

```
In [61]: #Name:Khushi Chandrashekhar Satpute
# Aim : To perform and Data analysis with Co-relation Matrix
#Roll no:43
#Section: B
#Subject: ET-II
```

```
In [63]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
```

```
In [65]: os.getcwd()
```

```
Out[65]: 'C:\\Users\\asus\\Downloads'
```

```
In [67]: os.chdir("C:\\Users\\asus\\Downloads")
```

```
In [69]: data=pd.read_csv("diabetes.csv")
```

```
In [71]: data.head()
```

```
Out[71]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

```
In [73]: data.tail()
```

```
Out[73]:
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

```
In [75]: data.shape
```

```
Out[75]: (768, 9)
```

```
In [77]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Pregnancies           768 non-null   int64
1   Glucose               768 non-null   int64
2   BloodPressure         768 non-null   int64
3   SkinThickness         768 non-null   int64
4   Insulin               768 non-null   int64
5   BMI                   768 non-null   float64
6   DiabetesPedigreeFunction 768 non-null   float64
7   Age                   768 non-null   int64
8   Outcome               768 non-null   int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
In [79]: data.isna()
```

Out[79]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
763	False	False	False	False	False	False	False	False	False
764	False	False	False	False	False	False	False	False	False
765	False	False	False	False	False	False	False	False	False
766	False	False	False	False	False	False	False	False	False
767	False	False	False	False	False	False	False	False	False

768 rows × 9 columns

In [81]:

data.isna().any()

Out[81]:

Pregnancies

Glucose

BloodPressure

SkinThickness

Insulin

BMI

DiabetesPedigreeFunction

Age

Outcome

dtype: bool

False

False

False

False

False

False

False

False

False

In [83]:

data.isna().sum()

Out[83]:

Pregnancies

Glucose

BloodPressure

SkinThickness

Insulin

BMI

DiabetesPedigreeFunction

Age

Outcome

dtype: int64

0

0

0

0

0

0

0

0

0

In [85]:

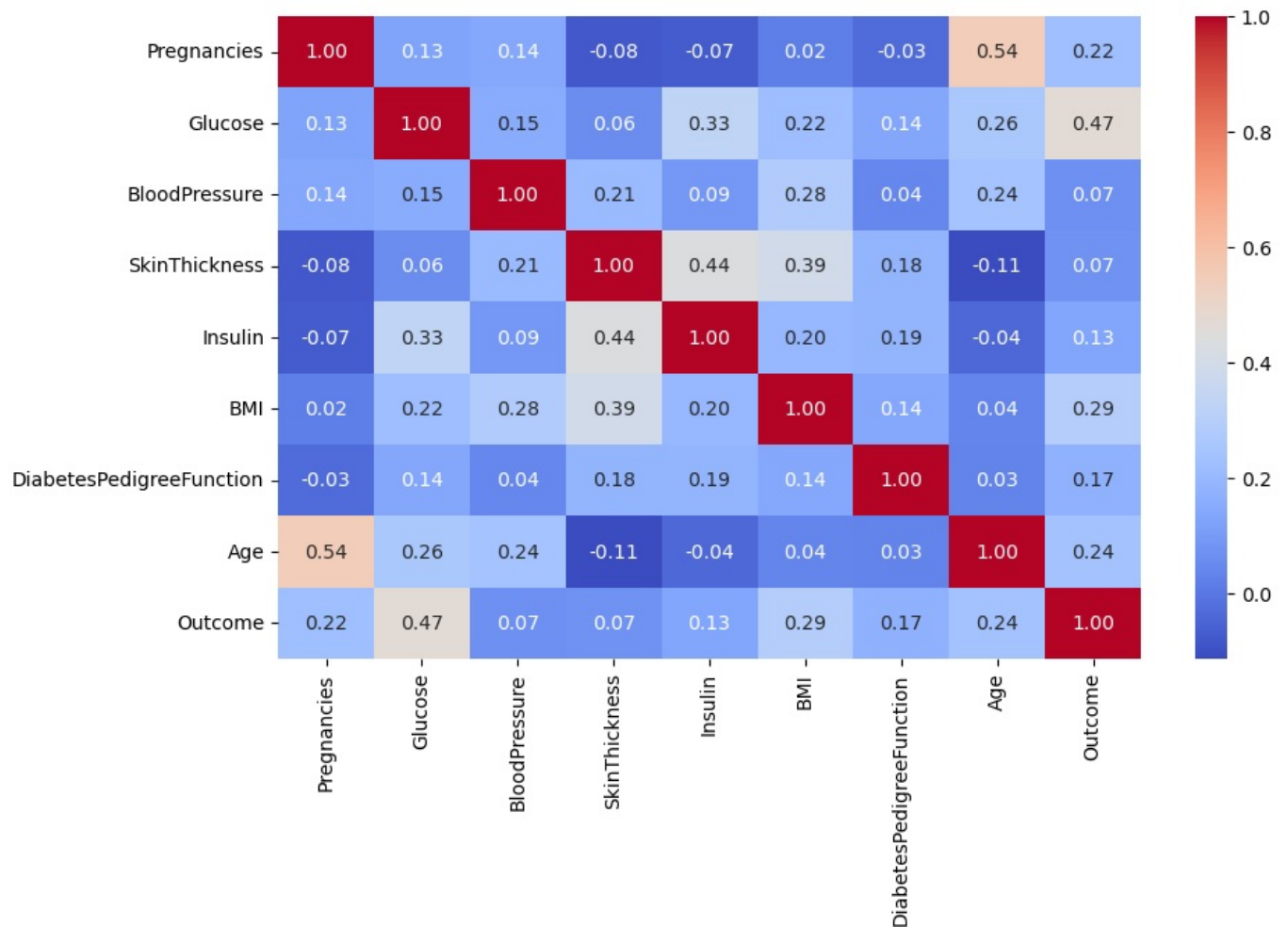
data.corr()

