Notes on Data and Methodology

Tommaso Pellegrinetti

2023-10-20

Jobs and Cities Included in the Analysis

The list of job positions comprising the sample for this analysis was generated with the assistance of ChatGPT and can be located in the file `files/jobs.txt`. More specifically, job titles were produced by ChatGPT in response to the following prompt:

For a personal data visualization projects, I want to explore differences in salaries across some European Cities. To do so, I need to define a set of job positions whose salaries will be collected from websites like Glassdoor, Indeed, or Monster. Keep in mind that I would like to research commonly used job titles across different industries and ensure consistency across locations and local language variations. In addition, I am primarily interested in the tertiary sector. Could you please provide me with a list of 100 job titles that in your opinion would be suitable for this kind of analysis?

In the process of data cleaning, four job titles out of 100 were excluded because of insufficient observations.

The resulting list of job titles covered in this analysis is as follows:

```
 [1] "account manager"                "accountant"
 [3] "administrative assistant"       "administrative coordinator"
 [5] "advertising manager"            "analytics manager"
 [7] "art director"                   "brand manager"
 [9] "business analyst"               "business development manager"
[11] "business intelligence analyst"  "business process analyst"
[13] "business systems analyst"       "cloud architect"
[15] "communications manager"         "content manager"
[17] "content strategist"             "contract administrator"
[19] "corporate development manager"  "customer service manager"
[21] "customer success manager"       "data analyst"
[23] "data engineer"                  "data scientist"
[25] "database administrator"         "designer"
```

```
[27] "developer"                        "digital marketing manager"
[29] "digital strategist"               "e-commerce manager"
[31] "editor"                           "event coordinator"
[33] "executive assistant"              "financial advisor"
[35] "financial analyst"                "financial controller"
[37] "front end developer"              "full stack developer"
[39] "general manager"                  "graphic designer"
[41] "human resources assistant"        "human resources generalist"
[43] "human resources manager"          "industrial designer"
[45] "information technology manager" "inside sales representative"
[47] "interior designer"                "investment analyst"
[49] "investment banker"                "it manager"
[51] "it project manager"               "logistics manager"
[53] "machine learning engineer"        "management consultant"
[55] "marketing assistant"              "marketing manager"
[57] "marketing specialist"             "media planner"
[59] "mobile developer"                 "network administrator"
[61] "operations analyst"               "operations manager"
[63] "pricing analyst"                  "procurement manager"
[65] "procurement specialist"           "product manager"
[67] "product marketing manager"        "program manager"
[69] "project coordinator"              "project manager"
[71] "public relations manager"         "quality assurance analyst"
[73] "quality control manager"          "real estate agent"
[75] "recruiter"                        "research analyst"
[77] "risk manager"                     "sales associate"
[79] "sales manager"                    "sales operations manager"
[81] "sales trainer"                    "scrum master"
[83] "seo specialist"                   "social media manager"
[85] "social media strategist"          "software engineer"
[87] "supply chain manager"             "system administrator"
[89] "tax specialist"                   "technical account manager"
[91] "ui designer"                      "ux designer"
[93] "vendor manager"                   "video editor"
[95] "web designer"                     "web developer"
```

It's worth noting that certain job titles only appear in specific countries, as documented in Figure Figure 1. For example, around 25 jobs appear in 11-15 cities, while only 12 jobs appear in 26-28 cities.

For personal interests, the analysis has been confined to cities in Western Europe. Initially, my objective was to gather data from the four/five largest cities in Italy, France, Germany, the UK, Austria, Spain, Switzerland, the Netherlands, and Poland. However, due to limited

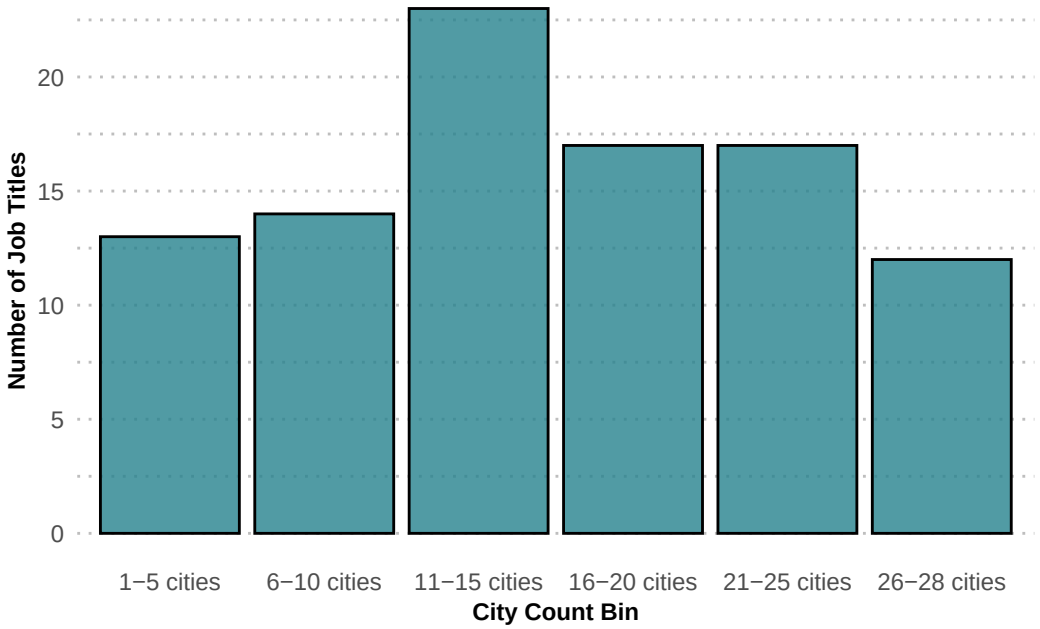

Figure 1: Histogram of the number of cities where a job role appears.

data availability, I was only able to collect data for some of those cities. In the process of data cleaning, I made the decision to exclude cities with fewer than 20 available job titles. It's worth noting that there is considerable variability in the number of job roles analyzed in each city, as illustrated in Figure Figure 2.

Limitations

It's essential to understand that the results of this analysis should not be regarded as representative of actual salaries in the selected cities. In fact, there are several significant limitations to this analysis that should be taken into consideration when interpreting the results. Firstly, it's important to note that the number of observations (i.e. the number of reported salaries) for each city-job data point is generally quite limited. Additionally, the sample of job roles selected is incomplete and non-representative of the whole population of salaries in those countries. Another potential source of bias stems from the fact that I exclusively searched for English job titles, even though some job positions are typically expressed in national languages. In conclusion, it is essential to underscore that the results of this analysis should be regarded as the outcome of a learning exercise rather than as definitive indicators of real-world salary patterns. Additional research and comprehensive data collection are necessary to refine our understanding of the intricate dynamics that shape compensation across Europe.

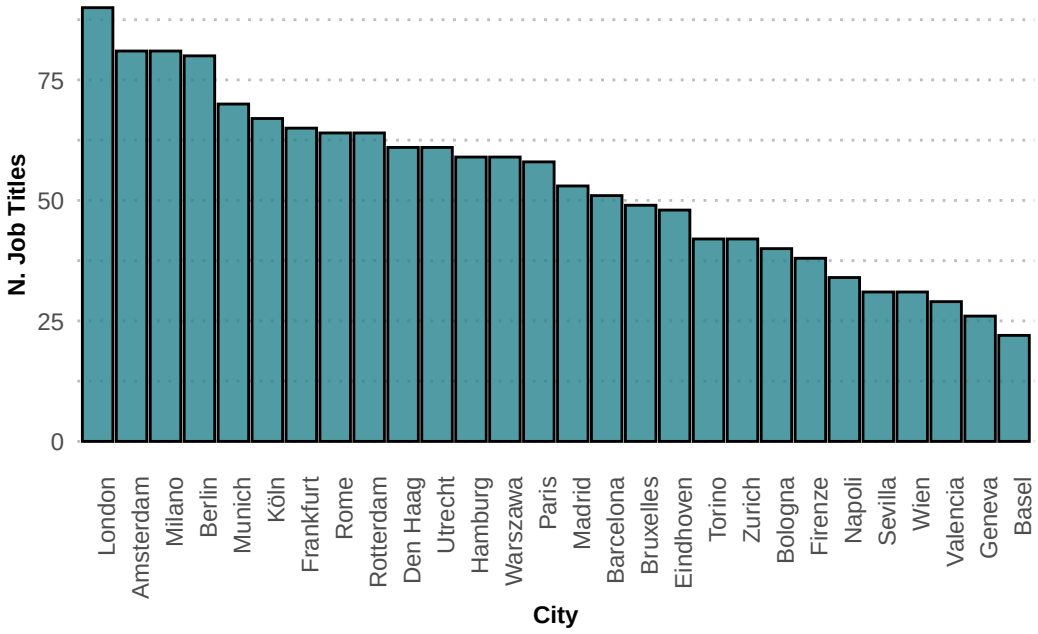

Figure 2: Distribution of Job Positions Collected Across Cities