Name: Akash Kulkarni

Roll No: 33241

Batch: K-10

## **Problem Statement:**

Create a review web scraper for ecommerce to fetch real time comments, reviews, ratings, comments, names

```
# import libraries
         import pandas as pd
         from bs4 import BeautifulSoup
         import requests
         # Connect to Website and pull in data
         URL = 'https://www.amazon.in/dp/B08T6THSMQ/ref=sspa_dk_detail_2?psc=1&pd_rd_i=B08T6THSMQ&pd_rd_w=9YTVi&pf_rd_p=
         headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chro
         page = requests.get(URL, headers=headers)
         print("Web Scraping Status : ", page.status_code)
         soup1 = BeautifulSoup(page.content, "html.parser")
         soup2 = BeautifulSoup(soup1.prettify(), "html.parser")
         df = pd.DataFrame()
        Web Scraping Status: 200
         raw profiles = soup1.find all("span", class ="a-profile-name")
         profiles = []
         for profile in raw profiles:
            profiles.append(profile.get_text().strip())
         profiles = profiles[0:8]
         df["profiles"] = profiles
In [4]:
         raw_dates = soup1.find_all("span", {'data-hook': "review-date"})
         dates = []
         for date in raw_dates:
             dates.append(date.get_text().strip())
         df["dates"] = dates
         raw title = soup1.find all("a", {"data-hook" : "review-title"})
         titles = []
         for title in raw title:
             titles.append(title.get text().strip())
         df["titles"] = titles
In [6]:
         raw stars = soup1.find all("i", {"data-hook" : "review-star-rating"})
         stars = []
         for star in raw_stars:
             stars.append(star.get_text().strip())
         df["stars"] = stars
         reviews_raw = soup1.find_all("div",class_="a-expander-content reviewText review-text-content a-expander-partial
         for review in reviews raw:
             reviews.append(review.get_text().strip())
         df["reviews"] = reviews
         raw_comments = soup1.find_all("span",class_="a-size-base a-color-tertiary cr-vote-text")
         comments = []
         for com in raw comments:
             comments.append(com.get_text().strip())
         df["comments"] = comments
In [9]:
         df
```

	profiles	dates	titles	stars	reviews	comments
0	Amazon Customer	Reviewed in India on 7 June 2021	Best Laptop under 40k	5.0 out of 5 stars	This laptop is perfect for home and student us	647 people found this helpful
1	Amazon Customer	Reviewed in India on 10 February 2021	Ryzen 3 3250U much better than core i3 10th	5.0 out of 5 stars	One of the best product at this price	263 people found this helpful
2	Amazon Customer	Reviewed in India on 7 February 2021	All in all	4.0 out of 5 stars	Very good After use of 9 months (updated)	246 people found this helpful
3	Shubhankar	Reviewed in India on 15 May 2021	Highly recommended. Go for it.	5.0 out of 5 stars	Best laptop under 40k. 1TB HDD, 256GB SSD, 8GB	75 people found this helpful
4	karthik.v	Reviewed in India on 7 February 2021	Good	5.0 out of 5 stars	The media could not be loaded.\n	65 people found this helpful
5	Kushal Mukherjee	Reviewed in India on 17 June 2021	Poor laptop for 41K - freezes with MS Teams	1.0 out of 5 stars	I bought this laptop for remote schooling of k	56 people found this helpful
6	Kushal Mukherjee	Reviewed in India on 18 June 2021	Does not work with microsoft teams app. it fre	1.0 out of 5 stars	I don't know how I missed the review given by	44 people found this helpful
7	Yaswanth	Reviewed in India on 2 May 2021	Unsure of the Configuration	1.0 out of 5 stars	This laptop is overall a really good package,	40 people found this helpful