lab2

October 18, 2024

1 DS 6600: Data Engineering: Lab 2

Jackson Miskill (jcm4bsq)

I did not use Generative AI for any portion of this assignment.

```python
[85]: import numpy as np
      import pandas as pd
      import requests
      import json
      from bs4 import BeautifulSoup
      import dotenv
      import os
      dotenv.load_dotenv()
```

```
[85]: True
```

2 1a)

```python
[12]: cands = pd.read_csv("data/cands22.txt", quotechar="|", header=None)
      cands.columns = ['Cycle', 'FECCandID', 'CID','FirstLastP',
                      'Party','DistIDRunFor','DistIDCurr',
                      'CurrCand','CycleCand','CRPICO','RecipCode','NoPacs']
      print(cands.shape)
      cands
```

```
(8928, 12)
```

```
[12]:       Cycle  FECCandID       CID              FirstLastP Party  \
      0      2022  H0AK00105  N00039029          Thomas Lamb (3)     3
      1      2022  H0AL01055  N00044245           Jerry Carl (R)     R
      2      2022  H0AL01063  N00044288         Wes Lambert (R)      R
      3      2022  H0AL01097  N00044750       James Averhart (D)     D
      4      2022  H0AL02087  N00030768          Martha Roby (R)     R
      ...     ...        ...        ...                      ...   ...
      8923   2022  S8WI00224  N00041929          Leah Vukmir (R)     R
      8924   2022  S8WV00119  N00041293  Paula Jean Swearengin (D)    D
      8925   2022  S8WV00127  N00035531         Evan Jenkins (R)     R
```

```
8926   2022   S8WV00135   N00041474            Bo Copley (R)      R
8927   2022   S8WV00143   N00012642       Patrick Morrisey (R)      R

     DistIDRunFor DistIDCurr CurrCand CycleCand CRPICO RecipCode NoPacs
0            AK01                                                    3N
1            AL01        AL01        Y         Y      I        RW
2            AL01                                                    RN
3            AL01                                                    DN
4            AL02                                                    RN
...           ...         ...       ...       ...    ...       ...
8923         WIS1                                                    RN
8924         WVS2                                                    DN
8925         WVS1                                                    RN
8926         WVS1                                                    RN
8927         WVS1                                                    RN

[8928 rows x 12 columns]
```

3 1b)

```
[13]: cand = pd.read_csv("data/cands22.txt", on_bad_lines='skip')
      cand
```

```
[13]:      |2022|  |H0AK00105|  |N00039029|            |Thomas Lamb (3)|  |3|  \
      0    |2022|  |H0AL01055|  |N00044245|              |Jerry Carl (R)|  |R|
      1    |2022|  |H0AL01063|  |N00044288|            |Wes Lambert (R)|  |R|
      2    |2022|  |H0AL01097|  |N00044750|         |James Averhart (D)|  |D|
      3    |2022|  |H0AL02087|  |N00030768|            |Martha Roby (R)|  |R|
      4    |2022|  |H0AL02137|  |N00045120|          |Will Dismukes (R)|  |R|
      ...   ...      ...         ...                        ...      ...
      8909 |2022|  |S8WI00224|  |N00041929|            |Leah Vukmir (R)|  |R|
      8910 |2022|  |S8WV00119|  |N00041293| |Paula Jean Swearengin (D)|  |D|
      8911 |2022|  |S8WV00127|  |N00035531|           |Evan Jenkins (R)|  |R|
      8912 |2022|  |S8WV00135|  |N00041474|              |Bo Copley (R)|  |R|
      8913 |2022|  |S8WV00143|  |N00012642|       |Patrick Morrisey (R)|  |R|

           |AK01| |    |  | | | |.1 | |.2  |3N| | |.3
      0    |AL01| |AL01|  |Y|    |Y|    |I|  |RW|  | |
      1    |AL01| |    |  | |    | |    | |  |RN|  | |
      2    |AL01| |    |  | |    | |    | |  |DN|  | |
      3    |AL02| |    |  | |    | |    | |  |RN|  | |
      4    |AL02| |    |  | |    | |    | |  |RN|  | |
      ...   ...    ...   ...    ...    ...   ...
      8909 |WIS1| |    |  | |    | |    | |  |RN|  | |
      8910 |WVS2| |    |  | |    | |    | |  |DN|  | |
      8911 |WVS1| |    |  | |    | |    | |  |RN|  | |
```

```
8912   |WVS1|  |    |  | |   | |   | |   |RN|   | |
8913   |WVS1|  |    |  | |   | |   | |   |RN|   | |

[8914 rows x 12 columns]
```

[15]: ```python
pd.read_csv("data/cands22.txt")
```

This is likely not an appropriate way to load the data. The first and main problem I see with this is that this read dropped some of the rows. The dataframe should be 8928 rows and this dataframe is 8914 rows. I am guessing that the `on_bad_lines='skip'` essentially just drops records with data configurations that it did not understand. The error that comes from trying to read in the data without any separate configuration is that `ParserError: Error tokenizing data. C error: Expected 12 fields in line 851, saw 13`. I think the reason for this is that somewhere (likely in a name), there is an extra comma (,) where that comma is being treated like the separator instead of a comma in that field. Therefore, the parser reads an extra field (column) instead of continuing over it.

## 4   2

[16]: ```python
def get_useragent():
        url = "https://httpbin.org/user-agent"
        r = requests.get(url)
        useragent = json.loads(r.text)['user-agent']
        return useragent

def make_headers(email='jcm4bsq@virginia.edu'):
    useragent = get_useragent()
    headers = {
        'User-Agent': useragent,
        'From': email
    }
    return headers

base = "https://api.nasa.gov/"
endpoint = "neo/rest/v1/feed"
key = os.getenv("NASA_API_KEY")
params = {"api_key": key,
          "start_date": "2024-10-08",
          "end_date":"2024-10-15"}
headers = make_headers()
r = requests.get(base + endpoint, params=params, headers=headers)
```

[17]: ```python
step1 = json.loads(r.text)

final_df = pd.DataFrame()
dates = list(step1['near_earth_objects'].keys())
```

```
for d in dates:
    temp = pd.json_normalize(step1['near_earth_objects'][d])
    final_df = pd.concat([final_df, temp])


print(final_df.shape)
final_df
```

(192, 18)

[17]:
```
          id neo_reference_id                name  \
0    2534676         2534676  534676 (2014 VK2)
1    3826858         3826858          (2018 QE)
2    3843427         3843427          (2019 OV3)
3   54105558        54105558          (2021 AG6)
4   54135427        54135427          (2021 GP3)
..       …               …               …
11  54487066        54487066          (2024 TU1)
12  54487280        54487280          (2024 TR2)
13  54487290        54487290          (2024 TA3)
14  54488494        54488494          (2024 TG5)
15  54489677        54489677          (2024 TW8)


                                  nasa_jpl_url  absolute_magnitude_h  \
0   https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 18.65
1   https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 27.70
2   https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 24.26
3   https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 26.88
4   https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 24.11
..                                             …                    …
11  https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 26.07
12  https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 25.56
13  https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 26.72
14  https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 27.61
15  https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm…                 27.03


    is_potentially_hazardous_asteroid  \
0                               False
1                               False
2                               False
3                               False
4                               False
..                                  …
11                              False
12                              False
13                              False
14                              False
```

```
15                        False

                                 close_approach_data  is_sentry_object  \
0   [{'close_approach_date': '2024-10-09', 'close_…             False
1   [{'close_approach_date': '2024-10-09', 'close_…              True
2   [{'close_approach_date': '2024-10-09', 'close_…             False
3   [{'close_approach_date': '2024-10-09', 'close_…             False
4   [{'close_approach_date': '2024-10-09', 'close_…             False
..                                                …               …
11  [{'close_approach_date': '2024-10-11', 'close_…             False
12  [{'close_approach_date': '2024-10-11', 'close_…             False
13  [{'close_approach_date': '2024-10-11', 'close_…             False
14  [{'close_approach_date': '2024-10-11', 'close_…              True
15  [{'close_approach_date': '2024-10-11', 'close_…             False

                                      links.self  \
0   http://api.nasa.gov/neo/rest/v1/neo/2534676?ap…
1   http://api.nasa.gov/neo/rest/v1/neo/3826858?ap…
2   http://api.nasa.gov/neo/rest/v1/neo/3843427?ap…
3   http://api.nasa.gov/neo/rest/v1/neo/54105558?a…
4   http://api.nasa.gov/neo/rest/v1/neo/54135427?a…
..                                                …
11  http://api.nasa.gov/neo/rest/v1/neo/54487066?a…
12  http://api.nasa.gov/neo/rest/v1/neo/54487280?a…
13  http://api.nasa.gov/neo/rest/v1/neo/54487290?a…
14  http://api.nasa.gov/neo/rest/v1/neo/54488494?a…
15  http://api.nasa.gov/neo/rest/v1/neo/54489677?a…

    estimated_diameter.kilometers.estimated_diameter_min  \
0                                            0.494943
1                                            0.007666
2                                            0.037373
3                                            0.011183
4                                            0.040046
..                                                  …
11                                           0.016239
12                                           0.020538
13                                           0.012038
14                                           0.007990
15                                           0.010437

    estimated_diameter.kilometers.estimated_diameter_max  \
0                                            1.106726
1                                            0.017141
2                                            0.083568
3                                            0.025006
4                                            0.089545
```

```
..                                                       …
11                                                0.036311
12                                                0.045924
13                                                0.026918
14                                                0.017867
15                                                0.023337


     estimated_diameter.meters.estimated_diameter_min  \
0                                          494.942761
1                                            7.665756
2                                           37.372743
3                                           11.182914
4                                           40.045616
..                                                 …
11                                          16.238839
12                                          20.537850
13                                          12.038018
14                                           7.990151
15                                          10.436502


     estimated_diameter.meters.estimated_diameter_max  \
0                                         1106.725658
1                                           17.141151
2                                           83.567994
3                                           25.005755
4                                           89.544719
..                                                 …
11                                          36.311148
12                                          45.924029
13                                          26.917826
14                                          17.866520
15                                          23.336729


     estimated_diameter.miles.estimated_diameter_min  \
0                                           0.307543
1                                           0.004763
2                                           0.023222
3                                           0.006949
4                                           0.024883
..                                                 …
11                                          0.010090
12                                          0.012762
13                                          0.007480
14                                          0.004965
15                                          0.006485


     estimated_diameter.miles.estimated_diameter_max  \
```

```
0                                          0.687687
1                                          0.010651
2                                          0.051927
3                                          0.015538
4                                          0.055640
..                                              …
11                                         0.022563
12                                         0.028536
13                                         0.016726
14                                         0.011102
15                                         0.014501

    estimated_diameter.feet.estimated_diameter_min  \
0                                       1623.828008
1                                         25.150118
2                                        122.613991
3                                         36.689351
4                                        131.383258
..                                              …
11                                        53.277033
12                                        67.381400
13                                        39.494810
14                                        26.214406
15                                        34.240494

    estimated_diameter.feet.estimated_diameter_max  \
0                                       3630.989809
1                                         56.237374
2                                        274.173218
3                                         82.039883
4                                        293.781896
..                                              …
11                                       119.131067
12                                       150.669390
13                                        88.313080
14                                        58.617195
15                                        76.564073

                                         sentry_data
0                                                NaN
1    http://api.nasa.gov/neo/rest/v1/neo/sentry/382…
2                                                NaN
3                                                NaN
4                                                NaN
..                                              …
11                                               NaN
12                                               NaN
```

```
13                                                   NaN
14   http://api.nasa.gov/neo/rest/v1/neo/sentry/544…
15                                                   NaN

[192 rows x 18 columns]
```

5   3)

```
[18]: key = os.getenv("GENIUS_API_KEY")
      headers = make_headers()
      headers['Authorization'] = "Bearer " + key
      headers
```

```
[18]: {'User-Agent': 'python-requests/2.32.3',
       'From': 'jcm4bsq@virginia.edu',
       'Authorization': 'Bearer
      v28gXb1_VUmcdbDYg1euMtA23I5diJzn953rhP2mAPiyMlrsqBiAON_SXaA3-XKf'}
```

```
[19]: base = "https://api.genius.com"
      endpoint = "/search"
      params = {"q": "Lizzo"}
      r = requests.get(base + endpoint, params = params, headers = headers)
      lizzo_id = json.loads(r.
        ↪text)['response']['hits'][0]['result']['primary_artist']['id']

      new_endpoint = f"/artists/{lizzo_id}/songs"
      r = requests.get(base + new_endpoint, params={"sort":"popularity", "per_page":␣
        ↪20}, headers=headers)

      lizzo_df = pd.json_normalize(json.loads(r.text)['response']['songs'])
      lizzo_df
```

```
[19]:     annotation_count          api_path                 artist_names  \
      0                  9  /songs/3241792                         Lizzo
      1                 15  /songs/4198226                         Lizzo
      2                  4  /songs/2438187                         Lizzo
      3                 12  /songs/7823996                         Lizzo
      4                 21  /songs/7066853             Lizzo & Cardi B
      5                 11  /songs/8616851                           SZA
      6                  3  /songs/3797161                         Lizzo
      7                  7  /songs/8080840                         Lizzo
      8                 11  /songs/4312315  Lizzo (Ft. Missy Elliott)
      9                  5  /songs/4286014                         Lizzo
      10                15  /songs/4312316                         Lizzo
      11                 1  /songs/4948719  Lizzo (Ft. Ariana Grande)
      12                 7  /songs/4464794                         Lizzo
```

```
13          4  /songs/9160872                         Lizzo
14         10  /songs/3241766      Charli xcx (Ft. Lizzo)
15         11  /songs/3721618                         Lizzo
16          5  /songs/7640097                         Lizzo
17          3  /songs/3201991                         Lizzo
18          5  /songs/4855379              KIDZ BOP Kids
19          5  /songs/8029110                         Lizzo


                                            full_title  \
0                                  Truth Hurts by Lizzo
1                                        Juice by Lizzo
2                                  Good As Hell by Lizzo
3                             About Damn Time by Lizzo
4                           Rumors by Lizzo & Cardi B
5                                           F2F by SZA
6                                         Boys by Lizzo
7                                       Grrrls by Lizzo
8                    Tempo by Lizzo (Ft. Missy Elliott)
9                             Cuz I Love You by Lizzo
10                              Like a Girl by Lizzo
11  Good as Hell (Remix) by Lizzo (Ft. Ariana Grande)
12                                Soulmate by Lizzo
13                                    Pink by Lizzo
14    Blame It on Your Love by Charli xcx (Ft. Lizzo)
15                                  Jerome by Lizzo
16                                 Special by Lizzo
17                                Water Me by Lizzo
18                        Truth Hurts by KIDZ BOP Kids
19                2 Be Loved (Am I Ready) by Lizzo


                           header_image_thumbnail_url  \
0   https://images.genius.com/713f196ed0f42f17fc2b…
1   https://images.genius.com/2fbb007529b4962cb0d8…
2   https://images.genius.com/9a2e5f20873abfc0b7ab…
3   https://images.genius.com/d69c5f3c6f3b7505bb86…
4   https://images.genius.com/07249b1846118211b81d…
5   https://images.genius.com/cde31e780d76c90305bc…
6   https://images.genius.com/91c735bc3cb0bfe91c4d…
7   https://images.genius.com/733bcd99684324716989…
8   https://images.genius.com/b735aaf3ce3992f757e4…
9   https://images.genius.com/78997f81e1b165ac69db…
10  https://images.genius.com/3eb5642187434d00e4f2…
11  https://images.genius.com/ad70c15d6cdb24007903…
12  https://images.genius.com/3eb5642187434d00e4f2…
13  https://images.genius.com/ec297b8459ca95ffeef0…
14  https://images.genius.com/6432cc927bb33993656b…
15  https://images.genius.com/3eb5642187434d00e4f2…
```

```
16    https://images.genius.com/733bcd99684324716989…
17    https://images.genius.com/57d9d235fbe4192ade54…
18    https://images.genius.com/1cb09ede2e20b2fd9425…
19    https://images.genius.com/733bcd99684324716989…

                                   header_image_url       id  \
0     https://images.genius.com/713f196ed0f42f17fc2b…  3241792
1     https://images.genius.com/2fbb007529b4962cb0d8…  4198226
2     https://images.genius.com/9a2e5f20873abfc0b7ab…  2438187
3     https://images.genius.com/d69c5f3c6f3b7505bb86…  7823996
4     https://images.genius.com/07249b1846118211b81d…  7066853
5     https://images.genius.com/cde31e780d76c90305bc…  8616851
6     https://images.genius.com/91c735bc3cb0bfe91c4d…  3797161
7     https://images.genius.com/733bcd99684324716989…  8080840
8     https://images.genius.com/b735aaf3ce3992f757e4…  4312315
9     https://images.genius.com/78997f81e1b165ac69db…  4286014
10    https://images.genius.com/3eb5642187434d00e4f2…  4312316
11    https://images.genius.com/ad70c15d6cdb24007903…  4948719
12    https://images.genius.com/3eb5642187434d00e4f2…  4464794
13    https://images.genius.com/ec297b8459ca95ffeef0…  9160872
14    https://images.genius.com/6432cc927bb33993656b…  3241766
15    https://images.genius.com/3eb5642187434d00e4f2…  3721618
16    https://images.genius.com/733bcd99684324716989…  7640097
17    https://images.genius.com/57d9d235fbe4192ade54…  3201991
18    https://images.genius.com/1cb09ede2e20b2fd9425…  4855379
19    https://images.genius.com/733bcd99684324716989…  8029110

    lyrics_owner_id lyrics_state                                  path  \
0           2366402     complete                 /Lizzo-truth-hurts-lyrics
1           5667167     complete                       /Lizzo-juice-lyrics
2            672397     complete                 /Lizzo-good-as-hell-lyrics
3           8280771     complete            /Lizzo-about-damn-time-lyrics
4           5748418     complete         /Lizzo-and-cardi-b-rumors-lyrics
5            498964     complete                         /Sza-f2f-lyrics
6            211404     complete                       /Lizzo-boys-lyrics
7           5748418     complete                     /Lizzo-grrrls-lyrics
8            104344     complete                      /Lizzo-tempo-lyrics
9            287691     complete             /Lizzo-cuz-i-love-you-lyrics
10          6351455     complete                /Lizzo-like-a-girl-lyrics
11          4509691     complete          /Lizzo-good-as-hell-remix-lyrics
12          6351455     complete                   /Lizzo-soulmate-lyrics
13          8280771     complete                       /Lizzo-pink-lyrics
14          4907679     complete  /Charli-xcx-blame-it-on-your-love-lyrics
15          4355603     complete                     /Lizzo-jerome-lyrics
16         11809232     complete                    /Lizzo-special-lyrics
17            93685     complete                   /Lizzo-water-me-lyrics
18          6837137     complete         /Kidz-bop-kids-truth-hurts-lyrics
```

