## DOCUMENTATION

Our data is structured around survey responses, with each row representing a set of aggregated results related to specific survey questions and demographic segments.

Our datasets had the following columns.

**LEVEL1ID:** These are categories used to describe the organization.

**SURVEYR:** Indicates the survey year, which for the provided rows is 2022.

**BYCOND:** Demographic condition applied

**DEMCODE:** Represents demographic codes, which could be used to segment the data according to specific demographic questions

**QUESTION:** Contains the question identifier, such as 'Q01', linking back to the survey questionnaire.

**Answer 1 to answer 7:** These columns represent aggregated response percentages for each response option to a survey question.

**MOST\_POSITIVE\_OR\_LEAST\_NEGATIVE:** An aggregated metric that represents the most positive or least negative responses, offering a quick overview of positive sentiment or agreement. (percentage)

**NEUTRAL\_OR\_MIDDLE\_CATEGORY:** Indicates the percentage of neutral responses.

**MOST\_NEGATIVE\_OR\_LEAST\_POSITIVE:** Counterpart to the most positive metric, highlighting the most negative or least positive responses.

**AGREE:** This could represent a combined measure of agreement, possibly summing up several positive response categories.

**SCORE5:** A score on a scale of 1 to 5, possibly an average rating based on the distribution of responses.

**SCORE100:** A normalized score out of 100

**ANSCOUNT:** The total count of responses for the question.

**KEYWORDS:** The keywords derived from the question in each entry.

**TYPE OF QUESTION:** The type of each question per entry categorized as MCQ, Negative, Positive, Suppressed .

**SENTIMENT:** The derived sentiment for each entry

The combined dataset is derived from the following subsets:

**Subset 2:** Results by how long people have worked there, their age, and if they’re supervisors, plus past survey comparisons.

**Subset 3:** Results by gender, LGBTQIA status, and diversity categories.

**Subset 6:** Results by where people work, their first language, and if they work in bilingual roles.

Questions are categorized into:

**INDICATORIDs:**

1. Employee engagement
2. Leadership
3. Workforce
4. Workplace
5. Workplace well-being
6. Compensation

**SUBINDICATORIDs:**

* Employee engagement
* Immediate supervisor
* Senior management
* Performance management
* Job fit and development
* Empowerment
* Work-life balance and workload
* Mobility and retention
* Organizational goals
* Organizational performance
* Diversity and inclusion
* Anti-racism
* Ethical workplace
* Physical environment and equipment
* Use of official languages
* Harassment
* Discrimination
* Duty to accommodate
* Future of work
* A safe and healthy workplace
* A psychologically healthy workplace
* Work-related stress
* Pay or other compensation issues
* Support to resolve pay or other compensation issues