

MAINFLOW SERVICES AND TECHNOLOGIES

TASK-5

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
df=pd.read_csv("C:\\Users\\glady\\OneDrive\\Desktop\\DS with Python\\heart.csv")

print(df.head())
print(df.tail())
print(df.columns.values)
print(df.isna().sum())
print(df.info())
df.hist(bins=90, grid=False, figsize=(20, 15))
plt.show()
print(df.describe())
print(df.target.value_counts())

df.target.value_counts().plot(kind='bar', color=["orchid", "salmon"])
plt.title("Heart Disease Values (VR)")
plt.xlabel("1 = Heart Disease, 0 = No Heart Disease")
plt.ylabel("Amount")
plt.show()

df.target.value_counts().plot(kind='pie', figsize=(8, 6), autopct='%1.1f%%', colors=["orchid", "salmon"])
plt.title("Distribution of Heart Disease (VR)")
plt.legend(["Disease", "No Disease"])
plt.show()

print(df.sex.value_counts())
df.sex.value_counts().plot(kind='pie', figsize=(8, 6), autopct='%1.1f%%', colors=["skyblue", "lightpink"])
plt.title("Male Female Ratio (VR)")
plt.legend(['Male', 'Female'], loc="upper right")
plt.show()

cross_tab = pd.crosstab(df.target, df.sex)
print(cross_tab)
sns.countplot(x='target', data=df, hue='sex')
plt.title("Heart Disease Frequency for Sex (VR)")
plt.xlabel("0 = No Heart Disease, 1 = Heart Disease")
plt.ylabel("Count")
plt.show()

df.cp.value_counts().plot(kind='bar', color=['salmon', 'lightskyblue', 'springgreen', 'khaki'])
plt.title("Chest Pain Type vs Count (VR)")
plt.xlabel("Chest Pain Type")
plt.ylabel("Count")
plt.show()

pd.crosstab(df.sex, df.cp).plot(kind='bar', color=['coral', 'lightskyblue', 'plum', 'khaki'])
plt.title("Type of Chest Pain for Sex (VR)")
plt.xlabel("0 = Female, 1 = Male")
plt.ylabel("Count")
plt.show()

sns.countplot(x='cp', data=df, hue='target', palette='Set2')
plt.title("Chest Pain Type vs Heart Disease (VR)")
plt.xlabel("Chest Pain Type")
plt.ylabel("Count")
plt.show()

sns.displot(x='age', data=df, bins=30, kde=True, color='blue')

plt.title("Age Distribution with Normal Distribution Curve (VR)")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
```

```
['age' 'sex' 'cp' 'trestbps' 'chol' 'fbs' 'restecg' 'thalach' 'exang'
 'oldpeak' 'slope' 'ca' 'thal' 'target']
```

```
age      0
sex      0
cp       0
trestbps 0
chol     0
fbs      0
restecg  0
thalach  0
exang    0
oldpeak  0
slope    0
ca       0
thal     0
target   0
```

```
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1025 entries, 0 to 1024
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	age	1025 non-null	int64
1	sex	1025 non-null	int64
2	cp	1025 non-null	int64
3	trestbps	1025 non-null	int64
4	chol	1025 non-null	int64
5	fbs	1025 non-null	int64
6	restecg	1025 non-null	int64
7	thalach	1025 non-null	int64
8	exang	1025 non-null	int64
9	oldpeak	1025 non-null	float64
10	slope	1025 non-null	int64
11	ca	1025 non-null	int64
12	thal	1025 non-null	int64
13	target	1025 non-null	int64

```
dtypes: float64(1), int64(13)
```

```
memory usage: 112.2 KB
```

```
None
```

	age	sex	...	thal	target
count	1025.000000	1025.000000	...	1025.000000	1025.000000
mean	54.434146	0.695610	...	2.323902	0.513171
std	9.072290	0.460373	...	0.620660	0.500070
min	29.000000	0.000000	...	0.000000	0.000000
25%	48.000000	0.000000	...	2.000000	0.000000
50%	56.000000	1.000000	...	2.000000	1.000000
75%	61.000000	1.000000	...	3.000000	1.000000
max	77.000000	1.000000	...	3.000000	1.000000

