



# Jobs Analysis and Recommendation

Dang Huu Phat    Phan Nguyen Huu Phong    Pham Quang Nhut

University of Information Technology

23th November 2024

# Agenda

- 1 **Introduction**
- 2 **Dataset**
- 3 **Exploratory Data Analysis**
- 4 **Model**
- 5 **Contribution and Futurework**
- 6 **Demo**

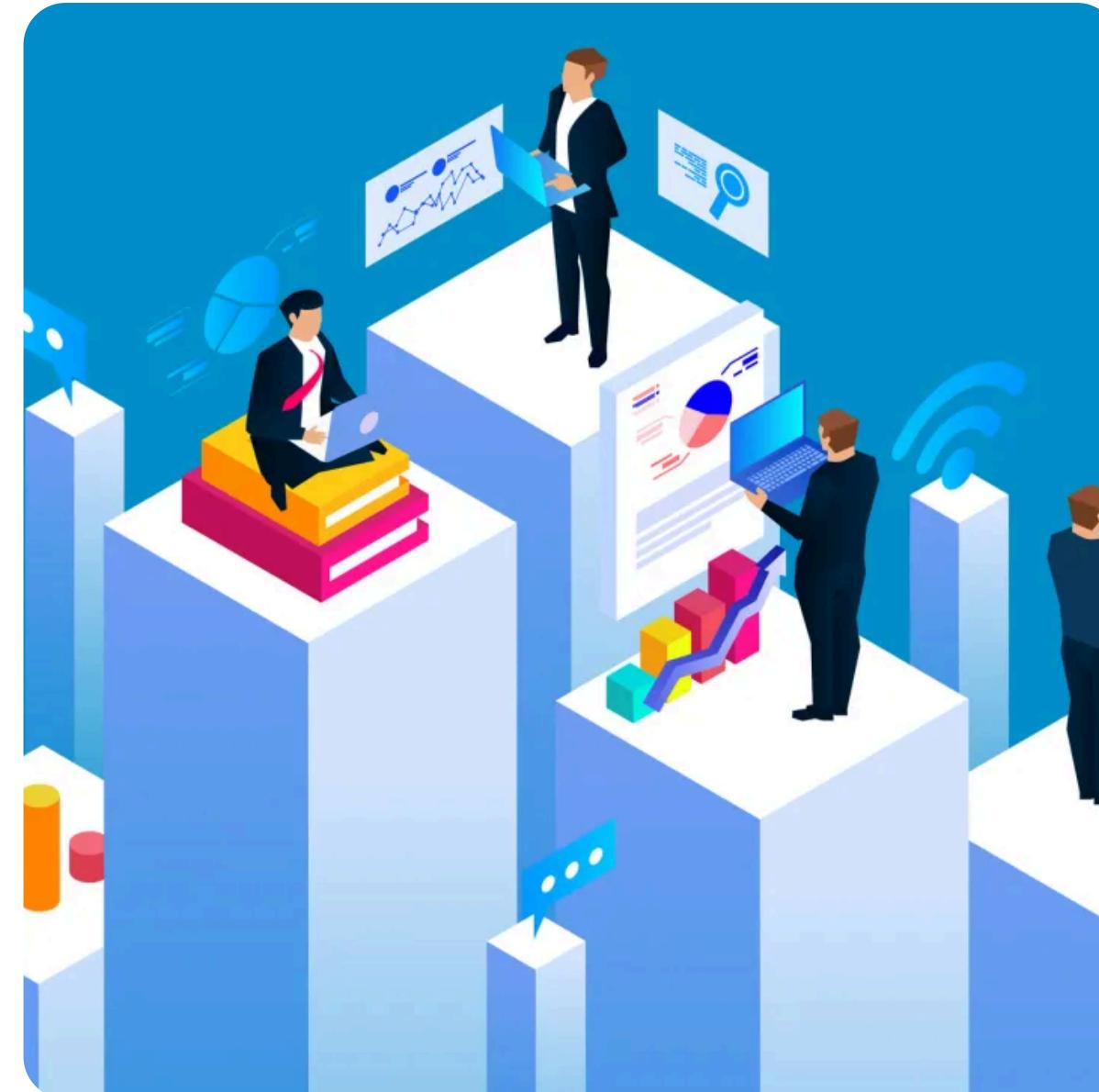
1

# INTRODUCTION



# Introduction

- Students struggling in choosing a career path
- Collect and analyze job-salary data to provide useful information



2

# DATASET

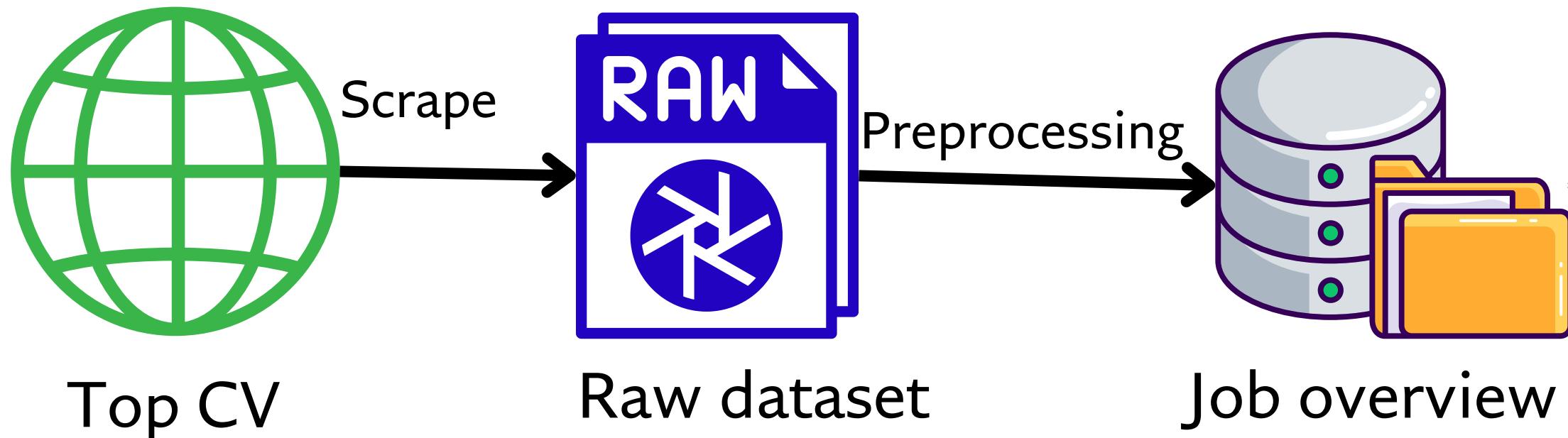


# Dataset

- Crawled from TopCV website [topcv việc làm mới nhất](#)
- Update 4/12/2024
- 2 file
  - job\_overview: Contains an overview of all jobs across all industries (complete data from topCV - updated on 04/12/2024)
  - IT\_job\_detail: Detailed information about IT job positions.

# Dataset

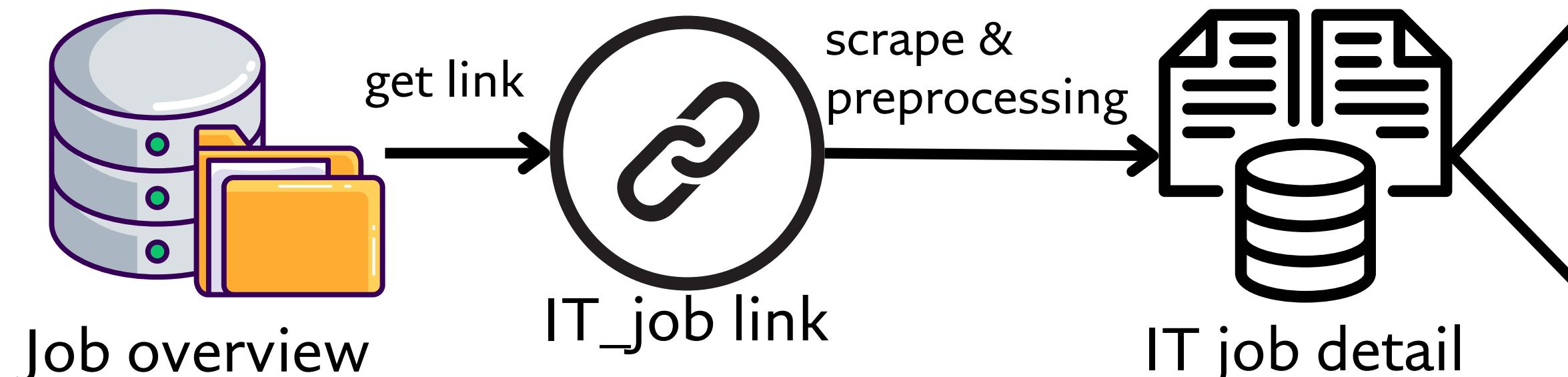
- Crawled from TopCV website [topcv việc làm mới nhất](#)
- Update 4/12/2024



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41742 entries, 0 to 41741
Data columns (total 12 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   Job Title        41742 non-null  object  
 1   Company Name     41742 non-null  object  
 2   Salary            41742 non-null  object  
 3   Location          41742 non-null  object  
 4   Experience        41742 non-null  object  
 5   Main Tags         41741 non-null  object  
 6   Additional Tags  20467 non-null  object  
 7   Job Link          41742 non-null  object  
 8   Posted Date       41742 non-null  object  
 9   Job ID            41742 non-null  int64   
 10  Job Position      41742 non-null  int64   
 11  Job Category      41742 non-null  object  
dtypes: int64(2), object(10)
```

# Dataset

- Crawled from TopCV website [topcv việc làm mới nhất](#)
- Update 4/12/2024



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5235 entries, 0 to 5234
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Title            4985 non-null    object 
 1   Salary           5021 non-null    object 
 2   Experience       4985 non-null    object 
 3   Description      4122 non-null    object 
 4   Requirements     4104 non-null    object 
 5   Benefits          4153 non-null    object 
 6   Company          4985 non-null    object 
 7   Level             4985 non-null    object 
 8   Vacancies         4985 non-null    object 
 9   Type              4985 non-null    object 
 10  Gender            4985 non-null    object 
dtypes: object(11)
memory usage: 450.0+ KB
```

# Dataset -Preprocessing

- Normalize Salary
  - Convert all salaries to **million VND**
  - Split into two columns: **min\_salary** and **max\_salary**
- Convert experience values into numerical format
- Posted Date:
  - The crawl date is Dec 4, 2024
  - Convert dates to the mm/dd/yy format

# Data Overview - Samples

## ● data\_overview

Job Title	Company Name	Location	Experience	Posted Date	Job ID	Job Position	Job Category	min_salary	max_salary	new_tags
Nhân Viên Kế Toán	Công ty cổ phần Thẩm định giá Big Việt Nam	Hà Nội	1	12/01/2024	1552536	27	kế toán, kiểm toán, thuế	8.0	10.0	Kế toán thuế; Tài chính; Kế toán; Thuế; Tài chính;...
Nhân Viên Kỹ Thuật Nhà Máy	CÔNG TY TNHH SẢN XUẤT, LẮP RÁP TUẤN NGHĨA	Hà Nội	1	11/06/2024	1404142	33	sản xuất	-1.0	-1.0	Bảo trì, sửa chữa khác
Chuyên Viên Vận Hành Lớp Học Online	CÔNG TY CỔ PHẦN AUTHENTIC EDUCATION HUB	Hà Nội	1	11/28/2024	1550198	10	dịch vụ khách hàng, vận hành	-1.0	-1.0	Vận hành sàn thương mại điện tử; Giáo dục / Đào...
Chăm Sóc Khách Hàng/ Nhân Viên Tư Vấn/ Telesal...	CÔNG TY CỔ PHẦN POWER MEMBERSHIP CARD	Nghệ An	0	11/06/2024	1526493	13	kinh doanh, bán hàng	8.0	15.0	Chăm sóc khách hàng
Marketing Facebook Ads - Không Yêu Cầu Kinh Ng...	Công Ty TNHH Dược Phẩm Linh Nam	Hà Nội	1	11/27/2024	1335615	25	marketing, pr, quảng cáo	6.5	8.0	Digital Marketing

# Data Overview - Samples

## ● IT\_job\_detail

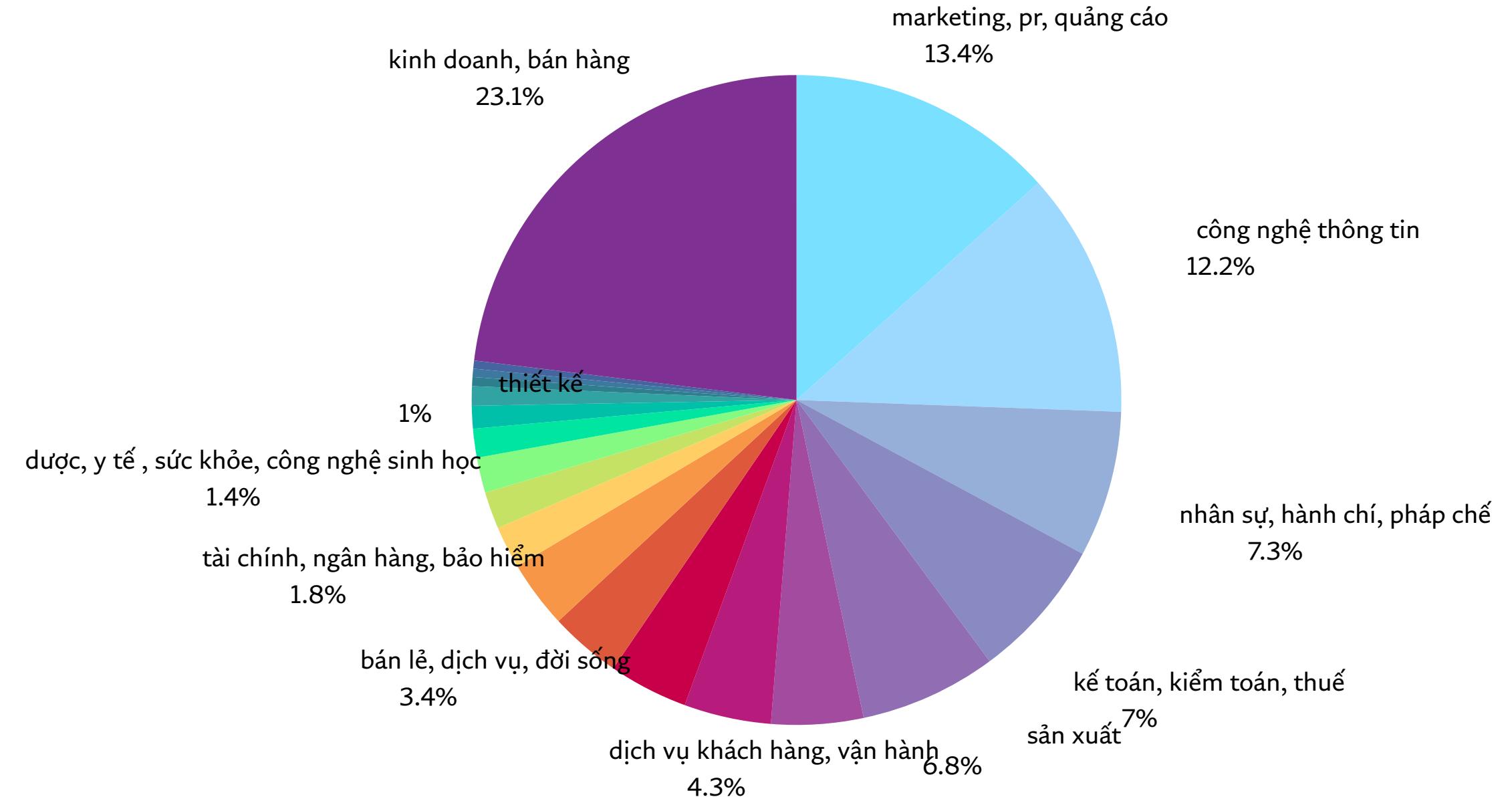
Title	Salary	Experience	Description	Requirements	Benefits	Company	Level	Vacancies	Type	Gender
Nhân Viên Kinh Doanh/SalesB2B/ Nhân Viên Tư VẤ...	12 - 40 triệu	1 năm	NaN	NaN	NaN	CÔNG TY TNHH CNV HOLDINGS	Nhân viên	7 người	Toàn thời gian	Không yêu cầu
Chuyên Viên Thiết Kế/Designer	12 - 14 triệu	1 năm	Thiết kế sản phẩm phục vụ cho việc quảng cáo o...	Sử dụng thành thạo phần mềm và công nghệ thiết...	1. Chế độ Thu nhập (11- 15 triệu): Lương cứng ...	CÔNG TY TNHH TAEHUI	Nhân viên	1 người	Toàn thời gian	Không yêu cầu
Chuyên Viên Kinh DoanhLĩnh vựcThương Mại Điện ... ...	10 - 50 triệu	Dưới 1 năm	NaN	Độ tuổi từ 22 - 36- Ứng viên có ít nhất 6 thán...	Lương cơ bản 10 triệu + thưởng + hoa hồng- Hoa...	CÔNG TY TNHH MEDIASTEP SOFTWARE VIỆT NAM	Nhân viên	5 người	Toàn thời gian	Không yêu cầu
IT Helpdesk	10 - 13 triệu	2 năm	Hỗ trợ người dùng xử lý các sự cố máy tính PC,...	Tốt nghiệp chuyên ngành công nghệ thông tin Đã...	Chế độ BHXH, BH24H, khám sức khoẻ định kỳ hàng...	CÔNG TY CỔ PHẦN TẬP ĐOÀN DẦU KHÍ AN PHAPRO Com...	Nhân viên	1 người	Toàn thời gian	Nam
Kỹ Sư Cầu Nối(ARTGenarate System)	40 - 60 triệu	1 năm	NaN	NaN	NaN	Công ty Cổ phần Mirabo	Nhân viên	1 người	Toàn thời gian	Không yêu cầu

3

# Exploratory Data Analysis

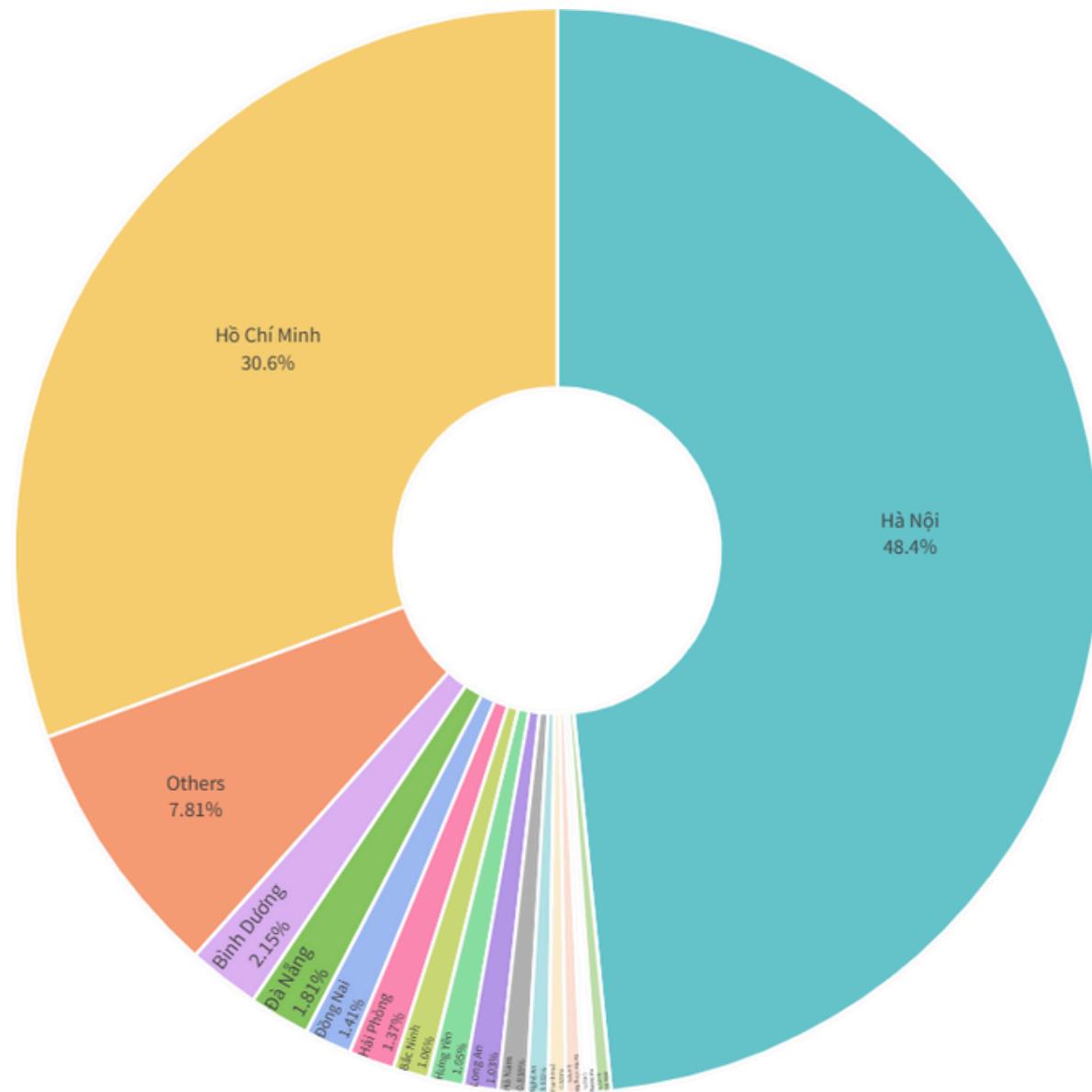


- Job category distribution (total 44230 Jobs)



