In [1]:
```
!python -m pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\manishaa\appdata\local\programs\pyth
on\python311\lib\site-packages (3.0.0)
Requirement already satisfied: numpy>=1.26.0 in c:\users\manishaa\appdata\local\progra
ms\python\python311\lib\site-packages (from pandas) (2.4.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\manishaa\appdata\roa
ming\python\python311\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: tzdata in c:\users\manishaa\appdata\local\programs\pyth
on\python311\lib\site-packages (from pandas) (2025.3)
Requirement already satisfied: six>=1.5 in c:\users\manishaa\appdata\roaming\python\py
thon311\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
[notice] A new release of pip is available: 24.0 -> 26.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [2]:
```
!python -m pip install scikit-learn
```

```
Requirement already satisfied: scikit-learn in c:\users\manishaa\appdata\local\program
s\python\python311\lib\site-packages (1.8.0)
Requirement already satisfied: numpy>=1.24.1 in c:\users\manishaa\appdata\local\progra
ms\python\python311\lib\site-packages (from scikit-learn) (2.4.2)
Requirement already satisfied: scipy>=1.10.0 in c:\users\manishaa\appdata\local\progra
ms\python\python311\lib\site-packages (from scikit-learn) (1.17.0)
Requirement already satisfied: joblib>=1.3.0 in c:\users\manishaa\appdata\local\progra
ms\python\python311\lib\site-packages (from scikit-learn) (1.5.3)
Requirement already satisfied: threadpoolctl>=3.2.0 in c:\users\manishaa\appdata\local
\programs\python\python311\lib\site-packages (from scikit-learn) (3.6.0)
[notice] A new release of pip is available: 24.0 -> 26.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [3]:
```
!python -m pip install seaborn
```

```
Collecting seaborn
  Using cached seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\manishaa\appdata\local
\programs\python\python311\lib\site-packages (from seaborn) (2.4.2)
Requirement already satisfied: pandas>=1.2 in c:\users\manishaa\appdata\local\programs
\python\python311\lib\site-packages (from seaborn) (3.0.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\manishaa\appdata\lo
cal\programs\python\python311\lib\site-packages (from seaborn) (3.10.8)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\manishaa\appdata\local\pro
grams\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.
3)
Requirement already satisfied: cycler>=0.10 in c:\users\manishaa\appdata\local\program
s\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\manishaa\appdata\local\pr
ograms\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.6
1.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\manishaa\appdata\local\pr
ograms\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.
4.9)
Requirement already satisfied: packaging>=20.0 in c:\users\manishaa\appdata\roaming\py
thon\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (26.0)
Requirement already satisfied: pillow>=8 in c:\users\manishaa\appdata\local\programs\p
ython\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (12.1.0)
Requirement already satisfied: pyparsing>=3 in c:\users\manishaa\appdata\local\program
s\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.3.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\manishaa\appdata\roami
ng\python\python311\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post
0)
Requirement already satisfied: tzdata in c:\users\manishaa\appdata\local\programs\pyth
on\python311\lib\site-packages (from pandas>=1.2->seaborn) (2025.3)
Requirement already satisfied: six>=1.5 in c:\users\manishaa\appdata\roaming\python\py
thon311\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn)
(1.17.0)
Using cached seaborn-0.13.2-py3-none-any.whl (294 kB)
Installing collected packages: seaborn
Successfully installed seaborn-0.13.2
[notice] A new release of pip is available: 24.0 -> 26.0
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [4]:
```python
import pandas as pd

