

Project -1

Heart failure Analysis

Project 1: Heart Failure Analysis

1. Load the file

```
[95] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
dataset=pd.read_csv("/content/archive(7).zip")
```

2. Print first 5 rows of data

```
[96] dataset.head()
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7	1
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8	1

Next steps: [Generate code with dataset](#) [View recommended plots](#) [New interactive sheet](#)

3. Print last 5 rows of data

```
[97] dataset.tail()
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
294	62.0	0	61	1	38	1	155000.0	1.1	143	1	1	270	0
295	55.0	0	1820	0	38	0	270000.0	1.2	139	0	0	271	0
296	45.0	0	2060	1	60	0	742000.0	0.8	138	0	0	278	0
297	45.0	0	2413	0	38	0	140000.0	1.4	140	1	1	280	0
298	50.0	0	196	0	45	0	395000.0	1.6	136	1	1	285	0

4. Basic cleaning of data for checking null values, missing values

```
# Check for null values
print(dataset.isnull().sum())
```

```
age                0
anaemia            0
creatinine_phosphokinase  0
diabetes           0
ejection_fraction  0
high_blood_pressure  0
platelets          0
serum_creatinine   0
serum_sodium       0
sex               0
smoking           0
time              0
DEATH_EVENT       0
dtype: int64
```

```
[99] #Check for missing values
dataset.isnull()
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
...
294	False	False	False	False	False	False	False	False	False	False	False	False	False
295	False	False	False	False	False	False	False	False	False	False	False	False	False
296	False	False	False	False	False	False	False	False	False	False	False	False	False
297	False	False	False	False	False	False	False	False	False	False	False	False	False
298	False	False	False	False	False	False	False	False	False	False	False	False	False

299 rows × 13 columns

Project -1

Heart failure Analysis

5. Info of the data

```
# Anaemia: 0-> NO || 1-> Yes
```

```
dataset['anaemia'] = dataset['anaemia'].replace({0: 'NO', 1: 'Yes'})  
print(dataset.head())
```

```
age  anaemia  creatinine_phosphokinase  diabetes  ejection_fraction  \  
0    75.0      NO                    582         NO                20  
1    55.0      NO                   7861         NO                38  
2    65.0      NO                    146         NO                20  
3    50.0      Yes                    111         NO                20  
4    65.0      Yes                    160         Yes                20  
  
high_blood_pressure  platelets  serum_creatinine  serum_sodium  sex  \  
0                Yes  265000.00                1.9            130  Male  
1                NO  263358.03                1.1            136  Male  
2                NO  162000.00                1.3            129  Male  
3                NO  210000.00                1.9            137  Male  
4                NO  327000.00                2.7            116  Female  
  
smoking  DEATH_EVENT  
0      No           Yes  
1      No           Yes  
2     Yes           Yes  
3      No           Yes  
4      No           Yes
```

```
# Diabetes: 0-> No || 1-> Yes
```

```
dataset['diabetes'] = dataset['diabetes'].replace({0: 'NO', 1: 'Yes'})  
print(dataset.head())
```

```
age  anaemia  creatinine_phosphokinase  diabetes  ejection_fraction  \  
0    75.0      NO                    582         NO                20  
1    55.0      NO                   7861         NO                38  
2    65.0      NO                    146         NO                20  
3    50.0      Yes                    111         NO                20  
4    65.0      Yes                    160         Yes                20  
  
high_blood_pressure  platelets  serum_creatinine  serum_sodium  sex  \  
0                Yes  265000.00                1.9            130  Male  
1                NO  263358.03                1.1            136  Male  
2                NO  162000.00                1.3            129  Male  
3                NO  210000.00                1.9            137  Male  
4                NO  327000.00                2.7            116  Female  
  
smoking  DEATH_EVENT  
0      No           Yes  
1      No           Yes  
2     Yes           Yes  
3      No           Yes  
4      No           Yes
```

