

Q

[x]


```
Python implementation: CPython
Python version       : 3.7.12
IPython version      : 5.5.0
```

```
pandas : 1.1.5
matplotlib : 3.2.2
seaborn : 0.11.2
google_play_scraper: 1.0.3
```



+ Code + Text

✓ RAM
Disk

Editing



```
[4] import json
import pandas as pd
from tqdm import tqdm

import seaborn as sns
import matplotlib.pyplot as plt

from pygments import highlight
from pygments.lexers import JsonLexer
from pygments.formatters import TerminalFormatter

from google_play_scraper import Sort, reviews, app

%matplotlib inline
%config InlineBackend.figure_format='retina'

sns.set(style='whitegrid', palette='muted', font_scale=1.2)
```



+ Code + Text

✓ RAM
Disk

Editing

```
[5] app_packages = [  
    'com.anydo',  
    'com.todoist',  
    'com.ticktick.task',  
    'com.habitrpg.android.habitica',  
    'cc.forestapp',  
    'com.oristats.habitbull',  
    'com.levor.liferpgtasks',  
    'com.habitsnow',  
    'com.microsoft.todos',  
    'prox.lab.calclock',  
    'com.gmail.jmartindev.timetune',  
    'com.artfulagenda.app',  
    'com.tasks.android',  
    'com.appgenix.bizcal',  
    'com.appxy.planner'  
]
```

```
[6] app_infos = []  
  
for ap in tqdm(app_packages):  
    info = app(ap, lang='en', country='us')  
    del info['comments']  
    app_infos.append(info)
```

100% | 15/15 [00:03<00:00, 4.68it/s]



+ Code + Text

✓ RAM
Disk

Editing



```
[7] def print_json(json_object):  
    json_str = json.dumps(  
        json_object,  
        indent=2,  
        sort_keys=True,  
        default=str  
    )  
    print(highlight(json_str, JsonLexer(), TerminalFormatter()))
```



+ Code + Text

✓ RAM Disk Editing

[8] print_json(app_infos[0])

```
{
  "adSupported": null,
  "androidVersion": "Varies",
  "androidVersionText": "Varies with device",
  "appId": "com.anydo",
  "containsAds": false,
  "contentRating": "Everyone",
  "contentRatingDescription": null,
  "currency": "USD",
  "description": "\u201c#1 to do list app out there\u201d - WSJ\u2019s Editor\u2019s Choice by Google\u2019s Choice by Google Over 30M people rely on Any.do to stay organized",
  "descriptionHTML": "\u201c#1 to do list app out there\u201d - WSJ\u2019s Editor\u2019s Choice by Google Over 30M people rely on Any.do to stay organized",
  "developer": "Any.do To-do list & Calendar",
  "developerAddress": "Any.do Inc.\u2019s Begin St. 22nd floor, Tel Aviv, Israel",
  "developerEmail": "feedback+androidtodo@any.do",
  "developerId": "5304780265295461149",
  "developerInternalId": "5304780265295461149",
  "developerWebsite": "https://www.any.do",
  "editorsChoice": false,
  "free": true,
  "genre": "Productivity",
  "genreId": "PRODUCTIVITY",
  "headerImage": "https://play-lh.googleusercontent.com/umHDP6pHYKbuh-WzrSu6_bgFmAWP4qf9WsktFVRWDCMmkIwAr7myOKka7A1VcjFoEQc",
  "histogram": [
    35010,
    12019,
    23561,
    56863,
    288360
  ],
  "icon": "https://play-lh.googleusercontent.com/zgOLUXCHkF91H8xuMTMLT17smwglPwS6jUlkVWF-cZRFjlv-Uvtman7DiHEi154fbEE",
  "inAppProductPrice": "$0.99 - $83.88 per item",
  "installs": "10,000,000+",
  "minInstalls": 10000000,
  "recentlyDeleted": null
}
```



