**GMDAC Data Analytics Consultant**

**Test**

4 June 2020

Estimated time for completion: 4-6 hours

**Principles**

* Don’t worry if you cannot complete all of the questions. It is better to submit whatever you have than submitting nothing.
* There is not one correct way of answering the tasks. Many different solutions are possible. We generally like tidyverse approaches but anything goes.

**Rules**

* You can submit one or more files. The fewer the better.
* Whatever you submit needs to be 100% reproducible

**Criteria**

* The less lines of code you need to answer the tasks, the better
* User friendliness – is what you produce easy to digest and comprehend for lay audiences?
* Creativity
* Sophistication
* Completeness

**Data**

The excel file contains data on 1000 university students at the Berlin University. Each row represents one student. Students are nested in courses and faculties. The second file contains data by faculty.

file 1 – student data

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Age | Age |
| Cob | Country of birth |
| Course | Current university course |
| Faculty | Enrolled at which faculty |
| gpa\_2010 | Grade Point Average in 2010 |
| gpa\_2011 | Grade Point Average in 2011 |
| gpa\_... | And so on |
| job | Does the student have a student job? |
| lifesat | Life satisfaction score |
| like | How much do you like your course? |
| relationship | Are you in a relationship? |
| sex | Male/female |
| term | What term are you in? |
| university | Location of University |

file 2 - Faculty data

|  |  |
| --- | --- |
| **Variable** | **Description** |
| profs | Average number of students per professor |
| salary | Average starting salary after graduation |
| cost | Average cost of the 5 year programme |

**Tasks**

**Data manipulation & Descriptive stats**

1. Create a summary table with a) the percentage of non-German students, b) average life satisfaction, c) percentage of students in a relationship, d) sex ratio, e) percentage of students over 30, f) average gpa in 2010 and g) average number of terms per faculty.
2. Create a visual showing differences in average life satisfaction by faculty and relationship status.
3. Combine faculty data with student data into one dataset. Show differences in the average cost of the career by faculty and job status.
4. Create a visual showing the relationship between life satisfaction and age.

**Modelling**

1. Check whether the relationship status has an effect on life satisfaction regardless of the number of terms, age, sex and the expected entry salary after university. If possible, visualize effects.
2. Test whether having a job has a negative effect on the average grade point average over the last 10 years. If possible, visualize effects.
3. Forecast the grade point average for next two year (i.e. 2021 and 2022). You can ignore model assumptions (anything goes). Extra points: Visualize different forecast scenarios.
4. For each year between 2010 and 2020, estimate the predicted probability of having a grade point average above 2 for students with a job vs. students without a job (controlling for sex and age). (hint: Looking for a loop).

**Visualization, user experience and presentation skills**

1. Create an application or visual where users can select the data they want to see and the breakdowns they are interested in (be creative, anything goes).
2. Record a max. 3 min screencast or audio of you walking the user through the app and explaining how it works and what the results are.

**Maps**

1. Create a map of the countries of birth of students. The map should provide information on how many students at the University of Berlin were born in each country.
2. Include in the map the location of the University of Berlin.

Good luck!