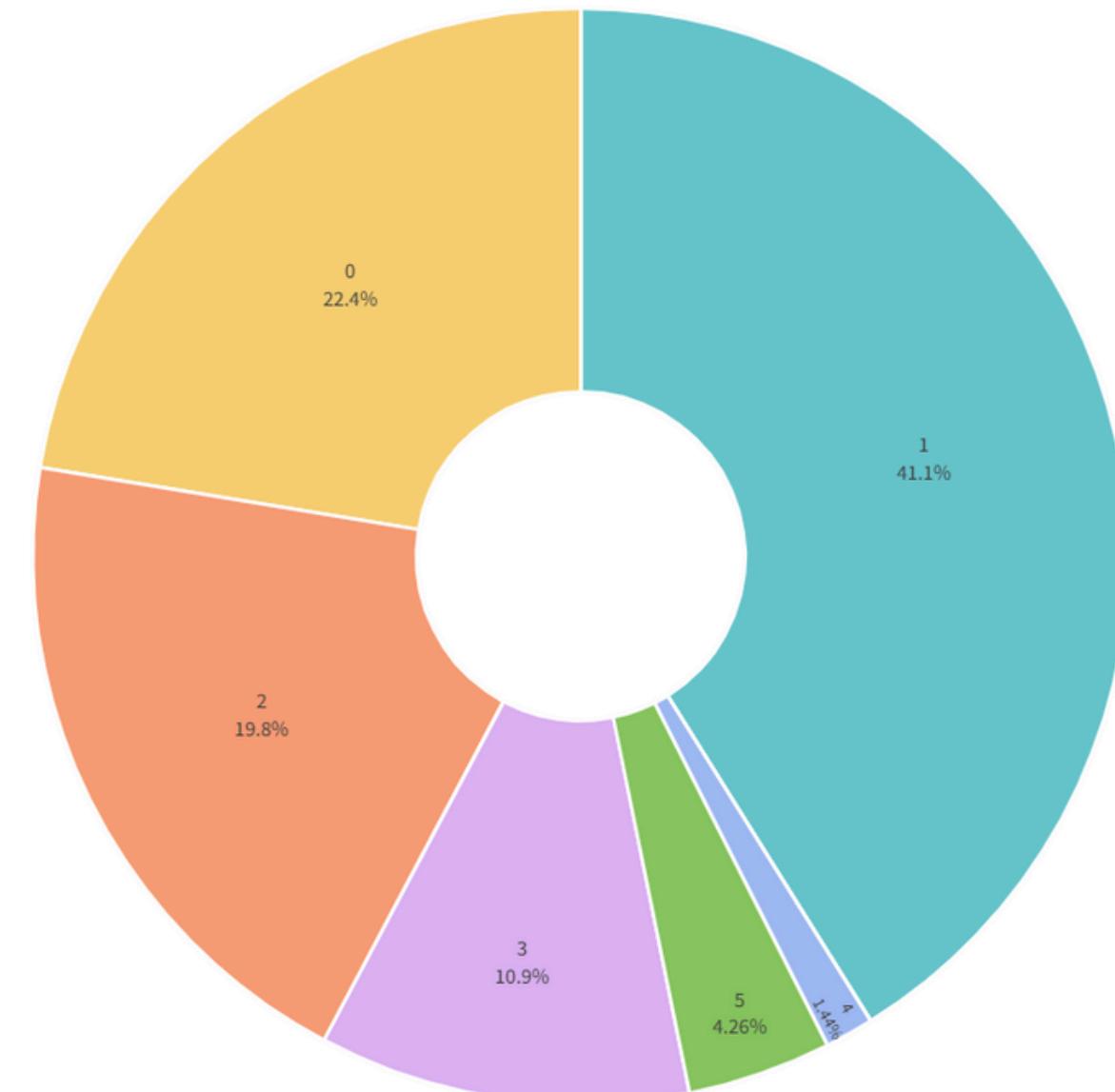
- 'Kinh doanh bán hàng' is the most prominent job category, followed by Marketing, IT.

# Distribution of jobs by Location



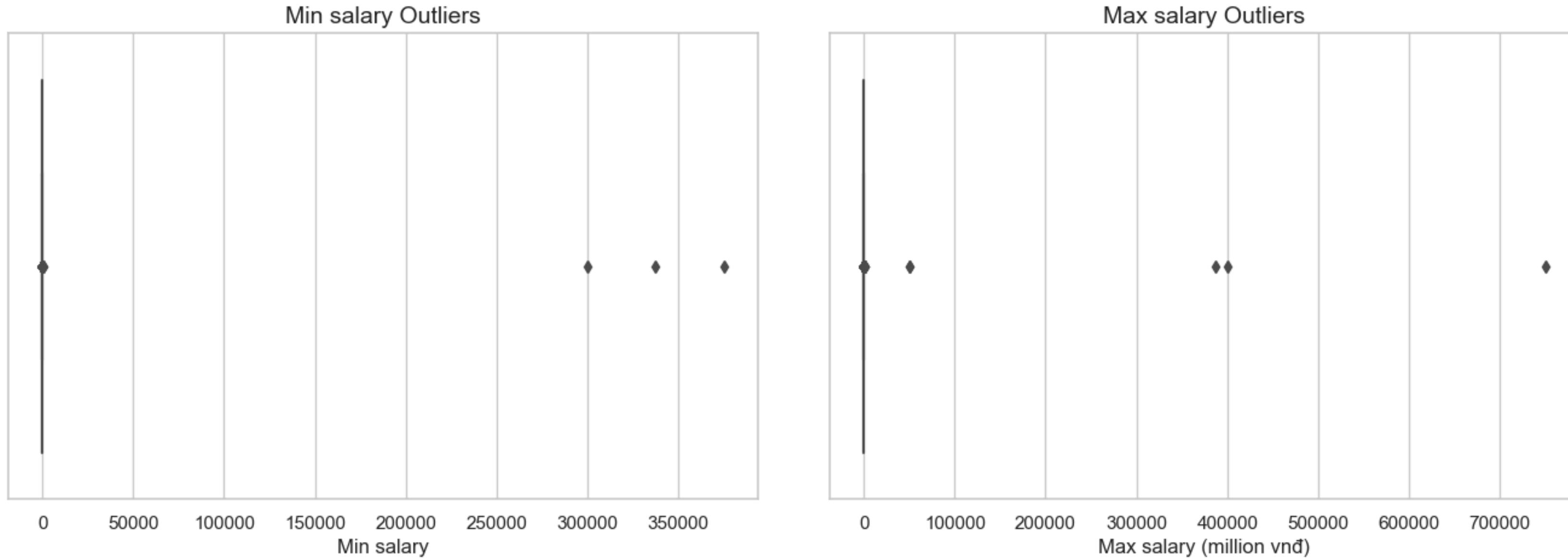
- Hanoi and Ho Chi Minh is the biggest job market

## ● Distribution of jobs by Year of Experience



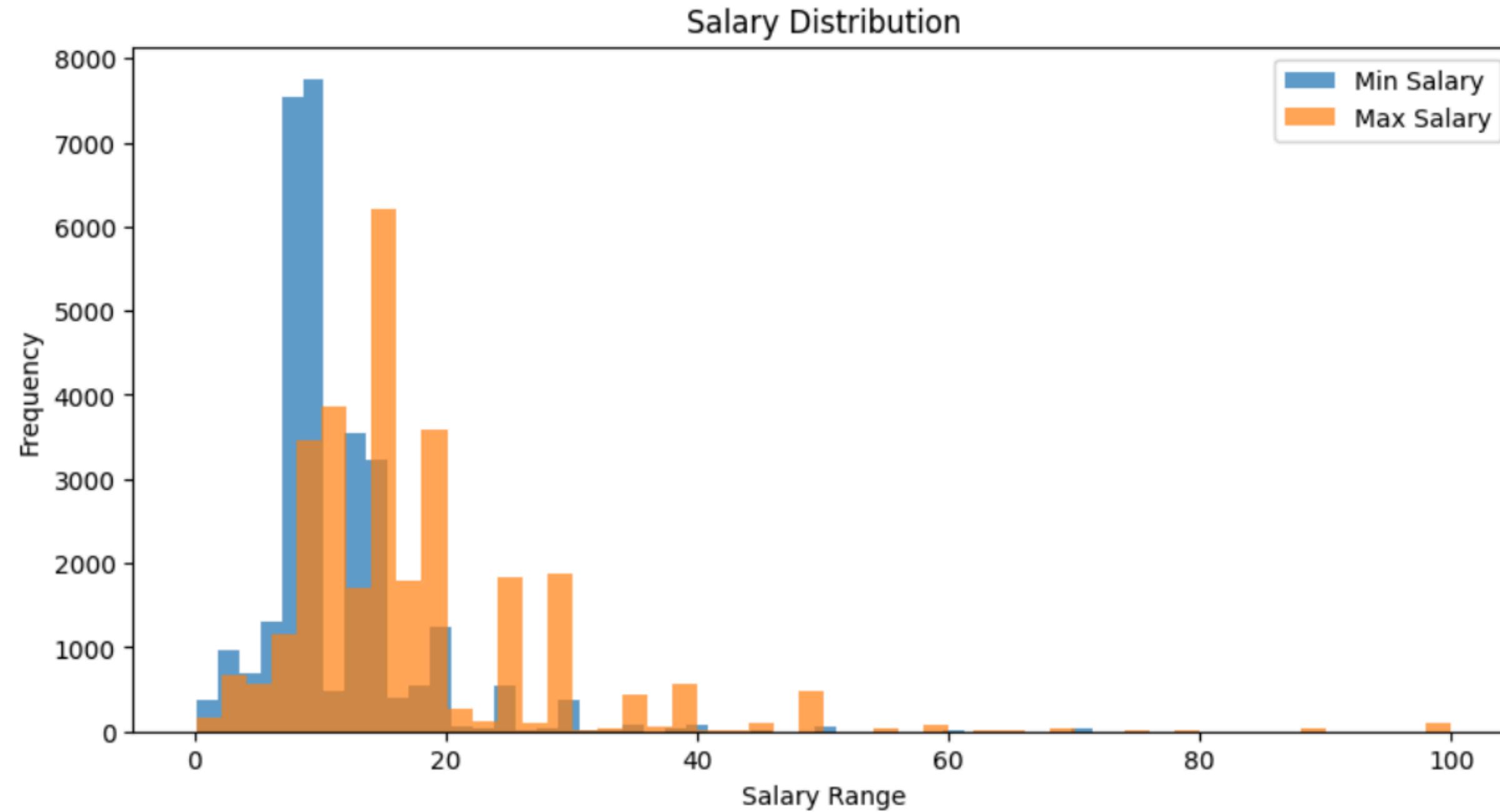
- Over half of the available jobs require less than 1 year of experience

# EDA



- Salary has outliers value  
→ remove these rows

- Distribution of min and max salary.