+ Code + Text

RAM  Disk 

Editing



```
[8] "moreByDeveloper": null,
    "offersIAP": true,
    "originalPrice": null,
    "price": 0,
    "privacyPolicy": "https://www.any.do/privacy",
    "ratings": 415813,
    "recentChanges": "What's new?\r\n- Filters! You can now filter your lists by tags\r\n- A new way to create and manage lists and projects\r\n- Redesigned menus to keep things simple",
    "recentChangesHTML": "What's new?<br>- Filters! You can now filter your lists by tags<br>- A new way to create and manage lists and projects<br>- Redesigned menus to keep things",
    "released": "Nov 10, 2011",
    "reviews": 35369,
    "sale": false,
    "saleText": null,
    "saleTime": null,
    "score": 4.3263693,
    "screenshots": [
      "https://play-lh.googleusercontent.com/cGoZDZ8BpHMKadBQeLaRwGQmKvTv3_nMkKwZkwT6V0B8EU-OvAfBX6XY3wD3wvy2lY0",
      "https://play-lh.googleusercontent.com/49E2H7BYGrJkbH1dx-e08GzGpgme8wcfh8crvMyLSbCNZHbYBAI3YpR1fwi1trpTtXoF5",
      "https://play-lh.googleusercontent.com/rwql50U-W5Vn34MKy0rrzVl2wSF8Xj-utk7ym2Aa4vCKgxUqyV6Y1YrduGkv01bug",
      "https://play-lh.googleusercontent.com/zN8p1cQhwIuZ2sqeawQkC6-iwVMSMwY4QZAtKEgu9rEV76BASXxX6_XZCF3em3yBSc3",
      "https://play-lh.googleusercontent.com/x7uQgU_nqe_82ehBrRYvP6OdCFgl5lCHVYaEXAvP3l7mhFa7Bdo3NHTkhvoZ79P5l2OH",
      "https://play-lh.googleusercontent.com/cZBjFbEngmqom0VDI5GBwe-KwZetnBxS4L-lumWlnEnq5tIvAAak6NqdvJaOp5_8CGg",
      "https://play-lh.googleusercontent.com/OCmp0hmEVO8NUFhSRWe7Zv8WcxDMhKR6Lu5U6fkr8SCL2Dibakr-dQog07R_Li4MRMq",
      "https://play-lh.googleusercontent.com/SLdWLPyeKvCtbxaPH_PUe9Q0Eblt7E1XXff6pnnhohkPhzG-FBdIqK7QhCDA9j2x_w",
      "https://play-lh.googleusercontent.com/Hv5TFdmXbJ7_mQ7diXgGLZ6XbDj5kN7lwnj_Cg10ym81Pw46PRKT4nDf1jnpMTvWVA",
      "https://play-lh.googleusercontent.com/galzc8DFPM_7lIvRv9GgnB9TrcAKd6H80Fke2-U0oR6m7W8GvCB79tv5CFOKu8yAX2T7",
      "https://play-lh.googleusercontent.com/fnde6nXNd_AmyM9UZRROU1WRkokoISG3AloM_fjL2Nck08-ovo6OX34czu2WeiQ7LPiM",
      "https://play-lh.googleusercontent.com/x1s1l6gwi4oxMpoejQccfLEKshCxj1-6ebTpg8JamQxzQtKDi7e0N_Hc_Hw3i01QUs",
      "https://play-lh.googleusercontent.com/b7F7N0e209j700TMI7wBNORAwjJN7hTlht7We1Dv5u30kqoZdkgo_9B_QjAfo_YcDnU",
      "https://play-lh.googleusercontent.com/2h1Q6gdn14CfEngt9MCgxAq1qz32TFAtaTzoolI7cG_VlVjV7zIasBFpadUYrg-1Z78",
      "https://play-lh.googleusercontent.com/4b_0L0ovdv0VBowvua4loQzycri8SMubOQN3HZCPGUuglo2Vz5Gb2n0368RMRJN0A",
      "https://play-lh.googleusercontent.com/ndr_2Hm3vabWzGY4-mUYqFwsW49KfovkhN5nbFpPG4Hv9fZjXQjfa3TYf-QYH832SUD8",
      "https://play-lh.googleusercontent.com/UZUL06XshlbT5o6pJksSqrh2_g-pR-w2BYXq2h1Yoo6pDFhKEjyivnjCavGfGKIwCw",
      "https://play-lh.googleusercontent.com/zPnaXyFo_LEMhgfwz_Olnv-Tcc3a1xy26vWThpSrChGaL-KzTHRLk3GBS48v7j-fg",
      "https://play-lh.googleusercontent.com/P-bsis080AcKF903GtrpamILuJhA3B-90050KYNuHoxfB7lWtb8K2ZVvNX38Y0SFI"
```




+ Code + Text

✓ RAM
Disk

Editing



```
[8] https://play-in.googleusercontent.com/sLowLpYekVtoXaPH_P0eXQ0E0IT/E1XXtrepnnnonKInZG-FB01QKZQnCUawJ2X_W ,
https://play-lh.googleusercontent.com/Hv5Tfdnxbj7_mQZdiXgG176XbDj5kN7lwnj_Cg10ym81Pw46PRKT4nDf1jnpHlVwVA",
https://play-lh.googleusercontent.com/galzc8DFPM_1lIyRv96gnB9TrcAkD6H80Eke2-U0oR6m7WBGvCB79tv5CEOKuByAX2T1",
https://play-lh.googleusercontent.com/fnde6nXnd_AmyM9U7RRou1WkoKOiSGJAloM_jl2NCK08-ovo6OX34c7u2We1Q7LP1M",
https://play-lh.googleusercontent.com/x1s116gwi4oxMpoejQfcccLfkshCxj1-6ebTpg8JamQxzQTKDi7e0W_Hc_Hw3i01Qus",
https://play-lh.googleusercontent.com/b7FN0ez09i700TmLIwbNO8Awjn7bTLhtZWe1DvSu30knoZdkgo_9R_QjAf0_YcDnU",
https://play-lh.googleusercontent.com/2hLQGgdn14CfEngt9MCgxAg2qz32IFAtaTzooiI7cG_VUvJVzIas0EpadUYrg-1Z78",
https://play-lh.googleusercontent.com/4h_0L0ovdrQVBoWvua4loQzycrIBSMuh0Q9N3HZCPGZUJglo2VzSGb2n0368RMWRJN0A",
https://play-lh.googleusercontent.com/ndr_2Hm3vakizgY4-mUYqFwsW49KfovHkN5nbFpPG4HUG9fzjXQjfa3TYf-OYNB32SUDB",
https://play-lh.googleusercontent.com/UZUL06Xshlbt5o6pjks5qrhJ_g-pR-wJBYXq2h1Yoo6pDFhKEjyivnjCavGfGKIWCw",
https://play-lh.googleusercontent.com/zPnaXYbEo_LEMhgfw_O1nv-IccJalxy26vMIbpSrChGaL-Kz1hBLk3GBs48v2j-fg",
https://play-lh.googleusercontent.com/P_b5jSp8QAcqKF903GtrpamTlU7hA7B-9005QkYnEUHoxfB7lWtBRK2ZVvnX78Y0SEI",
https://play-lh.googleusercontent.com/9Jv2TQ78nmEVuluW3JQ5ELDWeX_IuHIZX3NO6rFUTviCCKtkSVT8tCN208Zzq0Qge-y",
https://play-lh.googleusercontent.com/nnY_hlmuLBw10BFtOkKVglv2vXe3WngVaKaejPe0Na1yXbtEGNhCf_a6sstmwZkbFyQ",
https://play-lh.googleusercontent.com/-oge_75_vm-EKoQrmzPwSLO1inIFIPRSbcxZE_vv57VLIehBVA9TvjmoS1WRN1jhG85D",
],
"similarApps": [
  "com.splendapps.splendo",
  "com.todoist",
  "com.fenchrose.reflog",
  "todolist.scheduleplanner.dailyplanner.todo.reminders",
  "com.microsoft.todos"
],
"size": "Varies with device",
"summary": "Simple planner with to-do list, tasks, reminders, checklists, calendar & agenda",
"summaryHTML": "Simple planner with to-do list, tasks, reminders, checklists, calendar & agenda",
"title": "Any.do - To do list & Calendar",
"updated": 1636559792,
"url": "https://play.google.com/store/apps/details?id=com.anydo&hl=en&gl=us",
"version": "Varies with device",
"video": null,
"videoImage": null
}
```



