

# BREAST CANCER

## INFOGRAPHIC PREDICTIVE MODELS.

### DATASET

#### Data Source:

(2017 - 2021) update of the SEER Program of the NCI (<https://seer.cancer.gov/data/>)

### BACKGROUND

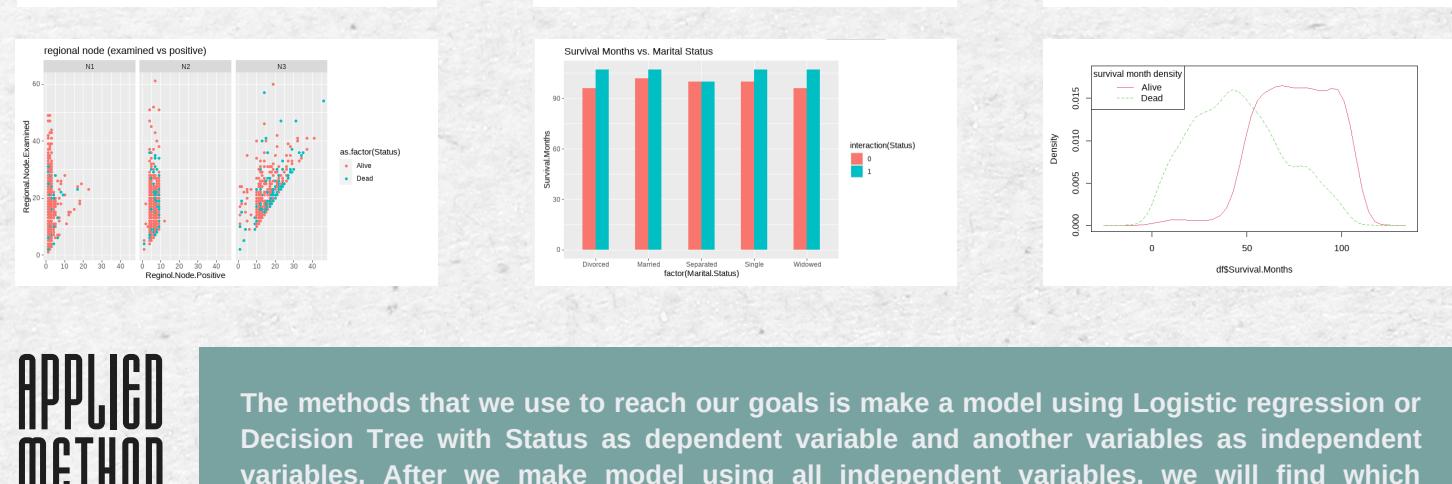
Breast cancer is a leading cause of death worldwide, claiming 685,000 lives annually. With 2.3 million new cases reported, early prevention and detection are crucial. Understanding the number of breast cancer survivors helps predict survival rates and guides treatment. Promoting awareness, regular screenings, and genetic counseling can reduce the likelihood of contracting breast cancer. Advances in research, accessible healthcare, and support networks are vital in combating this disease. By prioritizing early prevention efforts and improving prognosis, we aim to reduce breast cancer mortality rates and improve the quality of life for those impacted by this formidable illness.

### VARIABLE

16 VARIABLE

- T-stage
- N-stage
- X6th-stage
- Differentiate
- Grade
- A-stage
- Age
- Race
- Marital Status
- Tumor Size
- Regional Node (Examine)
- Regional Node (Positive)
- Survival Month
- Status
- Estrogen Status
- Progesterona Status

