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

```
df=pd.read_csv('db.csv')
df.head()
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	Diabetes_Pedigree	Ag
0	6	148.0	72.0	35.0	NaN	33.6	0.627	5
1	1	85.0	66.0	29.0	NaN	26.6	0.351	3
2	8	183.0	64.0	NaN	NaN	23.3	0.672	3
3	1	89.0	66.0	23.0	94.0	28.1	0.167	2
4	0	137.0	40.0	35.0	168.0	43.1	2.288	3

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.isnull().sum()
```

	0
Pregnant	0
Glucose	0
Diastolic_BP	35
Skin_Fold	227
Serum_Insulin	374
BMI	11
Diabetes_Pedigree	0
Age	0
Class	0

dtype: int64

```
df['Diastolic_BP']=df['Diastolic_BP'].fillna(df['Diastolic_BP'].mean())
df['Skin_Fold']=df['Skin_Fold'].fillna(df['Skin_Fold'].mean())
df['Serum_Insulin']=df['Serum_Insulin'].fillna(df['Serum_Insulin'].mean())
```

```
df['BMI']=df['BMI'].fillna(df['BMI'].mean())
df
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	Diabetes_Pedigree
0	6	148.0	72.0	35.00000	155.548223	33.6	0.627
1	1	85.0	66.0	29.00000	155.548223	26.6	0.351
2	8	183.0	64.0	29.15342	155.548223	23.3	0.672
3	1	89.0	66.0	23.00000	94.000000	28.1	0.167
4	0	137.0	40.0	35.00000	168.000000	43.1	2.288
...	...	...	...	...	...	...	...
763	10	101.0	76.0	48.00000	180.000000	32.9	0.171
764	2	122.0	70.0	27.00000	155.548223	36.8	0.340
765	5	121.0	72.0	23.00000	112.000000	26.2	0.245
766	1	126.0	60.0	29.15342	155.548223	30.1	0.349
767	1	93.0	70.0	31.00000	155.548223	30.4	0.315

768 rows × 9 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

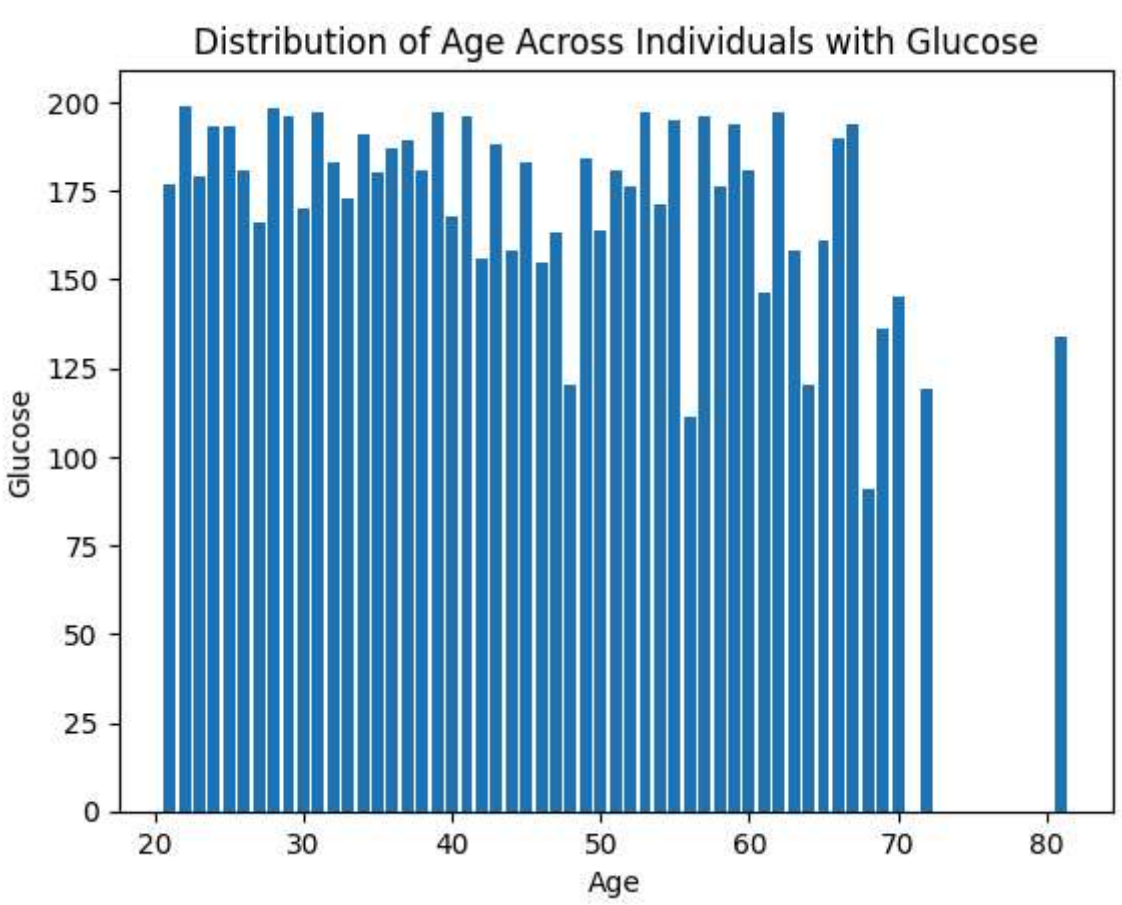
Correlation Btw the Data

```
df.corr()
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI
Pregnant	1.000000	0.127911	0.208522	0.082989	0.056027	0.021565
Glucose	0.127911	1.000000	0.218367	0.192991	0.420157	0.230941
Diastolic_BP	0.208522	0.218367	1.000000	0.192816	0.072517	0.281268
Skin_Fold	0.082989	0.192991	0.192816	1.000000	0.158139	0.542398
Serum_Insulin	0.056027	0.420157	0.072517	0.158139	1.000000	0.166586
BMI	0.021565	0.230941	0.281268	0.542398	0.166586	1.000000
Diabetes_Pedigree	-0.033523	0.137060	-0.002763	0.100966	0.098634	0.153400
Age	0.544341	0.266534	0.324595	0.127872	0.136734	0.025519
Class	0.221898	0.492928	0.166074	0.215299	0.214411	0.311924

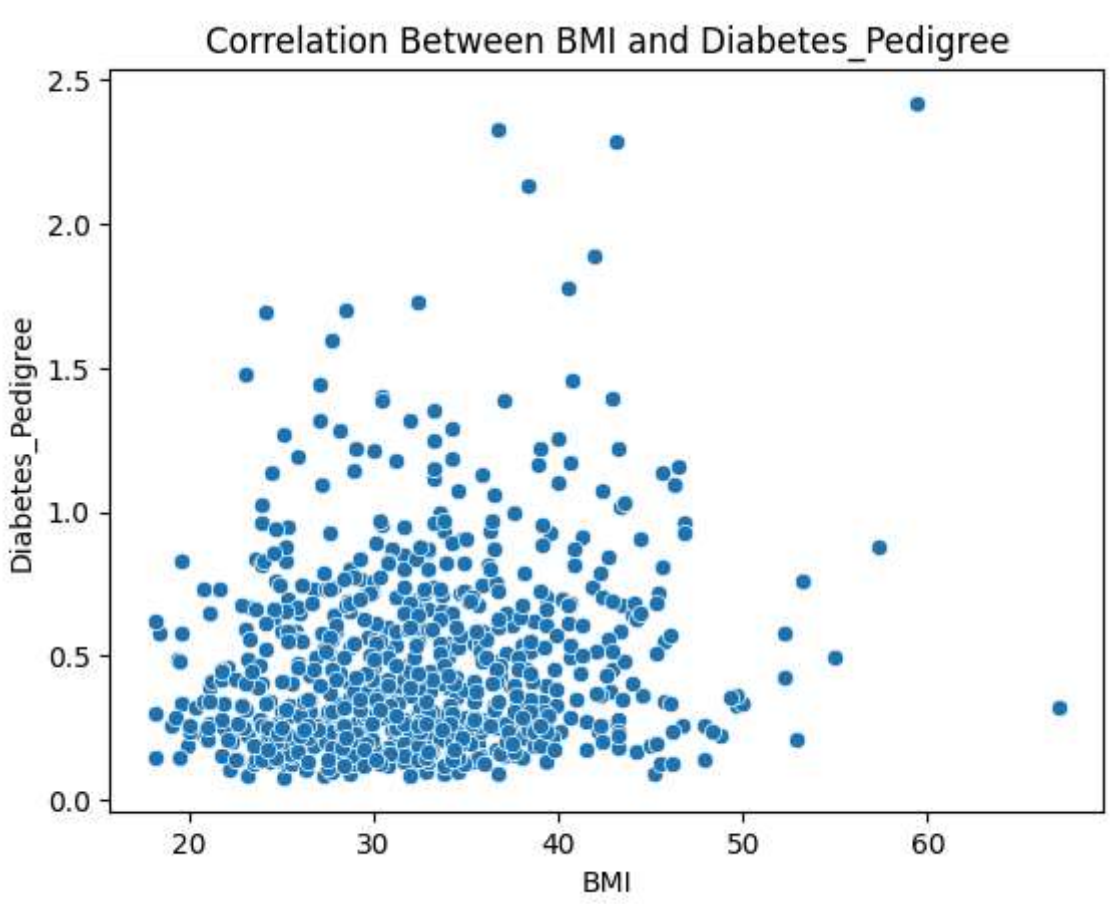
1.What is the distribution of age across individuals with Glucose in the dataset?

```
plt.bar(df['Age'],df['Glucose'])
plt.xlabel('Age')
plt.ylabel('Glucose')
plt.title('Distribution of Age Across Individuals with Glucose')
plt.show()
```



