

MSIS2607

Search

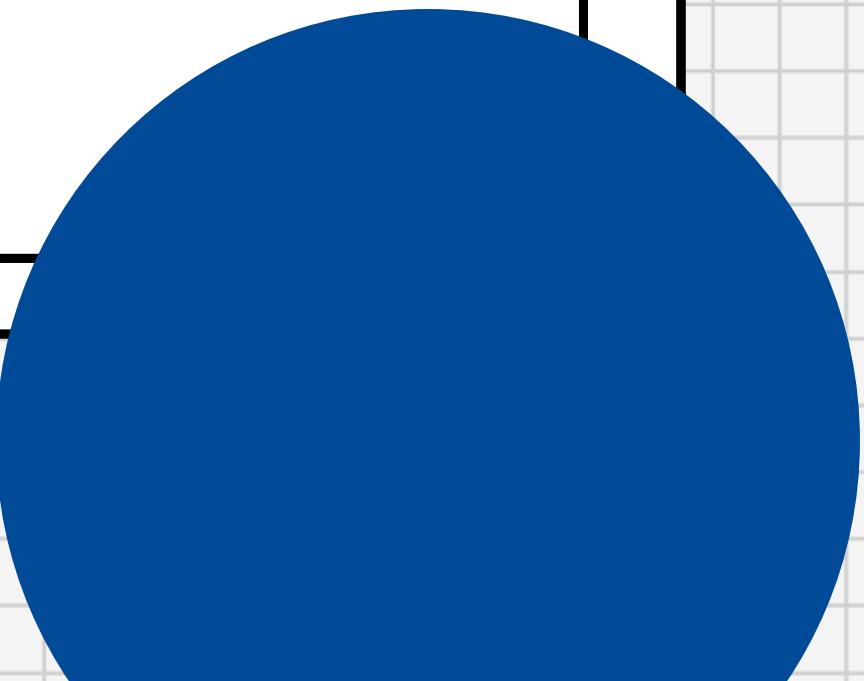


ooo

+

# Job Data Analysis

Team 5



# About the dataset



Columns of Dataset



**Company Profile**  
Job Portal Job Title  
Country Company  
Salary Range location  
**Company Size**  
Benefits Contact Job Id skills  
**Qualifications**  
Responsibilities Work Type  
longitude Contact Person Role  
**Experience**  
Preference latitude  
Job Posting Date  
**Job Description**

# Data Cleaning

▶ Checking for null values and handling them

▶ Checking for duplicates and handling them

▶ Splitting experience and salary range columns

▶ Cleaning on contact column

Contact	Cleaned_Contact
001-381-930-7517x737	+1-381-930-7517
461-509-4216	+1-461-509-4216
9687619505	+1-968-761-9505
+1-820-643-5431x47576	+1-182-064-3543
343.975.4702x9340	+1-343-975-4702



Checking for null values and handling them



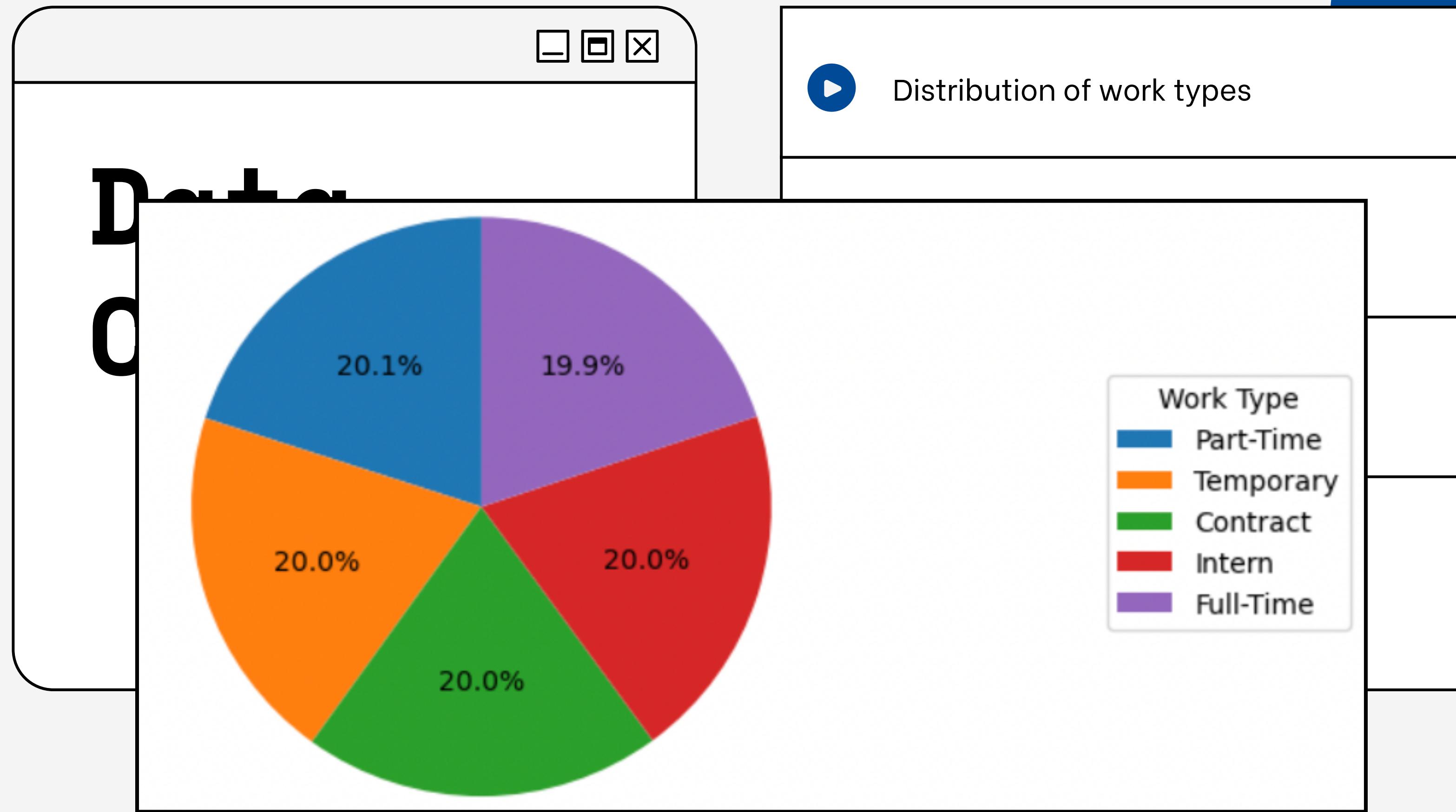
Checking for duplicates and handling them



Splitting experience and salary range columns



Cleaning on contact column

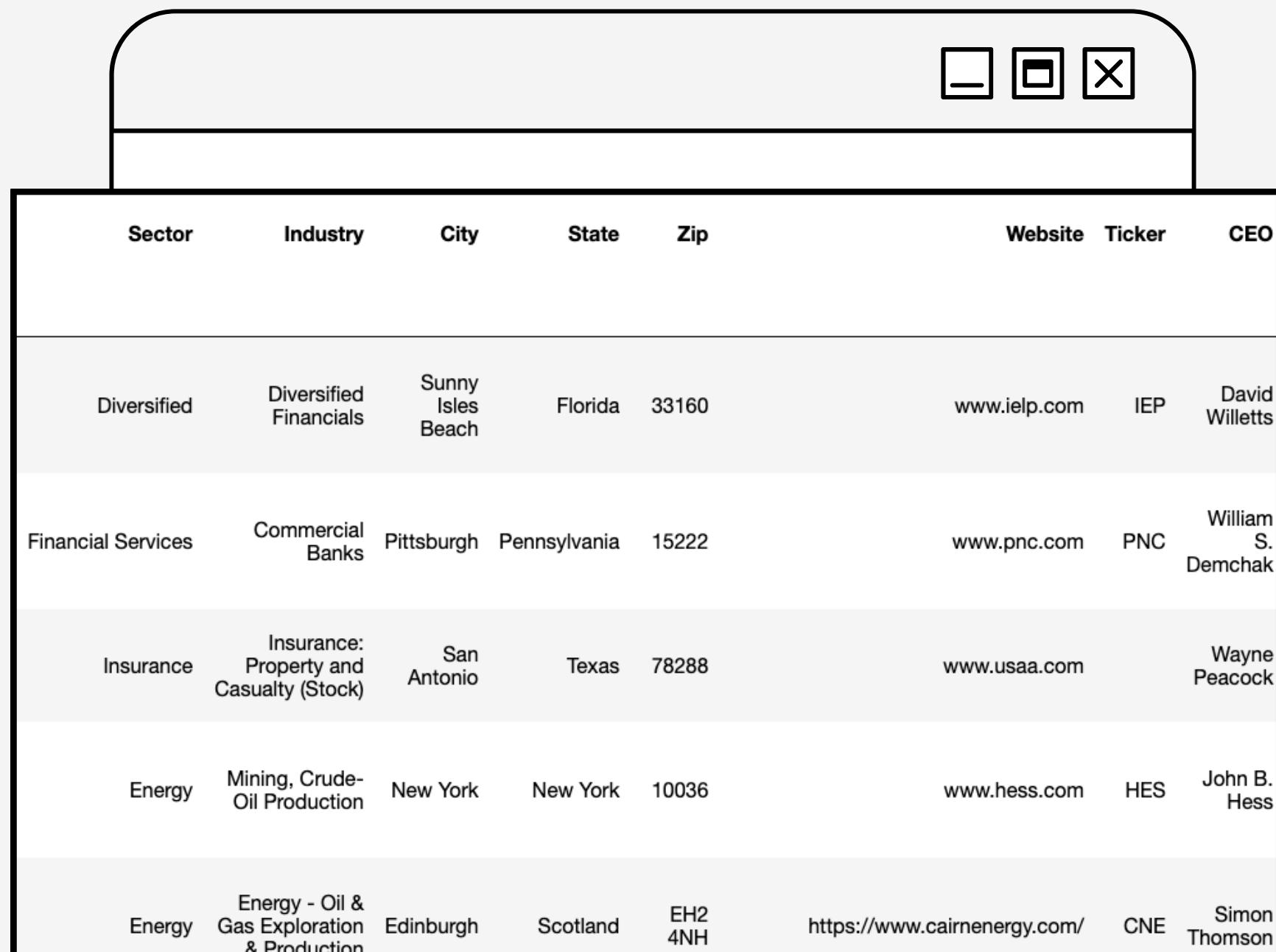


Count of each benefit:

Bonuses and Incentive Programs	404287
Flexible Work Arrangements	324021
Professional Development	323743
Retirement Plans	323674
Childcare Assistance	323605
Employee Assistance Programs EAP	323573
Employee Discounts	323396
Transportation Benefits	323295
Profit-Sharing	323295
Wellness Programs	323258
Relocation Assistance	323216
Tuition Reimbursement	323157
Parental Leave	323157
Health Insurance	323092
Employee Referral Programs	323079
Casual Dress Code	323079
Health and Wellness Facilities	323079
Legal Assistance	322869
Flexible Spending Accounts FSAs	322828
Financial Counseling	322828
Stock Options or Equity Grants	322676
Life and Disability Insurance	322538
Social and Recreational Activities	322538
Employee Recognition Programs	322287
Paid Time Off PTO	243130
Name: Benefits, dtype: int64	

Distribution of work types

Cleaning on benefits column



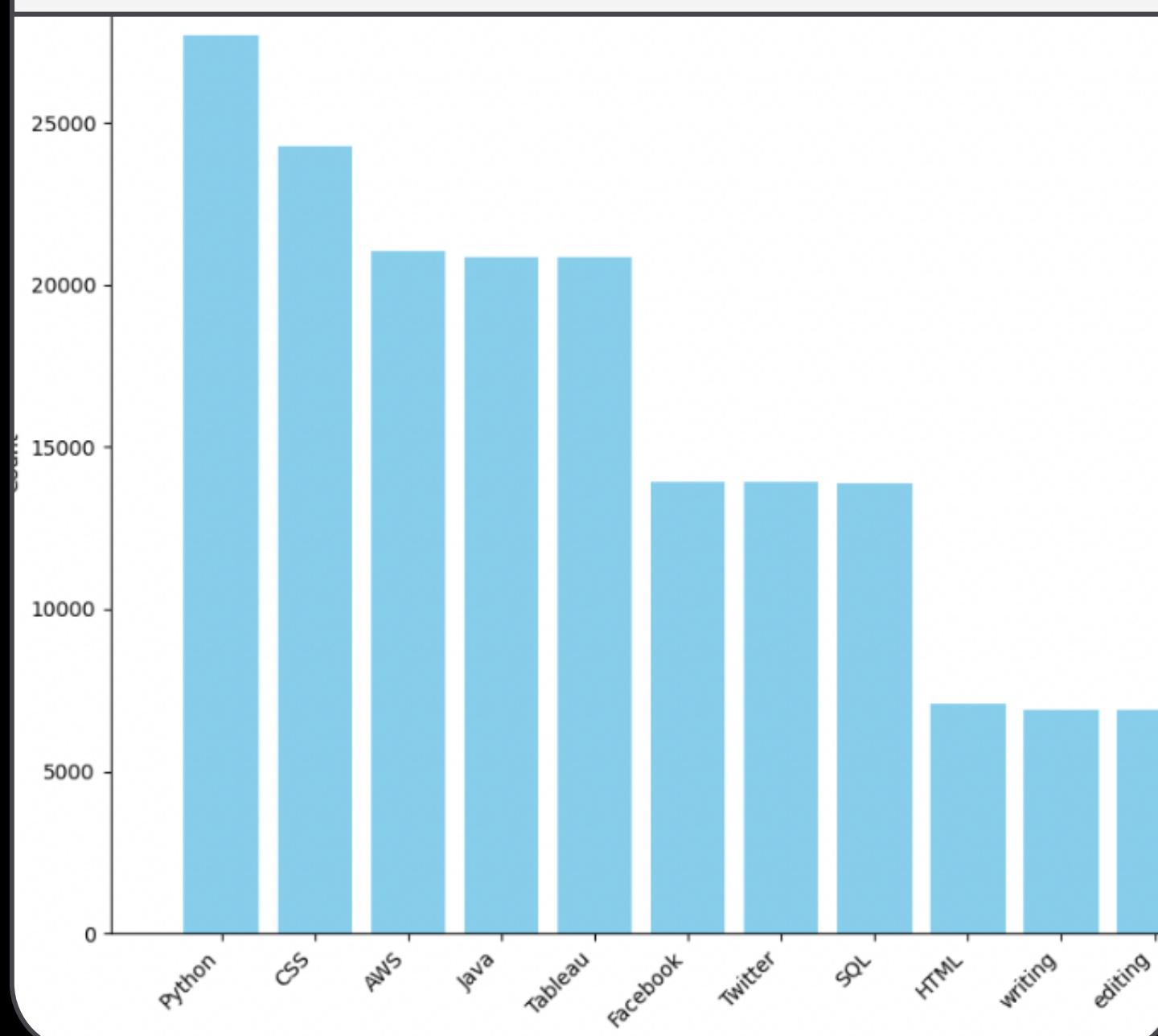
Sector	Industry	City	State	Zip	Website	Ticker	CEO
Diversified	Diversified Financials	Sunny Isles Beach	Florida	33160	www.ielp.com	IEP	David Willetts
Financial Services	Commercial Banks	Pittsburgh	Pennsylvania	15222	www.pnc.com	PNC	William S. Demchak
Insurance	Insurance: Property and Casualty (Stock)	San Antonio	Texas	78288	www.usaa.com		Wayne Peacock
Energy	Mining, Crude-Oil Production	New York	New York	10036	www.hess.com	HES	John B. Hess
Energy	Energy - Oil & Gas Exploration & Production	Edinburgh	Scotland	EH2 4NH	https://www.cairnenergy.com/	CNE	Simon Thomson

Distribution of work types

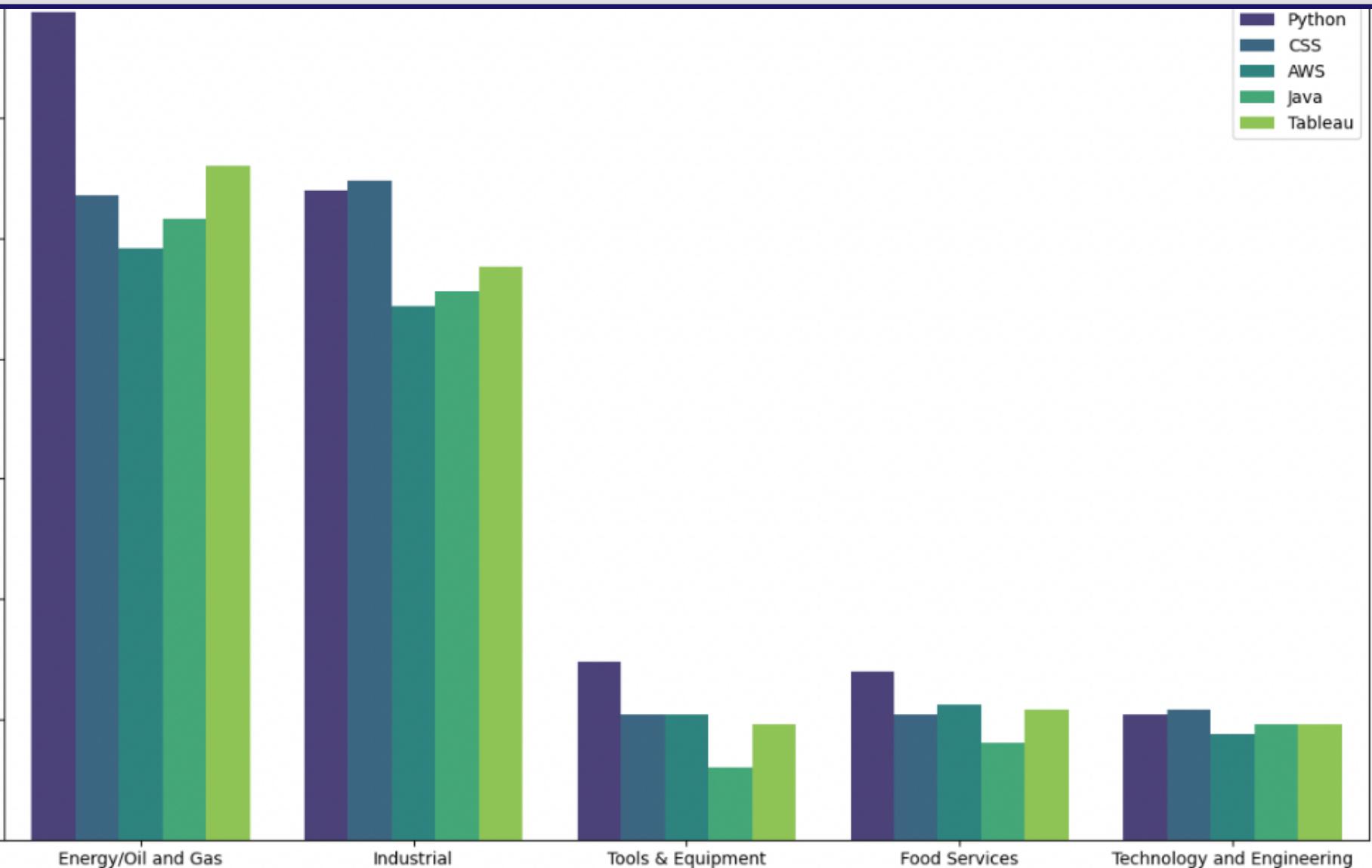
Cleaning on benefits column

Cleaning on company profile column

## Count of Buzzwords in Skills



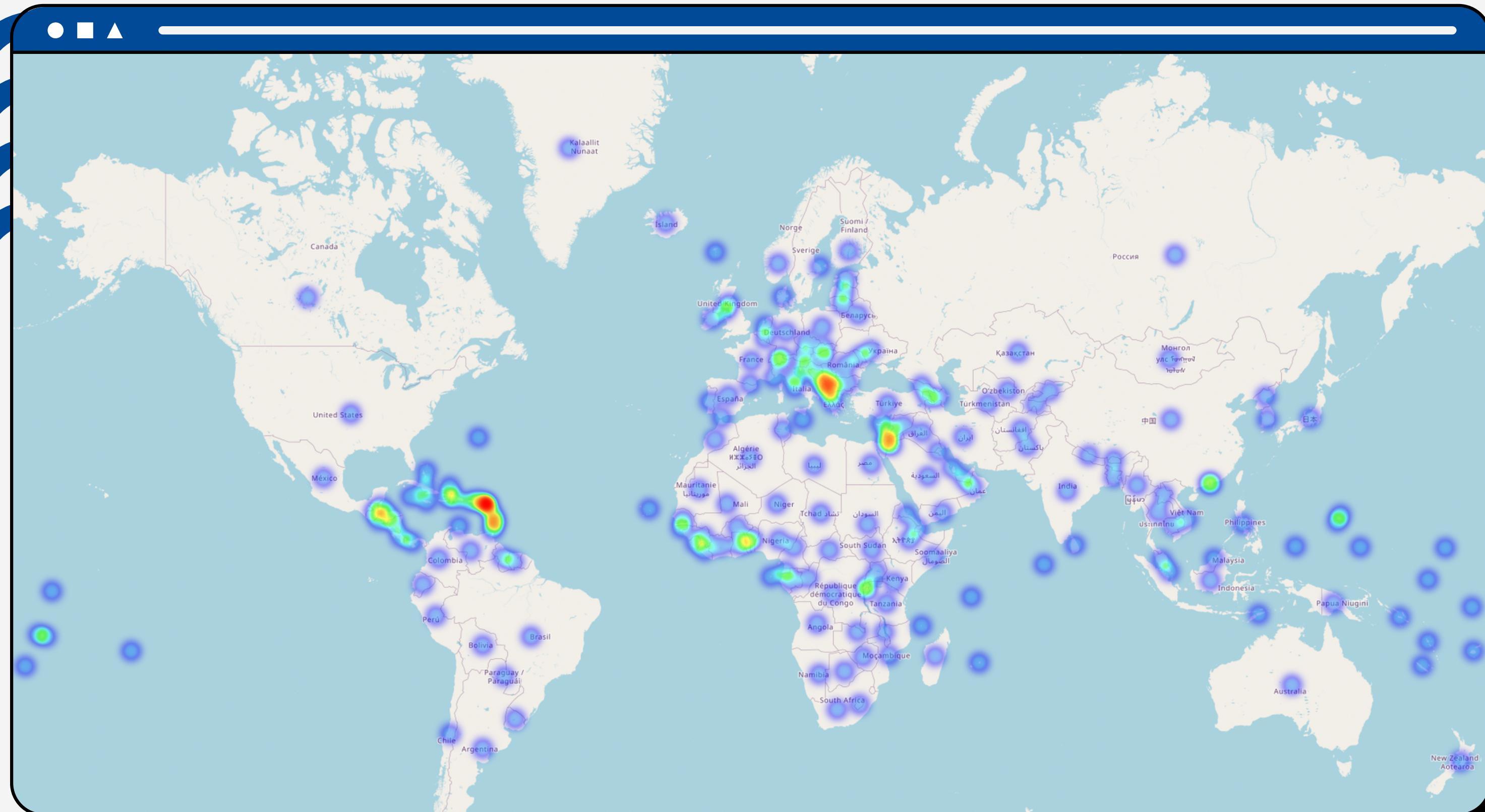
## Top 5 Buzzwords in Sectors



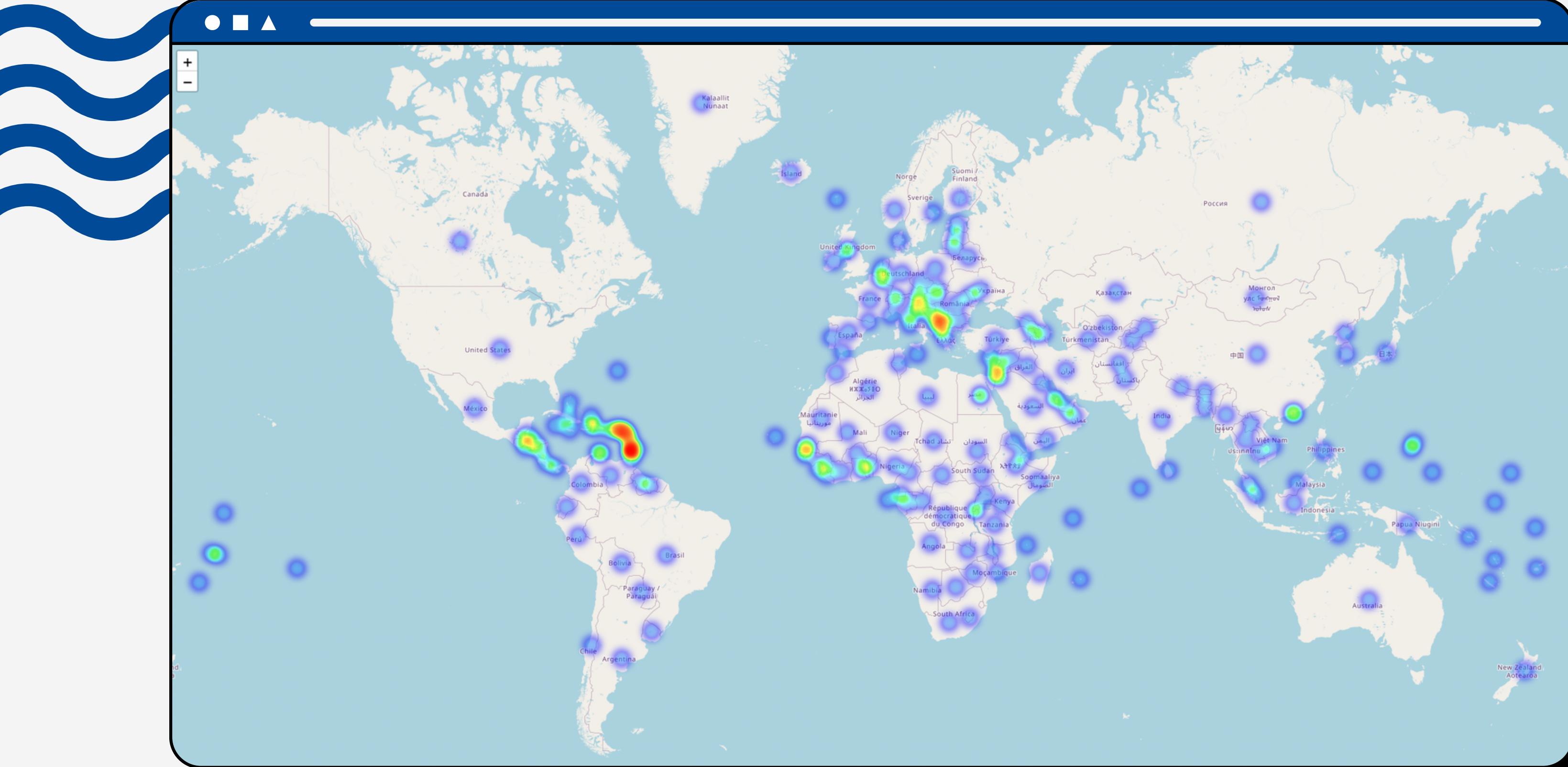
