



GLOBAL AI JOB MARKET & SALARY

COURSE: MIS 301
PRESENTED BY PANDAS
LECTURERS: Mr. Nguyen Tien Phat
Ms. Bui Nguyen Minh Ngoc

TEAM MEMBER

Name	ID
Dinh Nho Thanh Binh	2132300011
Nguyen Do Xuan Ai	2332300353
Bui Thu Thuy	1932300173



TABLE OF CONTENT



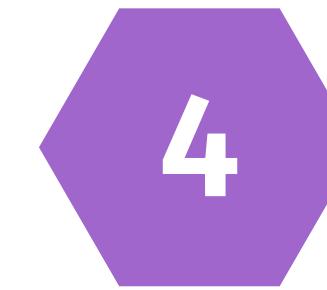
1 **Introduction**



2 **Data**



3 **Analysis**



4 **Conclusion**

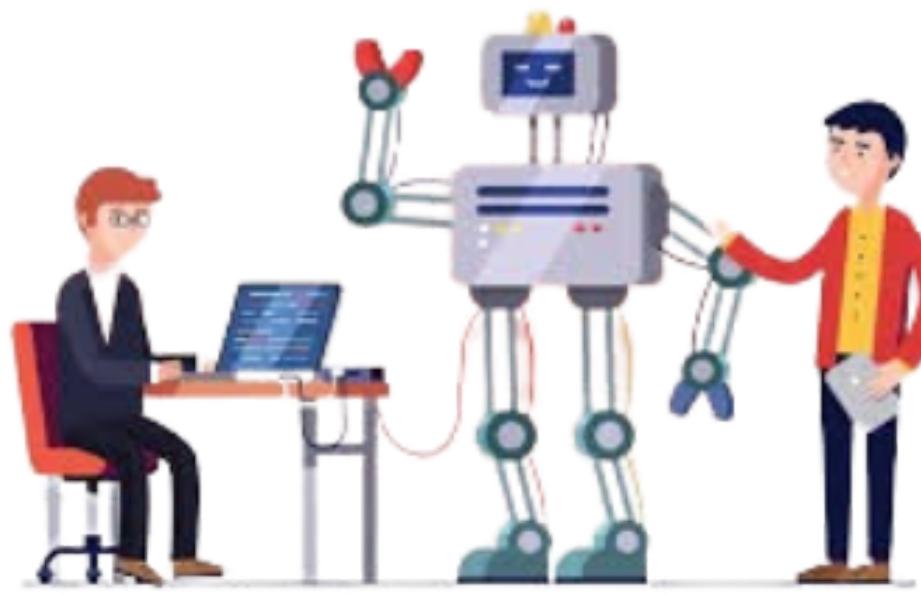
I. INTRODUCTION

Providing a comprehensive snapshot of the Artificial Intelligence (AI) job market

containing detailed information on various roles, salaries, required skills, and company demographics.

Offering a valuable opportunity to explore and understand the nuances of this rapidly evolving field

become increasingly important for individuals looking to enter or advance in the field, as well as for companies seeking to understand the talent landscape.

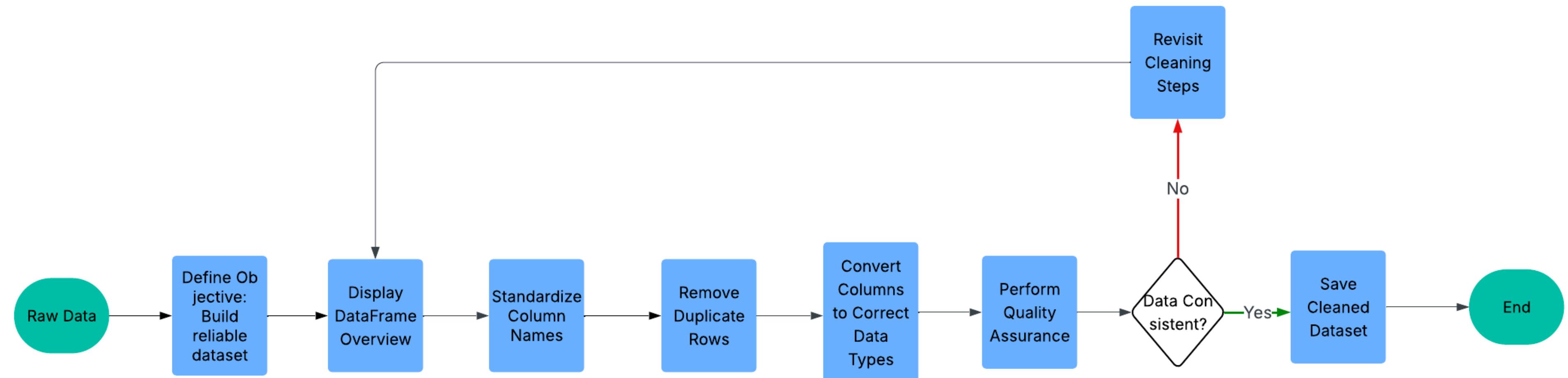


II. DATA

job_id	job_title	salary_usd	salary_currency	experience_level	employment_type	company_location	company_size	employee_residence	remote_ratio	required_skills	ed
AI00001	AI Research Scientist	90376.0	USD	SE	CT	China	M	China	50.0	Tableau, PyTorch, Kubernetes, Linux, NLP	AI
AI00002	AI Software Engineer	61895.0	USD	EN	CT	Canada	M	Ireland	100.0	Deep Learning, AWS, Mathematics, Python, Docker	AI
AI00003	AI Specialist	152626.0	USD	MI	FL	Switzerland	L	South Korea	0.0	Kubernetes, Deep Learning, Java, Hadoop, NLP	AI
AI00004	NLP Engineer	80215.0	USD	SE	FL	India	M	India	50.0	Scala, SQL, Linux, Python	NLP
AI00005	AI Consultant	54624.0	EUR	EN	PT	France	S	Singapore	100.0	MLOps, Java, Tableau, Python	AI

education_required	years_experience	industry	posting_date	application_deadline	job_description_length	benefits_score	company_name
Bachelor	9.0	Automotive	10/18/2024	11/07/2024	1076.0	5.9	Smart Analytics
Master	1.0	Media	11/20/2024	01/11/2025	1268.0	5.2	TechCorp Inc
Associate	2.0	Education	03/18/2025	04/07/2025	1974.0	9.4	Autonomous Tech
PhD	7.0	Consulting	12/23/2024	02/24/2025	1345.0	8.6	Future Systems
Master	0.0	Media	04/15/2025	06/23/2025	1989.0	6.6	Advanced Robotics