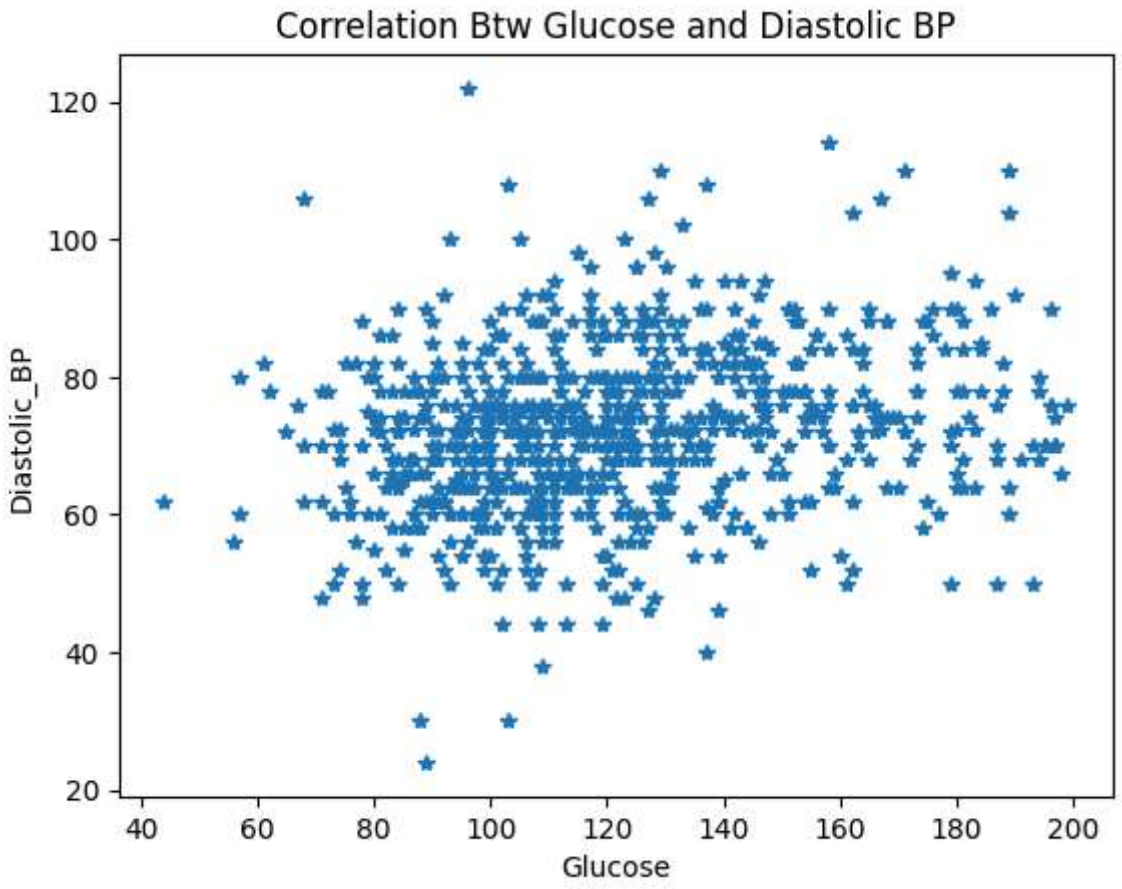
2.What is the correlation between body mass index (BMI) and Diabetes_Pedigree?

```
sns.scatterplot(x='BMI',y='Diabetes_Pedigree',data=df)
plt.xlabel('BMI')
plt.ylabel('Diabetes_Pedigree')
plt.title('Correlation Between BMI and Diabetes_Pedigree')
plt.show()
```



