

# Research Assistant's Task

## Data Collection on DOT Projects

**Overview.** We are assembling a nationwide, project-level database of state Department of Transportation (DOT) contracts to investigate issues such as infrastructure financing, project management, and bidding practices. Your work will expand this dataset by turning raw DOT records – bid lettings, project plans, contract summaries - into a clean, standardized panel ready for empirical analysis.

**Task.** You are asked to convert raw Ohio DOT records into a clean, standardized format. In the folder you will find two spreadsheets:

- “**Ohio\_2018\_Resurfacing\_PRR.xlsx**” is the master list of every resurfacing contract in Ohio in 2018.
- “**Ohio\_projects\_collected.xlsx**” is our template, already completed for one project so you can see the required layout.

Please choose **any five distinct projects** from the master list and, for each, locate the details needed to complete every blank field in the template. Specifically,

- project\_id: project id
- route: route
- mileage: length of the project
- lanes: number of lanes of the highway
- project\_duration\_days: duration of project in days
- eng\_estimate\_mils: engineer's estimated cost for project
- win\_bid\_mils: winning bid
- cost\_mils: actual cost of completing the project
- num\_bidders: number of bidders
- bidders\_list: names of all bidders (separated by commas or semicolons)

You are free to use any source that helps you identify these fields, including the bid-letting PDFs, attached plan sheets, and Ohio DOT's public bid-tab portal. The folder already contains the four PDFs that were used to create the example row; use them for guidance on where typical variables live. If a particular field truly is not available after a reasonable search, leave the cell blank.

When you are finished, rename the completed spreadsheet *Ohio\_projects\_collected\_clean.xlsx*. Write a short README (half a page to a page) that lists the five projects you selected (and why), describes the search steps and sources you relied on, and explains any judgement calls or unresolved data gaps.

**Deliverables.** Please submit a **single zipped folder** named RA\_task\_<your\_lastname>.zip. This zipped folder must contain: Ohio\_projects\_collected\_clean and your README. If you wrote any scripts to assist with cleaning or validation (optional), please include it in the folder.

Please return the zipped folder by **e-mail within 48 hours** of receiving this assignment. We will evaluate submissions based on accuracy, judgment, and transparency.