Gender Bias in Online Job Screenings

Zoe, Maya, & Anh

- What did we explore?
  - Gender bias in online job screenings
  - Influence of gender-coded language
- Why choose this?
  - We are the demographic that may be affected negatively
    - Women in STEM
    - Generation of online job screenings

3 Research Questions & Expected Findings

Do male and female candidates with similar qualifications have different job screening outcomes?

Does gender bias in job screening vary across industries (specifically gender-dominated fields)?

Are gender-coded words (e.g., “collaborative” vs. “assertive”) in job descriptions associated with differences in the gender distribution of job applicants?

Data

- ➔ 600 scraped Resumes from Java Developers/Architects Resumes page
- ➔ Created a new CSV file organizing all of the resumes
  - ◆ Each resume entry contains text under sections such as Summary, Technical Skills, and Professional Experience
  - ◆ Separated data, assigned random values, and created new data
- ➔ We randomly assigned gender (male/female) to each entry
- ➔ Determined how many “gendered” words were in each resume

Ios Developer Resume

4.00 /5 (Submit Your Rating)

Colorado Springs, CO

SUMMARY

- Around 12+ years of experience in experience in the entire process of the software development life cycle (SDLC) including design, implementation, testing and maintenance.
- Broad experience in different technology platforms: Web and Client/Server, Databases, Client - Server applications, IOS application development
- Experience with IOS application development using IOS SDK(IPAD/IPhone), Objective-C, REST, JSON, SqlLite and Xcode.
- Experience using C, C++, Cocoa Touch - UIKit (ViewControllers, Tableviews, gestures, view elements, alerts etc), Core Services - CoreData.
- Experience in developing web applications using Ruby On Rails, PhoneGap, JavaScript, JQuery, JQuery Mobile, Bootstrap, HTML5, XML, CSS3, mysql, SqlLite
- Experience with multiple life cycle methodologies and design methods like Waterfall, Agile, Scrum and Sprint.
- Working experience in using RESTful web-services to provide connections to back end services and handling data using parsers with formats like JSON and XML.
- Experience in deploying application to Heroku
- Strong knowledge in Application Programming under Windows, UNIX and Linux environment
- Excellent exposure to Version Control Systems like Git (git flow), Svn
- Strong concepts and fundaments in Agile Methodology, Object Oriented Analysis and Design, Best Practices

TECHNICAL SKILLS

Operating Systems: Mac OS, AIX 5.3, Sun Solaris 5.8, OS/390, Windows XP/NT

Databases: MySQL, Oracle 10g/9i, DB2 UDB, SQL Server, IMS/DB, NSQL

Programming Languages: Ruby, ROR, C, C++, Java, Unix Shell Scripting (Korn), Perl, AWK, PL/SQL, XML, Cobol, JCL, PL/I, Focus

Web Technologies: HTML5, CSS3, JQuery, Apache, Tomcat, XML, Javasript

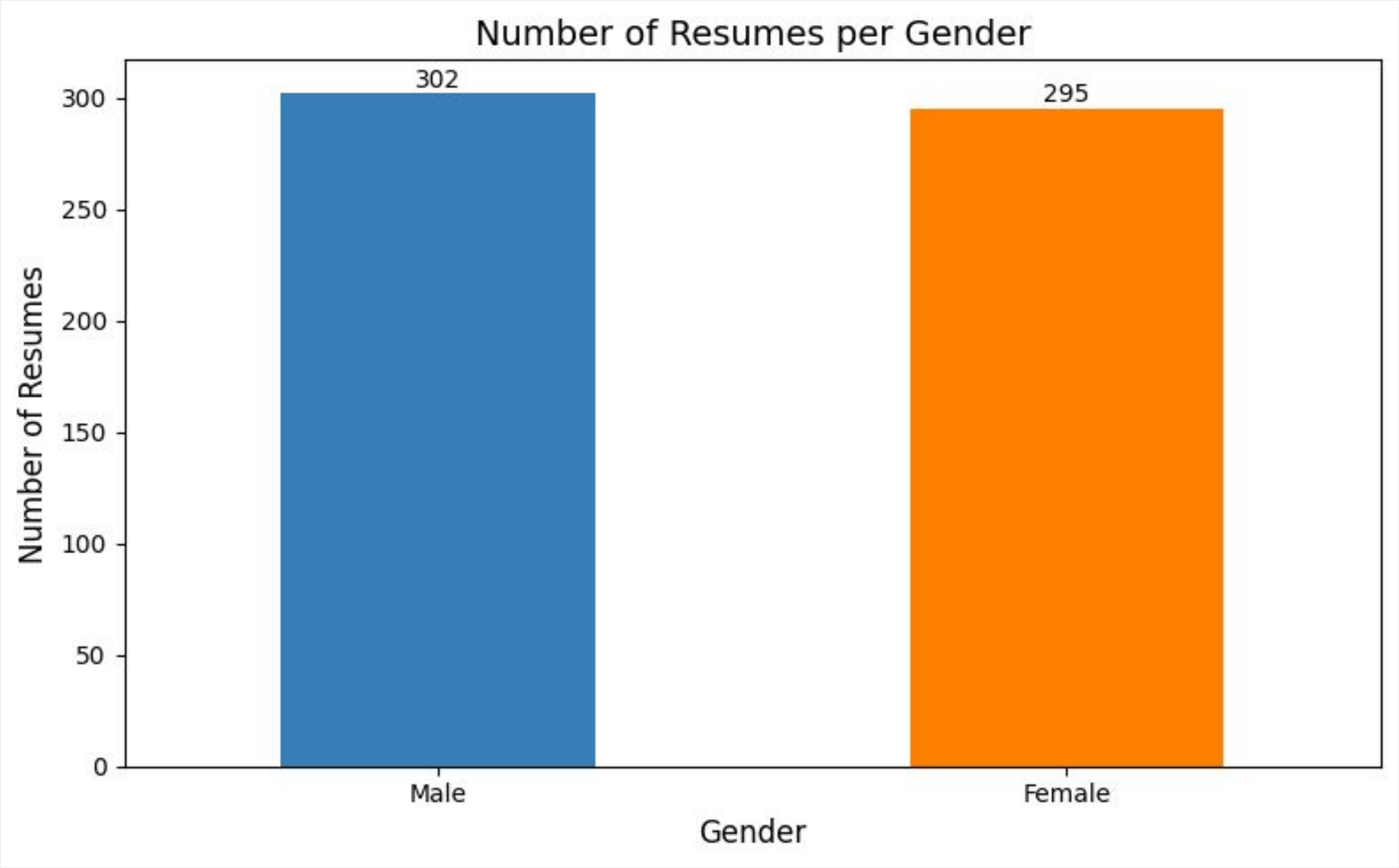
Mobile Technologies: IOS SDK, Object-c, XCode

Masculine prefixes	'active', 'adventurous', 'aggress', 'ambitio', 'analy', 'assert', 'athlet', 'autonom', 'battle', 'boast', 'challeng', 'champion', 'compet', 'confident', 'courag', 'decid', 'decision', 'decisive', 'defend', 'determin', 'domina', 'dominant', 'driven', 'fearless', 'fight', 'force', 'greedy', 'head-strong', 'headstrong', 'hierarch', 'hostil', 'impulsive', 'independen', 'individual', 'intellect', 'lead', 'logic', 'objective', 'opinion', 'outspoken', 'persist', 'principle', 'reckless', 'self-confiden', 'self-relian', 'self-sufficien', 'selfconfiden', 'selfrelian', 'selfsufficien', 'stubborn', 'superior', 'unreasonab'
Feminine prefixes	'agree', 'affectionate', 'child', 'cheer', 'collab', 'commit', 'communal', 'compassion', 'connect', 'considerate', 'cooperat', co-operat', 'depend', 'emotiona', 'empath', 'feel', 'flatterable', 'gentle', 'honest', 'interpersonal', 'interdependen', 'interpersona', 'inter-personal', 'inter-dependen', 'inter-persona', 'kind', 'kinship', 'loyal', 'modesty', 'nag', 'nurtur', 'pleasant', 'polite', 'quiet', 'respon', 'sensitiv', 'submissive', 'support', 'sympath', 'tender', 'together', 'trust', 'understand', 'warm', 'whin', 'enthusias', 'inclusive', 'yield', 'share', 'sharin'

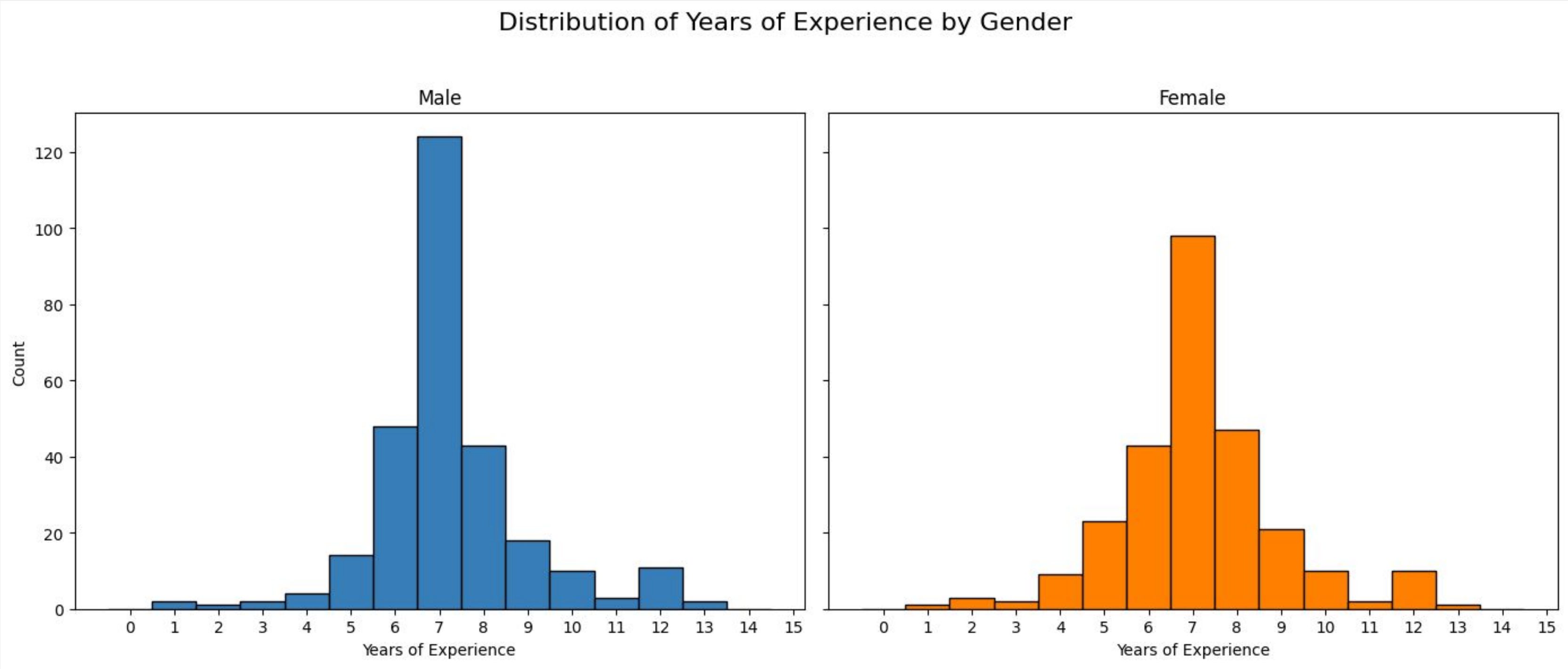
Data Analysis

- Applicants have 7 years of experience on average, with most between 4–10 years
- Number of past jobs is right-skewed, typically around 4 prior roles
- Experience and job history look similar across genders under random assignment
- Resumes contain more masculine-coded wording, regardless of assigned gender

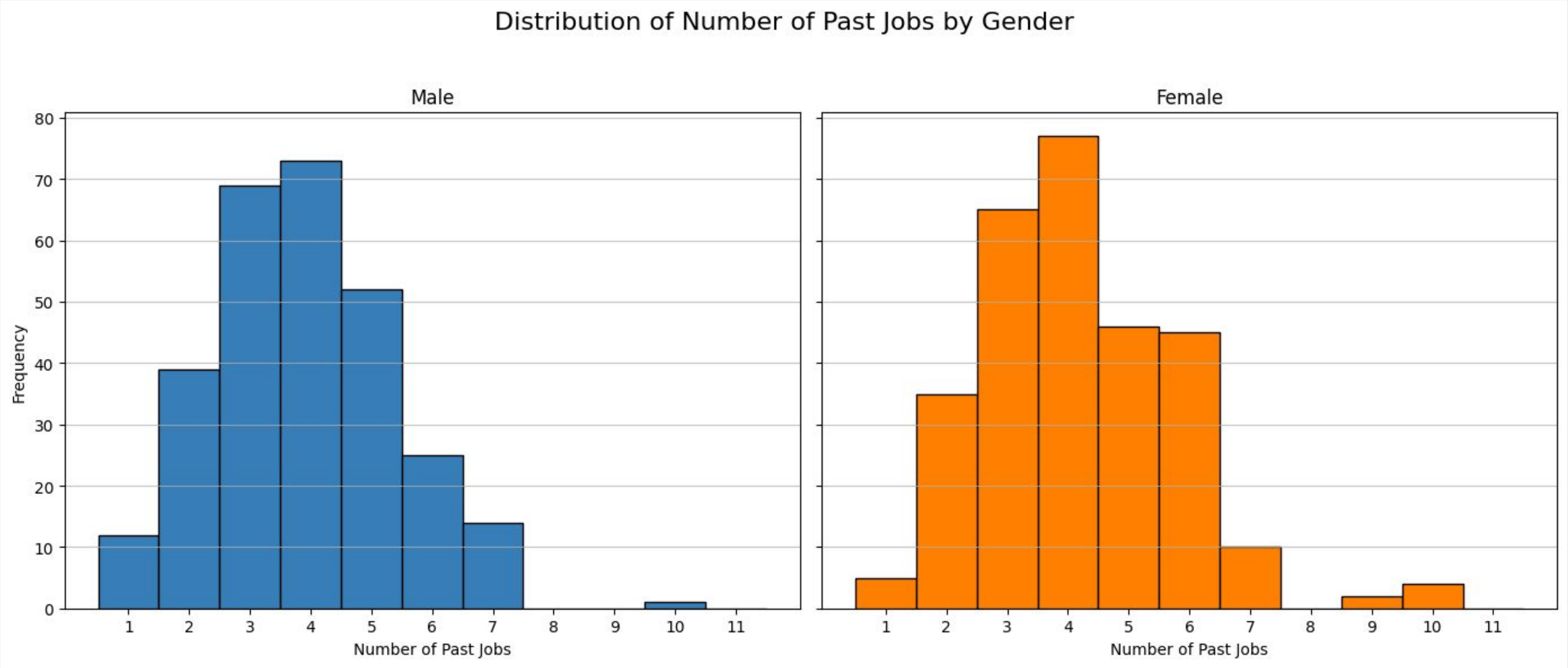
Profile of Applicants



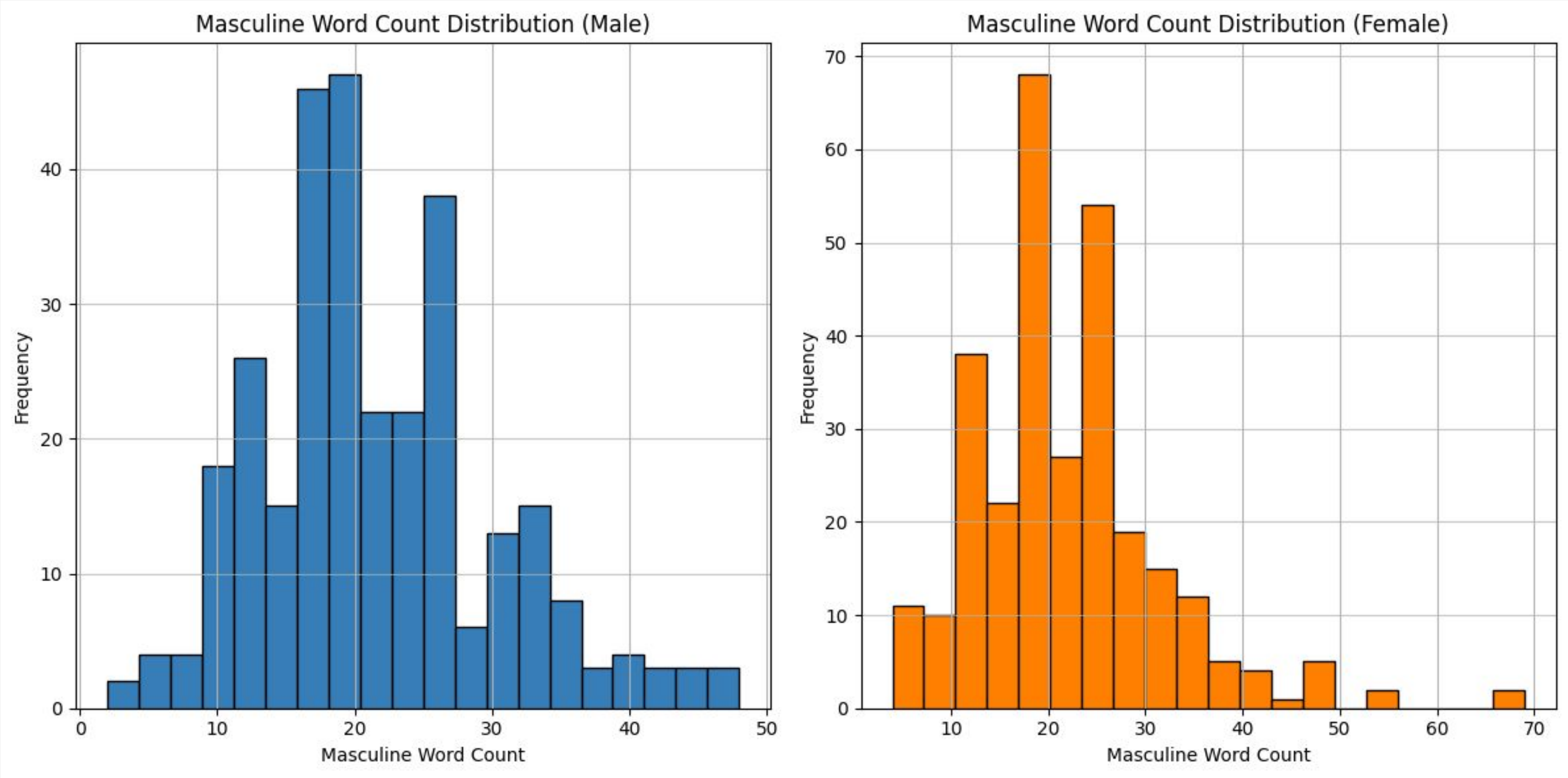
Profile of Applicants



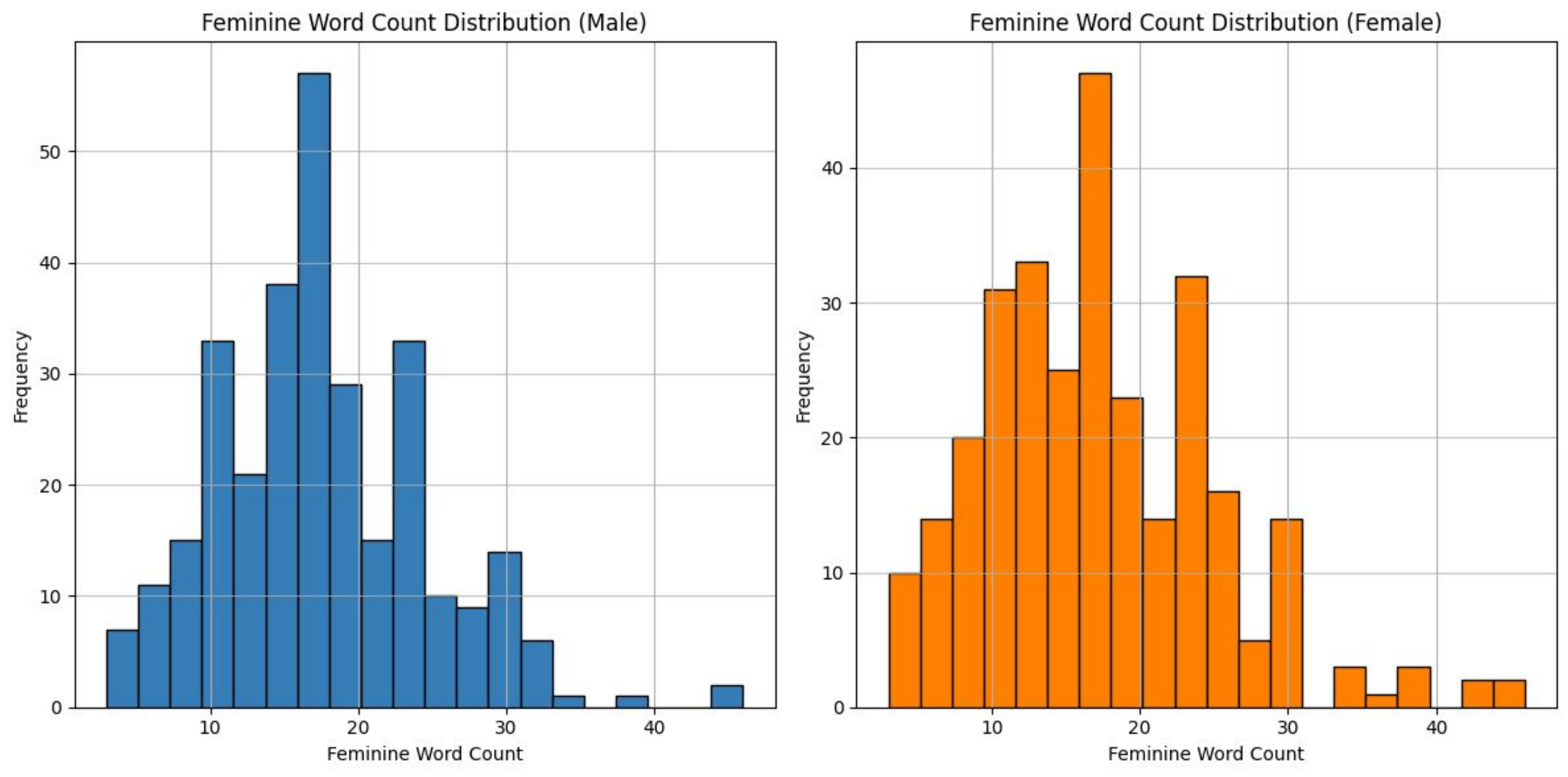
Profile of Applicants



Profile of Applicants



Profile of Applicants



Methods

- Algorithm combines semantic similarity (70%), years of experience (20%), and number of past jobs (10%)
  - Chosen to prioritize content relevance while still prioritizing basic measures of experience

- Experience features normalized with min–max scaling

- Semantic similarity computed using TF-IDF Vectorizer to match résumés to the description of the job posting
  - Measured if applicants' resumes matched against a set of keywords that suited the resume description

Methods

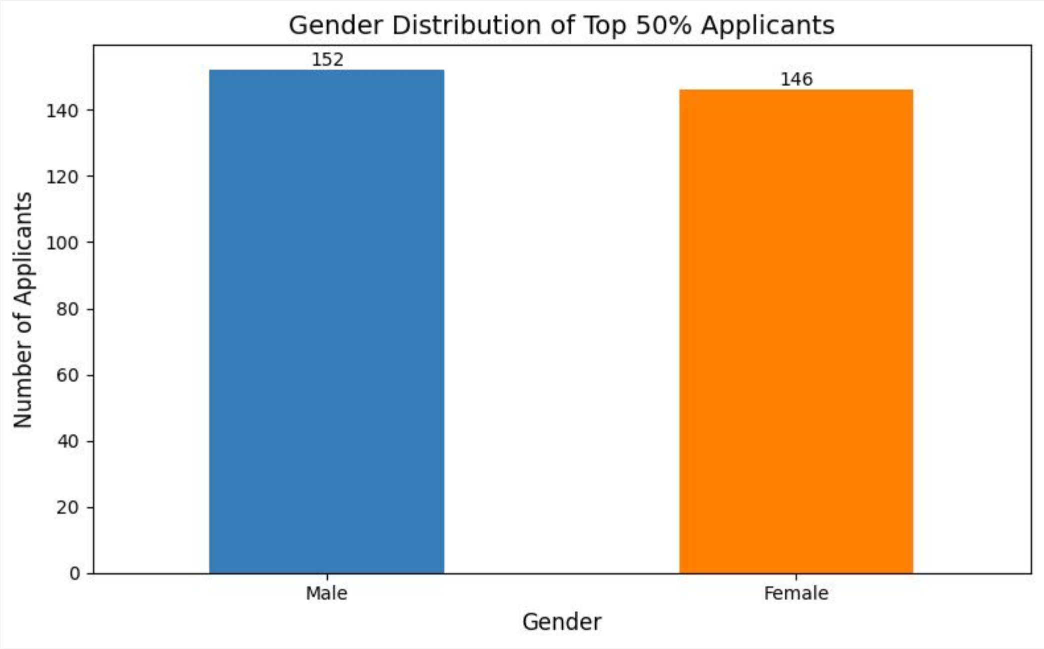
➔ Final score = (0.70 · Similarity) + (0.20 · Experience) + (0.10 · Past Jobs)

candidate_id	similarity_tfidf	years_experience	jobs_count	final_score	rank_final
122	0.04039557027	13	15	0.2716102325	1
228	0.02070510472	15	11	0.2644935733	2
227	0.02070510472	15	11	0.2644935733	2
90	0.0593893511	10	17	0.2549058791	3
92	0.0593893511	10	17	0.2549058791	3
91	0.0593893511	10	17	0.2549058791	3
99	0.0593893511	10	17	0.2549058791	3
97	0.0593893511	10	17	0.2549058791	3
98	0.0593893511	10	17	0.2549058791	3
96	0.05601773837	10	15	0.2425457502	4
434	0.1797376098	7	5	0.2391496602	5
433	0.1797376098	7	5	0.2391496602	5
252	0.05730242667	13	6	0.238445032	6
352	0.04444711499	8	21	0.2377796472	7

Results Summary

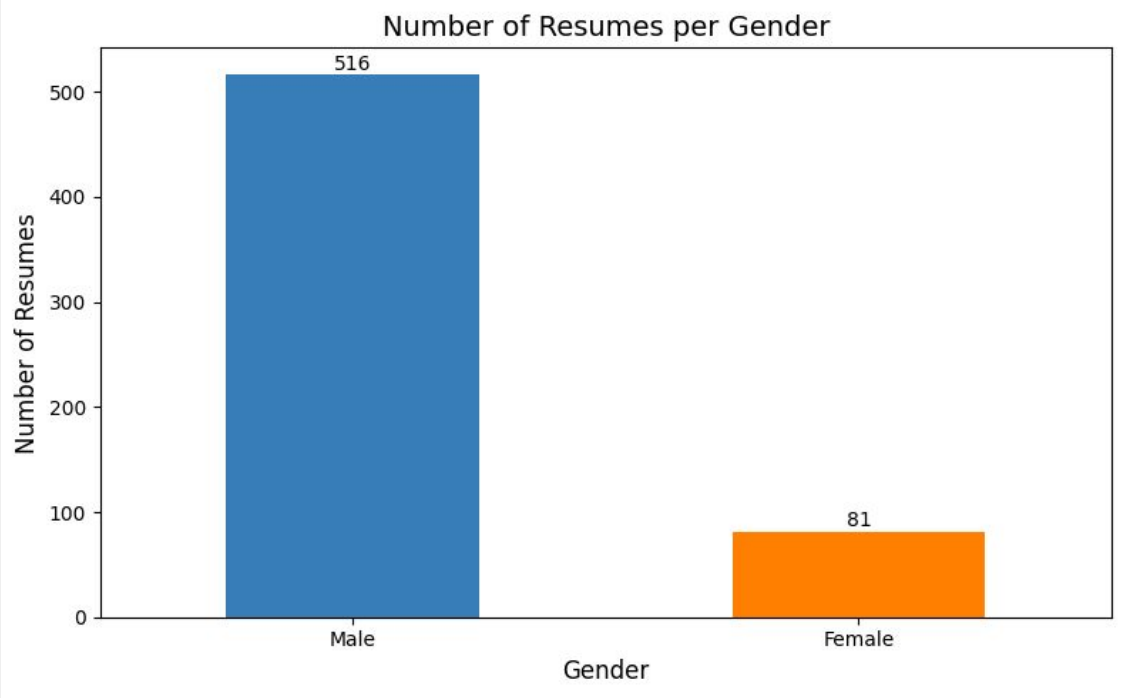
- ➔ Applicants ranked in descending score, demographic data kept only for descriptive analysis
- ➔ Algorithmic rankings under two gender conditions
- ➔ Comparison reveals how linguistic style affects fairness outcomes

Random Gender Assignment



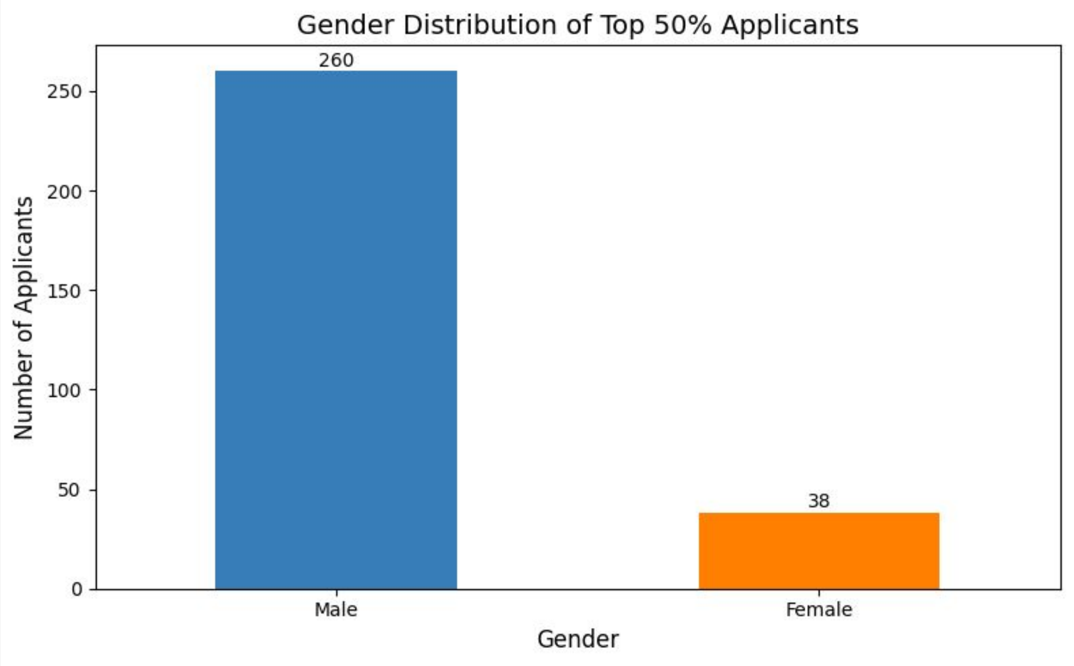
- ➔ Gender randomly assigned, dataset balanced
- ➔ Top 50% shows nearly equal male/female proportion
- ➔ Indicates algorithm behaves neutrally when gender ≠ resume features

Text-Based Gender Assignment



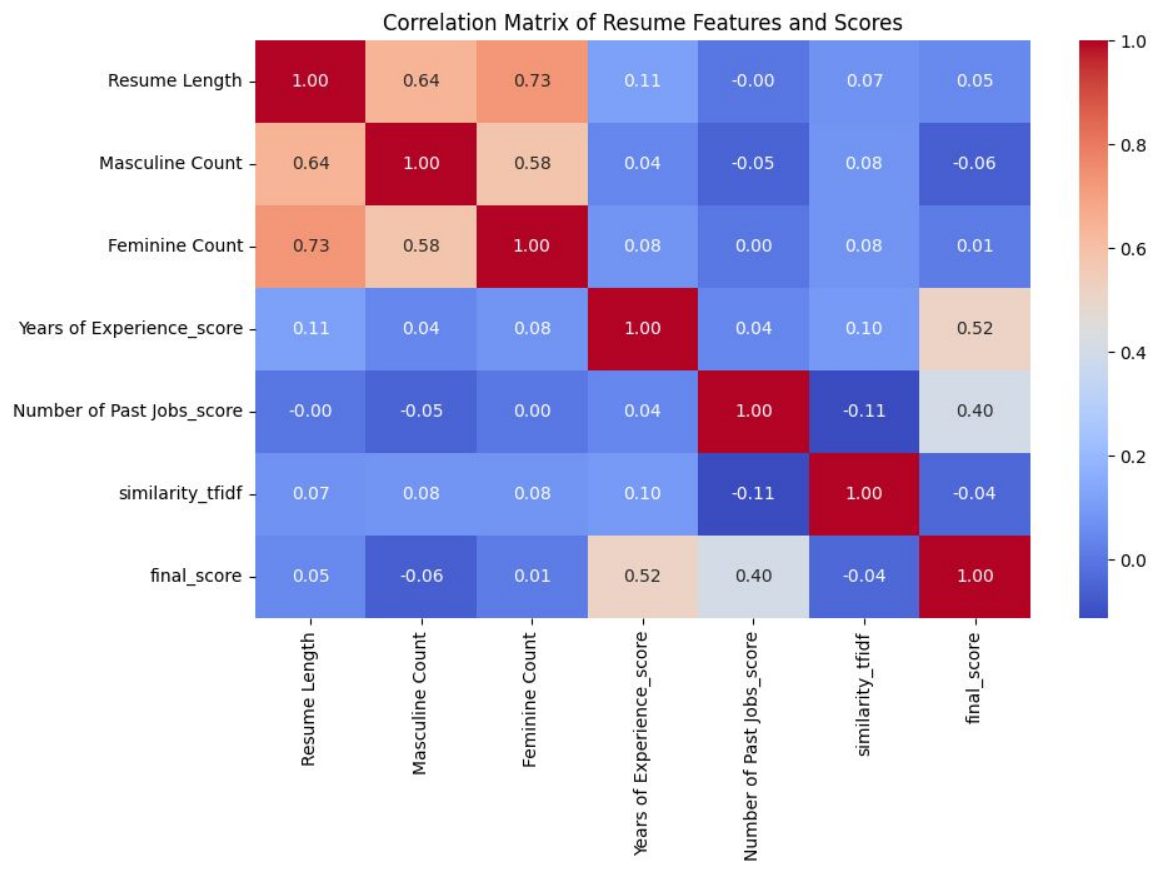
- ➔ Many resumes classified as male-coded due to linguistic patterns
- ➔ Masculine-coded resumes tend to be longer, more technical

Text-Based Gender Assignment



- ➔ As a result, male-coded resumes dominate the top 50% of applicants
  - ◆ Higher similarity scores

Verifying Hypothesis



Comparisons & Key Insights

- ➔ Random gender: rankings appear balanced
- ➔ Text-based gender: rankings skew toward male-coded resumes
- ➔ Differences arise from resume linguistic style, not algorithmic bias
- ➔ Masculine-coded writing aligns with scoring features → higher rankings

Discussion

- Conclusions
- Data Reliability and Validity
- Limitations for Generalization
- Ethical Considerations and Societal Implications

THANK YOU