import numpy as np
from sklearn.preprocessing import MinMaxScaler, StandardScaler
import seaborn as sns
import matplotlib.pyplot as plt
```

In [5]:
```python
df = pd.read_csv("diabetes.csv")
df.head()
```

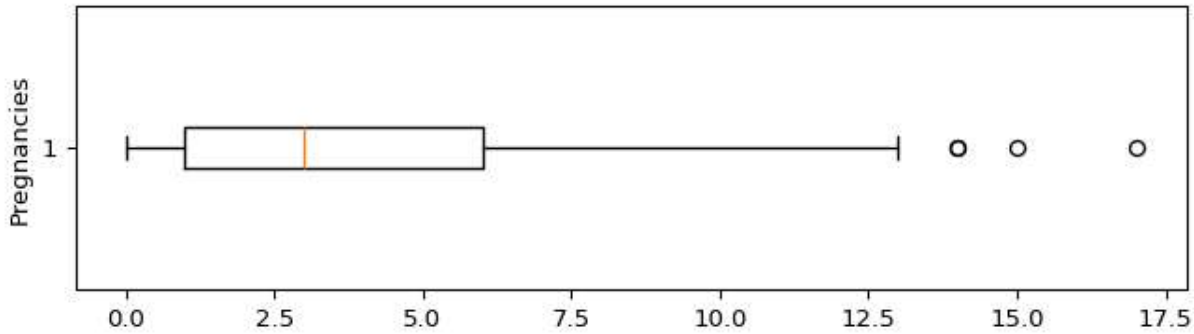
Out[5]:

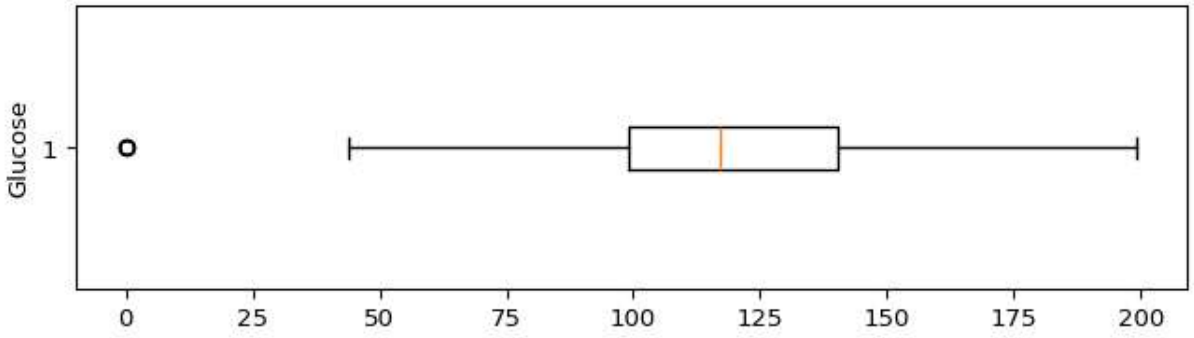
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFun
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	

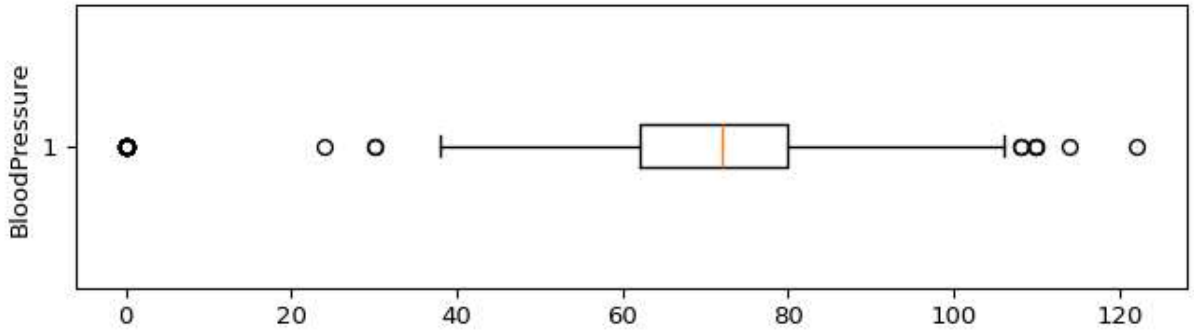
In [6]:
```
df.info()
print(df.isnull().sum())
```

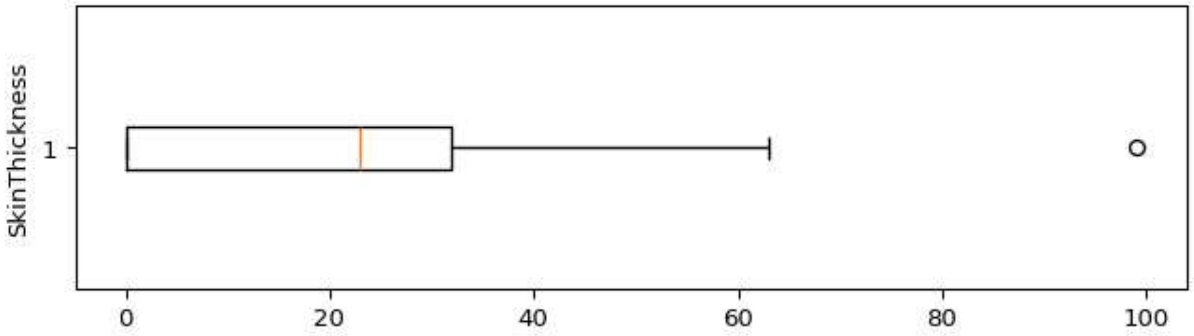
```
<class 'pandas.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    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
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

