# Dublin Business School

# Assessment Brief

# Assessment Details

|  |  |
| --- | --- |
| Module Title: | Advanced Data Analytics |
| Module Code: | B8IT109 |
| Module Leader: | Dr Shahram Azizi |
| Stage (if relevant): |  |
| Assessment Title: | CA One |
| Assessment Number (if relevant): |  |
| Assessment Type: |  |
| Restrictions on Time/Length : | Submission before deadline |
| Individual/Group: |  |
| Assessment Weighting: |  |
| Issue Date: |  |
| Hand In Date: |  |
| Planned Feedback Date: |  |
| Mode of Submission: | Online |

**Guideline:**

* This CA assesses students on core concept in descriptive analytics, discrete and continuous probability models and generalized linear models (regression analysis).
* All questions are mandatory.
* Use R/Rstudio to solve questions and perform analytics.
* Any submission after deadline will not be considered and scored.

Consider a real-world, relational dataset. This dataset must have at least 2 categorical and 2 continuous variables.

**Question 1 (35 Marks)**

* 1. Describe the dataset using appropriate plots/curves/charts,… (7)
  2. Consider one of continuous attributes, and compute central and. (8)
  3. For a particular variable of the dataset, use Chebyshev's rule, and propose one-sigma interval. Based on your proposed interval, specify the outliers if any. (10)
  4. Explain how the box-plot technique can be used to detect outliers. Apply this technique for one attribute of the dataset (10)

**Question 2 (35 Marks)**

1. Select four variables of the dataset, and propose an appropriate probability model to quantify uncertainty of each variable.  **(10)**

1. For each model in part (a), estimate the parameters of model. (10)

1. Express the way in which each model can be used for the predictive analytics, then find the prediction for each attribute. (15)

**Question 3 (30 Marks)**

From your dataset, specify your input and output variables, then:

1. Suggest an appropriate GLM to model **ouput** to **input** variables. (5)
2. Split the dataset into 80% as a trainset and 20% testset, then model the trainset by fitting your proposed GLM. (7)
3. Specify the significant variables on the **output** variable at the level of 𝛼=0.05 and explore the related hypotheses test. Estimate the parameters of your model. (8)
4. Predict the output of the test dataset using the trained model. And provide the functional form of the optimal predictive model. (6)
5. Propose the appropriate measure of performance to evaluate the model and compute it for your derived model. (4)