Welcome to the “Step Up” – Lloyds Banking Group live online career challenge

What will you need?

- A good wi-fi connection
- A notebook and pen
- A quiet environment where you will not be disturbed for 45 minutes!

What do we expect?

- Respectful and appropriate behaviour throughout this 45-minute briefing session
- A positive attitude and a willingness to learn
- There will be a number of opportunities throughout today’s presentation to ask questions using the chat function.

How will you benefit from completing this challenge?

- The opportunity to solve Lloyds Banking Group’s business problem
- To learn about how your curriculum learning at school is used in the “real world”
- Understand how valuable YOUR “Key skills” are when solving this problem and delivering the solution.
- Gain a virtual work experience certificate

Any technical issues during this session – please email us on info@digdata.online

Lloyds Banking Group Data Challenge

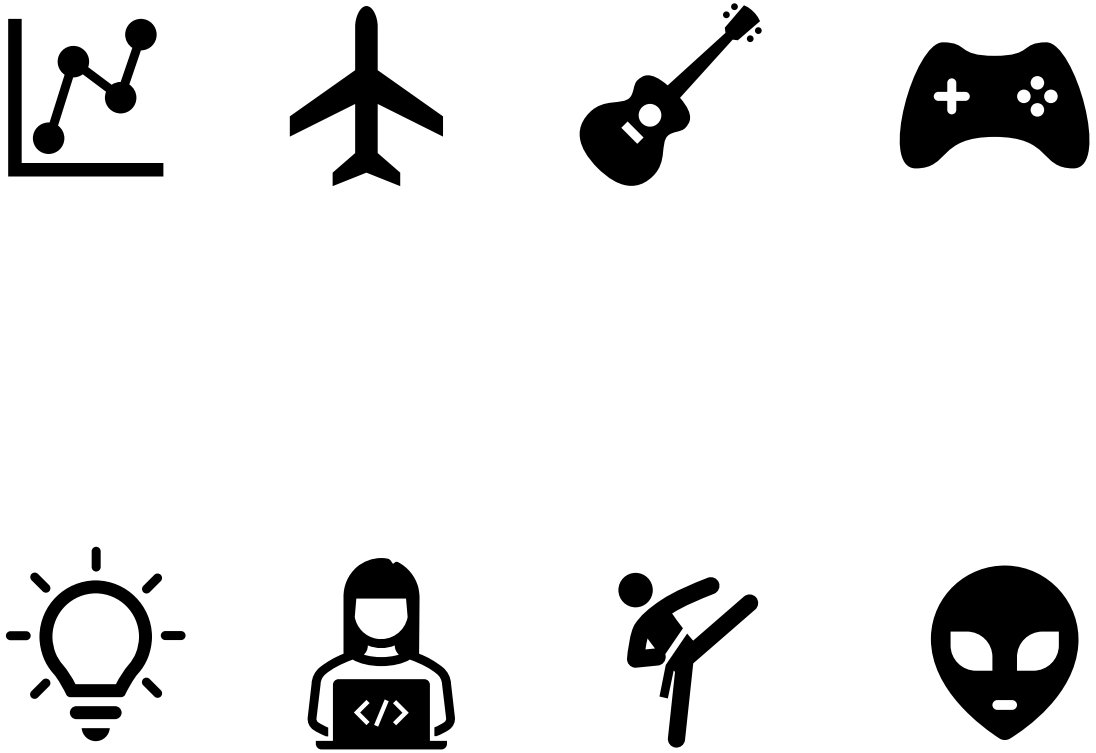
David Albizzati, Sukhvinder Panesar

9th November 2022

Meet the team

David Albizzati
Senior Manager

Sukhvinder Panesar
Lead Data Engineer



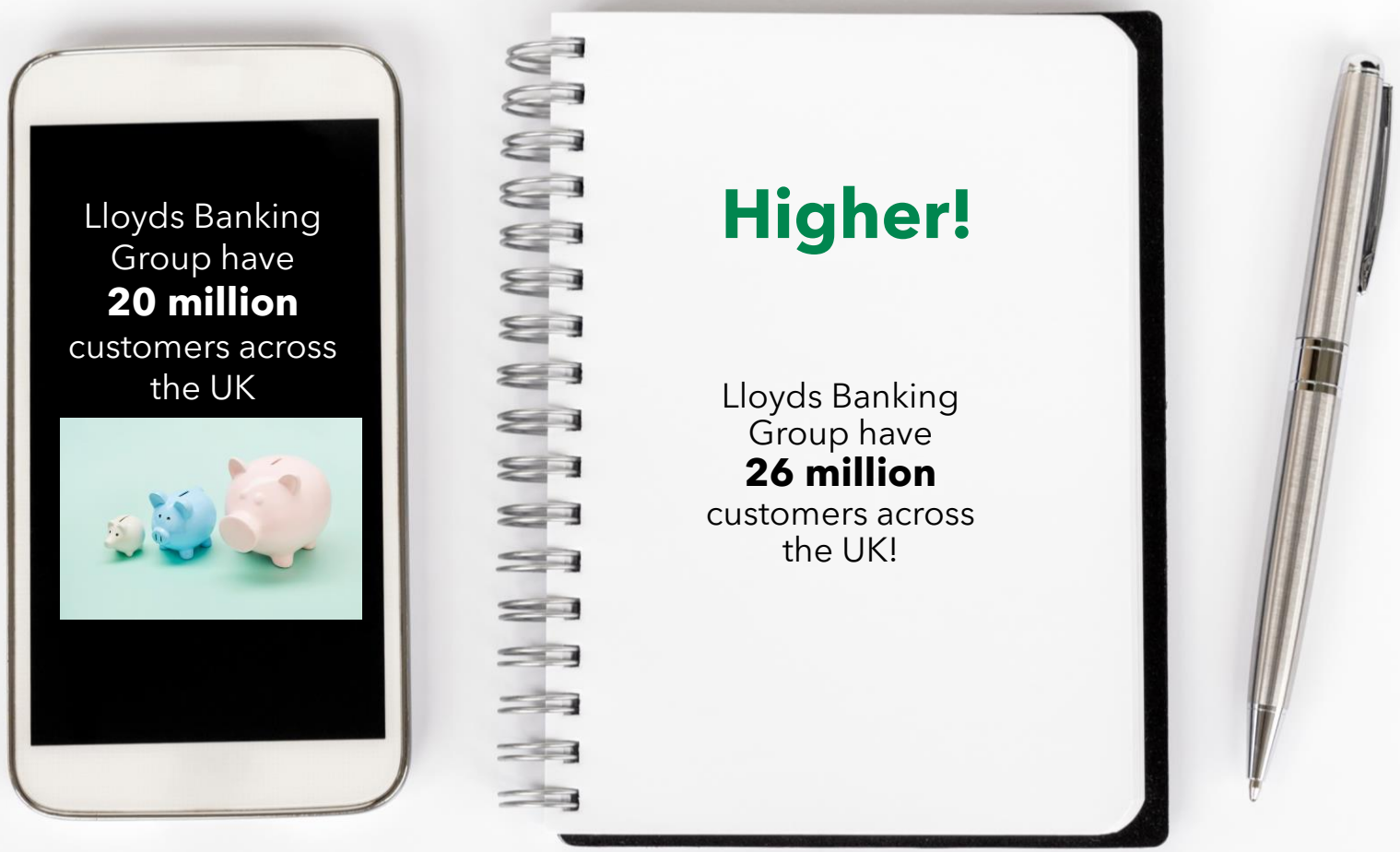
Elaine Corkery
Senior Manager – Data Science

↑ Icebreaker Challenge ↓

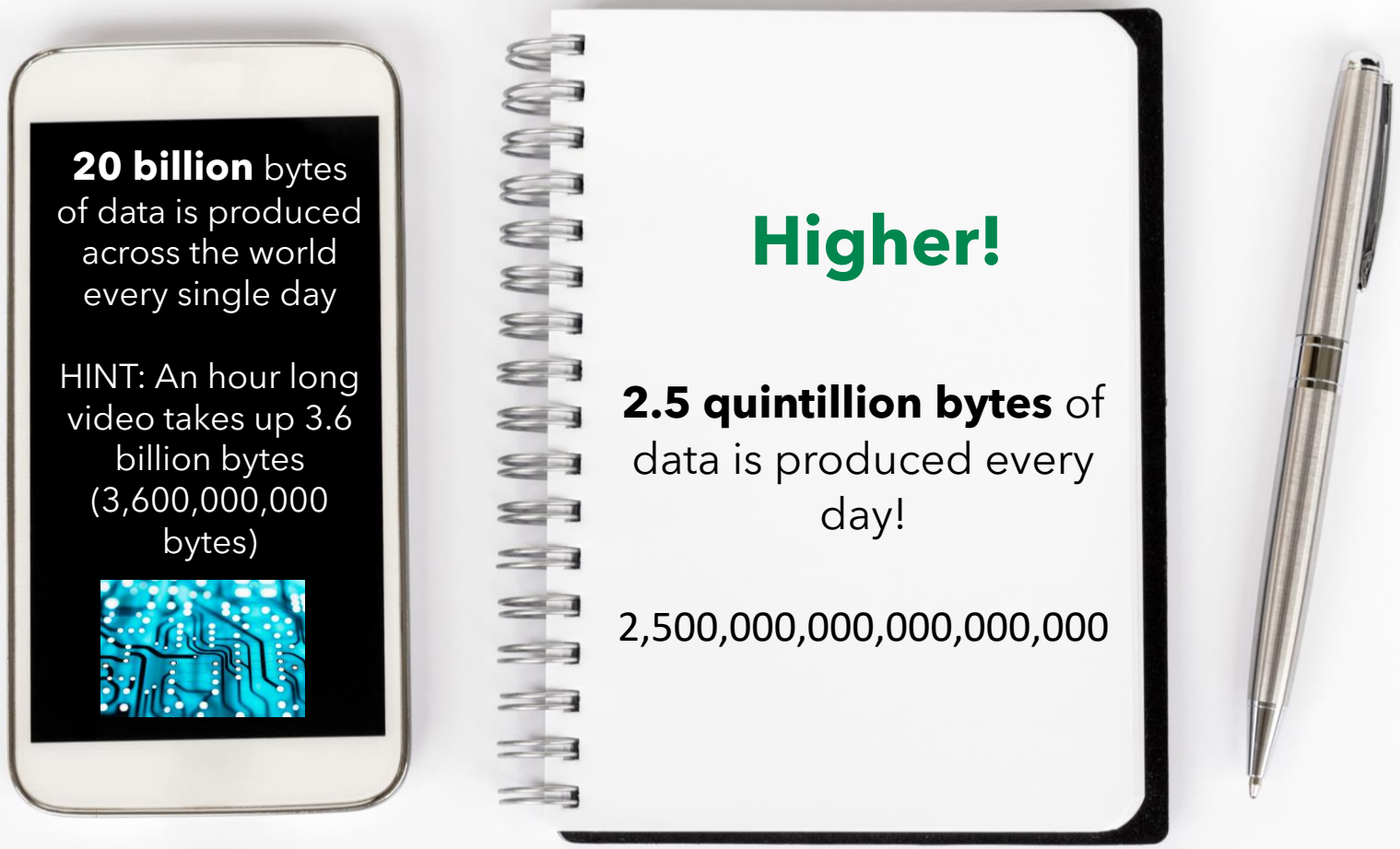
Higher or Lower?

Higher or Lower?

