

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0.0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2.0	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
42	F	NAP	115	211	0	ST	137	N	0.0	Up	0

Figure 1: Entire Dataset

# Dataset Size: (918, 12)

Figure 2: Size of the Dataset

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0

Figure 3: First 10 Rows of Dataset

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
63	M	ASY	140	187	0	LVH	144	Y	4.0	Up	1
63	F	ASY	124	197	0	Normal	136	Y	0.0	Flat	1
41	M	ATA	120	157	0	Normal	182	N	0.0	Up	0
59	M	ASY	164	176	1	LVH	90	N	1.0	Flat	1
57	F	ASY	140	241	0	Normal	123	Y	0.2	Flat	1
45	M	TA	110	264	0	Normal	132	N	1.2	Flat	1
68	M	ASY	144	193	1	Normal	141	N	3.4	Flat	1
57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
38	M	NAP	138	175	0	Normal	173	N	0.0	Up	0

Figure 4: Last 10 Rows of Dataset

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
 #   Column   Non-Null Count Dtype  
--- 
 0   Age      918 non-null   int64  
 1   Sex      918 non-null   object  
 2   ChestPainType 918 non-null   object  
 3   RestingBP 918 non-null   int64  
 4   Cholesterol 918 non-null   int64  
 5   FastingBS 918 non-null   int64  
 6   RestingECG 918 non-null   object  
 7   MaxHR    918 non-null   int64  
 8   ExerciseAngina 918 non-null   object  
 9   Oldpeak   918 non-null   float64 
 10  ST_Slope  918 non-null   object  
 11  HeartDisease 918 non-null   int64  
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB

```

Figure 5: Dataset Information

Statistic	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
count	918.0	918.0	918.0	918.0	918.0	918.0	918.0
mean	53.5109	132.3965	198.7996	0.2331	136.8094	0.8874	0.5534
std	9.4326	18.5142	109.3841	0.423	25.4603	1.0666	0.4974
min	28.0	0.0	0.0	0.0	60.0	-2.6	0.0
25%	47.0	120.0	173.25	0.0	120.0	0.0	0.0
50%	54.0	130.0	223.0	0.0	138.0	0.6	1.0
75%	60.0	140.0	267.0	0.0	156.0	1.5	1.0
max	77.0	200.0	603.0	1.0	202.0	6.2	1.0

Figure 6: Statistical View of Dataset

Column	Data Type
Age	int64
Sex	object
ChestPainType	object
RestingBP	int64
Cholesterol	int64
FastingBS	int64
RestingECG	object
MaxHR	int64
ExerciseAngina	object
Oldpeak	float64
ST_Slope	object
HeartDisease	int64

Figure 7: Data Types of Dataset Columns

Column	Null Count
Age	0
Sex	0
ChestPainType	0
RestingBP	0
Cholesterol	0
FastingBS	0
RestingECG	0
MaxHR	0
ExerciseAngina	0
Oldpeak	0
ST_Slope	0
HeartDisease	0

Figure 8: Null Values per Column

Column	Unique Count
Age	50
Sex	2
ChestPainType	4
RestingBP	67
Cholesterol	222
FastingBS	2
RestingECG	3
MaxHR	119
ExerciseAngina	2
Oldpeak	53
ST_Slope	3
HeartDisease	2

Figure 9: Number of Unique Values per Column

Column	% Nulls
Age	0.0
Sex	0.0
ChestPainType	0.0
RestingBP	0.0
Cholesterol	0.0
FastingBS	0.0
RestingECG	0.0
MaxHR	0.0
ExerciseAngina	0.0
Oldpeak	0.0
ST_Slope	0.0
HeartDisease	0.0

Figure 10: Percentage of Null Values per Column

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0.0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2.0	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
42	F	NAP	115	211	0	ST	137	N	0.0	Up	0

Figure 11: Dataset After Filling Missing Values with Mean

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
61	M	ASY	134	0	1	ST	86	N	1.5	Flat	1
63	M	ASY	185	0	0	Normal	98	Y	0.0	Up	1
66	M	ASY	150	0	0	Normal	108	Y	2.0	Flat	1
60	M	ASY	135	0	0	Normal	63	Y	0.5	Up	1
63	M	ASY	150	0	0	Normal	86	Y	2.0	Flat	1
57	M	ASY	140	0	0	Normal	120	Y	2.0	Flat	1
68	M	ASY	135	0	0	ST	120	Y	0.0	Up	1
61	M	ASY	150	0	0	Normal	117	Y	2.0	Flat	1
64	F	ASY	200	0	0	Normal	140	Y	1.0	Flat	1
60	M	ASY	140	293	0	LVH	170	N	1.2	Flat	1
56	M	ASY	120	0	0	ST	148	N	0.0	Flat	1
63	F	ASY	150	407	0	LVH	154	N	4.0	Flat	1
47	M	ASY	160	0	0	Normal	124	Y	0.0	Flat	1
59	M	ASY	135	234	0	Normal	161	N	0.5	Flat	0
53	M	ASY	142	226	0	LVH	111	Y	0.0	Up	0

Figure 12: Dataset Sorted by Third Column (ChestPainType)

Age	Sex_new	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0.0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2.0	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
42	F	NAP	115	211	0	ST	137	N	0.0	Up	0

Figure 13: Renamed Column 'Sex' → 'Sex\_new'

**Categorical Columns:**  
['Sex', 'ChestPainType', 'RestingECG', 'ExerciseAngina', 'ST\_Slope']

**Numerical Columns:**  
['Age', 'RestingBP', 'Cholesterol', 'FastingBS', 'MaxHR', 'Oldpeak', 'HeartDisease']

Figure 14: Categorical and Numerical Columns

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngin	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0.0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0.0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0.0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0.0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0.0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2.0	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0.0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat	1
42	F	NAP	115	211	0	ST	137	N	0.0	Up	0

Figure 15: Dataset After Dropping Outliers