In [103…

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
df= pd.read_csv(r'C:\Users\PC-chetan\Desktop\data files\data.csv')
```

In [104…

```
df
```

Out[104…

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_m
0	842302	M	17.99	10.38	122.80	1001.0	0.11
1	842517	M	20.57	17.77	132.90	1326.0	0.08
2	84300903	M	19.69	21.25	130.00	1203.0	0.10
3	84348301	M	11.42	20.38	77.58	386.1	0.14
4	84358402	M	20.29	14.34	135.10	1297.0	0.10
...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	0.11
565	926682	M	20.13	28.25	131.20	1261.0	0.09
566	926954	M	16.60	28.08	108.30	858.1	0.08
567	927241	M	20.60	29.33	140.10	1265.0	0.11
568	92751	B	7.76	24.54	47.92	181.0	0.05

569 rows × 33 columns

Exploratery Data Analysis

In [38]:

```
df.info ()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 33 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   id                       569 non-null    int64
 1   diagnosis                569 non-null    object
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
     _se                      569 non-null    float64
```

```
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
 32  Unnamed: 32              0 non-null      float64
dtypes: float64(31), int64(1), object(1)
memory usage: 146.8+ KB
```

In [39]:
```
df.isnull().sum()
```

Out[39]:
```
id                         0
diagnosis                  0
radius_mean                0
texture_mean               0
perimeter_mean             0
area_mean                  0
smoothness_mean            0
compactness_mean           0
concavity_mean             0
concave points_mean        0
symmetry_mean              0
fractal_dimension_mean     0
radius_se                  0
texture_se                 0
perimeter_se               0
area_se                    0
smoothness_se              0
compactness_se             0
concavity_se               0
concave points_se          0
symmetry_se                0
fractal_dimension_se       0
radius_worst               0
texture_worst              0
perimeter_worst            0
area_worst                 0
smoothness_worst           0
compactness_worst          0
concavity_worst            0
concave points_worst       0
symmetry_worst             0
fractal_dimension_worst    0
Unnamed: 32                569
dtype: int64
```

In [44]:
```
df.shape
```

Out[44]:
```
(569, 33)
```

In [41]:
```
df.head(10)
```

Out[41]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
0	842302	M	17.99	10.38	122.80	1001.0	0.1184
1	842517	M	20.57	17.77	132.90	1326.0	0.0847
2	84300903	M	19.69	21.25	130.00	1203.0	0.1096
3	84348301	M	11.42	20.38	77.58	386.1	0.1425
4	84358402	M	20.29	14.34	135.10	1297.0	0.1003
5	843786	M	12.45	15.70	82.57	477.1	0.1278
6	844359	M	18.25	19.98	119.60	1040.0	0.0946
7	84458202	M	13.71	20.83	90.20	577.9	0.1189
8	844981	M	13.00	21.82	87.50	519.8	0.1273
9	84501001	M	12.46	24.04	83.97	475.9	0.1186

10 rows × 33 columns

In [42]:
```
df.columns
```

Out[42]:
```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
       'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
      dtype='object')
```

In [43]:
```
df.dtypes
```

Out[43]:
```
id                         int64
diagnosis                 object
radius_mean              float64
texture_mean             float64
perimeter_mean           float64
area_mean                float64
smoothness_mean          float64
compactness_mean         float64
concavity_mean           float64
concave points_mean      float64
symmetry_mean            float64
fractal_dimension_mean   float64
radius_se                float64
texture_se               float64
perimeter_se             float64
area_se                  float64
smoothness_se            float64
compactness_se           float64
concavity_se             float64
concave points_se        float64
symmetry_se              float64
fractal_dimension_se     float64
radius_worst             float64
texture_worst            float64
perimeter_worst          float64
                         float64
```

```
smoothness_worst           float64
compactness_worst          float64
concavity_worst            float64
concave points_worst       float64
symmetry_worst             float64
fractal_dimension_worst    float64
Unnamed: 32                float64
dtype: object
```

In [45]:
```
df.index
```

Out[45]:
```
RangeIndex(start=0, stop=569, step=1)
```

In [46]:
```
df.describe()
```

Out[46]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
count	5.690000e+02	569.000000	569.000000	569.000000	569.000000	569.000000
mean	3.037183e+07	14.127292	19.289649	91.969033	654.889104	0.096360
std	1.250206e+08	3.524049	4.301036	24.298981	351.914129	0.014064
min	8.670000e+03	6.981000	9.710000	43.790000	143.500000	0.052630
25%	8.692180e+05	11.700000	16.170000	75.170000	420.300000	0.086370
50%	9.060240e+05	13.370000	18.840000	86.240000	551.100000	0.095870
75%	8.813129e+06	15.780000	21.800000	104.100000	782.700000	0.105300
max	9.113205e+08	28.110000	39.280000	188.500000	2501.000000	0.163400

8 rows × 32 columns

In [47]:
```
df.describe().T
```

Out[47]:

	count	mean	std	min	25%	5
id	569.0	3.037183e+07	1.250206e+08	8670.000000	869218.000000	906024.000
radius_mean	569.0	1.412729e+01	3.524049e+00	6.981000	11.700000	13.370
texture_mean	569.0	1.928965e+01	4.301036e+00	9.710000	16.170000	18.840
perimeter_mean	569.0	9.196903e+01	2.429898e+01	43.790000	75.170000	86.240
area_mean	569.0	6.548891e+02	3.519141e+02	143.500000	420.300000	551.100
smoothness_mean	569.0	9.636028e-02	1.406413e-02	0.052630	0.086370	0.095
compactness_mean	569.0	1.043410e-01	5.281276e-02	0.019380	0.064920	0.092
concavity_mean	569.0	8.879932e-02	7.971981e-02	0.000000	0.029560	0.061
concave points_mean	569.0	4.891915e-02	3.880284e-02	0.000000	0.020310	0.033
symmetry_mean	569.0	1.811619e-01	2.741428e-02	0.106000	0.161900	0.179
fractal_dimension_mean	569.0	6.279761e-02	7.060363e-03	0.049960	0.057700	0.061
radius_se	569.0	4.051721e-01	2.773127e-01	0.111500	0.232400	0.324
texture_se	569.0	1.216853e+00	5.516484e-01	0.360200	0.833900	1.108
perimeter_se	569.0	2.866059e+00	2.021855e+00	0.757000	1.606000	2.287
ea_se	569.0	4.033708e+01	4.549101e+01	6.802000	17.850000	24.530

	count	mean	std	min	25%	5
smoothness_se	569.0	7.040979e-03	3.002518e-03	0.001713	0.005169	0.006
compactness_se	569.0	2.547814e-02	1.790818e-02	0.002252	0.013080	0.020
concavity_se	569.0	3.189372e-02	3.018606e-02	0.000000	0.015090	0.025
concave points_se	569.0	1.179614e-02	6.170285e-03	0.000000	0.007638	0.010
symmetry_se	569.0	2.054230e-02	8.266372e-03	0.007882	0.015160	0.018
fractal_dimension_se	569.0	3.794904e-03	2.646071e-03	0.000895	0.002248	0.003
radius_worst	569.0	1.626919e+01	4.833242e+00	7.930000	13.010000	14.970
texture_worst	569.0	2.567722e+01	6.146258e+00	12.020000	21.080000	25.410
perimeter_worst	569.0	1.072612e+02	3.360254e+01	50.410000	84.110000	97.660
area_worst	569.0	8.805831e+02	5.693570e+02	185.200000	515.300000	686.500
smoothness_worst	569.0	1.323686e-01	2.283243e-02	0.071170	0.116600	0.131
compactness_worst	569.0	2.542650e-01	1.573365e-01	0.027290	0.147200	0.211
concavity_worst	569.0	2.721885e-01	2.086243e-01	0.000000	0.114500	0.226
concave points_worst	569.0	1.146062e-01	6.573234e-02	0.000000	0.064930	0.099
symmetry_worst	569.0	2.900756e-01	6.186747e-02	0.156500	0.250400	0.282
fractal_dimension_worst	569.0	8.394582e-02	1.806127e-02	0.055040	0.071460	0.080
Unnamed: 32	0.0	NaN	NaN	NaN	NaN	

In [49]:
```
df.corr()
```

Out[49]:

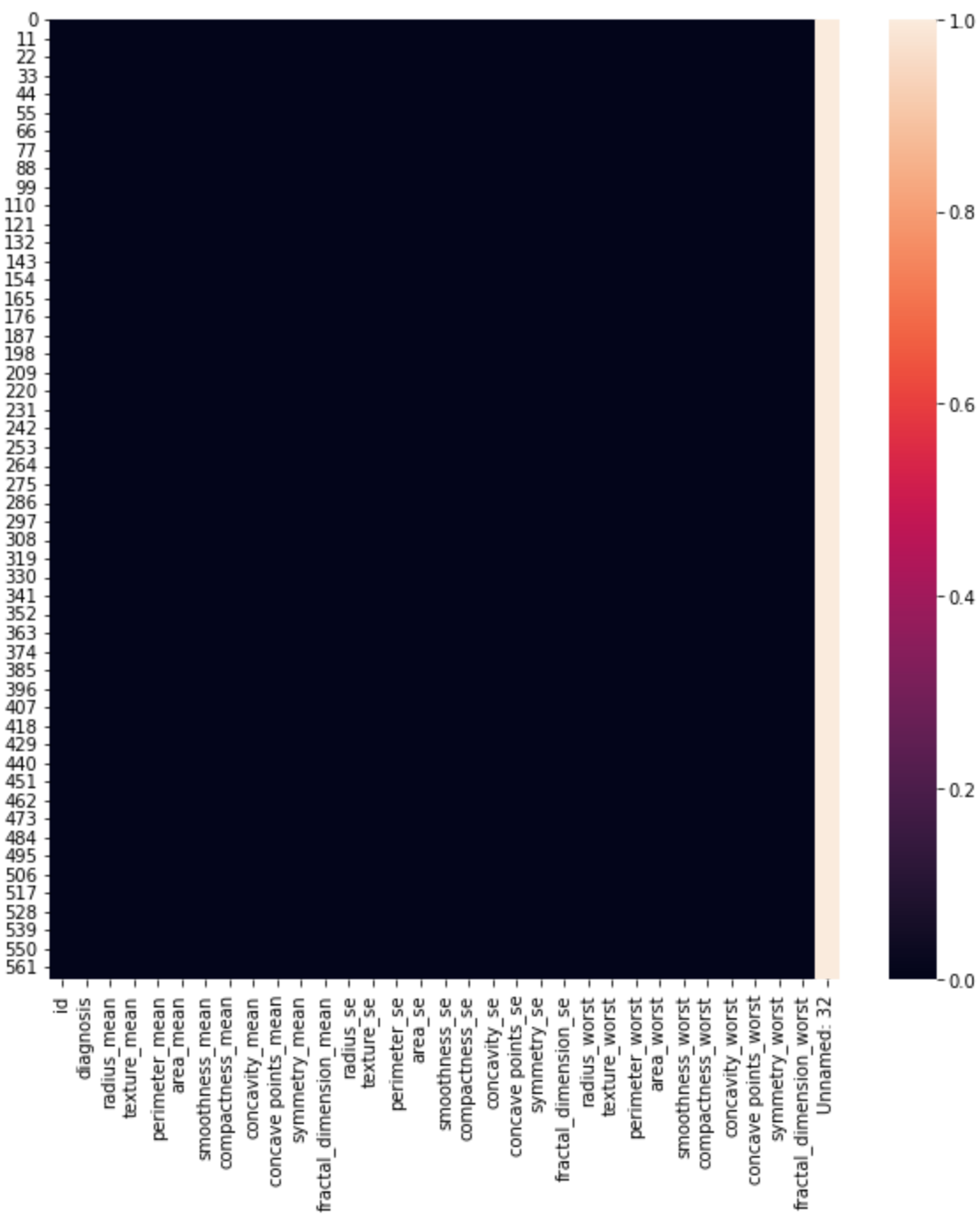
	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoot
id	1.000000	0.074626	0.099770	0.073159	0.096893	
radius_mean	0.074626	1.000000	0.323782	0.997855	0.987357	
texture_mean	0.099770	0.323782	1.000000	0.329533	0.321086	
perimeter_mean	0.073159	0.997855	0.329533	1.000000	0.986507	
area_mean	0.096893	0.987357	0.321086	0.986507	1.000000	
smoothness_mean	-0.012968	0.170581	-0.023389	0.207278	0.177028	
compactness_mean	0.000096	0.506124	0.236702	0.556936	0.498502	
concavity_mean	0.050080	0.676764	0.302418	0.716136	0.685983	
concave points_mean	0.044158	0.822529	0.293464	0.850977	0.823269	
symmetry_mean	-0.022114	0.147741	0.071401	0.183027	0.151293	
fractal_dimension_mean	-0.052511	-0.311631	-0.076437	-0.261477	-0.283110	
radius_se	0.143048	0.679090	0.275869	0.691765	0.732562	
texture_se	-0.007526	-0.097317	0.386358	-0.086761	-0.066280	
perimeter_se	0.137331	0.674172	0.281673	0.693135	0.726628	
area_se	0.177742	0.735864	0.259845	0.744983	0.800086	
smoothness_se	0.096781	-0.222600	0.006614	-0.202694	-0.166777	
compactness_se	0.033961	0.206000	0.191975	0.250744	0.212583	
concavity_se	0.055239	0.194204	0.143293	0.228082	0.207660	

	id	radius_mean	texture_mean	perimeter_mean	area_mean	smoot
concave points_se	0.078768	0.376169	0.163851	0.407217	0.372320	
symmetry_se	-0.017306	-0.104321	0.009127	-0.081629	-0.072497	
fractal_dimension_se	0.025725	-0.042641	0.054458	-0.005523	-0.019887	
radius_worst	0.082405	0.969539	0.352573	0.969476	0.962746	
texture_worst	0.064720	0.297008	0.912045	0.303038	0.287489	
perimeter_worst	0.079986	0.965137	0.358040	0.970387	0.959120	
area_worst	0.107187	0.941082	0.343546	0.941550	0.959213	
smoothness_worst	0.010338	0.119616	0.077503	0.150549	0.123523	
compactness_worst	-0.002968	0.413463	0.277830	0.455774	0.390410	
concavity_worst	0.023203	0.526911	0.301025	0.563879	0.512606	
concave points_worst	0.035174	0.744214	0.295316	0.771241	0.722017	
symmetry_worst	-0.044224	0.163953	0.105008	0.189115	0.143570	
fractal_dimension_worst	-0.029866	0.007066	0.119205	0.051019	0.003738	
Unnamed: 32	NaN	NaN	NaN	NaN	NaN	

32 rows × 32 columns

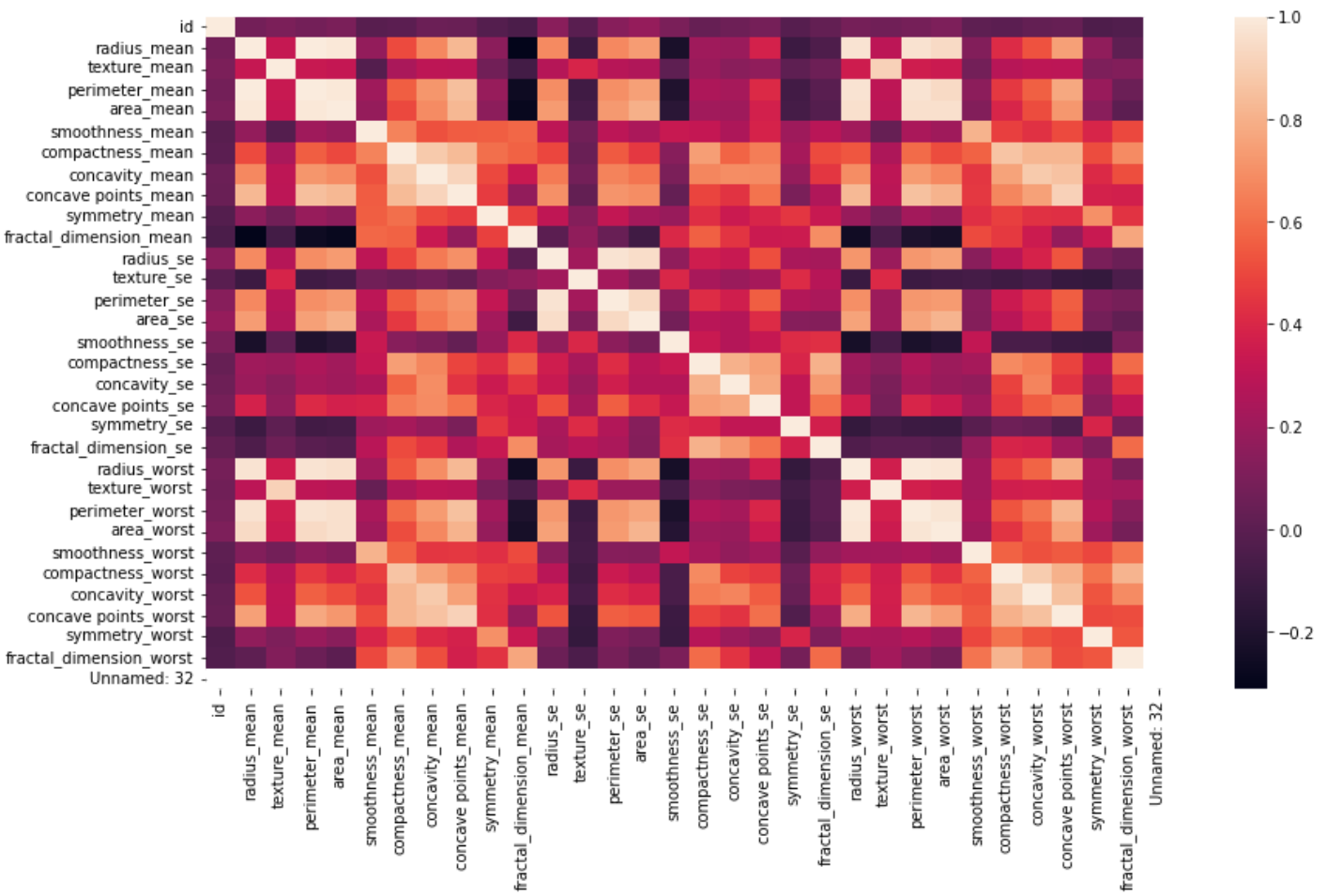
In [52]:

```
plt.figure(figsize=(10,10))
sns.heatmap(df.isnull())
plt.show()
```



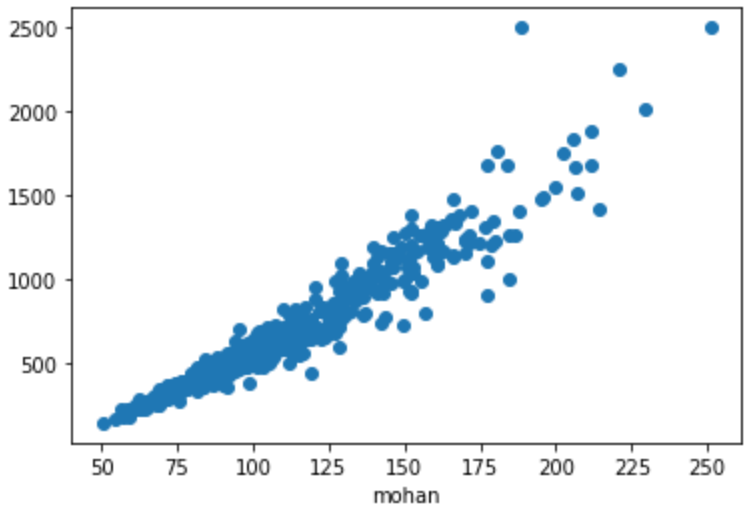
In [57]:

```
plt.figure(figsize=(14,8))
sns.heatmap(df.corr())
plt.show()
```



In [60]:

```
plt.scatter(df['perimeter_worst'], df['area_mean'])
plt.xlabel('mohan')
plt.show()
```

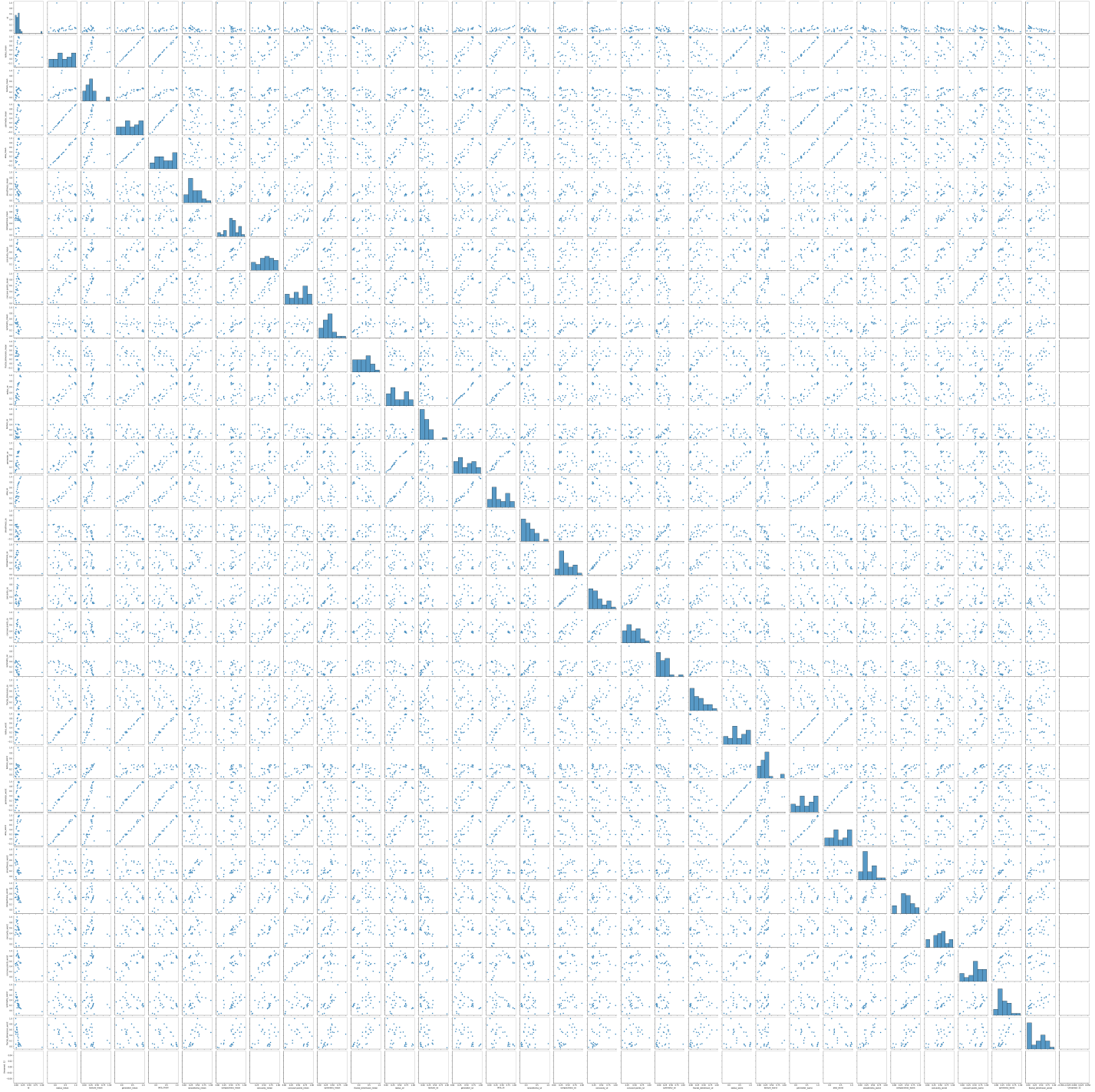


In [61]:

```
sns.pairplot(df.corr())
```

Out[61]:

```
<seaborn.axisgrid.PairGrid at 0x1e776124f40>
```



In [62]:
```
df
```

Out[62]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_m
0	842302	M	17.99	10.38	122.80	1001.0	0.11
1	842517	M	20.57	17.77	132.90	1326.0	0.08
2	84300903	M	19.69	21.25	130.00	1203.0	0.10
3	84348301	M	11.42	20.38	77.58	386.1	0.14
4	84358402	M	20.29	14.34	135.10	1297.0	0.10
...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	0.11
565	926682	M	20.13	28.25	131.20	1261.0	0.09
566	926954	M	16.60	28.08	108.30	858.1	0.08

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_m
567	927241	M	20.60	29.33	140.10	1265.0	0.11
568	92751	B	7.76	24.54	47.92	181.0	0.05

569 rows × 33 columns

In [105…
```
df.drop(['Unnamed: 32', 'id'] , axis = 1,inplace= True)
```

In [106…
```
df
```

Out[106…

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compa
0	M	17.99	10.38	122.80	1001.0	0.11840	
1	M	20.57	17.77	132.90	1326.0	0.08474	
2	M	19.69	21.25	130.00	1203.0	0.10960	
3	M	11.42	20.38	77.58	386.1	0.14250	
4	M	20.29	14.34	135.10	1297.0	0.10030	
...	...	...	...	...	...	...	
564	M	21.56	22.39	142.00	1479.0	0.11100	
565	M	20.13	28.25	131.20	1261.0	0.09780	
566	M	16.60	28.08	108.30	858.1	0.08455	
567	M	20.60	29.33	140.10	1265.0	0.11780	
568	B	7.76	24.54	47.92	181.0	0.05263	

569 rows × 31 columns

In [67]:
```
type_1 = df[df['diagnosis'] == 'M']
```

In [68]:
```
type_2 = df[df['diagnosis'] == 'B']
```

In [79]:
```
plt.scatter(type_2.radius_mean, type_2.texture_mean, c = 'r', label = 'keto', alpha=0.4)
plt.scatter(type_1.radius_mean, type_1.texture_mean, c = 'g', label = 'iyi', alpha= 0.5)
plt.xlabel = ('boby')
plt.ylabel = ('goby')
plt.title('grapgh')
plt.legend()
plt.show()
```