**Social Buzz project:**

What you'll learn

How to clean, model and analyze data to create valuable insights for the client

What you'll do

Identify which datasets will be required to answer the client’s business question

Clean the datasets and merge them to prepare the data for analysis

Determine the answer to the client’s business question

data sets - each data set contains different columns and values

A data model - this shows the relationships between all of the data sets, as well as any links that you can use to merge tables.

- focus on the analysis of sample datasets with visualisation to understand popularity of different content categories.

- 3 Client requirements: audit of big data practice, recommendations for IPO, analysis of popular content.

- client has a scale issue, and that they don't have enough resources to manage their business scaling problem. So they're looking for help with the management of their journey into such a large scale.

Business question example: which is to figure out the top 5 categories with the largest popularity.

\* You’ve identified Reaction, Content, and Reaction Types as our relevant data sets.

\*client wanted to see “An analysis of their content categories showing the top 5 categories with the largest popularity”.

\*As explained in the data model, popularity is quantified by the “Score” given to each reaction type.

\*therefore need data showing the content ID, category, content type, reaction type, and reaction score.

\*figure out popularity, we’ll have to add up which content categories have the largest score.

\*Before we begin to work with the data sets, we’ll need to ensure that the data is clean and ready for analysis…

to make sure you are using the right data to answer the business questions you’ll follow these steps:

Requirements gathering- first step is to use this data model to identify which datasets will be required to answer your business question

Data cleaning- Clean the data by:

\*removing rows that have values which are missing,

\*changing the data type of some values within a column, and

\*removing columns which are not relevant to this task.Think about how each column might be relevant to the business question you’re investigating. If you can’t think of why a column may be useful, it may not be worth including it.

Data modelling-

1. Create a final data set by merging your three tables together

We recommend using the Reaction table as your base table, then first join the relevant columns from your Content data set, and then the Reaction Types data set.

Hint: You can use a “VLookUp” formula

2. Figure out the Top 5 performing categories

Add up the total scores for each category.

Hint: You can use the “Sum If” formula

The end result should be one spreadsheet which contains:A cleaned dataset,The top 5 categories

Definitions of different data types:

String - Sequence of characters, digits, or symbols—always treated as text

UUID - Universally Unique Identifiers

Array - List with a number of elements in a specific order—typically of the same type

Integer - Numeric data type for numbers without fractions

Timestamp - Number of seconds that have elapsed since midnight (00:00:00 UTC), 1st January 1970 (Unix time)

Source: Direct extract

Storytelling:

How to connect business and data and how to communicate relevant and engaging insights in a presentation.

\* Choose data visualizations that best support the story you want to tell the client.

\* You should plan your presentation structure, ensure presentation answers the right questions.

PowerPoint

\* For each slide, think about:

1. Agenda: what will your presentation cover?

2. Project recap: what are the key points from the brief?

3. Problem: what is the problem that you answer in the presentation?

4. The analyst team: who's your team?

5. The process: how did you complete your analysis? Outline the analysis steps.

\* Populate 2 to 6 slides for each objective question. Use the data analyst presentation template from forage job simulation.

NB: information should be clear, concise, and summarise the most important points.

\* You can include extra valuable insights in your presentation that you found while doing your analysis. For example include these in your forage job simulation: remember the client wanted to know the top 5 content categories?

- how many unique categories are there?

- how many reactions are there for most popular categories?

- what was the month with the most posts made?

\* Provide answers to these questions in your presentation.

Presentation tips:

\* Clients may not be familiar with data, so try to present in a business friendly language.

\* You want to show that you understand how their business works and requirements. Use their terminology and language relevant to the task.

\* This your time to shine, show how great you are! Try to present your work with confidence and conviction. Or else it'll seem like you don't understand your results fully. Speaking about your results should feel natural when you speak about what you have done. 05111

\* A good presentation always takes a lot of practice or preparation, also helps build confidence. You can do a mock presentation and then ask for feedback.