II. DATA



II. DATA

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 19 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   job_id          15000 non-null  object  
 1   job_title        14983 non-null  object  
 2   salary_usd       14980 non-null  float64 
 3   salary_currency  14979 non-null  object  
 4   experience_level 14979 non-null  object  
 5   employment_type  14981 non-null  object  
 6   company_location 14981 non-null  object  
 7   company_size      14980 non-null  object  
 8   employee_residence 14976 non-null  object  
 9   remote_ratio      14976 non-null  float64 
 10  required_skills   14975 non-null  object  
 11  education_required 14976 non-null  object  
 12  years_experience 14978 non-null  float64 
 13  industry          14979 non-null  object  
 14  posting_date       14981 non-null  object  
 15  application_deadline 14981 non-null  object  
 16  job_description_length 14981 non-null  float64 
 17  benefits_score     14981 non-null  float64 
 18  company_name       14997 non-null  object  
dtypes: float64(5), object(14)
memory usage: 2.2+ MB
```

II. DATA

1. Standardize column names by stripping whitespace, converting to lowercase, and replacing spaces with underscores
2. Removing duplicate rows from dataframe.

```
df_jobs.duplicated().sum()
```

```
np.int64(0)
```

II. DATA

3. Converting column to the correct type.

Convert 'salary_usd' to numeric

Convert salaries to USD

eur_to_usd_rate = 1.16

gbp_to_usd_rate = 1.34

Convert 'posting_date' and
'application_deadline' to datetime objects

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 19 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   job_id          15000 non-null  object  
 1   job_title        14983 non-null  object  
 2   salary_usd       14980 non-null  float64 
 3   salary_currency  14979 non-null  object  
 4   experience_level 14979 non-null  object  
 5   employment_type  14981 non-null  object  
 6   company_location 14981 non-null  object  
 7   company_size      14980 non-null  object  
 8   employee_residence 14976 non-null  object  
 9   remote_ratio      14976 non-null  float64 
 10  required_skills   14975 non-null  object  
 11  education_required 14976 non-null  object  
 12  years_experience 14978 non-null  float64 
 13  industry          14979 non-null  object  
 14  posting_date       14981 non-null  object  
 15  application_deadline 14981 non-null  object  
 16  job_description_length 14981 non-null  float64 
 17  benefits_score     14981 non-null  float64 
 18  company_name       14997 non-null  object  
dtypes: float64(5), object(14)
memory usage: 2.2+ MB
```

II. DATA

3. Converting column to the correct type.

salary_usd	salary_currency	salary_usd	salary_currency	posting_date	application_deadline	posting_date	application_deadline
90376.0	USD	90376.00	USD	10/18/2024	11/07/2024	2024-10-18	2024-11-07
61895.0	USD	61895.00	USD	11/20/2024	01/11/2025	2024-11-20	2025-01-11
152626.0	USD	152626.00	USD	03/18/2025	04/07/2025	2025-03-18	2025-04-07
80215.0	USD	80215.00	USD	12/23/2024	02/24/2025	2024-12-23	2025-02-24
54624.0	EUR	63363.84	USD	04/15/2025	06/23/2025	2025-04-15	2025-06-23

4. Handling missing values

job_id	job_title	salary_usd	salary_currency	experience_level	employment_type	company_location	company_size	employee_residence	remote_ratio	required_skills
AI00001	AI Research Scientist	90376.0	USD	SE	CT	China	M	China	50.0	Tableau, PyTorch, Kubernetes, Linux, NLP

education_required	years_experience	industry	posting_date	application_deadline	job_description_length	benefits_score	company_name
Bachelor	9.0	Automotive	10/18/2024	11/07/2024	1076.0	5.9	Smart Analytics

Drop

Drop rows with missing values in key columns(job title, company name, salary usd, education required)

Fill “Unknown”

These columns contain categorical data (required skills, employee residence, company locations, industries, experience levels)

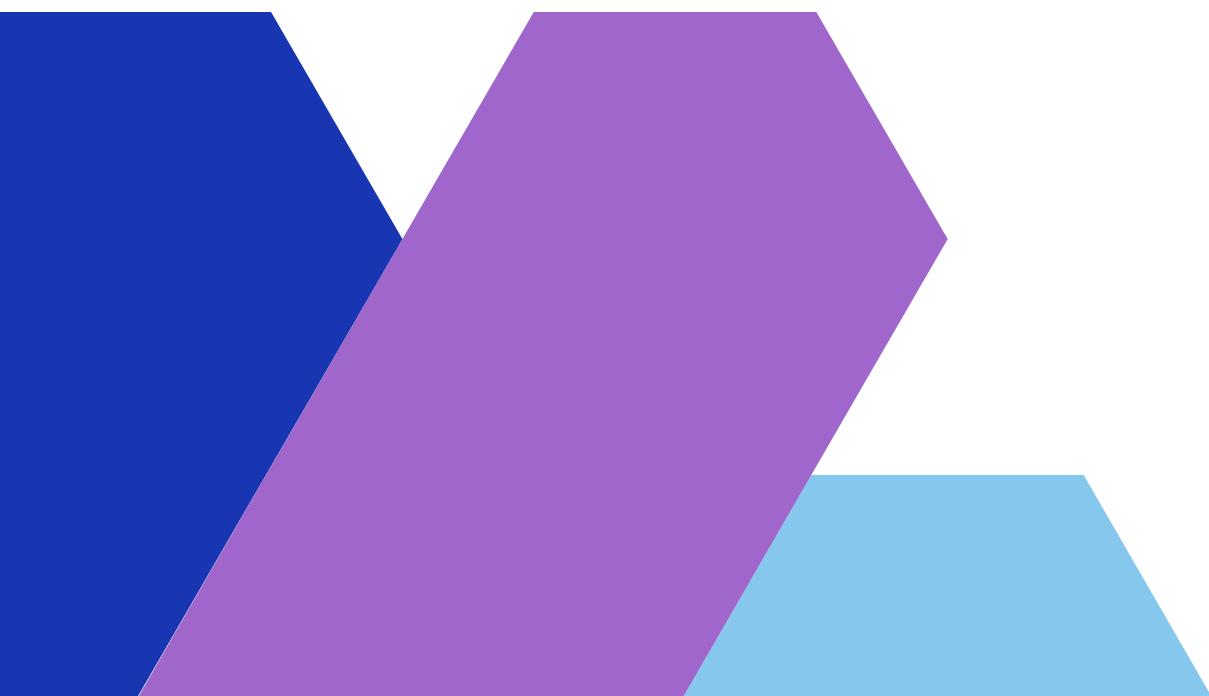
Fill mode

Fill missing values in specified columns with the mode of each column:(company size, employment type, salary currency)

