Product Requirements Document (PRD)

Project Title: Data Scraper From Web
Date: July 29, 2025

1. Overview

1.1 Purpose
The purpose of this project is to develop a Python-based web scraping tool to extract structured data from the targeted directory. The tool will automate data collection, allowing for easy access to details for further analysis or integration.

1.2 Scope
1.2 Scope
The scraper will:

- Access the main store directory
- Navigate through U.S. states and cities
- Collect all store names under each city
- Structure the output with only three columns:
    - State
    - City
    - Stores (comma-separated names of all stores in that city)
- Export the data into CSV and JSON formats

2. Goals

- Automate the collection of Target store information from all 50 U.S. states and territories.
- Structure the data in a machine-readable format (CSV/JSON) for further use.

3. Functional Requirements

FR1 Access the base store directory and retrieve all U.S. state URLs.

FR2 From each state page, retrieve all city names and corresponding store links.

FR3 For each city, extract store names and group them under the city.

FR4 Format output into rows with 3 fields: State, City, Stores (comma-separated).

FR5 Export the final dataset as `target_stores.csv` and `target_stores.json`.

FR6 Include error handling and retry logic for failed requests.

FR7 Log progress and handle interruption gracefully.

---

4. Non-Functional Requirements

NFR1 The scraper should avoid duplicate entries and incomplete data.

NFR2 Response delays and polite scraping intervals (rate limiting) should be implemented.
NFR3 Python 3.x must be used with common libraries (requests, BeautifulSoup, pandas, etc.).
NFR4 The code must be modular, well-documented, and easy to maintain.

---

5. Output Format

Example row from the final CSV/JSON:

State	City	Stores
California	Los Angeles	Target Downtown LA, Target Hollywood
Texas	Austin	Target South Austin, Target North Austin

---

6. Technical Stack

Component	Technology
Programming Language	Python 3.x
Libraries	`requests`, `BeautifulSoup`, `pandas`, `json`, `logging`
Output Format	CSV, JSON
Environment	Local machine or virtual environment

---

7. Deliverables

- Python script: `target_scraper.py`
- Output files:
  - `target_stores.csv`
  - `target_stores.json`
- Documentation:
  - `README.md` with setup, usage, and output info