# Tasks and Challenges (estimated completion time 4 hours)

If you have any questions concerning this task please feel free to reach out to us at any time.
You can contact us at: karriere@team-con.de

## 1 Probability of passing an exam

**Task 1. A**. You have a friend who is applying for a public job. The way the exam works is as follows:

- There are 69 topics.
- On the day of the exam, 5 topics are chosen at random in front of the students.
- Each student chooses a single topic from these 5 randomly selected topics.
- If your friend knows the topic she has chosen and passes the exam, she gets the job.

*Examples. There are 69 topics: T1, T2, ..., T69. On the day of the exam, the topics **T2, T8, T29, T50, T51** are randomly selected.*

- *Student A has prepared topics T1, T2, and T3. Student A will decide to take T2.*
- *Student B has prepared T50, T51. He will choose either of the two topics to take the exam.*
- *Student C has prepared T1, T3, T4, T5. Unfortunately, he has not prepared any of the randomly selected topics. Student C will fail.*

Your friend wants to know how likely she is to pass if she studies 20 of the 69 topics perfectly. Help her to calculate mathematically the probability of passing if she focuses only on studying 20 topics; assuming that she will pass if at least one of the topics she has studied is selected.

**Task 1. B.** Make a small script using Monte Carlo methods to prove that the theoretical probability you have calculated in the previous section (Exercise 1, part A) is empirically correct. Using your script, what's the probability of passing the exam if she prepares 15 topics?

## 2 Classification

**Task 2**. Using the provided dataset, in a Jupyter Notebook:

- Do an exploratory data analysis.
- Transform the data.
- Train a model which predicts the class of the object (column "class") [GALAXY, STAR].
- Test and explain the performance of your model.

You are free to choose the type of model. At every relevant step, add a comment explaining **why** you made that decision, **what you are trying to achieve**, etc… The more comments, the better.

*Variables description:*

*objid = Object Identifier*

*ra = J2000 Right Ascension (r-band)*

*dec = J2000 Declination (r-band)*

*u = better of deV/Exp magnitude fit (u-band)*

*g = better of deV/Exp magnitude fit (g-band)*

*r = better of deV/Exp magnitude fit (r-band)*

*i = better of deV/Exp magnitude fit (i-band)*

*z = better of deV/Exp magnitude fit (z-band)*

*run = Run Number*

*rerun = Rerun Number*

*camcol = Camera column*

*field = Field number*

*specobjid = Object Identifier*

*redshift = Final Redshift*

*plate = plate number*

*mjd = MJD of observation*

*fiberid = fiberID*

***class = object class (galaxy, star)***