Final datasets description

Final dataset: final\_course\_evals\_with UNI.xlsx

Columns

|  |  |
| --- | --- |
| Column name | Description |
| professor\_uni | Map of professor uni that was already in the course evals data, or a rough mapping which may be inaccurate. |
| professor\_fullname | Fullname field in the course evals data. Is not clean- sometimes has names reversed, or alternate names, or multiple prof names |
| lowercase | Converted to lowercase for better matching (standardizing) |
| enrollment\_total | From course evals data- original data itself- no manipulations. |
| eval\_respondents | No manipulations-- this was used to evaluate which bie\_course or bie\_professor should be kept based on the matches in tcdb |
| bie\_course | If multiple different values existed for same unique course id and same prof, this was manually chosen as to what best matched the tcdb. |
| bie\_professor | If multiple different values existed for same unique course id and same prof, this was manually chosen as to what best matched the tcdb. |
| average\_workload\_rating | No manipulations |
| response\_rate | No manipulations |
| clean\_course\_number | Cleaned the course\_number field in order to standardize it across datasets: Make sure it's a string, strip whitespace, and convert to uppercase; Extract the first letter and a 4-digit number at the end; removing special chars etc. |
| new\_unique\_course\_id | Re-did unique course\_id as the original unique\_course\_id field was sometimes messy—freshly combined (as strings) term\_number+ clean\_course\_number + section\_number |
| name\_1 | Professor fullname column was split into four possible names based on spaces between names. Names were cleaned to standardize: lowercase, no special characters, no whitespace |
| name\_2 | “ |
| name\_3 | “ |
| name\_4 | “ |
| fixed\_name |
| correct\_map | This is the correct and final UNI map for each professor based on all models and manual mapping. |

# UNI MAPPING TO COURSE EVALS

Cleaned and lowercased all names—and standardized names in all sheets to lower, without special chars etc.

Ran a bunch of models.

The ensemble gave the fewest missing.

Outputs: name\_ref\_df which has names and unis of pros from SIS and tycdb combined.

Course\_uni: course and uni matches from SIS and tcdb.

Wrote ensemble missing values to csv. Manually mapped the missing professors (ran pivot and got unique profs and mapped those without unis). After this, noticed that many courses that didn’t have a map was because it has multiple prof last names in professor\_fullname column.

Ran the multiple\_professor code to explode the columns such that each prof had own row for each course.

Then matched prof name (which was just a single name)—so to avoid error matched name+course\_unique\_id and mapped uni --- however, I think one iteration matched just the name and all other profs got that.

Wrote the missing sheet and manually mapped missing values here--- manually looked up each last name and tried to see if the same/similar course had that professor’s full name and uni—then mapped these two to the part-name.

CREATED NEW EXCELL ‘final merging ensemble manual and multiple manual’ this will eventually have the master-list of prof-uni and fixed\_name.

Combined: Combined the ‘ensemble course evals with unis manual mapping’ with ‘multiple\_manual mapping’.--- AAAH HERE WE ARE MISSING THE ROWS THAT GOT MAPPED IN MULTIPLE MAPPINGS---ADD THEM TOO- SOLVED-- ✅

Then put in fixed name for all in the combined sheet---- filtered out professor\_fullname that had a / and & as they have been exploded already.

Pivoted prof\_fixed name and uni—then took count of how manmy times that uni occurred--- sorted bin alphabetical order--- manually checked so that each Uni only had one professor name—however, flipped names having same uni were okay—just needed to make sure diff profs hadn’t accidently got same one assigned (because of course matching above).

After this, I added the mapped parts from multiple\_profs by filtering out the blanks (as the blanks would have been in the ‘missing’ csv). (these are pasted in the ‘combined’ sheet in blue). Manually added correct names by matching the name present. Eh Konik . =

|  |  |
| --- | --- |
| Kerri Konik | kk2229 |

From the SIS sheet. Used sis and tcdb.

Added to master list of names and unis.

Sorted the name, uni and count of uni columns into alphabetical by name now—to see if missing profs have an extra row that is filled.

XLOOKUP from this cleaned list and mapped fixed\_prof\_name for the new multi-prof mapped part of the COMBINED data (the one in blue).

Again, I filtered for all empty unis (the ones I removed because of it being mapped by course so wrong prof got uni if they co-taught a course) then mapped manually for those missing uni profs from SIS and tcdb.

NOW I will map the fixed name with the master list. There are still some unis missing from the name\_uni masterlist—but im unable to find these profs at all.

I take these blank profs, and manually search SIS and TCDB again.—names are reversed etc.--- adding manually.

FINALLY, copy and paste alllll other fields that were initially in course\_evals+relevant new cols to new sheet-----AND I will DELETE the ‘multiple’s from the fixed names column, and now we have one course per prof per term mapped with uni!!! UNI maps doneeee.

Actually, still a lot of names were unmapped—I mapped them manually using SIS—in the ‘final\_course\_evals\_with\_uni’ file itself. Added it to the master UNI sub-tab as well.

Something’s upp—there are more multiple than that were exploded. Im going to take the lsit from ensemble\_manual map sheet—which has all the rows we need—match the fullname with the lowercase fullname in the ‘final\_list’ (in the ‘final\_course\_evals\_with\_Uni’ which is now renamed “NOT FINAL’ but the ‘final’ tab is what was used to create new final\_course\_evals with uni) and assign it the fixed\_name. then I will remap. And RE-explode the multiples!

The new final\_course\_evals\_with UNI also has the masterlist of name-UNI -🡪 where Kiersten gave missing prof’s UNIs.

# Prof and course evals in Course evals