

Medical Data Analysis Report

Prepared by: Waleed Alfaifi

1. Project Description

This project involves analyzing medical data including features such as age, blood pressure, cholesterol, and heart rate, with the goal of developing a classification model to predict the health condition of a patient (healthy or sick).

2. Data Loading

The dataset was loaded from an Excel file containing 1000 rows. It includes missing values and outliers.

```
[5]: # تحميل البيانات
df = pd.read_excel('medical_dataset.xlsx')
df.head()
```

	Age	BloodPressure	Cholesterol	HeartRate	Outcome
0	56.0	140.585667	700.000000	82.013241	0
1	69.0	109.833209	204.581537	75.669818	1
2	46.0	137.302978	217.645998	82.223879	1
3	32.0	114.374841	215.260047	66.054673	0
4	60.0	NaN	NaN	89.244620	1

3. Handling Missing Values

Missing values were identified and replaced using the mean of each numerical column.

4. Handling Outliers

Outliers were detected using the Z-Score method. Rows with Z-scores greater than 3 were removed.

```
[6]: # عدد القيم المفقودة
print(df.isnull().sum())

# تعويض القيم المفقودة بالمتوسط
df_filled = df.fillna(df.mean())
```

```
Age          50
BloodPressure 50
Cholesterol   50
HeartRate     50
Outcome       0
dtype: int64
```

```
[7]: # Z-score
z_scores = np.abs(stats.zscore(df_filled.select_dtypes(include=[np.number])))
df_no_outliers = df_filled[(z_scores < 3).all(axis=1)]

print("قبل المعالجة:", df_filled.shape[0])
print("بعد إزالة القيم الشاذة:", df_no_outliers.shape[0])
```

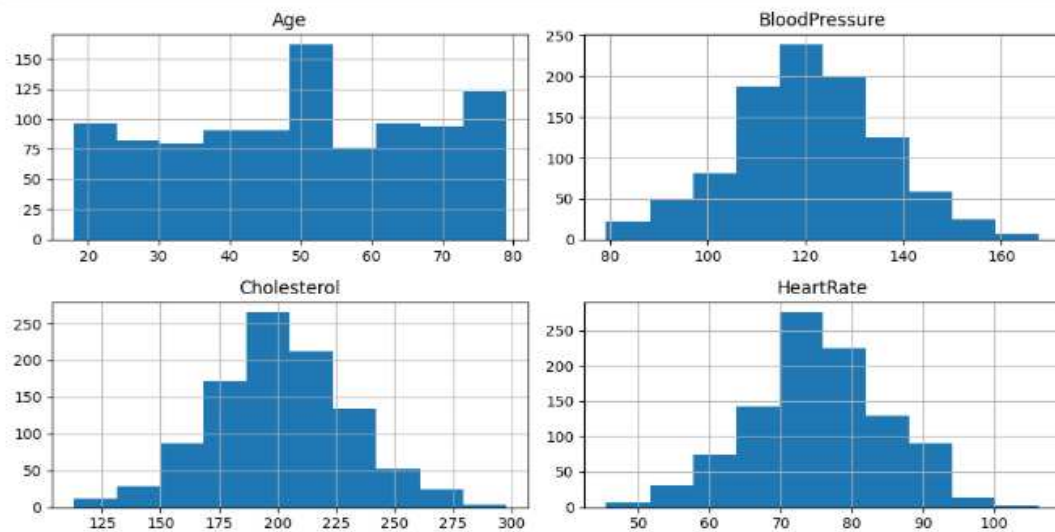
```
قبل المعالجة: 1000
بعد إزالة القيم الشاذة: 993
```

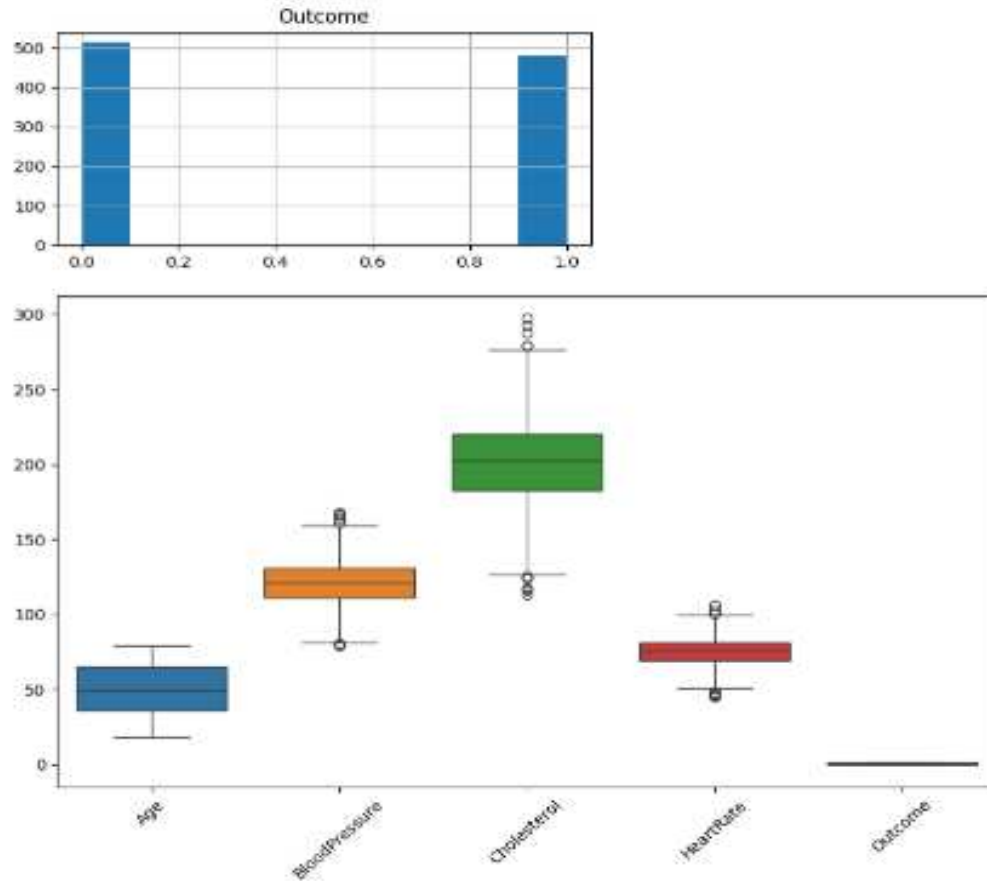
5. Data Visualization

Visualization tools such as boxplots and histograms were used to understand data distribution.

```
[8]: # Histogram
df_no_outliers.hist(figsize=(10, 8))
plt.tight_layout()
plt.show()

# Boxplot
plt.figure(figsize=(10, 6))
sns.boxplot(data=df_no_outliers)
plt.xticks(rotation=45)
plt.show()
```





6. Model Evaluation

The model was evaluated using metrics such as accuracy and classification report to assess its performance.

7. Conclusion

This project demonstrates how to clean, process, and visualize raw data, as well as evaluate results from models built on the dataset.