Welcome to the open raw data of the Developer Ecosystem Survey 2025!

**Note:** The data in the report is weighted. To replicate the results from the report, please apply the multiplier from the "weight" column in your calculations. Otherwise, the results will differ. Weights are applied to reduce bias and provide a more representative picture.

The folder includes the following files, each serving a specific purpose in the context of the survey data and its structure:

"2025\_sharing\_data\_outside.csv.zip": This is a CSV file compressed into a zip format, encompassing a wide dataset with 4740 columns and 24534 rows. Each row equates to an individual respondent, with each column representing either a response to a single-choice question or an option from a multiple-choice question. Special columns in this file include response\_id for the respondent's unique identifier and weight for data weighting to adjust for sampling biases by country, programming language preference, and relation to JetBrains.

"2025\_sharing\_data\_outside\_narrow.csv.zip": A zipped CSV file that provides data in a more standardized, database-like format, with rows corresponding to individual responses to single-choice questions or individual options from multiple-choice questions. When respondents have selected multiple options, these are represented as multiple rows. The file includes a total of four columns (respondent ID, weight, question ID, and answer) and over six million rows.

"DevEcosystem 2025 questions\_outside.csv": This CSV file lists all survey questions with their respective short names as they are referred to in the dataset. This file also identifies parent\_question\_id for tabular questions such as radiobutton grids or checkbox grids, connecting each item to its overarching "parent" question.

"Developer\_Ecosystem\_2025\_Survey\_Structure.pdf": A PDF document that contains the full text of the questions and answer options from the survey system, as well as a description of the logic behind the questions. For example, it explains that questions about the Java programming language were only asked to respondents who use Java.

"LICENSE.txt": A text file that includes the licensing terms (CC-BY-4.0 license) for using the data contained within the folder.

Additional details for a set of "Fiction" questions included in the dataset:

primary and rank: These columns are adjusted versions of the primary\_proglang and proglang\_rank questions. They have been modified to account for instances where respondents selected only one programming language in the proglang question, thus not being asked the subsequent ranking questions.

main: This question is a derivative of proglang\_rank, constructed to represent a scenario where the respondent is asked to choose their single most important programming language.

region: A column designed for reporting purposes that consolidates individual countries into broader regional categories.

salary\_group: This represents a classification within each region that segments individuals into three groups based on their salary range—high, medium, and low. The actual salary question has been excluded from the open dataset due to its sensitive nature.

**What Does It Mean When Data Is Published Under the CC-BY-4.0 License?**

With the Creative Commons Attribution 4.0 International Public License (CC BY 4.0), you're free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) the survey data published by JetBrains s.r.o., for any purpose, even commercially. All you need to do is give proper credit to the original source, provide a link to the license, and indicate if you made any changes. Also, you should not add any legal terms or technological measures that could restrict others from doing anything the license allows.

For scientific papers and blog posts that are based on, but not directly incorporating, these data, the same principles apply. You need to acknowledge JetBrains s.r.o. for the data your research or analysis is based on, clearly state that the data is used under CC BY 4.0, and describe how you've engaged with the data. Whether it's through analysis, commentary, or any other form of adaptation, this approach maintains transparency and encourages the sharing and expansion of knowledge, regardless of the medium through which your work is presented.