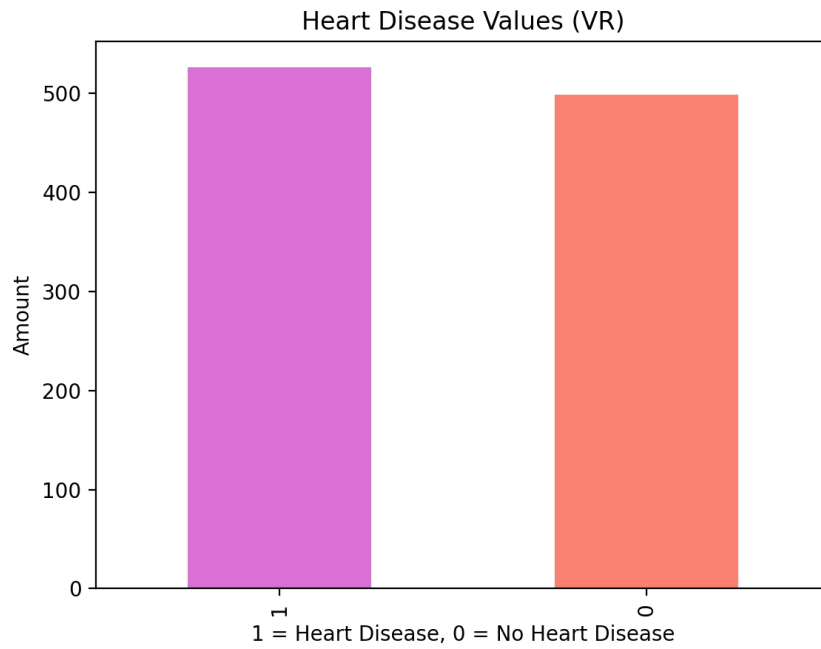
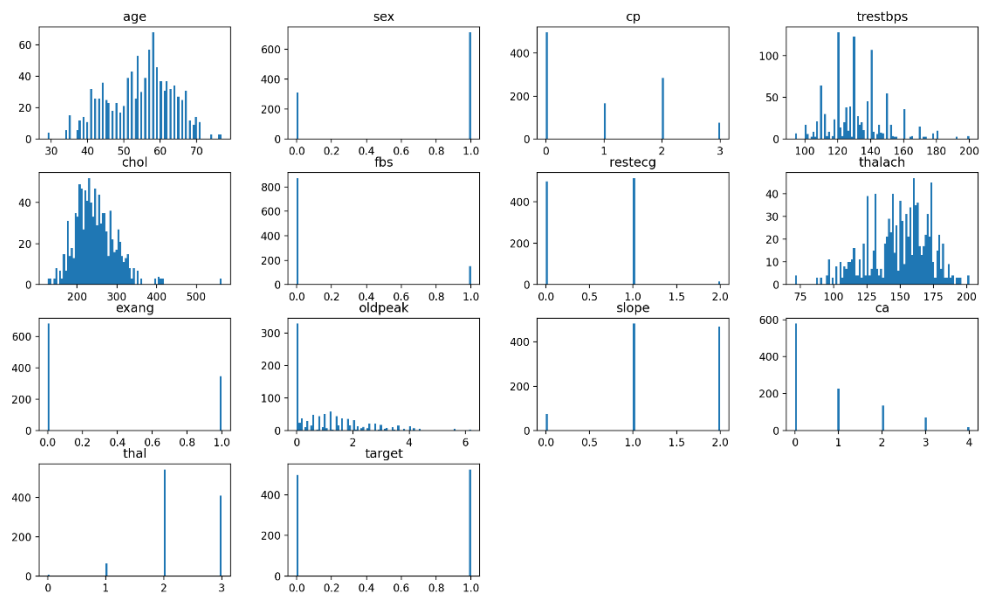
```
[8 rows x 14 columns]
```

```
target
```

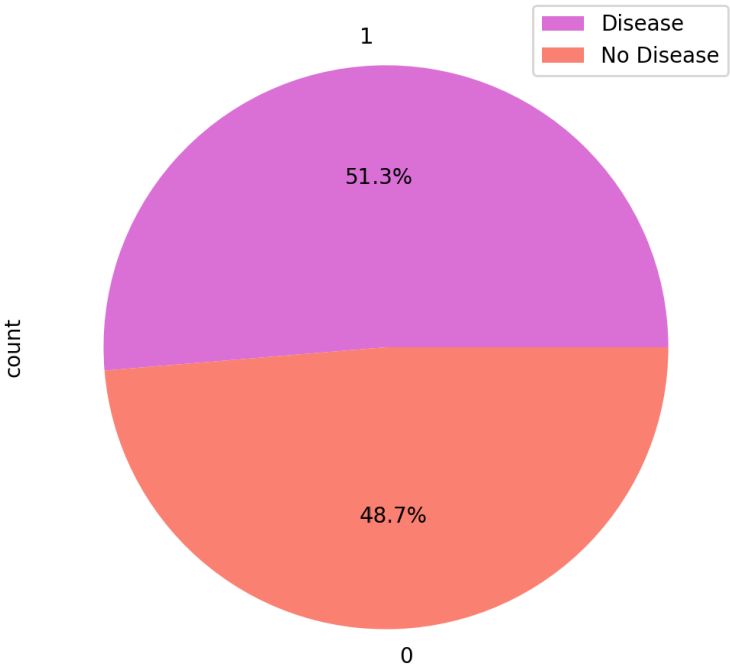
```
1    526
```

```
0    499
```

```
Name: count, dtype: int64
```



Distribution of Heart Disease (VR)



Male Female Ratio(VR)

