

```
# -*- coding: utf-8 -*-
```

```
"""EDA assingment HEART FALIUR.ipynb
```

Automatically generated by Colab.

Original file is located at

<https://colab.research.google.com/drive/1KUJQ8srKshwxggRARLh8NCKBlk7wc7hD>

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as st
import statsmodels.stats.proportion as sp

import matplotlib.ticker as mticker

import warnings
warnings.filterwarnings("ignore")

pd.read_csv("heart_failure_clinical_records_dataset.csv")

df=pd.read_csv("heart_failure_clinical_records_dataset.csv")

df

df.head()

df.tail()

df.info

df.shape

df.isnull().sum()

df.columns

df.dtypes

df.count()

df.columns

df.nunique()

df.value_counts()

df.duplicated().sum()

# Outliers values
plt.figure(figsize=(20, 15))
plt.boxplot(df, labels=df.columns)
plt.show()

df.age
```

#1.What is the distribution of age among heart failure patients in the dataset?

```
df.groupby('sex')['age'].unique()
```

```
plt.figure(figsize=(10, 7))
plt.hist(df.sex, histtype='bar')
plt.title('Gender of Patients')
plt.show()
```

```
sns.histplot(df.age,kde=True,color="r")
```

#2 How does the death rate vary with age?

```
df.columns
```

```
df.DEATH_EVENT.value_counts()
```

```
df.groupby('DEATH_EVENT')['age'].describe()
```

```
plt.figure(figsize=(10, 7))
death_age = df.groupby('DEATH_EVENT')['age'].mean()
myexplode = [0.1, 0]
plt.pie(death_age, autopct='%1f%%', labels=['Alive', 'Dead'], explode=myexplode, shadow=True)
plt.title('Average Ages In Death Cases')
plt.legend(death_age, loc='upper right')
plt.show()
```

#2' What is the percentage of male and female patients in the dataset

```
df.columns
```

```
df.sex.value_counts(normalize= True)
```

#4.how does the platlet caount vary different age groupes?

```
df.columns
```

```
df.groupby('platelets')['age'].unique()
```

```
df.groupby('platelets')['age'].describe()
```

```
df.platelets.value_counts()
```

```
plt.figure(figsize=(10, 7))
plt.hist(df.high_blood_pressure, histtype='bar')
plt.title('platelets')
plt.show()
```

#5Is there any corealation between crathine and sodium levels?

```
df.columns
```

```
plt.figure(figsize=(15, 10))
sns.heatmap(df.loc[:, ['serum_creatinine', 'serum_sodium']].corr(), annot=True)
plt.title('serum_creatinine', 'serum_sodium', loc='left')
```

```
plt.show()
```

#6. how does the prevalence of high blood pressure differ between male and female patients?

```
df.groupby('sex')['high_blood_pressure'].value_counts()
```

```
plt.figure(figsize=(10, 7))
plt.hist(df.sex, histtype='bar')
plt.title('Gender of Patients')
plt.show()
```

what's the relation between smoking habit and occurrence of heart failure??

```
df.columns
```

```
df.groupby('DEATH_EVENT')['smoking'].value_counts()
```

```
plt.figure(figsize=(10, 7))
death_smok = df.groupby('DEATH_EVENT')['smoking'].value_counts()
myexplode = [0.1, 0, 0.2, 0]
plt.pie(death_smok, autopct='%1f%%', labels=['Alive-non smoker', 'Alive-smoker', 'Dead-non smoker',
'Dead-smoker'], explode=myexplode, shadow=True)
plt.title('Smoker Patients In Death Cases')
plt.legend(death_smok, loc='upper left')
plt.show()
```

#8. are there any noticeable patterns in the distribution of death events across the different age groups??

```
df.groupby('DEATH_EVENT')['age'].describe()
```

```
plt.figure(figsize=(10, 7))
death_age = df.groupby('DEATH_EVENT')['age'].mean()
myexplode = [0.1, 0]
plt.pie(death_age, autopct='%1f%%', labels=['Alive', 'Dead'], explode=myexplode, shadow=True)
plt.title('Average Ages In Death Cases')
plt.legend(death_age, loc='upper right')
plt.show()
```

#9. is there any significant difference between patients with and without diabetes?

```
plt.figure(figsize=(10, 7))
plt.hist(df.diabetes, histtype='bar')
plt.title('Diabetic Patients')
plt.show()
```

```
plt.figure(figsize=(10, 7))
death_diab = df.groupby('DEATH_EVENT')['diabetes'].value_counts()
myexplode = [0.1, 0, 0.2, 0]
plt.pie(death_diab, autopct='%1f%%', labels=['Alive-non diabetic', 'Alive-diabetic', 'Dead-non
diabetic', 'Dead-diabetic'], explode=myexplode, shadow=True)
plt.title('Diabetic Patients In Death Cases')
plt.legend(death_diab, loc='upper left')
plt.show()
```

how does the serum creatinine level vary with the patient who survived and who didn't?

```
df.columns
```

```
df.groupby('DEATH_EVENT')['serum_creatinine'].mean()

plt.figure(figsize=(10, 7))
death_SC = df.groupby('DEATH_EVENT')['serum_creatinine'].mean()
myexplode = [0.1, 0]
plt.pie(death_SC, autopct='%1f%%', labels=['Alive-avg Serum Creatinine', 'Dead-avg Serum Creatinine'], explode=myexplode, shadow=True)
plt.title('Average values of Serum Creatinine In Death Cases')
plt.legend(death_SC, loc='upper left')
plt.show()

#complate
```