```
import pandas as pd
import matplotlib.pyplot as pl
import numpy as np
def clean_data(input_file, output_file):
    # Load the CSV data
    data = pd.read_csv(input_file)

    # basic introduction
    print("data info :")
    print(data.info())

    # Display summary statistics
    print("\nData Description:")
    print(data.describe())

    # Count the number of missing values
    miss = data.isnull().sum()
    print("\nmissing values:")
    print(miss)

    # Drop missing values
    data = data.dropna()

    # Save cleaned new CSV file
    data.to_csv(output_file, index=False)
    print(f"\nCleaned data saved to {output_file}")

# Example usage:
clean_data("assignment\\fsa-headcount-as-at-28-february-2017.csv", "cleaned_data.csv")
```

```
def line_plot(data, x_column, y_column, labels, title, x_la, y_la, title_font=12, label_fontsize=10, tick_font=5, ylabel_fontsize=10):
    # Create a line plot with multiple lines and proper labels
    for i in range(len(y_column)):
        pl.plot(data[x_column], data[y_column[i]], label=labels[i])

    pl.title(title, fontsize=title_font)
    pl.xlabel(x_la, fontsize=label_fontsize)
    pl.ylabel(y_la, fontsize=ylabel_fontsize)  # Adjust the fontsize for the y-axis label
    pl.xticks(rotation=45, fontsize=tick_font)
    pl.yticks(fontsize=tick_font)
    pl.legend(fontsize=label_fontsize)
    pl.tight_layout()
    pl.show()
def other_plot1(data, x_column, y_column, title, x_label, y_label):
    # Create another type of visualization (e.g., scatter plot)
    pl.scatter(data[x_column], data[y_column])
    pl.title(title)
    pl.xlabel(x_label)
    pl.ylabel(y_label)
    pl.show()

def other_plot2(data, category_column, value_column, title, figsize=(12, 10), category_fontsize=10, autopct_fontsize=10):
    unique_categories = data[category_column].unique()
    color = plt.cm.viridis(np.linspace(0, 1, len(unique_categories)))

    category_colors = dict(zip(unique_categories, colors))

    fig, ax = plt.subplots(figsize=figsize)  # Increase the figsize for a larger pie chart
    data_to_plot = data.groupby(category_column)[value_column].sum()
```

```
    wedges, texts, autotexts = ax.pie(data_to_plot, labels=data_to_plot.index, autopct='%1.1f%%',
                                      textprops={'fontsize': autopct_fontsize},
                                      colors=[category_colors[category] for category in data_to_plot.index])

    pl.title(title, fontsize=14)

    ax.legend(loc='upper right', labels=[f"{category}: {category_colors[category]}" for category in
unique_categories], fontsize=category_fontsize, bbox_to_anchor=(1.2, 1))

    pl.show()

data = pd.read_csv("assignment\\cleaned_data.csv")



# Line Plot
line_plot(data, 'Grade', ['HeadcountMale'], ['Male'], 'Male Headcount by Grade', 'Grade',
'Headcount')

# Scatter Plot
other_plot1(data, 'FTE_Male', 'FTE_Female', 'Scatter Plot of FTE', 'FTE Male', 'FTE Female')

# Pie Chart
other_plot2(data, 'Grade', 'HeadcountMale', 'Headcount by Grade')
```