BA Questions

# Day 0

**In Business.csv, do blanks mean the trainee has been kicked off the course?**

Yes, the trainee is no longer on the cause

**What is the MVP?**

A populated data store and a dashboard and the final presentation

**What are the scores out of (for programming languages)**

Those are self-rated scores so technically not out of anything. As a safe bet, go with 8 as maximum. Look at the data and see what the max and min is and choose whatever you think is best.

If a trainee has nulls, they have been removed but consider the length of course.

**Do you have any special requirements with how the data is stored? We are currently leaning towards Azure.**

Pipeline will provide data for analyst so choose whatever will make it easiest for analyst to deal with.

**What tools can we use to work collaboratively using Databases (e.g. E2C instance)**

Docker is local and use PyCharm to create the database.

**Determining trainees over several files**

Need to do data analysis and be sure that you can uniquely identify and match individuals over multiple files

# Day 1

**What values are essential in the talent applications? (Phone number, email etc)**

Want as much information ideally. Null values are ok.

**What about those with no record of being invited to an applicant day?**

**For missing values in Talent-MonthYear, can we fill in the value from the file name?**

Yes, the month field can be filled in using the filename.

Where invited\_date value is empty, you can attempt to cross-reference with another file to determine that date.

**How do you want us to deal with missing data. For example, some applicants do not have university or degree information. Is this important? Do you want to know this information?**

Want as much information ideally. Null values are ok.

**Cleaning Phone Numbers**

Ideally no additional characters (so remove -, and ()). We want it to be easily queryiable by analysts.

## Day 1 More Questions

**How did the old Sparta process work? Did they do an assessment first and then come in for a Sparta Day?**

1. Candidate info contact details collected
   1. Take self-assessment
   2. Invited to Sparta Day (psychometric, presentation)
   3. Interview
   4. Get the results

**Is the result field in Talent.json the result of the Sparta\_Day (where they do the tests or is it when they enter their self-scores)?**

This happens on the Sparta Day

It is the result of the Sparta Day. The Sparta Day is when they take the self-assessment.

# Day 2

**Is InvitedBy referring to being invited to a Sparta Day or invited to the program?**

InvitedDate is to do with the sparta day and not the program

Only date that will be different will be start date of the course

**Who are the people inviting applicants to Sparta? It appears to be non-spartans.**

The invitee is a talent co-ordiantior – members of staff within sparta

**If an applicant/candidate passes the Sparta Day, can we assume they will become a Spartan. It is possible that they pass and not join.**

**There appears to be one trainer per course from the dataset. Is it possible in the future that a course may have multiple trainers?**

In this situation, futureproofing is good so keep it in

**In the talent.json files, there is a date field. What is this referring to? The data suggests that this is the Sparta Day date.**

**The SpartaDay.txt field contains the date along with the day of date. Do you want us to store both the day and date? Sully thinks we should store them separately.**

Just store date as full date

**What should we do if there’s data that doesn’t pass validation and can’t be corrected? Null it?**

Any kind of data cleaning, we don’t just get rid of it, flag it, print it, show the product owners, double check before dropping anything

**Can an applicant be interested in one stream but be put in another?**

Yes

Date for when course starts

Results better in fact table than information table (Danny), Paula personally doesn’t mind

Review datatype for IDs

Quick extra questions:

## UNANSWERED

**Do we need to provide an interface for Data Analyst to get the data?**

No. They will query the database directly. We need to create a database and make this available to Data Analyst.

Pushing clean data to S3 is not a requirement but we are allowed to upload clean data to S3. However, data analysts will be accessing data via an SQL database, not S3.

**Or can we assume they will query the database directly? In other words do we need to create a Python interface? If so do we need to implement OOP?**

* We can have just functions. We do not *have to have* classes.
* We definitely need tests, whether we use TDDs or not. Paula recommends TDD, it will help in the long run.
* Need to consider packaging it up like we did for MongoDB.
* OOP means we don’t have to have everything in classes, just the things we need. So we can use a combination of both.

**What should we do with tying the Unique Key to the Spartan Database?**

Not stress essentially, as none of the current possible methods are completely futureproof, we’ve shown the duplicated names do not make it into the academy so we don’t need to worry about it. Assert in the presentation that, in the future, an applicantID that gets assigned at the start of the application process would be a massive improvement