DATA USE CASES

Inflation Monitoring through Web Scraping

How the world's most sophisticated consumer agencies are enhancing their monitoring process by incorporating web scraped data.

INTRODUCTION

Inflation. It's one of the most important and influential factors shaping every economy on Earth - but we don't need to tell you that.

Keeping track of inflation is a tricky job, involving a lot of moving pieces. At Scrapinghub, we're proud to power inflation monitoring systems across the world. Large commercial agencies trust us to provide accurate, timely data for their pricing indexes, and we're glad to provide this service backed by our industry-leading reliability guarantee.

Without reliability and an obsessive standard for quality, we wouldn't have won the trust of the world's most sophisticated state agencies and statistic offices. Extracting pricing and product data at such a scale as required by these clients is often an extremely challenging and highly specialized project - but one we've more experience than anyone in executing.

Due to the sensitive nature of the information in this case study, we've anonymized our clients and will avoid using any identifying information to respect their confidentiality.

In this case study, we will explore how web scraped data has streamlined the work of statistics offices and large consumer agencies in monitoring pricing and product data - specifically for the purpose of monitoring inflation.

We're excited to work with governments and private industry to power extremely sophisticated inflation monitoring programs.

Never before has the sheer quantity of such data been made available, and especially at such speed. We save our clients time, money, and headaches by taking care of the technical aspect of generating the data used to forecast, monitor, and make further use of inflation-informing data.

In this white paper, you'll get a behind-the-scenes look into how we design these projects to boost productivity and save time for some of the world's busiest agencies.

Contents:

HOW WE HELP

We aim to streamline the data collection processes for our clients, and while each project is different, those involved in inflation monitoring generally have a similar scope involving the acquisition of pricing data. By working with the world's best web data extraction experts, we create solutions empowering you to spend more time actually working and less time fussing out a minutiae of tech-related problems.

When it comes to inflation monitoring, there's four straightforward advantages to using Scrapinghub as a web data solution:

1. **Aggregate better pricing data -** Whether from insurance policies available in a particular region or the high-volume pricing data of everyday consumer goods, our clients can choose from niche goods to the entire consumer market as the target for data extraction.

2. **Monitor more accurately -** Clients that only want data from a particular niche have opted to receive consistent, accurate reports about price changes, alerting t**hem to potential changes in a good which may have price externalities across the economy.**

3. **Completely validate your pricing data -** Our clients sometimes work with billions of points of data each day. Web scraping serves as an important quality assurance tool to check aggregated pricing data against itself to ensure the highest accuracy and real-time consistency.

4. **Reliability guaranteed -** As agencies are often held to the highest standards by their governments, they need a 100% reliable supply of web data. If a data feed dropped for even a day it might have cascading consequences. Fortunately, we are the industry's most reliable and most accountable and have led a number of projects for inflation monitoring agencies already.

Partial to Total

Some clients want an entire data infrastructure built to integrate the data provided into their internal systems. Others might simply want the raw data. Either way, we build bespoke data solutions so that our clients can tighten their analytics and reporting accuracy by compounding their records with web scraped data.

One-off to Partnerships

Most of our clients in this area establish long-term partnerships with us, as we dedicate an internal team to ensuring complete satisfaction of the scope of your project - meaning you have the data you need, when you need it, without the hassle.

Governmental agencies, from the most simple to the most sophisticated, stand to benefit a great deal by bringing on an industry-leading third party data aggregator.

TURNING PRODUCT DATA INTO MARKET INTELLIGENCE

As businesses are increasingly moving online, even encouraged by state agencies to do so, having a system for capturing and monitoring their fast-moving pricing is more important than ever.

As an inflation monitoring agency, you face serious technical challenges as the internet continues to exponentially grow in size. Scrapinghub is your bulwark against these challenges and their associated headaches, and our team is the most experienced in the world.

Your Friend in Big Data

The sheer size and scale of the data used by your agency on a daily basis requires a highly scalable web scraping infrastructure. This infrastructure needs to be flexible, capable of scaling from hundreds of spiders to thousands in a matter of months, so that if you need to scape 9 billion products from 2,000 online stores each day, you can.

More importantly, as agencies monitoring inflation are tasked with providing actionable reports, the frequency and freshness of this data is critical. While there is value in being able to analyze historical pricing trends, etc,

the real value of this data is in its freshness. Our solutions satisfy a need for historicity and a need for up-to-date validation simultaneously - making it the ultimate source of online pricing data.

Quality and Safety

At Scrapinghub, we work with some the world's largest retailers. We understand how significant high quality pricing data, as for many of our clients, it is a foundational aspect of their entire business. For our clients in inflation monitoring, we guarantee the highest possible quality of data, as we fully understand how a drop in data quality would immediately filter through to reporting mechanisms, compromising an agency's reputation and potentially violating state-mandated requirements.

Reliability is another aspect of our quality-centric approach to web scraping. Agencies that monitor the economy are often held to the highest standards by their governments and need a 100% reliable supply of web data. Fortunately, we are the industry's most reliable and most accountable web scraping company, having led a number of projects for inflation monitoring agencies already. While data is in our DNA, reliability is as well - and it's also a clearly stated guarantee in our SLA.

Any Kind, Anytime.

No matter why or how you're monitoring or measuring inflation, the flexible nature of web scraping means that if it's online, we can get it for you.

Some of our clients need billions of data points - others, only a handful - and the diversity of data requested is tremendous. After consulting with our specialists, many agencies are surprised to find that they can supplement their data sources with such a rich variety of high quality information.

USE CASE 1: AN AGILE INSURANCE DATA AGGREGATOR

A state agency reached out to us looking for the collection of insurance pricing information across their country's market. This included both home, auto, and other forms of insurance. Price quotations for such products update on an irregular basis, and the websites hosting the information are protected by sophisticated systems and pagination, unfriendly to most scraping operations.

By partnering with Scrapinghub, the agency involved wanted pricing data from this industry on a monthly basis - no frills, no spider source codes.

The Data You Need - No Strings Attached

This client in particular simply wanted the data collected and provided and had no need for the spiders' source code. Furthermore, they only wanted the data, and didn't need an infrastructure built around its delivery.

Requesting around 100,000 records with around two dozen fields across the types of records involved, we were able to complete this project quickly and to their satisfaction.

Thwarting Barriers to Data

This client in question benefited from both the automation of their data collection as well as the intrepidness of our world-class crawling technology. The problem with many data crawls is that those performing the crawl do so in an obtrusive, unkind way. We combine our professionalism and understanding of peak server times with intelligent data-obscuration workarounds to ensure our clients get the data they need when they need it.

An insurance pricing index will be made up of hundreds of thousands to millions of records, so ensuring both consistency and reliability across these records is absolutely vital to the performance of the agency involved. Part of ensuring a steady stream of high quality data is making sure that intentionally unfriendly systems don't interrupt crawls and create blind spots in the data coverage so important to inflation monitoring.

Custom Crawls with Risk Profiles

Insurance is all about risk. When generating pricing quotations for our client, it wasn't a simple matter of crawling between web pages and scraping the data from a simple text format. Instead, we worked with the clients' risk profiles and developed a custom crawler which would retrieve quotations from each data source.

Five hundred risk profiles were used to generate records from the data sources requested, and we were happy to provide the highest quality of uninterrupted data to the client.

USE CASE 2: AUTOMATING RETAIL DATA COLLECTION FOR A PRICE INDEX

Prior to incorporating our web scraping solutions, a few of our clients had been spending an inordinate amount of time performing manual data extraction. As prices and averages in retail move often, this resulted in agencies wasting a lot of time and effort on tasks that were both time-sensitive and extremely mundane.

One client, part of a large country's national inflation monitoring service, needed high quality pricing data from its largest retailers, for integration into their CPI calculations.

Goodbye, Manual Extraction

The web is full of disparate, cluttered websites. Navigating and collecting information from these data sources manually is a nightmare - fortunately, by partnering with Scrapinghub, data collection is a breeze. For this client, who needed an immense amount of pricing data from quickly changing, complex websites, we developed crawlers to scrape and format the data needed in a highly structured, useful form.

Quickly after reaching out to us, we assembled a dedicated internal team whose goal was to provide them with a consistent, time-saving solution. Because we hire globally and have an elastic talent pool of the world's best data extraction specialists, we were able to quickly build, test, and deploy a solution for our client, saving them a tremendous amount of time and allowing their talented analysts and staff to spend more time on meaningful work.

Only the Data You Need

Vendor lock-ins are an annoyance, and it's absurd to pay for software that your office doesn't need when all you need is the data.

This client, like others, requested data extraction services to supplement a structured pricing index. In addition to quickly developing a robust solution for their data collection needs, we provide the service without any obnoxious hidden fees or software licenses. In fact, we're proud to make most of the tools we develop available for free as open source technology.

Agencies monitoring inflation are busy enough. Partnering with Scrapinghub removes the hassle, and the headache, from web data acquisition.

Any Kind of Data, Exactly When You Need It

Are you scraping insurance policies? High volume consumer goods? Web scraping is the ultimate service for extracting and structuring data used by the world's largest statistics offices and inflation monitoring agencies every day.

By developing crawlers and data extraction tools specifically for your individual project, we ensure only the highest degree of quality and reliability. The data you're using is important, and your conclusions drawn from it have a serious impact. That's why we're excited to have partnered with agencies around the world, with the aim of ensuring global economies are armed with the best tools available for steering their economy ahead.

PROJECT EXECUTION

We build robust, reliable, and scalable systems capable of scraping millions of records per day, while maintaining enough flexibility to accommodate a constantly-changing list of products and targets for extraction.

Working with our Solution Architecture team, our projects can avail of our entire stack of web scraping tools, resulting in a data extraction infrastructure bespoke to the client's needs.

Our Toolkit for Inflation Monitoring

A web scraping stack is all the individual components you need to scrape the web at scale: data extraction spiders, infrastructure configuration, proxy management, data quality assurance, and much more.

The web scraping stack used in aiding inflation monitoring includes:

- **Scrapy -** the open source Python web scraping framework developed by Scrapinghub's founders.
- **Scrapy Cloud -** Scrapinghub's dedicated hosting service specifically designed for deploying and managing web crawlers in the cloud.

- **Frontera -** the open source framework Scrapinghub developed to facilitate building a crawl frontier, helping to manage your crawling logic and sharing it between your web scraping spiders.
- **Crawlera -** Scrapinghub's single endpoint proxy solution that automates proxy management and ensures our customers can reliably scrape the web a scale.
- **Splash -** the lightweight scriptable browser developed by Scrapinghub that enables fast rendering of web pages that use JavaScript.
- **Spidermon -** the spider monitoring system developed by Scrapinghub to enable the real-time monitoring crawl status and detection of spider failures, bans, errors, etc.
- **GATF -** the automated data quality assurance framework developed by Scrapinghub to assist QA engineers in validating and monitoring the accuracy and coverage of datasets.

Bespoke Solutions

Each project has its own distinct challenges and requires its own dedicated approach. When we take on a client for a project such as this, we dedicate an entire internal team to its execution, drawing from our pool of the world's most talented web data extraction specialists.

The individual attention and care given to your project will ensure a lightning-fast turnaround time on any potential bug fixes or crawler repairs necessary to complete the project. Due to the international, remote nature of our teams, we're proud to admit that we work around the clock to see your problems solved.

OUR GUARANTEE

A key feature of our projects is the contractual commitment to reliability and data quality. In our Service Level Agreement with the customer, Scrapinghub explicitly guarantees:

Data Quality - Our SLAs guarantee a 99% accuracy rate as measured by the total number of correct items over the total number of items extracted. Scrapinghub ensures this level of data quality through a combination of robust spider QA during the development stage and continuous automatic/manual QA of the data by our dedicated Quality Assurance team.

Response Times - Our SLAs with enterprise clients also guarantees that any reported problem, JIRA ticket or spider break will be picked up within 12 hours. Ensuring 90% of issues (spider breaks, etc.) are fixed within a day. Due to the inevitable constant changes to the structure and function of a website, spiders often break every 2-3 months on average. As a result, in a project with over 4,500 spiders, our team has to update 25-40 spiders per day.

WHY SCRAPINGHUB?

For our inflation monitoring clientele, Scrapinghub is more than just a web scraping provider. The web scraping system we build and maintain is the backbone of daily activity, and without it, they wouldn't have the data infrastructure necessary to fulfill their reporting and analytic requirements.

Here are some of the main reasons consumer agencies across the globe choose to partner with Scrapinghub:

Reason #1 - Industry Experts

When working with Scrapinghub you are partnering with the industry's leading data extraction experts. Data extraction and web scraping is in our company DNA, all we have done for the past 8 years is find better ways to extract data from the web at scale. As a result, we've built a deep institutional knowledge that we are able to call upon when tackling a new project or problem. Our team of engineers know the major websites inside-out, know how to squeeze the highest performance out of Scrapy, the open source python web scraping framework developed by Scrapinghub, and have developed industry-leading technologies in the areas of proxy management, QA, etc. They knew that when they were working with Scrapinghub, they were working with the best in the business.

Reason #2 - Reliable Partner

Not only is Scrapinghub the leading web scraping expert in the market, it is the most established business in the market. Whereas most web scraping consultancy firms or providers are either small operations with 2-10 staff (mainly engineers) or risky venture-backed startups, Scrapinghub is a robust and established business (founded in 2010) that has moved past the risky "startup" phase where companies often close overnight with little warning for their customers. Scrapinghub has a diverse client portfolio, a strong leadership team, a rapidly growing team of engineers and support staff. This was a huge reason why the client (and many other companies) decided to outsource such a mission-critical component of their business to Scrapinghub. They knew that if they outsourced their web scraping to Scrapinghub they would have a partnership that would remain with them for the long term.

Reason #3 - Guaranteed Data Quality & Uptime

By working with Scrapinghub the client was guaranteed reliable and a high-quality data feed. Every customer contract comes with an SLA that explicitly guarantees near perfect data quality from their data feeds. They just needed to specify the data they wanted and Scrapinghub would do the rest. Not only that but when a spider eventually breaks due to website format changes, they knew that it would be fixed in a matter of hours. Enabling them to completely outsource the headache of monitoring and maintaining thousands of spiders.

Reason #4 - Access to the Best Talent

The huge advantage to working with Scrapinghub is the fact that our clients get access to the best web scraping experts in the world. The fact that Scrapinghub is 100% remote and hires globally, means we are able to hire only the best web scraping experts. Whereas if a company was based in Boston, or Edinburgh, for example, their talent pool would be limited by their geographical location. By working with Scrapinghub, companies get access to the world's best web scraping engineers.

Reason #5 - Elastic Pool of Engineers

The beauty of working with Scrapinghub versus managing web scraping in-house is the ability to cope with surges in demand. If a big customer requires a new data feed, then Scrapinghub can quickly pull in more experienced crawl engineers from other parts of the business to work on the project. Whereas if it was an internal project, the company would often only be able to bring in inexperienced engineers or hire external help. This flexibility was a huge advantage to this client as they were confident that no matter what new consumer brand client they signed as a customer, they would have a fully functioning data feed live within a matter of days for that client. Establishing them as the best in the business for product monitoring services.

CONCLUSION

As we've seen, web scraped pricing and product data is capable of saving time and improving the reporting capabilities of inflation monitoring agencies around the world.

No matter the size of the economy or the agency, monitoring inflation is a practice which requires extremely specific tools and extremely reliable data. Partnering with Scrapinghub ensures you get the data you need, when you need it, at the highest level of quality and reliability available.

As markets continue to grow, the amount of pricing and product data generated grows with it, opening new sources for those who monitor consumer goods to improve their analytics and develop more informed reports.

ScrapingHub is proud to offer the most robust web scraping solutions on the market, so if these use cases inspired you or you're curious about how web scraped product data can benefit your business, request a free consultation today.

ABOUT SCRAPINGHUB

Scrapinghub turns websites into data with industry leading technology and services.

Our solutions include:
- **Data on Demand** for any size scraping project with accurate, reliable data feeds at speed
- **Custom Data Extraction** with a team of the best web scraping engineers
- **Crawlera** for IP proxy management to crawl quickly and reliably
- **Scrapinghub Platform** to deploy and run your web crawlers

We have a strong reputation and deep domain expertise. Whatever data you need, and however you need it, we're the best team to deliver it!

ABOUT SCRAPINGHUB

Scrapinghub turns websites into data with industry leading technology and services.

Our solutions include:
- **Data on Demand** for any size scraping project with accurate, reliable data feeds at speed
- **Custom Data Extraction** with a team of the best web scraping engineers
- **Crawlera** for IP proxy management to crawl quickly and reliably
- **Scrapinghub Platform** to deploy and run your web crawlers

We have a strong reputation and deep domain expertise. Whatever data you need, and however you need it, we're the best team to deliver it!

Visit our website for more resources
www.scrapinghub.com