## Insight 1

From the above graphs, we can see maximum number of Buzzwords that we used in the skills columns. Using the top 5 buzzwords, we visualised the sectors where these skills were mostly used

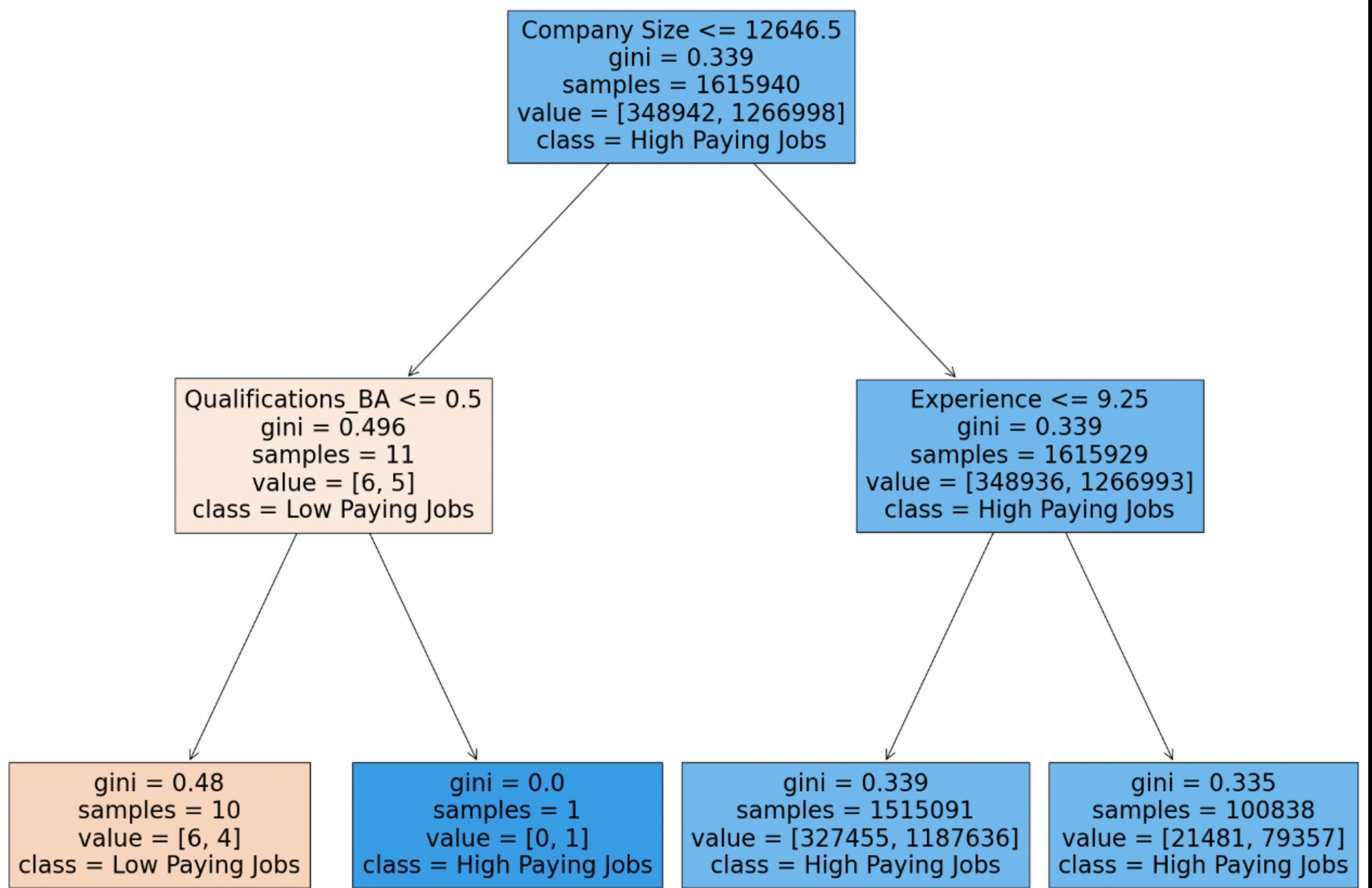
# **Insight 2: Distribution of Companies**



# Insight 2: Headquarters



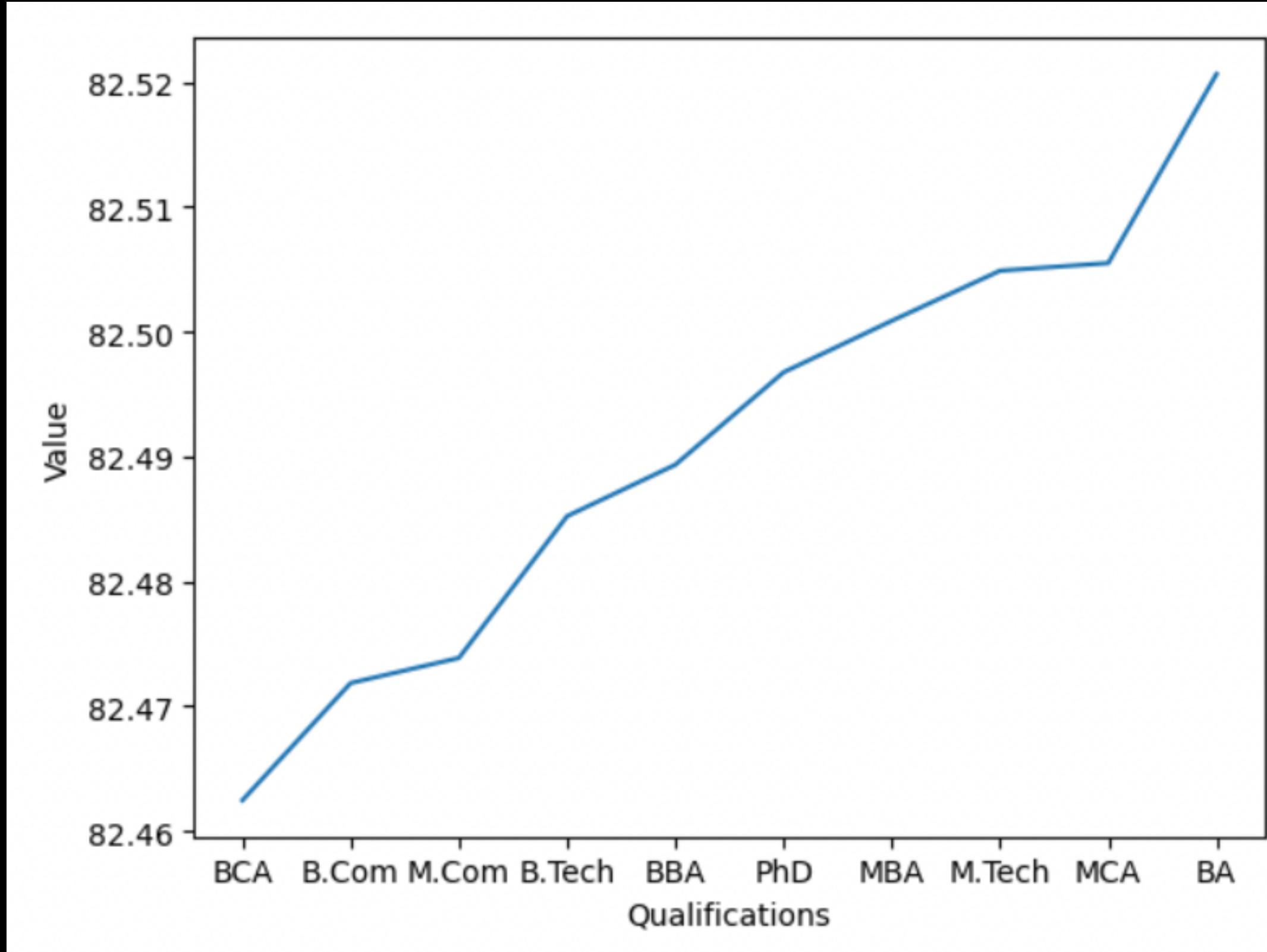
# Decision Tree



To know the factors affecting the average pay:

- 1) We performed one-hot encoding on categorical columns like Work Type, Qualifications and Preference.
- 2) Next, we applied a lambda function to each element in the 'Average Pay' column. The lambda function checks if the value x is greater than 75. If it is, it assigns the value 1.0; otherwise, it assigns the value 0.0.
- 3) We can see that for company size above 12646 and for qualifications as BA, we'll always get a high paying job

