Run the following code to import the data :

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.datasets import load_diabetes

data = load_diabetes()
df = pd.DataFrame(data['data'], columns=data['feature_names'])
df['target'] = data['target']
```

Your data is now in the variable df

The task:

1. Do basic descriptive analytics on the data
2. Write code to find percentage of Na values in each column
3. Convert the variable sex to a categorical variable where -0.044642 : 0 and 0.050680 is 1
4. Check for normal distribution of all variables and check for a linear relationship
5. Check for Multicollinearity and find out the columns that are collinear
6. Plot a linear regression model with BMI as the X variable and Target as the Y variable
7. Run a linear regression model on the data with ‘target’ as the dependent variable.
8. Find out the relationship between the correlation of all independent variables with the target variable and the coefficients of the same variables.
9. Run a separate Regression model for data with No collinearity(Remove variables with >5 vif) and compare results with the regression model with all the data
10. Print out MAE,MSE,RMSE,r2_score values for the above models and compare