Out[85]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	
Pregnancies	1.000000	0.129459	0.141282	-0.081672	-0.073535	0.017683	-0.033523	0
Glucose	0.129459	1.000000	0.152590	0.057328	0.331357	0.221071	0.137337	0
BloodPressure	0.141282	0.152590	1.000000	0.207371	0.088933	0.281805	0.041265	0
SkinThickness	-0.081672	0.057328	0.207371	1.000000	0.436783	0.392573	0.183928	-0
Insulin	-0.073535	0.331357	0.088933	0.436783	1.000000	0.197859	0.185071	-0
BMI	0.017683	0.221071	0.281805	0.392573	0.197859	1.000000	0.140647	0
DiabetesPedigreeFunction	-0.033523	0.137337	0.041265	0.183928	0.185071	0.140647	1.000000	0
Age	0.544341	0.263514	0.239528	-0.113970	-0.042163	0.036242	0.033561	1
Outcome	0.221898	0.466581	0.065068	0.074752	0.130548	0.292695	0.173844	0

In [87]:

plt.figure(figsize=(10,6))  
sns.heatmap(data.corr(),annot=True,cmap='coolwarm', fmt='.2f')  
plt.show()



```
In [88]: sns.distplot(data,bins=20)#Make new notebook for normal distribution include this in that notebook
plt.show()
```

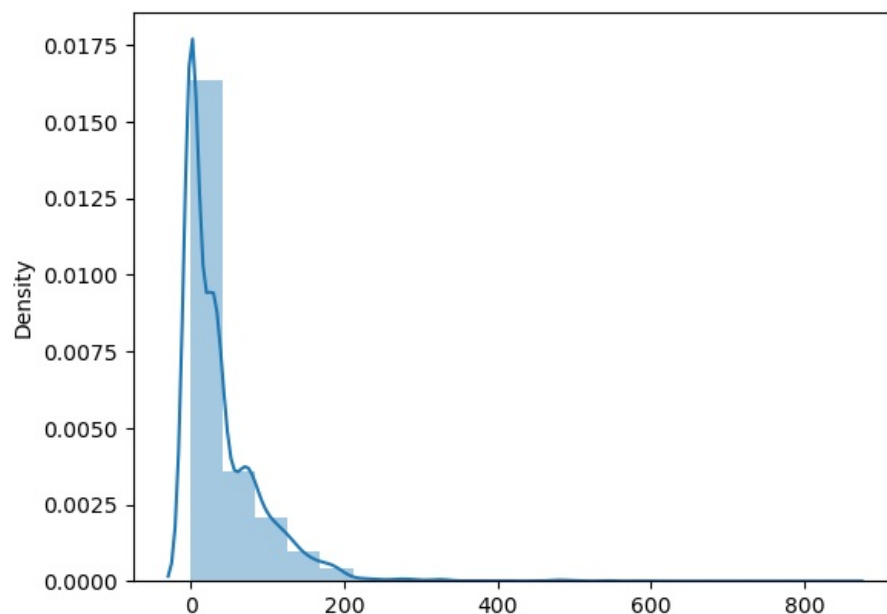
C:\Users\asus\AppData\Local\Temp\ipykernel\_17044\2386004695.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(data,bins=20)#Make new notebook for normal distribution include this in that notebook
```



In [ ]:

In [ ]: