

The **Final Project Instructions** are posted on the Piazza resources page:

- <https://piazza.com/illinois/summer2019/cs513/resources>

The direct link is:

- https://piazza.com/class_profile/get_resource/jv6qe5xlp4c4vf/jxgpafv3bnq4yw

Below are additional details, primarily concerning the **upload logistics**.

As a reminder, you are encouraged to form groups of 2 or 3 students, but it is OK to do an individual project as well. (Groups of 4 require permission by the instructor: post a private note on Piazza, explaining your rationale; then send email to ludaesch@illinois.edu with Subject: "CS-513 Final Project" alerting the instructor to your request for an exception).

What to submit:

- All the deliverables are to be uploaded **in a single zipped file** (the upload URL will be posted on Piazza).
- Naming your zip file: If you have a group of two students, name the file as follows:
 - **LASTNAME1-Firstname1_LASTNAME2-Firstname2.zip**
- Note that the ALL CAPS lastname, the Capitalized firstname, the dash "-" in between first- and lastname, and the underscore "_" between students.
- The main deliverable is your **final project report**, whose content should be as described in the final project instructions. Also check out the office hours, e.g., from July 13, 2019 on a Q&A session about the final project.
- The final project report should be named
 - **LASTNAME1-Firstname1_LASTNAME2-Firstname2_Final-Report.pdf**
- Note that it should be part of your zipped upload file (see above).
- Other than the Final Report, what else goes into the zip file? For each part of your project, provide adequate explanations (in the report) and additional evidence through uploading artifacts.
- In particular, if you're using the following tools ...
 - **OpenRefine**
 - include the **JSON recipe**
 - optionally (for groups of two or more): provide the **or2yw outputs** if you have used or2yw.
 - **SQL**
 - include the queries & views you're using for profiling the data and checking ICs

- include (summaries of) the results of those queries and views in the report itself.
- We encourage the use of SQLite, but you can also use other RDBMS systems
- **Python, R, etc.**
 - include your scripts as part of zip archive
- **YesWorkflow**
 - if you use YW, include the YW annotations file as part of your zip file.
- **Input (“dirty”) and output (“clean”) datasets:**
 - Do **not** include the input and output datasets directly in your zipfile, unless they require no more than 10MB space combined.
 - In particular, if you use the reference datasets (NYPL or Farmers Markets) do not include the input files.
 - Instead, include **links (URLs)** to input and output datasets by placing them in a Box folder you own and putting those links (URLs) prominently at the beginning of your final project report.
- Last not least, your **zip file size** should not exceed 20MB total.
- Reminder: The deadline for uploading is August 1st.