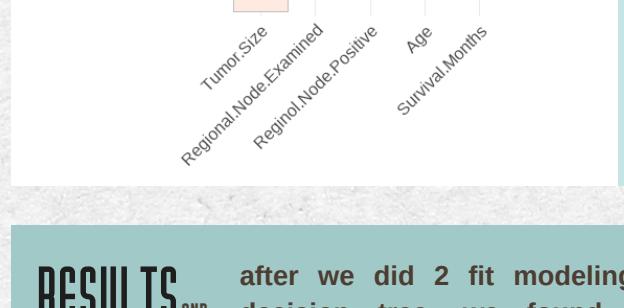
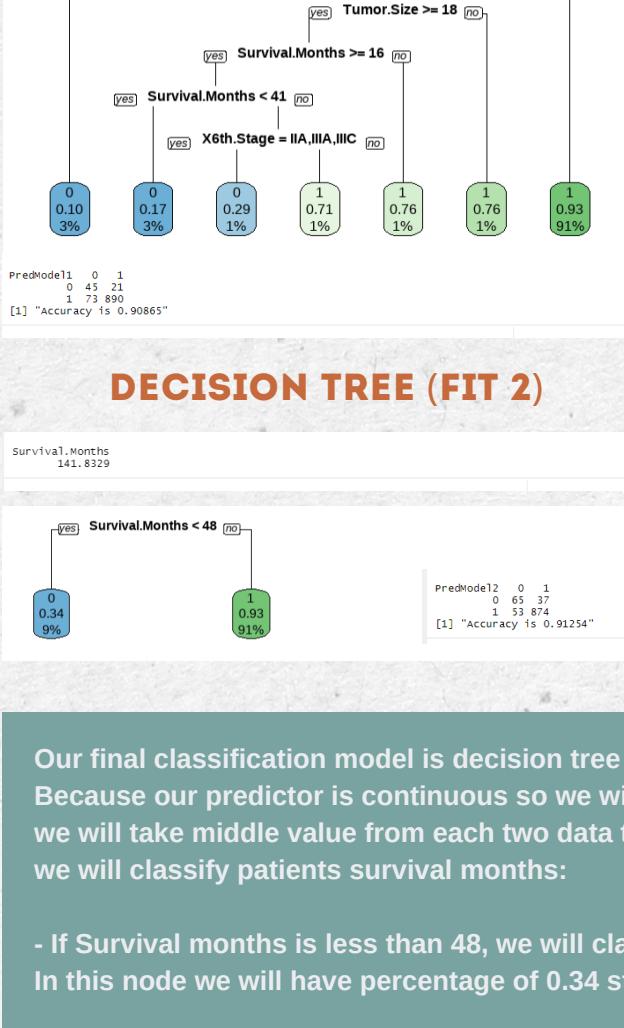
### EXPLORATORY DATA ANALYSIS (INSIGHT)



### APPLIED METHOD EXPLAIN

The methods that we use to reach our goals is make a model using Logistic regression or Decision Tree with Status as dependent variable and another variables as independent variables. After we make model using all independent variables, we will find which variables have a high correlation with Status then we will fit into our new model for better prediction

### DECISION TREE (FIT 1)

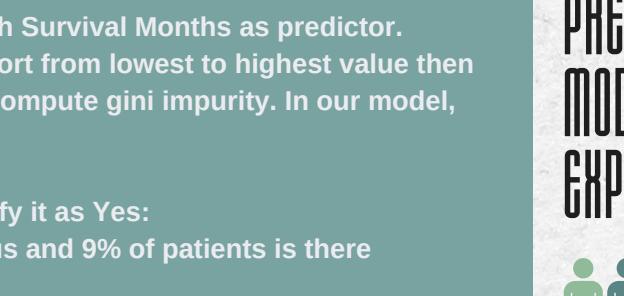


### CORRELATION HEATMAP

### RESULTS AND DISCUSSIONS

after we did 2 fit modeling on the decision tree, we found that the second model was more suitable to be used as our final model. In the second model (spesific predictor), we use the survival month variable as final predictor (as the variable of greatest importance), in which this model has greater results than model 1 (global predictor) with a result of 0.91254

### DECISION TREE (FIT 2)



### PREDICTIVE MODEL EXPLANATION

### PREDICTIVE MODEL CONCLUSION

Based on our analysis, there is very few independent variables that has a correlation with our dependent variables and on our plot above in exploratory data analysis there is some information that we can conclude from our plotting. Survival Months has highest relations with our dependent variable. So for our final model, we will use Decision Trees with Survival Months as our predictor to predict Status which has 91.2% accuracy which is pretty accurate.