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: Nukala Ayyappa Bharthwaj

Batch Id: DSWDMCOS 21012022

Topic: Survival Analytics

1. Business Problem
    1.1. Objective
    1.2. Constraints (if any)

2. Work on each feature of the dataset to create a data dictionary as displayed in the below image:

Name of Feature	Description	Type	Relevance
ID	Customer ID	Quantitative, Nominal	Irrelevant, ID does not provide useful information

    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:

3. Exploratory Data Analysis (EDA):
    3.1. Summary
    3.2. Univariate analysis
    3.3. Bivariate analysis

4. Model Building
    4.1 Build the model on the scaled data (try multiple options)
    4.2 Perform Survival analytics on the given datasets.

    4.3 Briefly explain the model output in the documentation.

5. 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
Sol:
Business Objective: To apply survival analysis on the patients data set.

Survival analysis: survival analysis is made for the given data set by taking the time variables as followup and event variable as event type.
Keplan-Meier Survival Plot for the given data set by taking the event type as 1 and out put as patient as follows.
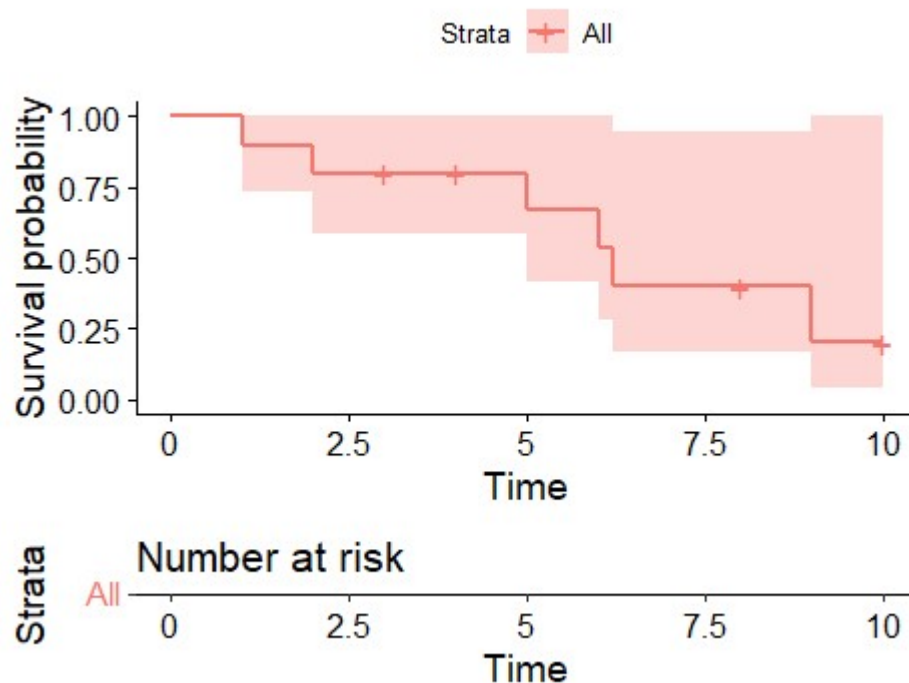
|	PatientID	Followup	Eventtype	Scenario
1	John	1.0	1	A
2	Jess	2.0	1	A
3	Ann	3.0	0	A
4	Mary	4.0	0	A
5	Frank	5.0	1	A
6	Steven	6.0	1	A
7	Andy	6.2	1	A
8	Elizabeth	8.0	0	A
9	Joe	9.0	1	A
10	Kate	10.0	0	A

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.
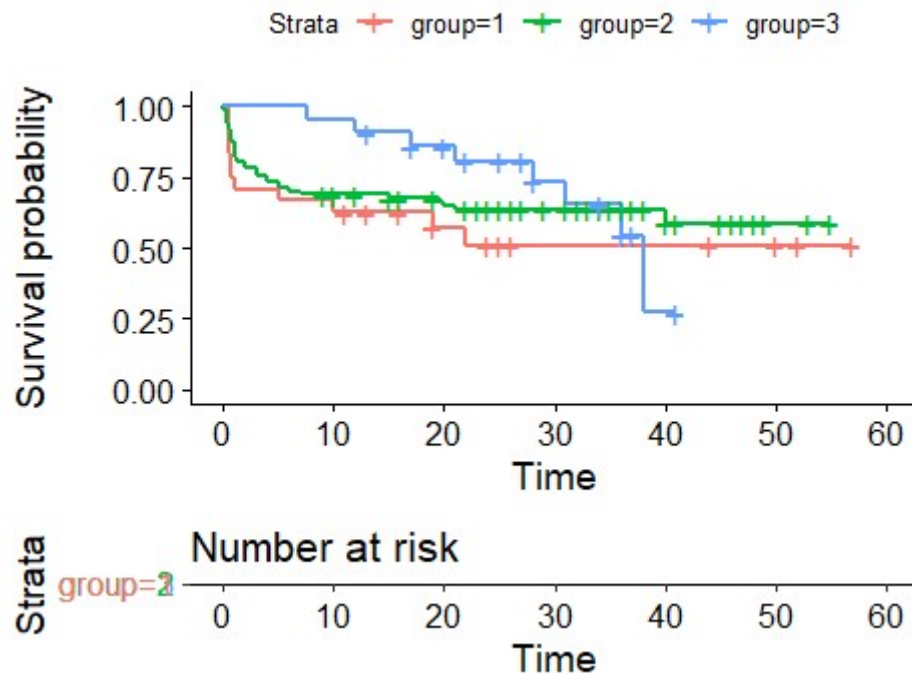
Sol:
Business Objective: To apply survival analysis on the ECG survival data set.

Survival analysis: survival analysis is made for the given data set by taking the time variables as survival time hr and event variable as alive.
Keplan-Meier Survival Plot for the given data set by taking the alive as 1 and out put as groups as follows.
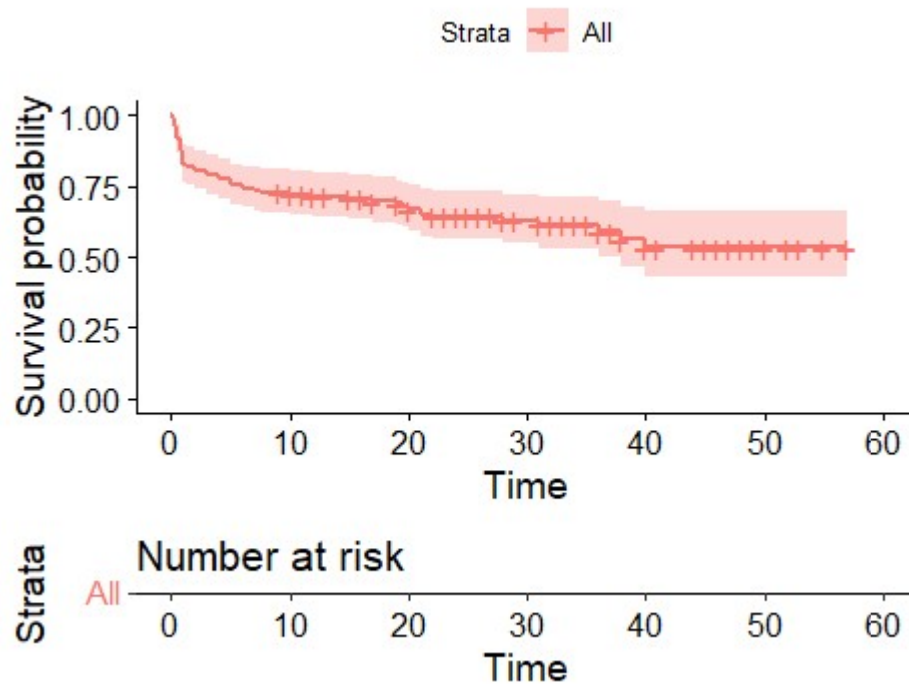
survival_time_hr	alive	age	pericardialeffusion	actionalshortenin	epss	lvdd	wallmotion-score	wallmotion-index	multi_sensor	name	group
11	0	71	0	0.26	9	4.6	14	1	1	name	1
19	0	72	0	0.38	6	4.1	14	1.7	0.588	name	1
16	0	55	0	0.26	4	3.42	14	1	1	name	1
57	0	60	0	0.253	12.062	4.603	16	1.45	0.788	name	1
19	1	57	0	0.16	22	5.75	18	2.25	0.571	name	1
26	0	68	0	0.26	5	4.31	12	1	0.857	name	1
13	0	62	0	0.23	31	5.43	22.5	1.875	0.857	name	1
50	0	60	0	0.33	8	5.25	14	1	1	name	1
19	0	46	0	0.34	0	5.09	16	1.14	1.003	name	1
25	0	54	0	0.14	13	4.49	15.5	1.19	0.93	name	1
10	1	77	0	0.13	16	4.23	18	1.8	0.714	name	1
52	0	62	1	0.45	9	3.6	16	1.14	1.003	name	1
52	0	73	0	0.33	6	4	14	1	1	name	1
44	0	60	0	0.15	10	3.73	14	1	1	name	1
0.5	1	62	0	0.12	23	5.8	11.67	2.33	0.358	name	1
24	0	55	1	0.25	12.063	4.29	14	1	1	name	1
0.5	1	69	1	0.26	11	4.65	18	1.64	0.784	name	1
0.5	1	70	1	0.07	20	5.2	24	2	0.857	name	1
22	1	66	0	0.09	17	5.819	8	1.333	0.429	name	1
1	1	66	1	0.22	15	5.4	27	2.25	0.857	name	1
0.75	1	69	0	0.15	12	5.39	19.5	1.625	0.857	name	1
0.75	1	85	1	0.18	19	5.46	13.83	1.38	0.71	name	1