SHOW CODE

Welcome to M

Please Note :

- Go through the questions below and solve using **Excel ONLY**
- Please ensure that you include all your worked out files in a folder, zip the same and upload the same as attachment while submitting
- In the absence of worked out files, your submission will stand **INVALID**

Question 1

- Refer the dataset : **iris.xlsx**
- Create a **summary table** to show count of observations for each class of the flower.
  Expected Outcome :

class	Count

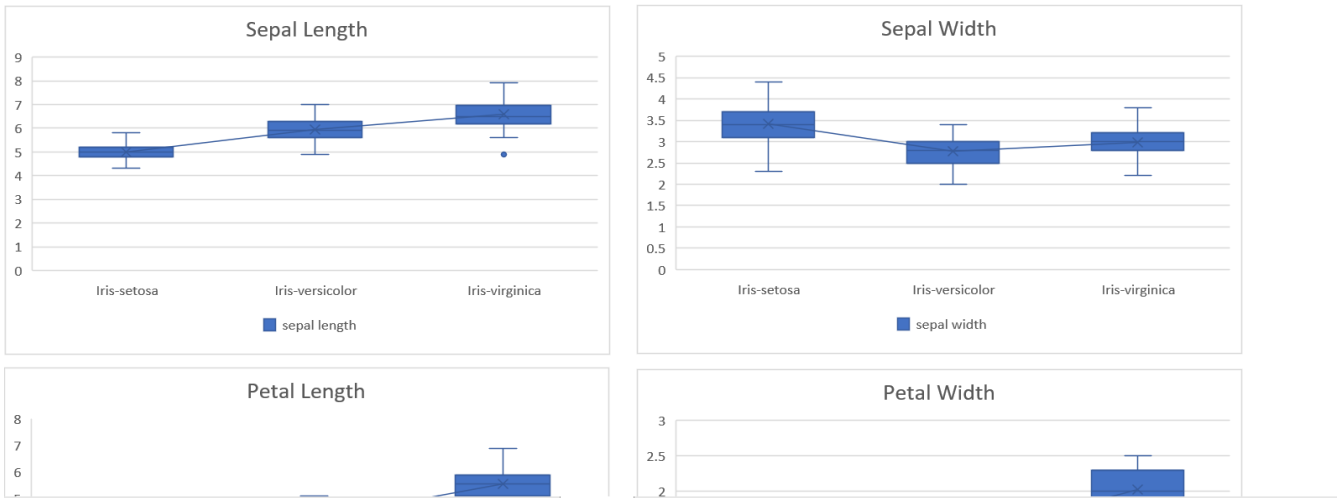
- Perform **Descriptive Statistics** on each of the 4 columns of sepal length, sepal width, petal length, petal width
  Expected Outcome :

sepal length	*sepal width*	*petal length*	*petal width*
Mean	Mean	Mean	Mean
Standard Error	Standard Error	Standard Error	Standard Error
Median	Median	Median	Median
Mode	Mode	Mode	Mode
Standard Deviation	Standard Deviation	Standard Deviation	Standard Deviation
Sample Variance	Sample Variance	Sample Variance	Sample Variance
Kurtosis	Kurtosis	Kurtosis	Kurtosis
Skewness	Skewness	Skewness	Skewness
Range	Range	Range	Range
Minimum	Minimum	Minimum	Minimum
Maximum	Maximum	Maximum	Maximum
Sum	Sum	Sum	Sum
Count	Count	Count	Count

Distribution of which measure is closest to a **standard normal distribution**?

Question 2

- Create a box-whiskers plot to show comparative distribution of each of the three classes across the 4 measures :
  Expected Outcome :



- Is the mean **sepal length** of **setosa** significantly different than that of **versicolor**?
  - Frame your **null hypothesis**
  - Perform a paired two-sample **t-test**
  - Do you have sufficient evidence to **reject** the null-hypothesis?

Question 3

- Refer the dataset : **advertisements.xlsx**
- Which of the 4 variables have distributions which look like a **uniform distribution**?
- Use **Data Analysis** package and calculation correlations between each-pair of variables
- Which **pair** has the strongest **correlation** and how much?
- Which **input variable** has the strongest **correlation** with the response variable (Sales)
- Which **advertisement channel** (amongst TV, Radio and Newspaper) would you suggest your marketing team to choose for maximum impact on Sales?
- Is the mean **sepal length** of **setosa** significantly different than that of **versicolor**?
  - Frame your **null hypothesis**
  - Perform a paired two-sample **t-test**
  - Do you have sufficient evidence to **reject** the null-hypothesis?

Question 4

- Refer the dataset : **popular_kids.xlsx**

- In the dataset **Popular Kids**, students in grades 4-6 were asked whether **good grades**, **athletic ability**, or **popularity** was most important to them.
- Prepare a **two-way** table separating the students by grade and by choice of most important factor :

Count of Sl.No.	Column Labels			
Row Labels	**4**	**5**	**6**	**Grand Total**
grades	63	88	96	247
popular	31	55	55	141
sports	25	33	32	90
Grand Total	**119**	**176**	**183**	**478**

- Do we really have any **relationship** between the **grades** and the **goal** which the students wish to achieve?
    - Is there a case that say, students in grade 6 would like to do better at studies which the students in lower grades would like to excel at Sports?
- Perform a **chi-square test** and determine whether the 2 variables grades and top_goal are independent of each other or not?