Output for Project 2

Step 1
Loading Packages from pandas, matplotlib, seaborn

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

Step 2
Loading data from gss.csv

```
df=pd.read_csv("gss.csv")
```

Step 3
Summary of the dataset

(Age refers to the age of the person whereas year refers to the year that they were born.)

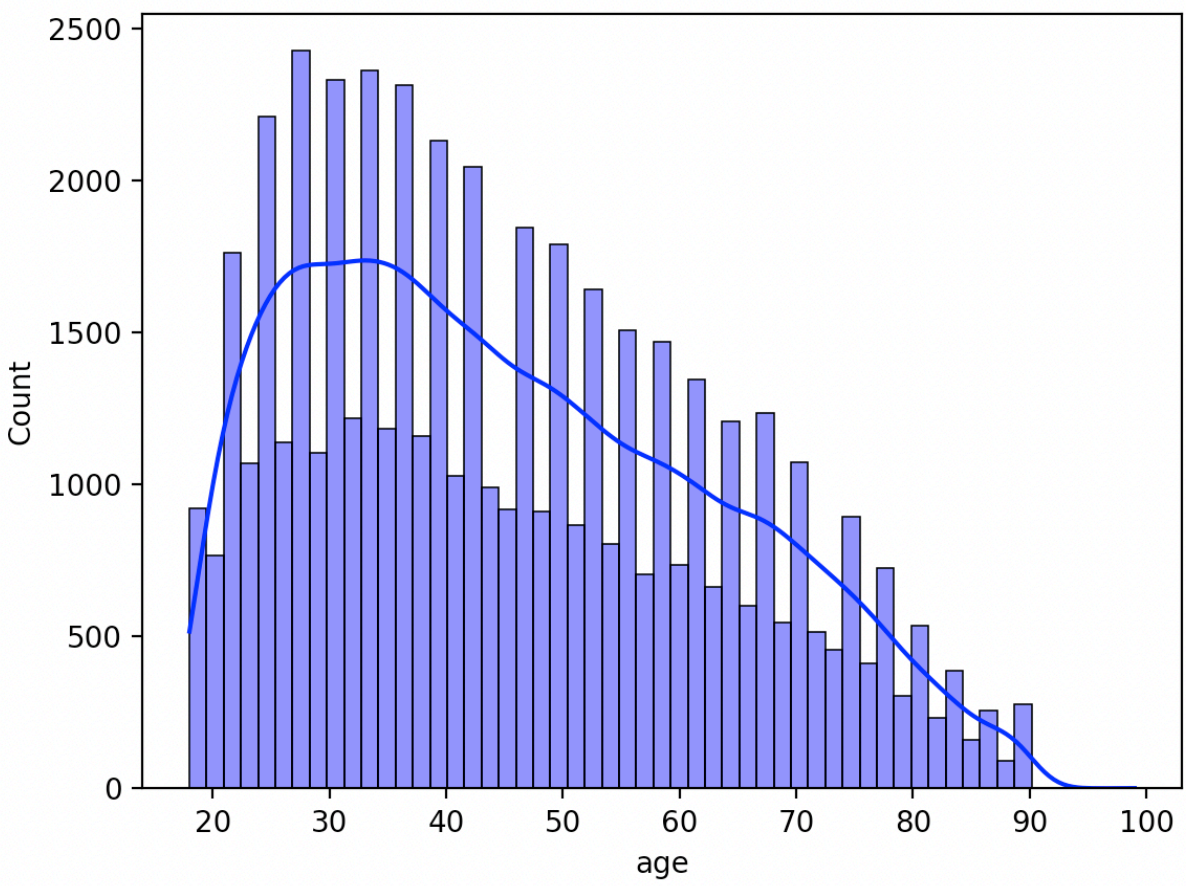
```
print(df.describe())
```

```
                age          year
count  53285.000000  53474.000000
mean      45.631285   1991.954894
std       17.472393     10.948351
min       18.000000   1973.000000
25%       31.000000   1983.000000
50%       43.000000   1993.000000
75%       59.000000   2002.000000
max       99.000000   2010.000000
```

Step 4

Visualization – Plot a histogram on density of Age

```
plot = sns.histplot(df["age"], kde=True, color="blue", label="Age")
plot.legend()
plt.show()
```



Step 5

Conclusion: The dataset includes mostly age group from 25 to 35. The mean age is approximately 46.