In [1]:

```
import pandas as pd
import matplotlib
df=pd.read_csv("question4.csv")
df
```

Out[1]:

	Course	Student
0	Yes	248
1	NO	11

In [2]:

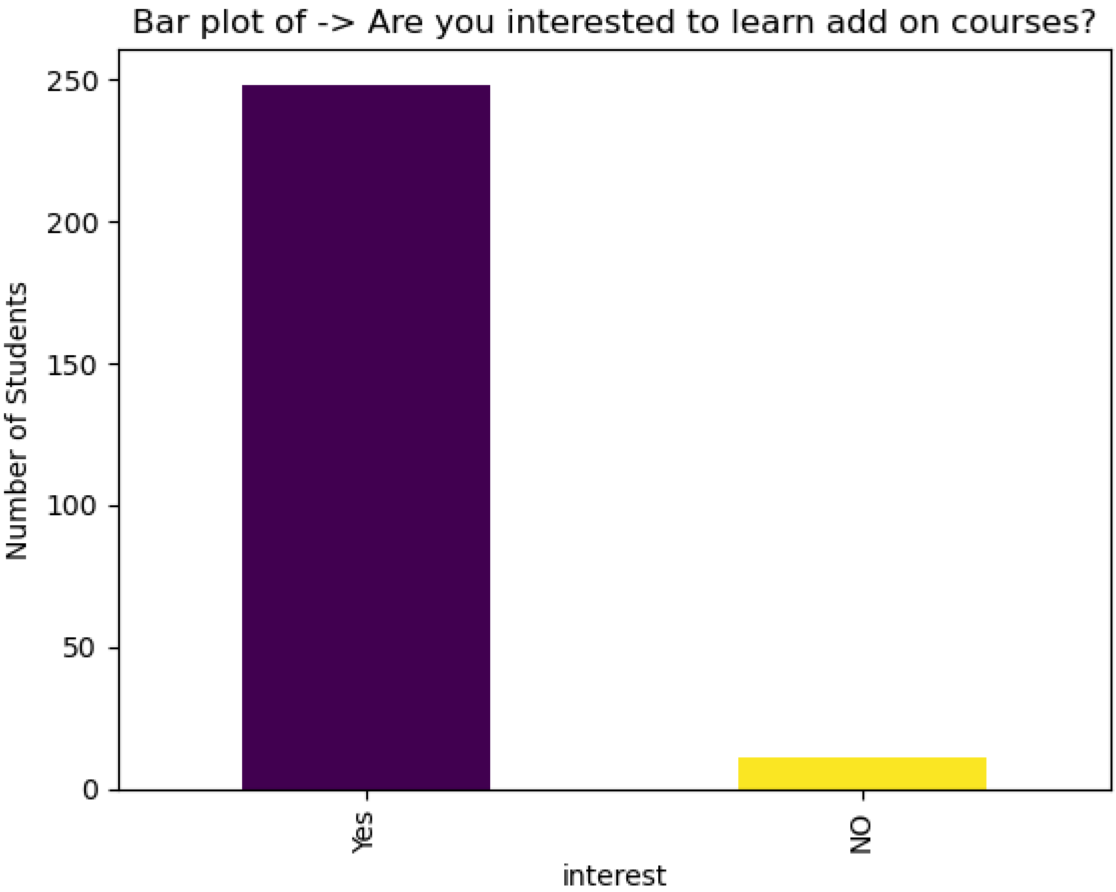
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import os

# Set OMP_NUM_THREADS environment variable to 1
os.environ["OMP_NUM_THREADS"] = "1"
# Number of clusters
k = 2
# Extract the 'student' column for clustering
X = df.drop('Course',axis=1)
# Apply k-means with explicit n_init
kmeans = KMeans(n_clusters=k, n_init=10, random_state=42)
predict = kmeans.fit_predict(X)
print(predict)
```

```
C:\Users\Ritesh\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1436: UserWarn
ing: KMeans is known to have a memory leak on Windows with MKL, when there are less c
hunks than available threads. You can avoid it by setting the environment variable OM
P_NUM_THREADS=1.
  warnings.warn(
[0 1]
```

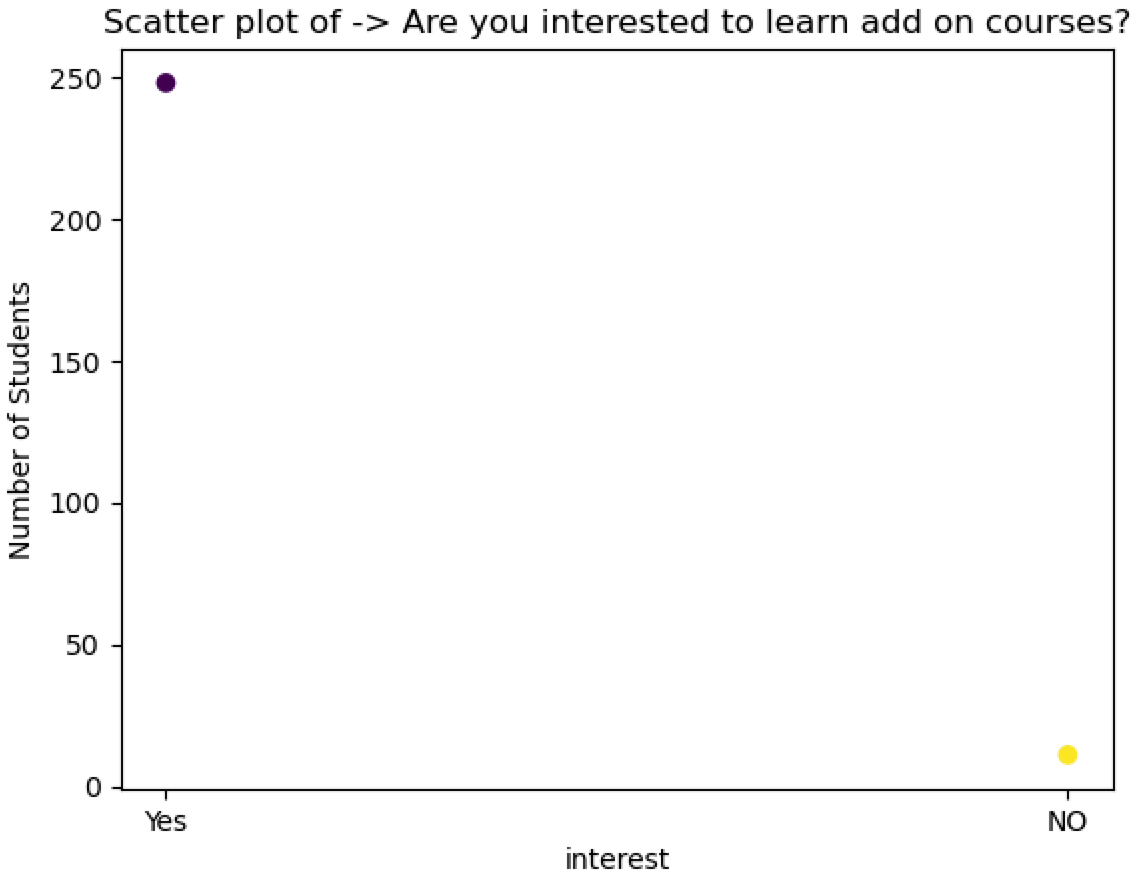
In [3]:

```
# Get unique colors for each cluster using a colormap
colors = plt.cm.viridis(predict/(k - 1))
# Bar plot
df.plot.bar(x='Course', y='Student', color=colors, legend=False)
plt.title('Bar plot of -> Are you interested to learn add on courses?')
plt.xlabel('interest')
plt.ylabel('Number of Students')
plt.show()
```



In [4]:

```
# Scatter plot
plt.scatter(range(len(df)), df['Student'], c=colors, marker='o')
plt.title('Scatter plot of -> Are you interested to learn add on courses?')
plt.xlabel('interest')
plt.ylabel('Number of Students')
plt.xticks(range(len(df)), df['Course'])
plt.show()
# Display cluster information
print("Clusters:")
print(df)
```



```
Clusters:
  Course  Student
0    Yes      248
1     NO       11
```

In [ ]: