Project Report

Data Exploration and Analysis of Upwork Job Market

Submitted by: Team Members

Hend Khaled

Mario Habib

Rowaina Reda

Suhaila Ahmed

Supervisor: Dr. Eman Raslan

Co-Supervisor: Eng. Yomna Ehab Gamal

Date: March 2025

1. Abstract

This project aims to analyze the job market on Upwork by extracting, processing, and visualizing data from various job postings. By leveraging web scraping techniques, we collected data from different categories, including AI, JavaScript, Android Development, and Data Analysis. The project involved multiple stages, such as data cleaning, encoding, and exploration, to derive meaningful insights into hiring trends, required skills, and market demand. Our findings provide valuable insights for freelancers and businesses looking to understand the current job market landscape on Upwork.

2. Acknowledgments

We would like to express our sincere gratitude to all team members for their dedication and contributions throughout this project.

A special thanks to ChatGPT for its valuable assistance, providing guidance and insightful suggestions throughout the research and analysis process.

We extend our deep appreciation to **Eng. Yomna Ehab Gamal** for her continuous support, instructions, and valuable feedback that helped refine our approach.

Our heartfelt thanks to **Dr. Eman Raslan** for her clear explanations and in-depth teaching of the subject, which greatly enhanced our understanding.

Finally, we are immensely grateful to our **ITI team** for their unwavering support, collaboration, and encouragement, which made this project a truly enriching experience.

Table of Contents

3. Introduction

Upwork is one of the leading freelancing platforms, connecting businesses with professionals across various domains. This project aims to explore hiring trends and skills demand by analyzing job postings across different tracks. The study provides insights into the most sought-after skills, job durations, and employment preferences.

Background: Upwork is one of the leading freelancing platforms, connecting businesses with professionals across various domains. With the increasing demand for remote work, understanding hiring trends and job market dynamics is essential for both freelancers and employers.

Problem Statement: Freelancers often struggle to identify in-demand skills and optimize their profiles to match market needs. Similarly, businesses face challenges in hiring the right talent. This study aims to bridge this gap by analyzing job postings to extract meaningful insights about hiring trends, skill demand, and job characteristics.

Objectives:

- To extract and analyze job postings from Upwork across different categories.
- To identify key hiring trends, including job durations, experience levels, and work hours.
- To determine the most in-demand skills for freelancers in various fields.
- To provide a data-driven perspective on how freelancers can enhance their profiles to align with market needs.

Scope:

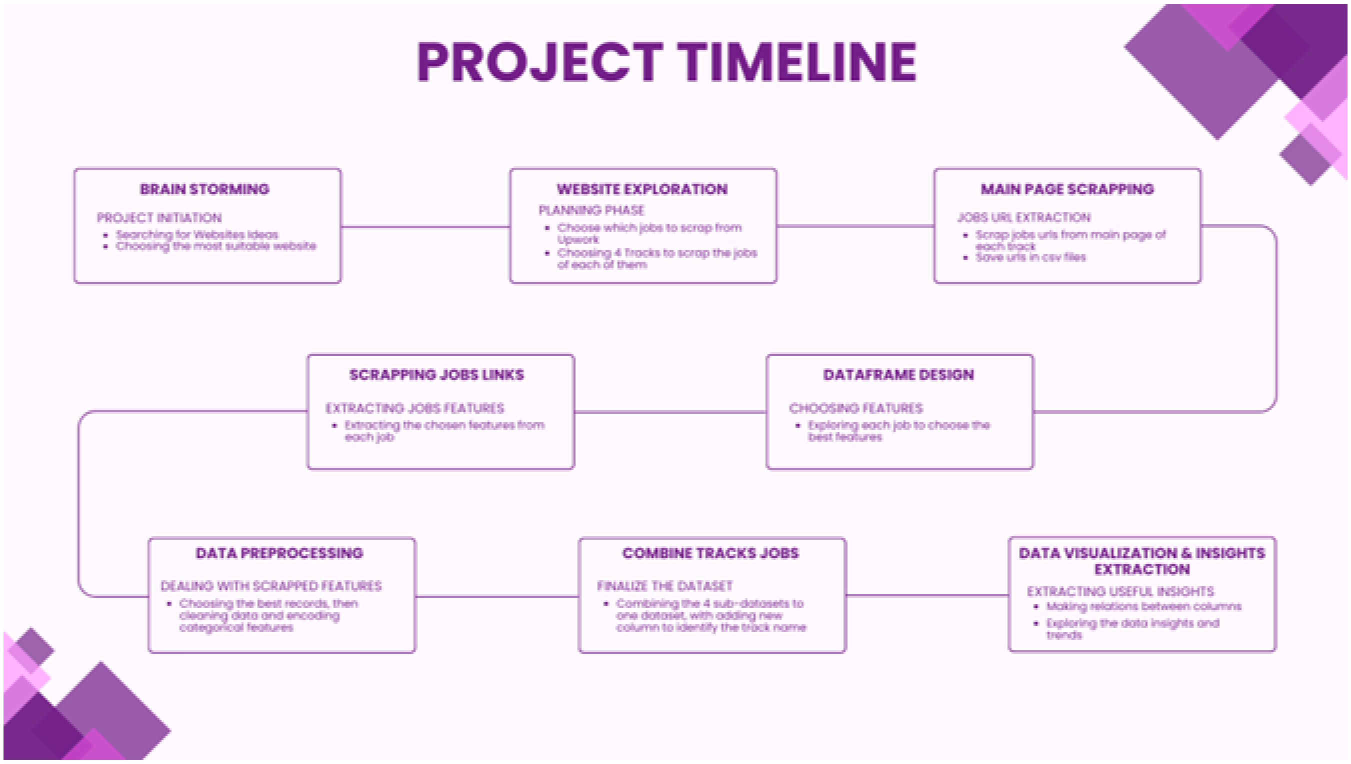
- **Included:** The study focuses on job postings from Upwork in AI, JavaScript, Android Development, and Data Analysis. It covers aspects such as job durations, required experience levels, skill demand, and hiring trends.
- **Excluded:** The study does not include job postings from other freelancing platforms or analyze employer-specific hiring behaviors beyond Upwork.

Methodology Overview:

- Web scraping techniques were used to extract job postings from Upwork.
- Data preprocessing steps, including cleaning, encoding, and handling missing values, were applied.

- Exploratory Data Analysis (EDA) was conducted to identify trends and insights.
- Data visualization techniques were used to represent findings in an interpretable manner.

Project Pipeline:



4. Project Workflow

Tasks were distributed among team members for efficiency, covering brainstorming, web scraping, data design, and exploration.

5. Literature Review

Summary of Previous Research: Research on freelancing platforms has emphasized the growing demand for digital skills. Studies suggest that job postings on platforms like Upwork reflect broader industry trends.

Relevant Theories and Models:

- Labor Market Theories: These helps explain job demand fluctuations.
- Web Scraping Techniques: Used to extract job postings.
- Data Mining & EDA: Applied to uncover patterns in hiring trends.

How This Project Differs:

While previous studies have analyzed freelancing trends, this project specifically focuses on Upwork and uses real-time data to generate up-to-date insights.

6. Methodology

Data Collection: The dataset was gathered using web scraping techniques from Upwork job postings. The following tracks were analyzed:

- AI
- JavaScript
- Android Development
- Data Analysis

Each team member contributed to brainstorming website ideas and executing the scraping tasks.

Tools & Technologies: The project utilized the following technologies:

- **Web Scraping:** Python (BeautifulSoup, Selenium)
- **Data Processing:** Pandas, NumPy
- **Data Visualization:** Matplotlib, Seaborn
- **Version Control:** GitHub

Algorithms & Techniques:

- Data Cleaning and Handling Missing Values
- Encoding Categorical Variables
- Exploratory Data Analysis (EDA)
- Data Visualization Techniques

7. Data Processing

Data preprocessing involved the following steps:

- **Cleaning and Handling Missing Values:**
    - The goal was to retain as many clean records as possible while ensuring each job profile has an equal number of records.
    - Records with missing or duplicate job titles were dropped, as these were critical for further analysis. Missing values in job titles were often due to private jobs, which may share common titles across multiple job profiles.

Records Remaining: 3265

- **Encoding Categorical Variables:**
    - Categorical variables such as job titles were encoded using a structured schema to convert text-based data into numerical values.
- **Combining Data Frames from Different Tracks:**
    - Multiple data frames, each representing a different job profile or track, were merged to create a unified dataset for analysis.

Challenges and Solutions:

- **Challenge 1: Data Consistency and Handling Large Datasets**
    - Issue: Ensuring uniformity across records and managing large amounts of data, especially when dealing with missing values and duplicates.
    - Solution: By cleaning and preprocessing the data efficiently, we ensured consistency and reduced the dataset size for further analysis.
- **Challenge 2: Extracting Relevant Information**
    - Issue: Identifying useful features and dealing with missing or unstructured data.
    - Solution: Through structured scraping and preprocessing techniques, we extracted relevant features and handled missing data accordingly.

Detailed Data Processing Steps:

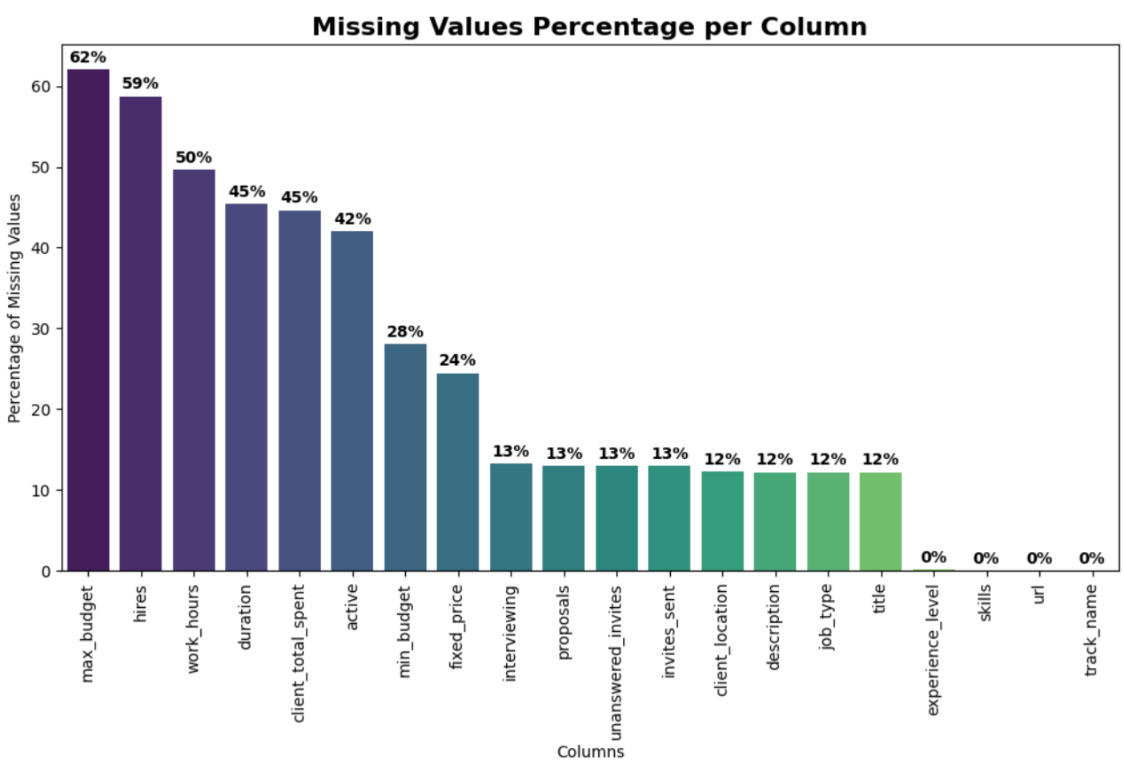
- **Missing and Duplicated Job Titles:**
  - We dropped records with missing or duplicate job titles, ensuring the dataset contained only the relevant job information.
- **Converting Budget Values:**
  - The **min_budget** and **max_budget** columns were in string format with symbols like $ and ,. These were removed, and the columns were converted to numerical values (floats).
  - **Note:**
    - The **min_budget** column represents the minimum budget for both Hourly and Fixed-Price jobs.
    - The **max_budget** column holds the maximum budget for Hourly Jobs and NaN values for Fixed-Price jobs, which explains the significant number of NaNs in this column.
- **Creating the Average Budget Column:**
  - We created a new **average_budget** column by calculating the average of **min_budget** and **max_budget**.
  - If **max_budget** is NaN (Fixed-Price job), the **min_budget** was used as the average.
  - After creating the **average_budget** column, both **min_budget** and **max_budget** columns were dropped, as they were no longer necessary.
- **Handling NaNs in Duration & Work Hours:**
  - Both **Duration** and **Work Hours** columns contained 1214 NaN values, mostly due to Fixed-Price jobs not having determined duration or working hours.
  - We filled the NaNs in **Duration** with the value Contract and in **Work Hours** with Flexible, ensuring consistency in these columns.
- **Converting Salary Values:**
  - We removed the $ and, symbols from the salary values.

- For values with ‘K’ (thousands) or ‘M’ (millions), we converted these to float and multiplied them by 1000 or 1,000,000, respectively. For other cases, the value was simply converted to a float.

- **Addressing Client Location Inconsistencies:**
  - Some client locations contained country abbreviations, such as "US" for United States or "UK" for United Kingdom.
  - We mapped these abbreviations to the full country names, ensuring consistent formatting across the dataset and improving the accuracy of the location analysis.

Number of Countries before Mapping: 159
Number of Countries after Mapping: 105

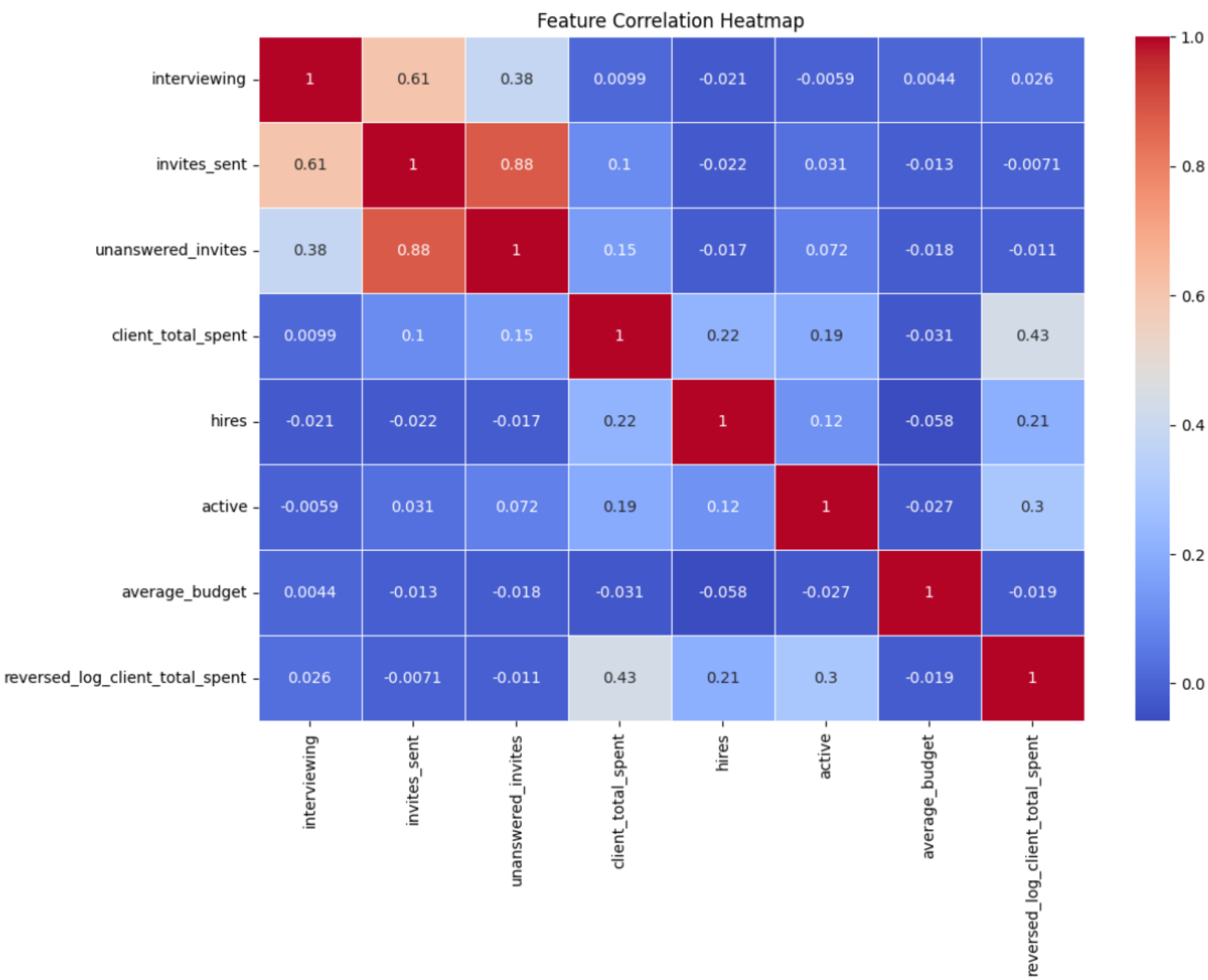
- **Selecting Final Records:**
  - To maintain balance across job profiles, we selected 2400 records in total, with 600 records per job profile.
  - Records with missing values in **hires**, **active**, **min_budget**, or **fixed_price** were dropped to ensure the dataset contained only complete and meaningful entries.



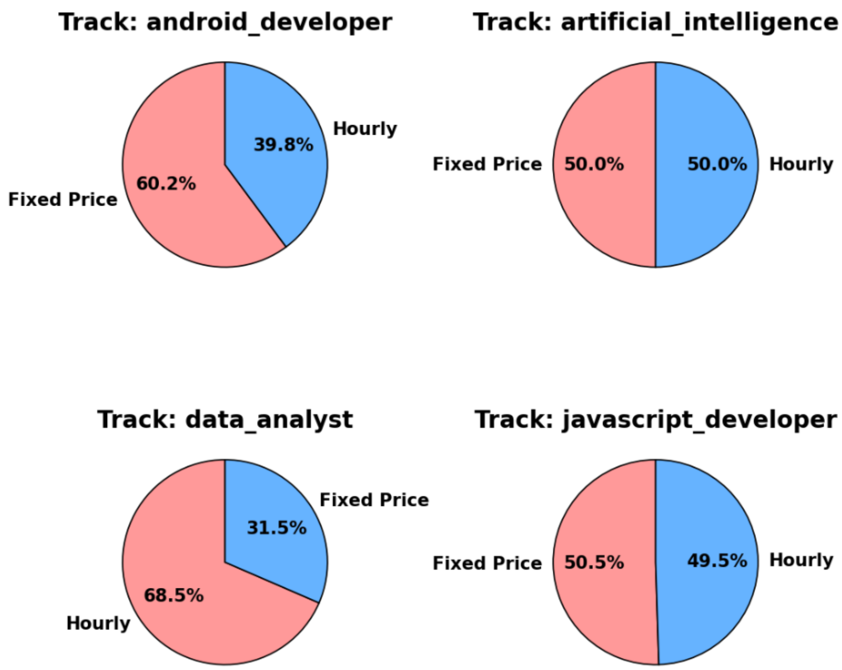
8. Results & Discussion

Findings

- **Correlation Between Features**:
  After analyzing the "interviews", "invites sent", and "unanswered invites" columns, it appears that most values in these columns are zeros, which creates a false impression of correlation between these columns.
  - **Insight**: There is no actual correlation between these columns when examined individually. The presence of zeros in the data distorts the interpretation.



- **Job Type Split Between Hourly & Fixed-Price**:
  The pie chart reveals that job types are evenly split between "Hourly" and "Fixed-Price".
  - **Insight**: Job types are equally split for AI engineers and JavaScript developers. For Android developers, the split is 60% Fixed-Price and 40% Hourly. For data analysts, 70% are Hourly and 30% Fixed-Price.

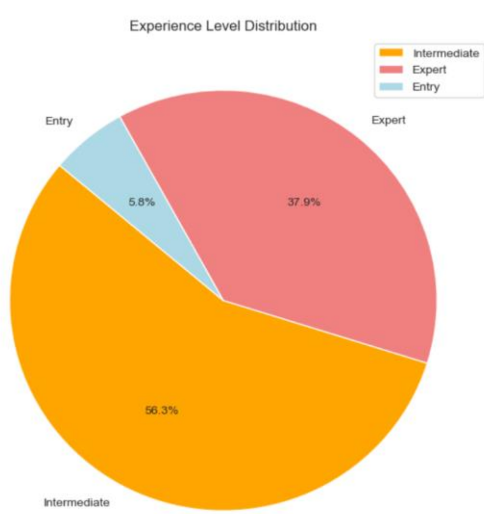


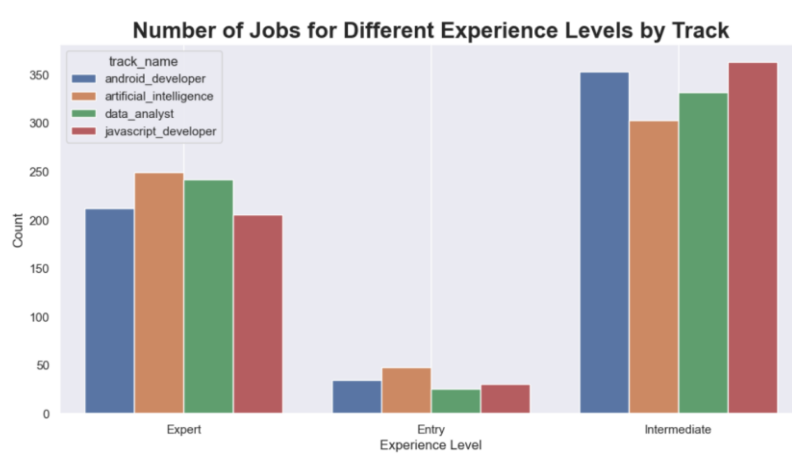
- **Relation Between Experience Level, Job Types & Job Profiles**:
  It was found that most jobs are for Intermediate levels, followed by Experts, and Entry-level positions have the fewest job postings.

  - **Insight**: Entry-level positions consistently have fewer available jobs across different job types, while Intermediate and Expert levels tend to have more available jobs, with Intermediate levels having the most postings.

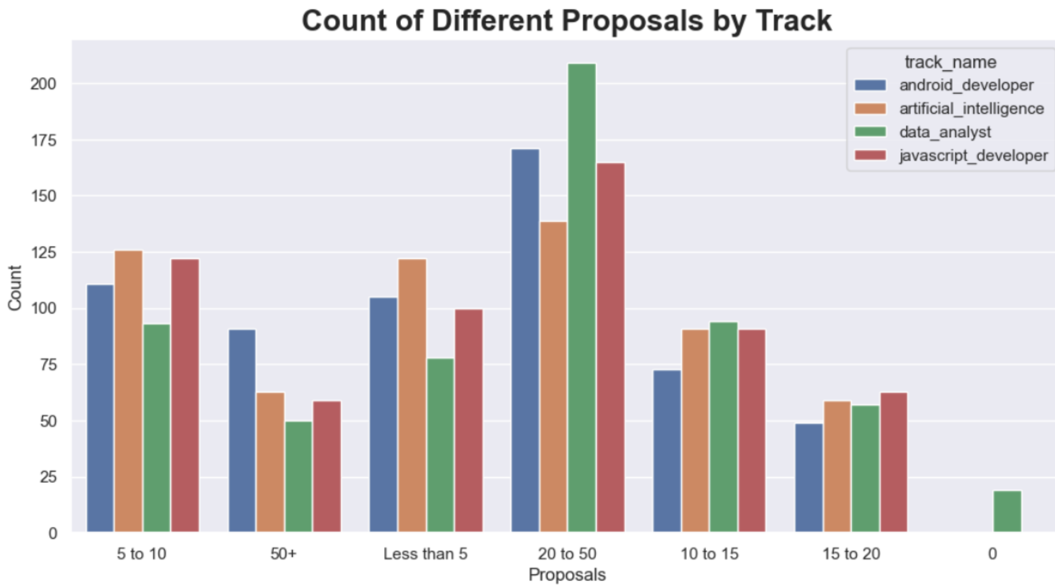
  - For Entry-level positions, Fixed-Price jobs outnumber Hourly jobs, while for Intermediate and Expert levels, Hourly jobs slightly surpass Fixed-Price jobs.

  - **Insight**: Having at least an Intermediate experience level significantly increases the number of job opportunities.

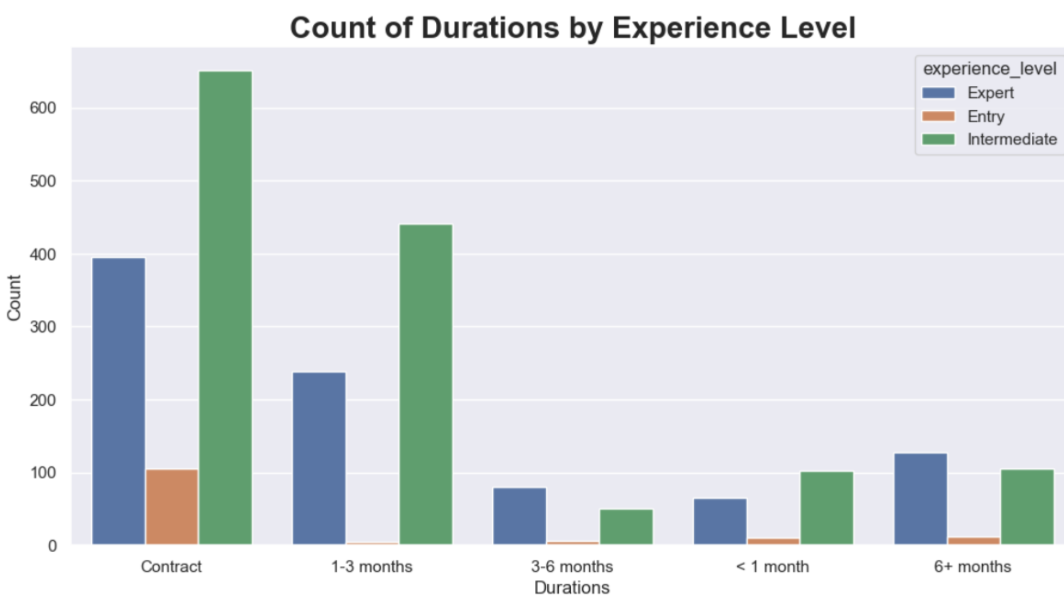




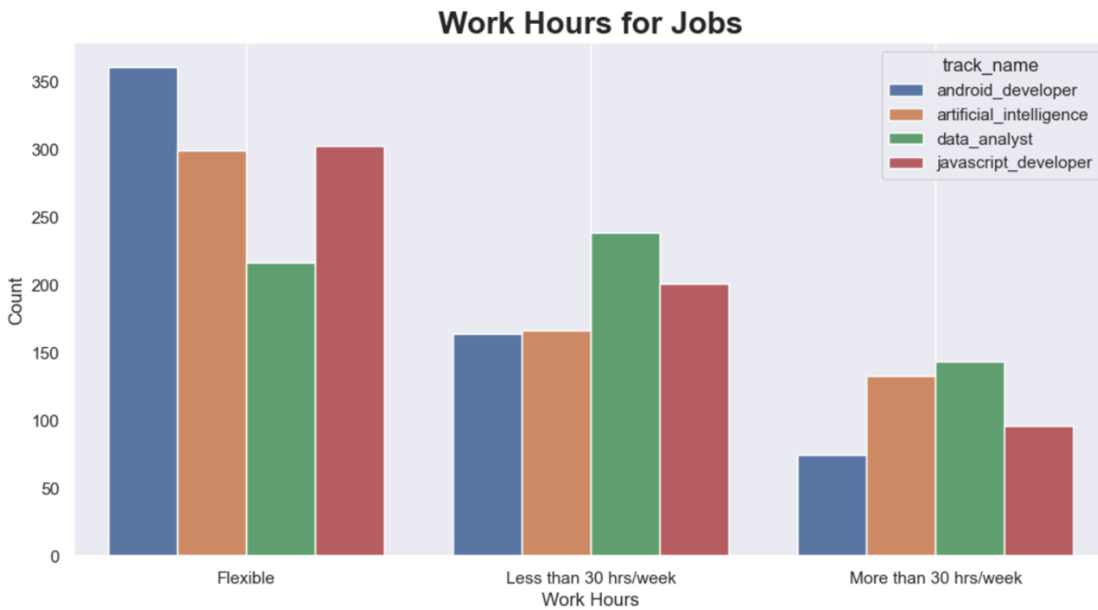
- **What Affects the Number of Proposals a Job Gets?**
  Most jobs receive between 20-50 proposals, regardless of job profile.
    - **Insight**: Hourly jobs tend to get more proposals than Fixed-Price jobs.



- **Does the Duration of Job Differ by Job Profile and/or Required Experience Level?**
  Most Hourly jobs require 1-3 months of work. Many Data Analyst jobs require a duration of 6+ months.
    - **Insight**: Entry-level jobs are predominantly contractual. For Intermediate and Expert levels, most jobs are either contractual or require 1-3 months of work.

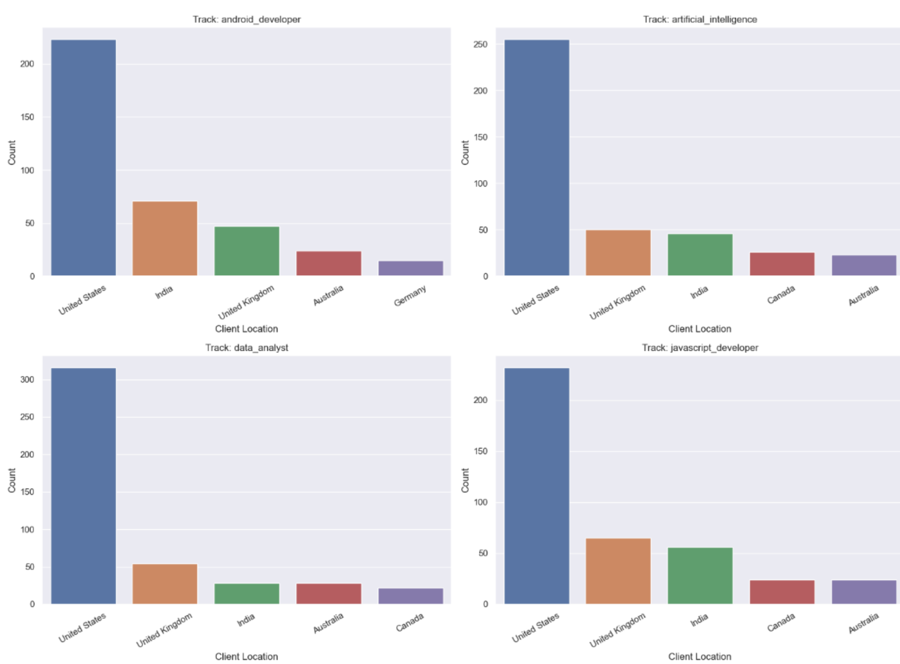


- **Do Work Hours Differ Depending on Job Profiles?**
  Most Hourly jobs require less than 30 hours per week.
  - **Insight**: Data Analysts have more Hourly jobs than Fixed-Price jobs. Android developers tend to have more Flexible work hours.

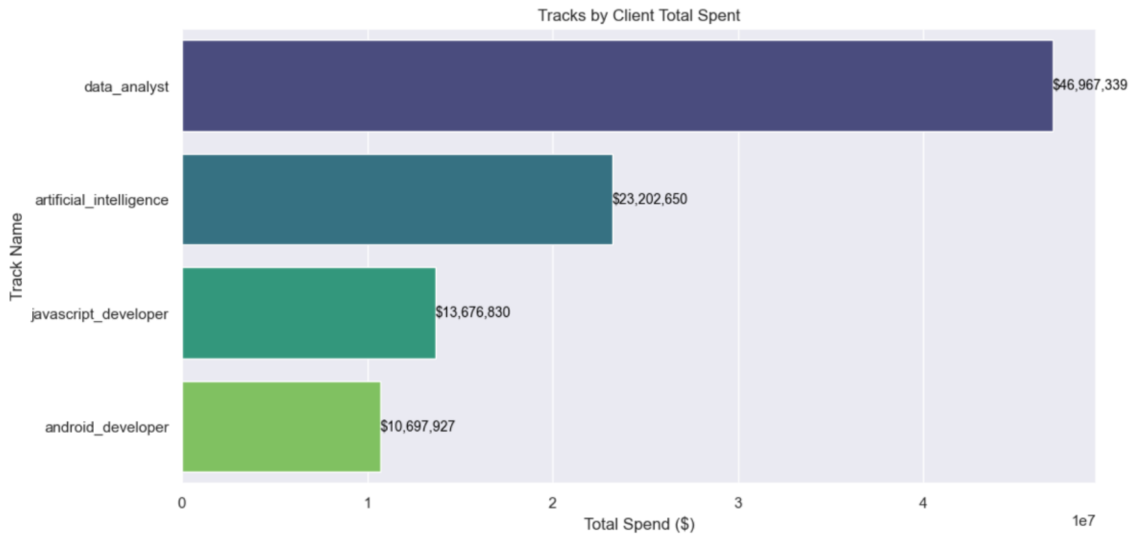


- **Are There More Clients from Certain Countries?**
  The top three client locations for all job profiles are:
  - United States
  - United Kingdom
  - India

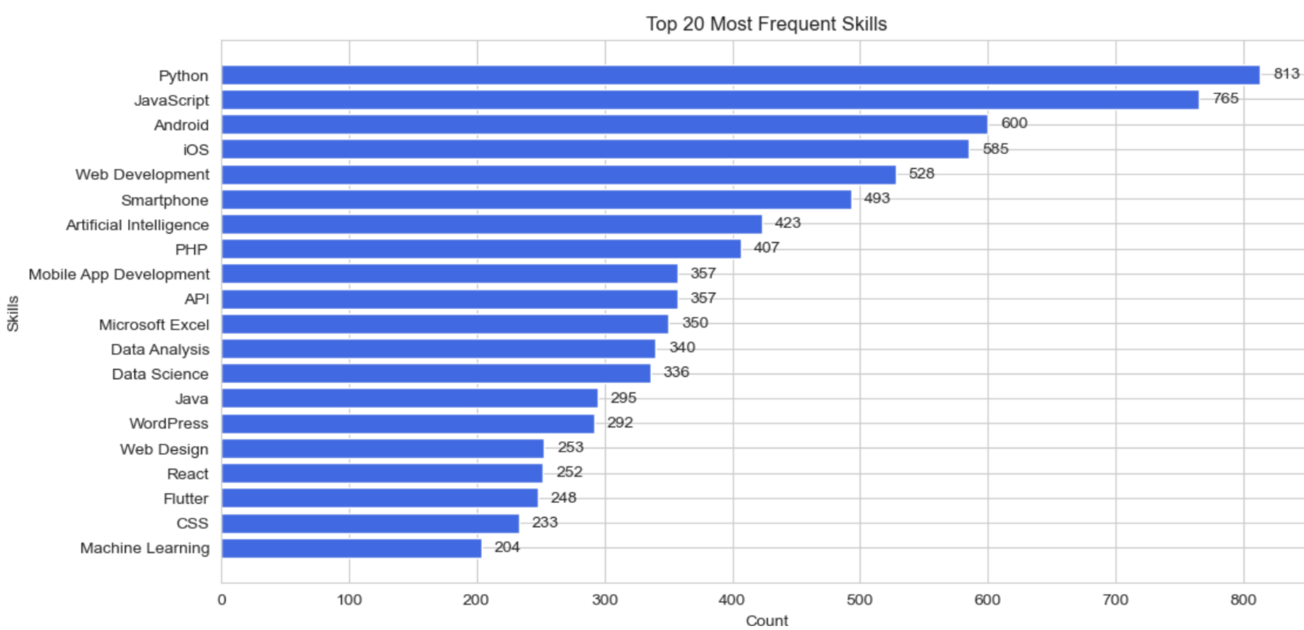
  The fourth and fifth places vary between Australia, Canada, and Germany.



- **By How Much Do Tracks Vary in Client Total Spending? By How Much Does Average Budget Vary?**
  Analyzing the mean and median budgets for Hourly and Fixed-Price jobs shows that there are significant outliers in the Fixed-Price category.
    - **Insight**: The Fixed-Price category is heavily skewed left with a few large outliers to the right. These outliers could be attributed to larger teams working on these projects rather than individuals.



- **By How Much Do Top Required Skills Differ for Each Job Profile?**



9. Challenges & Limitations

- **Challenges Faced**:
  The primary challenges in the project included missing data (zeros) in certain columns that made it difficult to accurately assess correlations. Additionally, the lack of data on time-related aspects (such as job start/end dates) and detailed information about the hiring process also posed challenges in drawing more comprehensive insights.

- **Constraints**:
  - No access to time data, making it difficult to analyze job duration trends more precisely.
  - Lack of information on the internals of the hiring process, which would have provided a better understanding of why certain job postings received more proposals than others.
  - No data about applying freelancers, limiting the scope of analysis regarding the competition for these jobs.

---

10. Conclusion and Future Work

- **Conclusion**:
  The analysis provided valuable insights into Upwork job trends, particularly in terms of job types, experience levels, and proposal patterns.
  Key takeaways:
  - Having at least an Intermediate experience level increases the chances of finding a job.
  - Hourly jobs tend to attract more proposals compared to Fixed-Price jobs.
  - AI Engineers should focus on NLP, Chatbot Development, APIs, and Model Deployment, while Data Analysts should emphasize Business Intelligence and Finance.

- **Future Work**:
  Future work may involve expanding the dataset to include more job profiles, tracking trends over a longer period, and implementing predictive models to forecast job demand.
  - **Suggestions for Improvements**:
    - Expanding the scope of the dataset to include more regions and freelance profiles.
    - Incorporating time-related data to better analyze job durations and proposal trends.
    - Implementing machine learning models to predict job proposal rates and success.

---

11. References

1. **Upwork**. (2023). *Upwork Job Postings Data*. Retrieved from https://www.upwork.com.
2. **Kaggle**. (2022). *Upwork Freelance Job Data*. Kaggle. Retrieved from https://www.kaggle.com.
3. **Pandas Documentation**. (2023). *Pandas: Python Data Analysis Library*. Retrieved from https://pandas.pydata.org/pandas-docs/stable/.
4. **Seaborn Documentation**. (2023). *Seaborn: Statistical Data Visualization*. Retrieved from https://seaborn.pydata.org.
5. **Matplotlib Documentation**. (2023). *Matplotlib: Python Plotting Library*. Retrieved from https://matplotlib.org.
6. **Scikit-learn Documentation**. (2023). *Scikit-learn: Machine Learning in Python*. Retrieved from https://scikit-learn.org.
7. **OpenAI**. (2023). *Exploratory Data Analysis: An Overview of Methods and Best Practices*. Retrieved from https://www.openai.com.
8. **Cheng, M., & Li, H.** (2020). *Trends in the Freelance Economy: An Analysis of Upwork and Freelancer Job Markets*. International Journal of Business and Management, 15(2), 40-50.
9. **Wang, X., & Zhang, Y.** (2019). *Predicting Job Success in Freelance Marketplaces*. Proceedings of the International Conference on Data Science, 123-134.
10. **Hadley Wickham, & Garrett Grolemund**. (2017). *R for Data Science*. O'Reilly Media.