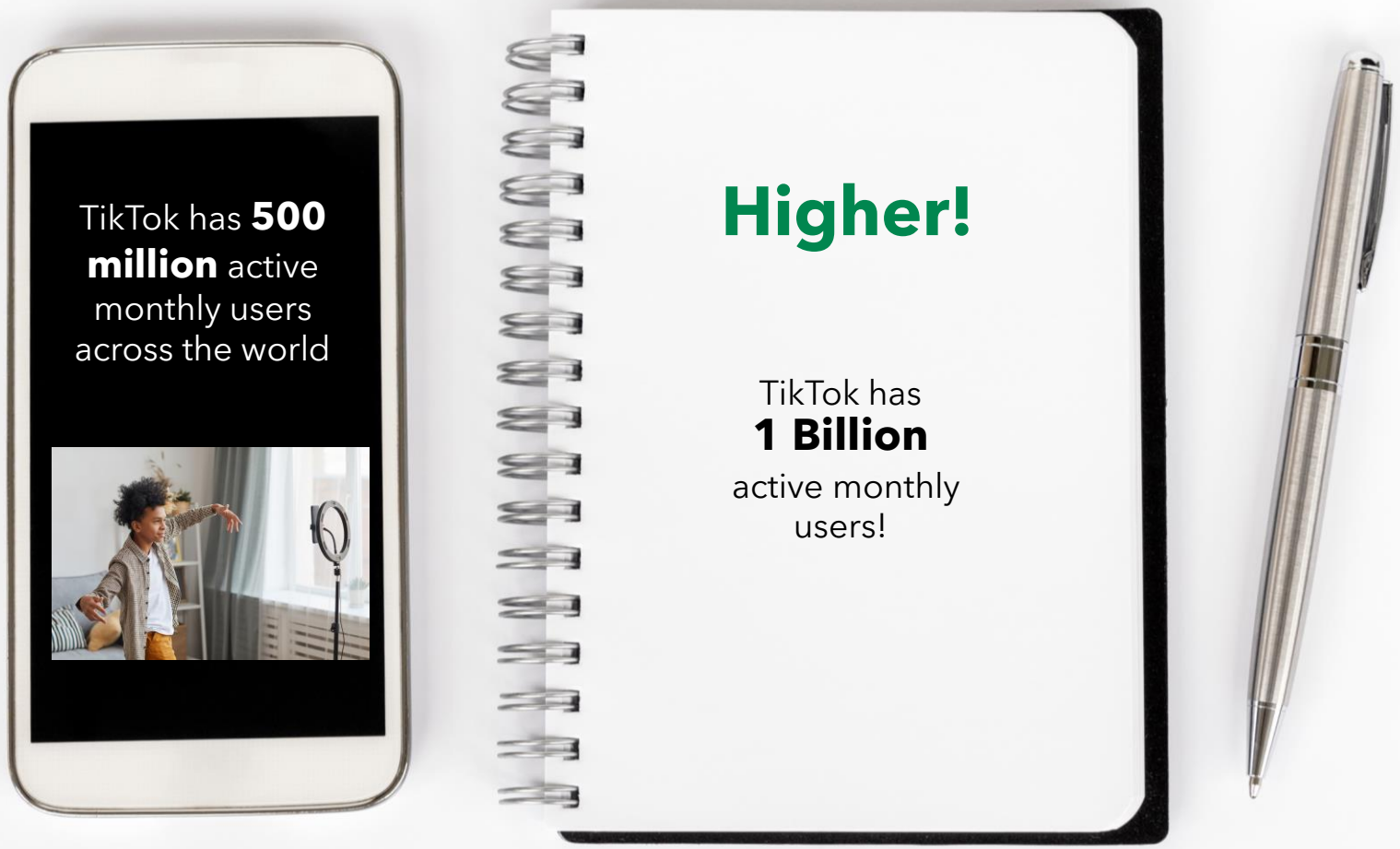
Higher or Lower?



Higher or Lower?



Higher or Lower?



Lloyds Banking Group

Who are we?

Who are Lloyds Banking Group?

UK's largest Retail Bank

Includes many household names such as Lloyds Bank, Bank of Scotland, Halifax & Scottish Widows

We have been serving businesses and communities of Britain since 1765

We supported 1,000 Olympic and Paralympic hopefuls through our Local Heroes Programme

Around 58,000 employees

26 million customers in the UK

What do we use data for?

Data is everywhere today, pretty much anything we do involves data

Some of the data Lloyds Banking Group gathers:

- Customer details
- Customer transaction data
- Text from online chats
- Audio from call centres

It's not just numbers that we work with!

What tools do we use?

There are a wide variety of data roles across the organisation and LBG provides colleagues with the right tools for each role. Such as:

- Tools to summarise, query and transform your data (e.g. Excel, SQL & BigQuery)
- Visualisation tools to help you tell a story using data (e.g. PowerBI, Tableau)
- Analytics tools (e.g. SAS, Python, Google Cloud Software)

Training and support is provided when upskilling on a new tool or technology.

