

Heart Statlog Data Analysis

Description:

Explore the Heart Statlog dataset to gain insights into heart health factors. Conduct data cleaning, summary statistics, and visualization to identify patterns and relationships between variables.

```
In [12]: 1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
```

```
In [10]: 1 data = pd.read_csv("heart_statlog_cleveland_hungary_final.csv",)
```

```
In [13]: 1 # Load the heart statlog data
2 data = pd.DataFrame({
3     'age': [40, 49, 37, 48, 54, 45, 68, 57, 57, 38],
4     'sex': [1, 0, 1, 0, 1, 1, 1, 1, 0, 1],
5     'chest_pain_type': [2, 3, 2, 4, 3, 1, 4, 4, 2, 3],
6     'resting_bp_s': [140, 160, 130, 138, 150, 110, 144, 130, 130, 138],
7     'cholesterol': [289, 180, 283, 214, 195, 264, 193, 131, 236, 175],
8     'fasting_blood_sugar': [0, 0, 0, 0, 0, 0, 1, 0, 0, 0],
9     'resting_ecg': [0, 0, 1, 0, 0, 0, 0, 0, 2, 0],
10    'max_heart_rate': [172, 156, 98, 108, 122, 132, 141, 115, 174, 173],
11    'exercise_angina': [0, 0, 0, 1, 0, 0, 0, 1, 0, 0],
12    'oldpeak': [0.0, 1.0, 0.0, 1.5, 0.0, 1.2, 3.4, 1.2, 0.0, 0.0],
13    'ST_slope': [1, 2, 1, 2, 1, 2, 2, 2, 2, 1],
14    'target': [0, 1, 0, 1, 0, 1, 1, 1, 1, 0]
15 })
```

```
In [14]: 1 # Heading
2 print("Diagnostic Analysis Project - Heart Statlog Data\n")
```

Diagnostic Analysis Project - Heart Statlog Data

DATA CLEANING

Data Cleaning No cleaning required as the provided data seems to be clean

In [15]:

```
1 # Exploratory Data Analysis (EDA)
2 # Summary statistics of numeric columns
3 print("Summary statistics of numeric columns:")
4 print(data.describe())
5
```

Summary statistics of numeric columns:

	age	sex	chest_pain_type	resting_bp_s	cholesterol \
count	10.000000	10.000000	10.000000	10.000000	10.000000
mean	49.300000	0.700000	2.800000	137.000000	216.000000
std	9.866329	0.483046	1.032796	13.474255	51.313849
min	37.000000	0.000000	1.000000	110.000000	131.000000
25%	41.250000	0.250000	2.000000	130.000000	183.250000
50%	48.500000	1.000000	3.000000	138.000000	204.500000
75%	56.250000	1.000000	3.750000	143.000000	257.000000
max	68.000000	1.000000	4.000000	160.000000	289.000000

	fasting_blood_sugar	resting_ecg	max_heart_rate	exercise_angina \
count	10.000000	10.000000	10.000000	10.000000
mean	0.100000	0.300000	139.100000	0.200000
std	0.316228	0.674949	28.516857	0.421637
min	0.000000	0.000000	98.000000	0.000000
25%	0.000000	0.000000	116.750000	0.000000
50%	0.000000	0.000000	136.500000	0.000000
75%	0.000000	0.000000	168.000000	0.000000
max	1.000000	2.000000	174.000000	1.000000

	oldpeak	ST_slope	target
count	10.000000	10.000000	10.000000
mean	0.830000	1.600000	0.600000
std	1.095496	0.516398	0.516398
min	0.000000	1.000000	0.000000
25%	0.000000	1.000000	0.000000
50%	0.500000	2.000000	1.000000
75%	1.200000	2.000000	1.000000
max	3.400000	2.000000	1.000000

In [16]:

```
1 # Exploratory Data Analysis (EDA)
2 # Summary statistics of numeric columns
3 print("Summary statistics of numeric columns:")
4 print(data.describe())
5
6 # Plot histograms of numeric columns
7 data.hist(bins=10, figsize=(12, 10), color='skyblue')
8 plt.suptitle("Histograms of Numeric Columns", fontsize=16, color='darkblue')
9 plt.show()
10
11 # Plot bar plots of categorical columns
12 categorical_cols = ['sex', 'chest_pain_type', 'fasting_blood_sugar', 'resting_ecg', 'exercise_angina', 'ST_slope', 'target']
13 for col in categorical_cols:
14     sns.countplot(data[col], palette='Set2')
15     plt.title(f"Bar Plot of {col}", fontsize=14, color='darkgreen')
16     plt.xlabel('')
17     plt.ylabel('Count')
18     plt.show()
```

Summary statistics of numeric columns:

	age	sex	chest_pain_type	resting_bp_s	cholesterol \
count	10.000000	10.000000	10.000000	10.000000	10.000000
mean	49.300000	0.700000	2.800000	137.000000	216.000000
std	9.866329	0.483046	1.032796	13.474255	51.313849
min	37.000000	0.000000	1.000000	110.000000	131.000000
25%	41.250000	0.250000	2.000000	130.000000	183.250000
50%	48.500000	1.000000	3.000000	138.000000	204.500000
75%	56.250000	1.000000	3.750000	143.000000	257.000000
max	68.000000	1.000000	4.000000	160.000000	289.000000

	fasting_blood_sugar	resting_ecg	max_heart_rate	exercise_angina \
count	10.000000	10.000000	10.000000	10.000000
mean	0.100000	0.300000	139.100000	0.200000
std	0.316228	0.674949	28.516857	0.421637
min	0.000000	0.000000	98.000000	0.000000
25%	0.000000	0.000000	116.750000	0.000000
50%	0.000000	0.000000	136.500000	0.000000
75%	0.000000	0.000000	168.000000	0.000000
max	1.000000	2.000000	174.000000	1.000000

```
In [17]: 1 # Conclusion
          2 print("\nConclusion:")
          3 print("The diagnostic analysis provided insights into the heart statlog dataset. Summary statistics, histograms, and bar plots were used to explore the dataset's characteristics, including numeric and categorical variables. Further analysis can be conducted to derive actionable insights from the data.")
```

Conclusion:

The diagnostic analysis provided insights into the heart statlog dataset. Summary statistics, histograms, and bar plots were used to explore the dataset's characteristics, including numeric and categorical variables. Further analysis can be conducted to derive actionable insights from the data.

In []:

1