# **Intro**

Thank you for your interest in a data engineering role at Tesla. We run a lean operation requiring every team member to take on responsibilities and demonstrate skills that cannot be limited to a single job function. While this is a technical assessment, a purely technical solution will not be sufficient to land you the role. Make sure you document your problem-solving process, including any assumptions you make and any information or learnings you discover while building your solution.

# **Goal**

Tesla Residential Energy uses various datasets to inform the execution of tasks that enable efficient installation of our products. One of these datasets is roof imagery data, sourced from both 3rd-party imagery providers and internal drone surveyors. This data is parsed into a geometry model of a customer’s roof which is used by Tesla designers in creating construction plans used for permitting and installing our solar products.

Attached you will find a folder with 10 anonymized geometry model json files. Your task is to construct a data pipeline that reads these files, corrects for any data integrity issues, and transforms the data into a data model of your choice (e.g. relational, nosql, star, etc.) that best satisfies the needs of your consumers.

The consumers of this data are as follows:

* Analyst: looking for aggregate statistics, roof type information
* Data Scientist: looking for feature-engineered quantities of various roof properties
* Software Engineering Team: looking for precise mounting plane angles to inform a tiling api for the Solar Roof product

Reviewers will be looking for a submission containing an executable bash script that builds a docker container, runs a python script in that container, and outputs at least one csv file that represents the transformation you would deliver to the consumers listed above. This could be several relational tables, a flattened json object, or whatever other data structure you decide would best meet consumers’ needs.

Please also include a short (<250 words) justification of your solution.

# **Suggestions**

* Explaining why you chose your solution matters more than what solution your chose.
* Comment your code – you are also being evaluated on the clarity and replicability of your solution.
* Do not assume data is perfect. Double check and clean data if necessary.
* It’s possible that the provided dataset includes versioning information; think about how you would handle modifications/updates to an original model