Project Proposal (20 points) The proposal cites at least two sources of data. (10 points)

The proposal includes the type of final production database (relational or non-relational) that the data will be loaded into. (5 points)

The proposal gives a relevant and succinct description (2–3 sentences) of findings. (5 points)

8 Types of data transformations

1. Data Filtering

Filtered CSV column data

1. Data Mapping

Changed names of states and measures

1. Data Deduplication

After creating analysis table, duplicate redundant columns were removed

4. Derived Variables

Derived variables or metrics compute attributes from other data points and variables.

‍Example: *You create a column “profit” in your orders table that is computed by subtracting taxes, shipping\_costs, and* [*COGS*](https://www.investopedia.com/terms/c/cogs.asp#:~:text=Cost%20of%20goods%20sold%20(COGS)%20refers%20to%20the%20direct%20costs,used%20to%20create%20the%20good.&text=Cost%20of%20goods%20sold%20is,as%20%22cost%20of%20sales.%22) *from your sales\_total for each row (aka order).*

5. Data Sorting or Ordering

None really

6. Joining data from multiple sources

Merge related data that was collected from multiple data sources.

Example: *Join Facebook Ads spend data, Google Ads spend data and Linkedin Ads spend data into a single table called digital\_advertising\_spend.*

*‍*7. Aggregating data

Data aggregation is a form of summarization. Aggregation is done by picking a relevant dimension and grouping a metric across that dimension.

It is often performed for metrics that are commonly used to speed up analytics.

8. Splitting data

When working with unstructured and semistructured data (such as strings, JSON, or XML files), you split the data before saving it into a table.

1. Identify the target tables schema. Prepare a list of tables and their schemas (columns, possible values, etc.) of how the data would ideally look once it has been cleaned and loaded into the target database or data warehouse.
2. Record the gap between target and incoming data. Contrast the ideal target state with the raw data you collected in the extraction step. Note down what transformation (types) would be needed to convert the raw data into the cleaned data.‍
3. Implement the changes needed to bridge the gap between extracted data and target data. Validate the transformations before setting them on autopilot with one of the ETL tools.