King Saud University
College of Computer and Information Sciences
Department of Information Technology
IT 362 – Data Science
2nd Semester 1446 H

Analyzing Freelancing Trends and Sustainability

Report

	Students Name	ID
Section #66850 Group #6	**Layan Alfawzan**	**444200793**
	Rose Mady	**444200107**
	Aljawharah Alwabel	**444200750**
	Arwa Almutairi	**444201055**

Supervised by: Dr.Reem Alqifari

TABLE OF CONTENTS

INTRODUCTION

Freelancing has become a popular career path due to its flexibility and the opportunities it provides across various fields. However, several questions arise regarding its sustainability, income stability, and the demographics of those engaging in this type of work.

Understanding freelancing trends can help clarify whether freelancing is a sustainable career choice, identify income patterns, and determine the demographics of freelancers. For example, it can help examine whether freelancing is a sustainable career path, identify countries with the highest concentration of skilled freelancers, and more. Which in return, answers our main question of the research, i.e. *Is freelancing a sustainable career choice, and what are the key factors influencing its viability?*

DATA SOURCES

- **Data sources:**

Freelancer Web Pages: Data scraped from Freelancer.com, includes freelancer profiles containing:

- Name
- Hourly Rate
- Skills
- Location
- Rating
- Reviews
- Total Earnings
- Bio

- **Details of the Data Collected:**

1. **Number of Observations:** Approximately 1000 observations.
2. **Features and Data Types:**

Feature	Data Type	Description
Freelancer Name	Qualitative	The freelancer's display name (Nominal).
Hourly Rate	Quantitative	Freelancer's rate per hour (Ratio).
Skills	Qualitative	Areas of expertise (Nominal).
Location	Qualitative	Country of residence/work (Nominal).
Rating	Quantitative	Average rating given to the freelancer based on client reviews and feedback.
Reviews	Quantitative	Number of client reviews (Ratio).
Total Earnings	Quantitative	Freelancer's overall earnings from the projects and contests they have completed on the site (Ratio).
Bio	Qualitative	Freelancer's background summary (Nominal).

3. Dataset Bias Evaluation:

- **Representation Bias:**
    - Potential bias if freelancers from certain regions or industries are over/underrepresented.
    - Scraping may unintentionally favor highly active profiles.

- **Measurement Bias:**
    - Hourly rates may not reflect actual earnings due to differences in project durations or frequency.
    - Skills listed may not fully represent freelancers' capabilities if profiles are outdated.

- **Historical Bias:**
    - The dataset might reflect inequalities in access to freelancing platforms (e.g., higher representation of freelancers from developed countries or certain industries).

OBJECTIVES

Questions to Be Answered Using the Dataset:

1. Is freelancing income sustainable as a career path?
2. Which regions/countries have the highest concentration of skilled freelancers?
3. What are the most in-demand skills in the freelancing market?
4. What is the average income for freelancers across different skill sets?
5. Is there a positive correlation between having multiple skills and higher earnings, or do specialized freelancers perform better?

METHOD

1. **Data Preprocessing:**

   The first step in preparing the dataset for analysis is cleaning the raw data obtained through web scraping. This includes removing duplicates, handling missing values, and addressing inconsistencies such as variations in skill names. Noise in the data, such as irrelevant information or HTML tags, will be removed to ensure a clean and structured dataset. Additionally, tokenization will be applied to break down text-based attributes, such as freelancer bios and skills, into meaningful components. This will help standardize skill categorization, improve text analysis, and ensure consistency in data representation.

2. **Data Transformation:**

   The transformation process will focus on **organizing the data into analyzable categories**.

   Skills will be normalized to group similar or overlapping terms (e.g., "Graphic Designer" and "Visual Designer").

   Regions or countries will categorize location data to support geographical analysis.

   Finally, the cleaned and categorized data will be stored in a structured format such as CSV or JSON for easy integration with analytical tools.

3. **Analysis Techniques:**

   Once the data is preprocessed and transformed, various analytical methods will be applied to answer the research questions.

   Descriptive analysis will summarize the dataset, providing insights into average hourly rates, review counts, and skill diversity.

   Correlation analysis will identify relationships between variables, such as the number of skills and hourly rates.

   Geographical analysis will map the distribution of freelancers by location, highlighting regions with the highest concentrations of skilled freelancers.

Skill demand analysis will determine the most common skills and their association with earnings.

Lastly, **regression analysis** will assess whether freelancers with multiple skills tend to have higher hourly rates compared to those who specialize in a single skill.

4. **Reporting insights:**

The final step involves synthesizing the findings into a comprehensive report that provides actionable conclusions based on the analyzed data. Key insights regarding freelancer income stability, skill demand, and regional variations will be clearly outlined. The report will also include recommendations for freelancers on optimizing their career paths, as well as potential implications for policymakers or organizations interested in supporting the freelancer economy. Limitations of the study and areas for further research will also be discussed to provide a well-rounded perspective.

CHALLENGES FACED AND RECOMMENDATIONS

During the data collection process, a key challenge was the dynamic nature of web scraping, where each code execution retrieved different data from the platform. This inconsistency made it difficult to maintain a stable dataset for analysis, as newly scraped data varied with each run.

To address this issue, the immediate solution was to save the scraped data into a structured dataset as soon as it was collected. By storing the data in a static format, i.e. CSV file, we ensured consistency in our analysis and prevented discrepancies caused by continuous data changes.

This approach allows us to work with a fixed dataset, ensuring that results remain reproducible even if the web scraping process is executed multiple times.