# Decision Tree



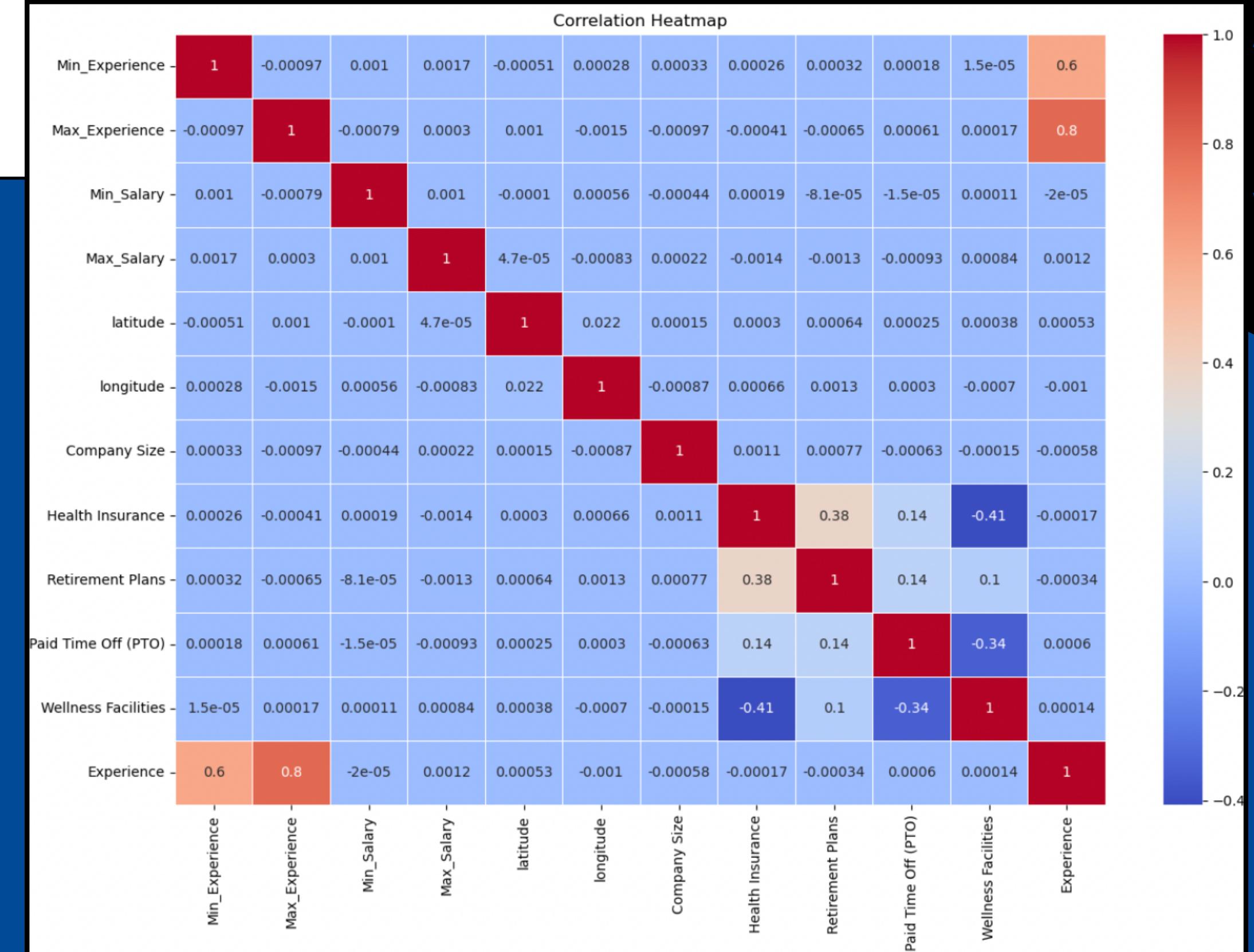
To know the factors affecting the average pay:

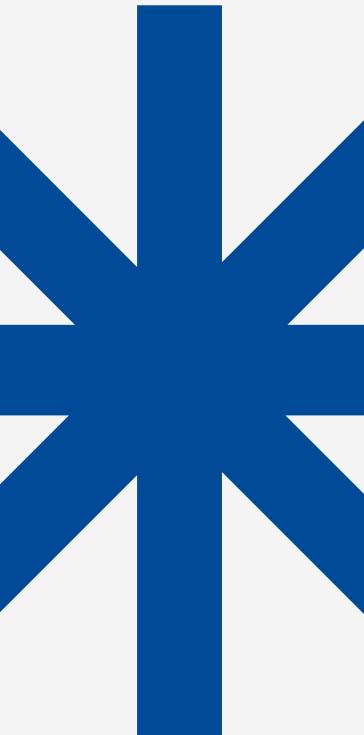
- 1) We pivoted the columns like 'Work Type', 'Qualifications', 'Preference' so that only numerical and binary values are present.
- 2) Next, we applied a lambda function to each element in the 'Average Pay' column. The lambda function checks if the value  $x$  is greater than 75. If it is, it assigns the value 1.0; otherwise, it assigns the value 0.0.
- 3) We can see that for company size above 12646 and for qualifications as BA, we'll always get a high paying job

