```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import scipy.stats as stats
```

```
df=pd.read_csv('/content/case_study_1.csv')
df.head()
```

	user_id	age_group	subscription_status	engagement_time
0	14451	18-34	subscribed	5.55
1	18386	under 18	subscribed	5.12
2	12305	35 and over	not_subscribed	4.25
3	17546	18-34	subscribed	8.54
4	15399	18-34	subscribed	12.12

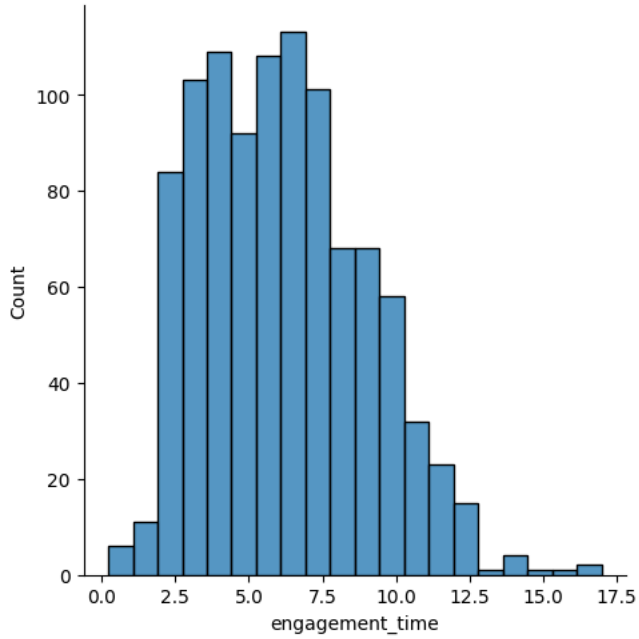
```
df.shape
```

```
(1000, 4)
```

```
#prob 1) is the avg app enagagement time significantly greater than the market avg(6 hrs)

print('sample mean:',np.round(df.engagement_time.mean(),2))
sns.displot(df.engagement_time)
plt.show()
```

```
sample mean: 6.18
```



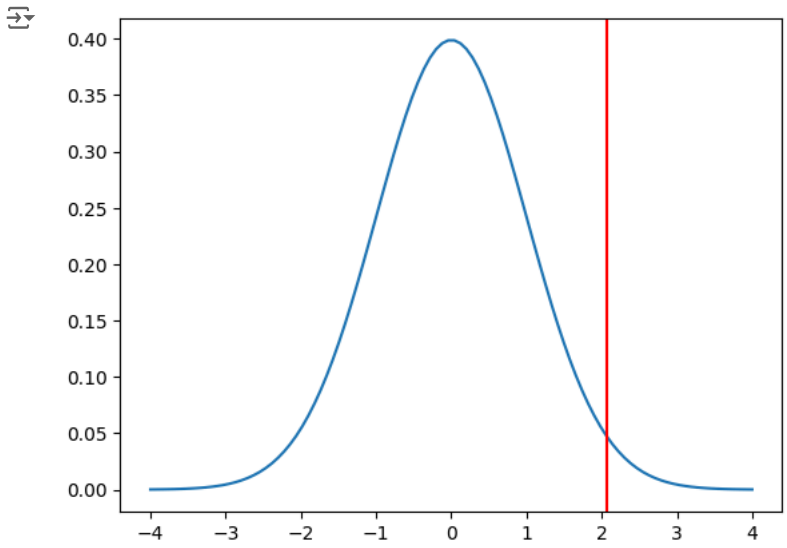

```
hyp_mean=6
t_stat, p_value=stats.ttest_1samp(df['engagement_time'],hyp_mean,alternative='greater')
print('t_stat:',round(t_stat,2))
print('p_value:',(p_value))
```

```
t_stat: 2.06
p_value: 0.019598877431817586
```

```
from scipy.stats import t

x=np.linspace(-4,4,100)
plt.plot(x,t.pdf(x,df=len(df)-1))
plt.axvline(x=t_stat,color='red')
```

```
plt.show()
```




```
if p_value<0.05:
  print('reject null hypothesis')
else:
  print('fai null hypothesis')

  # Since the null hypothesis is rejected we say that the mean app engagement is greater that the market avg
```

```
reject null hypothesis
```

```
#prob2) is the difference of proportions of subscribers and non-subscribers significantly different to conclude that a particular group is m

#pd.crosstab(index, columns, margins=False, normalize=False)
pd.crosstab(df.age_group,df.subscription_status,margins=True)
```

subscription_status	not_subscribed	subscribed	All
age_group			
18-34	103	262	365
35 and over	237	171	408
under 18	107	120	227
All	447	553	1000

```
pd.crosstab(df.age_group,df.subscription_status,normalize=True).plot(kind='bar',stacked=True)
```

```
<Axes: xlabel='age_group'>
```


```
contigency_table=pd.crosstab(df.age_group,df.subscription_status)
contigency_table
```

subscription_status	not_subscribed	subscribed
age_group		
18-34	103	262
35 and over	237	171
under 18	107	120


```
from scipy.stats import chi2_contingency

chi,p,dof,expected=chi2_contingency(contigency_table)
print('Test_Stats:',chi)
print('p_value',p)
print('deg_of_freedom:',dof)
print('expected_frequencies:\n',expected)
```

```
Test_Stats: 70.23716243606756
p_value 5.600076564450542e-16
deg_of_freedom: 2
expected_frequencies:
 [[163.155 201.845]
 [182.376 225.624]
 [101.469 125.531]]
```

```
if p<0.05:
  print('reject null hypothesis')
else:
  print('fai null hypothesis')
```

```
reject null hypothesis
```