

Treatment Explanation File

Introduction

This document provides a detailed explanation of the Python script used to extract and save review data from Trustpilot into a CSV file. The script, as requested in the project explanation, is divided into two main parts: the first part collects general information about ATM companies, and the second part gathers specific details about individual reviews for every company.

Part 1: Collecting General Company Information

- **Library imports:** The script begins by importing the necessary libraries: *csv*, *requests*, and *BeautifulSoup* from the *bs4* package. These libraries are essential for handling web requests, parsing HTML content, and writing data to a CSV file.
- **Defining the URL:** The URL of the Trustpilot category page is defined. This URL points to the category of interest from which the ATM company data will be extracted.
- **Sending a web request:** A request is sent to the specified URL to retrieve the HTML content of the page. The response is then parsed using *BeautifulSoup* to facilitate the extraction of relevant data.
- **Finding company listings:** The script searches for all company listings on the page. Each listing contains information about a company, such as its name, domain, number of reviews, and trust score.
- **Opening a CSV file:** A CSV file is opened in write mode to store the extracted data. The header row is written to the file, defining the columns for company name, domain, number of reviews, trust score, and review percentages for different star ratings.
- **Extracting company information:** For each company listing, the script extracts the company name, domain, number of reviews, and trust score.

Part 2: Collecting Detailed Review Information

- **Defining the company review URL:** The URL of the company's review page is defined. This URL will be used to access the individual reviews for the company. It is a combination of the root of the Trustpilot URL, plus the domain name retrieved in part 1.
- **Iterating through review pages:** The script iterates through the pages of reviews for the company. It sends a request to each page and parses the HTML content to find the reviews. It also retrieves the percentages of reviews for each star rating (1 to 5 stars).
- **Finding reviews:** The script searches for all review elements on the page. Each review contains information about the star rating, review comment, and whether the company has responded to the review.
- **Extracting review information:** For each review, the script extracts the star rating, review comment, and company response status. If the review has a low star rating (1 or 2 stars), it checks if the company has responded to the review.
- **Writing data to CSV:** The extracted review information is written to the CSV file, along with the general company information. This ensures that all relevant data is captured in a structured format.

Conclusion

The provided Python script effectively collects and saves review data from Trustpilot into a CSV file. By automating the extraction process, it facilitates the analysis of company reviews and responses, providing valuable insights into customer feedback and company performance.