```
# High blood pressure: 0-> No || 1-> Yes
```

```
dataset['high_blood_pressure'] = dataset['high_blood_pressure'].replace({0: 'NO', 1: 'Yes'})  
print(dataset.head())
```

```
age  anaemia  creatinine_phosphokinase  diabetes  ejection_fraction  \  
0    75.0      NO                    582         NO                20  
1    55.0      NO                   7861         NO                38  
2    65.0      NO                    146         NO                20  
3    50.0      Yes                    111         NO                20  
4    65.0      Yes                    160         Yes                20  
  
high_blood_pressure  platelets  serum_creatinine  serum_sodium  sex  \  
0                Yes  265000.00                1.9            130  Male  
1                NO  263358.03                1.1            136  Male  
2                NO  162000.00                1.3            129  Male  
3                NO  210000.00                1.9            137  Male  
4                NO  327000.00                2.7            116  Female  
  
smoking  DEATH_EVENT  
0      No           Yes  
1      No           Yes  
2     Yes           Yes  
3      No           Yes  
4      No           Yes
```

Project -1

Heart failure Analysis

```
[121] # Sex: 0-> Female || 1-> Male
dataset['sex'] = dataset['sex'].replace({0: 'Female', 1: 'Male'})
print(dataset.head())
```

```
age anaemia creatinine_phosphokinase diabetes ejection_fraction \
0 75.0 NO 582 NO 20
1 55.0 NO 7861 NO 38
2 65.0 NO 146 NO 20
3 50.0 Yes 111 NO 20
4 65.0 Yes 160 Yes 20

high_blood_pressure platelets serum_creatinine serum_sodium sex \
0 Yes 265000.00 1.9 130 Male
1 NO 263358.03 1.1 136 Male
2 NO 162000.00 1.3 129 Male
3 NO 210000.00 1.9 137 Male
4 NO 327000.00 2.7 116 Female

smoking DEATH_EVENT
0 No Yes
1 No Yes
2 Yes Yes
3 No Yes
4 No Yes
```

```
# Smoking: 0-> No || 1-> Yes
dataset['smoking'] = dataset['smoking'].replace({0: 'No', 1: 'Yes'})
print(dataset.head())
```

```
age anaemia creatinine_phosphokinase diabetes ejection_fraction \
0 75.0 NO 582 NO 20
1 55.0 NO 7861 NO 38
2 65.0 NO 146 NO 20
3 50.0 Yes 111 NO 20
4 65.0 Yes 160 Yes 20

high_blood_pressure platelets serum_creatinine serum_sodium sex \
0 Yes 265000.00 1.9 130 Male
1 NO 263358.03 1.1 136 Male
2 NO 162000.00 1.3 129 Male
3 NO 210000.00 1.9 137 Male
4 NO 327000.00 2.7 116 Female

smoking DEATH_EVENT
0 No Yes
1 No Yes
2 Yes Yes
3 No Yes
4 No Yes
```

```
# Death event: 0-> No || 1-> Yes
```

```
dataset['DEATH_EVENT'] = dataset['DEATH_EVENT'].replace({0: 'No', 1: 'Yes'})
print(dataset.head())
```

```
age anaemia creatinine_phosphokinase diabetes ejection_fraction \
0 75.0 NO 582 NO 20
1 55.0 NO 7861 NO 38
2 65.0 NO 146 NO 20
3 50.0 Yes 111 NO 20
4 65.0 Yes 160 Yes 20

high_blood_pressure platelets serum_creatinine serum_sodium sex \
0 Yes 265000.00 1.9 130 Male
1 NO 263358.03 1.1 136 Male
2 NO 162000.00 1.3 129 Male
3 NO 210000.00 1.9 137 Male
4 NO 327000.00 2.7 116 Female

