

Codebook

With our cleaned data, let's summarize all the information in our data.

work_year: Describes what year the salary came from. Ordinal variable with 3 levels. "Pandemic", "2023", "2024"

experience_level: Describes level of experience for the role. Ordinal variable with 4 levels. "EN"=Entry Level "MI"=Mid-Level "SE"=Senior "EX"=Executive

job_title: Describes the job title and what kind of work is done. Nominal variable with 6 levels. "DA"=Data Analyst "BI"=Business Intelligence "DE"=Data Engineer "DS"=Data Science "ML"=Machine Learning "Other"

salary_in_usd: The salary in USD which is a quantitative variable

remote_ratio: Describes the format of the job. Nominal variable with 3 levels. "In-Person", "Hybrid", "Remote"

company_location: Describes what kind of country this job came from. Nominal variable with 3 levels. "US" "First World" "Developing"

company_size: Describes the size of the company. Ordinal variable with 3 levels. "S"=Small "M"=Medium "L" = Large