



INTERVIEW ASSIGNMENT

Data Scientist

November 2022

1 OVERVIEW

This assignment is a practical application of the requirements of the Data Scientist role at JN Group and is designed to assess the candidate's ability to meet key requirements of the role. The Data Scientist will be required to support the organisation, and will serve as a technical analyst within the Data and Analytics team, responsible for using statistical methods, machine learning algorithms and other tools to analyze data and create predictive models.

2 INSTRUCTIONS

As a part of the interview process, a presentation is to be done outlining the results of the analysis done on the provided datasets (attached). The presentation will be delivered to a panel of judges on a date to be communicated with you.

Your assignment is to use one or both of the attached excel datasets (*cs-training* and *cs-test*) to conduct an analysis on the data provided. The *cs-training* file contains two sheets (a Data Dictionary and the Data Sheet) and includes hypothetical records of the credit history of 150,000 persons.

Since the role will require you to analyse and interpret data, **you are tasked with interpreting the data in the training sheet and communicating some of the key findings.**

The sheets were originally created in the context of a credit scoring competition that had the following guidelines:

“Credit scoring algorithms, which make a guess at the probability of default, are the method banks use to determine whether or not a loan should be granted. This competition requires participants to improve on the state of the art in credit scoring, by predicting the probability that somebody will experience financial distress in the next two years.”- Kaggle –Give Me Some Credit Competition.

As a bonus activity, if you can, use the *cs-test* data to formulate a probability of default, enter the information in column B. If you do so, be prepared to explain how you came up with the formula.

The following guidelines are to be observed:

- The presentation can be delivered utilizing Microsoft PowerPoint;
- It should include the rationale for chosen methodology for the data analysis and explanation of any statistical tools used to analyze the datasets;
- Include all relevant charts, tables or images to assist in depicting the key inferences and findings from the datasets;
- The presentation must be no longer than fifteen (15) minutes but can be shorter;

3 ASSESSMENT GUIDELINE

The following scoring guideline will be used to grade the presentation.

Assessment	Point
Presentation: Quality (Clarity, Arrangement, Use of Graphical Representation etc.)	10
Methodology: (explanation and rational for chosen method)	20
Analysis/ interpretation of results	30