

**Data Science Technical Task**

**Scope of task:**

This task is to test data science competencies from both technical skills and the ability to present and evaluate findings to several key stakeholders.

**Languages:**

**- Python**

**- R**

**- SQL**

**Packages:**

SciPy, Matlab, Tensorflow, Numpy, MatplotLib, Keras, Pandas

**Brief:**

Kids Insights are aiming to develop a product that will enable clients to gain further insights in the attitudes, behaviours and consumption patterns surrounding the films kids have watched, are looking forward to and the purchasing habits related to them.

By using these questions from the survey, and data science techniques, it is hoped that Kids Insights can support clients with their decision making and inform their strategies with enhanced insights and predictions.

**Objective and deliverables:**

Utilising the data set provided, please write an algorithm(s) that would help to yield in-depth analysis and insight into the film preferences of children.

You are required to identify any useful information, insight and patterns to inform several key decision makers within the business. You are also required to explain your approach and why the techniques deployed would help reach the goal of the brief.

Please note any findings must be suitable for both technical and non-technical staff throughout the business.

Note: external data is allowed to be utilised to help findings

Technical info for the data:

**The Data Sheet:**

The spreadsheet is comprised of 3 pages, with 2 years of data:

* “Questions” the 3 questions asked of participants, 2 are multiple choice, the other is a freeform cleaned answer.
* “Profiles” – basic information on the participants. Date of survey taken, Age, Gender and Location.
* “Answers” – List of answers for the 3 questions.

**The Data Model:**

The Data model consists of two basic tables, joined on the record\_id column. The answers table in the “Film Data” sheet also consists of a “question\_code” column to identify which of the 3 questions the answers are from.

In the case of the multiple-choice questions, more than one answer may appear for each record\_id and question code reflecting it being multiple choice. In the case of the cleaned freeform answers, a number of record\_ids may not have an associated answer, this is where cleaning routines have removed spoilt answers.