Case Study: Automated Web Scraping & Job Market Data Analysis

From Dynamic Websites to Structured Business Intelligence

Tools Used: Python • Jupyter Notebook • Requests • BeautifulSoup • Pandas • Matplotlib

Project Overview

This project demonstrates a complete web data extraction and analysis workflow using real job market data collected from the Profession.hu job portal. The objective was to automatically gather, clean, and structure unstructured HTML content and transform it into a usable analytical dataset.

Web scraping is a critical capability for companies that rely on market intelligence, competitive monitoring, pricing research, recruitment analytics, and trend tracking.

Business Problem

Many valuable business insights are locked inside websites that do not provide structured data exports. The goal of this project was to replace manual data collection with a fully automated pipeline capable of continuously extracting large volumes of job market information for analysis.

This enables faster decision-making, better labor market visibility, and scalable data acquisition.

Data Collection & Automation

The scraping process was built using automated HTTP requests and HTML parsing. Multiple job listing pages were programmatically accessed, and relevant job attributes such as job title, company, location, and posting details were extracted.

The scraper was designed to be reusable and scalable for continuous or scheduled data collection.

Data Cleaning & Structuring

Raw HTML outputs were transformed into a clean, structured dataset. This included removing irrelevant text, handling missing values, standardizing fields, and converting raw strings into analysis-ready formats.

The final dataset was stored in a tabular format suitable for analytics and reporting.

Exploratory Data Analysis (EDA)

Exploratory analysis was performed on the scraped job data to identify trends in demand, geographic distribution, and job posting patterns. Visualizations were created to uncover emerging market signals.

This allows stakeholders to quickly understand labor market dynamics based on real-time web data.

Visualization & Reporting

Key trends and distributions were summarized through professional charts and tables. These outputs can be directly integrated into business reports, dashboards, or labor market monitoring tools.

Interpretation & Business Insights

The scraped and analyzed data was translated into actionable insights such as high-demand job categories, location-based hiring trends, and posting frequency patterns. These findings provide direct value to recruitment firms, HR departments, and market analysts.

Skills Demonstrated

- Automated Web Scraping
- HTML Parsing & Data Extraction
- Data Cleaning & Structuring
- Exploratory Data Analysis (EDA)
- Trend Detection
- Data Visualization
- Business Insight Interpretation
- Reusable Data Collection Pipelines

Client Value

- Automated replacement of manual data collection
- Real-time access to market intelligence
- Scalable competitive and labor market monitoring
- Structured datasets ready for analytics
- Custom scraping solutions for ongoing use

Summary

This case study highlights the ability to extract valuable business data directly from websites and convert it into structured, decision-ready intelligence. The project demonstrates how automation, data engineering, and analytics can be combined to deliver continuous market insight at scale.