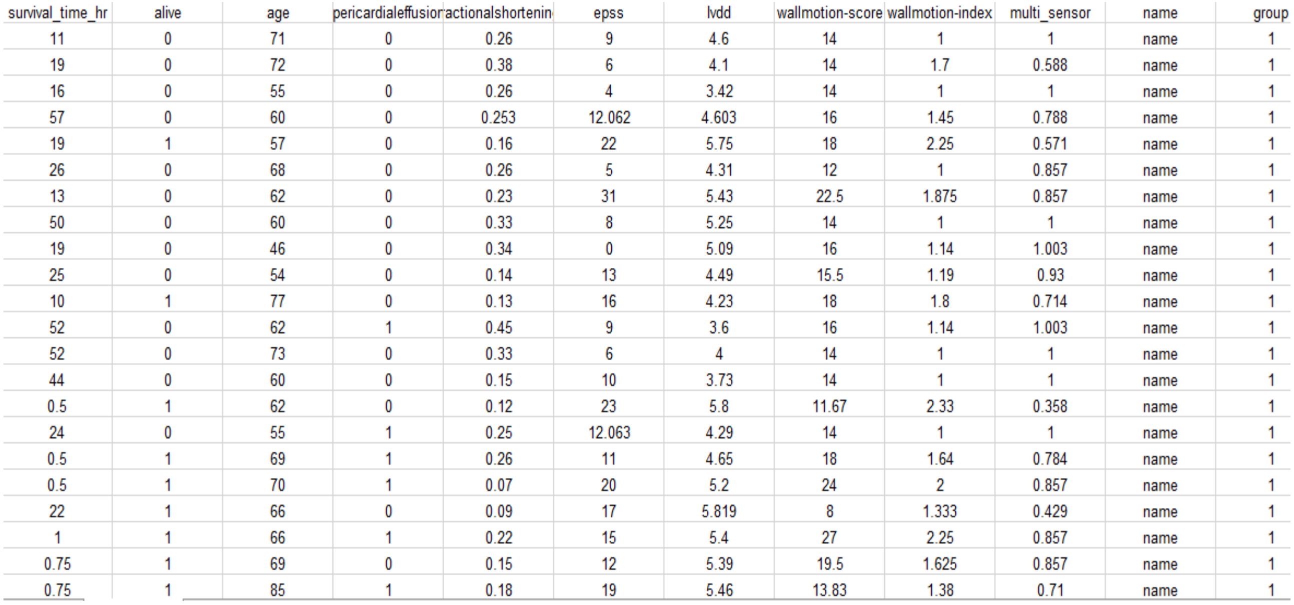
Topic: Survival Analytics

Name: ULLI VENKATA SAI KUMAR
Batch Id: 04072024HYD10AM
Topic: Survival Analytics

Problem Statement:

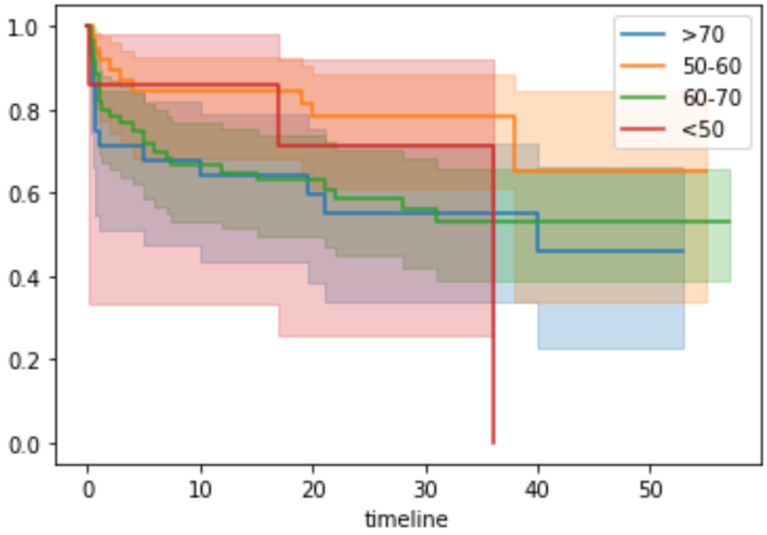
ECG of different age groups of people has been recorded. The survival time in hours after the operation is given and the event type is denoted by 1 (if dead) and 0 (if alive). Perform survival analysis on the dataset given below and provide your insights in the documentation.



survival_time_hr	alive	age	pericardialeffusior	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

```
import pandas as pd

pip install lifelines

import matplotlib.pyplot as plt

from lifelines import KaplanMeierFitter


# Load the dataset

data = pd.read_excel('ECG_Surv.xlsx')


# Show the first few rows of the dataset to understand its structure

data.head()
```

```
# Remove the 'name' column as it's not relevant for the analysis

data_cleaned = data.drop(columns=['name'])


# Create age groups (e.g., below 50, 50-60, 60-70, above 70)

bins = [0, 50, 60, 70, 100]

labels = ['<50', '50-60', '60-70', '>70']

data_cleaned['age_group'] = pd.cut(data_cleaned['age'], bins=bins, labels=labels, right=False)


# Check the processed data

data_cleaned.head()


# Kaplan-Meier estimator

kmf = KaplanMeierFitter()


# Plotting survival functions for different age groups

plt.figure(figsize=(10, 6))
```



```
for group in data_cleaned['age_group'].unique():

    group_data = data_cleaned[data_cleaned['age_group'] == group]

    kmf.fit(group_data['survival_time_hr'], event_observed=group_data['alive'], label=group)

    kmf.plot_survival_function()


plt.title('Kaplan-Meier Survival Curves by Age Group')

plt.xlabel('Survival Time (hours)')

plt.ylabel('Survival Probability')

plt.grid(True)

# Ensure legend shows the correct age groups

plt.legend(title="<50")
plt.show()
```