**TrustPilot data scraping**

£500 agreed fixed price to Edward Sims.

The key objective for the project is to identify companies on TrustPilot that are receiving negative reviews because of their poor customer service using WhatsApp, Messaging, SMS, Text, Webchat, etc.

Target Site – [www.uk.trustpilot.com](http://www.uk.trustpilot.com)

**Step 1** - Scrape the following categories

* Money & Insurance
* Travel & Vacation  
  Food, Beverages & Tobacco  
  Restaurants & Bars  
  Events & Entertainment  
  Beaty & Well-being
* Shopping & Fashion  
  Home & Garden  
  Vehicle & Transportation  
  Electronics & Technology  
  Animals & Pets
* Business services (for logistics)

Scroll all the way through all companies on each of the above categories.

**Step 2** – Select companies with an aggregate score of 3 stars or fewer, min of 1,000 customer reviews and operating in the UK.

Histogram of the distribution of companies and number of individual customer reviews. Then we can adjust the 1,000 figure.

Select all companies that have aggregate rating of 3 stars or fewer

(Example below, we would not select British Gas)

A screenshot of a company review

Description automatically generated

**AND**

The company has a minimum of customer 1,000 reviews on TrustPilot (see above screenshot)

**AND**

Contact

Companies must be operating in the UK from Contact on company review page.

(Example below)

A screenshot of a phone

Description automatically generated

**Step 4** – Of these selected companies, select those where the following are mentioned in individual customer reviews for any the companies with aggregate 3 star or less reviews:

1. Whatsapp
2. Whats app
3. Whats App
4. WhatsApp
5. Texted
6. Text
7. SMS
8. Message
9. Messaged
10. ~~Called~~
11. ~~Phone~~
12. ~~Phoned~~
13. ~~Call centre~~
14. ~~Contact centre~~
15. ~~Rang~~
16. ~~E-mailed~~
17. ~~E-mail~~
18. ~~E mail~~
19. ~~E mailed~~
20. ~~Email~~
21. ~~Emailed~~
22. ~~Wrote~~

All of 1 to 9 above need to be mentioned in at least 40% (which we may vary depending on number of companies generated) of the total number of reviews for each company. Could you do a histogram of the number of mentions for each company? Reminder that the The key objective for the project is to identify companies on TrustPilot that are receiving negative reviews because of their poor customer service using WhatsApp, Messaging, SMS, Text, Webchat, etc.

**In Step 4 ensure to exclude** **the responses from the reviewed individual companies for identifying the communication channels (eg WhatsApp etc), for example:**

A screenshot of a message

Description automatically generated

**Step 5 – Run scraping**

Generate a list by each of the categories in Step 1 that meet the above criteria.

**Step 6 – Provide an Excel output list**

Provide Excel document:

* Name of company
* Location
* Aggregate customer rating score
* Number of customer reviews
* Percentage number of mentions out of number of aggregate reviews for each company mentioning for each of the following (will be more than 40% refer back to step 4):
* Whatsapp
* Whats app
* WhatsApp
* Whats App
* Texted
* Text
* SMS
* Message
* Messaged
* Show 50 random examples the customer reviews that qualify for each company that scores an aggregate score of 3 stars or less

-----------

Programming so far:

from bs4 import BeautifulSoup

import requests

import pandas as pd

*# set list of companies - ideally this would be automated to go through a list of companies taken from the TrustPilot categories page*

list\_of\_companies = ['www.seescents.com', 'www.asda.com']

data = []

*# Set Trustpilot page numbers to scrape here*

from\_page = 1

to\_page = 5

for company in list\_of\_companies:

    trustpilot\_url = **f**"<https://uk.trustpilot.com/review/>{company}"

    print(trustpilot\_url)

    for i in range(from\_page, to\_page + 1):

        response = requests.get(trustpilot\_url+**f**"?page={i}")

        web\_page = response.text

        soup = BeautifulSoup(web\_page, "html.parser")

        for e in soup.select('article'):

            data.append({

                'review\_title':e.h2.text,

                'review\_date\_original': e.select\_one('[data-service-review-date-of-experience-typography]').text.split(': ')[-1],

                'review\_rating':e.select\_one('[data-service-review-rating] img').get('alt'),

                'review\_text': e.select\_one('[data-service-review-text-typography]').text if e.select\_one('[data-service-review-text-typography]') else None,

                'page\_number':i

            })

    pd.DataFrame(data)

data\_df = pd.DataFrame(data)

*# Additional columns as detailed in the brief should be provided.*

data\_df

has context menu