The Challenge

Challenge Overview

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

- **Task 1 – Data Strategy :** Understand and summarise the different behaviours or attributes between customers who paid back their loan and customers who did not
- **Task 2 – Data Science:** To use the historical data to design a process which predicts the likelihood of a new customer not paying back their loan

The Data

Task

Lloyds Banking Group has sent you a dataset of 18,324 customers who:

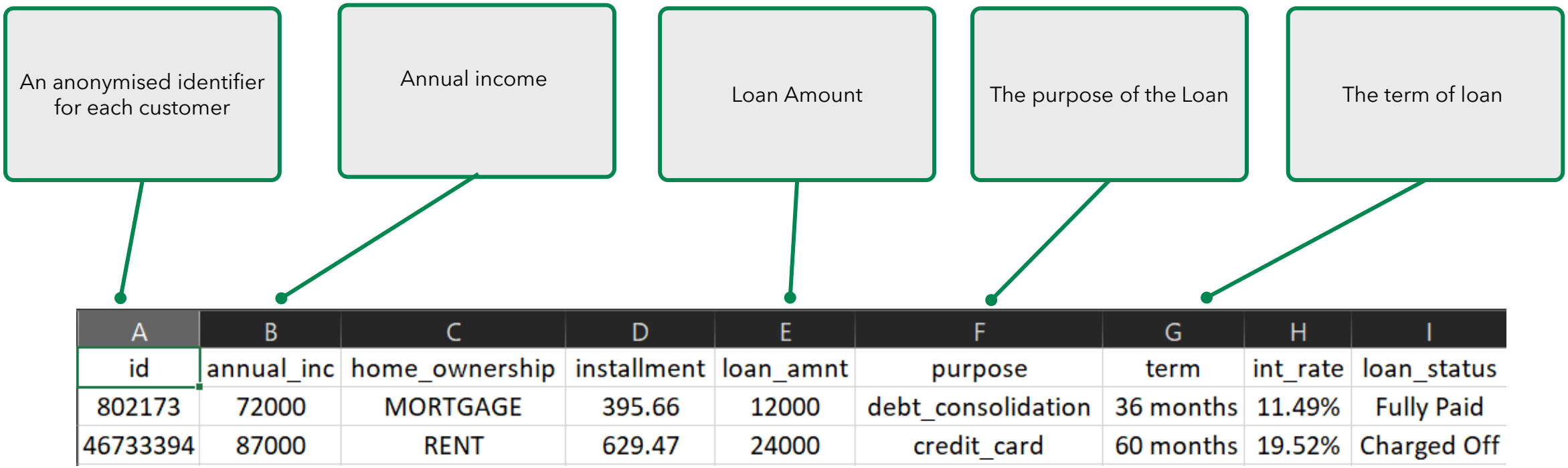
- Previously held a loan
- The status of that loan - did the customer pay back the loan or not:
  - Customers who paid back are categorised as 'Fully Paid'
  - Customers who did not pay back their loan are categorised as 'Charged-off'
- Other credit and product information that can be used to understand a customer's credit or financial behaviour

Please note the dataset is based on an American credit risk problem and as such some variables will be different. For example, annual income in US will be much higher compared to annual income of UK population. The premise and spirit of the problem statement is very relevant to the UK.

Understanding the data

The dataset will be provided to you in excel there are 31 columns which represent a customer attribute (variable).
A data dictionary which describes the type and definition of each variable will also be provided.

It will have data for customers on various transactional or credit attributes which looks like this:



A	B	C	D	E	F	G	H	I
id	annual_inc	home_ownership	installment	loan_amnt	purpose	term	int_rate	loan_status
802173	72000	MORTGAGE	395.66	12000	debt_consolidation	36 months	11.49%	Fully Paid
46733394	87000	RENT	629.47	24000	credit_card	60 months	19.52%	Charged Off

Data Strategy Task

Task 1 Data Strategy – Customer Profiling

Understand and summarise the different behaviours or attributes between customers who paid back their loans and customers who did not

The objective of this task is the create a visual summary which profiles the customers who pay back their loan and those who don't.

- Use simple data queries, using the variable 'loan status', to determine the different attributes or behaviours between customers who paid back their loan or not
- It is up to you to determine what data to include or variables to analyse and how to display the results. The key success criteria is that the summary should be clear, visual and easy for your non-technical business stakeholder to understand.
- Summarising your findings in a PowerPoint slide can be a helpful medium to tell your data story but please be as creative as you want!

Hints and tips – Considerations when approaching analysis

- Given you want to determine the difference between two groups of customers, it is important to use **the 'loan_status' variable** when completing your analysis
- You can use any tool you would like to complete this analysis. However, we recommend using at least one or two visuals/graphs to summarise the information to ensure it is engaging for the end user.
- The dataset contains both numerical and categorical variables. Use the accompanying data dictionary to determine the format of the variable and use the appropriate metrics for each type.
- If you are completing the analysis in excel. You can create formulas directly in a cell or use the drop and drag pivot table/chart functionality.

