



LLOYDS BANK: DIGDATA HOW-TO-GUIDE & BRIEF

This document contains all the information you will need to complete the Lloyds Bank live online career challenge. You will have received the following 4 files:

- Lloyds Bank live online career challenge – How to Guide (this document)
- PowerPoint presentation (plus a link to a recording of the live presentation), this includes hints and tips for the challenges
- Lloyds Bank live online career challenge – Dataset (Excel worksheet)
- Lloyds Bank live online career challenge – Data Dictionary (Excel worksheet)

LLOYDS BANK DIGDATA CHALLENGE – HOW-TO-GUIDE

This document contains a summary of the different files that you have access to.

LLOYDS BANK DIGDATA CHALLENGE – BRIEF & QUESTIONS

You can choose to complete either a data strategy or data science challenge. The PowerPoint presentation contains a full brief for each of the challenges along with some useful hints and tips!

Brief

Lloyds Banking Group is launching a new loans product. Prior to the launch they would like you to use historical customer data to:

- **Task A – Data Strategy** : Understand and summarise the different behaviours or attributes between customers who paid back their loan and customers who did not
- **Task B – Data Science**: To use the historical data to design a process which predicts the likelihood of a new customer not paying back their loan

LLOYDS BANK DIGDATA CHALLENGE – POWERPOINT PRESENTATION

This document contains the slides presented during the live presentation. You will also have access to a video recording of the live presentation.

LLOYDS BANK DIGDATA CHALLENGE – DATASET (EXCEL SPREADSHEET)

This spreadsheet contains some information about Lloyds Banking Group customers (c. 18,324) that have previously had a loan with the bank. There are 31 variables including a unique ID for each customer. Other variables include the final status of their loan and other credit and product information that can be used to understand a customer's credit or financial behaviour or attributes

Each row of data represents a view of a specific customer (a customer observation) and each column describes an attribute about that customer (a variable)

An accompanying data dictionary describes the variables and their format (character or numerical)

LLOYDS BANK DIGDATA CHALLENGE – DATA DICTIONARY (EXCEL SPREADSHEET)

- The data dictionary contains the file name and the number of row/observations (customers)
- The columns represent the fields/variables, and the data dictionary contains:
 - Field Number – which column represents the variable
 - Field Name – the variable name this is usual a shortened version of the description
 - Data Type – whether the variable is formatted as a character (e.g., contains information that is read as text by your analytical tool like employment title) or numerical
 - Field description – definition for variable - is especially useful when the field names are not intuitive

See illustration:

File Specification

File Name: data.csv
No. of Records: 18,324

Field Number	Field Name	Data Type	Description (optional)
1	id	Unique Identifier	A unique ID for the loan listing.
2	addr_state	Categorical Variable	The US state provided by the borrower in the loan application (address)
3	annual_inc	Numeric Variable	The annual income provided by the borrower during registration.
4	emp_length	Categorical Variable	Employment length in years. Possible values are between 0 and 10 where
5	emp_title	Categorical Variable	The job title supplied by the Borrower when applying for the loan.
6	home_ownership	Categorical Variable	The home ownership status provided by the borrower during registration
7	installment	Numeric Variable	The monthly payment owed by the borrower if the loan originates.
8	loan_amnt	Numeric Variable	The listed amount of the loan applied for by the borrower.
9	purpose	Categorical Variable	A category provided by the borrower for the loan request.