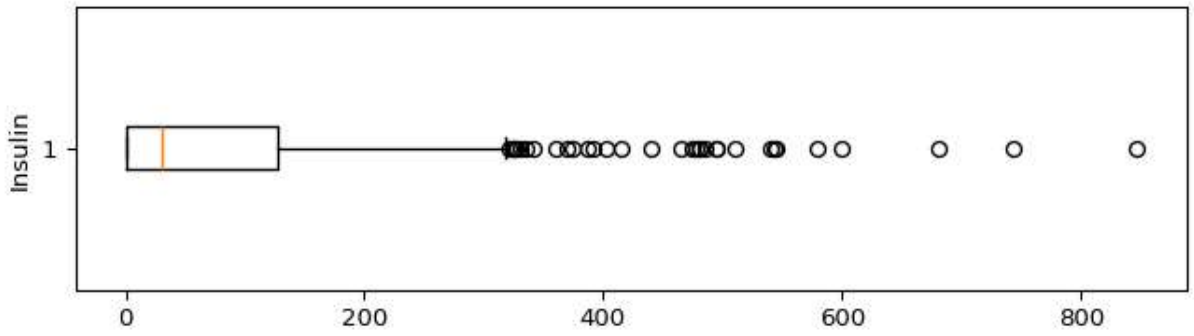
In [8]:
```
df.describe()

fig, axs = plt.subplots(len(df.columns), 1, figsize=(7, 18), dpi=95)
for i, col in enumerate(df.columns):
    axs[i].boxplot(df[col], vert=False)
    axs[i].set_ylabel(col)
plt.tight_layout()
plt.show()
```

