```python
import pandas as pd
import seaborn as sb
```

```python
df = pd.read_csv("/content/student_pass_fail_dataset.csv")
```

```python
print(df.head())
```

```
   hours_studied  attendance_percentage  assignments_submitted  pass_fail
0            0.5                     35                      0          0
1            1.0                     45                      1          0
2            1.5                     60                      2          0
3            2.0                     70                      3          1
4            2.5                     75                      3          1
```

```python
df.describe()
```

	hours_studied	attendance_percentage	assignments_submitted	pass_fail
count	15.000000	15.000000	15.000000	15.000000
mean	2.400000	67.533333	2.733333	0.666667
std	1.490446	23.311146	1.751190	0.487950
min	0.000000	20.000000	0.000000	0.000000
25%	1.250000	52.500000	1.500000	0.000000
50%	2.000000	70.000000	3.000000	1.000000
75%	3.500000	86.500000	4.000000	1.000000
max	5.000000	100.000000	5.000000	1.000000

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 4 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   hours_studied          15 non-null     float64
 1   attendance_percentage  15 non-null     int64
 2   assignments_submitted  15 non-null     int64
 3   pass_fail              15 non-null     int64
dtypes: float64(1), int64(3)
memory usage: 612.0 bytes
```

```python
df['hours_studied'].unique()
```

```
array([0.5, 1. , 1.5, 2. , 2.5, 3. , 3.5, 4. , 0. , 4.5, 5. ])
```

```python
from sklearn.preprocessing import LabelEncoder
```

```python
l=LabelEncoder()
```

```python
df['hours_studied']=l.fit_transform(df['hours_studied'])
```

```python
df['hours_studied'].unique()
```

```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  0,  9, 10])
```

```python
df
```

	hours_studied	attendance_percentage	assignments_submitted	pass_fail
0	1	35	0	0
1	2	45	1	0
2	3	60	2	0
3	4	70	3	1
4	5	75	3	1
5	6	80	4	1
6	7	85	4	1
7	8	90	5	1
8	0	20	0	0
9	9	95	5	1
10	10	100	5	1
11	2	50	1	0
12	4	65	2	1
13	4	55	2	1
14	7	88	4	1

```python
from sklearn.model_selection import train_test_split

X = df[['hours_studied', 'attendance_percentage', 'assignments_submitted']]

y = df['pass_fail']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, )
```
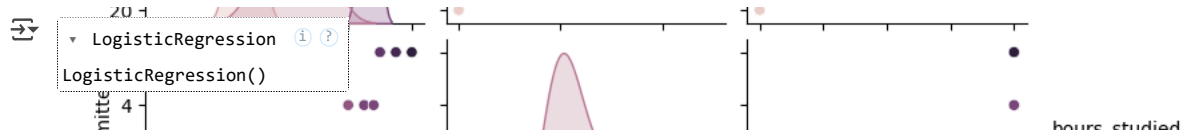
```python
sb.pairplot(df,hue='hours_studied')
```

```
<seaborn.axisgrid.PairGrid at 0x7d1653867a90>
```



```
from sklearn.linear_model import LogisticRegression
```

```
lr = LogisticRegression()
lr.fit(X_train, y_train)
```

▾ LogisticRegression ⓘ �ⓘ
LogisticRegression()

```
y_pred = lr.predict(X_test)
```

```
from sklearn.metrics import accuracy_score
```

```
accuracy_score(y_test,y_pred)
```

```
1.0
```

```
import pickle
```

```
pickle.dump(lr,open('students.pkl','wb'))
```

Start coding or generate with AI.