

IRisk Lab Data Discovery and Consolidation - Task 2 : Midterm Report

Week 7 Report

Oct. 10

Rohit Valmeekam

What Was Accomplished

- ❖ Learned Web Scraping and its applications for insurance datasets
 - Importance of Web Scraping for Insurance Datasets:
 - Market Analysis: Insurance companies can utilize web scraping to gather information on competitors, such as their premium rates, policies, and customer reviews. This data is essential for making informed decisions and staying competitive.
 - Risk Assessment: By scraping data from various sources, insurers can analyze geographic and demographic trends, helping them better understand and assess risks associated with different regions or customer profiles.
 - Fraud Detection: Real-time data scraping from social media or news outlets can be used to detect potential insurance fraud by identifying inconsistencies in claims and reported incidents.
 - Product Development: Data collected through web scraping can provide insights into customer needs and preferences, allowing insurance companies to design tailored products and services.
 - Python: Python is one of the most popular languages for web scraping due to its extensive libraries such as BeautifulSoup, Requests, and Scrapy.
 - Make HTTP Requests: `requests.get(url)`
 - Parsing HTML with BeautifulSoup:
 - Create a BeautifulSoup object: `BeautifulSoup(html_content, 'html.parser')`.
 - Use BeautifulSoup's functions to find and select specific HTML elements.
 - Selecting Elements:
 - `soup.find_all('replace_with_tag')`
 - JavaScript: Libraries like Puppeteer are ideal for websites heavily reliant on JavaScript for rendering content.
 - Web Scraping Tools: Commercial tools like Octoparse and open-source tools like WebHarvy can simplify the scraping process.

Challenges

- ❖ Learning what technologies to use with web scraping
- ❖ Learning what tags to search for specific insurance data