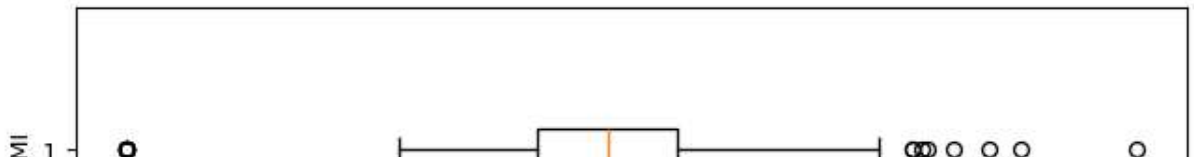

Pregnancies
1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

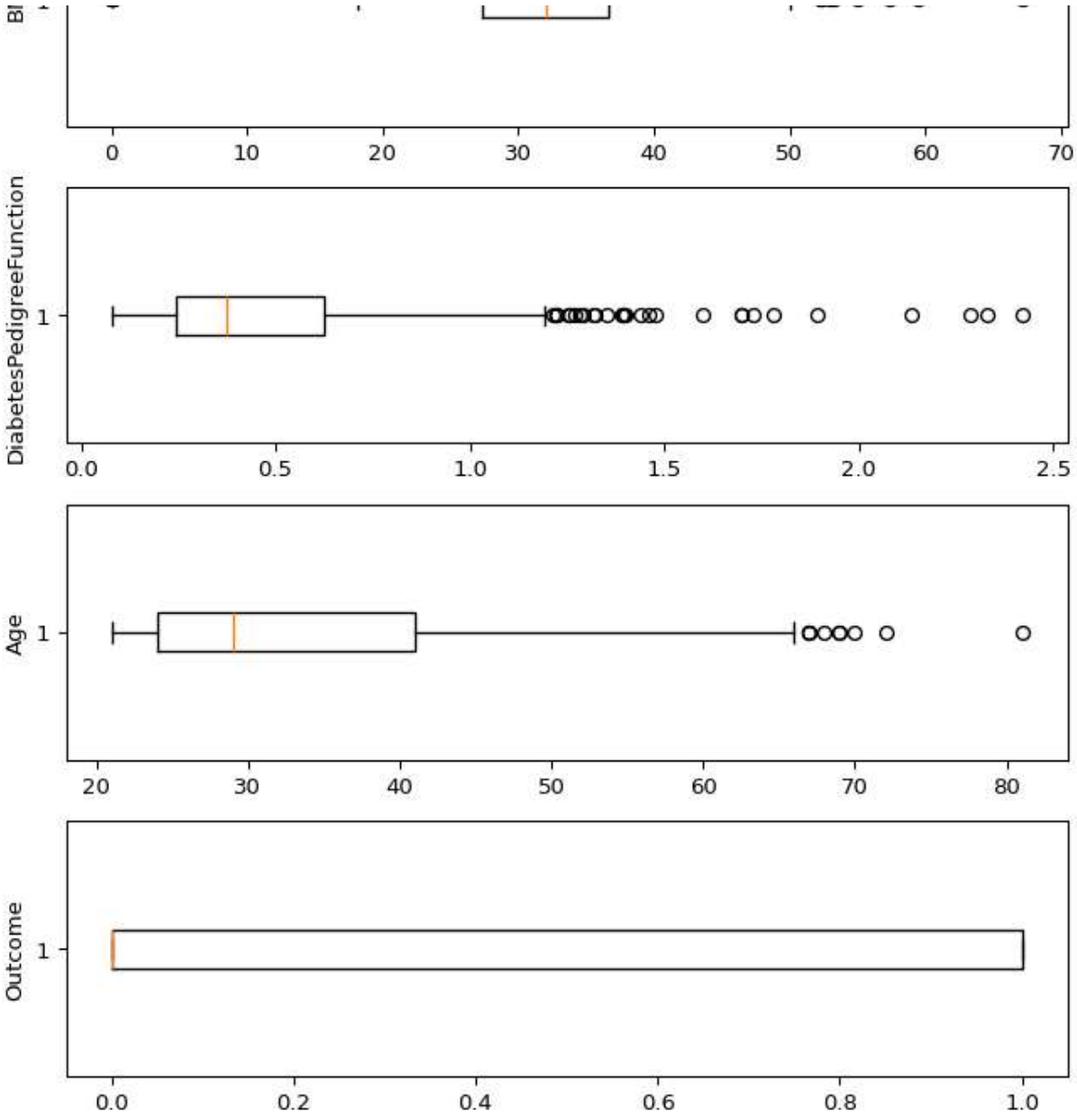

Glucose
1
0
25
50
75
100
125
150
175
200


BloodPressure
1
0
20
40
60
80
100
120


SkinThickness
1
0
20
40
60
80
100

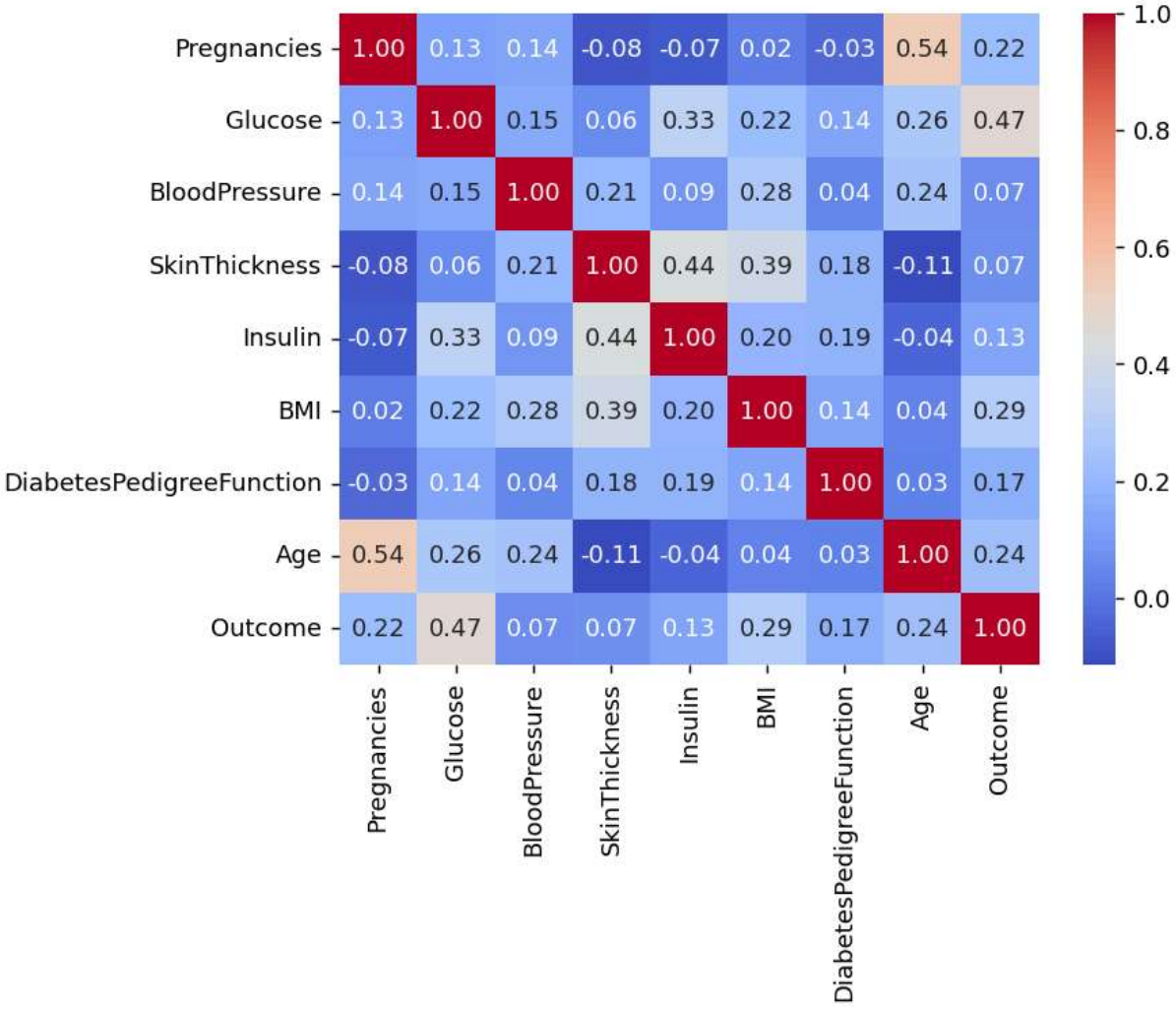

Insulin
1
0
200
400
600
800


1



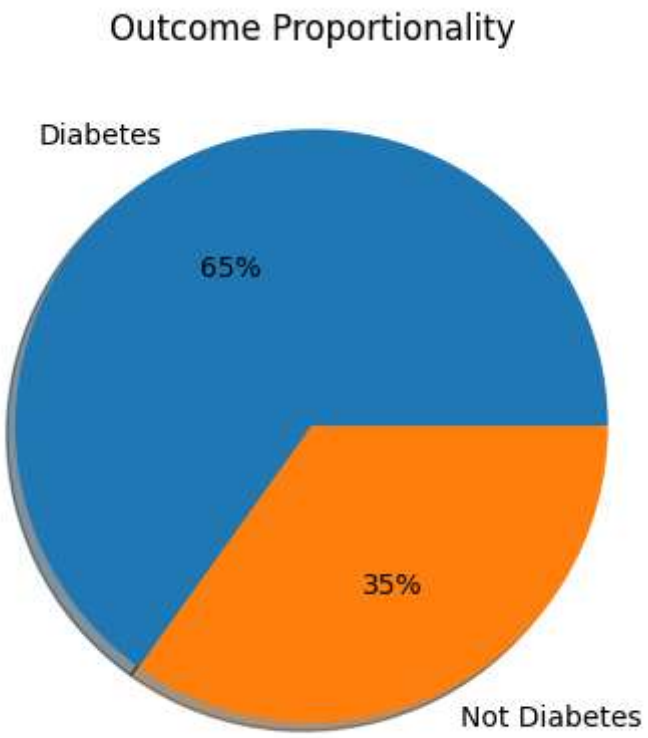
In [9]:
```
q1, q3 = np.percentile(df['Insulin'], [25, 75])
iqr = q3 - q1
lower = q1 - 1.5 * iqr
upper = q3 + 1.5 * iqr
clean_df = df[(df['Insulin'] >= lower) & (df['Insulin'] <= upper)]
```

In [10]:
```
corr = df.corr()
plt.figure(dpi=130)
sns.heatmap(corr, annot=True, fmt='.2f', cmap='coolwarm')
plt.show()
print(corr['Outcome'].sort_values(ascending=False))
```




```
Outcome                    1.000000
Glucose                    0.466581
BMI                        0.292695
Age                        0.238356
Pregnancies                0.221898
DiabetesPedigreeFunction   0.173844
Insulin                    0.130548
SkinThickness              0.074752
BloodPressure              0.065068
Name: Outcome, dtype: float64
```

In [11]:
```
plt.pie(df['Outcome'].value_counts(), labels=[
'Diabetes', 'Not Diabetes'], autopct='%.f%%', shadow=True)
plt.title('Outcome Proportionality')
plt.show()
```



In [12]:
```
X = df.drop(columns=['Outcome'])
y = df['Outcome']
```

In [13]:
```
scaler = MinMaxScaler()
X_normalized = scaler.fit_transform(X)
print(X_normalized[:5])
```

```
[[0.35294118 0.74371859 0.59016393 0.35353535 0.         0.50074516
  0.23441503 0.48333333]
 [0.05882353 0.42713568 0.54098361 0.29292929 0.         0.39642325
  0.11656704 0.16666667]
 [0.47058824 0.91959799 0.52459016 0.         0.         0.34724292
  0.25362938 0.18333333]
 [0.05882353 0.44723618 0.54098361 0.23232323 0.11111111 0.41877794
  0.03800171 0.        ]
 [0.         0.68844221 0.32786885 0.35353535 0.19858156 0.64232489
  0.94363792 0.2       ]]
```

In [ ]: