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

كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

Analyzing Freelancing Trends and Sustainability

Report

<table>
<tr><th></th><th>Students Name</th><th>ID</th></tr>
<tr><td rowspan="4">Section #66850
Group #6</td><td>Layan Alfawzan</td><td>444200793</td></tr>
<tr><td>Rose Mady</td><td>444200107</td></tr>
<tr><td>Aljawharah Alwabel</td><td>444200750</td></tr>
<tr><td>Arwa Almutairi</td><td>444201055</td></tr>
</table>

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?*

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?

DATA DESCRIPTION

- ## Data source:

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

- ## Data collection:

```python
# Collect and parse first page
response = requests.get('https://www.freelancer.com/freelancers/1')
print(response.status_code)

soup = BeautifulSoup(response.text, 'html.parser')
print (soup.prettify())
```

We first fetch the HTML content of the pages to the attributes we are interested in about freelancers by requesting the page, and then we parsed the content using BeautifulSoup.

- ## Screenshots:

Raw Data (unstructured):

Bio

	Freelancer Name	Hourly Rate	Skills	Location	Rating	Reviews	Total Earni	Bio
2	artdjuna	$30 USD per hour	Logo Design, Photoshop, Illustration, Corporate Identity, Website Design	Indonesia	4.9	171	8.3	Trying to do the best I can and will make my clients satisfied, and I have experience since 2011..
3	kkwebart09	$15 USD per hour	Graphic Design, Photoshop Design, Logo Design, Website Design, HTML	India	4.9	17	6.4	I'm a professional web designer and web developer with 5 years of experience. I specialize in creating visu
4	enervell	$50 USD per hour	Node.js, AngularJS, PostgreSQL, JavaScript, PHP	Colombia	5	27	6.3	I am a seasoned Full Stack Developer with over 9 years of experience in crafting robust and secure softwar
5	sharminnaharm	$10 USD per hour	Logo Design, Graphic Design, Photoshop, Illustrator, Illustration	Bangladesh	4.9	138	6.8	Hello everyone! As-salamu alaykum,
6	ancineha	$15 USD per hour	Graphic Design, Website Design, User Interface / IA, PHP, CSS	India	5	71	7.7	Hello,
7	Kubragull	$50 USD per hour	3D Rendering, Interior Design, 3D Modelling, Building Architecture, Home Design	Turkey	4.5	55	6.4	Welcome to my page!
8	mdsadikulislamss	$10 USD per hour	Graphic Design, Illustrator, Photoshop, Business Card Design, Flyer Design	Bangladesh	5	15	5.5	I'm Sadikul, one of the top online graphic designer in the freelance community. Mostly I'm expert in Logo, F
9	trivediheenalove	$20 USD per hour	AJAX, API, Codeigniter, eCommerce, Express JS	India	5	4	5.5	We are providing B2B, B2C and SAAS based application development, We have an expert team in Front En
10	RMMcontractors	$50 USD per hour	PHP, MySQL, Laravel, HTML, Data Entry	Ireland	5	7	5.5	Hi there! Iâ€™m Ronan, your trusted IT consultant.

Processed Data (structured):

	Freelancer	Hourly Rate	Skills	Location	Rating	Reviews	Total Earnir	Bio	Skills Count
2	artdjuna	0.035088	2	31	0.98	0.024446	0.83	trying less	2
3	kkwebart0\(	0.016291	1	30	0.98	0.00243	0.64	developer	1
4	enervell	0.06015	1	17	1	0.00386	0.63	developer	1
5	sharminnal	0.010025	1	6	0.98	0.019728	0.68	price branc	1
6	ancineha	0.016291	5	30	1	0.01015	0.77	contact int	5
7	Kubragull	0.06015	3	64	0.9	0.007863	0.64	high less c	3
8	mdsadikuli	0.010025	1	6	1	0.002144	0.55	less flyer g	1
9	trivediheen	0.022556	3	30	1	0.000572	0.55	based rese	3
10	RMMcontra	0.06015	2	32	1	0.001001	0.55	based expe	2

- ## Feature summary:

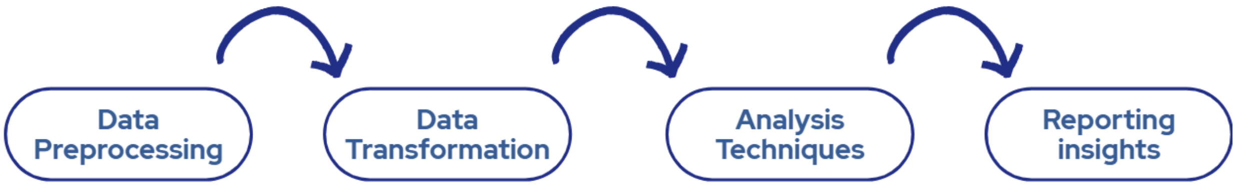
1. **Number of Observations:** 1000 observations.
2. **Number of features:** 8 features.
3. **Data Types:**

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).

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).

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.

KEY EDA INSIGHTS AND HOW EACH OBJECTIVE/QUESTION WAS ADDRESSED USING THE DATA

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

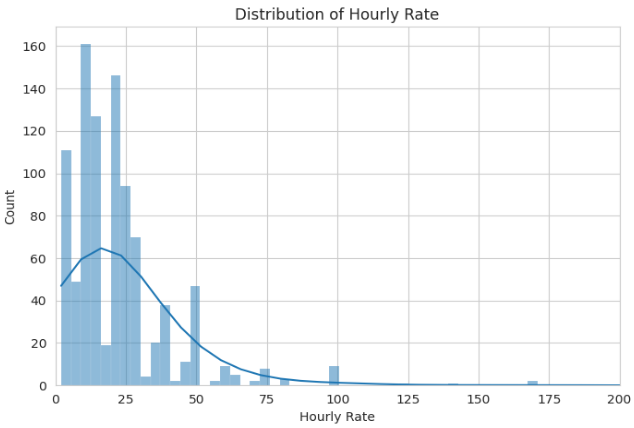


Figure 1:Histogram of Hourly Rate

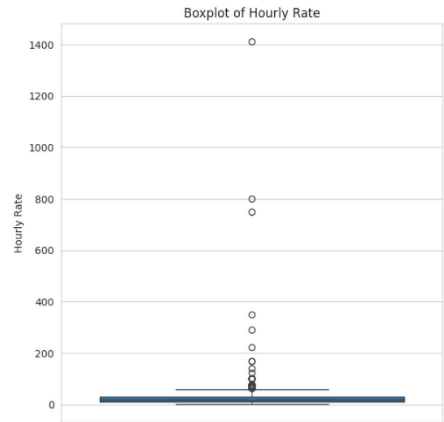


Figure 2: Boxplot of Hourly Rate

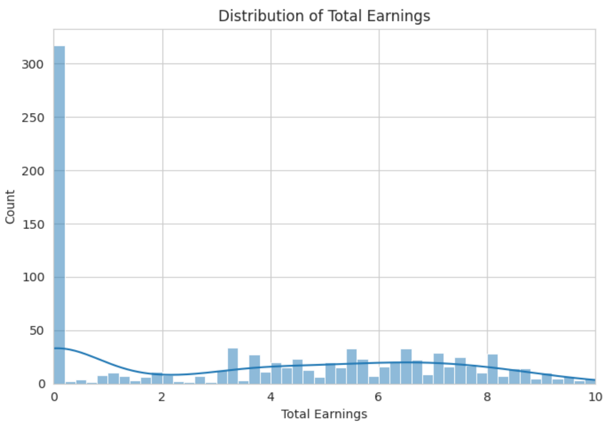


Figure 3: Histogram of Total Earnings

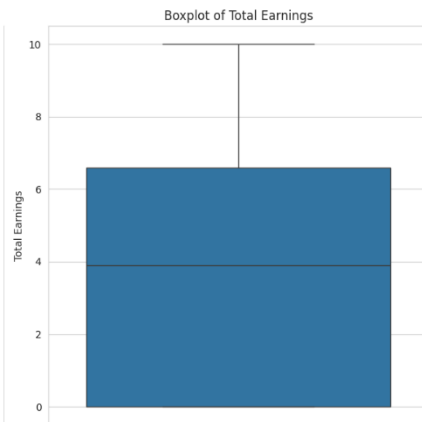


Figure 4: Boxplot of Total Earnings

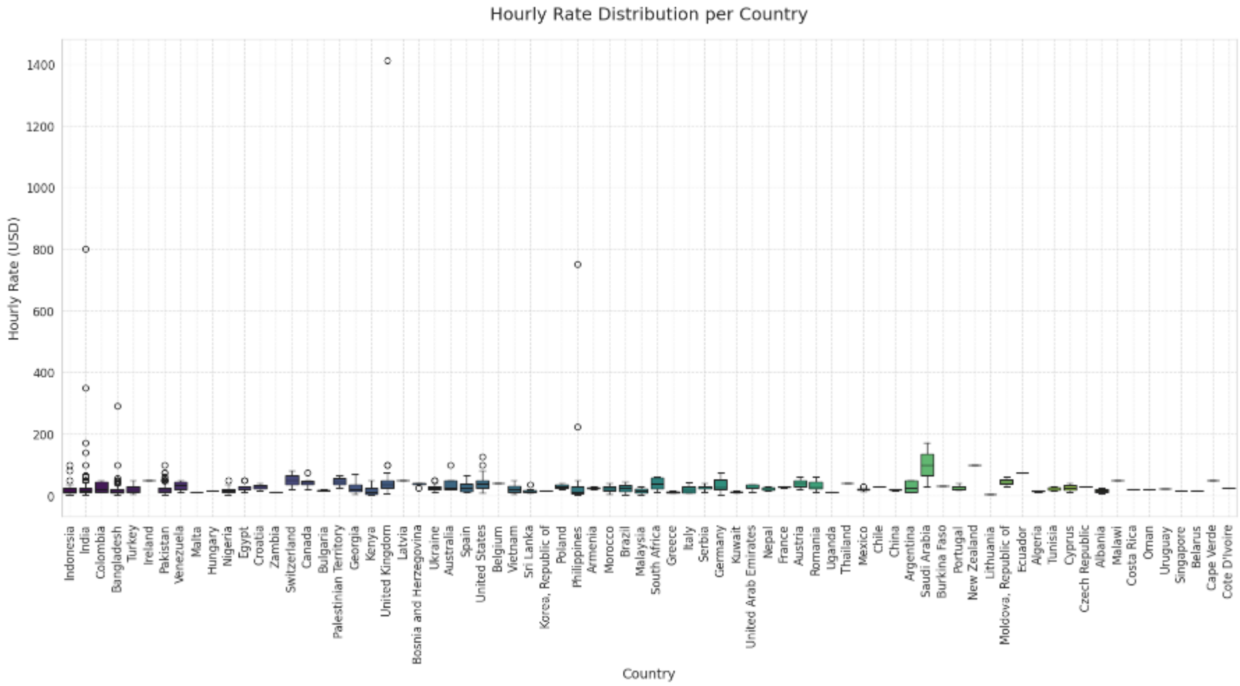


Figure 5: Boxplot of Hourly Rate by Country

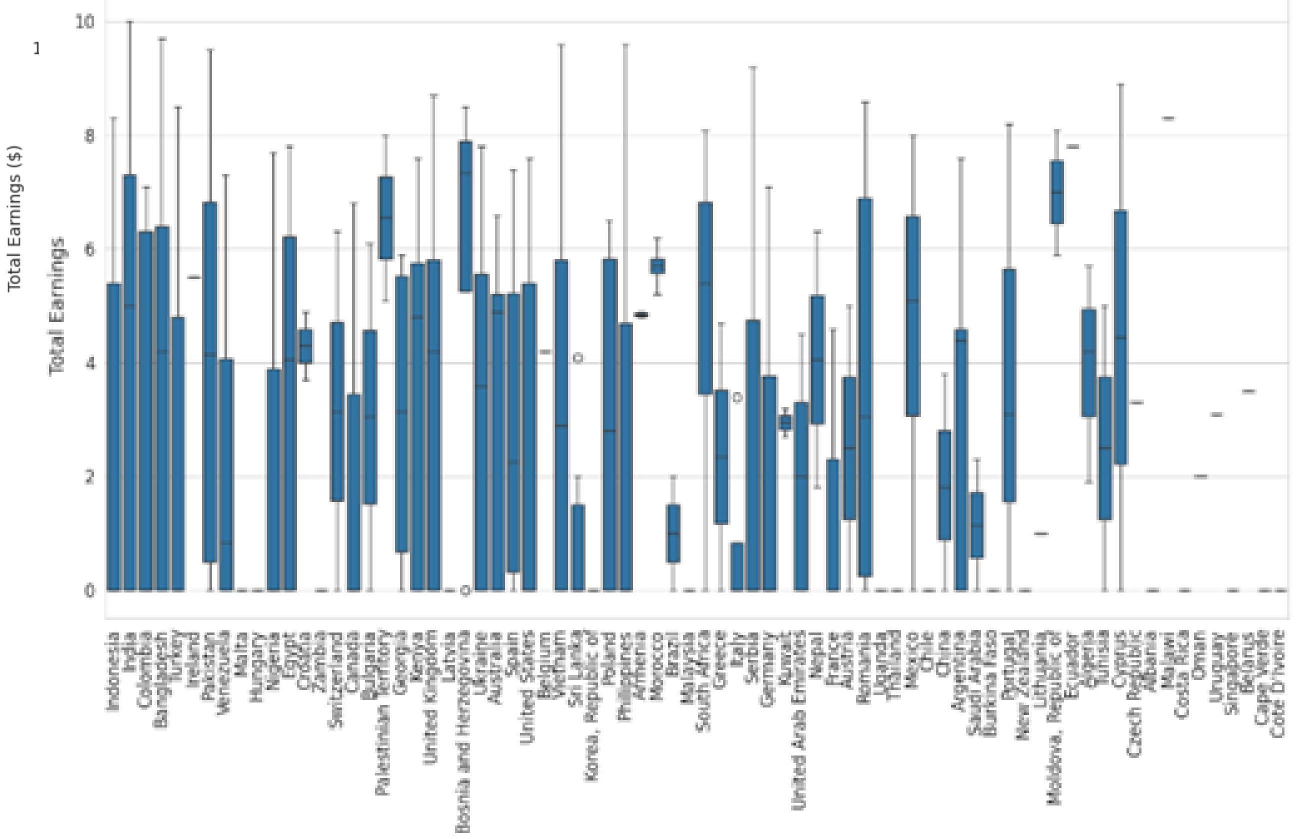


Figure 6: Histogram of Total Earnings

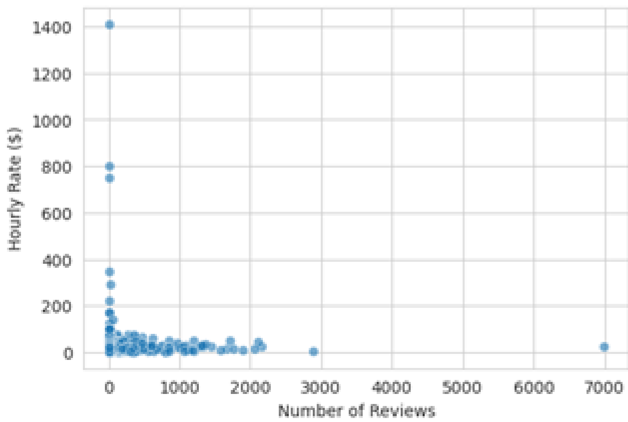


Figure 9: Scatter Plot: Reviews vs. Hourly Rate

Figure 10: Correlation Matrix Heatmap

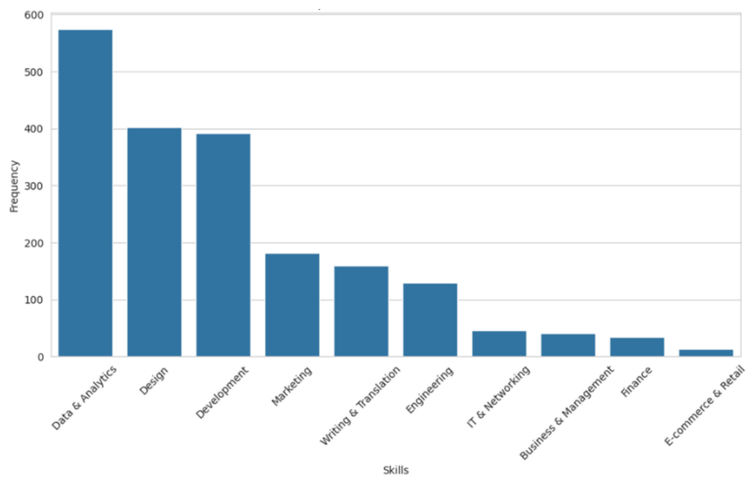


Figure 11: Top 10 Most In-Demand Skills

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.

DATA PROCESSING AND MAIN TECHNIQUES APPLIED

Normalization:

Normalization scales numerical columns like Hourly Rate, Rating, Reviews, and Total Earnings to a range of 0 to 1 using Min-Max Scaling. This ensures features with different scales contribute equally, preventing bias in algorithms like clustering or classification. This step prepares the dataset for accurate and fair analysis when constructing modeling algorithms.

Encoding:

Encoding converts categorical data into numeric values for analysis. We used LabelEncoder on the Location column to transform categorical location names into numerical labels.

For the Skills column, we focused on the count of skills by splitting the comma-separated values and defaulting to 0 for missing entries. This prepares the dataset for machine learning or further 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:

Freelancers' average income is influenced by their skill set, hourly rate, and rating.

- Independent Variable: Skill set, hourly Rate, Rating.

- Dependent Variable: average Income

MODELLING

Regression Models:

We used clustering to test hypothesis 5: Freelancers' average income is influenced by their skill set, hourly rate, and rating.

- Independent Variable: Skill set, hourly Rate, Rating.
- Dependent Variable: average Income

We started by preparing the data and selecting wanted attributes, then split data for training and testing (80 for testing and 20 for training) then train data and evaluate it.

1- Baseline model

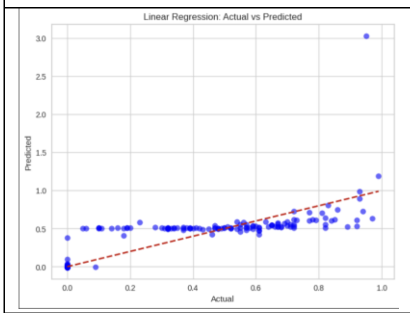
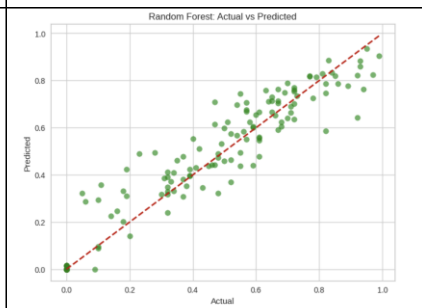
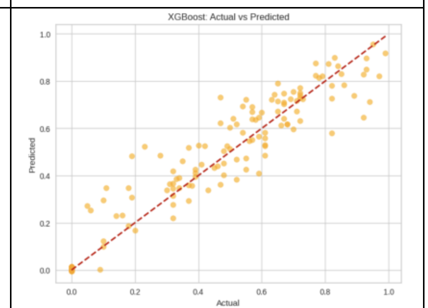
This model assumes a linear relationship between the predictors and earnings. It helps us set a performance benchmark.

2- Random Forest regression

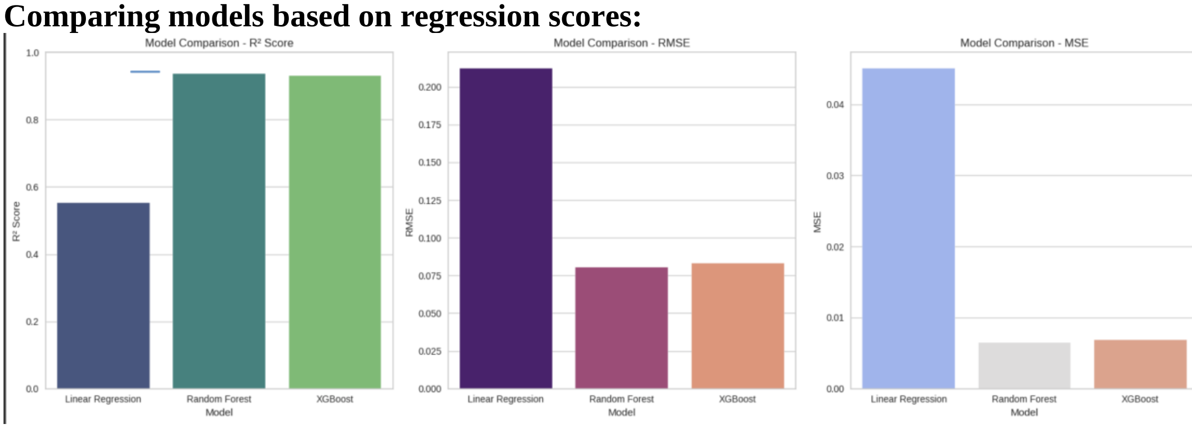
Random Forest builds multiple decision trees and averages their predictions, making it robust against overfitting and capable of modeling nonlinear patterns

3- Gradient Boosting Regression (XGBoost)

XGBoost is a boosting method that sequentially builds trees, each trying to correct the errors of the previous one. It is often more accurate and efficient than Random Forest, especially in structured datasets.

Baseline model	**Random Forest regression**	**Gradient Boosting Regression (XGBoost)**
		
Linear Regression R^2: 0.5537 Linear Regression RMSE: 0.2123 Linear Regression MSE: 0.0451	Random Forest Regression R^2: 0.9358 Random Forest Regression RMSE: 0.0805 Random Forest Regression MSE: 0.0065	XGBoost Regression R^2: 0.9316 XGBoost Regression RMSE: 0.0831 XGBoost Regression MSE: 0.0069

Comparing models based on regression scores:



Key Findings: Random Forest is the Top Performer:

- Highest R^2 (0.936) and lowest errors (RMSE=0.080, MSE=0.0065).
- Marginally outperforms XGBoost in accuracy and consistency.

XGBoost is a Close Second wiht near-identical to RF (R^2=0.932, RMSE=0.083) but slightly less precise.

Linear Regression is Underpowered:Poor fit (R^2=0.554) and high errors (RMSE=0.212), failing to capture nonlinear relationships.

Clustering Models:

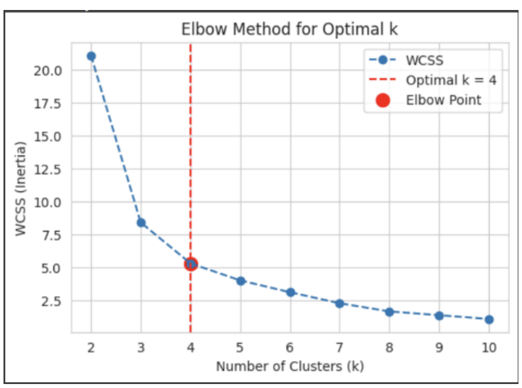
We used clustering to test hypothesis 1: Freelancers with a higher number of reviews are more likely to acquire new clients, leading to increased total earnings.

- Independent Variable: number of reviews
- Dependent Variable: total earnings

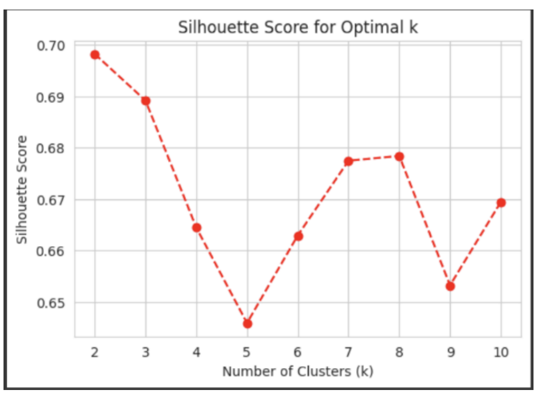
1- K-means algorithm

K-Means is an unsupervised machine learning algorithm used for clustering data points into groups based on their similarity. It works by iteratively assigning data points to k clusters, where k is a predefined number. The algorithm minimizes the distance between points within a cluster while maximizing the distance between clusters.

i. Determine the optimal number of clusters (k): methods like the Elbow Method and Silhouette Score are used, which help evaluate the quality of clustering.

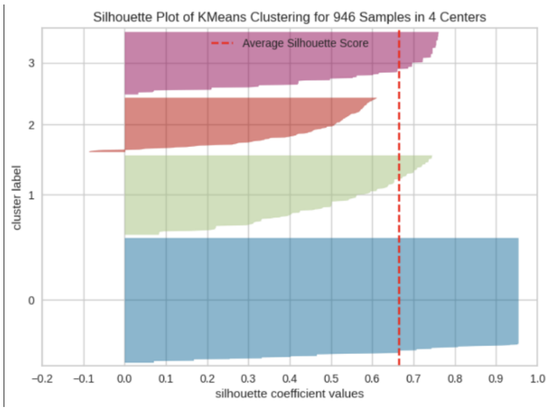


The Elbow Method helps determine the best number of clusters by plotting Within-Cluster Sum of Squares (WCSS) for different k values. The elbow point (where the WCSS curve starts flattening) suggests the best k.
We can observe that the optimal number of clusters (k = 4)

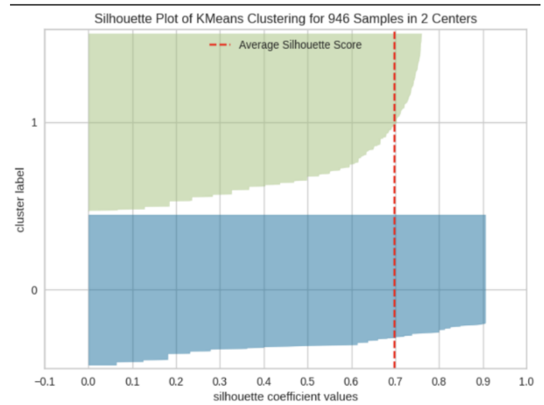


The Silhouette Score measures how well-defined clusters are, with higher scores indicating better separation. A graph of silhouette scores for different k values helps confirm the best clustering choice.
We can observe that the highest Silhoutte score is ~0.698 for (k = 2)

ii. K visualization and evaluation: performing K-means clustering using (k=2, and k-4) and visualize the resulting clusters. To evaluate the performance of the clustering.

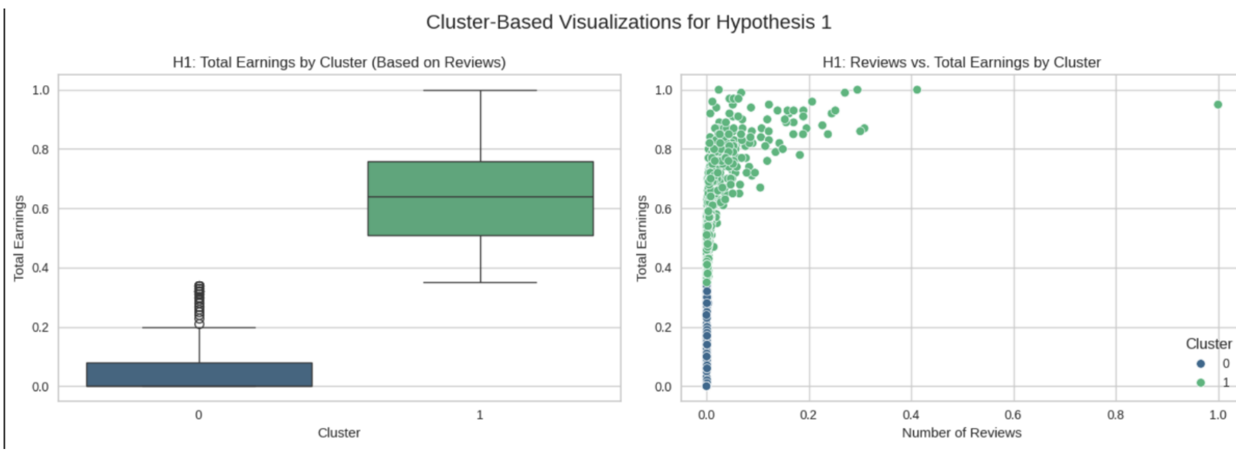


WSS: 5.285959333426594
Average Silhouette Score: 0.6642172179400736



WSS: 21.078687664960075
Average Silhouette Score: 0.6981923091630428

Based on the metrics we've analyzed, including WSS (Within-Cluster Sum of Squares) and the average Silhouette score, we have determined that K = 2 is the most suitable choice for our clustering model.

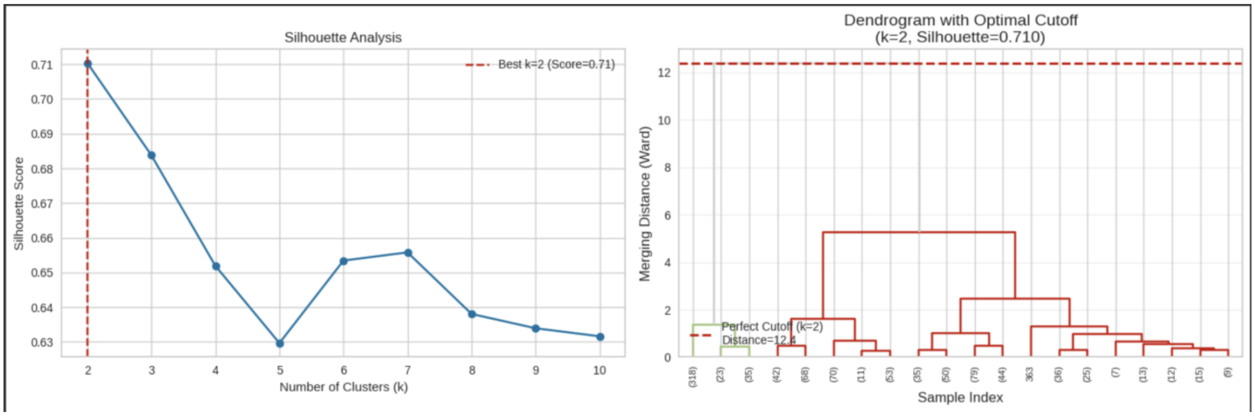


Cluster 0 likely represents freelancers who have built credibility and client trust (through reviews), leading to increased income. Meanwhile, Cluster 1 reflects less experienced or less visible freelancers with limited traction in the market. This strongly supports the hypothesis.

2- Hierarchal clustering

Hierarchical clustering is a method of clustering that builds a hierarchy of clusters in a tree-like structure (dendrogram). Unlike K-means, it does not require specifying the number of clusters beforehand. It is useful for understanding relationships between data points and discovering patterns.

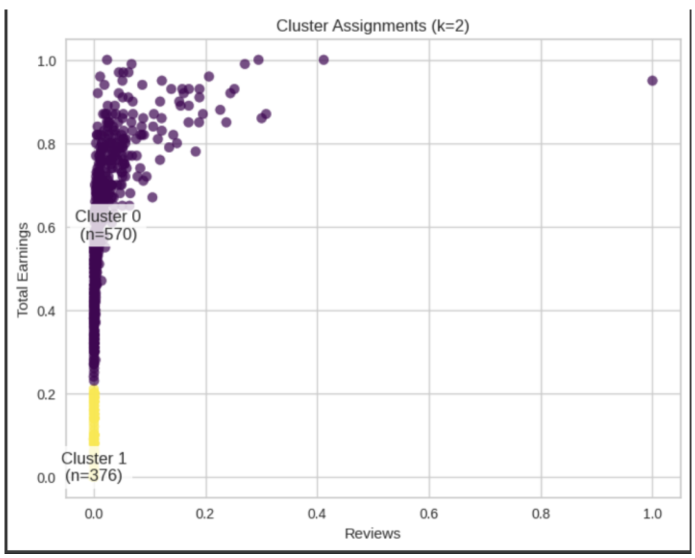
i. Determine the optimal number of clusters: using silhouette score as quantitative measurement tree-like diagram (dendrogram), marking where to cut the tree based on the best k dynamically.



silhouette score is highest at k=2 with a score of 0.71 (close to 1) indicates strong clustering—points in each group are similar, while different groups stay apart, meaning the data naturally splits into two well-separated groups.

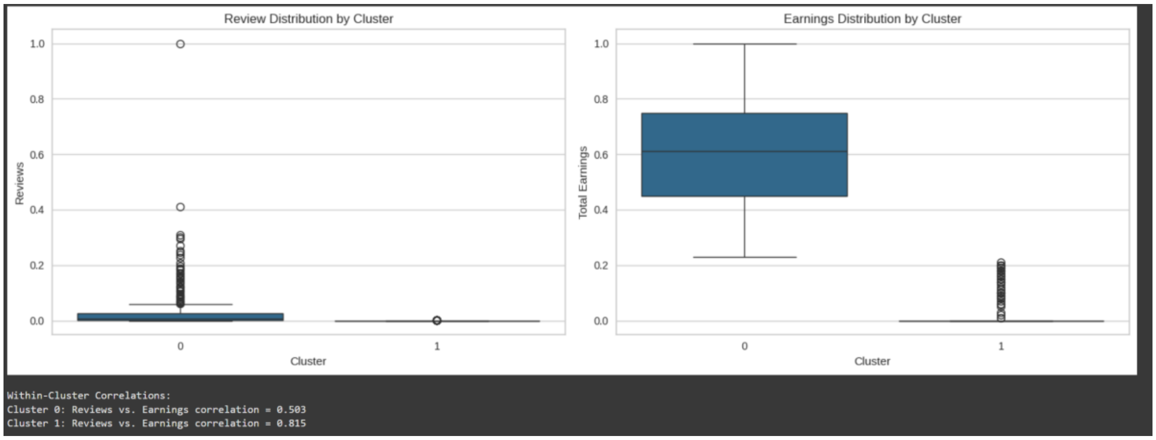
Also, The dendrogram (tree diagram) shows a clear split into two main branches before smaller merges. The red cutoff line (at height 15) intersects just two major clusters, matching the silhouette result.

ii. Cluster visualization: based on dendogram



The clustering results show a clear distinction between high-earning freelancers (who have reviews) and low-earning freelancers (who have none).

iii. Validation: using boxplots to show how Reviews and Earnings differ between clusters, and we measured correlation between them.

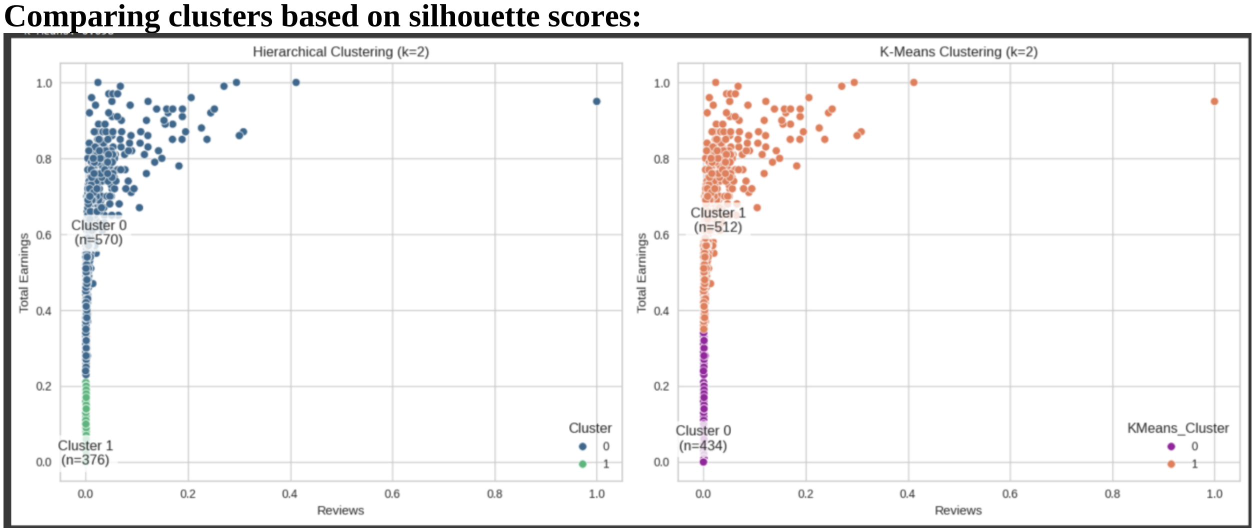


The within-cluster correlations suggest the following:

Cluster 0 (correlation = 0.503): There is a moderate positive relationship between reviews and earnings. As the number of reviews increases, earnings tend to increase as well, though the correlation is not very strong.

Cluster 1 (correlation = 0.815): There is a strong positive relationship between reviews and earnings. In this cluster, the more reviews a person has, the significantly higher their earnings tend to be, indicating reviews are a strong indicator of earnings in this group.

Comparing clusters based on silhouette scores:



Method	Silhouette Score	Correlation (Cluster 0)	Correlation (Cluster 1)	Objects per Cluster	Interpretation
Hierarchical Clustering	0.710	0.503	0.815	750 in Cluster 0, 376 in Cluster 1	Excellent separation (score > 0.7). Well-defined clusters with minimal overlap.
K-Means Clustering	0.698	0.5029	0.8152	512 in Cluster 0, 434 in Cluster 1	Strong separation (approaching 0.7). Slightly more overlap than hierarchical, but still good structure

- Silhouette Scores: Hierarchical Clustering has a higher silhouette score of 0.710, indicating better cluster separation and well-defined clusters with minimal overlap. In contrast, K-Means has a silhouette score of 0.698, suggesting strong separation but slightly more overlap between clusters.

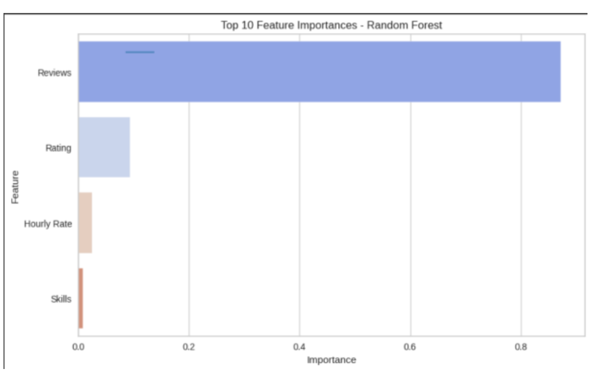
- Cluster Sizes: Hierarchical Clustering results in imbalanced clusters, with 750 objects in Cluster 0 and 376 in Cluster 1, while K-Means produces more evenly distributed clusters, with 512 objects in Cluster 0 and 434 in Cluster 1.
- Correlation between Reviews and Total Earnings: Both clustering methods show similar correlations between Reviews and Total Earnings. Hierarchical Clustering yields correlations of ~0.503 for Cluster 0 and 0.815 for Cluster 1, while K-Means shows correlations of ~0.5029 and 0.8152, indicating consistency in the relationship between these variables across both methods.

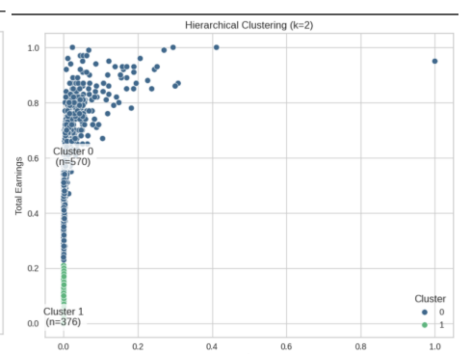
Overall, Hierarchical Clustering is preferred due to its better separation and more distinct clusters, while K-Means provides a more balanced cluster distribution but with slightly more overlap.

RESULTS AND DISCUSSION

- Is freelancing income sustainable as a career path?

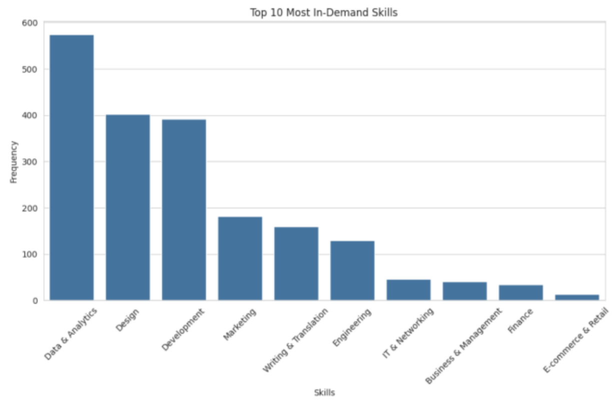
  Yes, based on hypothesis 1, 2 which is tested by random forest regression, the freelancer income is highly predictable ($R^2 > 0.93$) based on skill set, hourly rate, and r**ating**, suggesting sustainability *if* these factors are optimized.





- Which regions/countries have the highest concentration of skilled freelancers?
  Freelancers are distributed across many countries, with the highest concentration in **India (309 freelancers) and Pakistan (160 freelancers).**

- What are the most in-demand skills in the freelancing market?



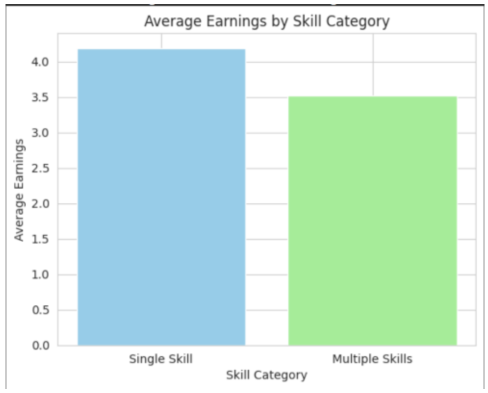
  The bar chart shows that Data and Design lead the freelancing market, showing strong demand for data and creative skills. Development and Marketing also remain key, while Writing, Engineering, Business, Finance, and Retail highlight the market's diversity.

- What is the average income for freelancers across different skill sets?

```
Skills
AI/ML               5.234862
Content Writing     4.706250
Data Analysis       5.169892
Graphic Design      5.207692
Web Dev             4.885263
Name: Total Earnings, dtype: float64
```

  Done using statistical and mathematical calculations.

- Is there a positive correlation between having multiple skills and higher earnings, or do specialized freelancers perform better?
  specialized freelancers perform better in earning with slight difference.



CONCLUSION AND FUTURE WORK

Conclusion

In this study, we analyzed the freelancers' skill sets, hourly rates, ratings, and reviews to understand their overall performance. The results from regression models such as Linear Regression, Random Forest, and XGBoost revealed that factors like reviews, ratings, and hourly rates significantly influence earnings performance. Additionally, clustering analysis helped identify distinct groups of freelancers, providing insights into which profiles tend to be more successful. Clustering revealed patterns that highlighted how freelancers with higher numbers of reviews attract more clients, contributing to better earnings performance. On the other hand, regression models quantified the relationship between these factors and earnings. This combination of approaches reinforces the importance of building a strong reputation and client trust in the freelance market.

Future Work

Future research could involve using actual income data, as opposed to earnings scores, for a more precise understanding of freelancer income. Expanding the dataset with features such as years of experience, client feedback, and geographical location could improve model predictions. Moreover, more advanced clustering that could provide richer insights. Long-term analysis or time-series forecasting models could help predict freelancers' earnings trajectories more accurately, and exploring deep learning techniques could enhance the overall prediction power of the models.