

Travel Product Feeds

Client: A popular travel portal from USA

Client: Site-specific crawl and extraction

Challenge: The client was looking to include a data layer into its current set-up that would allow continuous free flowing feeds free of “noise” so that the team could only focus on the other aspects of their travel portal like the marketing and promotion. They wanted to use the travel data aggregated from a list of sites to fuel the database beneath their website.

The Solution: The client provided us with the list of sources to be crawled and the data points required. The extraction was to be done on daily basis which meant fresh data sets have to be provided everyday. Our team set up crawlers to fetch the required data fields from the source sites provided by the client. This use case comes under our [site specific crawl](#) offering since the websites in the list had different structuring and design. The client needed the extracted data in [CSV format](#) and be uploaded to their S3 servers. The initial setup was complete in a few days and the crawlers started delivering data immediately. About 2 million records were delivered to the client during the first crawl.

Benefits:

- The complex technical aspects of data extraction were taken care of by us
- It took only a few days for the initial setup after which data started flowing consistently
- We set up monitoring for the source websites to ensure proper functioning of the crawler
- Our advanced tech stack handled huge amounts of data effortlessly
- The client was able to enrich their travel portal with an enormous number of listings within a short period of time