Row Labels	No of Customers
Charged Off	3906
Fully Paid	14418
Grand Total	**18324**

Hints and tips – Things to be aware of

- In real world data problems some variables have **missing information**, it is important that you treat the data correctly. Helpful questions to ask when considering missing values:
  - What proportion of the variable has missing values?
  - How reliable is the variable if there is a significant % of customers with missing information?
  - Does a missing value have an interpretable meaning and therefore would be useful information if handled correctly? For example, the absence (*missing*) of recent missed payments on a lending product is likely to indicate good credit worthiness.
- When working with numerical variables it is important to **consider the distribution/shape of the data**. Please be aware, outliers can skew a summary statistic like an average value. Therefore, consider if you need to clean your data beforehand or choose an appropriate metric for your analysis.

Data Science Task

Task 2 – Predictive Model

Use the data to design a process which predicts the likelihood of a new customer not paying back their loan

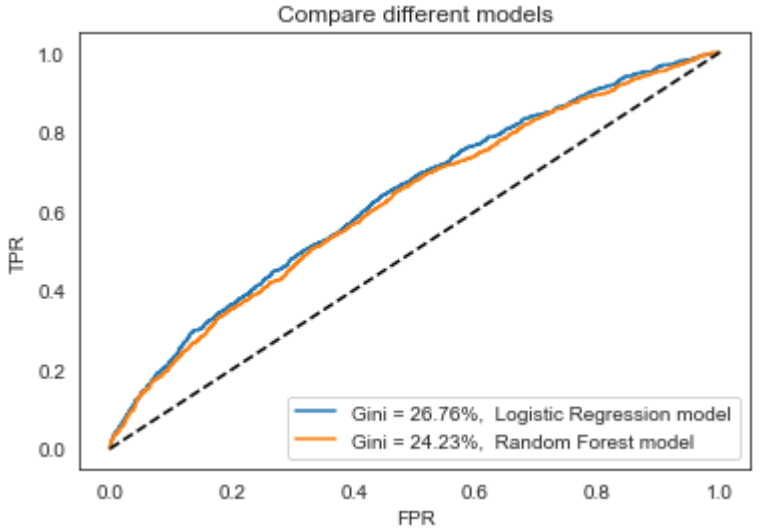
The objective of this task is the create a model predicting the likelihood of a customer paying back their loan:

- Building a model in whatever environment you feel comfortable (e.g. Python, Excel, R, etc.)
- Use whatever approach you see appropriate (e.g. logistic regression, decision trees, random forest, etc.)
- Evaluate the performance of your model
- Explain your model (e.g. what are the most predictive features and do they make logical sense)

Task 2 – Model outputs

Once you have built the model, assess the model outputs, here are some suggestions:

- Investigate what features are driving your model prediction (e.g. most weighted coefficients)
  - Do these features seem reasonable?
  - Is their relationship with the target variable as expected?
  - Are the features statistically significant?
- Further assess the performance of model (e.g. confusion matrix, Gini, etc.)
- If you have used different modelling techniques how do they compare with each other? Do they produce similar results (in terms of performance and most predictive features)?



Telling the story

Playback your results

- The key to using data in the commercial world is **telling a story** and providing insight.
- Data and its insights can be presented in a number of ways using tables, charts or even **using visuals to tell your story** and really bring it to life.
- You can produce your results as **creatively as you like** with graphs, visuals, infographics – or even through written form.
- Be clear on the **"so what"** from all your analysis – so what does this mean to the **customer**, the **Bank**, **employees**?
- Make your findings and recommendations really **clear and simple** to understand.

Next Steps

Find us on social media
@digdatagroup

Next Steps

What happens now?

Digdata will send you:

- Briefing document
- Data set (Excel worksheet)
- Data Dictionary (Excel worksheet)
- Recorded session link

When is the Lloyds Banking Group career challenge deadline?

Midnight on Friday 18th November

How can you present your solution?

- Infographics
- Graphs / charts
- Code
- Text / bullet points

Send all of your work to info@digdata.online and you will receive your virtual work experience certificate within 7 days.

Thank You!

Good Luck!