

```
In [125...]
```

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
In [126...]
```

```
dataset=pd.read_csv('cancer.csv')
dataset.head()
```

```
Out[126]:
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...	t
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...	
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...	

5 rows × 33 columns

```
In [127...]
```

```
x=dataset.iloc[:,2:32].values
x
```

```
Out[127]:
```

```
array([[1.799e+01, 1.038e+01, 1.228e+02, ..., 2.654e-01, 4.601e-01,
       1.189e-01],
       [2.057e+01, 1.777e+01, 1.329e+02, ..., 1.860e-01, 2.750e-01,
       8.902e-02],
       [1.969e+01, 2.125e+01, 1.300e+02, ..., 2.430e-01, 3.613e-01,
       8.758e-02],
       ...,
       [1.660e+01, 2.808e+01, 1.083e+02, ..., 1.418e-01, 2.218e-01,
       7.820e-02],
       [2.060e+01, 2.933e+01, 1.401e+02, ..., 2.650e-01, 4.087e-01,
       1.240e-01],
       [7.760e+00, 2.454e+01, 4.792e+01, ..., 0.000e+00, 2.871e-01,
       7.039e-02]])
```

```
In [128...]
```

```
y=dataset.iloc[:,1].values
y
```

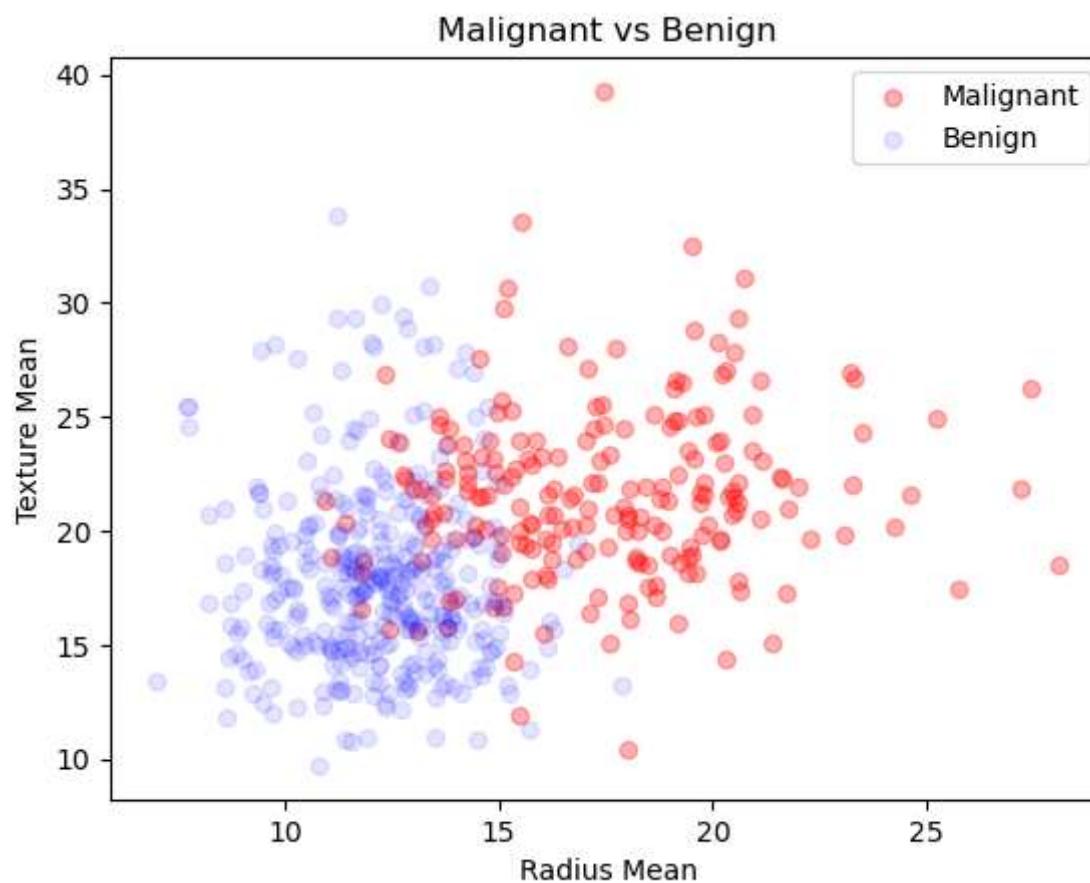
```
In [129...]: dataset.shape
```

```
Out[129]: (569, 33)
```

```
In [130... M=dataset[dataset.diagnosis == "M"]
```

```
In [131... B=dataset[dataset.diagnosis == "B"]
```

```
In [132... plt.title("Malignant vs Benign")
plt.xlabel("Radius Mean")
plt.ylabel("Texture Mean")
plt.scatter(M.radius_mean, M.texture_mean, color='red', label="Malignant", alpha=0.3)
plt.scatter(B.radius_mean, B.texture_mean, color='blue', label="Benign", alpha=0.1)
plt.legend()
plt.show()
```

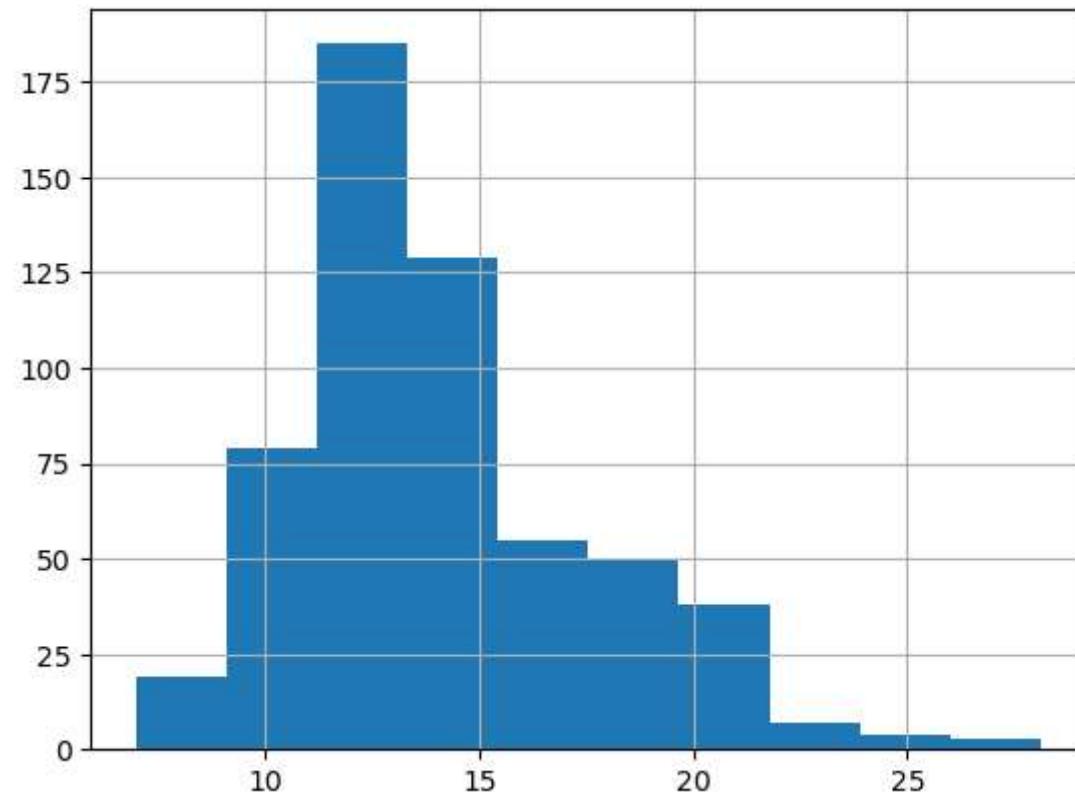


```
In [133... dataset.info()
```

```
<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 
 17  compactness_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
```

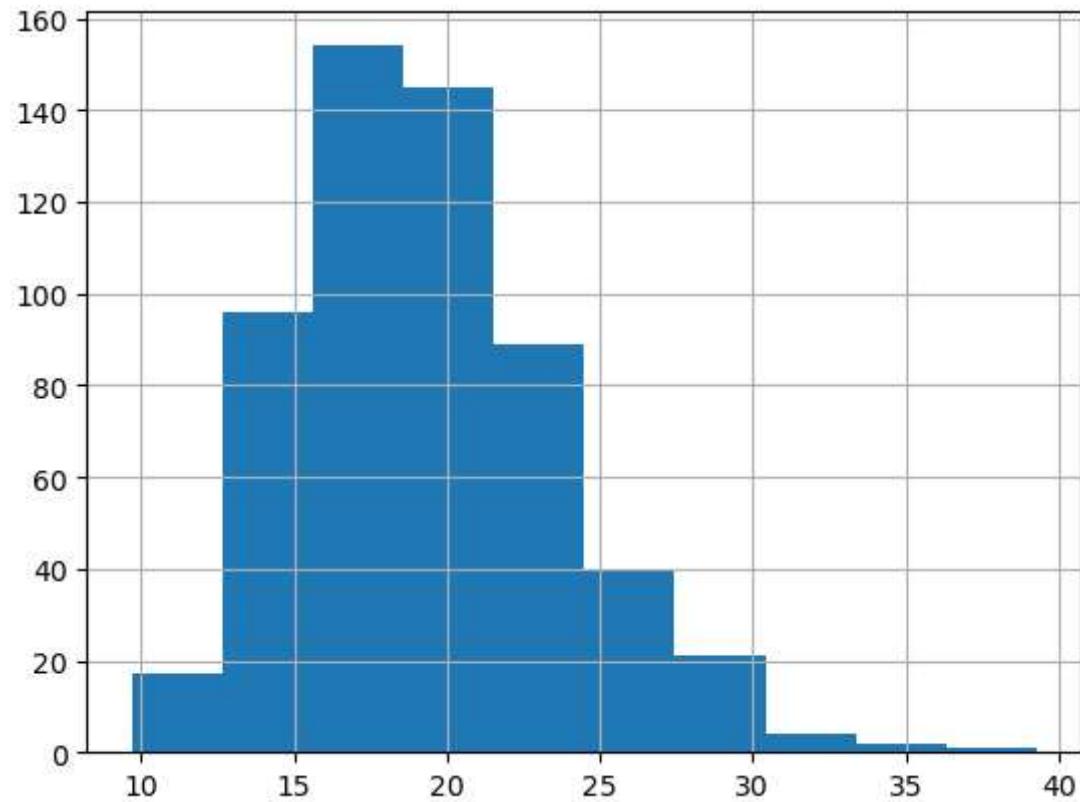
In [134...]

```
dataset['radius_mean'].hist()
plt.show()
```



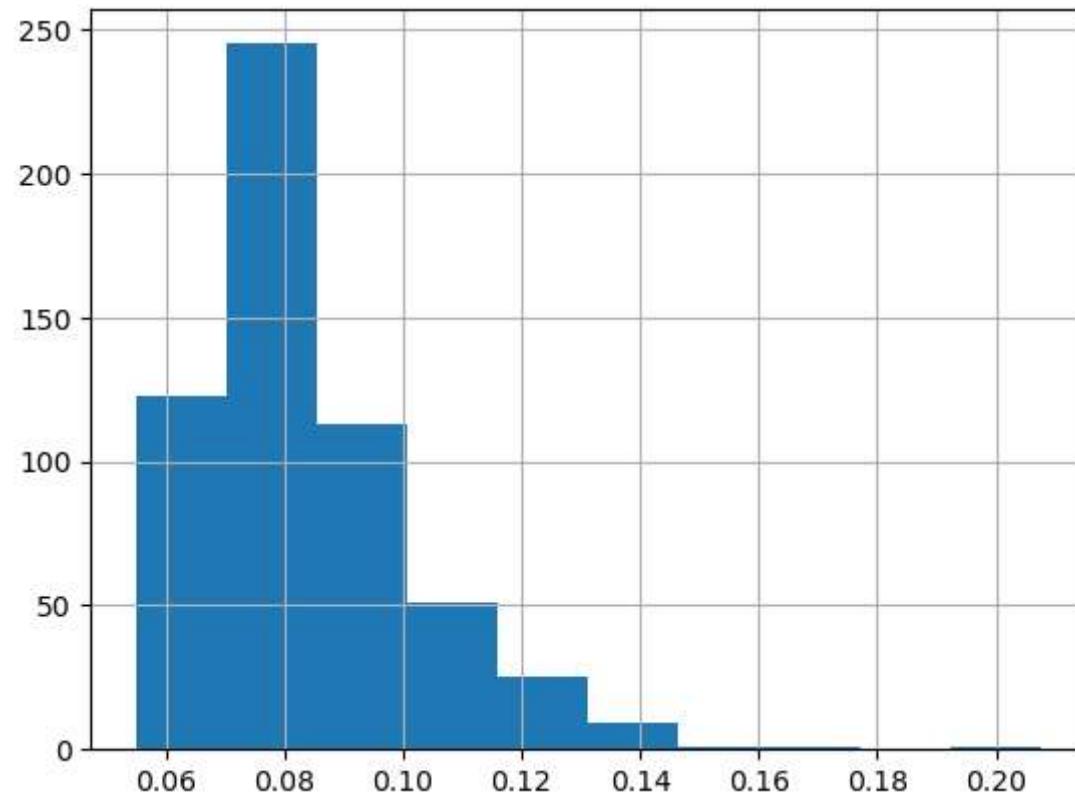
In [135...]

```
dataset['texture_mean'].hist()  
plt.show()
```



In [136]:

```
dataset['fractal_dimension_worst'].hist()  
plt.show()
```



```
In [137...]  
#dataset.boxplot(column='')  
#plt.show()
```

```
In [138...]  
dataset.isnull().sum()
```

```
Out[138]: 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 [139...]  
from sklearn.preprocessing import LabelEncoder  
e_y = LabelEncoder()  
y = e_y.fit_transform(y)
```

```
In [140...]  
y
```

```
In [141]: from sklearn.model_selection import train_test_split  
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25, random_state = 0)
```

In [142]: x_train

```
Out[142]: array([[1.185e+01, 1.746e+01, 7.554e+01, ..., 9.140e-02, 3.101e-01,
   7.007e-02],
   [1.122e+01, 1.986e+01, 7.194e+01, ..., 2.022e-02, 3.292e-01,
   6.522e-02],
   [2.013e+01, 2.825e+01, 1.312e+02, ..., 1.628e-01, 2.572e-01,
   6.637e-02],
   ...,
   [9.436e+00, 1.832e+01, 5.982e+01, ..., 5.052e-02, 2.454e-01,
   8.136e-02],
   [9.720e+00, 1.822e+01, 6.073e+01, ..., 0.000e+00, 1.909e-01,
   6.559e-02],
   [1.151e+01, 2.393e+01, 7.452e+01, ..., 9.653e-02, 2.112e-01,
   8.732e-02]])
```

```
In [143...]  
from sklearn.preprocessing import StandardScaler  
sc = StandardScaler()  
x_train = sc.fit_transform(x_train)  
x_test = sc.transform(x_test)
```

```
In [165...]  
from sklearn.linear_model import LogisticRegression  
classification = LogisticRegression()  
classification.fit(x_train, y_train)
```

```
Out[165]:  
LogisticRegression()  
|-----|
```

```
In [166...]  
y_pred = classification.predict(x_test)
```

```
In [167...]  
y_pred
```

```
Out[167]:  
array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1,  
     0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0,  
     0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,  
     1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,  
     1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0,  
     0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0,  
     0, 0, 0, 0, 1, 1, 0, 0, 0, 1])
```

```
In [168...]  
y_test
```

```
Out[168]:  
array([1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1,  
     0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0,  
     0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,  
     1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0,  
     1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1,  
     0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,  
     0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1])
```

```
In [169...]  
from sklearn.metrics import confusion_matrix  
cm = confusion_matrix(y_test, y_pred)
```

```
In [170...]  
print("Logistic Regression Accuracy ~{:.2f}".format((cm[0,0]+cm[1,1])/cm.sum()*100))
```

Logistic Regression Accuracy ~95.80

```
In [171...]  
from sklearn.neighbors import KNeighborsClassifier  
classification = KNeighborsClassifier(n_neighbors = 5)  
classification.fit(x_train, y_train)
```

```
Out[171]: ▾ KNeighborsClassifier  
          KNeighborsClassifier()
```

```
In [172]: y_pred = classification.predict(x_test)
```

```
In [173]: y_pred
```

```
In [174]: y_test
```

```
In [175]: cm = confusion_matrix(y_test, y_pred)
```

```
In [176]: print("Fitting KNN algo Accuracy ~{:.2f}%".format((cm[0,0]+cm[1,1])/cm.sum()*100))
```

Fitting KNN algo Accuracy ~95.10

```
In [177...]: from sklearn.naive_bayes import GaussianNB  
classification = GaussianNB()  
classification.fit(X_train, Y_train)
```

Out[177]: ▾ GaussianNB

```
In [178]: y pred = classification.predict(x test)
```

In [179]: y pred

```
In [180...  y_test
```

```
In [181]: cm = confusion_matrix(y_test, y_pred)
```

```
In [182...    print("Naive Bayes Accuracy ~{:.2f}").format((cm[0,0]+cm[1,1])/cm.sum()*100))
```

Naive Bayes Accuracy ~91.61

In []: