

# Decision Support Systems

Final Exam – DSC - A

Dated: 23rd Jan 2025

Dr. Humera Noor

# Important

- Please ensure that you copy the link to your solutions in the correct submissions sheet:
  - Roll # 1-40:  
[https://docs.google.com/spreadsheets/d/1mCAF5q9meYIMJ\\_87FSmeURtYlcICYRQNaNcahxDxSGc/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1mCAF5q9meYIMJ_87FSmeURtYlcICYRQNaNcahxDxSGc/edit?usp=sharing)
  - Roll # 41-80:  
[https://docs.google.com/spreadsheets/d/1JftZTuU4MFTZZvf8q4NMNHMB54\\_KU2r9iGpOAjy3JqE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1JftZTuU4MFTZZvf8q4NMNHMB54_KU2r9iGpOAjy3JqE/edit?usp=sharing)
  - Roll # 81-120: <https://docs.google.com/spreadsheets/d/1j17VcArjOusNpMhISbzw-ITXt5SD0s-tyxwnu8IQ40c/edit?usp=sharing>
- Caution: Edit your cells only, not for others in the class
- Please also ensure that you give me an edit access to your submissions:  
[humeranoor@gmail.com](mailto:humeranoor@gmail.com)

# Data Analytics

Question 1

# [DSC-A] Question 1 – Data Analytics – 10 marks

page 1 of 3

- You will work on `forbes_billionaires_geo.csv` dataset for this question.
- Please import the csv file in a google / Excel sheet.
- Perform the given operations in your sheet (create new tabs within the same file) and answer the following questions
- You will not use any programming language like python or tool like Tableau / PowerBI for this question.

# [DSC-A] Question 1 – Data Analytics – 10 marks

## page 2 of 3

- **Objective:** Identify how the total net worth is distributed across countries and understand the contribution of self-made individuals to the net worth in each country.
- **Specify the following:**
  - Which country has the highest total net worth?
  - What percentage of the net worth in each country is contributed by self-made individuals?
  - Are there any countries where non-self-made individuals dominate the net worth?

# [DSC-A] Question 1 – Data Analytics – 10 marks

## page 3 of 3

- Visualize the total net worth aggregated by country to identify the top 3 wealthiest countries in the dataset.
- Create a new tab in your sheet with data of billionaires of these top 3 countries with an age below 35. Make sure the headings are there as well. Name this tab as `young_billionaires`.

# Expert Systems

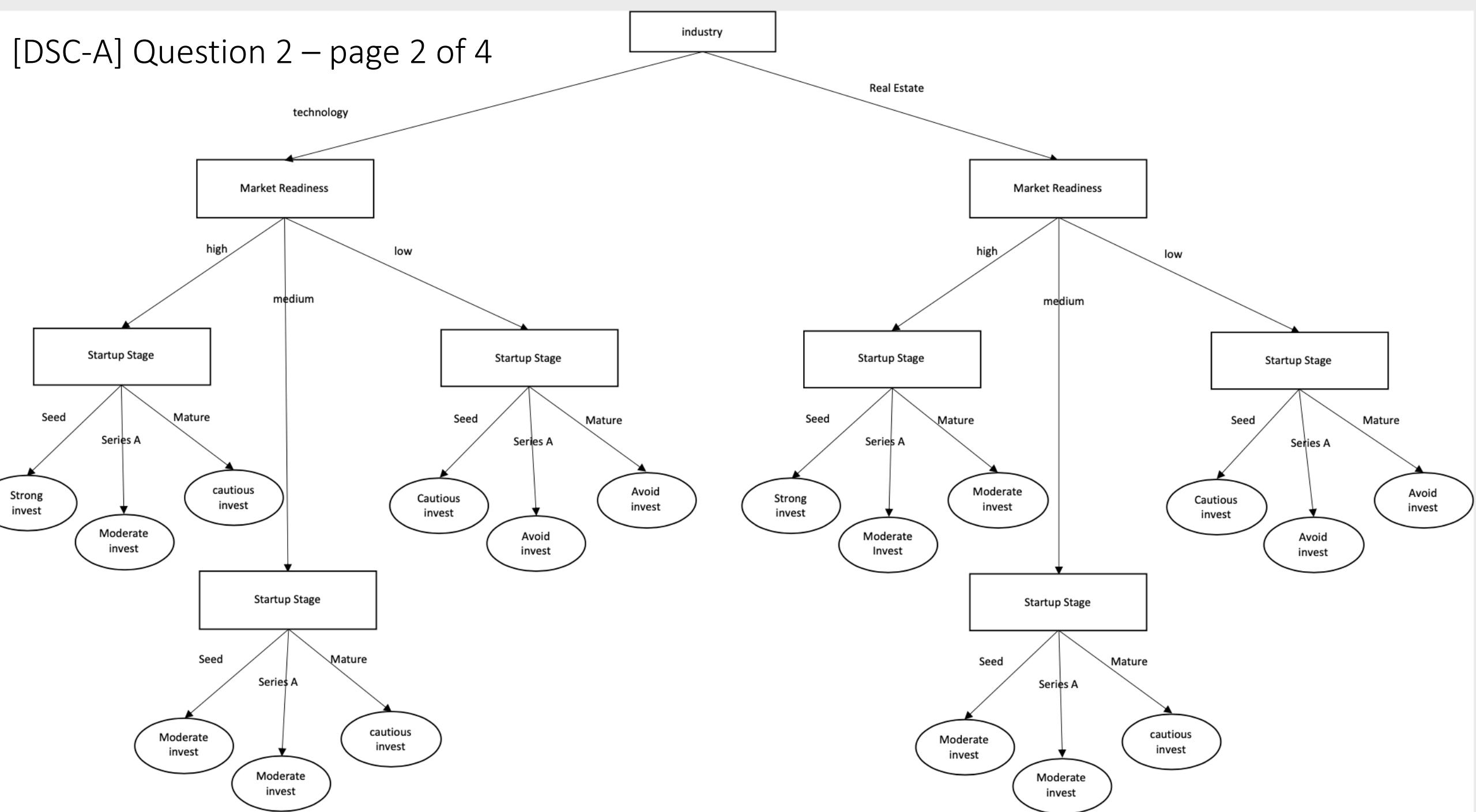
Question 2

# [DSC-A] Question 2 – Expert Systems – 10 marks

page 1 of 4

- You will use a colab or jupyter notebook for this question
- You can also program on your local machine
- Develop an expert system to model the following Investment recommendation system

# [DSC-A] Question 2 – page 2 of 4



# [DSC-A] Question 2 – Expert Systems – 10 marks

page 3 of 4

- Run the code with the following input and print results
- If the last digit of your 8-digit matriculation no. is an even number:
  - Investment in a series A technology startup with low market readiness
- If the last digit of your 8-digit matriculation no. is an odd number:
  - Investment in a mature real estate startup with high market readiness

[DSC-A] Question 2 – Expert Systems – 10 marks  
page 4 of 4

- Re-run the code with an input not covered by the the Expert system – in this case, update your code for the output to be „no suggestion“

# ChatGPT for Decision Support

Question 3

# [DSC-A] Question 3 – ChatGPT – 10 marks

page 1 of 2

- Login to chatgpt (or equivalent tool) and follow the instructions below for interaction

# [DSC-A] Question 3 – ChatGPT – 10 marks

## page 2 of 2

- Copy all the rows (with heading) from young\_billionaires in question 1.
- Bring this data to chatgpt and ask it to analyse the data – please don't upload the file there.
- Next ask chatgpt to Identify how the total net worth is distributed across countries and understand the contribution of self-made individuals to the net worth in each country.
- How does this answer compare to your original answer in Q1.
- Ask ChatGPT to create one meaningful visualization from the insights it generated (it should create the chart, not just describe the steps).

# ML - Video Recording

Question 4

# [DSC-A] Question 4 – ML-Video – 10 marks

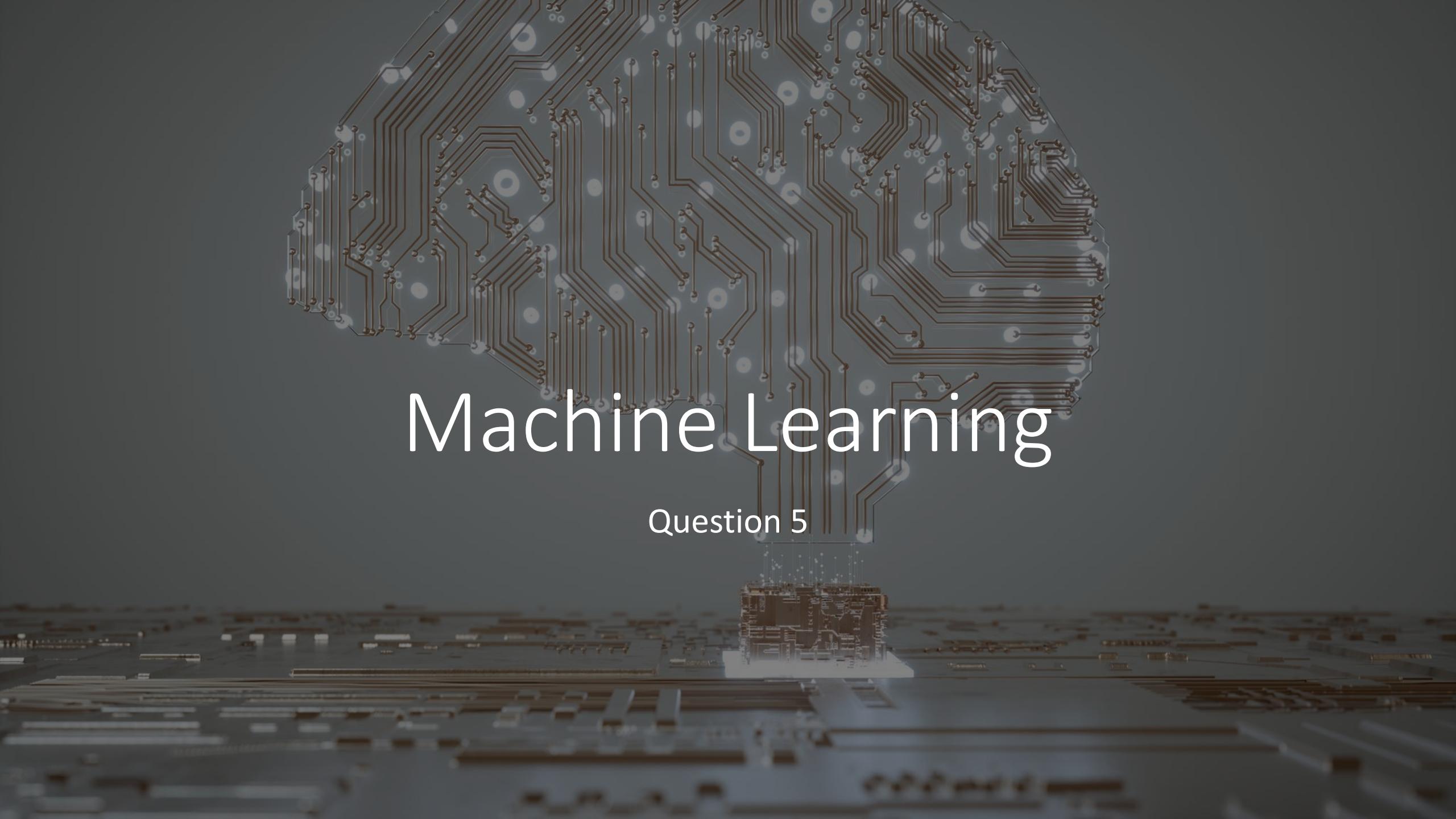
## page 1 of 2

- In this question you'll record a video of yourself and your screen and answer the following questions
- Please limit your video to 1-3 minutes

# [DSC-A] Question 4 – ML-Video – 10 marks

## page 2 of 2

- Referring to the Assignment 2 and 3, briefly give overview of your dataset and the motivation behind picking it
- Pick up your favourite visual in the assignment and describe it.



# Machine Learning

Question 5

# [DSC-A] Question 5 – Machine Learning – 10 marks

page 1 of 4

- You will use a colab or jupyter notebook for this question
- You can also program on your local machine
- For the cleanup, you can use either Excel/Google sheet or python
- You will work on the file: titanic.csv
- (optional) if you want to read about the titanic dataset, you can do it here: <https://www.kaggle.com/competitions/titanic/overview>
- The file has the parameters given on next page:

Variable	Definition	Key
passengerID	Passenger ID	
survived	Survival (output variable)	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
Name	Name of passenger	
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

# [DSC-A] Question 5 – Machine Learning – 10 marks

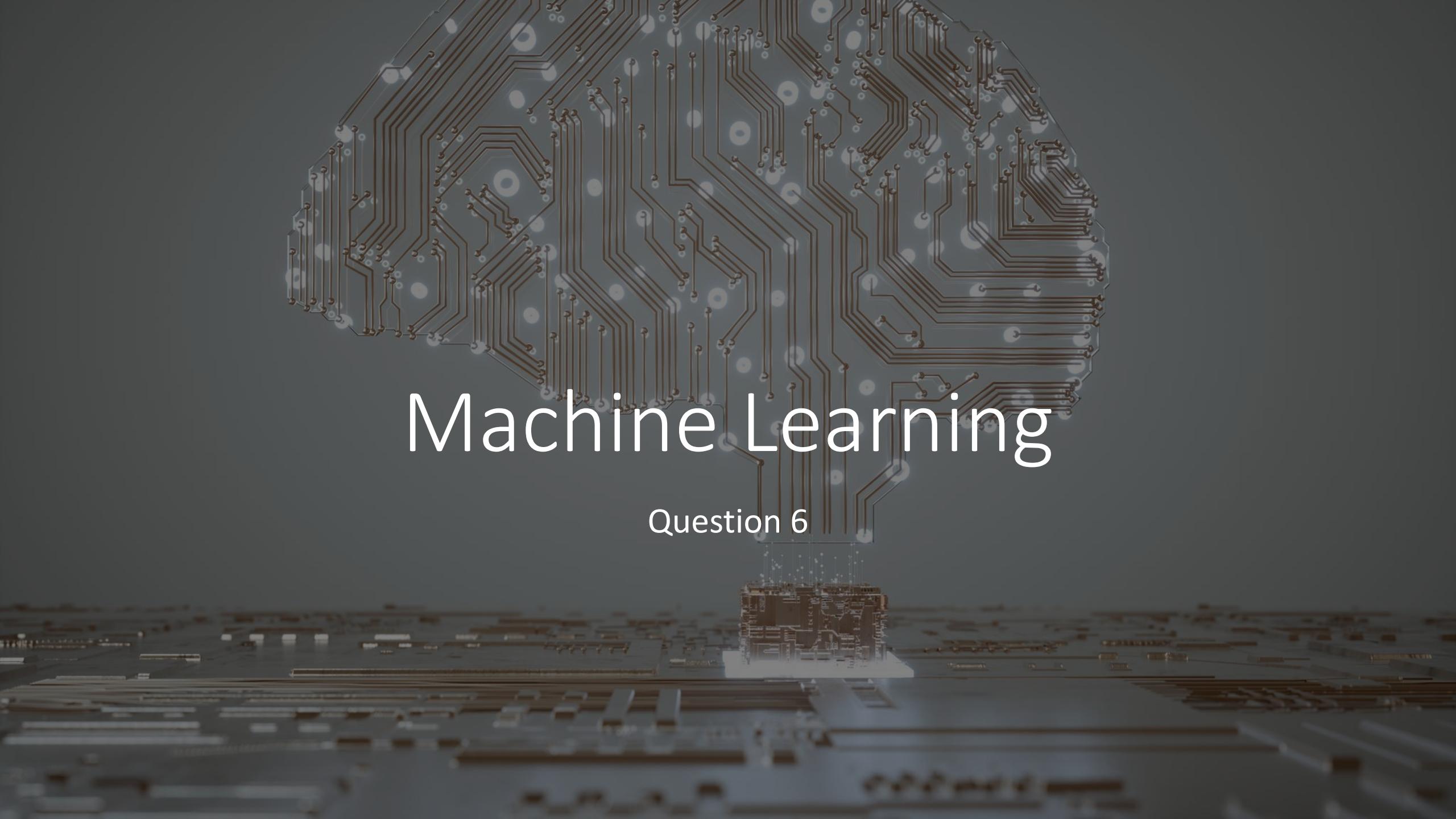
## page 3 of 4

- Use the tool of your choice to cleanup titanic.csv as follows:
  - Fix Non-numeric parameters:
    - You need to either remove them or convert to numbers. Decide on the basis of their importance whether they're needed to predict the final outcome (survived) or not.
  - Handle missing values
    - Get average or other suitable alternative
- The outcome of this step will be a clean dataset with no non-numeric columns and no missing values
- You will work on this data moving forward

# [DSC-A] Question 5 – Machine Learning – 10 marks

page 4 of 4

- Load the data in your python script
- Describe the data and generate stats
- Create at least three insightful visuals
- Prepare the data for training the model
  - Separate out the input and output parameters
  - Split the data in train and test sets



# Machine Learning

Question 6

# [DSC-A] Question 6 – ML – 10 marks

page 1 of 1

- Continue with the titanic dataset you split in Q5.
- Continue within the same notebook
- Train three different models to predict survival of passengers. Pick the best model and refine it further. Use k-folding.
- Show performance metrics and compare the models' performances as a boxplot.
- Please follow the best practices when writing the code.



Wind up and Submit

# Important

- Please ensure that you copy the link to your solutions in the correct submissions sheet:
  - Roll # 1-40:  
[https://docs.google.com/spreadsheets/d/1mCAF5q9meYIMJ\\_87FSmeURtYlcICYRQNaNcahxDxSGc/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1mCAF5q9meYIMJ_87FSmeURtYlcICYRQNaNcahxDxSGc/edit?usp=sharing)
  - Roll # 41-80:  
[https://docs.google.com/spreadsheets/d/1JftZTuU4MFTZZvf8q4NMNHMB54\\_KU2r9iGpOAjy3JqE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1JftZTuU4MFTZZvf8q4NMNHMB54_KU2r9iGpOAjy3JqE/edit?usp=sharing)
  - Roll # 81-120: <https://docs.google.com/spreadsheets/d/1j17VcArjOusNpMhISbw-ITXt5SD0s-tyxwnu8IQ40c/edit?usp=sharing>
- Caution: Edit your cells only, not for others in the class
- Please also ensure that you give me an edit access to your submissions:  
[humeranoor@gmail.com](mailto:humeranoor@gmail.com)