+ Code + Text

✓ RAM
Disk

Editing



```
[9] def format_title(title):
    sep_index = title.find('.') if title.find('.') != -1 else title.find('-')
    if sep_index != -1:
        title = title[:sep_index]
    return title[:10]

fig, axs = plt.subplots(2, len(app_infos) // 2, figsize=(14, 5))

for i, ax in enumerate(axs.flat):
    ai = app_infos[i]
    img = plt.imread(ai['icon'])
    ax.imshow(img)
    ax.set_title(format_title(ai['title']))
    ax.axis('off')
```

Any.do



Todoist



TickTick



Habitica



Forest



Habit Trac



Do It Now



HabitNow D



Microsoft



Sectograph



TimeTune



Artful Age



Tasks



Business C





+ Code + Text

✓ RAM
Disk

Editing



```
[10] app_infos_df = pd.DataFrame(app_infos)
      app_infos_df.to_csv('apps.csv', index=None, header=True)
```

```
[11] app_reviews = []

for ap in tqdm(app_packages):
    for score in list(range(1, 6)):
        for sort_order in [Sort.MOST_RELEVANT, Sort.NEWEST]:
            rvs, _ = reviews(
                ap,
                lang='en',
                country='us',
                sort=sort_order,
                count= 200 if score == 3 else 100,
                filter_score_with=score
            )
            for r in rvs:
                r['sortOrder'] = 'most_relevant' if sort_order == Sort.MOST_RELEVANT else 'newest'
                r['appId'] = ap
            app_reviews.extend(rvs)
```

100% | | 15/15 [01:11<00:00, 4.79s/it]

+ Code + Text

RAM Disk Editing ^

✓ [12] print_json(app_reviews[0])

```
{
  "appId": "com.anydo",
  "at": "2022-01-10 21:19:53",
  "content": "I had this app a LONG time ago, but recently I'm in need of location based reminders. So I figured I'd give it another try. On a positive note, I've never seen so many re",
  "repliedAt": "2022-01-11 15:51:10",
  "replyContent": "For any questions feel free to contact us at www.any.do/contact\_form. We would be happy to get an updated review once you get to know the app better :)\n\n",
  "reviewCreatedVersion": "5.15.3.0",
  "reviewId": "gp:A0qpT0NgR1qnD3AZbHvK36-Bb04pMkokJS23T1UvdVI3dMvvyJfnqr73SSvjMu3Plr0j4jutpgeYzu3r2J0QdQ",
  "score": 1,
  "sortOrder": "most_relevant",
  "thumbsUpCount": 2,
  "userImage": "https://play-lh.googleusercontent.com/a/AATXA7z7voZ4D9rF0Nx5ssdYt619gzIvbwcx0-vbFfX0=mo",
  "userName": "Glenn Anderson"
}
```

repliedAt and replyContent contain the developer response to the review. Of course, they can be missing.

How many app reviews did we get?

✓ [13] len(app_reviews)

17004



+ Code + Text



Editing

Let's save the reviews to a CSV file:

```
[14] app_reviews_df = pd.DataFrame(app_reviews)
      app_reviews_df.to_csv('reviews.csv', index=None, header=True)
```

Summary

We now have a dataset with more than 15k user reviews from 15 productivity apps.

Next, we're going to use the reviews for sentiment analysis with BERT.