```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.impute import KNNImputer
from google.colab import files

def clean_data(df):
    # Remove Duplicates
    df = df.drop_duplicates()

    # Handle Missing Values using KNN Imputer
    numeric_cols = df.select_dtypes(include=[np.number]).columns
    imputer = KNNImputer(n_neighbors=3)
    df_imputed = pd.DataFrame(imputer.fit_transform(df[numeric_cols]),
columns=numeric_cols)

    # Restore non-numeric columns
    for col in df.select_dtypes(exclude=[np.number]).columns:
        df_imputed[col] = df[col].values

    return df_imputed

def plot_graphs(raw_df, cleaned_df):
    numeric_cols = raw_df.select_dtypes(include=[np.number]).columns

    for col in numeric_cols:
        plt.figure(figsize=(10, 4))
```

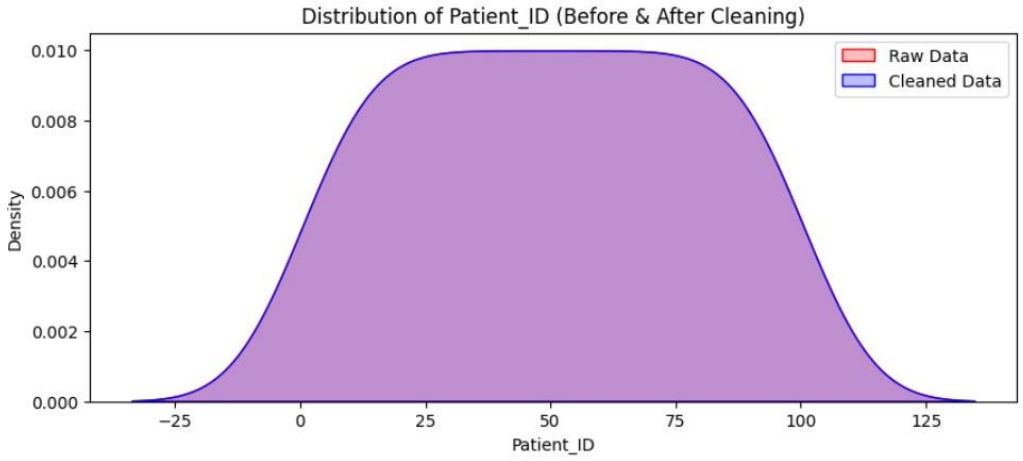
```
    sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
    sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True,
color='blue')
    plt.title(f"Distribution of {col} (Before & After Cleaning)")
    plt.legend()
    plt.show()

# Upload CSV File
uploaded = files.upload()

for filename in uploaded.keys():
  df = pd.read_csv(filename)

  print("Raw Data Preview:")
  print(df.head())

  cleaned_df = clean_data(df)

  print("\nCleaned Data Preview:")
  print(cleaned_df.head())

  # Save Cleaned Data
  cleaned_filename = "cleaned_data.csv"
  cleaned_df.to_csv(cleaned_filename, index=False)

  # Download the cleaned CSV file
  files.download(cleaned_filename)
```

```
# Plot graphs for comparison

plot_graphs(df, cleaned_df)
```

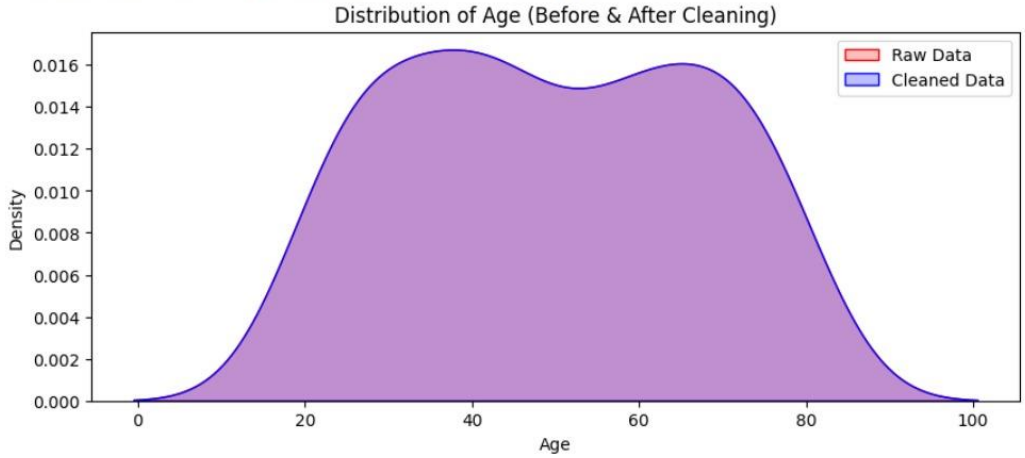
OUTPUT

```
Choose Files healthcare_data.csv
• healthcare_data.csv(text/csv) - 3975 bytes, last modified: 4/4/2025 - 100% done
Saving healthcare_data.csv to healthcare_data (1).csv
Raw Data Preview:
   Patient_ID  Age  Gender  Blood_Pressure  Heart_Rate  Glucose_Level  \
0           1   58  Female             NaN        87.0          130.0
1           2   71    Male           132.0       116.0          117.0
2           3   48  Female             NaN        84.0           73.0
3           4   34    Male             NaN       109.0          104.0
4           5   62  Female           102.0        82.0          118.0

       Diagnosis
0       Diabetes
1  Heart Disease
2       Diabetes
3       Diabetes
4       Diabetes

Cleaned Data Preview:
   Patient_ID   Age  Blood_Pressure  Heart_Rate  Glucose_Level  Gender  \
0         1.0  58.0      143.666667        87.0          130.0  Female
1         2.0  71.0      132.000000       116.0          117.0    Male
2         3.0  48.0      143.333333        84.0           73.0  Female
3         4.0  34.0      149.000000       109.0          104.0    Male
4         5.0  62.0      102.000000        82.0          118.0  Female

       Diagnosis
0       Diabetes
1  Heart Disease
2       Diabetes
3       Diabetes
4       Diabetes
```

```
`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
<ipython-input-2-185f48fa946f>:29: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True, color='blue')
```

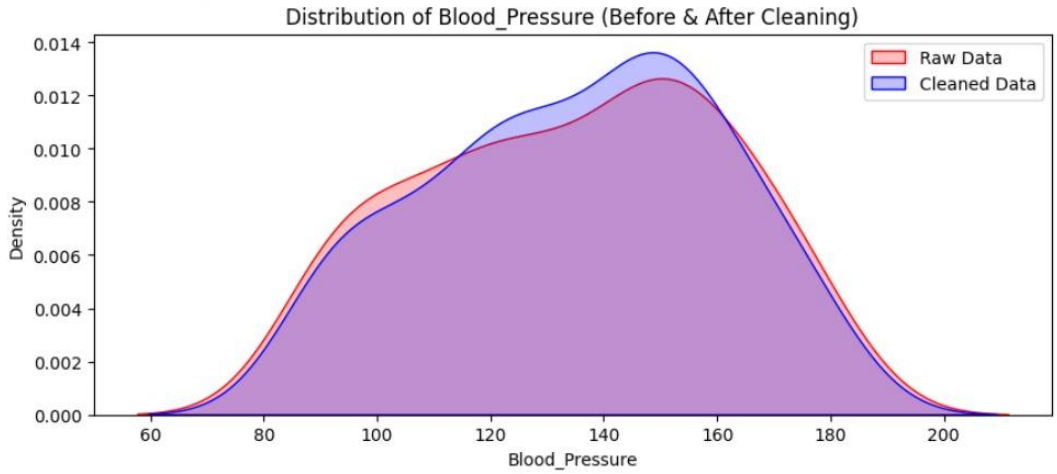


```
`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
<ipython-input-2-185f48fa946f>:29: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True, color='blue')
```



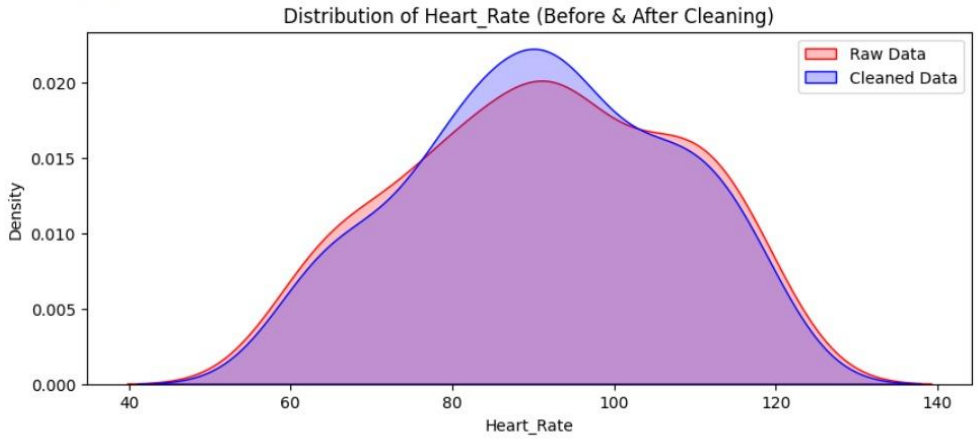

```
<ipython-input-2-185f48fa946f>:28: FutureWarning:
```

```
`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
<ipython-input-2-185f48fa946f>:29: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True, color='blue')
```




```
<ipython-input-2-185f48fa946f>:28: FutureWarning:
```

```
`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
<ipython-input-2-185f48fa946f>:29: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True, color='blue')
```




```
`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(raw_df[col].dropna(), label='Raw Data', shade=True, color='red')
<ipython-input-2-185f48fa946f>:29: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(cleaned_df[col].dropna(), label='Cleaned Data', shade=True, color='blue')
```