Streamlining Job Search: Leveraging SERP API's Google Jobs API and NLTK in Python for Job Listings Analysis

This project aimed to streamline the process of job search by leveraging the SERP API's Google Jobs API to extract job listings data for analysis and exploring the Natural Language Processing Toolkit in Python. The project involved various methodologies, including data extraction, exploration, cleaning, natural language processing, sentiment analysis, named entity recognition, and data visualization.

To achieve this objective, the author developed a scalable and easily modifiable script to extract the latest job listings data from the SERP API's Google Jobs API. The analysis was performed in a separate notebook using Python, Pandas, Matplotlib, and the Natural Language Toolkit (NLTK).

Out of the 2400 job postings collected over 20 days, only 240 unique listings were obtained. The NLTK was used to preprocess job descriptions and extract keywords using word tokenization. The analysis revealed that 57.50% of the unique listings had keyword hits. The NLTK concordance methods were used to extract salary information from job descriptions, and salary information was found in 49.17% of listings, with a count of 118.

The NLTK was also used to perform sentiment analysis and summarize job descriptions by an average of 74%. Additionally, the analysis provided insights into job titles, with "Data Analyst" being the most common role (12.10% of all listings captured), followed by "Operations Analyst" (9.95%). The analysis revealed that 18% of the listings matched senior roles, with a count of 22, and 14.17% of the captured jobs were via LinkedIn. Only 19 listings were posted by companies that had a perfect 5 rating.

In conclusion, this project successfully leveraged data extraction, natural language processing, and data visualization to generate insights from job listings data. The NLTK techniques proved effective in extracting meaningful insights from the textual data, while data visualization aided in presenting the findings in an easily understandable format.