

## CODEBOOK

### IL181 Gender inequality in online marketplaces

#### Project description:

This project seeks to investigate whether women have to undersell themselves in online job marketplaces to compete against men. We focus on the online job platform Freelancer.com to gather talent listings of individuals offering their services and compare the hourly rate offered by men vs. women, controlling for factors like job function, experience, job completion rate, etc.

#### Codebook description:

This codebook corresponds to the dataset after matching-specific preprocessing was carried out. Many of the variables are now bucketed, and some have been condensed into new features that are derived from others. This version of the dataset is the one that was used for matching. The corresponding data file is "skills\_certifications\_categorized\_female\_treatment.csv".

## VARIABLES

Name	Column label	Description	Format	Collection method	Missing values
Search query	search_query	The search query that was used to retrieve the user profile. For example, "software engineer" or "copywriter". This is kept only for reference, as the Freelancer search feature isn't that great. Do not use this value as an occupation indicator.	String	User inputted	No
Person name	name	The name of the person. Only the first name is provided.	String	Scraped	No
Gender	gender	The gender of the person -- can be "male" or "female". Male = 0, female = 1.	Integer	Derived	No
Location size	location_size	The size of the location where the user is based, with size being judged by the population of the place. 0 = extra small (smallest), 5 = extra extra large (biggest).	Integer	Derived	No
Join date	join_date_from_earliest	The date the user joined the Freelancer.com platform, represented as the number of days after the earliest date in the dataset. For example, a value of "10" indicates the user joined 10 days after the earliest join date in the dataset. The earliest join date has a value of 0.	Number	Scraped	No

Hourly rate	hourly_rate	The hourly rate charged by the user to perform their services. All rates are in USD.	Number	Scraped	No
Pay grade	pay_grade	The amount of money earned by the user doing projects in the skill/category they belong in. It is a number from 0 to 10, and increases the more money the freelancer earns in the category.	Number	Scraped	No
Average rating	avg_rating	The user's average rating received by the people they have worked for. It is a number from 0 to 5. The higher the number, the better the rating. Some users may have a rating of 0, which means they have not been reviewed yet.	Number	Scraped	No
Number of reviews	num_reviews	The number of reviews that the user has received.	Number	Scraped	No
Number of recommendations	num_recommendations	The number of recommendations that have been written on the user's profile by people that have contracted them before.	Number	Scraped	No
Percentage of jobs completed	pct_jobs_completed	The percentage of jobs the user has completed successfully. It is a number from 0 to 100.	Number	Scraped	Yes
Percentage of jobs on budget	pct_on_budget	The percentage of jobs performed by the user that have stayed within the agreed-upon budget. It is a number from 0 to 100.	Number	Scraped	Yes
Percentage on time	pct_on_time	The percentage of jobs completed on time by the user. It is a number between 0 and 100.	Number	Scraped	Yes
Repeat hire rate	repeat_hire_rate	The rate at which this user has been hired again by individuals who have contracted with them. The number is a percentage from 0 to 100.	Number	Scraped	Yes
Preferred freelancer	verification_preferred_freelancer	Whether the user has received the "preferred freelancer" verification from Freelancer.com. This is described by the website as an "elite group" of the best freelancers.	Boolean	Scraped	No
Identity verified	verification_identity_verified	Whether the user has verified their identity on the platform. This consists of a government-issued ID check and a proof of address.	Boolean	Scraped	No
Payment verified	verification_payment_verified	Whether the user has been verified to have a valid payment method	Boolean	Scraped	No
Phone verified	verification_phone_verified	Whether the user has been verified to have a valid phone number	Boolean	Scraped	No
Email verified	verification_email_verified	Whether the user has been verified to have a valid email address	Boolean	Scraped	No
Facebook connected	verification_facebook_connected	Whether the user has been verified to have a Facebook account	Boolean	Scraped	No

Badges	badge_[badge_name]	Some users have badges they have earned on the platform listed next to their username on their profile. These indicate whether they have been verified, have some kind of professional membership, amongst others. N/A indicates the user does not have the badge listed on their profile.	Boolean	Scraped	Yes
Engineering skills	engineering_skills	The number of engineering skills listed on a user's profile.	Integer	Derived	No
Writing skills	writing_skills	The number of writing skills listed on a user's profile	Integer	Derived	No
Technical programming skills	technical_programming_skills	The number of technical programming skills listed on a user's profile	Integer	Derived	No
Language translation skills	language_translation_skills	The number of language translation skills listed on a user's profile	Integer	Derived	No
Finance and accounting skills	finance_accounting_skills	The number of finance and accounting skills listed on a user's profile	Integer	Derived	No
Management skills	management_skills	The number of management skills listed on a user's profile	Integer	Derived	No
Marketing and business skills	marketing_business_skills	The number of marketing and business skills listed on a user's profile	Integer	Derived	No
Performance arts skills	performance_arts_skills	The number of performance arts skills listed on a user's profile	Integer	Derived	No
Design skills	design_skills	The number of design skills listed on a user's profile	Integer	Derived	No
Teaching and training skills	teaching_training_skills	The number of teaching and training skills listed on a user's profile	Integer	Derived	No
Miscellaneous skills	miscellaneous_skills	The number of miscellaneous skills listed on a user's profile	Integer	Derived	No
Language certifications	language_certifications	The number of language certifications listed on a user's profile	Integer	Derived	No
Freelancer certifications	freelancer_certifications	The number of freelancer certifications listed on a user's profile	Integer	Derived	No
General skill certifications	general_skill_certifications	The number of general skill certifications listed on a user's profile	Integer	Derived	No
Programming certifications	programming_certifications	The number of programming certifications listed on a user's profile	Integer	Derived	No

