# Executive Summary - Soko Glam Web Scraping Feasibility

## Project Overview

Target Website: https://sokoglam.com

Analysis Date: October 7, 2025

Total Products: 781

Business Type: Korean Beauty E-commerce (Shopify-based)

## Key Findings

### Difficulty Assessment: EASY (3/10)

Soko Glam represents an optimal scraping target with minimal technical barriers and excellent data accessibility.

### Critical Success Metrics

* ✅ \*\*100% HTTP Success Rate\*\* using real browser headers
* ✅ \*\*Complete Product Data\*\* available in JSON-LD format
* ✅ \*\*No CAPTCHA/JavaScript Challenges\*\* encountered
* ✅ \*\*Standard Protection Level\*\* (Shopify + Cloudflare)
* ✅ \*\*Predictable URL Structure\*\* via public sitemap

## Technical Recommendation: HTTP-First Approach

### Why HTTP Scraping is Optimal

1. \*\*Perfect Success Rate\*\*: 8/8 test products successfully scraped via HTTP
2. \*\*Rich Data Access\*\*: Complete product information in structured JSON-LD format
3. \*\*Fast Performance\*\*: Average 2.5 seconds per product page
4. \*\*Cost Effective\*\*: No browser automation overhead required

### Browser Automation NOT Required

Testing confirms that browser automation would be unnecessary complexity given the 100% HTTP success rate with authentic headers.

## Business Impact Analysis

### Operational Efficiency

* \*\*Data Collection Time\*\*: 6-8 minutes for complete 781-product catalog
* \*\*Infrastructure Costs\*\*: Minimal - datacenter proxies sufficient
* \*\*Maintenance Overhead\*\*: Low - stable Shopify platform
* \*\*Scalability\*\*: Excellent - standard HTTP scaling patterns apply

### Data Quality Assessment

**Complete Product Information Available:**

* Product names, descriptions, pricing
* Brand information and categorization
* Stock availability and variants
* Customer ratings and review counts
* High-resolution product images

### Competitive Advantages

* \*\*Faster Time-to-Market\*\*: Immediate implementation possible
* \*\*Lower Resource Requirements\*\*: HTTP-only approach reduces complexity
* \*\*Reliable Data Pipeline\*\*: Shopify's structured data ensures consistency
* \*\*Scalable Architecture\*\*: Easy to expand to additional Shopify stores

## Implementation Roadmap

### Phase 1: Core Setup (1-2 days)

* Configure HTTP client with extracted browser headers
* Set up datacenter proxy rotation (OxyLabs/Bright Data)
* Implement JSON-LD data extraction pipeline
* Deploy rate limiting (2-3 requests/second)

### Phase 2: Production Deployment (1 day)

* Full 781-product catalog scraping
* Data validation and quality checks
* Monitor success rates and response times
* Implement error handling and retries

### Phase 3: Ongoing Operations (ongoing)

* Daily sitemap monitoring for new products
* Monthly browser header refresh
* Performance monitoring and optimization

## Risk Assessment: LOW RISK

### Minimal Risk Factors

* Standard Shopify/Cloudflare protection (not aggressive)
* No anti-bot escalation observed during testing
* Public sitemap indicates crawl-friendly approach
* HTTP success eliminates detection risks

### Risk Mitigation

* \*\*Rate Limiting\*\*: 2-3 RPS prevents overload detection
* \*\*Header Rotation\*\*: Monthly updates maintain authenticity
* \*\*Proxy Rotation\*\*: Datacenter IPs rotated every 50-100 requests
* \*\*Monitoring\*\*: Continuous success rate tracking

## ROI Projections

### Implementation Costs (Low)

* \*\*Development Time\*\*: 2-3 days for complete implementation
* \*\*Infrastructure\*\*: $50-100/month for datacenter proxies
* \*\*Maintenance\*\*: Minimal ongoing oversight required

### Data Value (High)

* \*\*781 Products\*\* with complete specifications
* \*\*Real-time Pricing\*\* and inventory data
* \*\*Customer Reviews\*\* and ratings
* \*\*Competitive Intelligence\*\* for Korean beauty market

## Strategic Recommendations

### Immediate Actions

1. \*\*Prioritize Soko Glam\*\* for quick wins given easy difficulty
2. \*\*Use as Reference Architecture\*\* for other Shopify-based targets
3. \*\*Leverage for Market Research\*\* in Korean beauty segment
4. \*\*Scale Approach\*\* to competitor analysis

### Future Opportunities

* \*\*Expand to Similar Sites\*\*: Apply methodology to other K-beauty retailers
* \*\*Competitive Monitoring\*\*: Track pricing and inventory changes
* \*\*Market Analysis\*\*: Trend identification across product categories
* \*\*Brand Intelligence\*\*: Monitor new product launches and popularity

## Conclusion

Soko Glam offers an exceptional opportunity for data collection with minimal technical barriers and maximum data quality. The 100% HTTP success rate, combined with rich structured data and predictable architecture, makes this an ideal target for immediate implementation.

Recommendation: Proceed with full implementation using HTTP-first approach with datacenter proxies.