

# Web Scraping and Analysis - Wuzzuf Jobs in Egypt

## YAS-timated Errors Team

### 1. Introduction

In this project, we scraped all job listings from [Wuzzuf.net](https://wuzzuf.net), a leading Egyptian job portal. We extracted information such as job title, company, publication time, location, job description, experience level, experience years, job type, and work shift.

The final cleaned dataset helped us explore broader job market trends in Egypt.

### 2. Data Extraction

- **Total Scraped Jobs:** 1960 listings
- **Initially Captured Fields:**
  - job\_title – Position title

company – Employer name

- location – Full address
- job\_skills – Mixed string with job type, experience, and skills
- jobs\_type – Work mode (e.g., Full TimeRemote)
- published\_time – Relative post date (e.g., “4 days ago”)
- **These fields were semi-structured, especially job\_skills, which combined multiple pieces of information in a single string.**

### 3. Data Cleaning

Post-scraping, we applied several cleaning steps using pandas and regex to ensure consistency and usability:

- **Dropped duplicates and null value**
- **Normalized text** (e.g., converting to lowercase, trimming whitespace)
- **Standardized published\_time into datetime format** for consistency
- **Parsed job\_skills field using regex to extract:**
  - experience\_level (e.g., "Entry Level")
  - experience\_years (e.g., "1 - 3 Yrs")
  - job\_description (keywords/skills list)
- **Split location into: city and work\_location**
- **Standardized categorical values in jobs\_type** (e.g., Remote, Hybrid)

These steps prepared the data for accurate analysis and visualization.

### **Data After Cleaning:**

- job\_title – Job position title
- company – Employer name
- published\_time – Date of posting
- work\_location, city – Location details
- job\_description – Summary of role, skills, and qualifications
- experience\_level – Career level (Entry, Experienced, etc.)
- experience\_years – Required experience duration
- program – Work model (On-site, Remote, Hybrid)
- shift – Job type (Full-time, Part-time, etc.)

## **4. Key Findings and Insights**

### **3.1 Top Hiring Companies**

A standout observation is the dominance of "Confidential" postings (254 jobs), signaling that many employers prefer to remain anonymous.

Among public companies, Intouch CX (25 jobs) and ECC Group (16 jobs) emerge as the leading hirers, reflecting active recruitment or expansion.

Other names like Alshaya and GB Corp show moderate but notable hiring efforts.

**Insight:** Confidential hiring is highly prevalent, but notable companies in customer experience and retail are also actively growing.

### **3.2 Most Common Job Titles**

The Egyptian job market appears heavily skewed toward accounting roles, with "Accountant" dominating (73 job postings), far ahead of other titles.

**Insight:** Accounting remains a backbone of hiring needs, suggesting strong, stable demand for finance professionals

### **3.3 Experience Level Trends**

- **Experience Required:**
  - **Entry Level:** 52%
  - **Experienced:** 44%
  - **Senior Management:** 4%

**Insight:** Over half of the opportunities target fresh graduates and juniors, offering an open door for newcomers into the data and finance job market.

### 3.4 Required Experience Years

Most jobs target "Experienced" candidates (1473 postings), followed by a decent number of "Entry Level" roles (466). Senior Management opportunities are rare (48), and a noticeable portion of jobs (346) did not clearly define the experience level.

#### **Insight:**

The experience level breakdown suggests that the market favors those with prior experience, while also offering opportunities for newcomers.

### 3.5 Work Location Trends

- **Top Cities:**

- Cairo
- Giza
- Alexandria

#### **Insight:**

Greater Cairo dominates the job market, with over 65% of postings concentrated in Cairo alone, reflecting strong centralization of opportunities.

### 3.6 Job Program and Shift Patterns

- **Work Program:**

- On-site 2037
- Hybrid 160
- Remote 136

This suggests that the Egyptian job market (where the data is focused) still heavily favors traditional on-site work arrangements. While there's some presence of hybrid and remote options, they remain relatively uncommon compared to the traditional model.

- **Shift:**

- Full Time: 97.6%
- Part Time: 0.57%
- Freelance: 0.35%
- Internship: 1.48%

#### **Insight:**

The overwhelming majority of opportunities are Full-Time, reflecting the market's preference for permanent staffing. Internships, Part-Time, and Freelance roles are much rarer, highlighting limited flexibility for non-traditional employment seekers.

## 5. Data Visualizations

### 5.1 Top 20 Job Titles

- **Visualization:** A horizontal bar chart was created to display the Top 20 Job Titles based on the number of job postings.
- **Key Observations:**
  - The job market seems heavily tilted towards accounting-related roles, with "Accountant" leads in frequency.

### 5.2 Top Hiring Companies

- **Visualizations:**
  1. Bar Chart showing the Top 10 Hiring Companies by number of postings.
  2. Word Cloud created from company frequency data.
- **Key Observations:**
  - Confidential postings dominate significantly, reflecting employer anonymity in job listings.
  - Among named companies, Intouch and ECC group dominate, leading with the highest hiring activity.
  - The word cloud highlights the major employers, with larger font sizes indicating higher frequencies of job postings.

### 5.3 Job Types Distribution

- **Visualization:** A double pie chart was plotted:
  - First Pie Chart: Displays the overall distribution of job types.
  - Second Pie Chart: Focuses on smaller categories for clarity.
- **Key Observations:**
  - Full-Time roles make up the overwhelming majority of job postings.
  - Internship, Part-Time, and Freelance positions are comparatively rare, reflecting the traditional employment structure in Egypt's data market.

### 5.4 Experience Level Distribution

- **Visualization:** A horizontal bar plot was used to show the distribution of job postings by experience level, and a lollipop plot to show distribution of experience by level.
- **Key Observations:**
  - The majority of job postings are targeted at professionals with 2 to 5 years of experience, indicating demand for early-career professionals.
  - Experienced-level roles are the most demanded, but there's a surge in entry-level roles too.

## 5.5 Top Locations and Cities

- **Visualization:** A treemap was created to visualize the top job postings by city.
- **Key Observations:**
  - Cairo, Maadi, and Nasr City are the top work locations, with Cairo alone accounting for over 65% of all job postings.
  - Cairo's high job concentration confirms regional market centralization.

## 5.6 Top Skills and Tools

- **Visualization:** A Word Cloud and bar chart were created to show the most common skills and tools mentioned in job descriptions.
- **Key Observations:**
  - Engineering, Sales, Marketing, Accounting, and Finance were the most frequently mentioned hard skills.
  - Communication skills emerged as a key soft skill, with strong emphasis across job postings.

## 5.7 Job Postings Over Time

- **Visualization:** A line plot was used to visualize job postings over time, showcasing trends in job posting frequency.
- **Key Observations:**
  - The number of job postings shows a significant spike around 30 days ago, followed by a steady increase over time.

## 5.8 Sankey Diagram: Flow from Location → Program → Job Type

- **Visualization:** A Sankey plot was used to show the flow of job postings from location to program to job type.
- **Key Observations:**
  - The majority of job listings flow from locations like New Cairo and Cairo to "On-site" and "Full Time" job types, indicating a strong preference for traditional work arrangements.

## 6. MongoDB Storage

To ensure scalability and flexibility for future data exploration, the cleaned job listings dataset was stored in a MongoDB database.

### Steps Taken:

- Connected to MongoDB using the pymongo library in Python.
- Created a new database named "WuzzufJobs".
- Created a collection inside the database named "JobListings".
- Inserted all cleaned and structured job records into the collection as documents.

### Benefits:

- Allows efficient querying and retrieval of job listings based on multiple fields.
- Supports future updates or expansions if more jobs are scraped later.
- Enables integration with web apps or dashboards without needing to reload static files.

## 7. Streamlit App Development

An interactive web application was built using Streamlit to explore the job market insights in real-time.

### Features of the App:

- Overview page displaying key statistics (total jobs, top cities, top companies).
- Dynamic filtering of job listings by City, Experience Level, and Work Type.
- Interactive visualizations, including:
  - Top 20 Job Titles (Bar Chart)
  - Top Hiring Companies (Bar Chart and Word Cloud)
  - Job Types Distribution (Pie Chart)
  - Top Skills and Tools (Word Frequency Cloud)
- Search functionality to explore specific job roles.

### Purpose:

- Provides an easy-to-use interface for users to navigate the Egyptian job market.
- Allows dynamic exploration without needing technical skills or data knowledge.

### Streamlit app link:

<https://blank-app-c70gcgu4tzt.streamlit.app/>

## 8. Conclusion

In this project, we successfully completed the Web Scraping and Analysis of Wuzzuf Jobs in Egypt.

Through advanced scraping, cleaning, storing, and visualization techniques, we uncovered major patterns such as:

- The dominance of Accounting, Sales, and Customer Service roles.
- A job market that heavily favors Full-Time, On-Site work.
- A strong demand for Entry-Level and Early-Career professionals.
- A clear concentration of opportunities in Greater Cairo.
- An essential mix of technical and soft skills across job postings.

By storing the final dataset in MongoDB and building an interactive Streamlit app, we made the project scalable, flexible, and user-friendly for future use.

This project not only demonstrated our technical skills in web scraping and data analysis but also our ability to transform raw web data into actionable insights for real-world decision-making.

### Team Members:

- Yassin Mohamed Hassan Elsayed      23011622
- Abdelrhman Khaled Mohamed      23011315
- Selvia Nasser Bekhit Mansour      23011293