Task:

Write a scraper that extracts the google reviews from a given Google Search term (use "aspria berlin ku'damm" for your testing) using Scrapy or any other comparable technology. The output should be a csv, containing the following information for all reviews of the target:
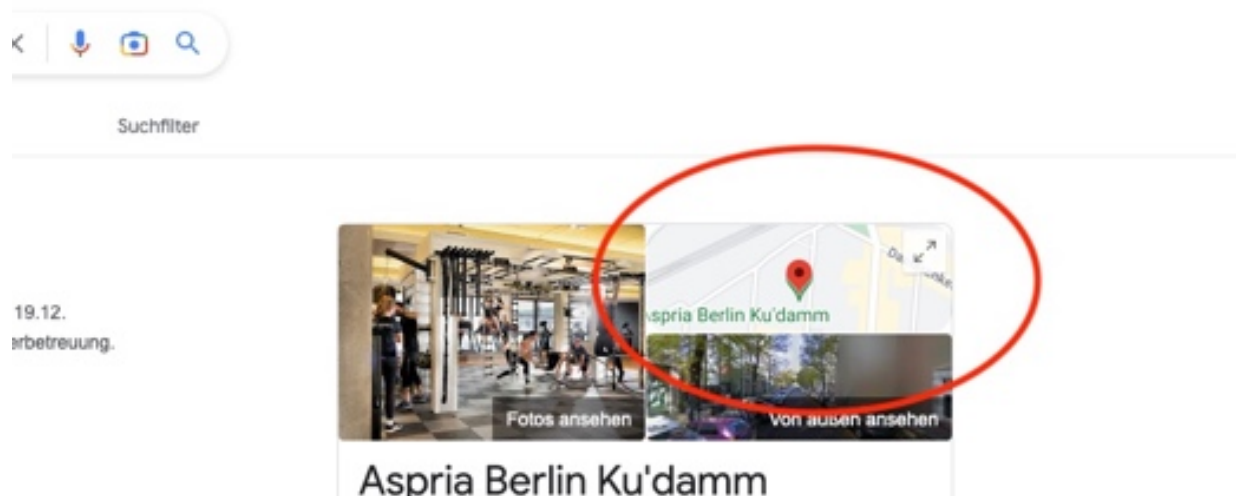
Reviewer Name (text)
Review content (text)
Full review link (text)
Rating (int)
Review Time Information (text)
Did the shop owner reply (bool)
Reply text from Owner (text)

Example of a full review URL:
https://www.google.com/maps/reviews/data=!4m8!14m7!1m6!2m5!1sChZDSUhNMG9nS0VJQ
0FnSURPODd2c0V3EAE!2m1!1s0x0:0x2faf0f02eacd864e!3m1!1s2@1:CIHM0ogKEICAgIDO87vsE
w%7CCgsI6d-8lgYQ2KOUMw%7C?hl=de&shorturl=1

The review URL for each review has to begin with
https://www.google.com/maps/reviews/data.

You find the data request with the relevant information as follows:
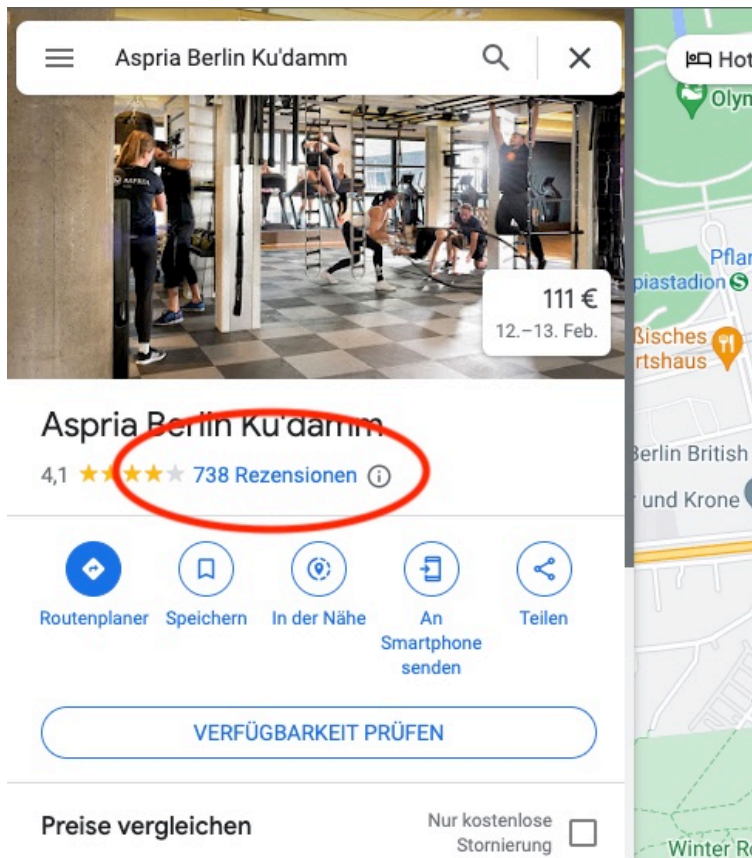
When clicking on "Rezensionen" the data is loaded:

The relevant call is a get request:

4,1

5
4
3
2
1

729 Berichte

Rezensionen werden nicht überprüft.
Google überprüft keine Rezensionen von anderen
Reisewebsites. Die Richtlinien für Rezensionen
findest du auf der jeweiligen Website.

[✎ Rezension schreiben]

Paare
3,9

Familien
4,0

Geschäftsreisende
4,0

Alleinreisende
3,9

[🔍] [⇅ Sortieren]

Alle zimmer 99 frühstück 40

sauna 41 wellness 35 reception 22

schwimmbad 13 handtücher 12

waschbecken 8 publikum 8 ruheräume 7

M Michelle Isenberg
27 Rezensionen

5/5 vor 4 Monaten bei G Google
Wir waren zu zweit für eine Nacht im Hotel und waren
rundum zufrieden!
Das Zimmer ist wirklich extrem groß, sehr schön
eingerichtet und sauber.
In vielen Bewertungen haben wir negatives über das
Frühstück gelesen - das ...
Mehr

👍 1   ⬗ Teilen

Antwort vom Inhaber vor 4 Monaten
Ohhh Michelle, da machst uns glücklich mit deinem
Feedback.
Und du hast das richtig gesehen, wir stellen unser
Team - insbesondere an der Rezeption- grade neu
zusammen. Alle hoch motiviert und offensichtlich
konzentriert :) danke für dein Verständnis und wir
freuen uns dich bald wieder begrüßen zu dürfen.

Elements   Console   Sources   **Network**   Performance   Memory   Application   Security   Lighthouse   Recorder ⬤   Performance insights ⬤

● ⊘ ▽ 🔍   ☐ Preserve log   ☐ Disable cache   No throttling ▼ 📶   ⬆ ⬇

Filter   ☐ Invert   ☐ Hide data URLs   All   **Fetch/XHR**   JS   CSS   Img   Media   Font   Doc   WS   Wasm   Manifest   Other   ☐ Has blocked cookies   ☐ Blocke

5000 ms   10000 ms   15000 ms   20000 ms   25000 ms   30000 ms   35000 ms   40000 ms   45000 ms   50000 ms   55000 ms   60000 ms   65000 ms   70000 ms   75

Name
☐ exm=sc2,per,mo,lp,ti,ds,stx,bom...
☐ exm=sc2,per,mo,lp,ti,ds,stx,bom...
☐ exm=sc2,per,mo,lp,ti,ds,stx,bom...
☐ ee=cEt90b:ws9Tlc;iFQyKf:vfuNJ...
☐ ee=cEt90b:ws9Tlc;iFQyKf:vfuNJ...
☐ listentitiesreviews?authuser=0&h...
☐ place?authuser=0&hl=de&gl=de...

× | **Headers** Payload Preview Response Initiator Timing Cookies

▼ **General**

Request URL: https://www.google.com/maps/preview/review/listentitiesreviews?authuser=0&hl=de&gl=de&pb=!1m2!1y5163465
2!1s2riYY72VE8KAxc8PkI6oQA!7e81

Request Method: GET

Status Code: ⬤ 200

Remote Address: 172.217.19.68:443

Referrer Policy: origin

▼ **Response Headers**

alt-svc: h3=":443"; ma=2592000,h3-29=":443"; ma=2592000,h3-Q050=":443"; ma=2592000,h3-Q046=":443"; ma=2592000,h3-Q043

cache-control: private, max-age=0

content-disposition: attachment; filename="f.txt"

content-encoding: br

content-length: 8817

content-security-policy-report-only: object-src 'none';base-uri 'self';script-src 'nonce-JRjaTi7IVaQKJxr4XUcUtQ' 'strict-
-uri https://csp.withgoogle.com/csp/gws/maps-tactile

content-type: application/json; charset=UTF-8