Case Study: Web Scraping & Market Mapping of Czech Beauty Businesses

From Unstructured Listings to Structured Market Intelligence

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

Project Overview

This project demonstrates an end-to-end automated web scraping and data structuring workflow focused on the Czech beauty services market. The objective was to collect, clean, and organize business listing data from multiple online sources and transform it into a structured dataset suitable for market analysis.

This type of solution is widely used in local market intelligence, competitive analysis, lead generation, and location-based business research.

Business Problem

Beauty and wellness businesses are often listed across fragmented, unstructured web sources with no centralized database available. The goal of this project was to automate the discovery and extraction of beauty service provider data to replace manual research with a scalable data collection solution.

Data Collection & Automation

Automated scraping scripts were built to systematically crawl listing pages and extract key attributes such as business name, service type, city, contact details, and online presence.

The pipeline was designed for repeatable execution, enabling ongoing data collection and updates.

Data Cleaning & Structuring

Raw scraped content was cleaned, standardized, and converted into structured tabular format. This included handling duplicates, correcting inconsistent text fields, filling missing values, and validating geographic information.

Exploratory Data Analysis (EDA)

Exploratory analysis was performed to understand the geographic distribution of beauty businesses, market density by city, and dominant service categories.

Key descriptive statistics and visualizations were produced to support market sizing and competitive landscape assessment.

Interpretation & Market Insights

The structured dataset enabled insights into regional competition levels, business concentration, and potential market gaps. These findings can directly support business expansion, franchise planning, and targeted marketing strategies.

Skills Demonstrated

- Automated Web Scraping
- HTML Data Extraction
- Data Cleaning & Standardization
- Duplicate Detection & Removal

- Exploratory Data Analysis (EDA)
- Market Mapping
- Data Visualization
- Reusable Scraping Pipelines

Client Value

- Fully automated market data collection
- Replacement of manual market research
- Up-to-date competitive landscape overview
- City-level market density insights
- Scalable solution for continuous monitoring

Summary

This case study demonstrates the ability to transform fragmented, unstructured business listings into a clean, searchable market intelligence dataset. By combining automation, data engineering, and analytical thinking, the project delivers actionable insights for localized business strategy and expansion planning.