

TASK-1

PROBLEM STATEMENT:- Create a barchart or histogram to visualize distribution of a categorical or continous variable


Dataset Used:Heart Failure Prediction Dataset

Source:[Kaggle](#)

Description:- Cardiovascular diseases (CVDs) stand as the foremost global cause of mortality, claiming approximately 17.9 million lives annually, amounting to 31% of all deaths worldwide. Among these fatalities, four out of five are attributed to heart attacks and strokes, with one-third transpiring prematurely in individuals below 70 years of age. Heart failure, often stemming from CVDs, is a prevalent occurrence. To address the urgent need for early detection and management, especially for those at high risk due to factors like hypertension, diabetes, or hyperlipidemia, leveraging machine learning models can provide invaluable assistance.

```
import pandas as pd
import matplotlib as pyplot
```

```
data=pd.read_csv('heart.csv')
data
```



	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up	
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	
2	37	M	ATA	130	283	0	ST	98	N	0.0	Up	
3	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	
4	54	M	NAP	150	195	0	Normal	122	N	0.0	Up	
...	
913	45	M	TA	110	264	0	Normal	132	N	1.2	Flat	
914	68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	
915	57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	
917	38	M	NAP	138	175	0	Normal	173	N	0.0	Up	

918 rows × 12 columns

Next steps: [View recommended plots](#)

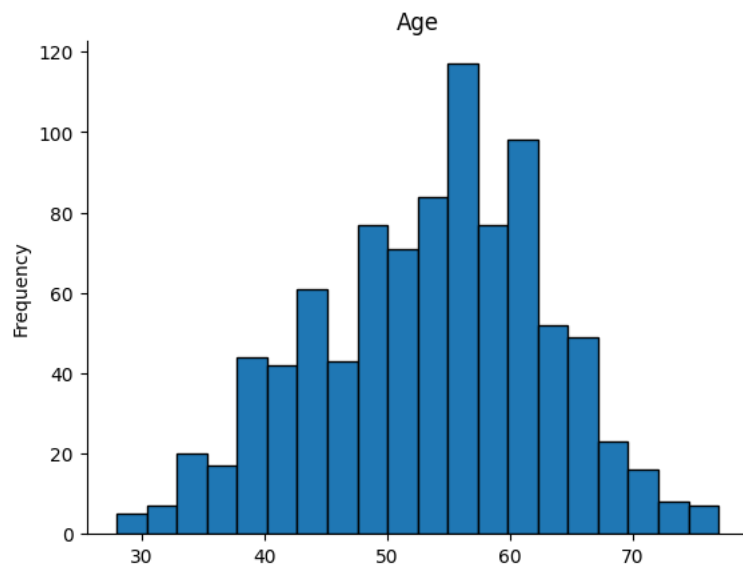
```
print(data.describe())
```

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000
mean	53.510893	132.396514	198.799564	0.233115	136.809368	0.887364	0.553377
std	9.432617	18.514154	109.384145	0.423046	25.460334	1.066570	0.497414
min	28.000000	0.000000	0.000000	0.000000	60.000000	-2.600000	0.000000
25%	47.000000	120.000000	173.250000	0.000000	120.000000	0.000000	0.000000
50%	54.000000	130.000000	223.000000	0.000000	138.000000	0.600000	1.000000
75%	60.000000	140.000000	267.000000	0.000000	156.000000	1.500000	1.000000
max	77.000000	200.000000	603.000000	1.000000	202.000000	6.200000	1.000000

VISUALISATION

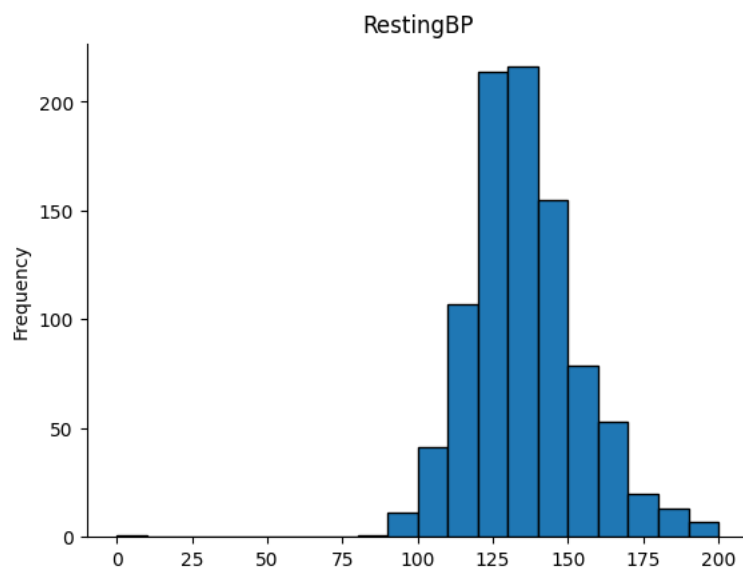
```
#Frequency of Age groups

data['Age'].plot(kind='hist', bins=20, title='Age',edgecolor='black')
plt.gca().spines[['top', 'right']].set_visible(False)
```



#Frequency of RestingBP

```
data['RestingBP'].plot(kind='hist', bins=20, title='RestingBP', edgecolor="black")
plt.gca().spines[['top', 'right']].set_visible(False)
```



#Distribution of Age by Heart Disease Status

```
heart_disease_yes = data[data['HeartDisease'] == 1]
heart_disease_no = data[data['HeartDisease'] == 0]

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

plt.hist(heart_disease_yes['Age'], bins=20, alpha=0.5, color='blue', label='Heart Disease = Yes')
plt.hist(heart_disease_no['Age'], bins=20, alpha=0.5, color='orange', label='Heart Disease = No')

plt.title('Distribution of Age by Heart Disease Status')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.legend()
plt.show()
```

