

Ghulam Abbas

Data Engineer

(+92) 336 8459456 • LinkedIn://[Abbas](#) • Gmail:// [Abbas](#)

ABOUT ME

I am a software professional with extensive knowledge of data methodologies, skilled at thriving in fast-paced agile environments. With three years of experience in data scraping, analysis, and modelling, I have successfully delivered results for multiple teams across the organization. I am committed to achieving objectives and promoting effective collaboration across teams.

EXPERIENCE

WALEE

Onsight / Islamabad, Pakistan

Data Engineer

(DEC 2021 – Present)

A highly motivated Web Scraping Professional with 3+ years of experience extracting data from various websites. Expertise includes developing scripts for data collection, creating and managing data pipelines, and utilizing advanced web scraping tools. Proficient in working with APIs, HTML, and CSS to extract and analyse data effectively. Possesses strong analytical and problem-solving skills, with a proven ability to streamline data collection methods and minimize analysis errors.

- Automated repetitive tasks using programming languages such as Python and R, significantly reducing the time during the analytical process.
- Designed and implemented scripts to extract data from multiple websites while maintaining robust data pipelines.
- Developed web scraping solutions to extract, clean, and store data in databases for further analysis.
- Analysed extracted data to identify patterns and trends, presenting findings in clear and actionable reports.
- Optimized web scraping processes to enhance the speed and accuracy of data extraction.
- Created and maintained a comprehensive data pipeline for data extraction, cleaning, and storage.
- Leveraged tools like BeautifulSoup and Scrapy to efficiently scrape data from websites.
- Utilized APIs, HTML, and CSS to accurately locate and extract website data.

WEBLIFY

Onsite / Gulzar-e-Quaid, Pakistan

Data ENGINEER

(Dec 2020 – Dec 2021)

- Successfully developed and executed web scraping scripts to collect and organize data from various US-based hospital websites.
- Extracted patient-related details essential for medical billing processes, ensuring accuracy and compliance with relevant standards.
- Utilized advanced tools and technologies such as Python, Scrapy, and Selenium for data extraction and processing.

- Processed large datasets to generate structured formats, including JSON, CSV, and Excel, for seamless integration into medical billing systems.
- Focused on HIPAA-compliant practices to maintain the confidentiality of sensitive patient information.

HA LINKS

Rawalpindi, Pakistan

Data Scraper

(Dec 2019 – Dec 2020)

I previously worked as a data scraper, specializing in extracting call centre-related data such as phone numbers, addresses, emails, and other essential information. This data was organized and provided in various formats, including Excel and CSV files.

TECHNICAL SKILLS

Languages: Python, SQL, Bash Scripting.

Web Scraping and Data Extraction:

- Proficient in web scraping using Python libraries like **Scrapy**, **Beautiful Soup**, and **Selenium**.
- Experience with handling **pagination**, **dynamic content**, and **JavaScript-heavy websites**.
- Expertise in extracting structured and unstructured data into formats like **JSON**, **CSV**, **Excel**, and **databases**.
- Skilled in bypassing protections like **CAPTCHA** and **rate limiting** using tools like **Proxies** and **User Agents**.

Programming:

- Proficient in **Python** (including Object-Oriented Programming and script development).
- Knowledge of **SQL** for querying and manipulating databases.
- Experience in **bash scripting** for task automation.

Data Engineering:

- Building **ETL pipelines** to process and transform large datasets.
- Proficient in working with **databases** (MySQL, PostgreSQL, MongoDB).
- Knowledge of **data warehouse** tools like AWS Redshift or Google Big Query.
- Familiar with **Big Data tools** like Hadoop or Apache Spark.

APIs:

- Expertise in working with **REST APIs** and handling **JSON/XML data**.
- Building custom integrations using APIs.

Cloud Platforms:

- Experience in deploying projects on **AWS** (e.g., EC2, S3, Lambda) or **GCP**.

Automation Tools:

- Proficient in using **Python automation frameworks** for repetitive tasks.

- Knowledge of **Airflow** or other workflow automation tools.

Version Control:

- Proficient in using **Git** for version control and collaboration.

Soft Skills

- Strong analytical and problem-solving skills.
- Attention to detail in data handling and reporting.
- Excellent communication and team collaboration.

PROJECTS

Developed the **Data2U Tool** for extracting and processing data from various websites, including **Schneider Electric** and **Siemens**.

Scraped data from prominent social websites such as **Glassdoor** and **Indeed**, focusing on extracting relevant information for analysis.

AWARDS & CERTIFICATIONS

[Python 3.7 - Programming Language Basics](#) (Udemy): 2024:

- Acquired skills in automating real-world tasks using Python, focusing on scripting, data processing, and task automation, instructed by Freedom 10x Ventures.

Python Internship / Data Science / Dec 2020 - Dec 2021, Islamabad

Ghulam Abbas pursued a Python internship, mastering data science concepts, web scraping, pipelines, and modelling. He completed an in-depth course covering Python from basics to advanced levels, acquiring hands-on experience with web scraping tools and technologies. Though he did not follow a traditional software degree path, Ghulam achieved expertise in Python through dedicated learning and practical application, building a strong foundation in programming and data-driven solutions.

EDUCATION

502 Model College, Lalazar Rawalpindi

(2014 – 2016)

Intermediate

National University of modern Languages

(Feb 2017 – Mar 2021)

Pursued Peace and Conflict Studies at NUML University, gaining in-depth knowledge of international law and social sciences. Enhanced understanding of global issues and conflict resolution. Acquired valuable social and cultural insights through interactive learning and collaboration.

LANGUAGES

English (**B2**), Urdu(**C2**).