# **Tyro Payments – Data Science Test**

This task is intended to test your basic data science and machine learning modelling skills. Candidates are asked to submit all code used in analysis/training/testing/visualisation etc. Ideally Python, R, SQL and/or Scala with any required libraries/frameworks will be used but we are happy to accept other languages if desired. Keep in mind that we want to see your process and experiments so something like a notebook might be desirable but feel free to use whatever format you believe to be most effective.

You have been provided with several data files that contain information about some fictional customers. The schemas and data dictionaries are detailed below.

‘**transactions\_customer.csv**’ contains a monthly aggregation of transaction volume for some customers.

|  |  |  |
| --- | --- | --- |
| Field | Data Type | Description |
| date | Date | Date of first day of each month |
| customer\_id | String | Unique customer identifier |
| industry | Integer | Code for 10 industries, ranging from 1 to 10 |
| location | Integer | Code for 10 locations, ranging from 1 to 10 |
| monthly\_amount | Numeric | Total transaction amount for customer in given month |
| daily\_average\_amount | Numeric | Average transaction amount per day for customer in given month |
| daily\_sd\_amount | Numeric | Standard deviation of the daily transaction amount per customer in given month |

‘**customer\_start\_terminate.csv**’ contains customer tenure information.

|  |  |  |
| --- | --- | --- |
| Field | Data Type | Description |
| customer\_id | String | Unique customer identifier |
| startDate | Date | Date the customer join Tyro |
| firstTransactionDate | Date | Date the customer start transact |
| terminatedDate | Date | Date the customer/tyro terminate the contract |

## **Tasks**

1. Build a model that predicts the likelihood of a customer defaulting.
2. Determine the key factors that influence this outcome.
3. Assume you are the lead on this credit risk modelling project. Create a short presentation (<1-5 mins) on your results and findings for the executive team who want to launch a new lending product using your model. Don’t feel you need to use the entire 5 minutes if you feel you can convey the information in less time. We will ask you to present this if you progress to the final technical interview round.

**To submit:**

1. All code, results, external data sets and recommendations for future improvement for tasks 1 and 2. Please make sure to document your reasons for the decisions you make. We want to understand your thought process.
2. Your presentation for task 3.