

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv("kidney_disease.csv")

# 1. Handling Missing Values

# Check missing values
print(data.isnull().sum())

# Fill numerical columns with median
numerical_cols = data.select_dtypes(include=['int64', 'float64']).columns
for col in numerical_cols:
    data[col] = data[col].fillna(data[col].median())

# Fill categorical columns with mode
categorical_cols = data.select_dtypes(include=['object']).columns
for col in categorical_cols:
    data[col] = data[col].fillna(data[col].mode()[0])

# 2. Handling Outliers

# Using IQR method
Q1 = data[numerical_cols].quantile(0.25)
Q3 = data[numerical_cols].quantile(0.75)
IQR = Q3 - Q1

data_clean = data[~((data[numerical_cols] < (Q1 - 1.5 * IQR)) |
(data[numerical_cols] > (Q3 + 1.5 * IQR))).any(axis=1)]

# 3. Data Visualization

# Histogram
data_clean[numerical_cols].hist(figsize=(12,8))
plt.suptitle("Distribution of Numerical Features")
plt.show()

# Boxplot

```

```
plt.figure(figsize=(12,6))
sns.boxplot(data=data_clean[numerical_cols])
plt.xticks(rotation=90)
plt.title("Boxplot for Outlier Detection")
plt.show()

# Correlation heatmap
plt.figure(figsize=(10,8))
sns.heatmap(data_clean[numerical_cols].corr(), cmap="coolwarm", annot=True)
plt.title("Correlation Heatmap")
plt.show()
```