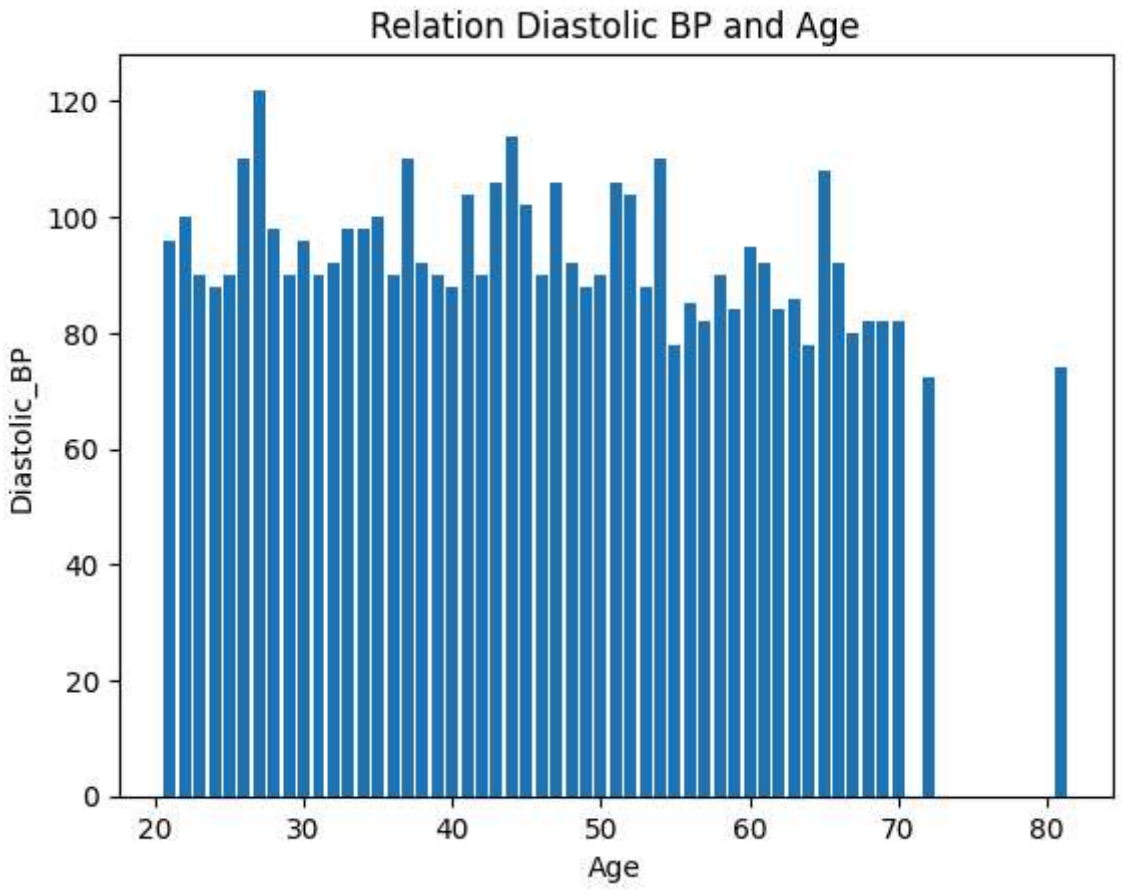
3.What correlation Btw Glucose and Diastolic BP

```
plt.plot(df['Glucose'],df['Diastolic_BP'],'*')
plt.xlabel('Glucose')
plt.ylabel('Diastolic_BP')
plt.title('Correlation Btw Glucose and Diastolic BP')
plt.show()
```



4.What was the Relation Diastolic BP and Age

```
plt.bar(df['Age'],df['Diastolic_BP'])
plt.xlabel('Age')
plt.ylabel('Diastolic_BP')
plt.title('Relation Diastolic BP and Age')
plt.show()
```



5.Relation Btw Pregnant and Glucose

```
plt.axes(projection='3d')
plt.scatter(df['Pregnant'],df['Glucose'],df['Age'])
plt.xlabel('Pregnant')
plt.ylabel('Glucose')
plt.title('Relation Btw Pregnant and Glucose')
plt.show()
```

6.Relation Between Pregnant and Diastolic BP

```
plt.axes(projection='3d')
plt.scatter(df['Pregnant'],df['Diastolic_BP'],df['Age'])
plt.xlabel('Pregnant')
plt.ylabel('Diastolic_BP')
plt.title('Relation Between Pregnant and Diastolic BP')
plt.show()
```