

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

# Load your dataset (replace 'data.csv' with your dataset file)
data = pd.read_csv('data.csv')

# Display the first few rows of the dataset
print("First few rows of the dataset:")
print(data.head())

# Summary statistics
print("\nSummary statistics:")
print(data.describe())

# Data types and missing values
print("\nData types and missing values:")
print(data.info())

# Correlation matrix
correlation_matrix = data.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

# Histograms for numerical variables
numerical_columns = data.select_dtypes(include=[np.number]).columns.tolist()
data[numerical_columns].hist(figsize=(12, 10), bins=20)
plt.suptitle('Histograms of Numerical Variables', y=0.92)
```

```
plt.show()
```

```
# Boxplots for numerical variables
```

```
plt.figure(figsize=(12, 10))
```

```
for i, column in enumerate(numerical_columns):
```

```
    plt.subplot(3, 3, i+1)
```

```
    sns.boxplot(data[column])
```

```
    plt.title(column)
```

```
plt.suptitle('Boxplots of Numerical Variables', y=0.95)
```

```
plt.tight_layout()
```

```
plt.show()
```

```
# Pairplot for numerical variables
```

```
sns.pairplot(data[numerical_columns])
```

```
plt.suptitle('Pairplot of Numerical Variables', y=1.02)
```

```
plt.show()
```

```
# Count plot for categorical variables
```

```
categorical_columns = data.select_dtypes(exclude=[np.number]).columns.tolist()
```

```
plt.figure(figsize=(12, 8))
```

```
for i, column in enumerate(categorical_columns):
```

```
    plt.subplot(2, 2, i+1)
```

```
    sns.countplot(data[column])
```

```
    plt.title(column)
```

```
    plt.xticks(rotation=45)
```

```
plt.suptitle('Count Plot of Categorical Variables', y=0.95)
```

```
plt.tight_layout()
```

```
plt.show()
```