For the first step of the project, reviews of companies with over 10,000 reviews should have been extracted. The scraped information should contain related information such as the reviewer's name, country, review title, text, rating (in number of stars), and whether the company replied to the review.

For the web-scraping, the Trustpilot website was chosen as a source of reviews. To limit the reviews, the category "Travel & Vacations" was selected. The comparison analysis of the range of categories showed that "Travel & Vacations" contains a wide range of companies having a higher number of reviews and also the companies with a lower rating, which is important for data variability.

Extraction of the data was performed with the BeautifulSoup Python package for parsing HTML and XML data.

Here are some reasons why BeautifulSoup was chosen:

1) Truspilot uses clear and predictable pagination URLs (e.g. `?page=2`, `?page=3`), which makes it easy to iterate through pages without the need to simulate clicks and AJAX requests.
2) The review data is already present in the raw HTML response. No JavaScript rendering is needed to reveal it, so there's no need for tools like Selenium or Playwright.
3) BeautifulSoup is lightweight and fast when dealing with static pages.

The company with the biggest number of reviews, "Viator", has 260,646 reviews at the moment. It would take a couple of hours to scrape that much data, therefore, it was decided to cut the number of reviews that are going to be extracted. To make the data more variable for future sentiment analysis and avoid saving only positive comments, scraping was limited to 20 pages per company. Not only does it increase variety, but it also filters the comments for the most recent.

To get the list of companies with a review number higher than 10,000, the data from the reviews page of the chosen category was extracted. It was placed in the elements with <a> tags with the attribute "name=business-unit-card". Therefore, all such elements were searched with the "find_all" function.

```
cards = soup.find_all("a", attrs={"name": "business-unit-card"})
```

Further, the related information as the company name, number of reviews, and the website (it will be used to construct the link to the page from where the reviews will be extracted). At the end, we received this list:

1. Viator.com - 260646 reviews - www.viator.com
2. JustFly - 194180 reviews - justfly.com
3. Vegas.com - 163448 reviews - www.vegas.com
4. Allianz Partners USA - 126680 reviews - www.allianztravelinsurance.com
5. Vrbo - 117753 reviews - www.vrbo.com
6. FlightHub - 110842 reviews - flighthub.com
7. ASAP Tickets - 97096 reviews - www.asaptickets.com

8. Priceline – 95379 reviews – www.priceline.com
9. Way – 75701 reviews – www.way.com
10. CheapFareGuru – 70924 reviews – cheapfareguru.com
11. AirTkt – 70488 reviews – airtkt.com
12. SmartFares – 59271 reviews – www.smartfares.com
13. Guest Reservations – 44893 reviews – guestreservations.com
14. Ship Sticks – 43384 reviews – shipsticks.com
15. Reservation Counter – 41983 reviews – reservationcounter.com
16. Headout – 39221 reviews – headout.com
17. parksleepfly.com – 39038 reviews – parksleepfly.com
18. AARDY – 35330 reviews – aardy.com
19. Global Airport Parking – 30278 reviews – globalairportparking.com
20. CheapOair.com – 26256 reviews – www.cheapoair.com
21. Outdoorsy – 25825 reviews – www.outdoorsy.com
22. Bravofly – 24680 reviews – www.bravofly.com
23. Great Value Vacations – 24153 reviews – greatvaluevacations.com
24. SkyLux Travel – 23293 reviews – www.skyluxtravel.com
25. EF Educational Tours – 21528 reviews – www.eftours.com
26. TrustedHousesitters – 20898 reviews – www.trustedhousesitters.com
27. Flyus.com – 20690 reviews – flyus.com
28. BookIt.com – 18656 reviews – bookit.com
29. Vayama.com – 17620 reviews – www.vayama.com
30. BudgetAir.com – 16842 reviews – www.budgetair.com
31. RVshare – 16647 reviews – rvshare.com
32. Airport Van Rental – 16396 reviews – airportvanrental.com
33. Wisecars – 16277 reviews – www.wisecars.com
34. Mozio – 15646 reviews – www.mozio.com
35. Newark Airport Long Term Parking – 15521 reviews – newarklongtermparking.com
36. wizzair.com – 15388 reviews – www.wizzair.com
37. Vacasa – 15140 reviews – vacasa.com
38. OurBus – 14052 reviews – ourbus.com
39. CarDelMar – 13890 reviews – www.cardelmar.com
40. Blue Ridge Mountain Rentals – 13379 reviews – www.blueridgerentals.com
41. Globus – 13043 reviews – globusjourneys.com
42. Rush My Passport – 12660 reviews – rushmypassport.com
43. EaseMyTrip – 12501 reviews – easemytrip.com
44. Air Park Laguardia Airport Long Term – 12488 reviews – airparkparking.com
45. LatinOFare – 12453 reviews – latinofare.com
46. Expresso Airport Parking – 12438 reviews – www.expressoparking.com
47. Reservation Desk – 12036 reviews – reservationdesk.com
48. Mango Tours – 11921 reviews – www.mangotours.com
49. CheapAir.com – 11814 reviews – www.cheapair.com
50. Passports and Visas.com – 11790 reviews – www.passportsandvisas.com
51. Transavia – 11758 reviews – www.transavia.com
52. Greyhound Bus – 11132 reviews – www.greyhound.com

53. GalaHotels.com – 10639 reviews – www.galahotels.com
54. HotelsOne – 10568 reviews – www.hotelsone.com
55. Hotels.com – 10508 reviews – www.hotels.com
56. FlyDealFare.com – 10380 reviews – www.flydealfare.com
57. Business-Class.com – 10283 reviews – business-class.com
58. Indian Eagle – 10238 reviews – indianeagle.com
59. Expedia – 10149 reviews – www.expedia.com

Total companies with over 10,000 reviews: 59