**UTSA Data Analytics & Visualization Bootcamp**

Final Project Proposal:

|  |  |
| --- | --- |
|  |  |
| **Project Title:** | Cancer Research Data  Finding correlation between different variable such as Age, Gender, Air Pollution, Alcohol use, Occupational Hazards, Genetic Risk, chronic Lung Disease, Balanced Diet, Obesity, Smoking, Passive Smoker |
|  |  |
| **Team members:** | Lorraine Gomez  Shwet ‘Sunny’ Bhatt |
|  |  |
| **Project Description:** | Our team will gather different datasets 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. |
|  |  |
| **Dataset(s) to be used:** | <https://www.kaggle.com/datasets/mukhazarahmad/worldwide-cancer-data>  <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>  <https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics> |
|  |  |
| **Rough Breakdown of Tasks:** | Topic idea generation  Documentation of project proposal  Perform ETL process to clean and load data into Database  Create ML model by using Scikit Learn  Tableau to show visuals  Manage GitHub  Create HTML using JavaScript |