All Tasks

* Clean data – Remove NAs, participants (moderators + guests?) – need to ask about this first
* **Remove the date from ‘first join’ and ‘last leave’ columns and make this a new separate column**
* Clean up the names in the ‘sessionType column’
* **Make a new column for whether it’s the first or second session of the week**
* **Make a new column for whether it’s one or two hours long. Based on how long the moderator is there for.**
* **Make a new column for whether it’s a morning or evening session**
* **Create function that takes each course from the original dataset and creates a new dataframe with just that course**
* Identify any strange sessions + ask Brittany about this
* Filter data by those who have consented
* Visualise the following for each course:
* How many times participants joined in a given session – line graph. Grouped by students vs moderators + evening vs morning session
* Total time participants spent in a given session – grouped by students vs moderators and evening vs morning session
* When a participant first joined + when they last left – grouped by students vs moderators and by morning vs evening session
* How many participants joined a given session – grouped by students vs moderators. Frequency on the y-axis + session on the x-axis. Color = role
* The number of participants who ALWAYS joined either the morning OR evening session, those who joined BOTH each week + those who joined a mix of both sessions across the course
* How many sessions a given participant attended in total
* Change all of this code so they become functions that can be used with the entire dataset
* Connect with Github so I can handover to Pawel
* Write assumptions at the top
* For anomalies, add another column for whether they were long or short
* Make summary table in wide format, pivot\_wider

Sprint 1:

* Create function that takes each course from the original dataset and creates a new dataframe with just that course
* Remove the date from ‘first join’ and ‘last leave’ columns and make this a new separate column
* Make a new column for whether it’s a morning or evening session
* Make a new column for whether it’s one or two hours long
* Check if there are any weird columns E.g. Person was only there for a few seconds or forgot to leave. Filter by the first column for ‘0’ and check if anyone was there for under 5 minutes total.

Sprint 2:

* Make a new column for whether it’s the first or second session of the week
* Hide columns that aren’t necessary
* Clean names of columns
* Make note of anomalous sessions
* Tell Pawel about things that have come up already
* Create a small summary of what I’ve learnt about the dataset
* Message Brittany about setting up meeting to talk about potential anomalies + asking her questions about her course.

Sprint 3:

* For people like ‘Lamis’, code something that means his alternative names are registered as part of the same entry + total time in session combines the total time values for these different entries.
* What to do about Sophie & Elaine who only go to one welcome session but aren’t course staff? Assign as moderators
* Meet with Brittany to talk about anomalies + questions I had
* Clean data – Remove NAs + anomalous values based on what Brittany said.

Sprint 4: Visualisations

Summary:

Note: The following summary is based off of one course (HEIN110372023-4).

* Som

With this course, it was very easy to determine which week a session was run as the sessions were clearly labelled with the week, type of session and the time e.g. ‘Week 2 [..] Tutorial 10 am’. In terms of the total time a participant joined a session, most of the time the dataframe registers the participant as the same individual despite them joining multiple times and calculates the total time accordingly. However, there are certain instances where a participant is registered as different entries for the same session. For example during the Week 6 Data Visualisation and Storytelling morning tutorial, the student ‘Lamis Barakat’ first joins the session using their integrated account. His total time for the session is registered as 6 seconds. However, there is another entry directly under titled ‘lamis’ who joined as a Guest and does not have a UUN who stayed for 52 minutes. These alternative spellings and versions of a person’s name need to be considered.

There are many instances of attendees joining as a guest. This can cause issues as the participant can then type in their own username, such as ‘lamis’, and may not be registered as the same individual. A lot of guest participants only include their first names so will need to create a key where just first names and full names are registered as the same person. This is easy to do by filtering dataframe with AttendeeType = Guest and matching names. Guest users also don’t have UUNs. Furthermore, some guest users are not students, for example in the Course Welcome morning session both Elaine Mowat and Sophie Marion de Procé are registered as guests and were assigned the participant role. It is difficult to differentiate between a student and non-student in this case unless you know the staff members.

Going through the ‘anomalous’ cases, which were entries that had total times of less than 10 minutes or more than 90 minutes for an allotted 1 hour session, there are nuances to be considered. With people who only joined for a few seconds this may be due to accidentally starting the meeting before the allotted time or due to technical issues. With technical issues ensure that the student has not joined as a guest and been registered as a different user. Some students also joined for a few seconds just before the session ended, which could suggest they mixed up timings due to timezones or being busy.

In cases where students have stayed for a longer amount of time, this may be due to forgetting to leave the call or continuing discussions once moderators have left. To determine this it is important to look at how long the moderators stayed for and whether other students stayed for a similar amount of time. For example, in the Week 3 data ownership and management tutorial (Morning) the moderators stayed for 1hr 10mins whilst multiple students stayed for 1hr 30mins and left around the same time, suggesting their discussion continued after the moderators left. Contrastingly, in Week 4 data quality and standards tutorial 10 am session (morning), one student stayed for over 2 hours with everyone already leaving 30 minutes before. This suggests the student forgot to leave the call. If some of these anomalies have been identified, it is best to talk to the course leads to understand them.

Questions for Brittany (and Pawel)?

* **Should I filter by those who have consented at this point?**
* Are the rows in the table below anomalies? E.g. Why is Helen Wu appearing as two separate entries for the same session + why does she have a student number in one row but not another, it’s these rows without student numbers where she’s a guest where it recognises her as different participants.
* Tutorials are 1 hour?
* Elaine + Sophie go to one welcome session + are registered as guests, assign role as moderators?
* Beginning of session has important
* Lots of people in different timezones
* Joined just to ask question at the end
* Count everyone as attending because they still intentionally joined.

Anomalies:

|  |  |
| --- | --- |
| **Session** | **Anomalies found (defined as the total time being below 10 minutes or above 90 mins in a session that was supposed to be 1 hour)** |
| Course Welcome (morning) | Nisha Daniel stayed for 1hr 36 mins. This is not an anomaly as he left when Brittany left. |
| Course Welcome (evening) | **Paolo** stayed for 2 mins. Seemed to have joined midway during the session so doesn’t seem to them accidentally joining before or after the session. Internet difficulties |
| Week 2 data science tutorial 10 am session | A lot of people stayed for over 90mins – must have just been a longer tutorial that week as Brittany also stayed longer? **Khuliena** stayed for an extremely long time of over 2hrs + much later after Brittany had already left, did they forget to leave? No one else stayed this long.  **Helen Wu** has 2 entries and stayed for 17 seconds in one entry + 43 seconds in another. Do not seem to be due to accidentally joining too early or later as they joined after 10am + before 11am. She was assigned as a ‘Guest’ rather than ‘Integration’ and has no UUN. |
| Week 3 data ownership and management tutorial (Morning) | This tutorial also was much longer with lots of students staying around 90 mins despite Brittany staying for 70mins.  Bright & Rachael stayed for over 2hrs + Brittany had left much earlier – does not seem to be due to forgetting to leave as they left at the same time?  Tom Fowles, Zainab, Jaclyn, Alexander, David + Zhuang also left around 30 mins after Brittany had already left  Seems like students were having their own discussion after moderators left? |
| Week 3 data ownership and management tutorial (evening) | The 3 moderators in this tutorial all stayed around 1 hr between 6pm – 7pm.  Lots of students stayed around 90mins but left at a similar time (7:30pm). Joanne + Michael stayed for over 2hrs but left at the same time (8:12pm) so this does not seem to be due to them accidentally staying on the call.  Students were having discussions after moderators left? |
| Week 4 data quality and standards tutorial 10 am session (morning) | The 1 moderator (Brittany) stayed for 1hr 35 mins between 10am-11:38am  **Khuliena** stayed until 12:20 for over 2hrs – forgot to leave? Everyone else had already left by 11:38am.  Seems like most people stayed for a longer discussion until 11:38 but only Khuliena stayed for much longer, suggesting accidentally didn’t leave. |
| Week 4 data quality and standards tutorial 6 pm session (evening) | There were three moderators who stayed for around 70mins between 6pm-7:08pm  Daniel Williams stayed for 6 seconds and joined around 17:57 – technical issues?  Patrick Njeri stayed for 1hr 51 mins – forgot to leave? No one else stayed this long. |
| Drop in session (Week 4) evening | There were two students who stayed for 52 seconds and 2mins 36 seconds respectively (Lamis + Zhuang). Both joined within the allotted time of 6-7pm – technical issues? |
| Week 5 data handling and analysis tutorial 10am session | The 2 moderators (including Brittany) stayed from 10am to 11:22am latest.  Rachael + Bernice left soon after Brittany so not anomalies  Bobby stayed for 7 mins however only stayed between 9:55 – 10:02am – technical issues?  David stayed for over 2 hours until 12:15pm + Nasser stayed for over 2 hours until 12:10pm – did they forget to leave or were they having a discussion together? |
| Week 5 data handling and analysis tutorial 6pm session | The three moderators stayed for 1hr between 6pm-7pm  A lot of students stayed for over 90mins until 7:38pm.  Rachael + Patrick stayed for over 2 hours until 8pm but both left around the same time – seemed to be having a discussion together?  There were two students who stayed for 1 min 25 secs and 2min 35 secs respectively (Anisa + Adeyemi) – Anisa joined within allotted time for session (6:10pm) whilst Adeyemi joined just as session finished at 7pm? Did Anisa have technical issues + Adeyemi timezone issues? |
| Week 6 data visualisation tutorial (morning) | One student (Lamis) stayed for 6 seconds within allotted time – technical issues? Lamis seems to then have joined using their Guest account ‘lamis’ + stayed for full session.  Nasser is registered as 3 different entries using their Guest + Integration accounts – technical issues? Rejoined 6 times using their Integration account. |
| Week 7 good coding practice tutorial 10 am session | Lots of people stayed for just over 90 mins but Holly Tibble + Brittany were there as moderators until 11:30am.  There is another entry for Holly Tibble joining for 3 seconds – joined call early accidentally at 9:37am? |
| Week 8 data science in HSC tutorial (morning) | Lots of people stayed for over 90 mins but John Wilson was there as a moderator. Don’t seem to be anomalies as those people all left around same time.  Bernice stayed for only 1 min 30 secs – accidentally joined at 9:47 am? |
| Week 8 data science in HSC tutorial (evening) | The moderators stayed for around an hour between 6 – 7:15pm latest.  Two people stayed for over 2 hours (Michael + Joanne) but left around the same time of 8:22pm – had their own discussion so doesn’t seem to be accidental |
| Drop in session Week 9 | Shiva stayed for 12 seconds – accidentally joined at 11:53am? |
| Week 9 DDI Tutorial 10 am session | Zainab stayed for 36 seconds, but joined during the allotted time at 10:48am – timezone issues? |
| Week 9 DDI Tutorial 6pm session | Moderators stayed for one hour between 6-7pm.  Annisah + Michael stayed for almost 2 hours but left at the same time of 7:50pm – does not seem to be an anomaly but rather they were having a discussion together |

Interesting questions for exploration:

* Do they do 3 stars and a wish before or after a session
* What KIND of session do they do the 3 stars before or after (e.g. tutorial, live session, etc)
* Do the same people always attend the session at the same time (i.e. morning or evening)
* Are they more likely to do the reflections before or after the first, second or third session of the week?
* Were participants more likely to attend a session in a certain week?

Data within the dataset:

* Number of participants in each session
* Name + username of participant
* Participant’s role + attendee type – can look at number of students vs moderators in given session?
* First join + last leave – can remove the date and make this a new column so you just get the time?
* Total time the participant was in the session
* Number of times the participant joined the session
* Course code
* Type of session – clean up the names for this