Project Phase-II Report Outline and Checklist

We strongly encourage you to follow the Phase-II report outline below, as it aligns well with the checklist and grading rubric.

The title / header of your Phase-I report should list (i) the **Team-ID** of your group and (ii) for each group member the **name** and **Illinois email address**.

Report Outline / Checklist

The **project instructions** (separate document!) describe what exactly you should do as part of Phase-II of the project. The following outline should be followed for your Phase-II report:

- ☐ **1. Description of Data Cleaning Performed**
  - ☐ Identify and describe all (high-level) **data cleaning steps** you have performed. (**20 points**)
  - ☐ For each high-level data cleaning step you have performed, explain its **rationale**. Was the step really required to support use case U1? Explain. If not, explain why those steps were still useful. (**20 points**)
- ☐ **2. Document data quality changes**
  - ☐ **Quantify** the results of your efforts, e.g., by providing a summary table of changes: Which columns changed? How many cells (per column) have changed, etc.? (**10 points**)
  - ☐ Demonstrate that **data quality has been improved**, e.g., by devising IC-violation reports (answers to denial constraints) and showing the difference between number of IC violations reported before and after cleaning. (**10 points**)
- ☐ **3. Create a workflow model**
  - ☐ A visual representation of your overall (or "**outer**") workflow W1, e.g., using a tool such as YesWorkflow. At a minimum, you should identify key inputs, outputs, and steps of the workflow, along with dependencies between these. Key phases and steps of your data cleaning project may include, e.g., data profiling, data loading, data cleaning, IC violation checks, etc. Explain the design of W1 and why you've chosen the tools that you have in your overall workflow. (**10 points**)
  - ☐ A detailed (possibly visual) representation of your "**inner**" data cleaning workflow W2 (e.g., if you've used `OpenRefine`, you can use the `OR2YW` tool). (**10 points**)
- ☐ **4. Conclusions & Summary** (**10 points**)
  - ☐ Please provide a concise summary and conclusions of your project, including lessons learned.
  - ☐ Reflect on how work was completed. You should **summarize the contributions of each team member** here (for teams with >= 2 members).
- ☐ **5. Submission of supplementary materials in a single ZIP file** (**10 points**)
  - ☐ Workflow Model

- [ ] Operation history
    - [ ] OpenRefine Recipe
    - [ ] Other scripts, provenance files
- [ ] Queries
- [ ] Original (“dirty”) and Cleaned datasets
    - [ ] Please provide a **accessible** Box folder link in a plain text file: DataLinks.txt