smoking DEATH_EVENT
0 No Yes
1 No Yes
2 Yes Yes
3 No Yes
4 No Yes
```

Project -1

Heart failure Analysis

6. Some info on the dataset

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype  
---  -
0   age                                   299 non-null   float64
1   anaemia                              299 non-null   object  
2   creatinine_phosphokinase             299 non-null   int64   
3   diabetes                             299 non-null   object  
4   ejection_fraction                   299 non-null   int64   
5   high_blood_pressure                 299 non-null   object  
6   platelets                            299 non-null   float64  
7   serum_creatinine                     299 non-null   float64  
8   serum_sodium                        299 non-null   int64   
9   sex                                  299 non-null   object  
10  smoking                              299 non-null   object  
11  DEATH_EVENT                          299 non-null   object  
dtypes: float64(3), int64(3), object(6)
memory usage: 28.2+ KB
```

7. Remove un needed data- time column

```
if 'time' in dataset.columns:
    dataset = dataset.drop('time', axis=1)
print(dataset.head())
```

```
age anaemia creatinine_phosphokinase diabetes ejection_fraction \
0  75.0      NO                      582      NO                20
1  55.0      NO                      7861     NO                38
2  65.0      NO                      146      NO                20
3  50.0     Yes                      111      NO                20
4  65.0     Yes                      160      Yes                20

high_blood_pressure platelets serum_creatinine serum_sodium sex \
0      Yes  265000.00          1.9          130   Male
1      NO  263358.03          1.1          136   Male
2      NO  162000.00          1.3          129   Male
3      NO  210000.00          1.9          137   Male
4      NO  327000.00          2.7          116  Female

smoking DEATH_EVENT
0      No          Yes
1      No          Yes
2     Yes          Yes
3      No          Yes
4      No          Yes
```

8. Description of the data

```
dataset.describe()
```

	age	creatinine_phosphokinase	ejection_fraction	platelets	serum_creatinine	serum_sodium
count	299.000000	299.000000	299.000000	299.000000	299.000000	299.000000
mean	60.833893	581.839465	38.083612	263358.029264	1.39388	136.625418
std	11.894809	970.287881	11.834841	97804.236869	1.03451	4.412477
min	40.000000	23.000000	14.000000	25100.000000	0.50000	113.000000
25%	51.000000	116.500000	30.000000	212500.000000	0.90000	134.000000
50%	60.000000	250.000000	38.000000	262000.000000	1.10000	137.000000
75%	70.000000	582.000000	45.000000	303500.000000	1.40000	140.000000
max	95.000000	7861.000000	80.000000	850000.000000	9.40000	148.000000

Project -1

Heart failure Analysis

9. Shape of the data

```
[127] dataset.shape
```

```
(299, 12)
```

10. Find how many gender, high blood pressure, diabetes, smoking, death event records are there(value counts)

```
# Gender value counts
print("Gender Value Counts:\n", dataset['sex'].value_counts())

# High blood pressure value counts
print("\nHigh Blood Pressure Value Counts:\n", dataset['high_blood_pressure'].value_counts())

# Diabetes value counts
print("\nDiabetes Value Counts:\n", dataset['diabetes'].value_counts())

# Smoking value counts
print("\nSmoking Value Counts:\n", dataset['smoking'].value_counts())

# Death event value counts
print("\nDeath Event Value Counts:\n", dataset['DEATH_EVENT'].value_counts())
```

```
Gender Value Counts:
sex
Male      194
Female    105
Name: count, dtype: int64

High Blood Pressure Value Counts:
high_blood_pressure
NO      194
Yes     105
Name: count, dtype: int64

Diabetes Value Counts:
diabetes
NO      174
Yes     125
Name: count, dtype: int64

Smoking Value Counts:
smoking
No      203
Yes      96
Name: count, dtype: int64

Death Event Value Counts:
DEATH_EVENT
No      203
Yes      96
Name: count, dtype: int64
```

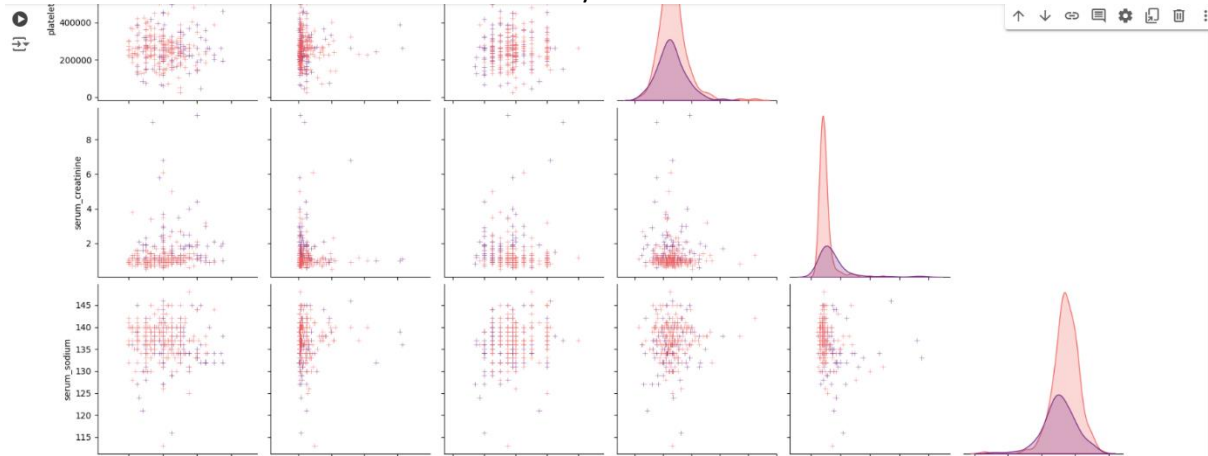
1. Show the relationship of the whole dataset(with relation to death event) using pairplot

```
sns.pairplot(dataset, hue='DEATH_EVENT', palette='magma', diag_kind='kde', markers='+', height=3, aspect=1, corner=True, plot_kws={'alpha': 0.7})
plt.show()
```



Project -1

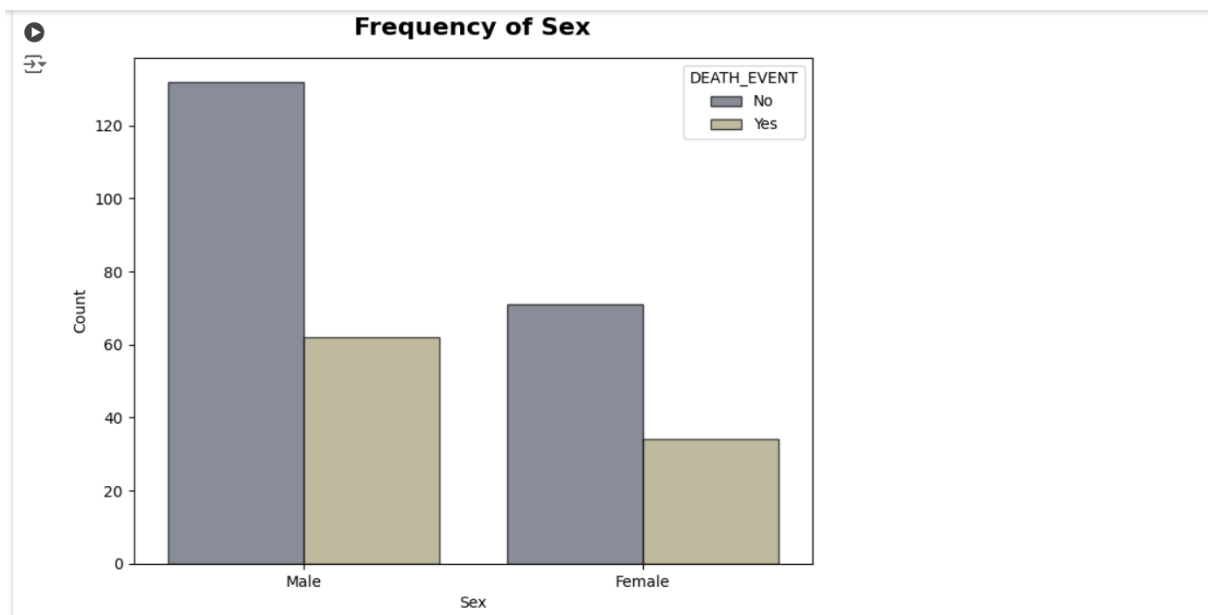
Heart failure Analysis



2. Showing the relationship between categorical variable "Sex" and its frequency using bar plot

```
plt.figure(figsize=(8, 6))
sns.countplot(x='sex', data=dataset, palette='cividis', order=dataset['sex'].value_counts().index, hue='DEATH_EVENT', dodge=True, hue_order=['No', 'Yes'], edgecolor='black', linewidth=1)
plt.title('Frequency of Sex', fontsize=16, fontweight='bold', pad=15)
plt.xlabel('Sex')
plt.ylabel('Count')
plt.show()
```

+ Code + Text



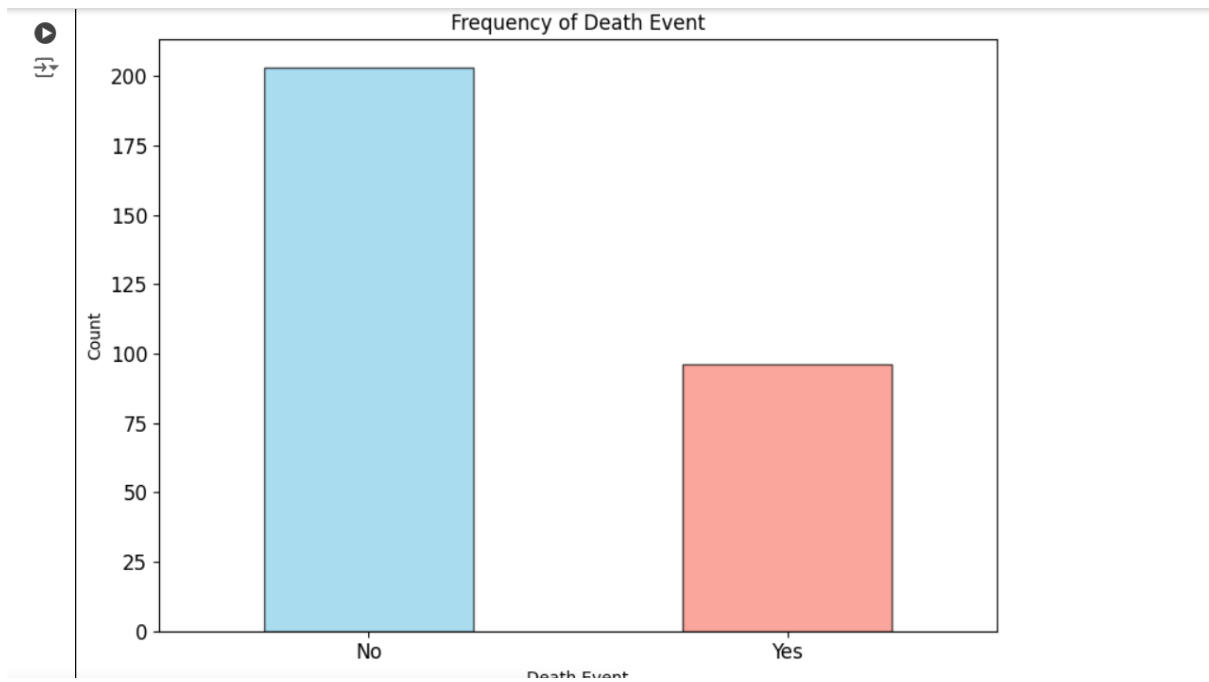
3. Showing the relationship between categorical variable death event and its frequency

```
[132] death_event_counts = dataset.groupby('DEATH_EVENT')['DEATH_EVENT'].count()
plt.figure(figsize=(8, 6), dpi=100, facecolor='w', edgecolor='k', frameon=True, linewidth=0.5, tight_layout=None)
death_event_counts.plot(kind='bar', color=['skyblue', 'salmon'], edgecolor='black', linewidth=1, alpha=0.7, fontsize=12, rot=0)
plt.title('Frequency of Death Event')
plt.xlabel('Death Event')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

Project -1

Heart failure Analysis

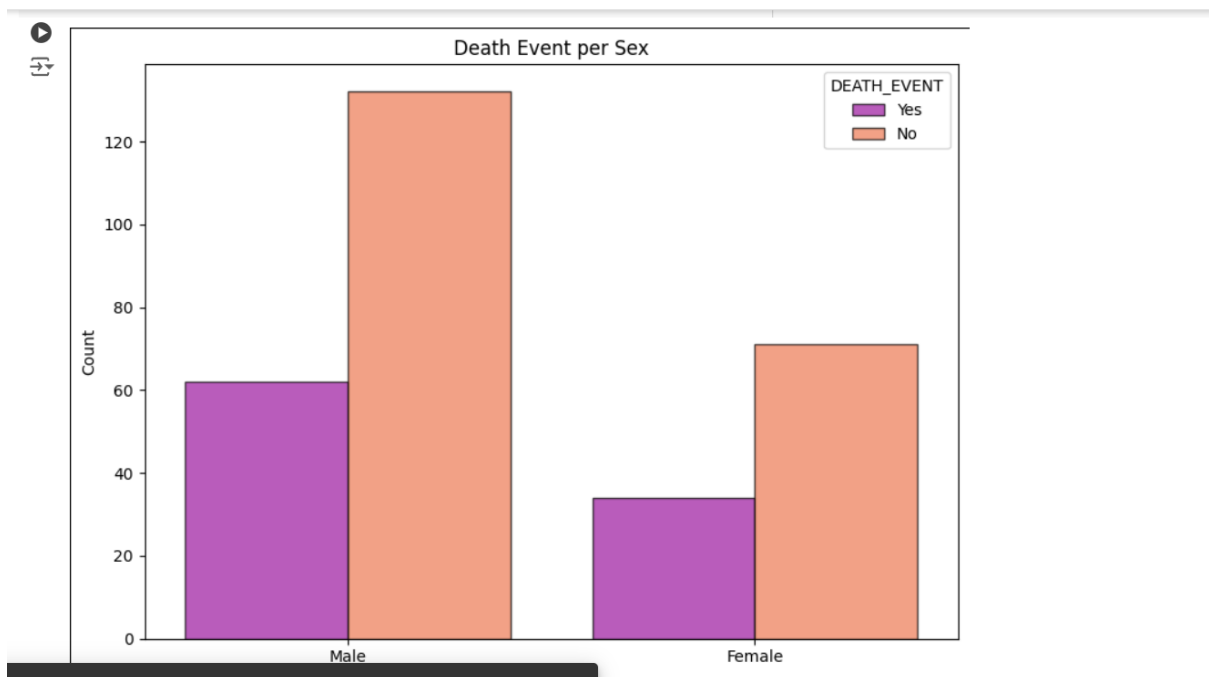
+ Code + Text



4. Death event per each sex using bar plot

```
plt.figure(figsize=(8, 6), dpi=100, facecolor='w', edgecolor='k', frameon=True, linewidth=0.5, tight_layout=None)
sns.countplot(x='sex', hue='DEATH_EVENT', data=dataset, palette='plasma', edgecolor='black', linewidth=1, alpha=0.7, saturation=1)
plt.title('Death Event per Sex')
plt.xlabel('Sex')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

