The aim of this project was to develop and hone skills in data extraction from APIs, data analysis using natural language processing, and data visualization. The assumption was that this would make searching for and applying for new jobs faster and more efficient. It is also assumed that using Google jobs will give me access to a wide range of listing sites.

The project's objective was to explore the potential of SERP API's Google Jobs API for acquiring job listings data for analysis, design a repeatable, scalable, and easily modifiable script to pull the latest job listings data, and analyze the data using natural language processing and data visualization.

The project utilized various methods, including data extraction, data exploration, data cleaning, natural language processing, sentiment analysis, named entity recognition, and data visualization. The technologies used in the project were Python, SERP API's Google Jobs API, Pandas, Matplotlib, and the Natural Language Tool Kit (NLTK).

The author developed a repeatable, scalable, and easily modifiable script to pull the current job listings data from the SERP API's Google Jobs API. The script utilized two separate API calls, and the analytical work was performed in a separate notebook. The author explored Python's Natural Language Tool Kit along with Pandas methods and functions to navigate and follow up on the job listings.

NLTK methods and functions were employed in this project to extract keywords and salary information from job descriptions, perform sentiment analysis, and identify named entities.

Overview of NLTK Methods

The analysis revealed that 48% of listings had keyword matches with a count of 440, while 52% of listings did not have keyword matches with a count of 480.

The NLTK summarizer brought the average length of job descriptions down to 973, which is a reduction of 77% in length on average.

The average sentiment score was mean 0.978115, and the min was -.92, and the max was 1. The 25% of this data fell at 0.9957 so there was not much variance in this at all making the data not as helpful as some of the other NLTK methods.

Using the concordance functionality, NLTK was able to find the salary mentioned in 44% of job descriptions pulled from the API, with a count of 406. This result was close in count to the salary information gathered form the Google jobs listing endpoint which had 417 results although not all of these results overlapped.

The overview of the rating information gathered from the Google jobs listings endpoint showed that the average rating out of 5 was 3.75, the min was 1.6, and the max was 5. Of the 554 company ratings gathered, only 3 listings were posted by companies that had a perfect 5 rating.

Overall, this project successfully achieved its objective of exploring the potential of SERP API's Google Jobs API for acquiring job listings data for analysis, designing a repeatable, scalable, and easily modifiable script to pull the latest job listings data, and analyzing the data using natural language processing and data visualization. The NLTK techniques helped in generating meaningful insights from the textual data, and the data visualization aided in presenting the findings in an easily understandable format. This project provided a solid foundation for further research into using data extraction and natural language processing to generate insights from job listings data.

Top of Form

Bottom of Form