Cricket is a sport that is played by people in different parts of the world. The dataset used for this assignment include details of scores, records, total number of matches played by players of India against England or vice versa. Project aims to unfold that portion of cricket statistics where we compare performance of teams (India, England) by taking 20 years of data as a sample (04 Mar 2000 - 04 Mar 2022). The data used in this project has been sourced from ESPN Cricinfo and data includes details of matches played by teams in their home ground, away from home and in neutral venues.

Name of the Field	Description
Player	Name of the player
Country	Name of the country
Matches Played	Number of matches played by the player
Runs	Number of runs scored by player
High Score	High score of a player in all matches played by him
Batting Avg	Batting average of a player
Wickets	Number of wickets taken by a player
Bowling Avg	Bowling average of a player
Catches	Number of catches by a player

Applying lessons learnt from Module-1:

1. Find the top 25% of the High Score by England and India?

P(z<z1) = 0.75

Z1 = 0.67

Mean = 73.46

SD = 61.64

Top 25% of High Score = 100.57

Area to the left of z:	0.67						
z=	0.439913	OR, if you need x:	mean=	73.457	st.dev=	61.636	x= 100.5715

Applying lessons learnt from Module-2:

1. We can construct the 95% confidence interval for the mean of catches by Indian Players.

Mean = 7.338

SD = 12.156

Degrees of freedom = n-1 = 65-1 = 64

For 95% confidence interval, t0.025,64 = 1.998

Thus, we are 95% confident that the mean of catches by players of Indian team is between 4.33 and 10.35

		IF variance =	
CONFIDENCE INTERVAL FOR MU		**THEN** st. dev =	
x-bar =	7.338		
s =	12.156		
n =	65		
$t_{a/2}$ =	1.998		
s of x-bar =	***1.51***		
Lower Conf. Limit LCL =	***4.33***		
Upper Conf. Limit UCL =	***10.35***		

Applying lessons learnt from Module-3:

We will test the following hypothesis:

1. **The mean of runs scored by top 40 players (both India & England) is greater than 525 (α=0.05).**

Step 1: State the hypotheses in plain English.

- **Null Hypothesis:** The mean of runs scored by top 40 players is less than or equal to 525.
- **Alternative Hypothesis:** The mean of runs scored by top 40 players is greater than 525.

Step 2: Select the appropriate statistical measure.

In this case, the parameter of interest will be the mean of runs scored by top 40 players.

μ= mean of runs scored by top 40 players.

Step 3: Determine whether the hypothesis should be one-sided or two-sided.

From the given information, we want to test if the mean of runs scored by top 40 players is greater than 525 or not. Thus, the alternative hypothesis will be one sided, and this will be a one tailed test.

Step 4: State the hypotheses using the appropriate statistical measure.

The resulting hypotheses would be as follows.

H0: $\mu \leq 525$

Ha: $\mu > 525$

Step 5: Specify the level of the test.

The level of the test is specified in the problem as 0.05

Step 6: Select the appropriate test statistic.

Here, Population standard deviation is not known. We will have to use sample standard deviation for our calculations. So, we use t-test statistic for this case.

Step 7: Determine the critical value.

Degrees of freedom = n – 1 = 40-1 = 39

Using the t Table, $t_{0.05,39} = 1.685$

Rejection Rule – We reject H0 if the computed t value is greater than the t critical value 1.685 Otherwise, we fail to reject H0.

Step 8: Compute the test statistic.

x-bar = 1019.025, S=636.002, and n=40

t = 4.9127 (Calculated using excel)

The sample mean is 4.9127 standard deviations more than the hypothesized value of the mean.

mu=	525	mu=	403
x-bar=	1019.025	x-bar=	432
s=	636.002	s^2=	841
n=	40	n=	16
t=	***4.9127***	***t=***	***4***
exact p-value (1-tail)=	8.25E-06	exact p-value (1-tail)=	0.00058
exact p-value (2-tail)=	1.65E-05	exact p-value (2-tail)=	0.001159

Step 9: Make the decision.

Since the test statistic falls in the rejection region. We reject the null hypothesis.

Step 10: State the conclusion in terms of the original question.

There is sufficient evidence to conclude that the mean of runs scored by top 40 players is greater than 525.

Applying lessons learnt from Module-4:

We will test the following hypothesis:

1. **The mean bowling average of Indian players is greater than the mean bowling average of England Players (α=0.05).**

Step 1: State the hypotheses in plain English.

- **Null Hypothesis:** The mean bowling average of Indian players is less than or equal to the mean bowling average of England players.
- **Alternative Hypothesis:** The mean bowling average of Indian players is greater than the mean bowling average of England players.

Step 2: Select the appropriate statistical measure.

μ1= The mean bowling average of Indian players

μ2= The mean bowling average of England players

Step 3: Determine whether the hypothesis should be one-sided or two-sided.

We are interested in knowing whether the mean bowling average of Indian players is greater than mean bowling average of England players. Thus, the alternative hypothesis is one-sided, and the test is one tailed test.

Step 4: State the hypotheses using the appropriate statistical measure.

H0: $\mu 1 - \mu 2 \leq 0$

Ha: $\mu 1 - \mu 2 > 0$

Step 5: Specify the level of the test.

The level of test is specified in the problem as $\alpha = 0.01$

Step 6: Select the appropriate test statistic.

Here, Population standard deviation is not known. We will have to use sample standard deviation for our calculations. So, we use t-test statistic for this case.

Step 7: Determine the critical value.

We assume that the population variances are equal.

It is a one tail test and the degrees of freedom = (77+65-2) = 140

Given, Alpha = 0.05

Therefore, Critical Value = **1.6558**

We will reject the null hypothesis if the computed value of the test statistic is greater than or equal to 1.6558

Step 8: Compute the test statistic.

	mu1-mu2=	0					
	x-bar1=	30.362	t for C.I. =	1.6558			
	x-bar2=	23.7					
	s1=	32.663					
	s2=	29.399					
	n1=	65					
	n2=	77				Lower	Upper
For Unequal Pop. Variances:	*calc t=*	***1.267***	stand error=	5.257	CI=	-2.043	15.367
For Equal Pop. Variances:	*calc t=*	***1.279***	stand error=	5.210	CI=	-1.965	15.289

Step 9: Make the decision.

We fail to reject null hypothesis because the test statistic value doesn't fall in the rejection region at $\alpha=0.05$.

Step 10: State the conclusion in terms of the original question.

There is no sufficient evidence at the 0.05 level to conclude that the mean bowling average of Indian players is greater than the mean bowling average of England players.