In [1]: ```python
import pandas as pd
```

In [4]: ```python
df = pd.read_csv("titanic.csv")
df.head()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Na
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Na
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Na

In [6]: ```python
inputs = df[['PassengerId','Survived','Pclass','Sex','Age','Fare']]
inputs.head()
```

Out[6]:

| | PassengerId | Survived | Pclass | Sex | Age | Fare |
|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | male | 22.0 | 7.2500 |
| 1 | 2 | 1 | 1 | female | 38.0 | 71.2833 |
| 2 | 3 | 1 | 3 | female | 26.0 | 7.9250 |
| 3 | 4 | 1 | 1 | female | 35.0 | 53.1000 |
| 4 | 5 | 0 | 3 | male | 35.0 | 8.0500 |

```
In [40]: inputs_n = inputs.drop('Survived' , axis = 'columns')
         inputs_n.head()
```

Out[40]:

	PassengerId	Pclass	Sex	Age	Fare
0	1	3	male	22.0	7.2500
1	2	1	female	38.0	71.2833
2	3	3	female	26.0	7.9250
3	4	1	female	35.0	53.1000
4	5	3	male	35.0	8.0500

```
In [41]: target = inputs.Survived
         target.head()
```

```
Out[41]: 0    0
         1    1
         2    1
         3    1
         4    0
         Name: Survived, dtype: int64
```

```
In [34]: from sklearn.preprocessing import LabelEncoder
```

```
In [35]: le_sex = LabelEncoder()
```

```
In [43]: inputs_n['sex_n'] = le_sex.fit_transform(inputs_n['Sex'])
         inputs_n.head()
```

Out[43]:

	PassengerId	Pclass	Sex	Age	Fare	sex_n
0	1	3	male	22.0	7.2500	1
1	2	1	female	38.0	71.2833	0
2	3	3	female	26.0	7.9250	0
3	4	1	female	35.0	53.1000	0
4	5	3	male	35.0	8.0500	1

```
In [47]: import math
```

```
In [51]: mean_sex =math.floor(inputs_n.Age.mean())
         mean_sex
```

Out[51]: 29

```
In [ ]:
```

In [68]:
```python
inputs_new = inputs_n.drop('Sex', axis = 'columns')
inputs_new.head()
```

Out[68]:

	PassengerId	Pclass	Age	Fare	sex_n
0	1	3	22.0	7.2500	1
1	2	1	38.0	71.2833	0
2	3	3	26.0	7.9250	0
3	4	1	35.0	53.1000	0
4	5	3	35.0	8.0500	1

In [70]:
```python
inputs_new.Age.fillna(mean_sex, inplace = True)
inputs_new.head(10)
```

Out[70]:

	PassengerId	Pclass	Age	Fare	sex_n
0	1	3	22.0	7.2500	1
1	2	1	38.0	71.2833	0
2	3	3	26.0	7.9250	0
3	4	1	35.0	53.1000	0
4	5	3	35.0	8.0500	1
5	6	3	29.0	8.4583	1
6	7	1	54.0	51.8625	1
7	8	3	2.0	21.0750	1
8	9	3	27.0	11.1333	0
9	10	2	14.0	30.0708	0

In [74]:
```python
from sklearn.model_selection import train_test_split
```

In [82]:
```python
x_train , x_test, y_train ,y_test = train_test_split(inputs_new , target, test_si
len(x_train)
```

Out[82]: 712

In [83]:
```python
from sklearn import tree
```

In [64]:
```python
model = tree.DecisionTreeClassifier()
```

In [84]:
```python
model.fit(x_train, y_train)
```

Out[84]: DecisionTreeClassifier()

In [85]:
```python
model.score(x_test , y_test)
```

Out[85]: 0.7150837988826816

In [80]:
```python
model.predict([[9 ,3,35,72, 0]])
```

Out[80]: array([1], dtype=int64)

In [ ]: