```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
df=pd.read_csv('/content/Untitled spreadsheet – Sheet1.csv')
print(df)
print(df.isnull().sum())
df_cleaned = df.dropna()
print(df_cleaned)
#Numeric Descriptors
#Mean of Glycan Molecules
mean_glycan_concentration = df.groupby('Glycan')['Glycan Concentration'].mean().reset_index() #mean
print(mean_glycan_concentration )
#Maximum Glycan Concentration (Peak)
peak_glycan_concentration = df.groupby('Glycan')['Glycan Concentration'].max().reset_index(name='Peak Glycan Concentra
print(peak_glycan_concentration)
#Area Under the Curve(AUC)
def calculate_auc(glycan_df): #AUC
    auc = np.trapz(glycan_df['Glycan Concentration'], glycan_df['Time (hours)'])
    return auc
auc_values = df.groupby('Glycan').apply(calculate_auc).reset_index(name='AUC')
print(auc_values)
#Combine the descriptors into a single DataFrame
descriptors = pd.merge(mean_glycan_concentration, auc_values, on='Glycan')
descriptors = pd.merge(descriptors, peak_glycan_concentration, on='Glycan')
print(descriptors)
```

```
      Site    Glycan  Count1  Count2  Count3  Count4  subject  Time (hours)  \
0        A  Glycan19       0       0       0       0     1001             1
1        A   Glycan8       3       3       0       0     1001             1
2        A   Glycan6       2       2       1       0     1001             1
3        A   Glycan9       3       3       1       0     1001             1
4        A   Glycan7       2       2       2       0     1001             1
...    ...       ...     ...     ...     ...     ...      ...           ...
1867     C  Glycan16       4       4       2       0     1006           168
1868     C  Glycan17       4       4       3       0     1006           168
1869     C  Glycan18       4       4       4       0     1006           168
1870     C  Glycan13       4       4       3       1     1006           168
1871     C  Glycan19       0       0       0       0     1006           168

      Area Percent  Protein Concentration  Glycan Concentration
0              3.1                  87500                2712.5
1              4.7                  87500                4112.5
2              3.0                  87500                2625.0
3              8.2                  87500                7175.0
4              6.4                  87500                5600.0
...            ...                    ...                   ...
1867           6.5                  31600                2054.0
1868          13.3                  31600                4202.8
1869           2.7                  31600                 853.2
1870           3.5                  31600                1106.0
1871          12.7                  31600                4013.2

[1872 rows x 11 columns]
Site                     0
Glycan                   0
Count1                   0
Count2                   0
Count3                   0
Count4                   0
subject                  0
Time (hours)             0
Area Percent             0
Protein Concentration    0
Glycan Concentration     0
dtype: int64
      Site    Glycan  Count1  Count2  Count3  Count4  subject  Time (hours)  \
0        A  Glycan19       0       0       0       0     1001             1
1        A   Glycan8       3       3       0       0     1001             1
2        A   Glycan6       2       2       1       0     1001             1
3        A   Glycan9       3       3       1       0     1001             1
4        A   Glycan7       2       2       2       0     1001             1
...    ...       ...     ...     ...     ...     ...      ...           ...
1867     C  Glycan16       4       4       2       0     1006           168
1868     C  Glycan17       4       4       3       0     1006           168
1869     C  Glycan18       4       4       4       0     1006           168
1870     C  Glycan13       4       4       3       1     1006           168
1871     C  Glycan19       0       0       0       0     1006           168
```

```
   Area Percent  Protein Concentration  Glycan Concentration
0           3.1                  87500                2712.5
1           4.7                  87500                4112.5
2           3.0                  87500                2625.0
3           8.2                  87500                7175.0
4           6.4                  87500                5600.0
```