The dataset used for this project contains 318,477 job advertisements. Each row in the dataset represents a unique job posting and includes various features such as job title, company name, location, industry classification, salary range, and job requirements.

The data is provided in CSV format and includes over 13 columns, with a combination of categorical (object) and numerical (int64) data types. Below are some key columns and their respective types:

* Title (object): The job title listed in the advertisement
* Company (object): Name of the company posting the job
* Location / Area (object): Geographic information
* Classification / SubClassification (object): Industry and job category
* Requirement (object): Brief job requirement summary
* FullDescription (object): Detailed job description text
* LowestSalary / HighestSalary (int64): Estimated salary range
* JobType (object): Employment type such as Full time, Contract, etc.
* Date, Source, index (object or int64): Metadata fields

Most fields are of type "object" (textual or categorical), while "LowestSalary" and "HighestSalary" are numeric fields with type "int64". Some columns have significant missing values — for example, Company has approximately 12,000 missing entries, and Location, Classification, and SubClassification each have over 100,000 missing entries.

For the core analysis, the following columns were retained:  
Title, Company, Requirement, LowestSalary, HighestSalary

Other columns such as Location, Area, Classification, SubClassification, FullDescription, and JobType were used only for descriptive statistics or summary purposes, and not included in predictive modeling due to the large number of missing values or limited relevance.