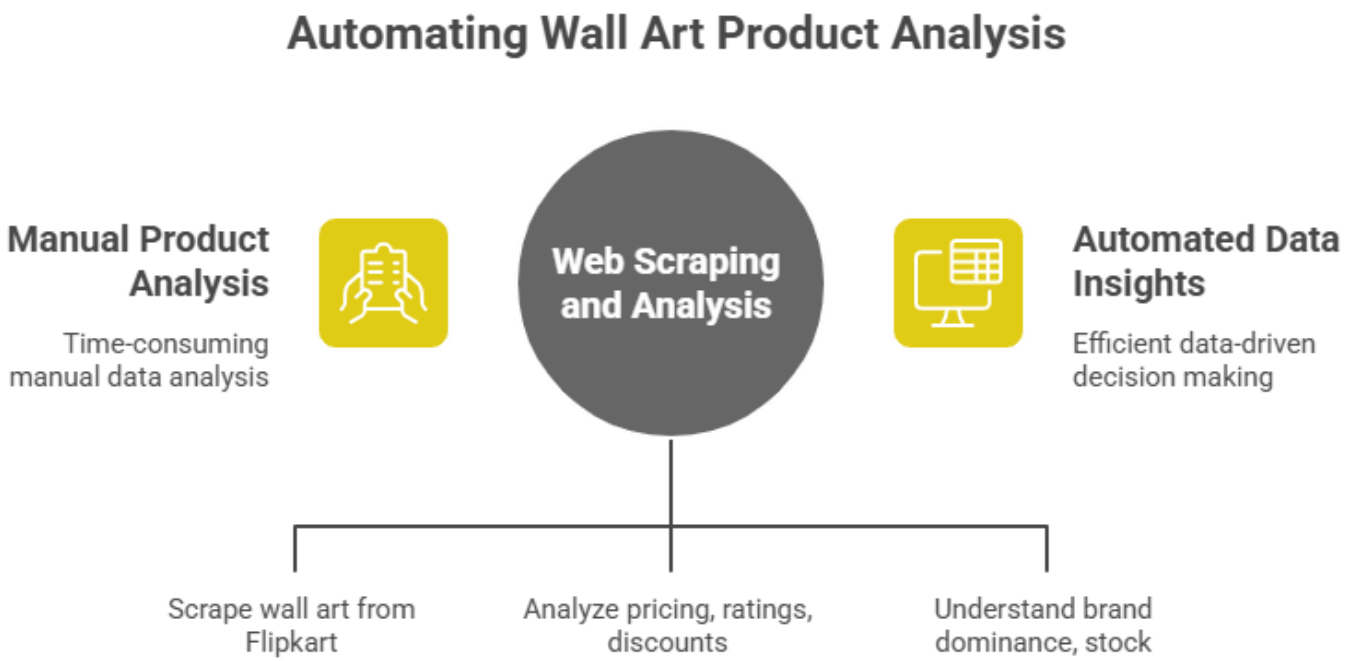
Web Scraping Flipkart Wall Art



Problem Statement

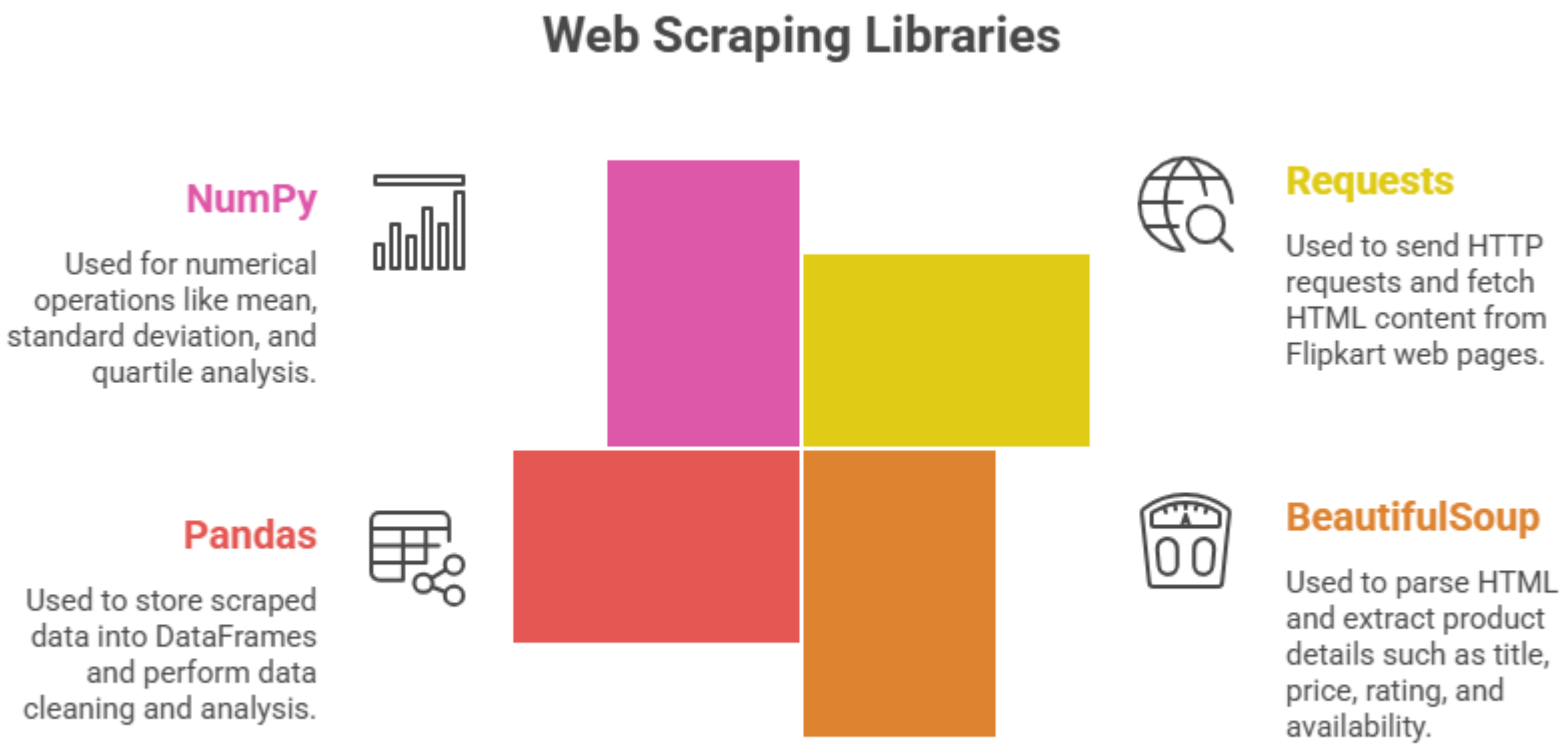
Why automate wall-art analytics on Flipkart?

- **Current analytical bottlenecks** : Extracting price, discount, rating, and inventory signals by hand demands excessive human hours and is prone to inconsistencies.
- **Scope of automated extraction** : The scraper will target the first five pagination windows—approximately one hundred distinct SKUs—capturing the full spectrum of listings presented to end-users.

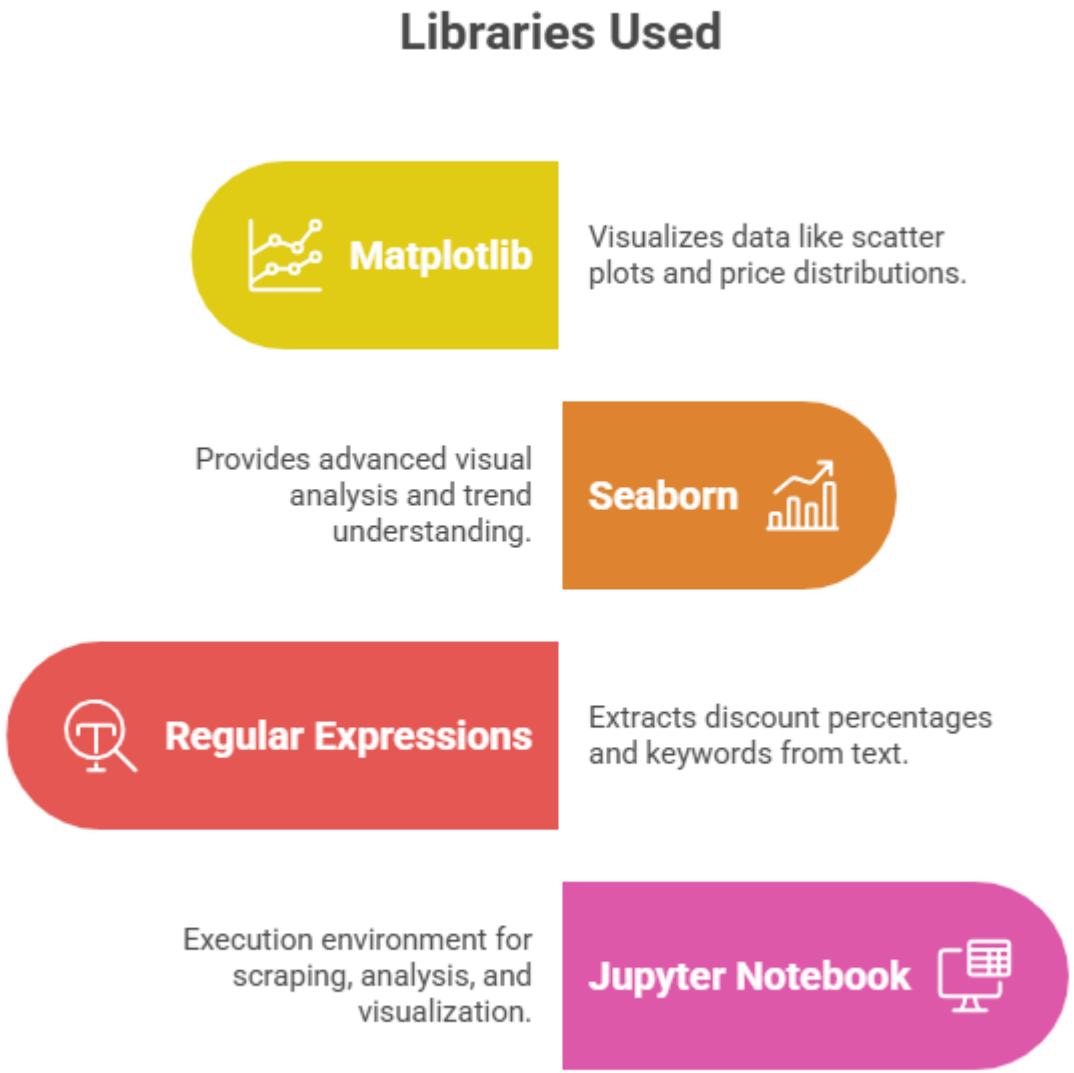
Key Objectives :

- Automate data collection from multiple product pages
- Analyze pricing and discount patterns
- Identify brand dominance and rating trends
- Assess stock availability patterns

Core Libraries for Data Collection



Visualization & Analysis Tools



Data Collection Pipeline

- **URL Identification**

  Identify Flipkart wall art category URLs and construct pagination parameters for systematic page traversal.

- **HTTP Requests**

  Send GET requests to each product listing page using the Requests library with appropriate headers.

- **HTML Parsing**

  Parse retrieved HTML content using BeautifulSoup to navigate the DOM structure efficiently.

- **Data Extraction**

  Extract product attributes: brand names, current prices, original prices, customer ratings, and stock availability.

Data Processing & Analysis Workflow

- **Data Storage**

  Store extracted product information in Python lists, then convert to a structured Pandas DataFrame for systematic analysis.

- **Missing Value Handling**

  Identify incomplete records and assign NaN values where product attributes like ratings or prices are unavailable.

- **Data Cleaning**

  Remove currency symbols (₹), commas, and other formatting characters from price columns to enable numerical operations.

- **Statistical Analysis**

  Calculate descriptive statistics, perform correlation analysis, and create visualizations to extract actionable insights from the dataset.

Technical Obstacles in Execution

Data Structure Challenges	Technical Constraints
Inconsistent HTML: Product listing structures varied significantly across different items, requiring flexible parsing logic.	**Request Blocking:** Risk of IP blocking and CAPTCHA challenges when making rapid successive requests to Flipkart servers.
Missing Attributes: Many products lacked complete information—ratings, original prices, or delivery dates were frequently absent.	**Brand Inconsistency:** Brand names appeared in various formats and spellings across different product listings.
Mixed Listings: Sponsored products appeared alongside organic results, complicating pure market analysis.	**Incomplete Discounts:** Discount percentages weren't available for all products, limiting promotional analysis.
Hidden Elements: Stock availability and delivery dates weren't consistently visible in the HTML structure.	**Pagination Issues:** Accessing pages beyond certain thresholds became unreliable, with inconsistent response structures.

Market Insights & Strategic Findings

- The wall art market is dominated by low-cost products, with most items priced under ₹500 and only a few premium-priced outliers.
- All products in the analyzed category are currently in stock, resulting in 0% unavailability and indicating strong inventory management.
- The WALLSTOXX brand does not offer discounts on any of its products, indicating a consistent full-price strategy.
- The scatter plot shows no strong relationship between discount percentage and customer ratings, indicating that product quality matters more than price reductions.
- The top 10 wall art brands all maintain high average ratings above 4, with ADVIK leading, indicating strong customer satisfaction and close market competition.

- Customer sentiment is overwhelmingly positive, with most products receiving 4–5 star ratings and very few negative reviews.
- Average delivery delay is around 2 days in Mumbai, highlighting generally reliable delivery with room for minor improvements.
- Premium products show significantly higher customer loyalty than entry-level products, as reflected by higher average ratings.
- Consistent brands dominate:

  Brands like PVL, Ajay, SHRI, saf appear multiple times.
- Heavy discounts drive top ratings :

  1.Discounts mostly between 56% – 90%

  2.Several top-rated products have >80% off


THANK YOU!