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 (Ratio).
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:**
    - 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.

PRIMARY DATA EDA

This section presents the results of exploratory data analysis (EDA) based on the scraped freelancer profiles dataset. The analysis focuses on uncovering patterns and insights related to freelancer earnings, ratings, reviews, skills, and geographical distribution. The findings are organized to address our five research questions directly.

IS FREELANCING INCOME SUSTAINABLE AS A CAREER PATH? (QUESTION 1)

- Total Earnings distribution is highly skewed, with many freelancers earning low amounts and a few earning very high totals (Figure 3: Histogram of Total Earnings).
- Boxplot of Total Earnings also highlights the wide range and presence of high outliers (Figure 4: Boxplot of Total Earnings).
- The scatter plot of Rating vs. Total Earnings (Figure 8) shows that freelancers with higher ratings tend to have higher total earnings, suggesting that reputation and quality are crucial for income sustainability.

WHICH COUNTRIES HAVE THE HIGHEST CONCENTRATION OF SKILLED FREELANCERS? (QUESTION 2)

- Location analysis shows that most freelancers are concentrated in India, Pakistan, Bangladesh, Indonesia, and the USA (Figure 4: Bar chart of Freelancer Locations).
- This reflects the globalized nature of the freelancing market, with significant representation from South Asia and Southeast Asia in particular.

WHAT IS THE AVERAGE INCOME FOR FREELANCERS ACROSS DIFFERENT SKILL SETS? (QUESTION 3)

- Hourly Rate distribution shows that most freelancers charge between $10 and $50 per hour, with a few charging above $100 (Figure 1: Histogram of Hourly Rate).
- Boxplot of Hourly Rate confirms the presence of significant outliers, indicating that some freelancers set much higher rates than the average (Figure 2: Boxplot of Hourly Rate).
- Although we examined "Skills Count vs. Total Earnings" (Figure 7), total earnings vary widely for each skill count, indicating that income is not solely dependent on the number of skills.

IS THERE A POSITIVE CORRELATION BETWEEN HAVING MULTIPLE SKILLS AND HIGHER EARNINGS, OR DO FREELANCERS SPECIALIZING IN ONE SKILL PERFORM BETTER? (QUESTION 4)

- Scatter plot of Skills Count vs. Total Earnings (Figure 7) shows that earnings vary widely for all skill counts. Some freelancers with fewer skills still earn high amounts, suggesting that having more skills does not always lead to higher earnings.
- Correlation matrix (Figure 9) confirms that Skills Count was not directly analyzed in the matrix, but relationships between Hourly Rate, Reviews, Rating, and Total Earnings were explored.
- This indicates that both specialization and skill diversification can be effective, depending on the value of the skill in the market.

WHAT ARE THE MOST IN-DEMAND SKILLS IN THE FREELANCING MARKET? (QUESTION 5)

- Skills Count vs. Total Earnings (Figure 7) shows that freelancers with more skills do not always earn more, as some with fewer skills also achieve high earnings.
- The number of skills alone is not a strong indicator of higher income; other factors like skill type and quality may influence earnings.
- (Figure 11) shows the Top 10 Most In-Demand Skills, led by Data & Analytics, Design, and Development, followed by Marketing and Writing & Translation.
- These top skills represent the most competitive areas in the freelancing market.
- The analysis highlights where opportunities for specialization may exist based on market demand.

Overall, the analysis shows that hourly rates and total earnings are highly variable among freelancers. Higher ratings and more reviews are associated with higher total earnings, emphasizing the importance of client satisfaction and reputation. The data also shows that earnings vary by country, with some locations showing higher earning ranges than others. Finally, having more skills does not necessarily guarantee higher earnings, as many freelancers with fewer skills also report significant earnings.

Figures

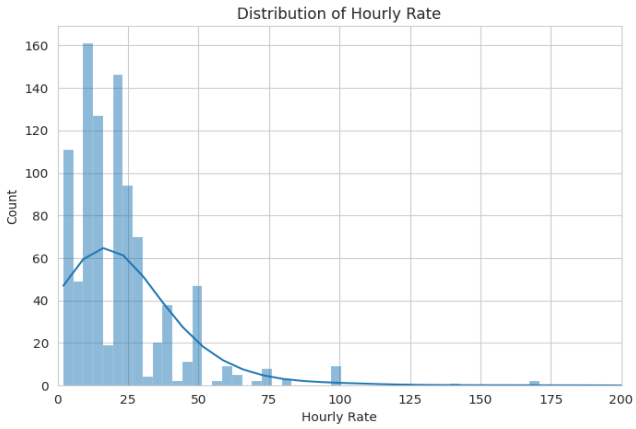

Figure 1:Histogram of Hourly Rate

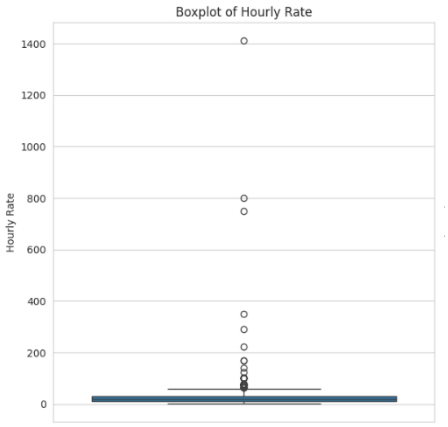

Figure 2: Boxplot of Hourly Rate

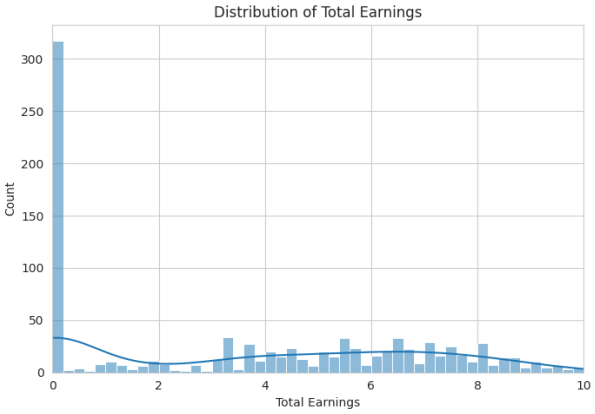

Figure 3: Histogram of Total Earnings

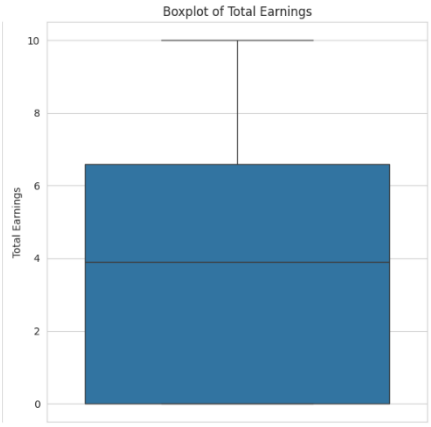

Figure 4: Boxplot of Total Earnings

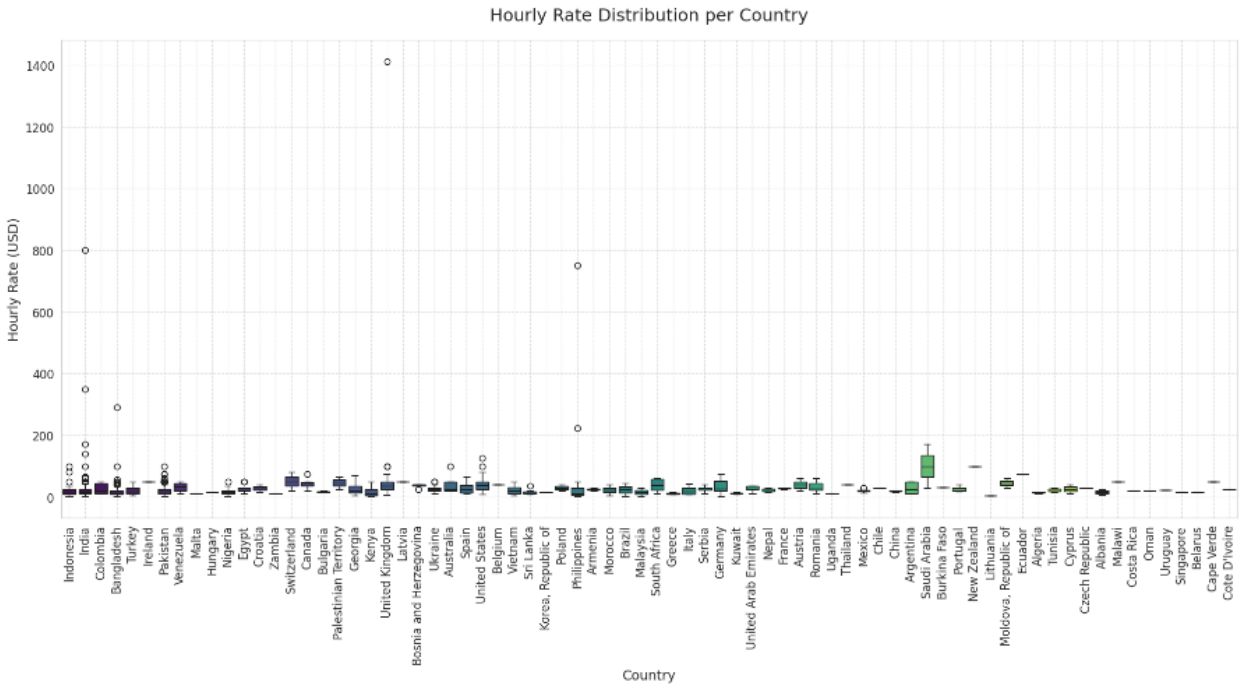

Figure 5: Boxplot of Hourly Rate by Country

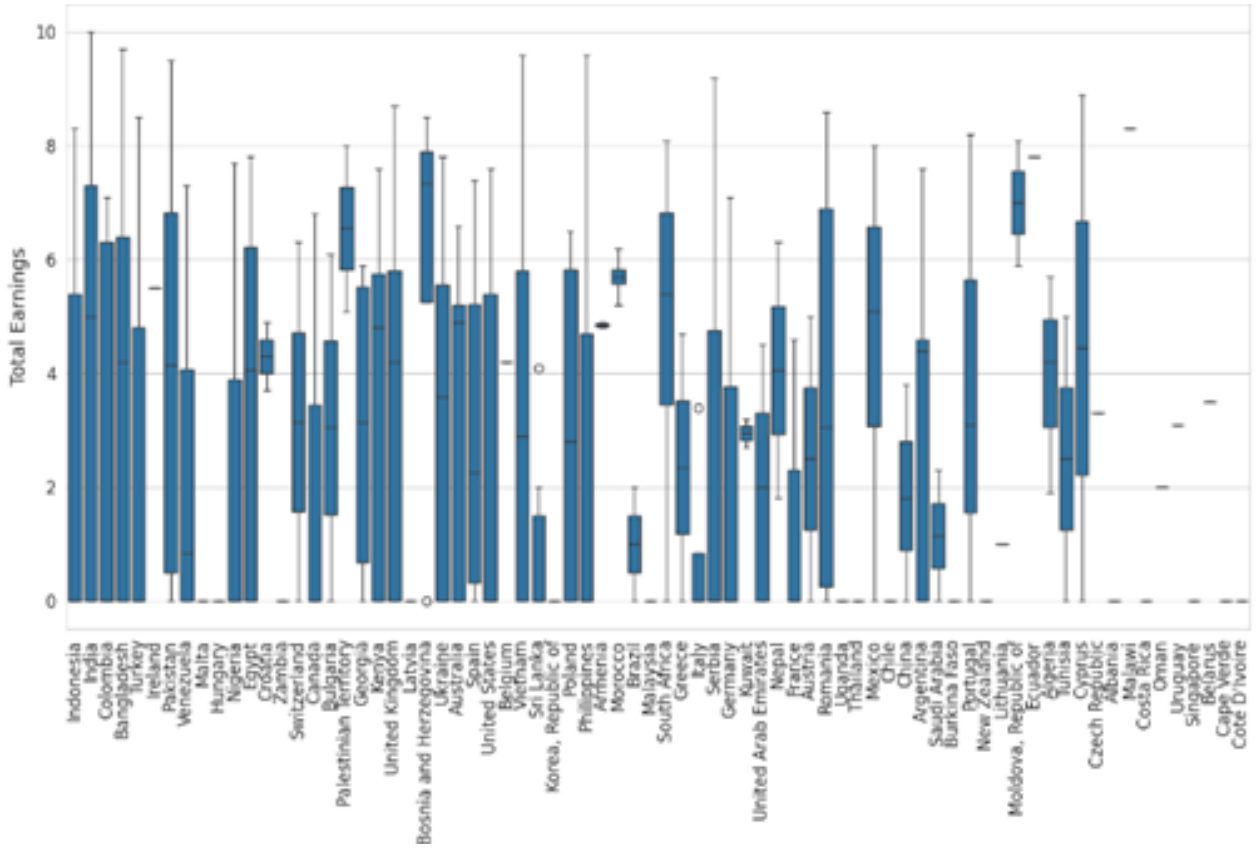

Figure 6: Histogram of Total Earnings

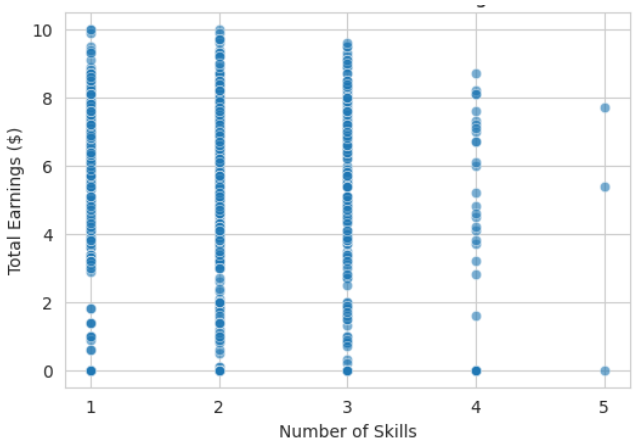

Figure 7: Scatter Plot: Skills Count vs. Total Earnings

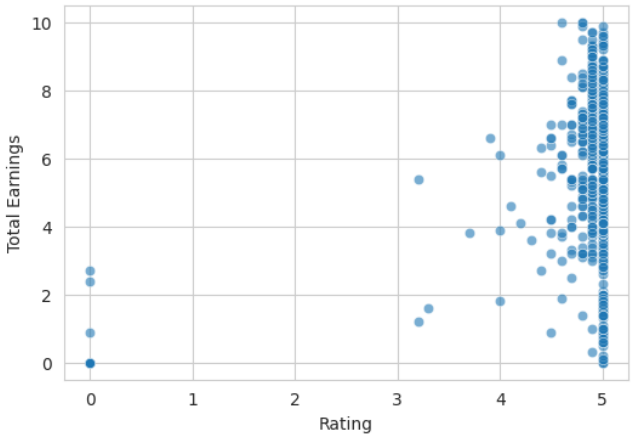

Figure 8: Scatter Plot: Rating vs. Total Earnings

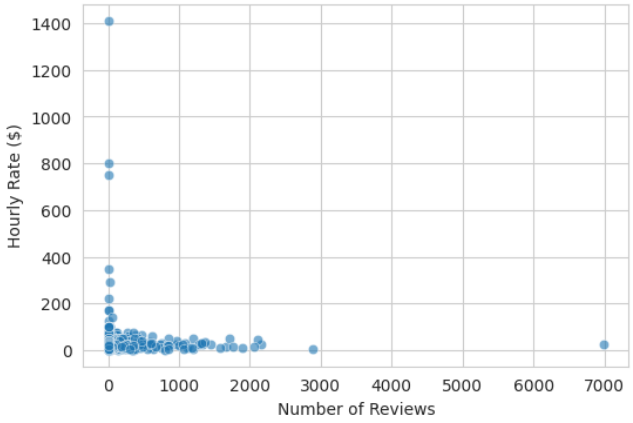

Figure 9: Scatter Plot: Reviews vs. Hourly Rate

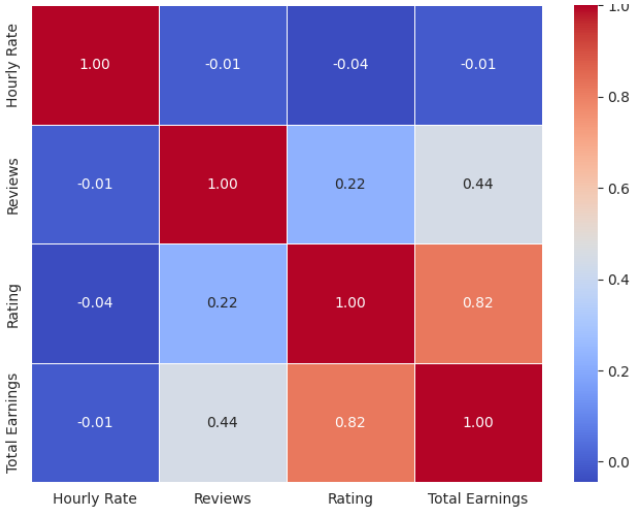

Figure 10: Correlation Matrix Heatmap

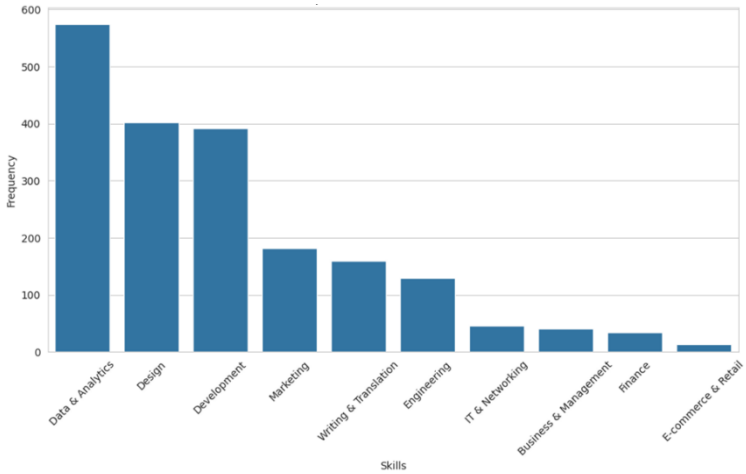

Figure 11: Top 10 Most In-Demand Skills

SECONDARY DATA

Use of Secondary Data:

Secondary data was not needed as the primary data collected through web scraping includes all necessary information for the analysis. The dataset contains detailed attributes such as hourly rates, total earnings, ratings, reviews, skills, and locations, which are sufficient to explore earnings patterns, skill demand, and market trends. All attributes of interest were available on the scraped website, making external datasets unnecessary for this analysis.

SUMMARY OF NEW INSIGHTS AND HYPOTHESES

New Insights:

- **Ratings as a Strong Indicator of Success:** Higher ratings strongly correlate with higher total earnings (0.82 correlation). Freelancers with a rating close to 5.0 tend to earn more, suggesting that client satisfaction and positive feedback are major drivers of success in the freelancing market.
- **Limited Influence of Hourly Rates:** Hourly rate has negligible correlation with earnings, reviews, and ratings. This suggests that charging higher rates does not guarantee higher earnings or better reviews. Factors such as reputation, consistency, and skill specialization play a more prominent role in determining success.
- **Global Variations in Freelancer Earnings and Hourly Rates:** There is a significant variation in hourly rates and earnings across different countries. Developed countries, like the United States, the United Kingdom, and Switzerland, generally have higher hourly rates and earnings, while countries with more competitive freelancing markets, such as India and Pakistan, tend to have lower rates and earnings. This highlights the influence of economic conditions and market competition on freelance pricing and income levels.
- **Freelancer Location and Market Demand:** Freelancers are spread across many countries, with a notable concentration in India and Pakistan. This points to a global freelancing market with highly competitive pricing in some regions. Additionally, Data & Analytics and Design fields dominate the freelancing market, reflecting high demand for these skills.

Hypotheses Generated:

Hypothesis 1:

Freelancers with a higher number of reviews are more likely to acquire new clients, leading to increased total earnings.

- Independent Variable: number of reviews

- Dependent Variable: total earnings

Hypothesis 2:

Freelancer earnings and hourly rates are influenced by a country's economic conditions, where freelancers in countries with higher GDP per capita tend to charge higher rates and earn more.

- Independent Variable: Economic conditions (measured by GDP per capita).

- Dependent Variable: Freelancer earnings and hourly rates.

Hypothesis 3:

A small percentage of top freelancers dominate earnings and reviews, creating barriers for new freelancers to compete.

- Independent Variable: Freelancer rank based on earnings and reviews

- Dependent Variable: Market share (earnings and review distribution)

Hypothesis 4:

Freelancers specializing in a single skill tend to earn higher profits than those with multiple skills due to the demand for deep expertise in specific fields.

- Independent Variable: Number of skills listed by a freelancer

- Dependent Variable: Freelancer income levels