```
19          11809232     complete     /Lizzo-2-be-loved-am-i-ready-lyrics

     …  stats.pageviews  primary_artist.api_path  \
0    …          2229533           /artists/84514
1    …           699931           /artists/84514
2    …           638364           /artists/84514
3    …           508740           /artists/84514
4    …           502919         /artists/2844720
5    …           342148           /artists/41579
6    …           319587           /artists/84514
7    …           237053           /artists/84514
8    …           220352           /artists/84514
9    …           171415           /artists/84514
10   …           138718           /artists/84514
11   …           119403           /artists/84514
12   …           117608           /artists/84514
13   …           109781           /artists/84514
14   …            91193           /artists/45349
15   …            88372           /artists/84514
16   …            81257           /artists/84514
17   …            68681           /artists/84514
18   …            63547          /artists/353676
19   …            58165           /artists/84514


                    primary_artist.header_image_url  primary_artist.id  \
0    https://images.genius.com/1b66a53cb1b8259201ae…               84514
1    https://images.genius.com/1b66a53cb1b8259201ae…               84514
2    https://images.genius.com/1b66a53cb1b8259201ae…               84514
3    https://images.genius.com/1b66a53cb1b8259201ae…               84514
4    https://images.genius.com/ba5f21e6d4cd24f34180…             2844720
5    https://images.genius.com/35f251a9b697812fc6fa…               41579
6    https://images.genius.com/1b66a53cb1b8259201ae…               84514
7    https://images.genius.com/1b66a53cb1b8259201ae…               84514
8    https://images.genius.com/1b66a53cb1b8259201ae…               84514
9    https://images.genius.com/1b66a53cb1b8259201ae…               84514
10   https://images.genius.com/1b66a53cb1b8259201ae…               84514
11   https://images.genius.com/1b66a53cb1b8259201ae…               84514
12   https://images.genius.com/1b66a53cb1b8259201ae…               84514
13   https://images.genius.com/1b66a53cb1b8259201ae…               84514
14   https://images.genius.com/4aa5e7149cf52e9329fa…               45349
15   https://images.genius.com/1b66a53cb1b8259201ae…               84514
16   https://images.genius.com/1b66a53cb1b8259201ae…               84514
17   https://images.genius.com/1b66a53cb1b8259201ae…               84514
18   https://images.genius.com/3e7f6dc58c92991638c8…              353676
19   https://images.genius.com/1b66a53cb1b8259201ae…               84514


                    primary_artist.image_url  \
```

```
0    https://images.genius.com/4ad3443f4e0ba7c59916…
1    https://images.genius.com/4ad3443f4e0ba7c59916…
2    https://images.genius.com/4ad3443f4e0ba7c59916…
3    https://images.genius.com/4ad3443f4e0ba7c59916…
4    https://images.genius.com/770c8757c098c591df22…
5    https://images.genius.com/c4afa603906452118e2c…
6    https://images.genius.com/4ad3443f4e0ba7c59916…
7    https://images.genius.com/4ad3443f4e0ba7c59916…
8    https://images.genius.com/4ad3443f4e0ba7c59916…
9    https://images.genius.com/4ad3443f4e0ba7c59916…
10   https://images.genius.com/4ad3443f4e0ba7c59916…
11   https://images.genius.com/4ad3443f4e0ba7c59916…
12   https://images.genius.com/4ad3443f4e0ba7c59916…
13   https://images.genius.com/4ad3443f4e0ba7c59916…
14   https://images.genius.com/de46f1e0502570b3e911…
15   https://images.genius.com/4ad3443f4e0ba7c59916…
16   https://images.genius.com/4ad3443f4e0ba7c59916…
17   https://images.genius.com/4ad3443f4e0ba7c59916…
18   https://images.genius.com/b493d74f772d44eb9f4a…
19   https://images.genius.com/4ad3443f4e0ba7c59916…

   primary_artist.is_meme_verified primary_artist.is_verified  \
0                            False                        True
1                            False                        True
2                            False                        True
3                            False                        True
4                            False                       False
5                            False                        True
6                            False                        True
7                            False                        True
8                            False                        True
9                            False                        True
10                           False                        True
11                           False                        True
12                           False                        True
13                           False                        True
14                            True                        True
15                           False                        True
16                           False                        True
17                           False                        True
18                           False                       False
19                           False                        True

   primary_artist.name                        primary_artist.url  \
0             Lizzo            https://genius.com/artists/Lizzo
1             Lizzo            https://genius.com/artists/Lizzo
2             Lizzo            https://genius.com/artists/Lizzo
```

```
3                 Lizzo                 https://genius.com/artists/Lizzo
4         Lizzo & Cardi B  https://genius.com/artists/Lizzo-and-cardi-b
5                   SZA                   https://genius.com/artists/Sza
6                 Lizzo                 https://genius.com/artists/Lizzo
7                 Lizzo                 https://genius.com/artists/Lizzo
8                 Lizzo                 https://genius.com/artists/Lizzo
9                 Lizzo                 https://genius.com/artists/Lizzo
10                Lizzo                 https://genius.com/artists/Lizzo
11                Lizzo                 https://genius.com/artists/Lizzo
12                Lizzo                 https://genius.com/artists/Lizzo
13                Lizzo                 https://genius.com/artists/Lizzo
14            Charli xcx             https://genius.com/artists/Charli-xcx
15                Lizzo                 https://genius.com/artists/Lizzo
16                Lizzo                 https://genius.com/artists/Lizzo
17                Lizzo                 https://genius.com/artists/Lizzo
18         KIDZ BOP Kids         https://genius.com/artists/Kidz-bop-kids
19                Lizzo                 https://genius.com/artists/Lizzo

    primary_artist.iq
0               509.0
1               509.0
2               509.0
3               509.0
4                 NaN
5              6448.0
6               509.0
7               509.0
8               509.0
9               509.0
10              509.0
11              509.0
12              509.0
13              509.0
14             4446.0
15              509.0
16              509.0
17              509.0
18                NaN
19              509.0

[20 rows x 38 columns]
```

6   4a

The `robots.txt` file from `genius.com` does not have anything in it that says that we cannot scrape anything from the website. In fact, they have specific permissions in their file that says that Claude and ChatGPT can scrape their data. I am assuming this is because they want those bots to use

the data on their website. It looks like they just don't want you to scrape data generated from their application (comments, metadata, email notifications, etc).

```python
[63]: def scrape_genius(url="https://genius.com/Lizzo-good-as-hell-lyrics"):
          r = requests.get(url)
          gumbo = BeautifulSoup(r.text, 'html.parser')
          list_of_lyrics = gumbo.body.find_all("div", attrs={"data-lyrics-container":
       ↪"true"})

          part1 = list_of_lyrics[0]
          part2 = list_of_lyrics[1]
          part3 = list_of_lyrics[2]

          lyrics = []
          for l in part1.contents:
              if str(l)[0] != "<":
                  lyrics.append(l)

          for l in part2.contents:
              if str(l)[0] != "<":
                  lyrics.append(l)

          for l in part3.contents:
              if str(l)[0] != "<":
                  lyrics.append(l)

          return lyrics
```

```python
[64]: other_tracks = gumbo.find_all("ol", attrs={"class":
      ↪"AlbumTracklist__Container-sc-123giuo-0 kGJQLs"})

      list_of_tracks = other_tracks[0].find_all("li")
      print(len(list_of_tracks))
      all_tracks = {}
      for i in range(0, len(list_of_tracks)):
          try:
              print("Track", i)
              lyrics = scrape_genius(list_of_tracks[i].a['href'])
              all_tracks[i] = lyrics
          except:
              continue

      print(all_tracks)
```

```
15
Track 0
Track 1
Track 2
```

Track 3
Track 4
Track 5
Track 6
Track 7
Track 8
Track 9
Track 10
Track 11
Track 12
Track 13
Track 14
{1: ['[Intro]', '[Verse 1]', 'Bad bitch, diamonds in my collarbone (Yee, yee)', 'Buy my whip by myself, pay my rent by myself', '[Pre-Chorus]', 'Got nothing to prove (Got nothing to prove)', "But I'ma show you how I do", '[Chorus]', 'Go and throw it (Like a girl)', "'Cause I run it (Like a girl)", 'Run it, run it (Like a girl)', 'I work my femininity', 'I make these boys get on their knees', 'Now watch me do it, watch me do it', "Look it, look it, I'ma do it", 'Like a girl (Like a girl)', 'Like a girl (Like a girl)', '[Verse 2]', 'Uh, put me on a pedestal, bet on me, bet I will', '[Pre-Chorus]', 'Got nothing to prove (Got nothing to prove)', "But I'ma show you how I do (But I'ma show you how I do)", '[Chorus]', "Find me up in Magic City bustin' hundreds by the bands", 'And I throw it (Like a girl)', 'Go and throw it (Like a girl)', "'Cause I run it (Like a girl)", 'Run it, run it (Like a girl)', 'I work my femininity', 'I make these boys get on their knees', 'Now watch me do it, watch me do it', "Look it, look it, I'ma do it", 'Like a girl (Like a girl)', 'Like a girl (Like a girl)', '[Bridge]', 'Look, so if you fight like a girl, cry like a girl', 'If you feel like a girl, then you real like a girl', 'Do your thing, run the whole damn world', 'If you fight like a girl, cry like a girl', 'Do your thing, run the whole damn world', 'Do your thing, run the whole damn world', '[Chorus]', '(Yeah, yeah, yeah, yeah, yeah, yeah)', 'Go and throw it (Like a girl)', "'Cause I run it (Like a girl) (Yeah, I run it like a girl)", 'Run it, run it (Like a girl)', 'I work my femininity (Yeah, I work it baby)', 'I make these boys get on their knees (Get on your knees, yeah)', 'Now watch me do it, watch me do it', "Look it, look it, I'ma do it (Watch me do it)", 'Like a girl (Like a girl)', 'Like a girl (Like a girl, yeah)'], 2: ['[Verse 1]', 'Louis down to my drawers', "I be drippin' so much sauce", 'Lit up like a crystal ball', "That's cool, baby, so is you", "That's how I roll", '[Pre-Chorus]', ' ', ' ', "I'm like chardonnay, get better over time ", 'Heard you say I'm not the baddest, bitch, you lie (Haha)', '[Chorus]', "It ain't my fault that I'm out here makin' news", '[Verse 2]', ' ', 'Hold up, nigga, please', "Don't make me have to take your bitch, shit ", '[Pre-Chorus]', ' ', ' ', ' ', "Heard you say I'm not the baddest, bitch, you lie ", '[Chorus]', "It ain't my fault that I'm out here makin' news", '[Bridge]', '[Chorus]', "It ain't my fault that I'm out here makin' news"], 3: ['[Intro]', 'True story', "No glory, let's go", '[Verse 1]', '[Pre-Chorus 1]', "It's a me, myself kinda attitude", '[Chorus]', "'Cause I'm my own soulmate (Yeah, yeah)", 'I know how to love me (Love me)', "I know that I'm always gonna hold me down", "Yeah, I'm my own soulmate (Yeah, yeah)", "No, I'm never lonely

(Lonely)", 'Look up in the mirror like "Damn, she the one"', '[Post-Chorus]', 'One, one, one, one, one, one (The one)', 'Like "Damn, she the one"', 'One, one, one (The one)', 'Bad bitch in the mirror like "Yeah, I\'m in love"', 'Love, love, love, love, love, love (In love)', 'Like "Yeah, I\'m in love"', 'