

# INFLUENCERMARKETING - WEB SCRAPE TEST

Wesley Souza Rocha

## THEORETICAL TEST

1. One of the biggest challenges are the constant changes on the scrapes via those tools and having to adapt on them. Scraping tools can be very easy to use but the challenge will be which tools suits you will need and what the best for your use case. The other problem is not using proxies while scraping. Sometimes the tools are working good but due the IP block we may think that something is going wrong with the tools. So my advice is to use proxies while scraping some sites. We can use OctoParse or Scraper API. If the scraping is not working my advice is to choose professional tools such as scrapy or parsehub (you can use the others two that i have told before). And the other problem is if it not still works when you are using those professional tools, make sure that you are using updated ones.

2. Data mining is basically extracting data from a source and analyzing it. For example, using API to extract some information and figuring out how do people do to increase their numbers in social media.

3. I did use once one called tempMonp to extract some data answers from a website to improve my research about fluids.

4. HTTP is like a big library, you say you want to access a library (server) in a city (domain or a higher domain), if you specify a rack or row you can do that as well (path). Also you can send questions at the end of the HTTP url, like asking some librarian something e.g. if has books about Java or Python. He works as hypertext, transfer and protocol loosely.

5. Proxy servers; It protects the computer networks by regulating request. It grants you a protection when you are receiving files all over the world when you navigate in some websites or programs. And they are important to scraping process because it can prevent responding to potentially hazardous request or disclosing personal information.

6. The Artificial Intelligence (AI) started as a subfield of computer science focusing on solving problems that humans can but computers can't (e.g. image recognition). Machine

Learning (ML) is a specific approach to AI, that use statistical algorithms called by neural network models, which is to enable machines to acquire skills by getting better over time. (e.g. execute tasks answer repositories, sorting algorithms...). Data science is basically a market decision making using data. Also is the domain of using advanced analytics strategies and scientific principles to yield worthwhile knowledge from data for business decision-making, strategic planning, and different benefits. Its extremely important to businesses for improve their marketing and lead competitive advantages. Is important to know that machine learning is more attached to engineering, for maintenance and production, while data science is into data analysis and predictions.

7. Feature engineering is basically making new data available to models. if you design it right, then you can reuse the code. To be more specific, raw data contain information of a house like length, height, width and after a simple feature engineering you can have a better information such as volume (height x width x length) which better represents the dimension of a house.

8. Basically, the difference between them is that data engineer is more attached into designs, constructs, tests, architecture and maintains big data from various resources. While data scientist is attached in apply statistics, helps organizations to translate big data making them more readable/understandable and a market decision making using data.

9. Data pipeline is basically a bunch of activities that extract data from various sources and move it to a destination. So for the beginning, there are 5 steps to design a data pipeline which is: Data Sources, Ingestion, Transformation, Destinations, Monitoring. So you need to know the process which will define what, where and how data are collected and transformed meaning that you will need to manually pick every data source, table, field, etc. Using marketing words, this means that you will must need a meeting agenda, a proposal, a seller responsible, negotiation and then the deal close.

## FINAL CONSIDERATIONS

### PRATICAL TEST

For this part, in my opinion, we didn't have much time to code the scraper and store all the information from the data. So i made a simple fluxogram just to clear the idea before start. Im more into data apresentation, analys and worksheets. Its important to a company have a setor like this. I did enjoy the questions it had great points.