Fill median

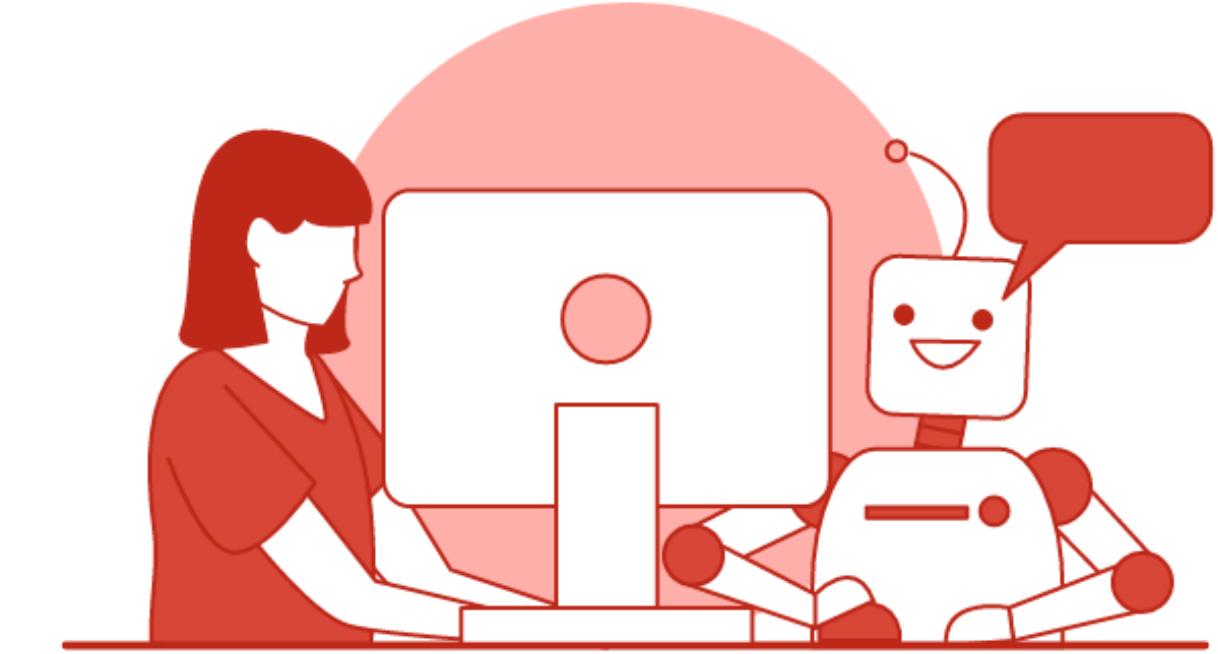
Fill missing numerical values with median of each column(remote_ratio, years_experience, benefits score, job description length, posting date, application deadline)

III. Business Problem Framing & Analytical Objectives



Business Context

- Strategic talent advisory firm in hyper-competitive AI sector
- Clients: AI-hiring companies | Candidates: AI professionals
- Market complexity: skills + experience + dynamic demand
- Average salary data = insufficient for elite-level advisory



Business Problem



Our ability to provide superior advice is limited by a one-dimensional view of the AI job market. We must build a 360° Market Intelligence Framework to answer critical questions for both clients and candidates.

Must build proprietary framework to answer:

- Clients: How to make an irresistible offer?
- Candidates: How to build a top-tier career?

Required Insights

What We Must Understand Before Advising



01

Core Technical Skills

Beyond buzzwords, what are the specific, recurring technical skills that employers consistently demand?

02

Remote Work Trends

What is the actual prevalence of on-site, hybrid, and fully remote work models in the AI field?

03

Demand Hotspots

Which industries are the most aggressive in hiring AI talent?

04

Elite Employers

Beyond just salary, who are the most attractive employers in each key region when considering a holistic package of compensation, benefits, and work-life flexibility?

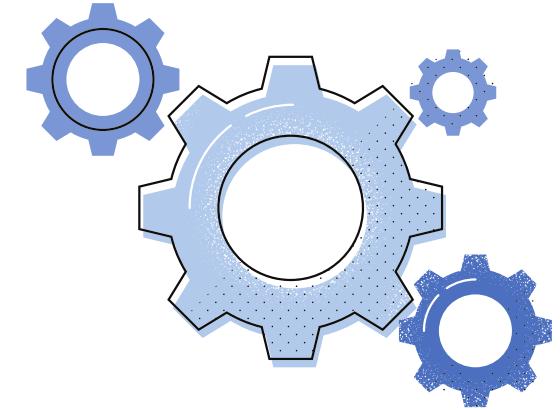
05

Salary Drivers

How, precisely, do factors like years of experience and level of education quantitatively impact an individual's earning potential?