+ Code + Text



Project -1

Heart failure Analysis

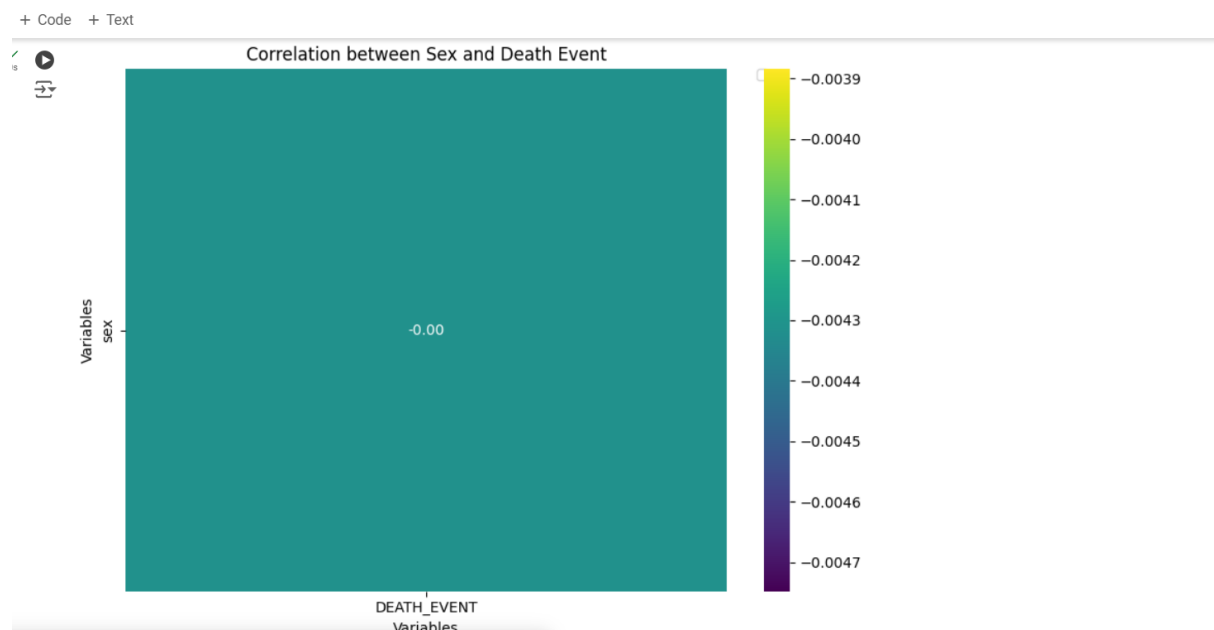
5. Sex correlated with death rate using heatmap

```
correlation_matrix = dataset.apply(lambda x: pd.factorize(x)[0]).corr()

sex_death_correlation = correlation_matrix.loc['sex', 'DEATH_EVENT']

plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix[['DEATH_EVENT']].loc[['sex']], annot=True, cmap='viridis', fmt=".2f")
plt.title('Correlation between Sex and Death Event')
plt.xlabel('Variables')
plt.ylabel('Variables')
plt.tight_layout()
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0)
plt.show()

print(f"Correlation between Sex and Death Event: {sex_death_correlation}")
```



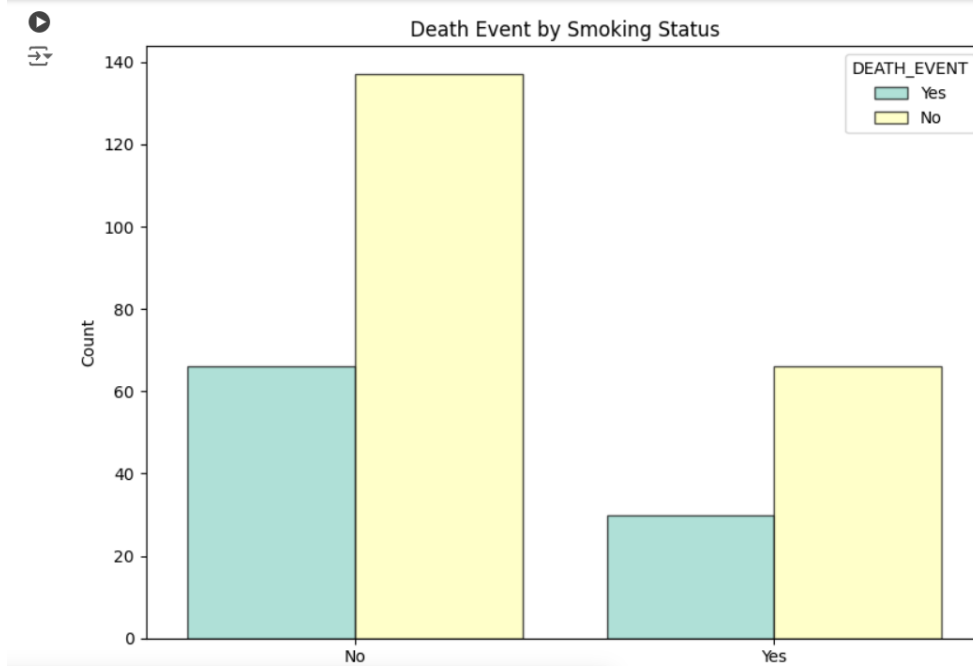
6. Smoking against death using bar plot

```
import matplotlib.pyplot as plt
plt.figure(figsize=(8, 6))
sns.countplot(x='smoking', hue='DEATH_EVENT', data=dataset, palette='Set3', edgecolor='black', linewidth=1, alpha=0.7, saturation=1)
plt.title('Death Event by Smoking Status')
plt.xlabel('Smoking')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```


Project -1

Heart failure Analysis

+ Code + Text



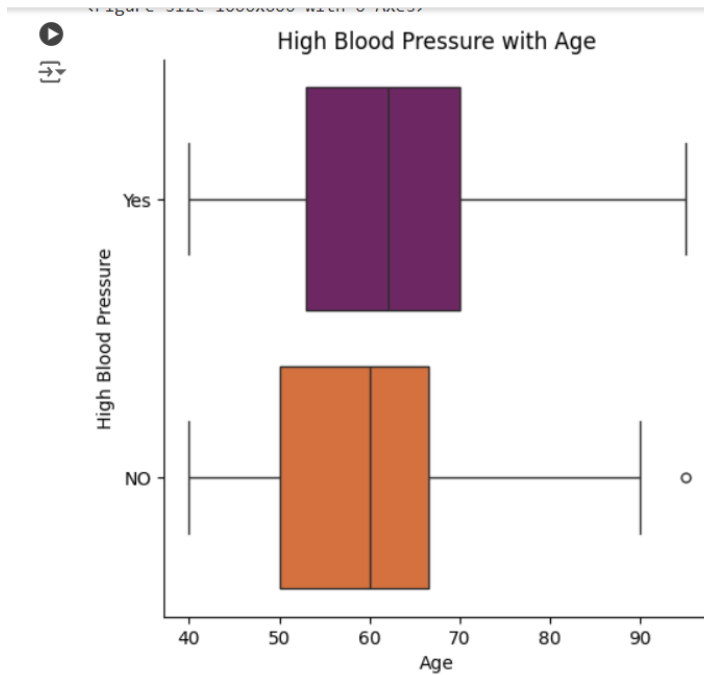
7. High blood pressure with age using catplot

```
import matplotlib.pyplot as plt
plt.figure(figsize=(10, 6), dpi=100, facecolor='w', edgecolor='k', frameon=True, linewidth=0.5, tight_layout=None)
sns.catplot(x='age', y='high_blood_pressure', data=dataset, kind='box', palette='inferno')
plt.title('High Blood Pressure with Age')
plt.xlabel('Age')
plt.ylabel('High Blood Pressure')
plt.show()
```

<ipython-input-136-1d0286d8683e>:5: FutureWarning:
Passing 'palette' without assigning 'hue' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.

sns.catplot(x='age', y='high_blood_pressure', data=dataset, kind='box', palette='inferno')
<Figure size 1000x600 with 0 Axes>

+ Code + Text



Project -1

Heart failure Analysis