Tests for data scientist position in Audience Origin team at **choreograph**

There is a set of 3 questions for the Data Scientist position at **choreograph**. A candidate is expected to do two of them with a freedom to choose which ones. The exercise should not take more than 2 hours per question. The goal of the exercise is to assess the thinking process of the candidate more than an expertise in specific DS, ML or AI technology.

The results should be presented on slides (Power Point, pdf, etc.) with introduction, outcomes and comments sections. There is a preference to use python and its tools stack in answering those questions. It is expected to attach a notebook(s) or a script file(s) along the slides with outcomes. The results should be sent at least a day before the interview day via email. The second stage interview will take around 1 hour, potentially a bit longer. In the first part a candidate will present the results after which we will have a conversation and questions session.

1. Analysis of two types of weights for one market.
   1. **Intro:** There are two variables ‘TA\_Weight’ and ‘TA\_Weight\_NCCS’. Both are correctly calculated weights that can be used to weight the survey results. However, they were created for a different target audience.
      1. ‘TA\_Weight’ was calculate for ‘Age - Gender’ and ‘Region’,
      2. ‘TA\_Weight\_NCCS’ was for ‘Age - Gender - Region Zone - Social Grade’
   2. **Task**: The task is to estimate and comment on the impact of those two weights on media consumption variables ‘DIA\_xx\_NET\_1’. Calculating the *weighting efficiency score* for each weight column may help you.
   3. **Files**: “test\_data\_weights.sav”, “loading\_sav\_example.ipynb”
2. Covid Attitudes vs. Covid Financial Pressure.
   1. **Intro**: We run a survey to measure attitudes and financial measures taken during Covid in 2020 in two waves of surveys across 53 markets. Attitudes are stored in ‘QP07\_xx’ vars and financial measures taken are captured in ‘CAT05\_xx’ variables.
   2. **Task**: Segment data using Covid Attitudes and Financial Measures.
   3. **Files**: “test\_data\_covid\_attitudes\_and\_purchase\_recency.sav”, “loading\_sav\_example.ipynb”
3. Quality of Purchase answers across several markets.
   1. **Intro**: Respondents are not always reliable and trustworthy of the answers they give in surveys, we sometimes call them” flatliners” or “speeders” because of the way they fill in a questionnaire.
   2. **Task**: Explore the quality of recency of purchase decisions ( ‘CAT02\_xxx’ variables) declared for tens of categories of products across 53 markets. Propose a risk index for respondents based on their answers to those questions and group markets based on that index.
   3. **Files**: “test\_data\_covid\_attitudes\_and\_purchase\_recency.sav”, “loading\_sav\_example.ipynb”