

Abdulaziz Almuhaidib

T5 Tuwaiq DS Bootcamp

30 September 2021

Project Proposal 1

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

The project determines whether the patient has survived respiratory cancer.
important to help the healthcare industry to make more accurate decisions.

- Who benefits from exploring this question or building this model/system?
Health industry (Public Health Authority).

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

SEER dataset 1973-2014, RESPIR data. I have submitted a request to NIH and obtained the dataset.

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

Attribute	Description
Disease	Name of the disease in the ICD-O.
Age	Age at diagnosis.
Year at Diagnosis	Year at diagnosis.
Gender	Gender.
Grade	Category based on the appearance of tumor.
Marital Status	Marital status at diagnosis.
Primary Site	Site in which the primary tumor originated.

Race	Race
Regional Nodes Examined	Number of regional lymph nodes removed and examined.
Regional Nodes Positive	Number of regional lymph nodes that contained metastases.
Summary Stage	Defined according to the spread of cancer.
Tumor size	Size in mm.
Survivability	Whether the patient has survived the cancer or not.

- If modeling, what will you predict as your target?

Survivability

Tools:

- How do you intend to meet the tools requirement of the project?

Python libraries: r , glob, NumPy, Pandas, Matplotlib, and the rest of libraries we covered in the bootcamp.

- Are you planning in advance to need or use additional tools beyond those required?

May be Public Tableau.

MVP Goal:

- What would a [minimum viable product \(MVP\)](#) look like for this project?

Cleaned Dataset and some visualization.