**UTSA Data Analytics & Visualization Bootcamp**

Final Project Proposal:

|  |  |
| --- | --- |
|  |  |
| **Project Title:** | Breast Cancer Research Data |
|  |  |
| **Team members:** | Lorraine Gomez  Shwet ‘Sunny’ Bhatt |
|  |  |
| **Project Description:** | Our team will gather different datasets train and test models and find correlation with pre-existing conditions and various demographics such as Age, Gender, Air Pollution, geographic location that may increases the likelihood of any specific kinds of cancer. We will further explore breast cancer data based on diagnosis.  We will build a model using supervised machine learning to make prediction using the features and analyze the model as specified in the description.  Our team will test Logistic regression, random forest classifier or KNN algorithms. We will use accuracy as our classification method for metrics. |
|  |  |
| **Dataset(s) to be used:** | * <https://seer.cancer.gov/seerstat/> * <https://aiplanet.com/challenges/data-sprint-31-breast-cancer-detection/75/overview/about> * <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29> * <https://www.kaggle.com/datasets/reihanenamdari/breast-cancer> |
|  |  |
| **Rough Breakdown of Tasks:** | Topic idea generation  Documentation of project proposal  Perform ETL process to clean and load data into Database  Create ML model and use 2-3 different algorithms  Tableau to show visuals  Manage GitHub |