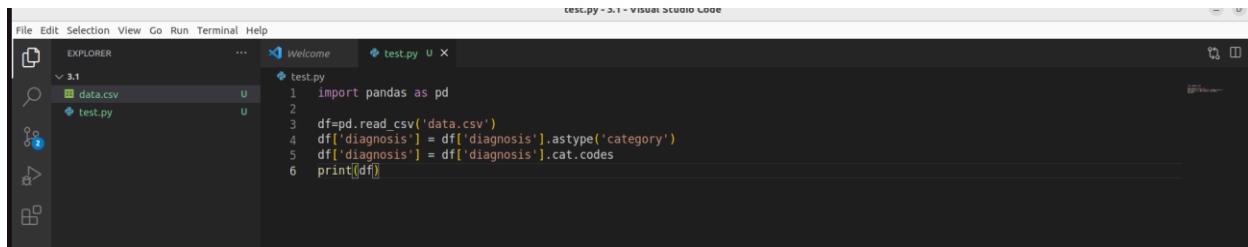# Assignment 3.1

**Arshad Shiwani (2303.KHI.DEG.026)**

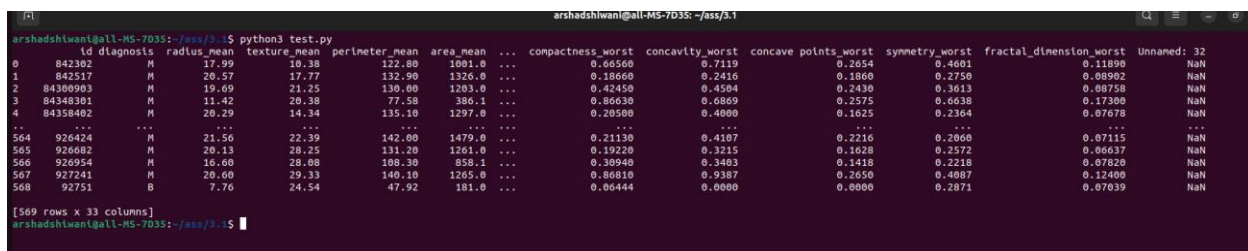**Syed Muhammad Hammad Irshad(2303.KHI.DEG.032)**

**Code:**

```python
import pandas as pd


df=pd.read_csv('data.csv')
df['diagnosis'] = df['diagnosis'].astype('category')
df['diagnosis'] = df['diagnosis'].cat.codes
print(df)
```



Before encode diagnosis:



After encode  diagnosis

```
arshadshiwani@all-MS-7D35:~/ass/3.1$ python3 test.py
           id  diagnosis  radius_mean  texture_mean  perimeter_mean  area_mean  ...  compactness_worst  concavity_worst  concave points_worst  symmetry_worst  fractal_dimension_worst  Unnamed: 32
0      842302          1        17.99         10.38          122.80     1001.0  ...            0.66560           0.7119                0.2654          0.4601                  0.11890          NaN
1      842517          1        20.57         17.77          132.90     1326.0  ...            0.18660           0.2416                0.1860          0.2750                  0.08902          NaN
2    84300903          1        19.69         21.25          130.00     1203.0  ...            0.42450           0.4504                0.2430          0.3613                  0.08758          NaN
3    84348301          1        11.42         20.38           77.58      386.1  ...            0.86630           0.6869                0.2575          0.6638                  0.17300          NaN
4    84358402          1        20.29         14.34          135.10     1297.0  ...            0.20500           0.4000                0.1625          0.2364                  0.07678          NaN
..        ...        ...          ...           ...             ...        ...  ...                ...              ...                   ...             ...                      ...          ...
564    926424          1        21.56         22.39          142.00     1479.0  ...            0.21130           0.4107                0.2216          0.2060                  0.07115          NaN
565    926682          1        20.13         28.25          131.20     1261.0  ...            0.19220           0.3215                0.1628          0.2572                  0.06637          NaN
566    926954          1        16.60         28.08          108.30      858.1  ...            0.30940           0.3403                0.1418          0.2218                  0.07820          NaN
567    927241          1        20.60         29.33          140.10     1265.0  ...            0.86810           0.9387                0.2650          0.4087                  0.12400          NaN
568     92751          0         7.76         24.54           47.92      181.0  ...            0.06444           0.0000                0.0000          0.2871                  0.07039          NaN

[569 rows x 33 columns]
arshadshiwani@all-MS-7D35:~/ass/3.1$
```