

Exploring the Thai Job Market Through the Lens of Natural Language Processing and Machine Learning

PIER Workshop

June 28th , 2024

Nuttapol Lertmethaphat
Nuarpear Lekfuangfu
Pucktada Treeratpituk

What We Have: Our Scraped Data

- 2 Large Online Job Markets
- Weekly scraped from 2020Q4 – 2023Q3

Demand-side

Over 1.2M+ job posts
Over 16,000 companies
Diverse job levels with from unskilled to c-level
Include salary, benefits, required education, experience, and others

Supply-side

Over 1.1M+ job seekers
Mostly first jobbers with experience less than 3 years
Aged 22-25
Bachelor's degree
61% are females
Includes expected salary & desired job fields

What We Have: Our Scraped Data

Demand-Side

company_name	job_title	contract_type	education_level	experience	salary_min	salary_max	age_max	working_hours	dayoff
กลุ่มบริษัทจันทฉฉฉฉฉ	R&DManager(FlexiblePackaging)ประจำโรงงานสินค้า	FULL_TIME	ปริญญาตรี - ปริญญาโท	5 - 5 ปีขึ้นไป	70000.0	130000.0	99.0	08:00 - 17:00	วันเสาร์
เซงเซงรวย	พนักงานPCออกบูชสินค้าสามารถออกต่างจังหวัดได้	FULL_TIME	ต่ำกว่า ม.6 ขึ้นไป	2ปีขึ้นไป	10500.0	17000.0	35.0	08:00 - 17:00	วันอาทิตย์
ทีเอสเอ็มกรุป(ประเทศไทย)จำกัด	พนักงานขายหน้าร้าน(แก้แ้วเจ็ตออกก้าลังกาย)เซ็นทรัล...	FULL_TIME	ม.6 - ปริญญาตรี	ไม่ระบุ	11500.0	25000.0	35.0	ไม่ระบุ	NaN
บริษัทมั่งคั่งผุดจำกัด	DigitalMarketing	FULL_TIME	ปริญญาตรี	3 - 5 ปีขึ้นไป	0.0	0.0	30.0	08:00 - 18:00	วันเสาร์
ฉัตรสยามลีและวัสดุภัณฑ์	พนักงานขับรถส่งของ(ปึกัฉพ)	FULL_TIME	ต่ำกว่า ม.6 - ปวส.	ไม่ระบุ	13000.0	15500.0	99.0	08:00 - 17:00	วันอาทิตย์

Supply-Side

gender	age	job_detail_position_1	expected_salary	education_level_1	education_institution_name_1	education_major_1	education_gpa_1	exp_position_1	exp_position_2	exp_position_3
ชาย	37.0	ล่ามภาษาจีน	30000	วรรณกรรมและวารสารศาสตร์ (College of Literature...	มหาวิทยาลัยฉฉฉฉ ประเทศจีน (Chongqing Univers...	วรรณศิลป์ (Literature and Arts - 文学)	3.66	เจ้าหน้าที่ประสานงานต่างประเทศ (ภาษาจีน)	ล่ามภาษาจีน-ไทย	เจ้าหน้าที่ประสานงานต่างประเทศ (ล่ามภาษาจีน-ไท...
หญิง	25.0	งานด้านเอกสารธุรการ	14000	ปริญญาตรี รัฐศาสตร์บัณฑิต	มหาวิทยาลัยสงขลานครินทร์	การปกครองท้องถิ่น	3.78	ลูกจ้างเหมาบริการพัฒนาตำบลแบบบูรณาการ	ลูกจ้างโครงการอว.สร้างงานเฟส2	ลูกจ้างโครงการบัตรสวัสดิการแห่งรัฐ 2565
ชาย	34.0	ควบคุมระบบบำบัดน้ำเสีย	15000	วท.บ วิทยาศาสตร์สิ่งแวดล้อม	มหาวิทยาลัยเกษตรศาสตร์	วิทยาศาสตร์สิ่งแวดล้อม	2.19	เจ้าหน้าที่คลังสินค้า	ควบคุมคุณภาพคอนกรีต	นักวิทยาศาสตร์สิ่งแวดล้อม
หญิง	33.0	ผู้ช่วยผู้จัดการบริษัท	19000	ปริญญาตรี	University of Bahrain	วิศวกรรมเคมี	2.25	ช่วยงานในตำแหน่งบรรณารักษ์ห้องสมุดวิทยาลัยเทคน...	ครูติว O-net เคมีให้กับนักเรียนชั้นมัธยมศึกษาป...	เจ้าหน้าที่ธุรการ ผู้ช่วยตำแหน่งศึกษานิเทศก์ภาษา...
หญิง	29.0	พนักงานให้บริการ	12000	ป.6	โรงเรียนหนองบัว	ทั่วไป	4.00	เซลล์ขายฝ่ายประมวลผล	พนักงานเสิร์ฟฟงดบิลที่เลี้ยงหน้าครัว	ฝ่ายผลิต

What We Want: Our Objectives

Categorizing the scraped data by occupations so we can ...

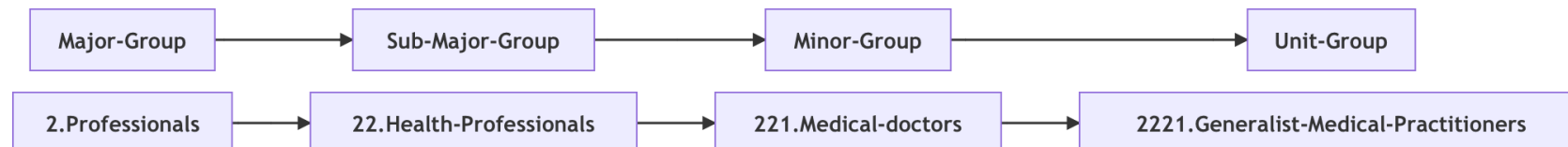
- Explore the untapped data
- Utilize *Volume & Velocity*
- More timely than the survey-based ones
- Beveridge Curve



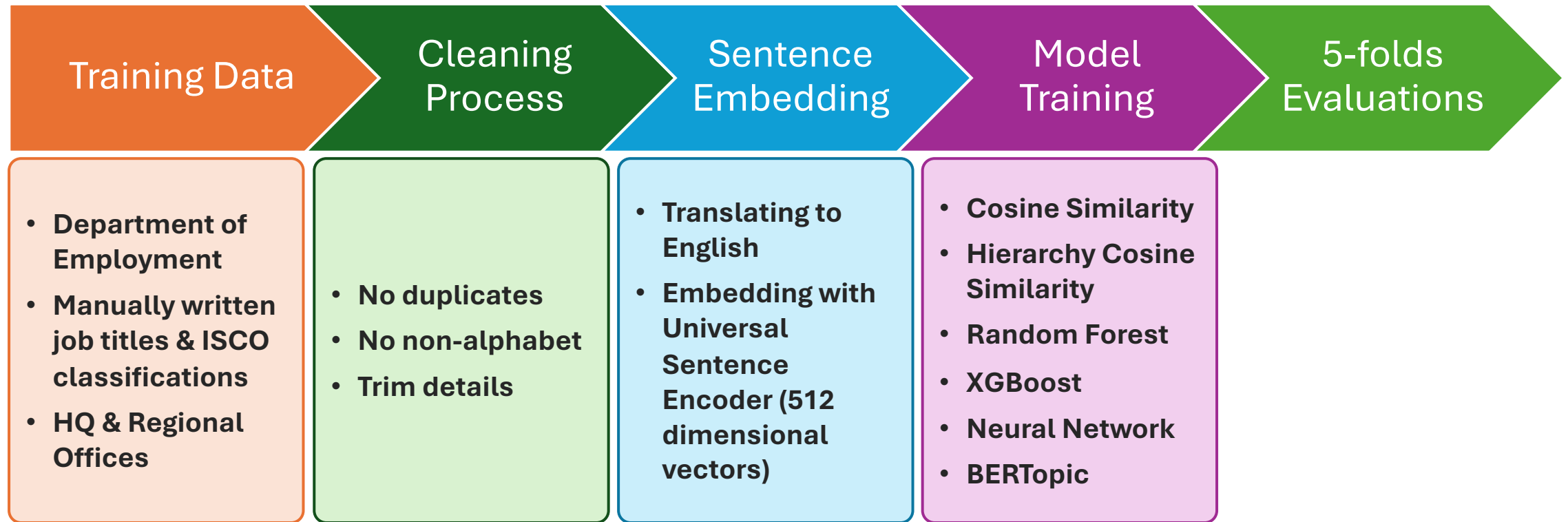
How We Want: ISCO Code

Major Groups	Sub Major Groups	Minor Groups	Unit Groups	Skill level
1 Managers	4	11	31	3 and 4
2 Professionals	6	27	92	4
3 Technicians and Associate Professionals	5	20	84	3
4 Clerical Support Workers	4	8	29	2
5 Service and Sales Workers	4	13	40	2
6 Skilled Agricultural, Forestry and Fishery Workers	3	9	18	2
7 Craft and Related Trades Workers	5	14	66	2
8 Plant and Machine Operators, and Assemblers	3	14	40	2
9 Elementary Occupations	6	11	33	2
0 Armed Forces Occupations	3	3	3	1, 2 and 4
Total number of groups	43	130	436	

Example:



What We Do: Our Methodology



Sentence Embedders

วิศวกรออกแบบเครื่องกล,
วิศวกรออกแบบระบบปรับอากาศ,
ภูมิสถาปนิก,
Architect/สถาปัตย์

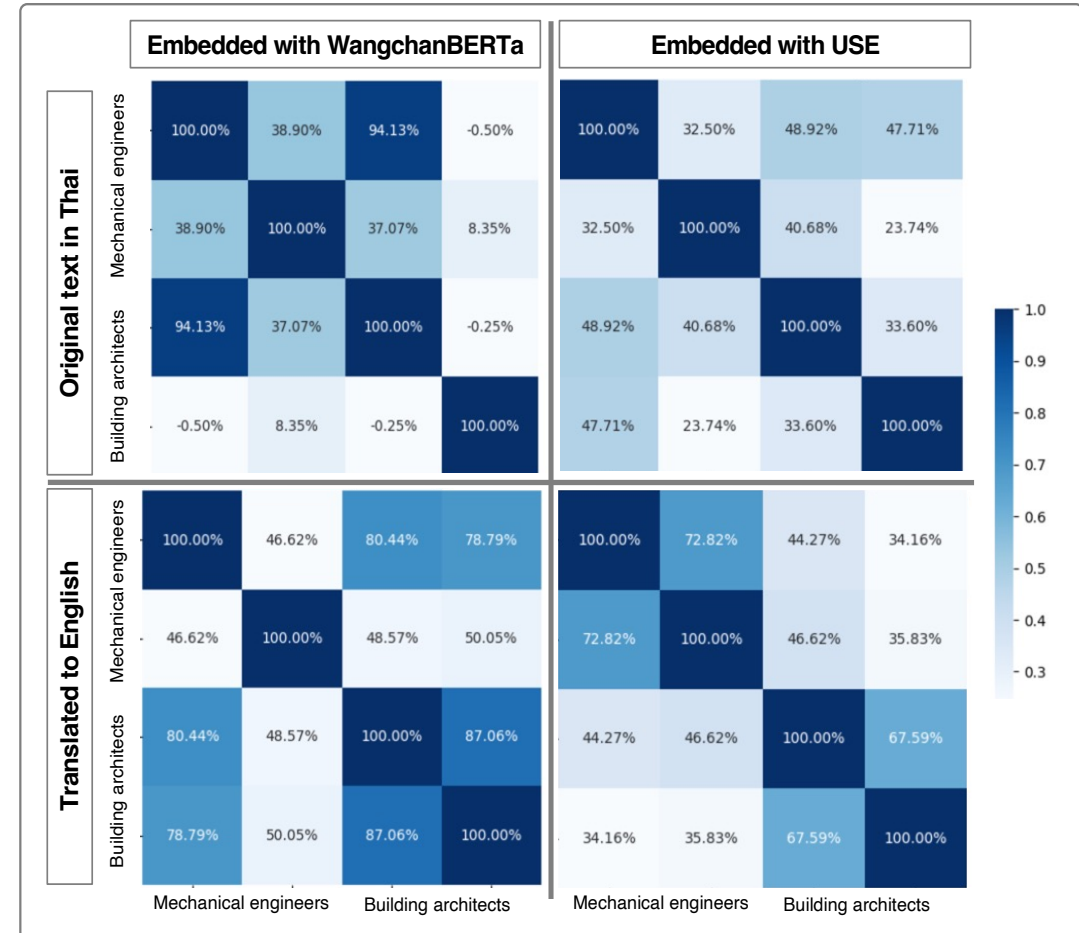
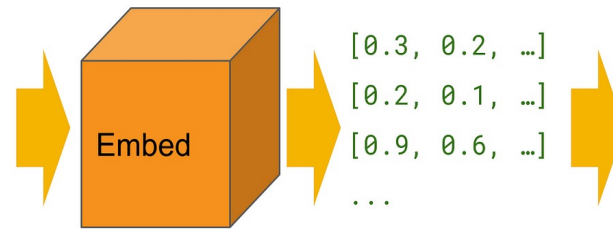
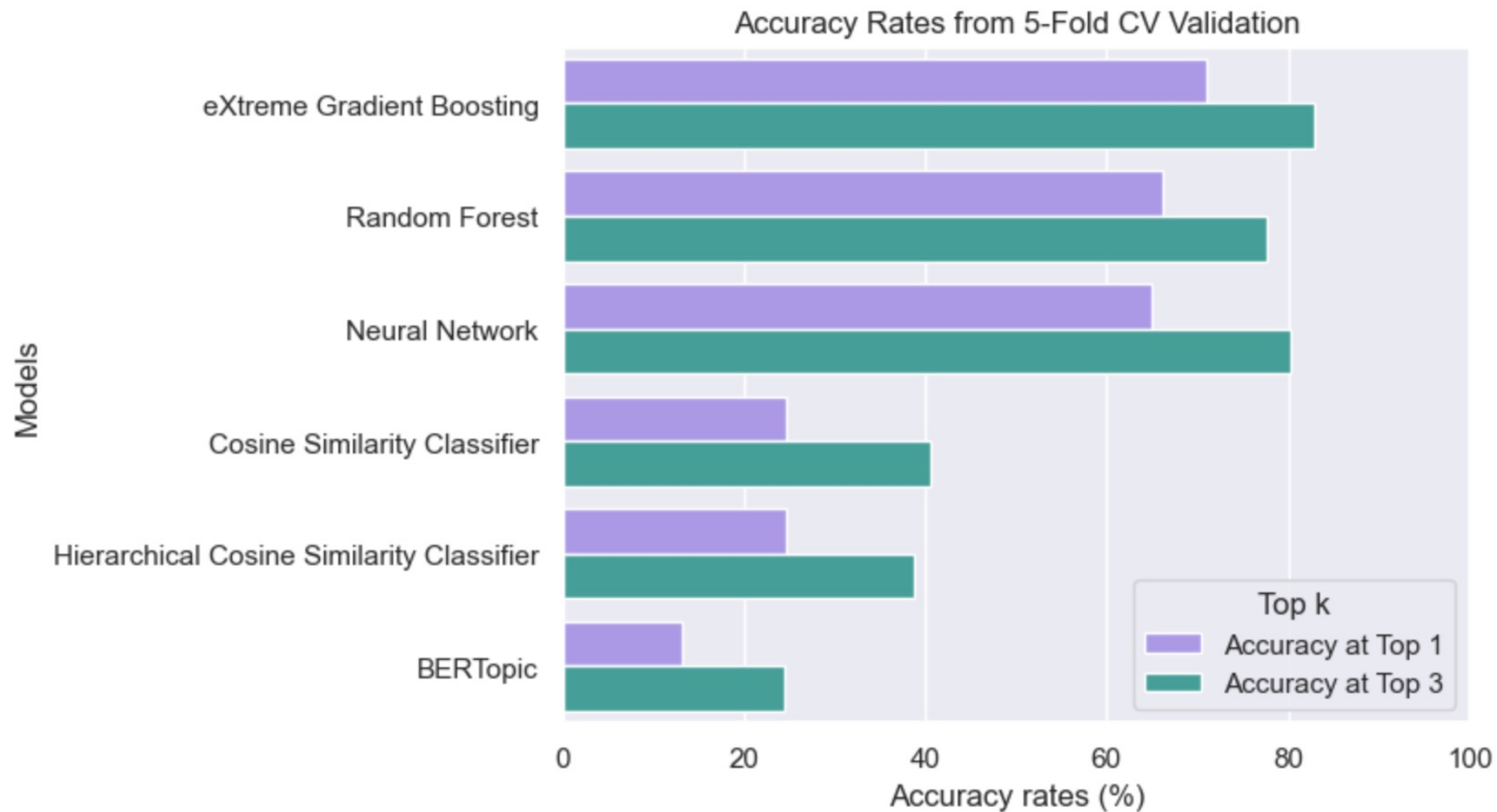
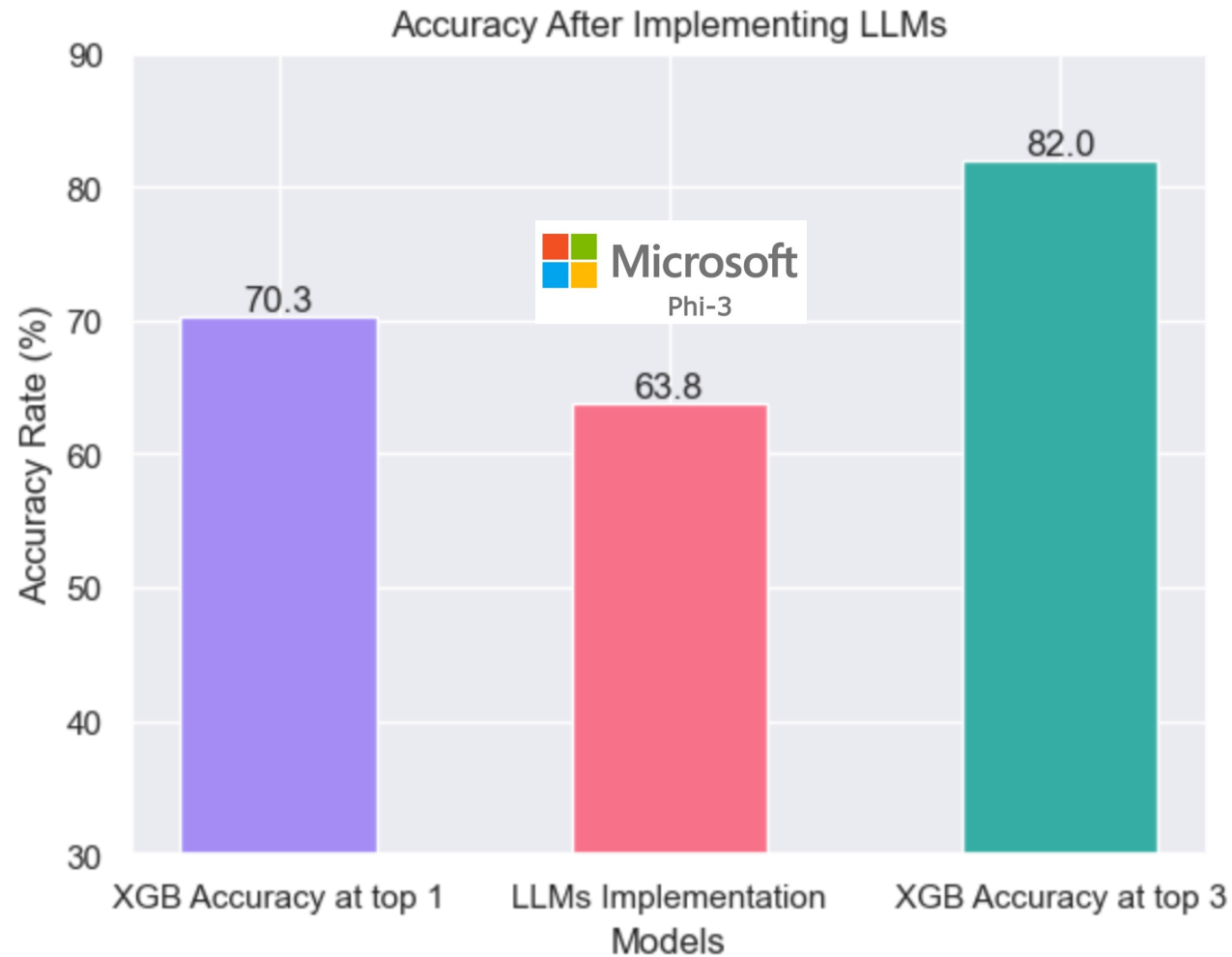


Figure 1: Cosine Similarity Matrices at Stage One

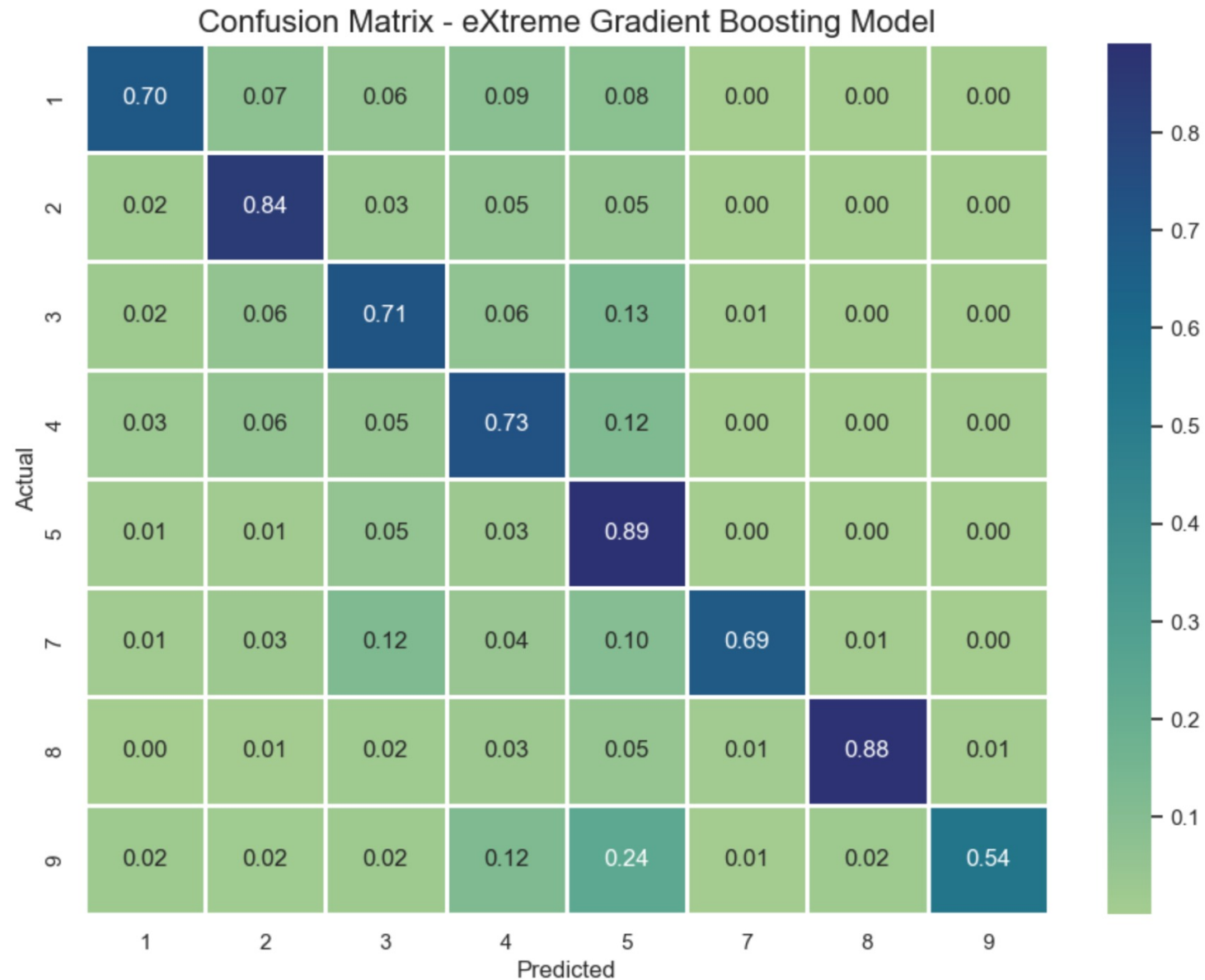
Results from 5-fold cross validation



Improving the Results With LLMs



Confusion Matrix



Example of Model's Confusions



Between Major Grounds

text	actual	actual_title	pred	actual_pred
Building Manager	1219	Business services and administration managers ...	5153	Building caretakers
Housekeeper and Gardener	3115	Mechanical engineering technicians	5152	Domestic housekeepers
Cook, Thai, Chinese food	9629	Elementary workers not elsewhere classified	5120	Cooks

Accounting Jobs' Hierarchy

text	actual	actual_title	pred	actual_pred
Accounts officer	4419	Clerical support workers not elsewhere classified	2411	Accountants
Accounting Officer	3313	Accounting associate professionals	2411	Accountants
Accounting officer for receiving and disbursement	3313	Accounting associate professionals	2411	Accountants
Accountant and Budget Officer	5244	Contact centre salespersons	3313	Accounting associate professionals
Administrative accounting, Vibhavadi office	5244	Contact centre salespersons	4311	Accounting and bookkeeping clerks

Human Error Also Clouds The Metric

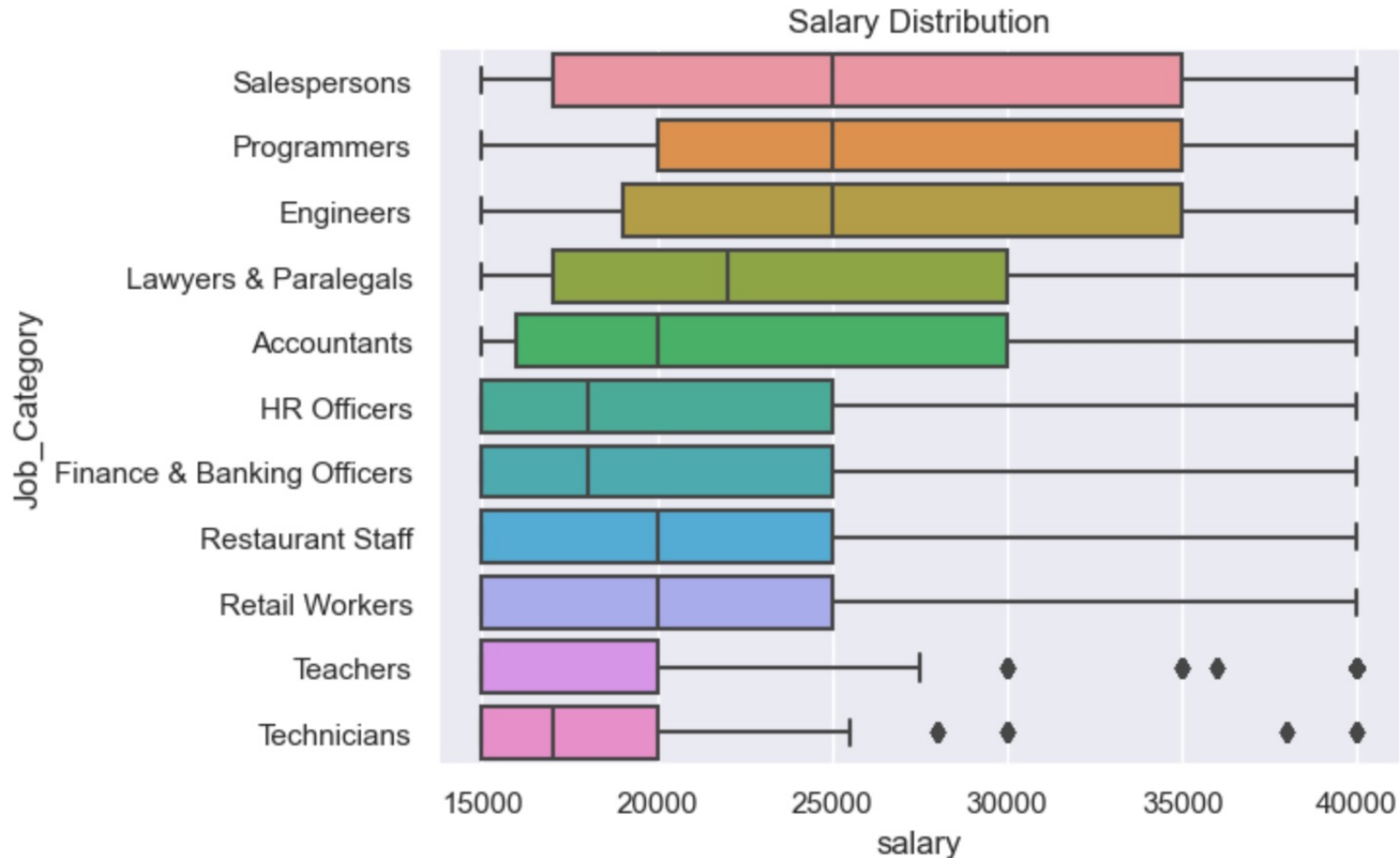


	text	actual	actual_title	pred	actual_pred
	Recruitment Support	3333	Employment agents and contractors	4416	Personnel clerks
	Spaghetti Shop Manager Factory	5244	Contact centre salespersons	1412	Restaurant managers
	Kitchen Manager	4416	Personnel clerks	5120	Cooks
	Property Appraisal Officer	5244	Contact centre salespersons	3315	Valuers and loss assessors
	Human resource management and general management	4416	Personnel clerks	1212	Human resource managers
	Accounting Officer Account Payable	4311	Accounting and bookkeeping clerks	2411	Accountants
	Social Media Content Editor Officer Knowledgea...	3322	Commercial sales representatives	2166	Graphic and multimedia designers
	Foreman Concrete slabs Ready made	3341	Office supervisors	3123	Construction supervisors
	Nanny, semi-housekeeper	5131	Waiters	5311	Child care workers
	Machine Repair Technician	7231	Motor vehicle mechanics and repairers	3115	Mechanical engineering technicians

Stylized Facts from Scraped Data

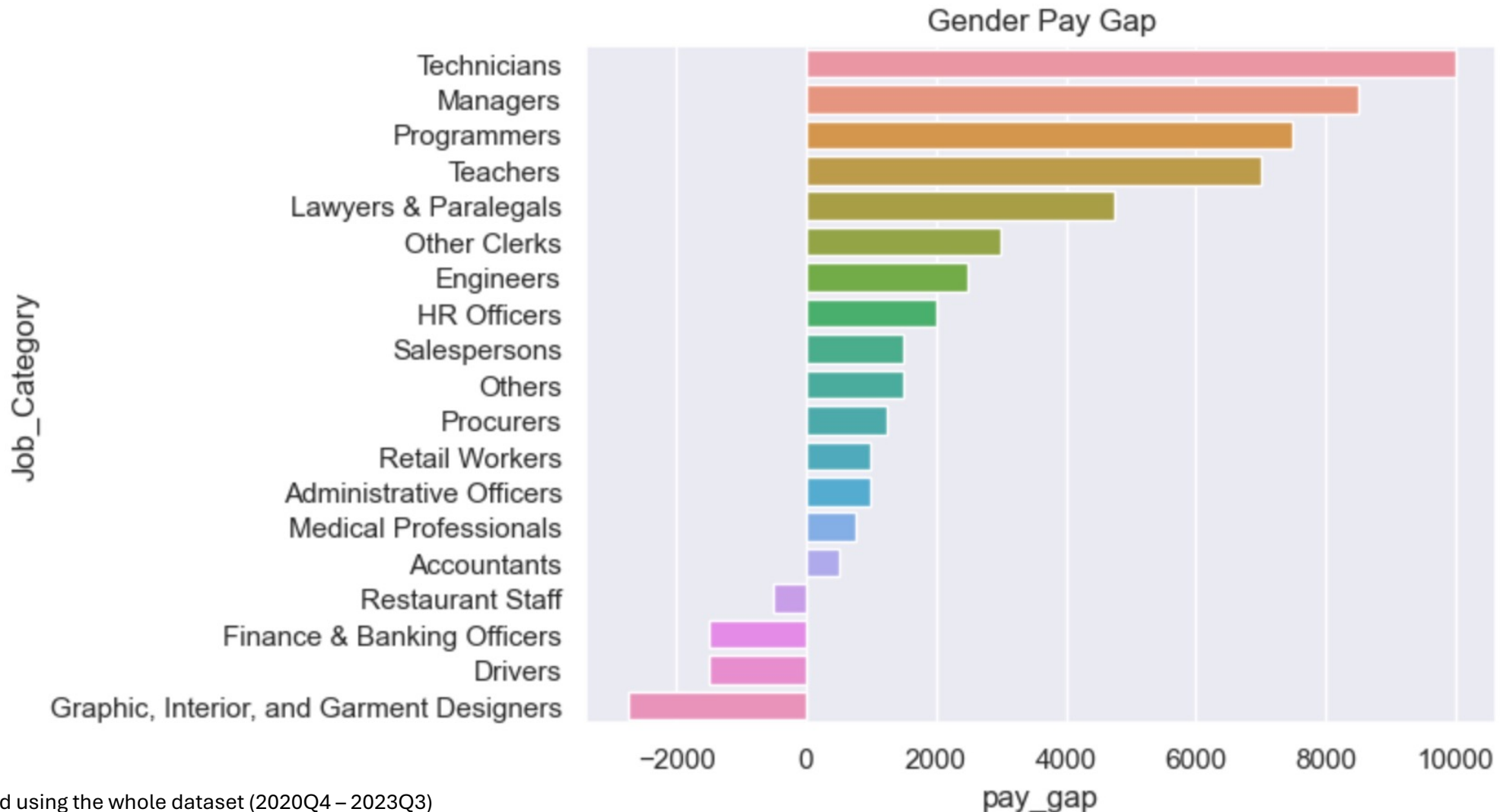
Stylized Facts 1: Salary Can Vary Largely in Some Fields

Pay Variation in Major Job Categories



Stylized Facts 2: Gender Pay Gap in Wage Offer Persists Across Industries

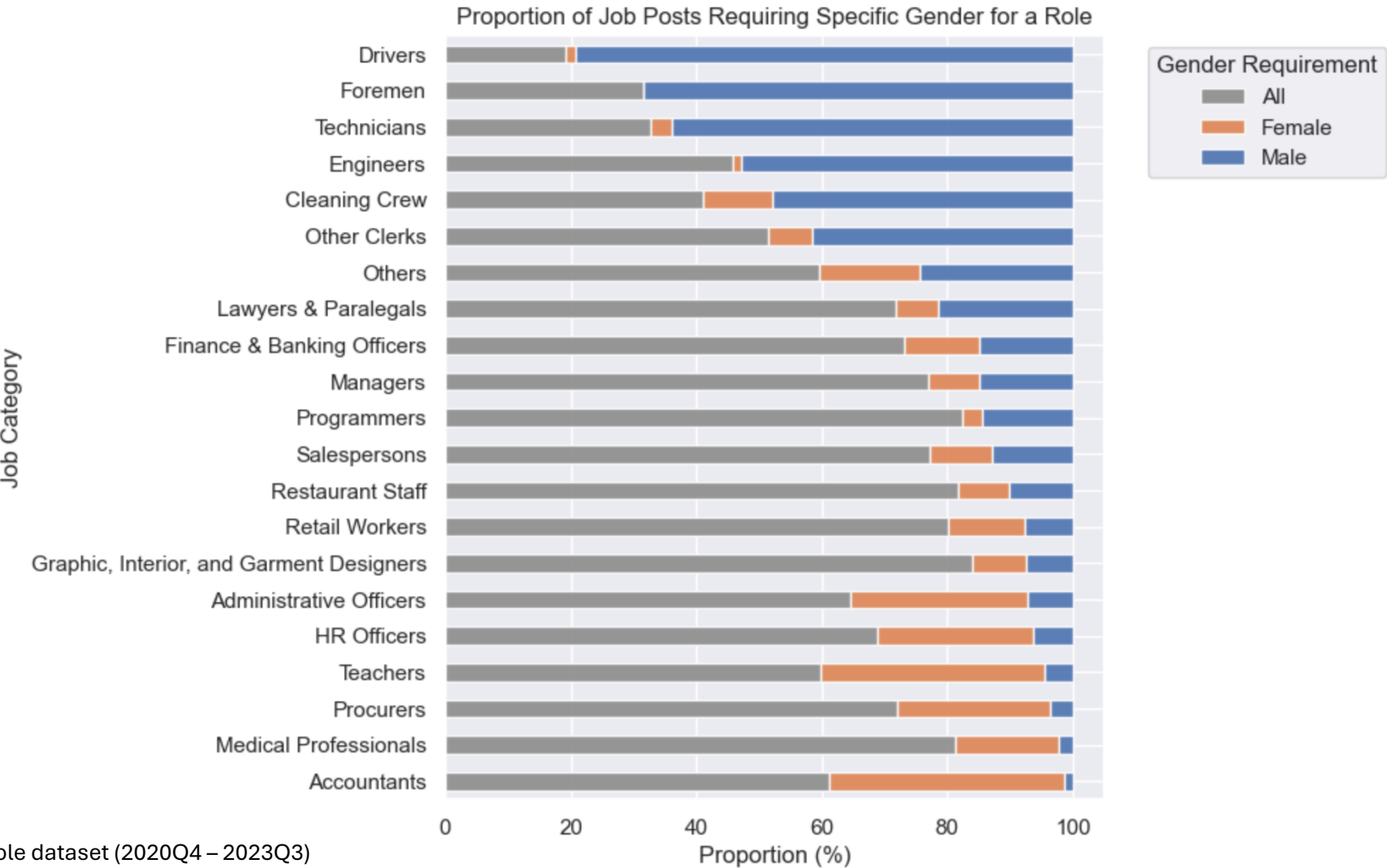
Men still get generally paid higher in many major occupations



Stylized Facts 3: Gender Discrimination in Job Ads

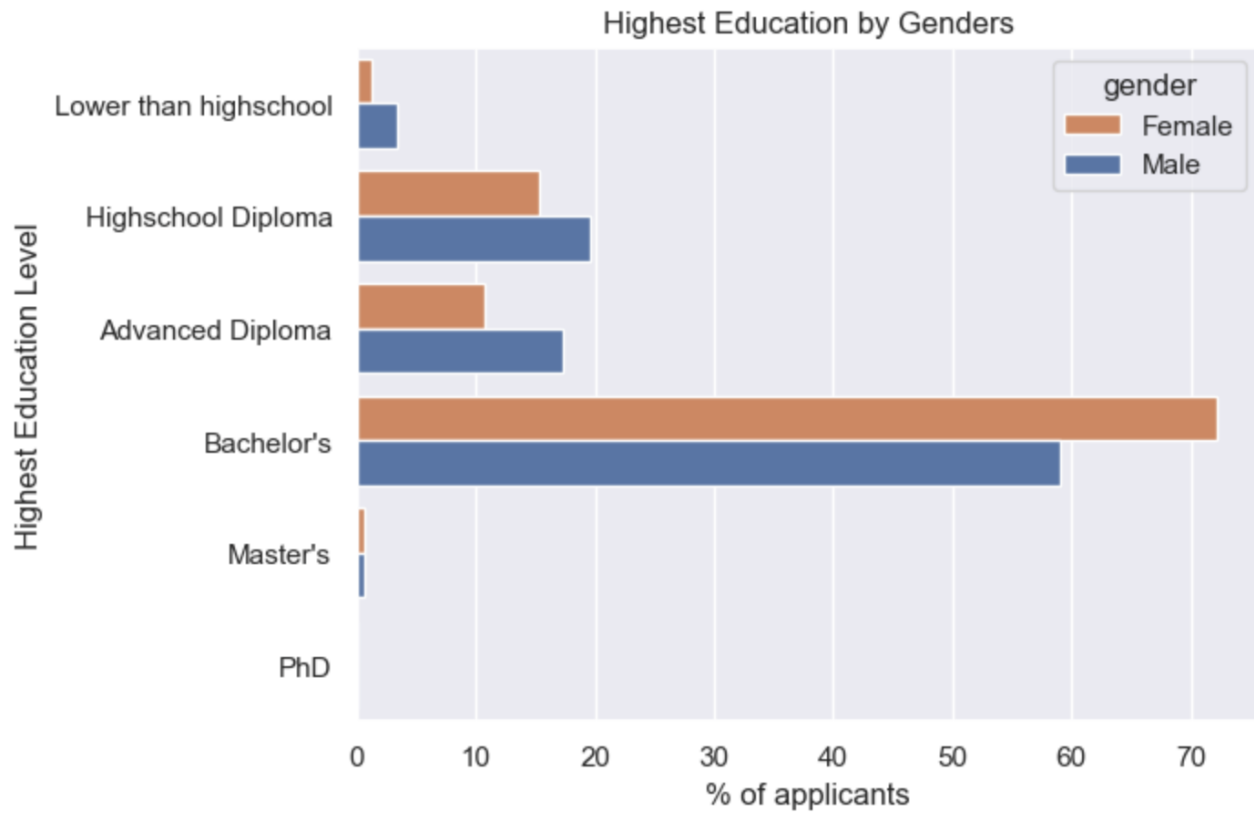
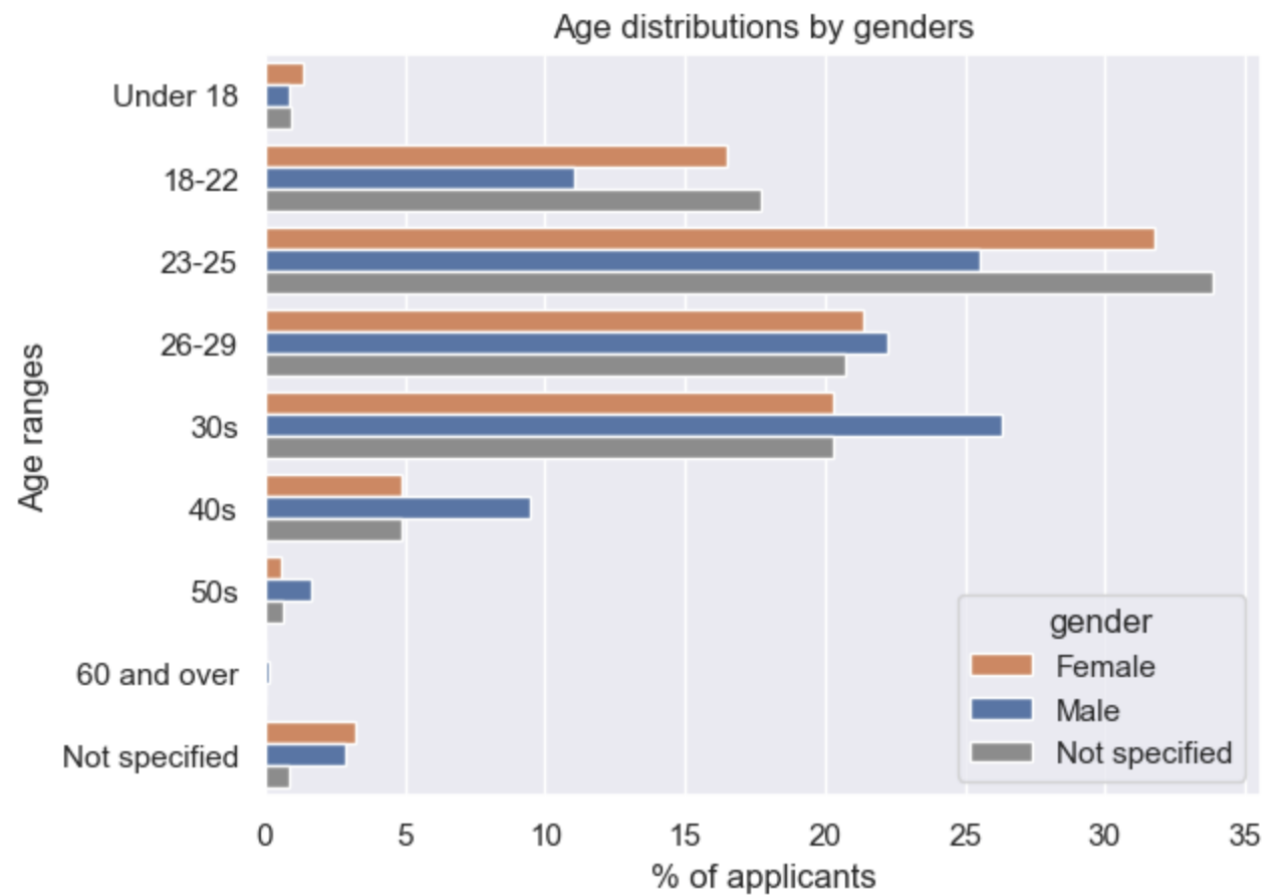
Persists Across Industries

Many fields prefer men over women as specifically required in job ads

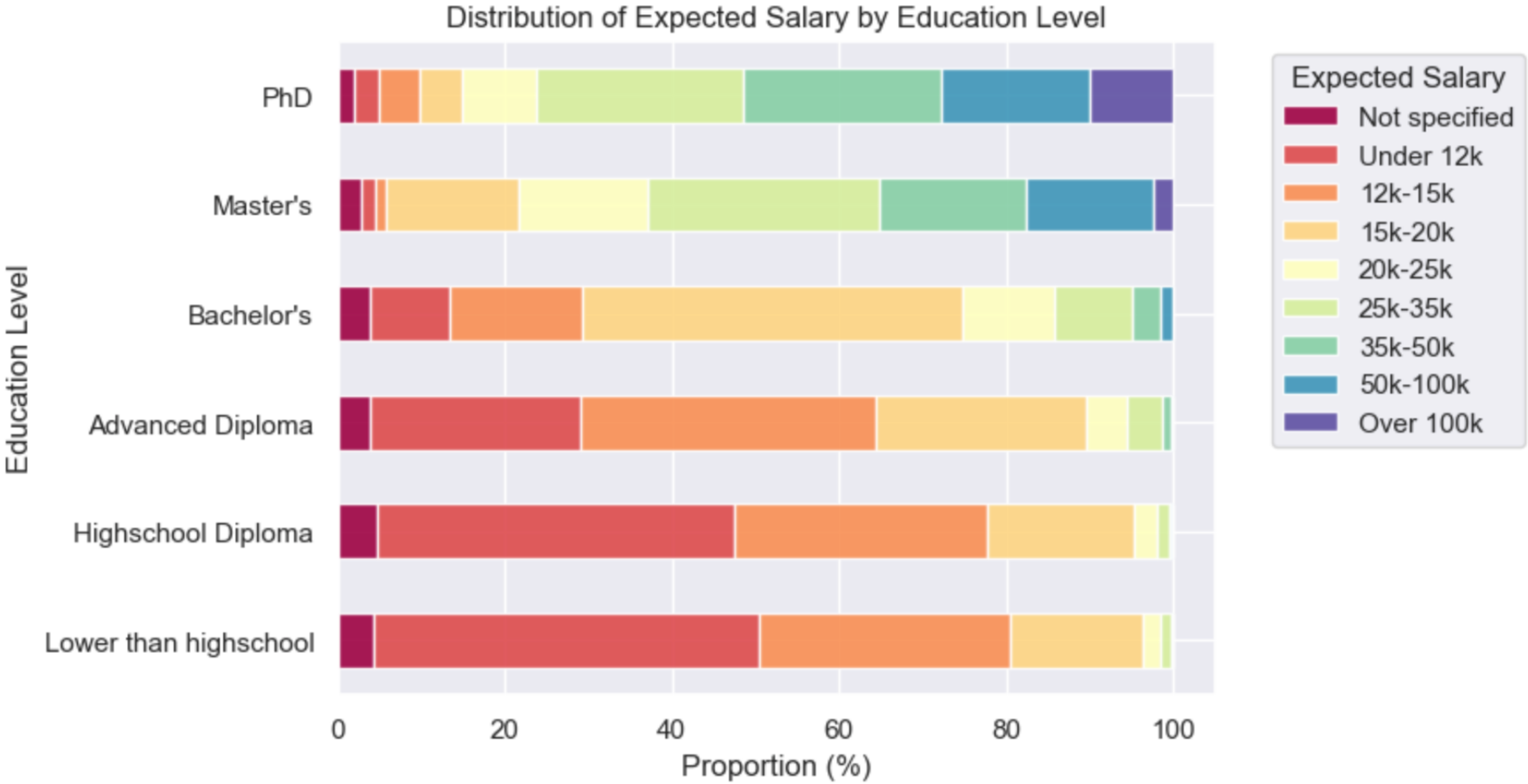


Note: Calculated using the whole dataset (2020Q4 – 2023Q3)

Ongoing Analysis: Resume Data



Ongoing Analysis: Resume Data



Conclusion



- Some Limitations Remain ...
 - Some occupations are naturally excluded from online job platforms (agriculture, factory workers, SMEs restaurant workers, etc.)
 - Difficult to know if they (Employee & -yer) match each other
 - Lack of transparency in compensation from employers
- What's Next?
 - Demand-supply analysis (Beveridge Curve)