Analytical Objectives

Translating Insights → Measurable Objectives



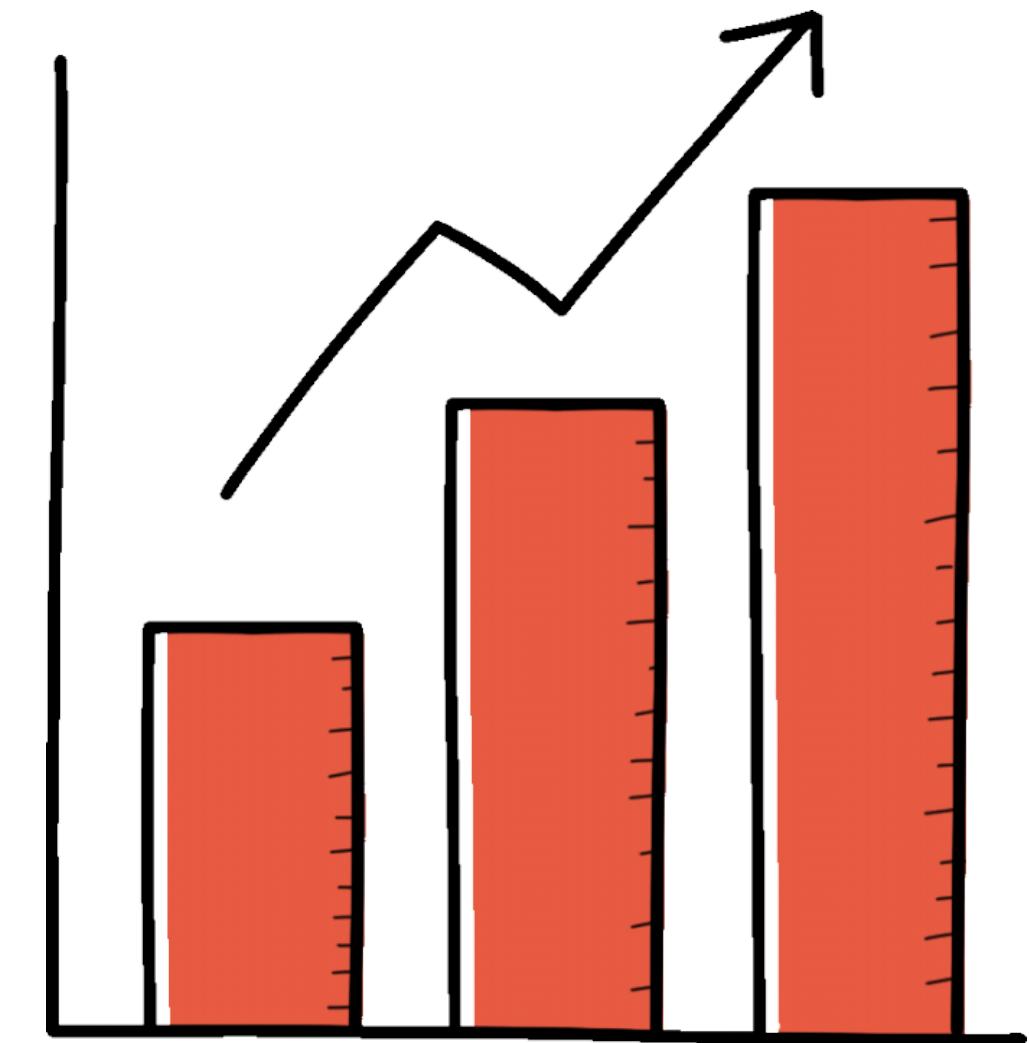
Insight	Objective
Core Technical Skills	Identify & rank Top 10 in-demand Skills
Remote Work Trends	Quantify prevalence & salary impact
Demand Hotspots	Identify top hiring industries
Elite Employers	Identify & rank the Top 5 Most Attractive Employers by Location.
Salary Drivers	Quantify impact of experience years & education levels

Data, Functions, Visuals Alignment

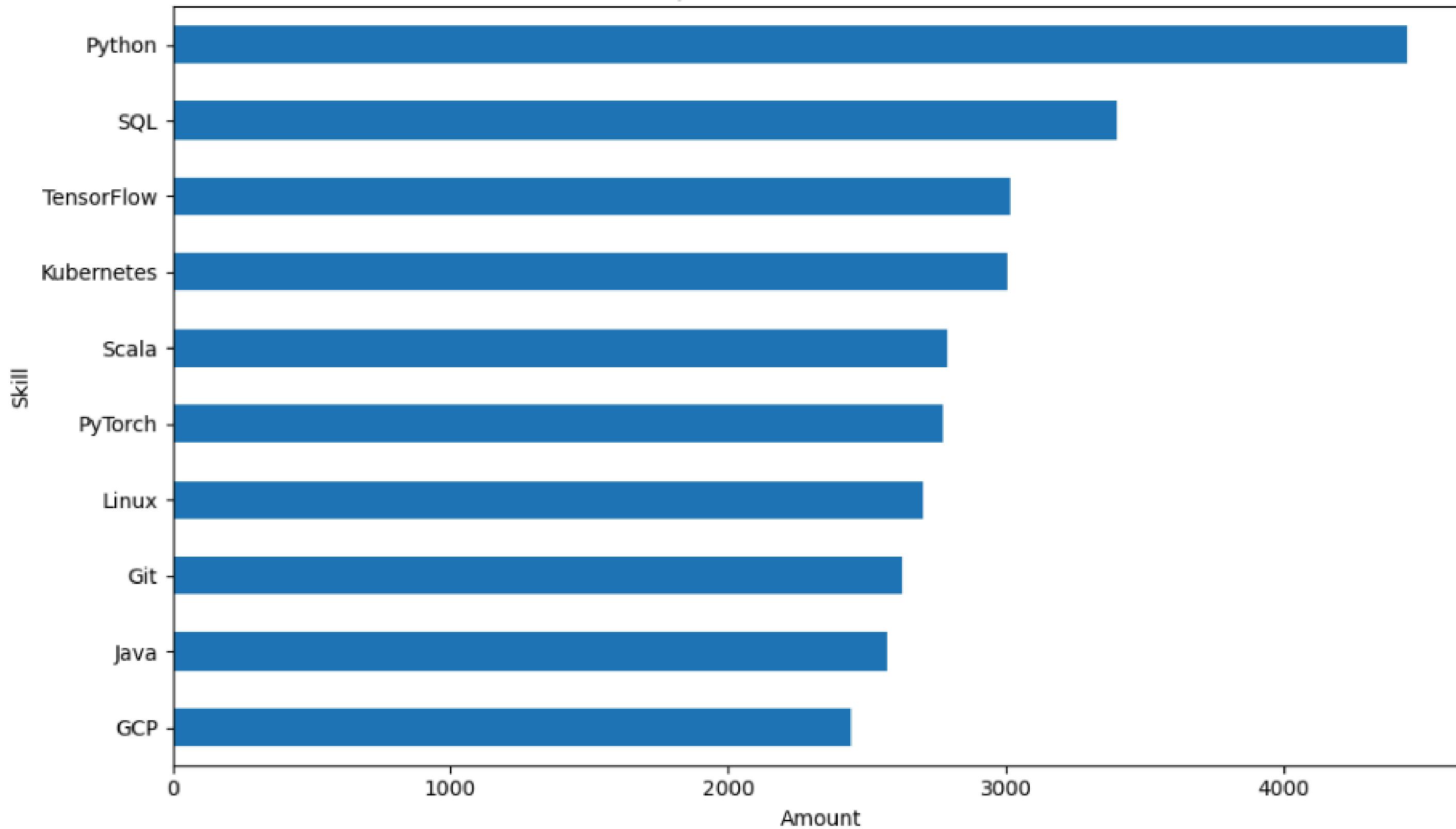
Translating Insights → Measurable Objectives

Analytical	Required Data	Method / Function	Visualization
Core Skills	required_skills	COUNT(), RANK()	Horizontal Bar Chart
Remote Work	remote_ratio, salary_usd	COUNT(), %, AVG()	Donut + Bar Combo
Industries	industry	COUNT(), RANK()	Treemap
Employer Score	company_name, salary, benefits, remote_ratio	Normalize → Weighted Avg. → Rank	Ranked Bar Chart
Salary Drivers	years_exp, edu_level, salary_usd	Corr., Regression, GROUP BY	Line chart + Box Plots

III. Exploratory Data Analysis (EDA) & Visualisation

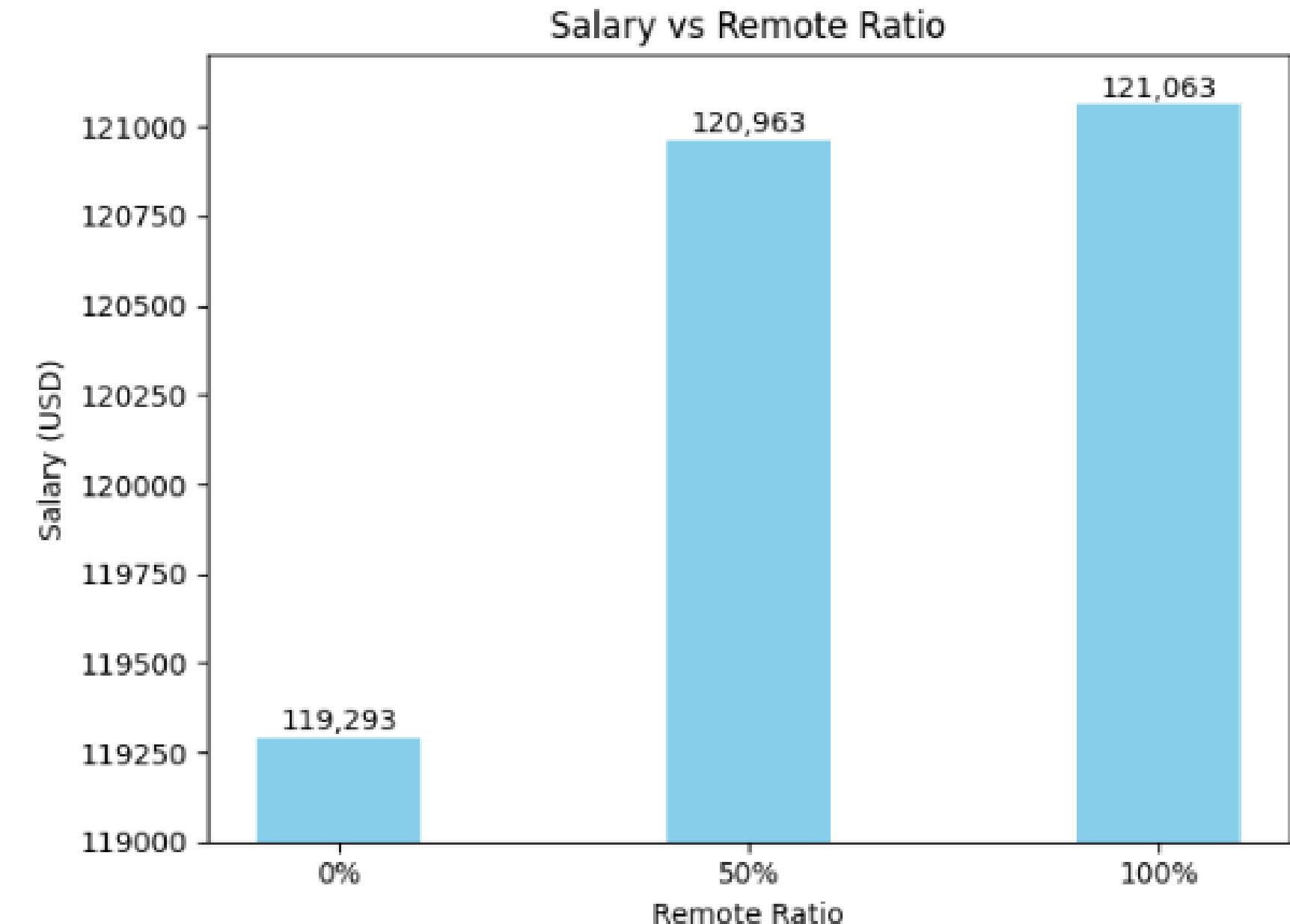
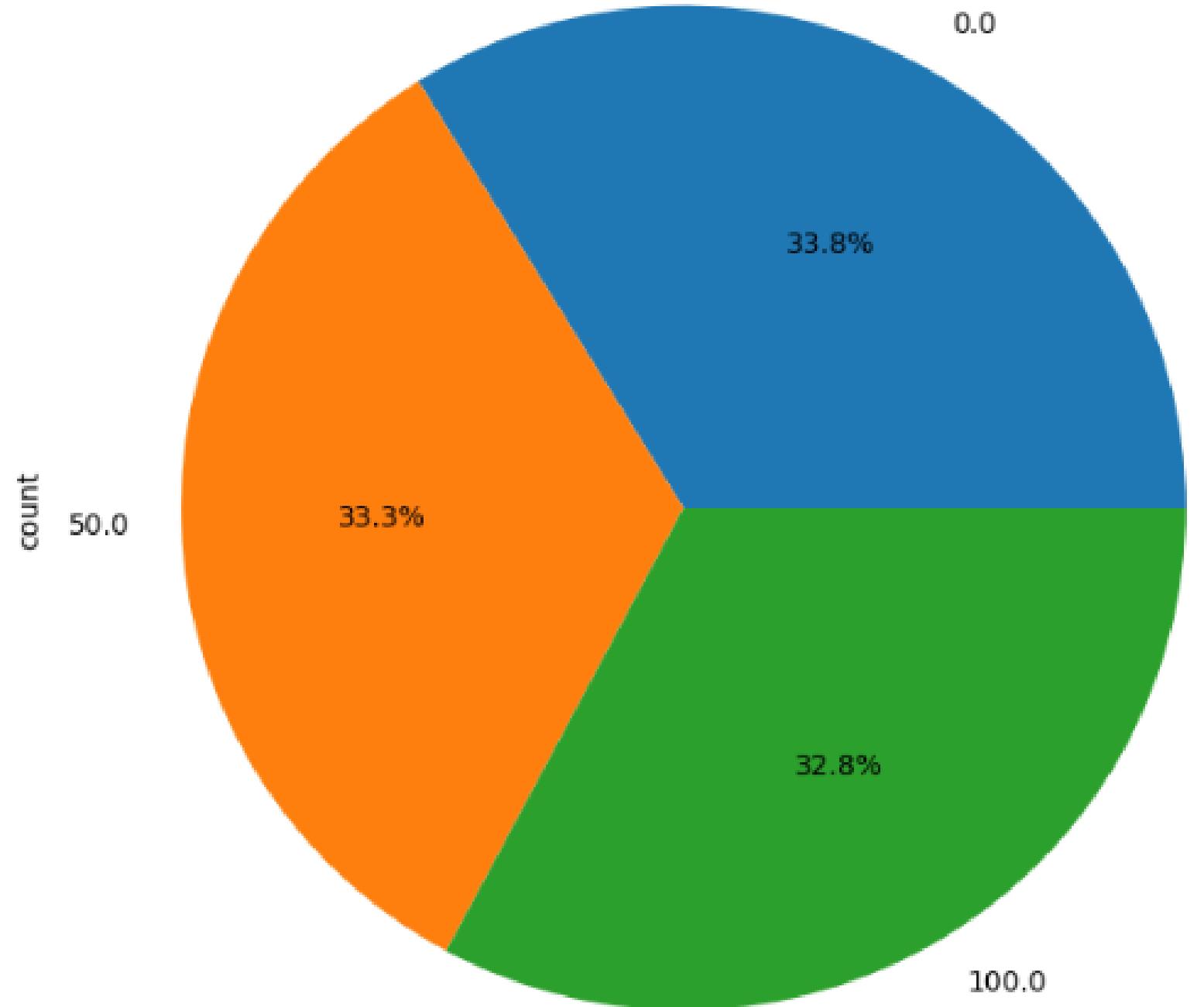


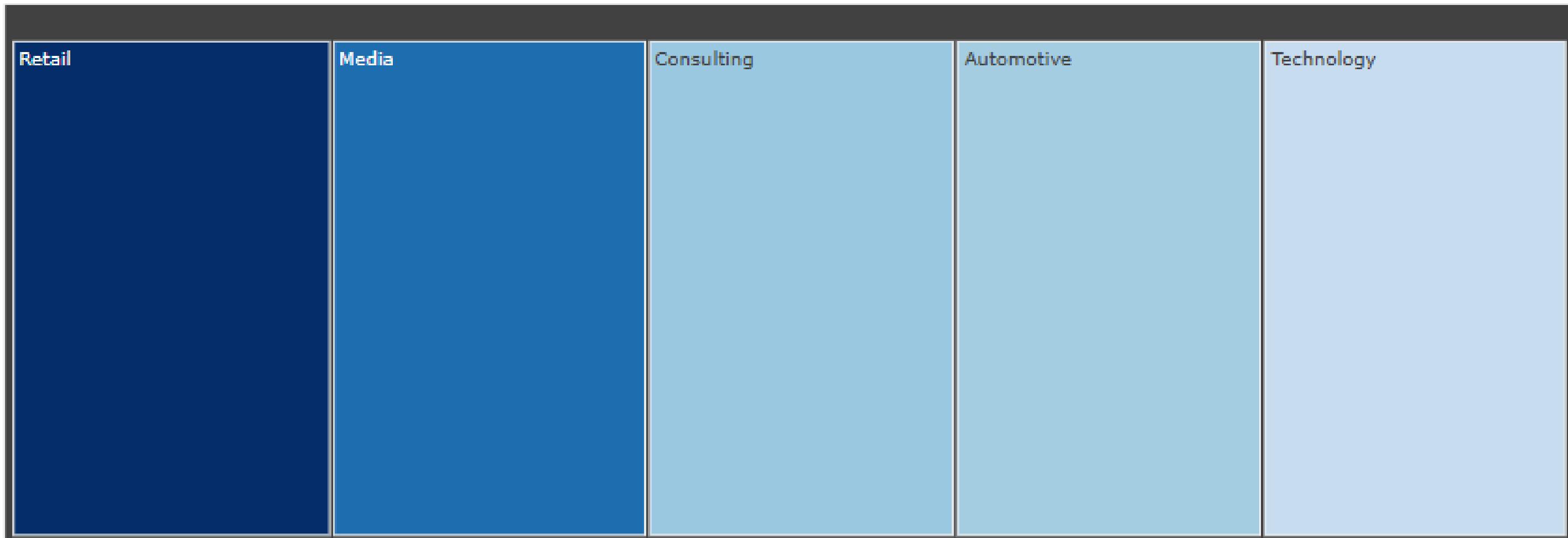
Top 10 Core Technical Skills



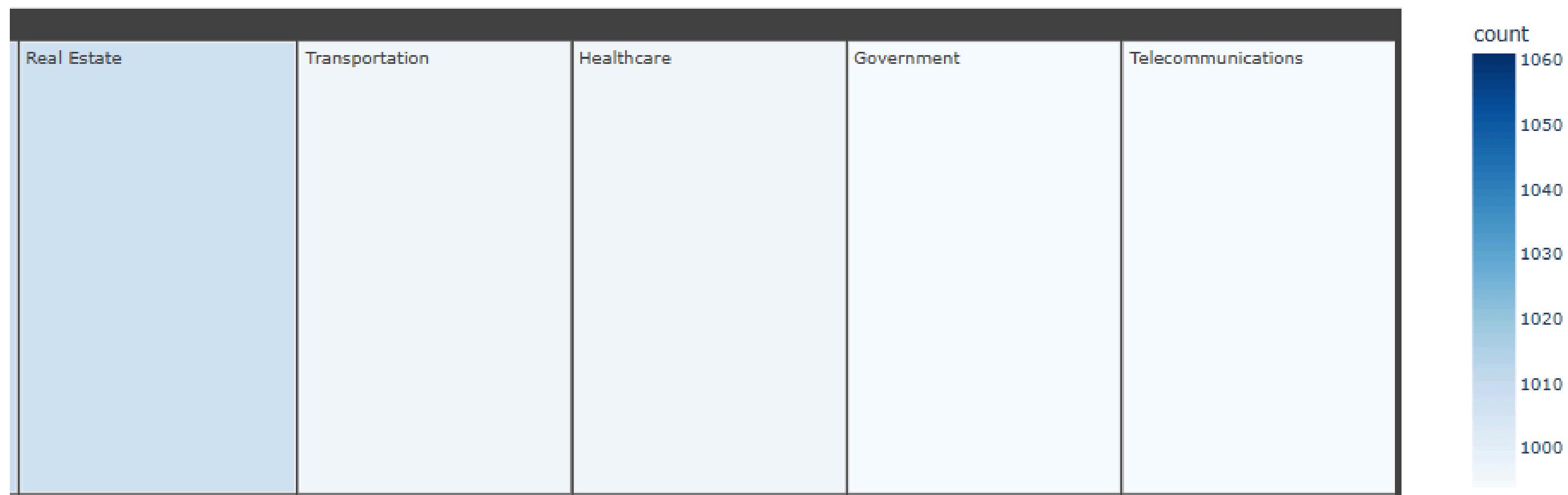
1st: Core Technical Skills

2nd: Remote Work Trends

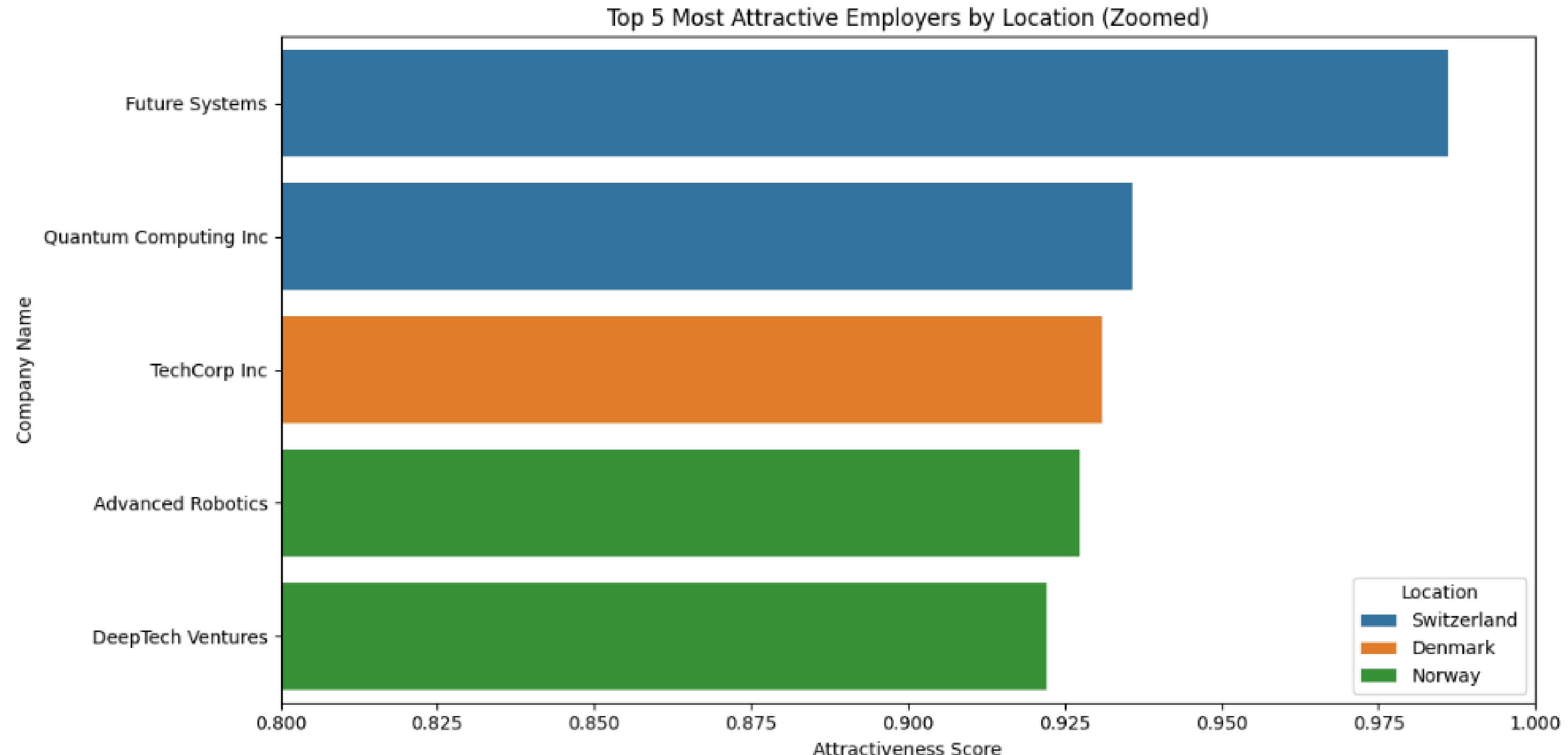




3rd:
Demand
Hotspots

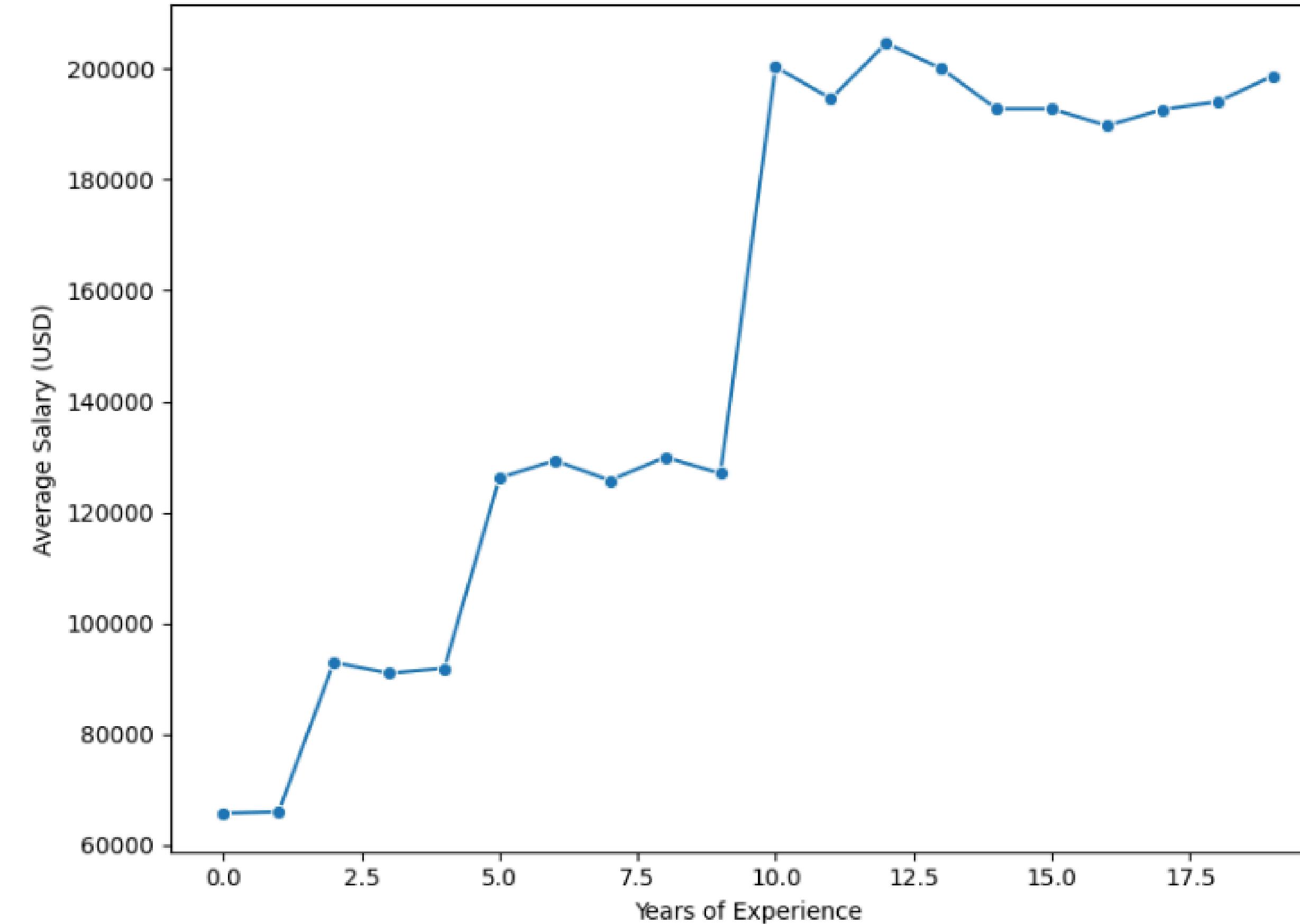


4th: Elite Employers

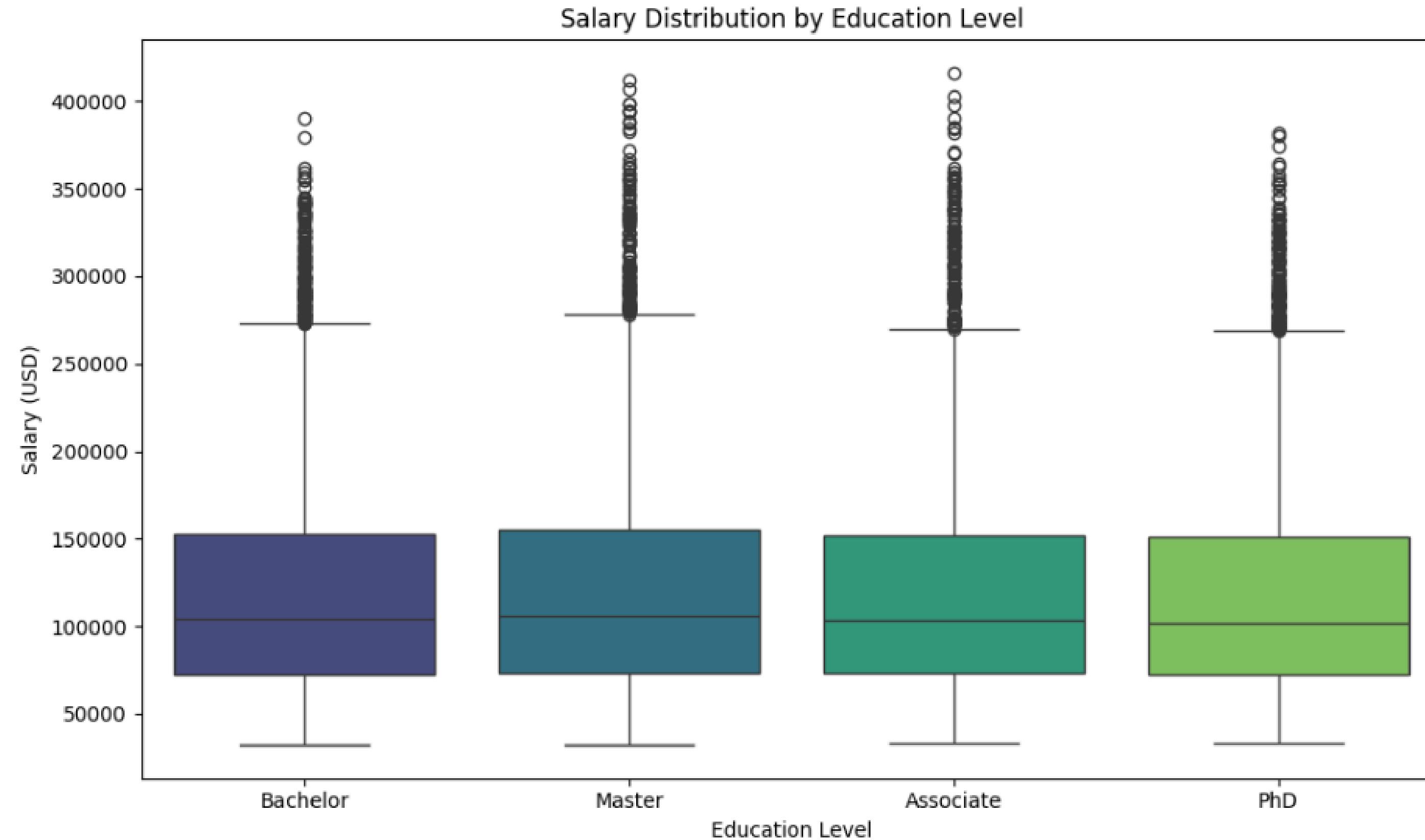


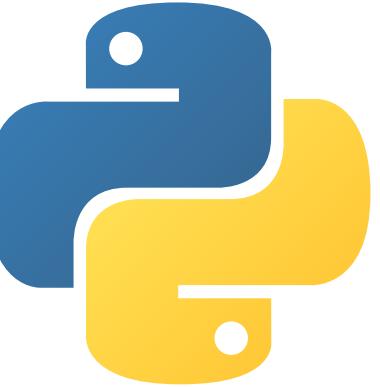
5th: Salary Drivers

Average Salary by Years of Experience



5th: Salary Drivers





**THANK YOU
FOR LISTENING**

