

Project 4 Data Collection

Ben Goldstone

4/21/2023

Data Set

```
library(readr)
datascience_salaries <- read_csv("~/CSVs/ds_salaries.csv")

## Rows: 3755 Columns: 11

## -- Column specification -----
## Delimiter: ","
## chr (7): experience_level, employment_type, job_title, salary_currency, empl...
## dbl (4): work_year, salary, salary_in_usd, remote_ratio

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

head(datascience_salaries)

## # A tibble: 6 x 11
##   work_year experience_level employment_type job_title      salary salary_currency
##   <dbl> <chr>           <chr>           <chr>      <dbl> <chr>
## 1    2023 SE             FT             Principal D~ 80000 EUR
## 2    2023 MI             CT             ML Engineer  30000 USD
## 3    2023 MI             CT             ML Engineer  25500 USD
## 4    2023 SE             FT             Data Scient~ 175000 USD
## 5    2023 SE             FT             Data Scient~ 120000 USD
## 6    2023 SE             FT             Applied Sci~ 222200 USD
## # ... with 5 more variables: salary_in_usd <dbl>, employee_residence <chr>,
## #   remote_ratio <dbl>, company_location <chr>, company_size <chr>
```

Columns

Data Science Job Salaries Dataset contains 11 columns, each are:

work_year: The year the salary was paid.

experience_level: The experience level in the job during the year. (SE:Senior, EN:Entry level, EX:Executive level, MI:Mid/Intermediate level)

employment_type: The type of employment for the role.

job_title: The role worked in during the year.

salary: The total gross salary amount paid.

salary_currency: The currency of the salary.

salaryinusd: The salary in USD.

employee_residence: Employee's primary country of residence in during the work year as an ISO 3166 country code.

remote_ratio: The overall amount of work done remotely.

company_location: The country of the employer's main office or contracting branch.

company_size: The median number of people that worked for the company during the year.

Data Source

This data was sourced from AI-jobs.net. AI-Jobs.net seems to connect directly to employers for posting information on jobs. This dataset is from Kaggle datasets.

The data does not seem to have a direct source for it's data besides specifying that they "Provide a direct link between candidates and employers."