# 

# Topic: Survival Analytics

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name:** Nithin Dsouza

**Batch Id:**  05012021-10AM

**Topic: Survival Analytics**

1. **Business Problem**
   1. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.**

**Using R and Python codes perform:**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building**

**4.1Build the model on the scaled data (try multiple options)**

**4.2Perform Survival analytics on the given datasets.**

**4.3Briefly explain the model output in the documentation.**

1. **Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.**

# Note:

The assignment should be submitted in the following format:

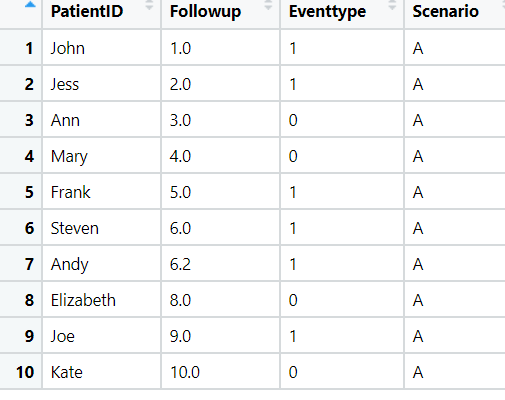
* R code
* Python code
* Code Modularization should be maintained
* Documentation of the model building (elaborating on steps mentioned above)

**Business problem**: the following dataset contains PatientID Follow up Event type

Scenarios. survival analysis model has to be built on the patient’s ID

\***important note**: -R and python code to be done

1. **Business Problem**
   1. **Objective:** Maximize the accuracy in estimating time to event for a group of individuals.
   2. **Constraints (if any):** Kaplan meier estimate cannot be used for multivariate analysis.



|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| PatientID | Name of patient | Categorical | Relevant, Provides useful information. |
| Followup | No of follow up | Continuous, Ratio | Relevant, Provides useful information. |
| Eventtype | type of event | Discrete, count | Relevant, Provides useful information. |
| Scenario | Scenerio fo patient | Categorical | Relevant, Provides useful information. |

**Problem Statement: -**

ECG of different age group of people has been recorded, the survival time in hours after operation and the event(death) occurred is denoted by 1 and 0 represent still alive. Perform survival analysis on the dataset given below and provide your insights in the documentation.

A large room

Description automatically generated

1. **Business Problem**
   1. **Objective:** Maximize the accuracy in estimating time to event for a group of individuals.
   2. **Constraints (if any):** Kaplan meier estimate cannot be used for multivariate analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| survival\_time\_hr | survival time per hour | Continuous, Interval | Relevant, Provides useful information. |
| alive | No of patient alive | Categorical,binary | Relevant, Provides useful information. |
| age | age of patient | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| pericardialeffusion | pericardialeffusion of patient | Categorical,binary | Irrelevant, Doesn’t Provides useful information. |
| fractionalshortening | fractionalshortening | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| epss | epss value | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| lvdd | lvdd value | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| wallmotion-score | wallmotion-score of patient | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| wallmotion-index | wallmotion-index of patient | Continuous, Ratio | Irrelevant, Doesn’t Provides useful information. |
| multi\_sensor | multi\_sensor value | Continuous, Interval | Irrelevant, Doesn’t Provides useful information. |
| name | Name of patient | Categorical | Irrelevant, Doesn’t Provides useful information. |
| group | group that which person belong to | Continuous, Ratio | Relevant, Provides useful information. |