# Correlation

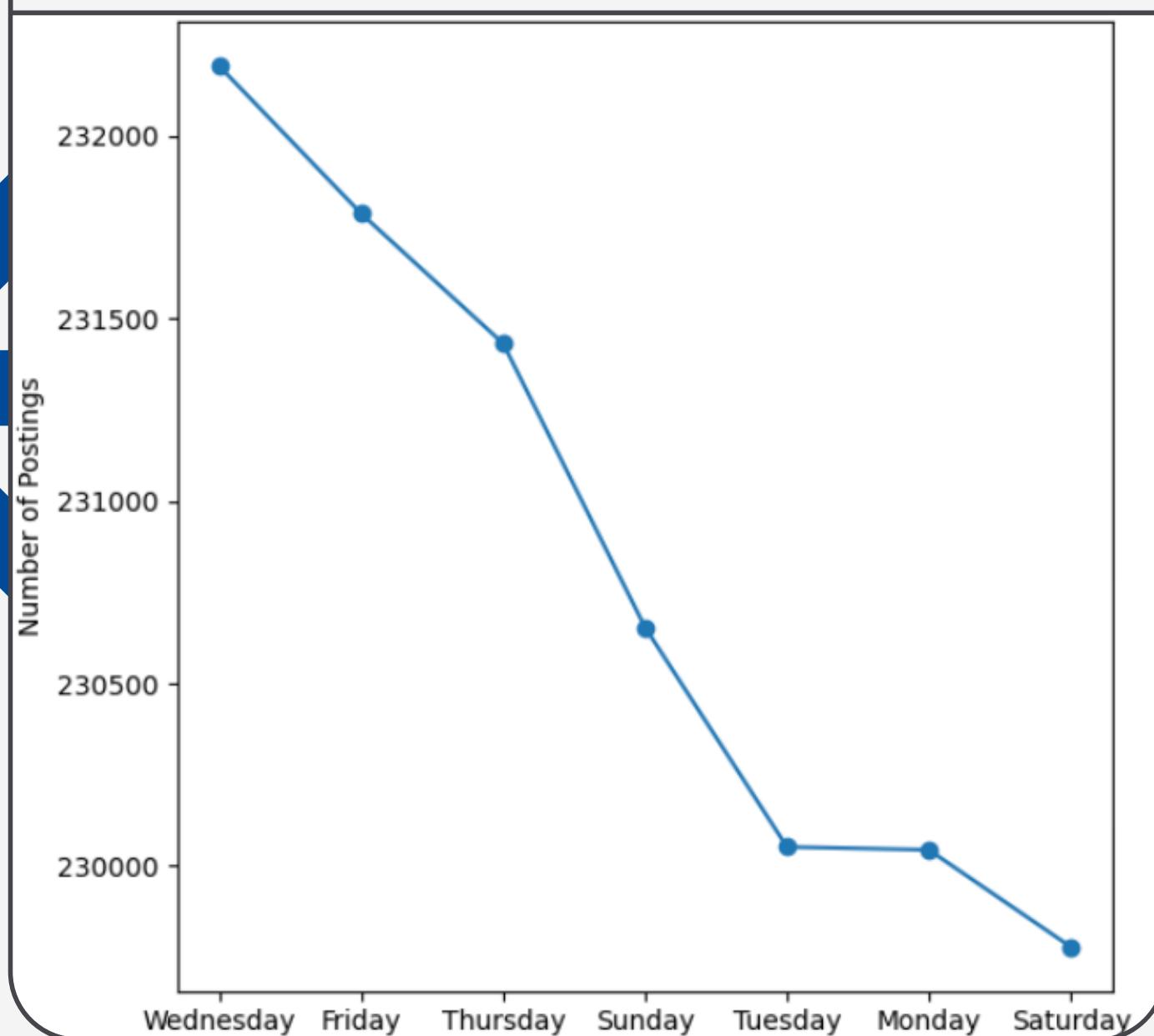
- 1) Data Preparation
- 2) Correlation Calculation
- 3) Identifying top correlations
- 4) Visualisation

As we can see, there isn't a high correlation between any of the columns so there is no need to drop any of them.

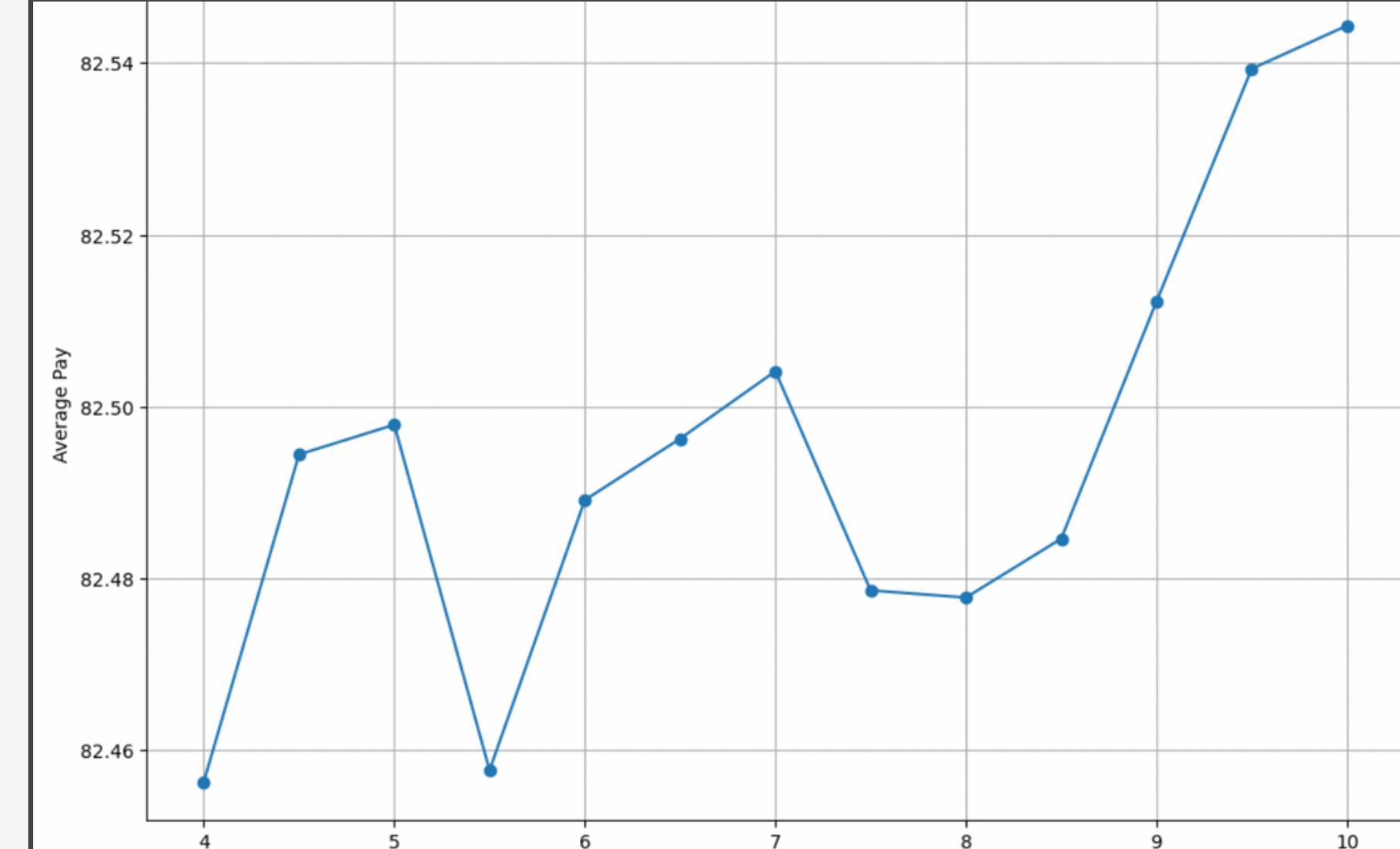




Job Postings by day of the week



X — Relation between Average Pay and Experience



## Extra Insights

- 1) We can see the peaks of when to apply for jobs
- 2) We can see the trend of how Experience affects the Average Pay