- Data is skew and need some transformation

# EDA

- Top 5 job categories with the highest maximum salaries.



- IT salary is slightly higher than other job categories

- Top 5 Location with the highest average salaries.



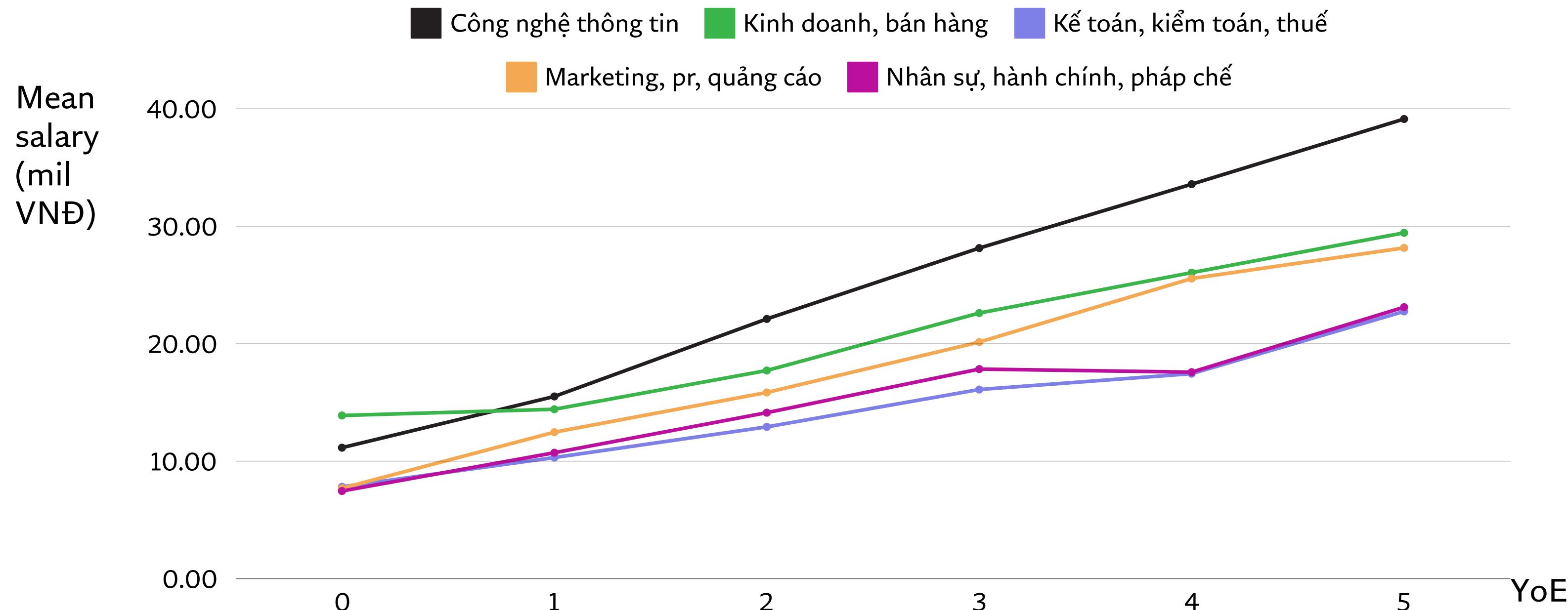
- Working abroad may offer the higher salaries, followed by provinces with limited jobs

- Average Salary by years of Experience



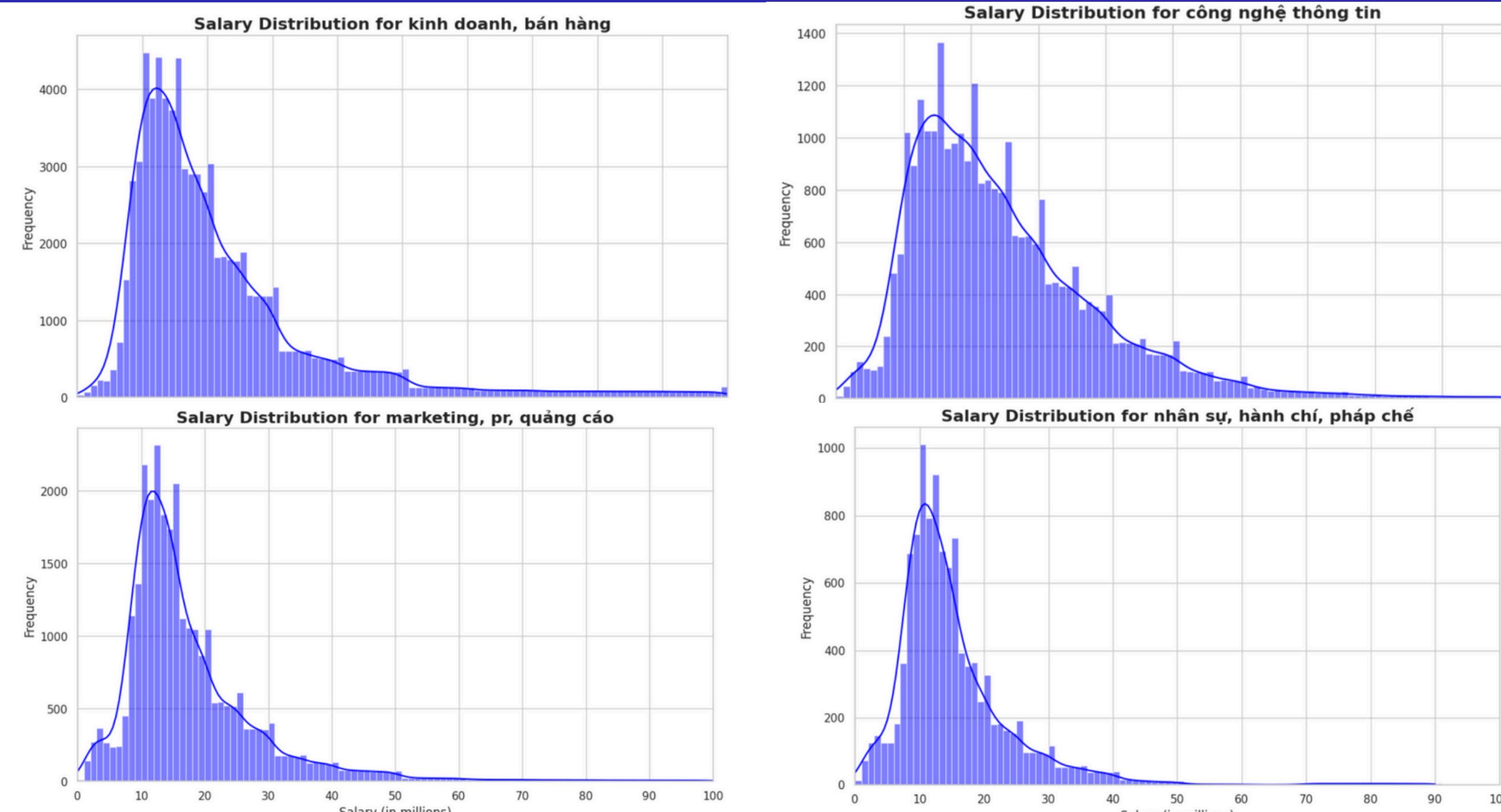
- More experience -> higher salary but the rising rate sinks over years

- Average Salary by years of Experience



- Rate of salary growth can differ significantly across job categories

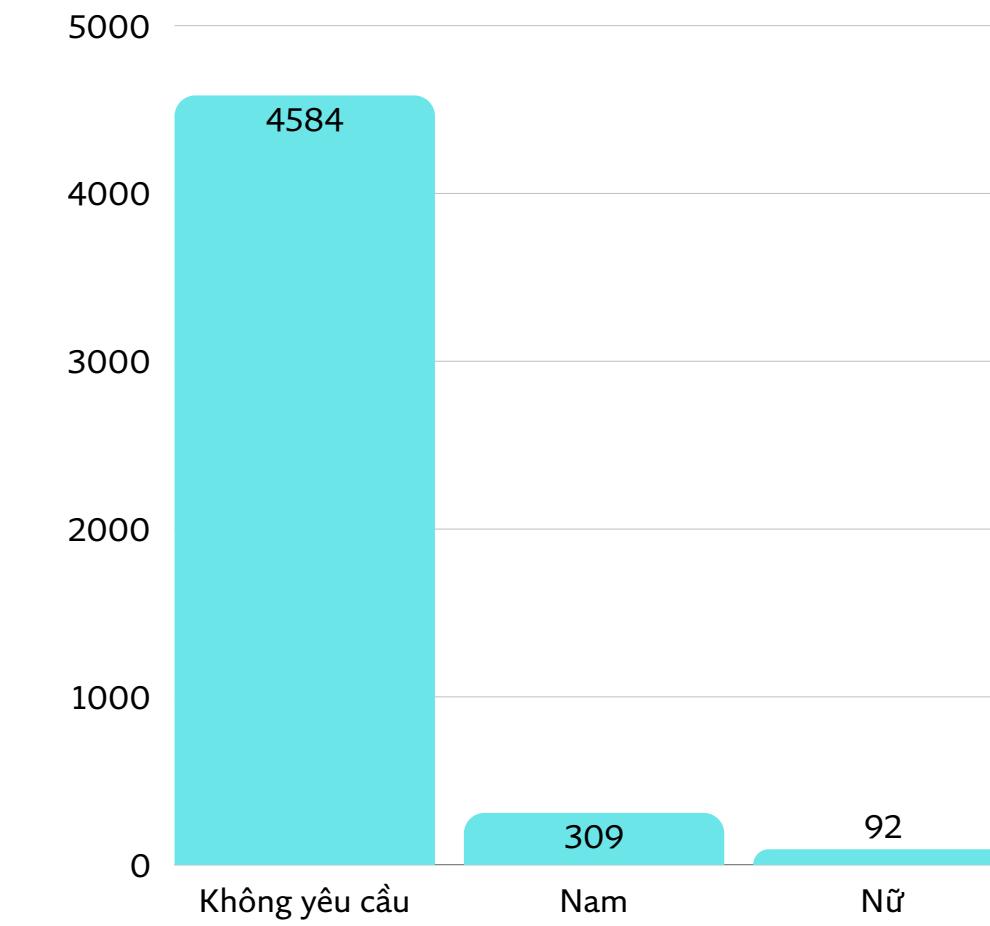
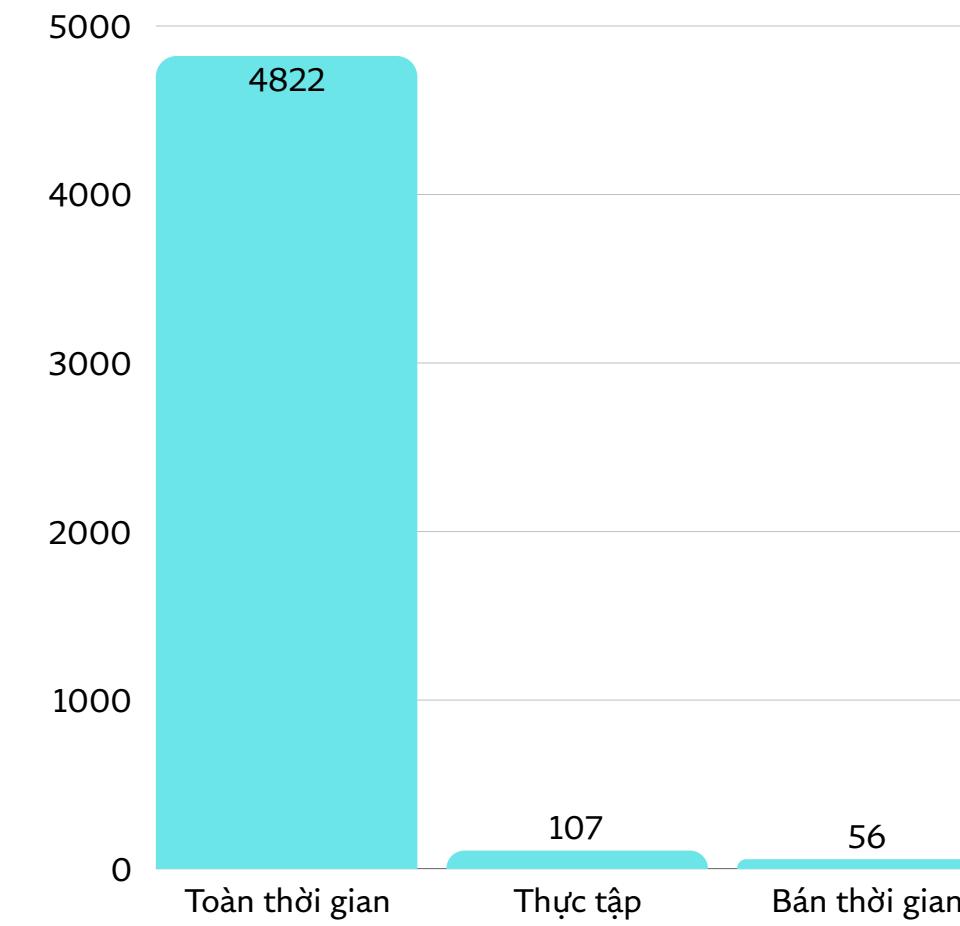
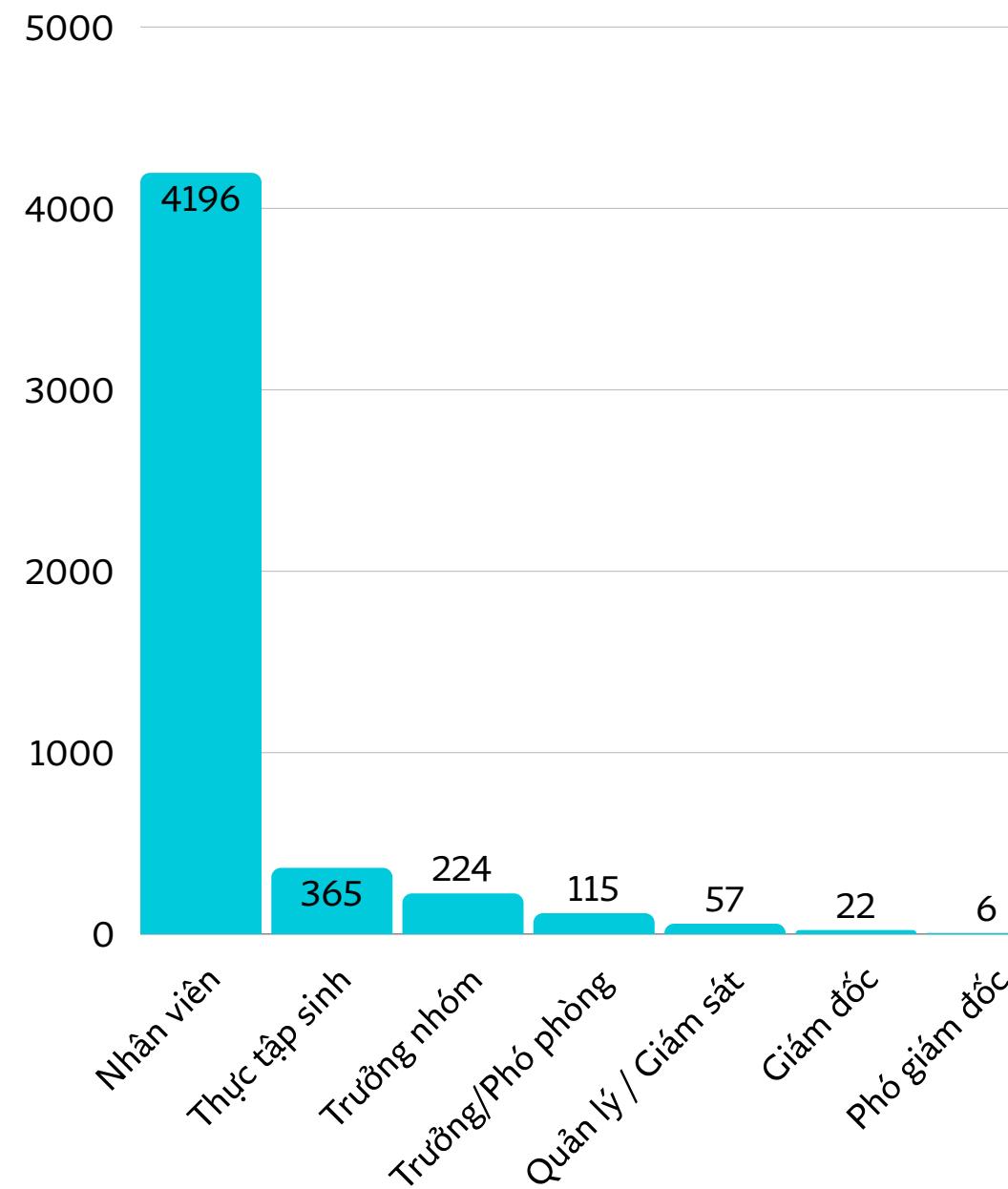
# EDA



- IT's salary vary more than the others -> new sector and higher salary potential
- Marketing's salary vary less -> mature sector and lower salary potential

# EDA - IT detail job

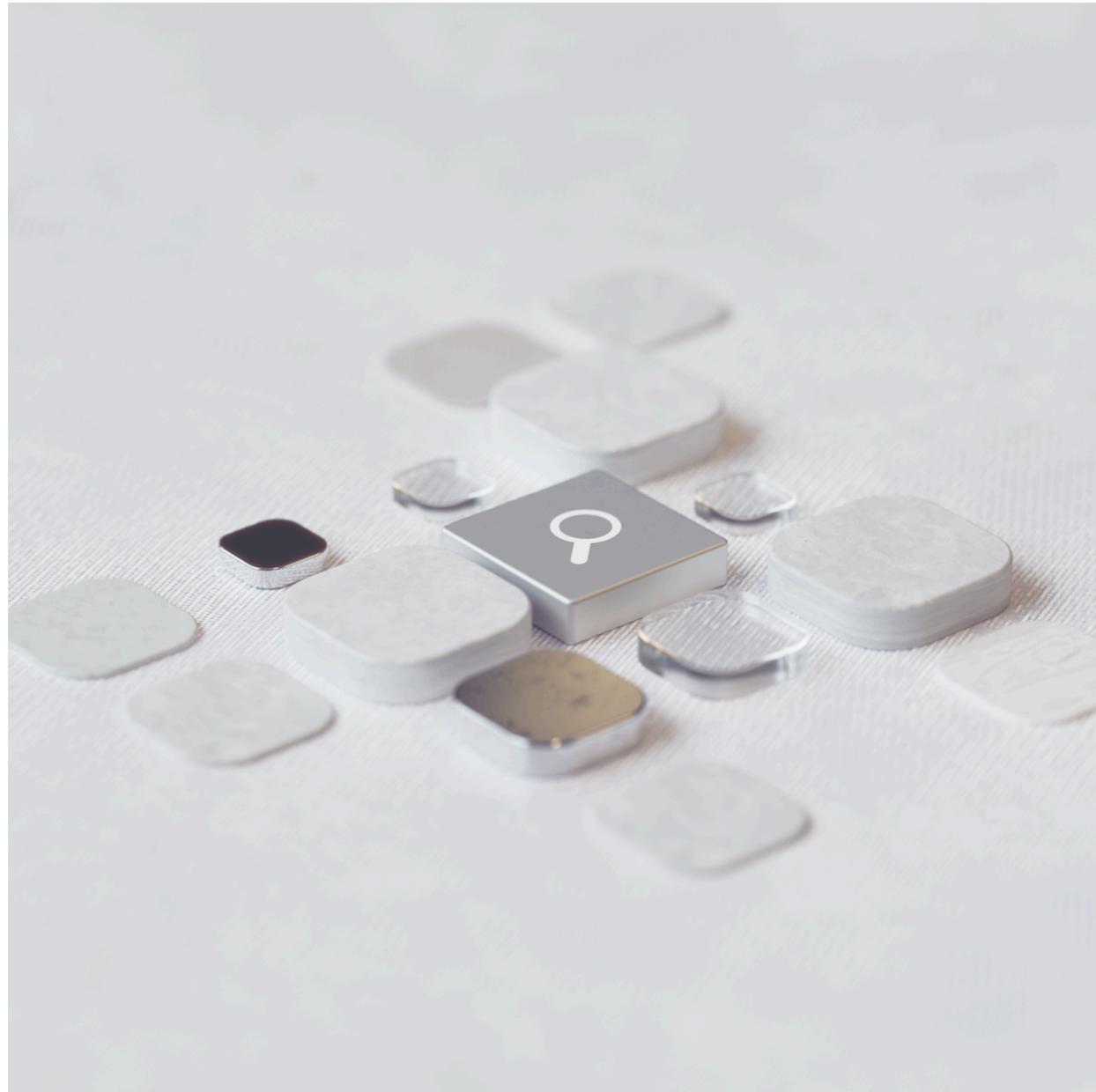
- Total 4985 jobs - with total 11065 number of recruits



- Most jobs is looking for full-time employees and do not require a specific gender

4

# MODEL

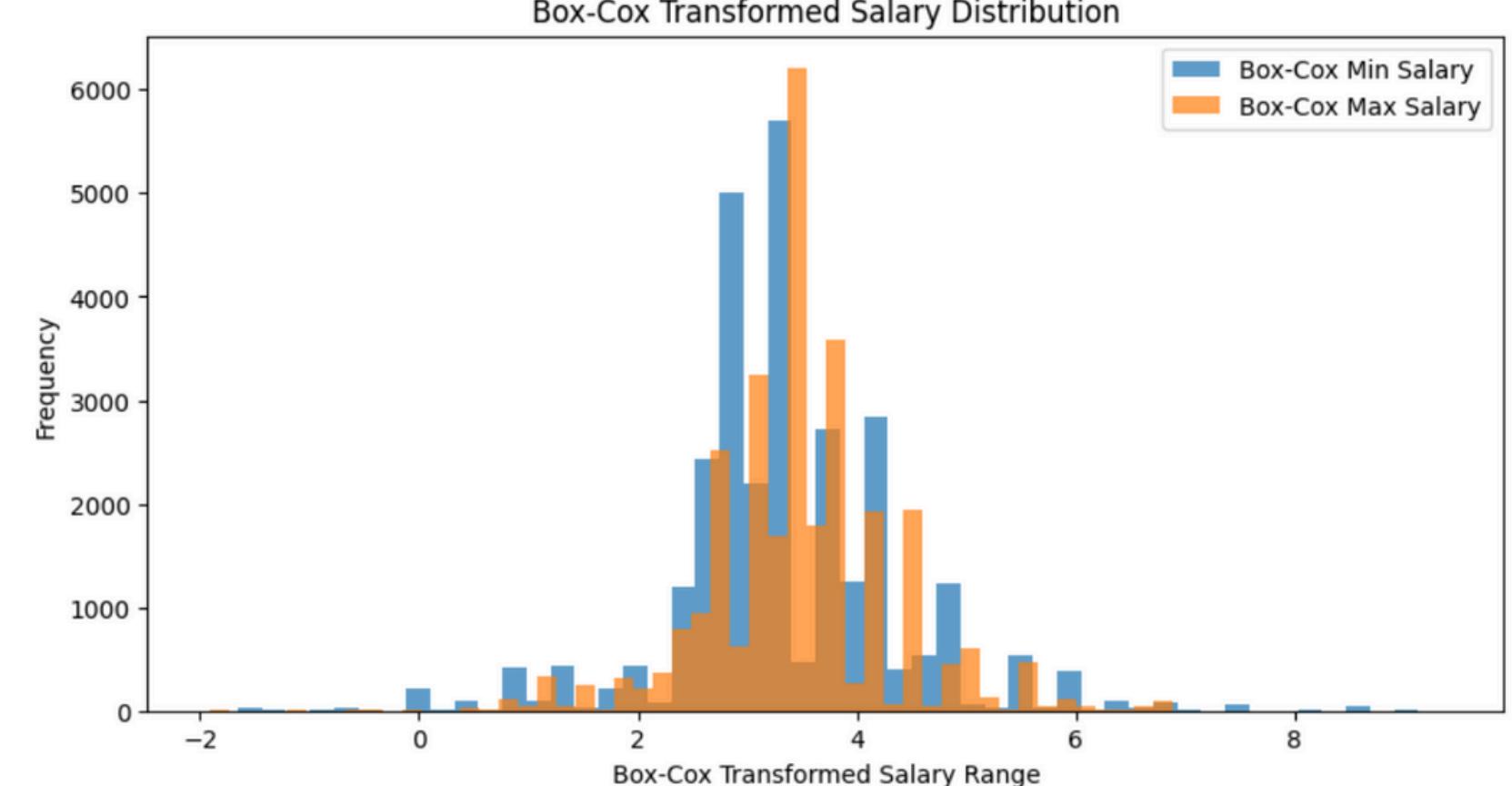
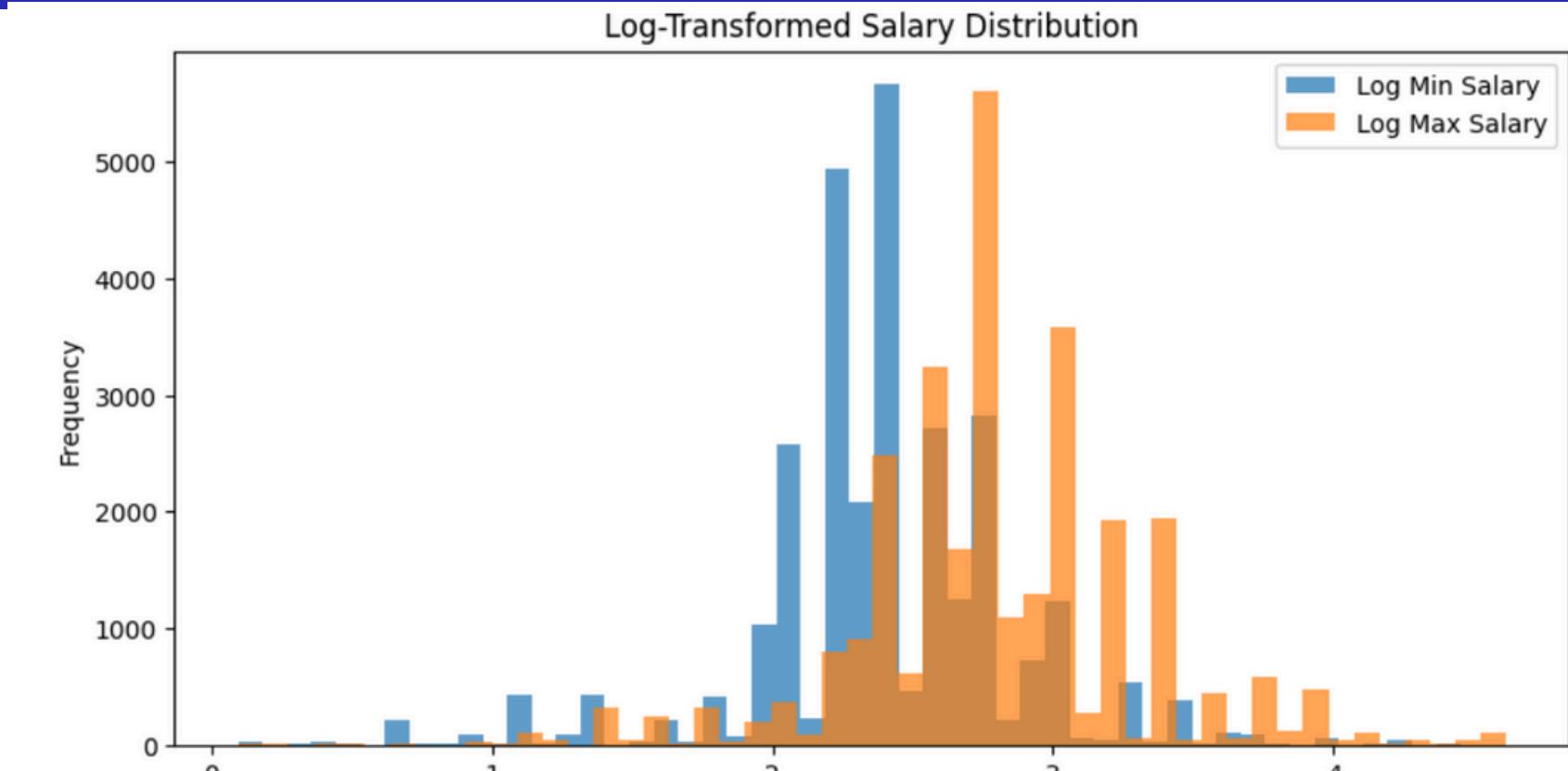
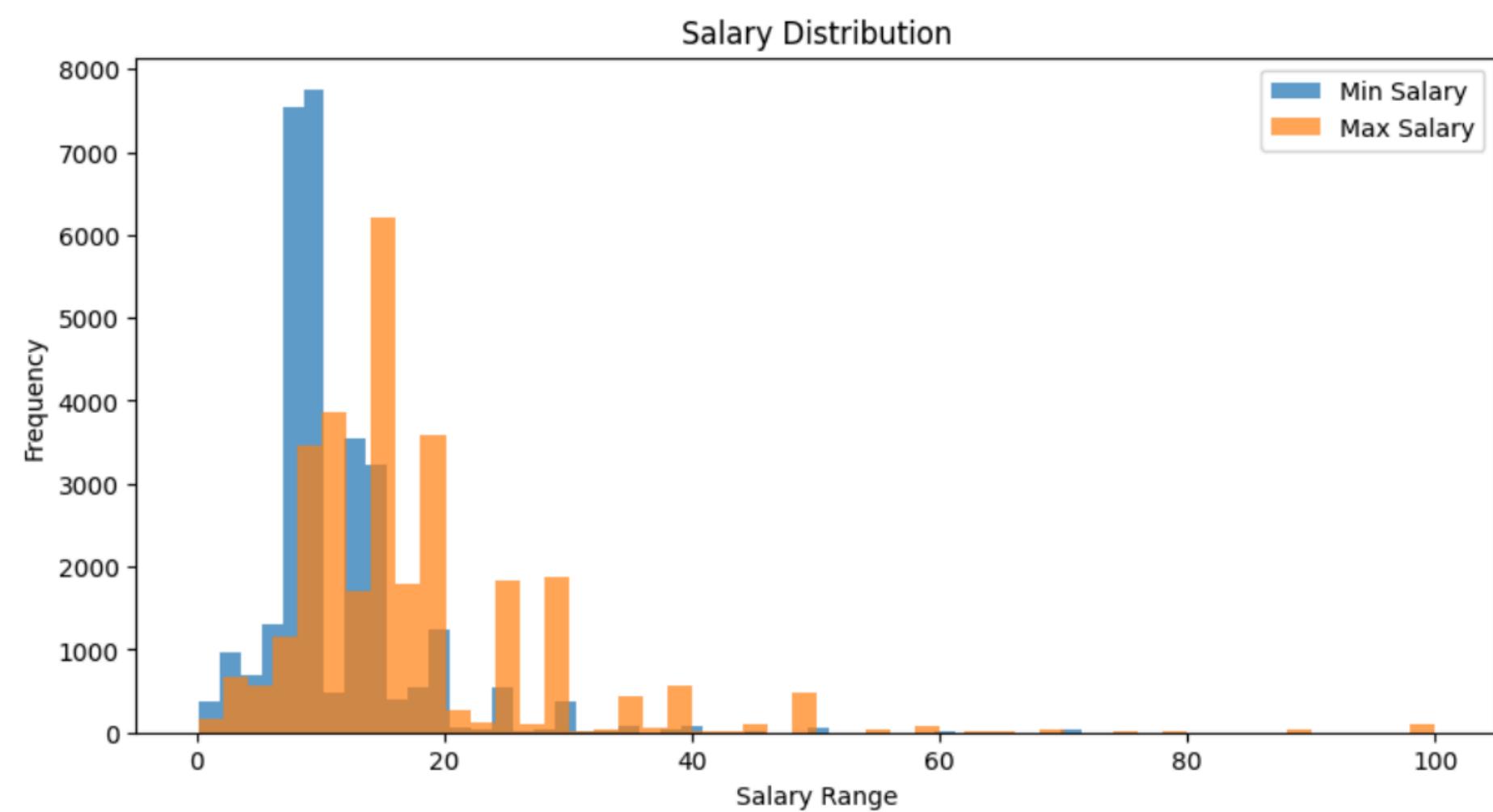


# Salary prediction model

- Build models to predict min salary and max salary for a job posting based on features such as **location**, **job category**, **experience** and skills.

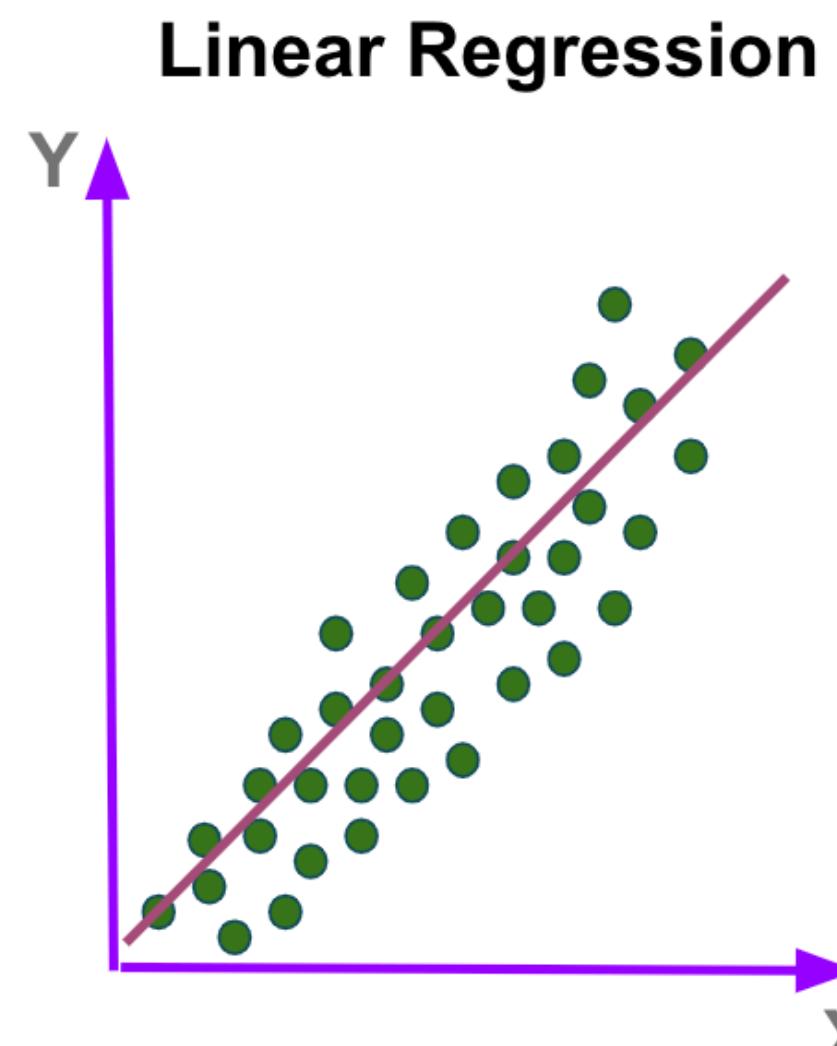


# Salary transformation



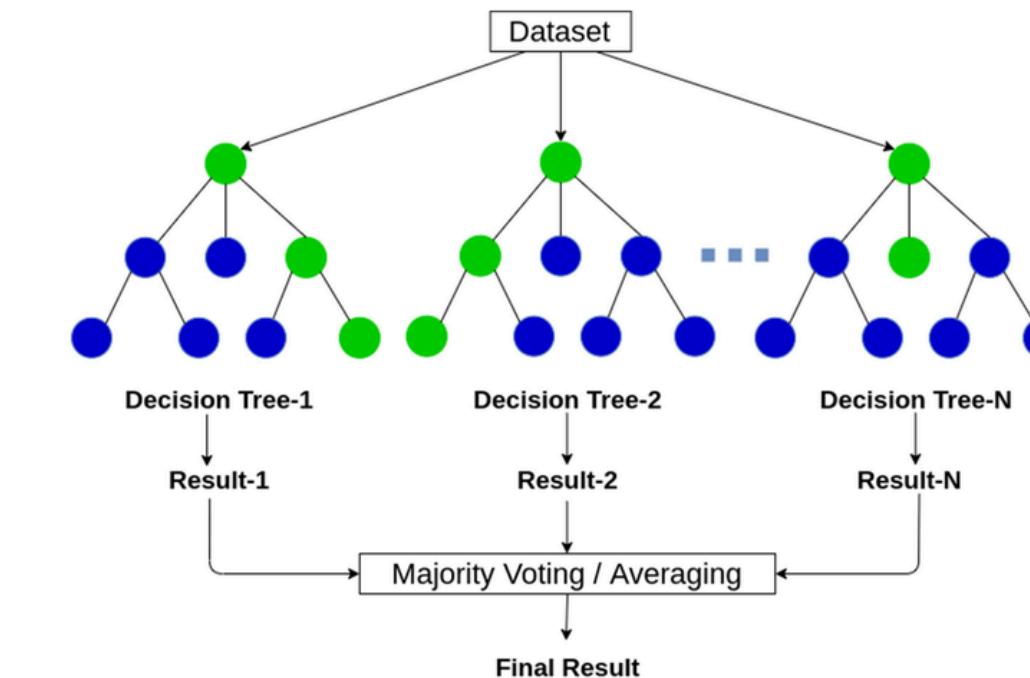
# Salary prediction model

- Machine Learning Model: **Linear Regression** and **RandomForest**.



- Metrics: **MAE**

## Random Forest



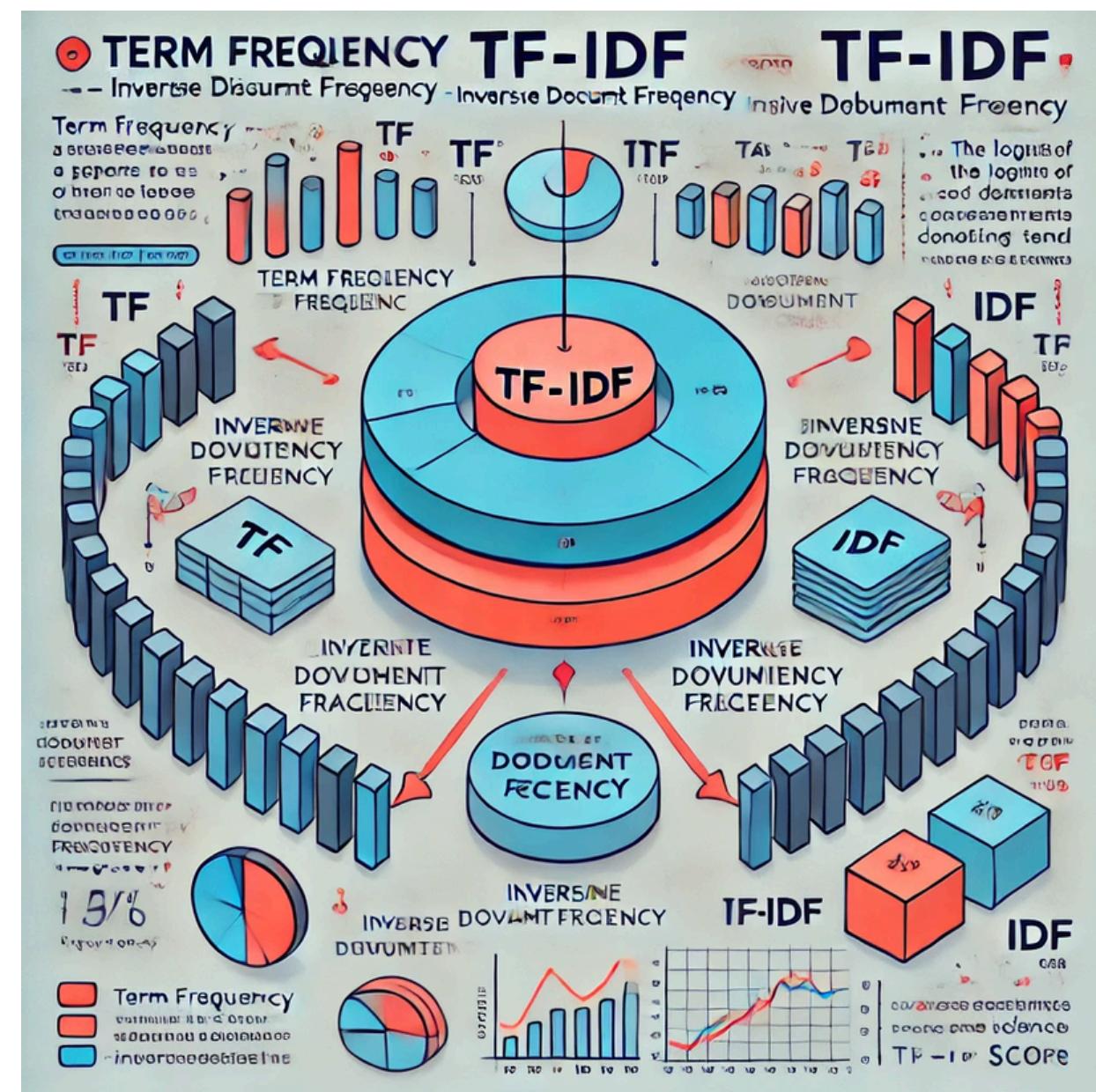
# Experiments on salary prediction model

Model	Transformation	Min Salary	Max Salary
Linear Regression	Original	3.14	6.19
	log	3.11	5.96
	boxcox	3.07	5.95
Random Forest	Original	3.08	6.13
	log	3.04	5.89
	boxcox	3.03	5.91

Results on salary prediction model (MAE)

# Job recommendation model

- Build a **IT job recommendation model** based on **TF-IDF vectorization**.
- Use information fields: a description, requirements, and benefits.



# TF-IDF vectorization

- The **TF-IDF vectorization**: transforms text data into a vector based on word's importance

**Term Frequency X Inverse Document Frequency**

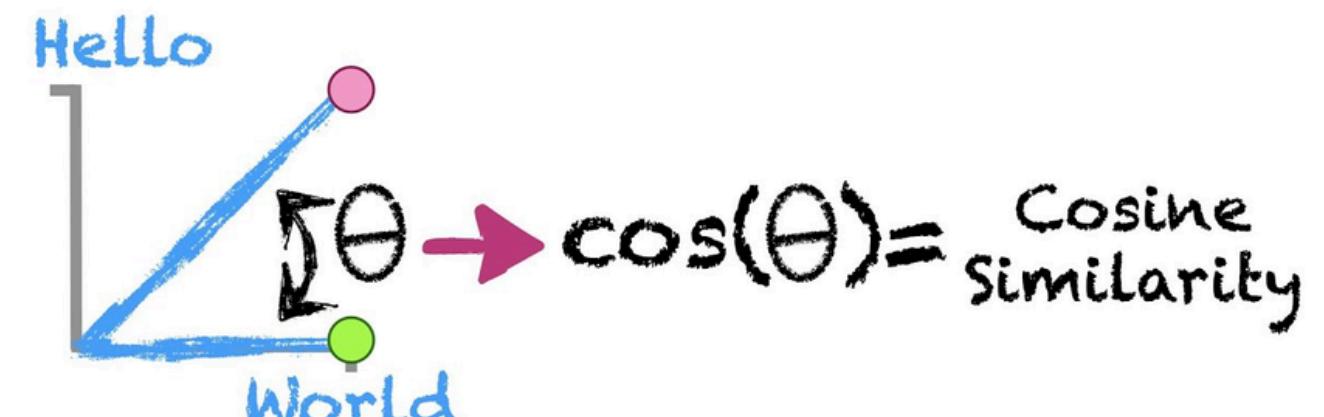
$$w_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

**Text1:** Basic Linux Commands for Data Science

**Text2:** Essential DVC Commands for Data Science

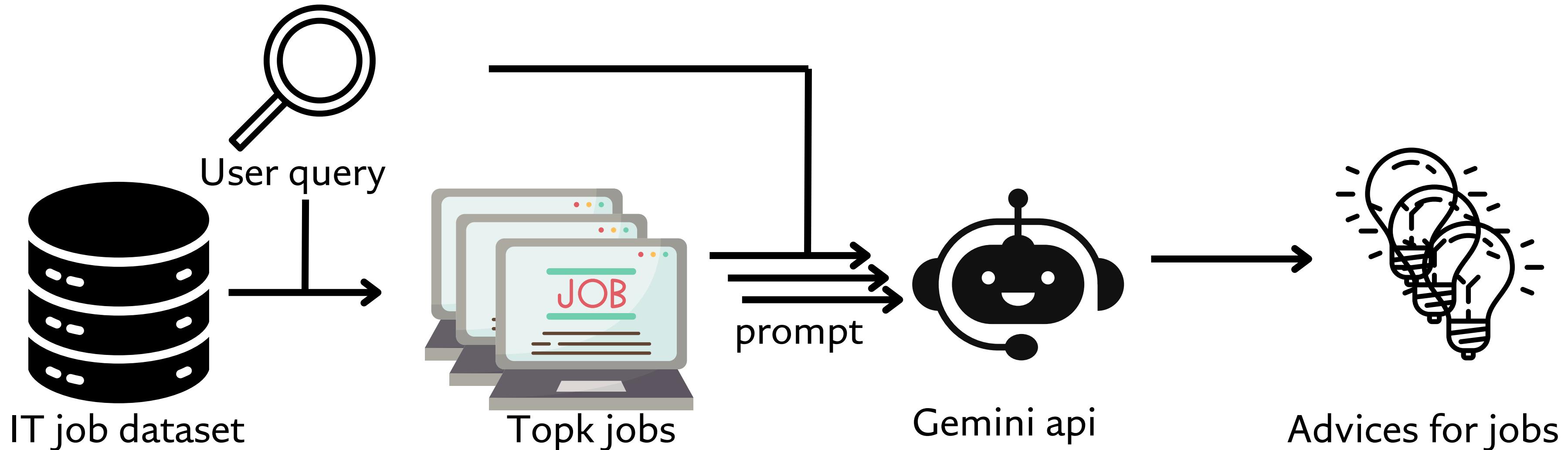
	basic	commands	data	dvc	essential	for	linux	science
Text 1	0.5	0.35	0.35	0.0	0.0	0.35	0.5	0.35
Text 2	0.0	0.35	0.35	0.5	0.5	0.35	0.0	0.35

**Cosine Similarity...**



**...Clearly Explained!!!**

# Pipeline



```
prompt = f"""
    Base on job requirements and my experiences, give an usefull
    advice, what I need to improve to meet job requirements.\njob
    requirements: {job_requirements}.\nMy experiences:
    {experiences_query}.\nAdvice:
    ...
    """
```

5

# CONTRIBUTION FUTURE WORK



# Contribution

- Explore the current job market.
- Explore salaries by industry and years of experience
- Recommend jobs that match the user's experience and preferences.
- Provide personalized advice for each job based on individual experience

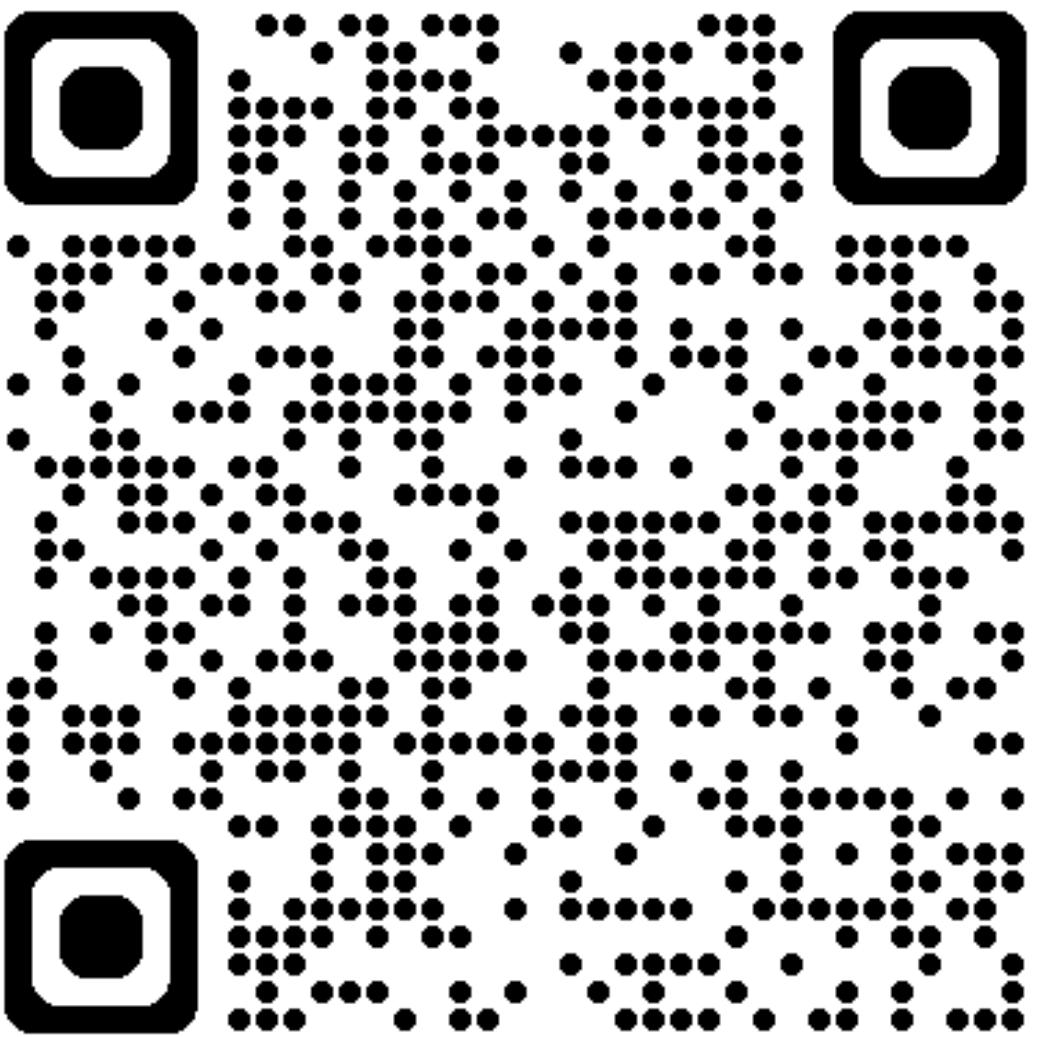
# Future work

- Update data weekly for the most accurate statistics and recommendations.
- Allow CV uploads
- Extract content from the CV, and recommend jobs without requiring manual input for each field

6

# DEMO

job analysis recommendation



# Thank you!