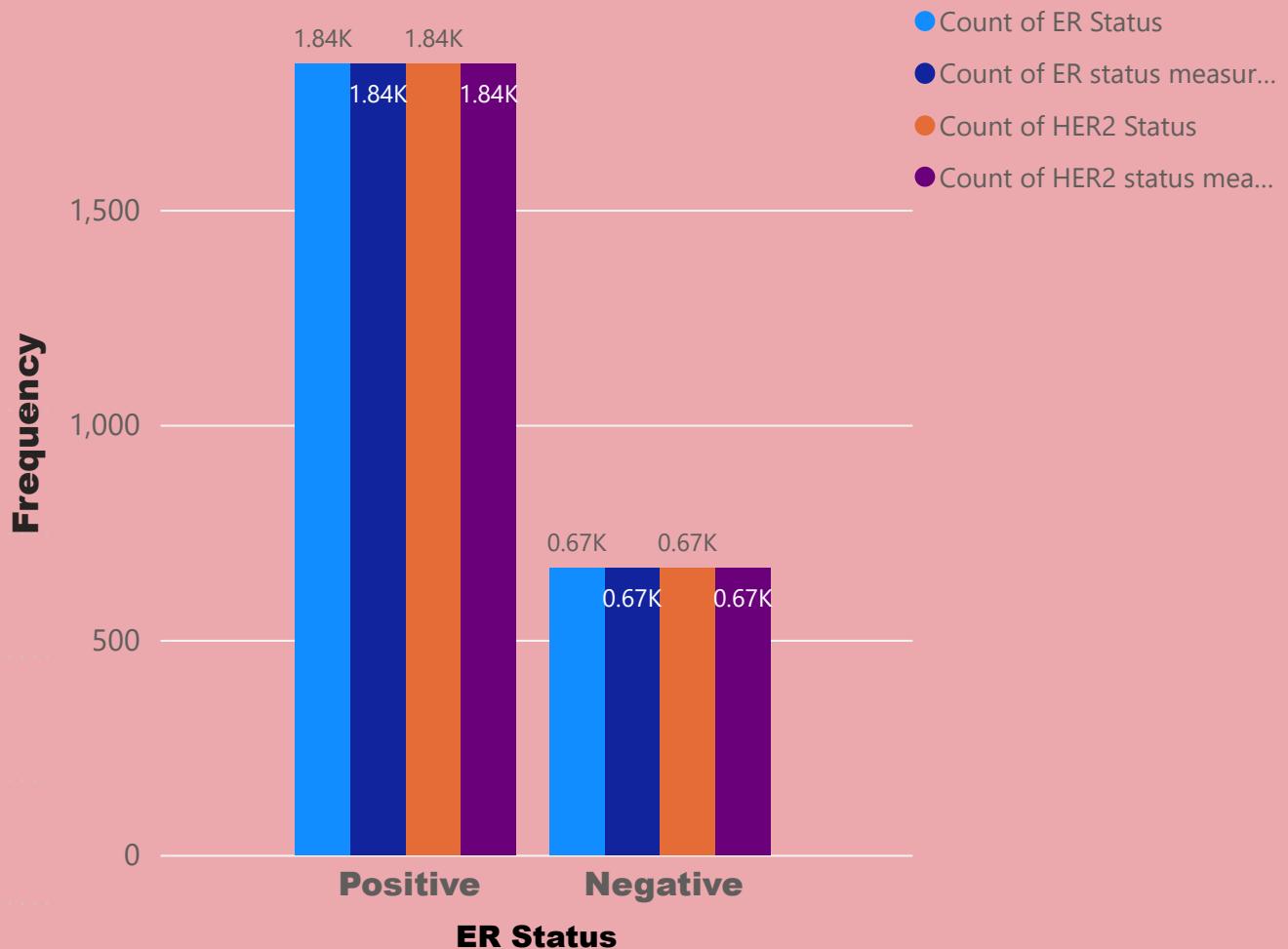
Count of Hormone Therapy by Hormone Therapy



Count of PR Status by PR Status

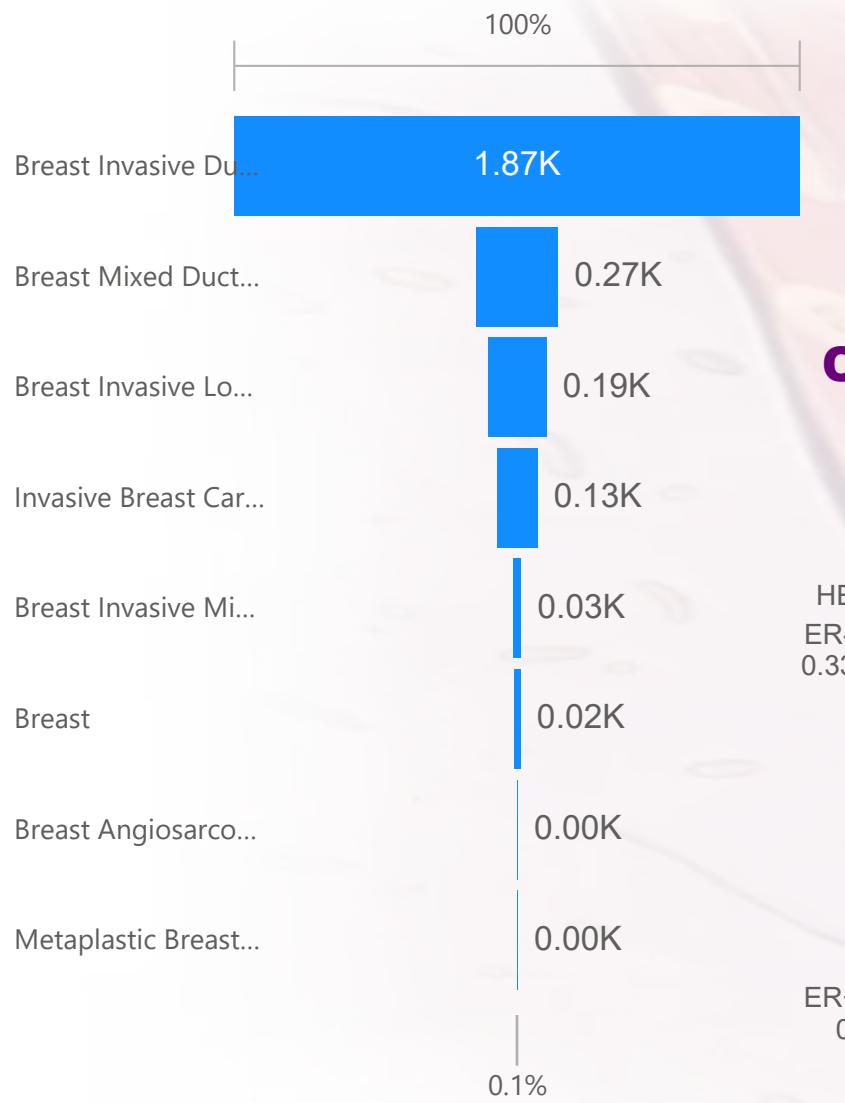


Distribution of ER Status by ER status and IHR

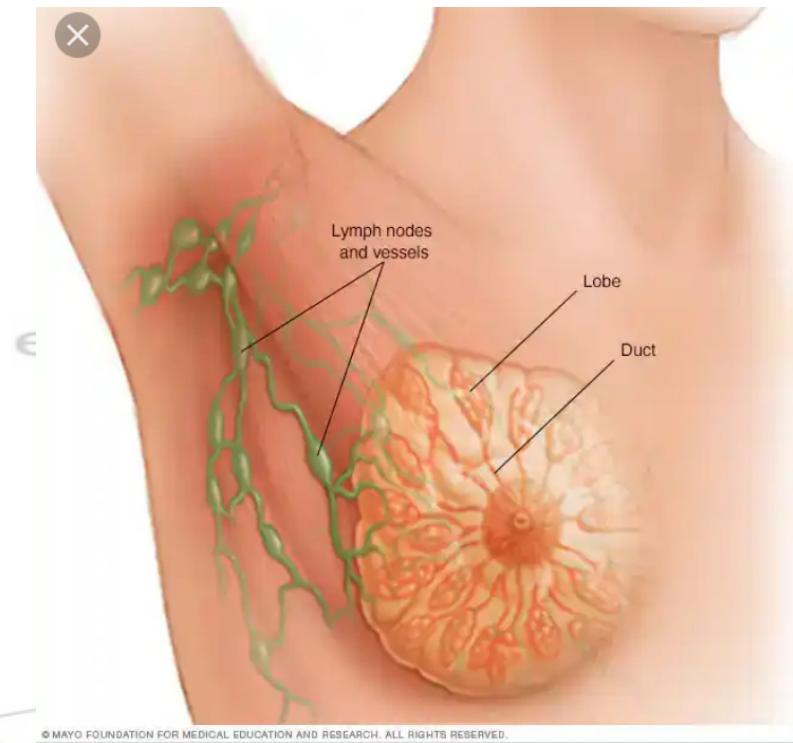
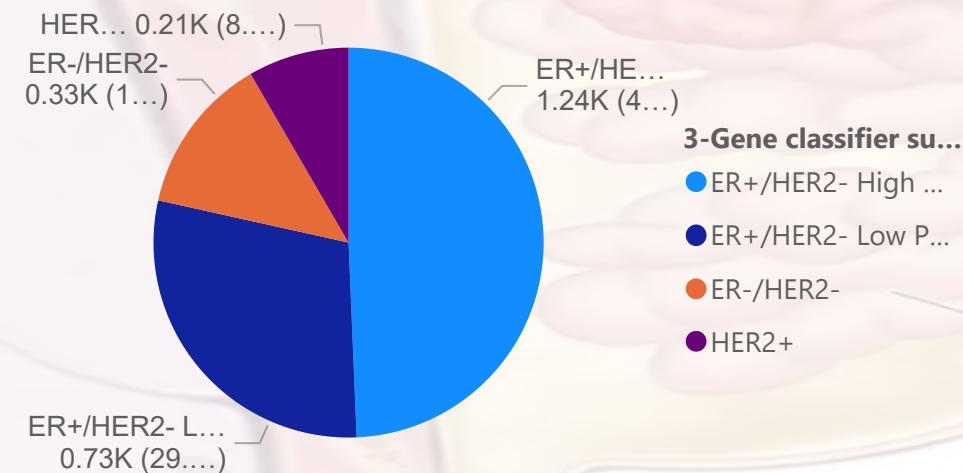


# Breast Cancer and Surgery Type

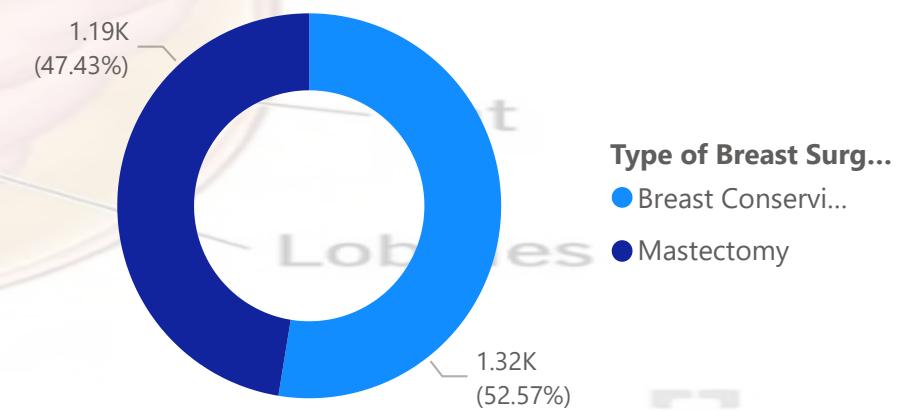
Count of Cancer Type Detailed



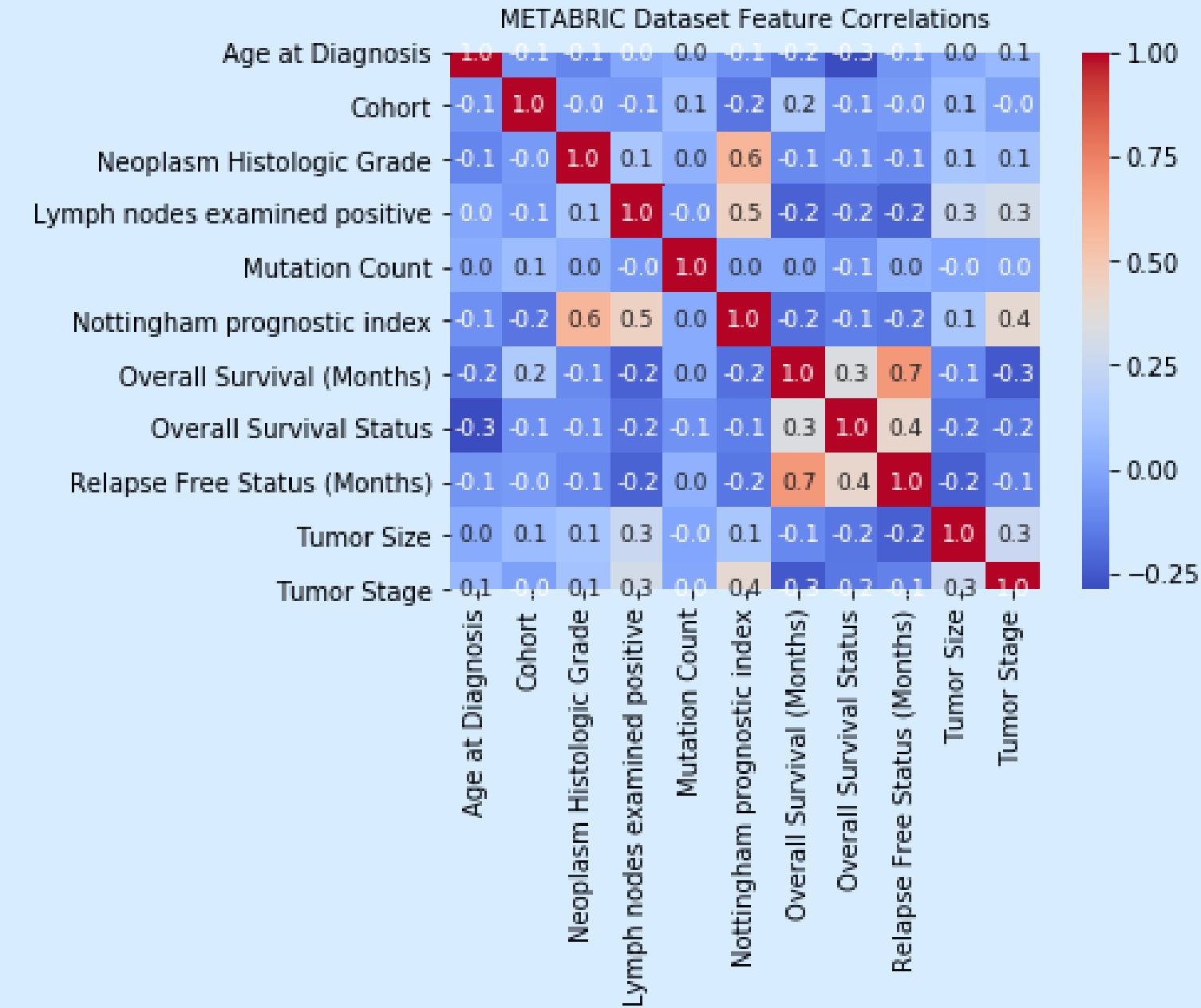
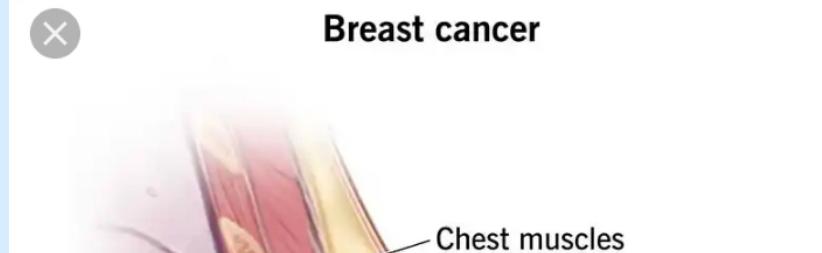
Count of 3-Gene classifier subtype



Count of Type of Breast Surgery



# Correlation Heat-Map for important Variables in the dataset



# Survival Probability Analysis Chart: Kaplan-Meler Estimator

Kaplan-Meier Estimators Survival Functions

