

# Data Science Assignment

September 9, 2019

## Introduction

This assignment has been created with the purpose of giving the candidate an opportunity to demonstrate their strengths and skills. We have tried to keep the assignment as general as possible, as we merely want to observe the candidate's programming, analysis, and logical reasoning.

The assignment will be split into two main sections. Part i will consist of an extensive exploration of the dataset provided. Part ii will be to use what you have learnt in part 1 and build a predictive model on a specified feature of the dataset. You will also have the opportunity to provide feedback and a self reflection on the task. For example, you may like to describe other approaches you may have taken had you had the time.

For part i in particular, you are free to use whichever approach you feel most comfortable with, or find most suitable. For example, you could treat this as an exploratory task and provide an annotated jupyter notebook; treat it as a customer relations task and provide a visual dashboard; or even provide a program that is tailored towards a technical end-user, such as an object/library that performs functions on this dataset, fully documented. We do ask, however, that what you provide is reproducible on our end.

## What we are looking for from candidates

The type of work we carry out at Auspex can vary a lot, so we would like a candidate that can thrive in various environments. Whilst it's great to shine in one area of the data science domain, we do look for candidate's to have very good general skills as well. The list below should outline the type of skills we are looking for in this task, and skills that we believe are of utmost importance to being a good data scientist.

- **Programming ability** - We would like you to produce your code in a professional and efficient manner. Please make your code as simple and readable as possible. If this means breaking up large, multi-task functions or not calling 5 methods on top of each object returned in one-line, then so be it. **Messy code will not fare well.**

- **Creativity** - This applies mainly to part i of the assignment. The dataset we have provided contains a lot of features, and there is a copious amount of insight to be gleaned here before any modelling is carried out. We expect some nice visuals (always good for clients), interesting statistics (helps select and configure models), and anything else you think is important.
- **Clarity** - For the written analysis and explanation of your work, we are looking for clear, understandable language. This will help show that you fully understand what you are doing and bolster any findings you have.

## Part i

### Description

In the data directory, you will find a csv file called *loans.csv*. Each sample within this dataset describes the characteristics of a loan to a borrower. You will also find a text file giving a brief description for each feature, *data\_dictionary.csv*.

### Exploratory Data Analysis Task

Please provide an exploratory analysis of this dataset. We would usually treat this section as a narrative, providing an explanation for each piece of analysis we produce, with any observations we can glean from this. We expect part of this exploration to be a basis for the following section.

#### Basic examples

- How has the average interest rate of a loan varied over time?
- What is the distribution of sub-grades, and can we easily segment borrowers based on prime and sub prime loans?

#### Advanced examples

- Are we able to visualise a low-dimensional split in the dataset when the target variable is the grade?

## Part ii

### Predictive Model

Once you are happy with the results you have for *Part i*, we would like you to create a model that can predict either the *grade* or *int\_rate* feature.

For this section, you will need to partition the data into a training set and testing set,

presenting your results for each set. We would like the candidate to choose the best model for the dataset in whichever way they feel is best. It is worth noting here how important it is to explain in a clear manner why and what you have chosen to do. For example, you may find that your validation result and test result are quite different, we expect the candidate to analyse this and provide an explanation. The more you can explain your actions, the better.

## 1 Section 2

***Optional*** - Please provide a reflection on the analysis you have provided. We understand that you may not have had as much time to complete this task as you'd have liked, so we encourage you to describe what you would have done differently had you had the time.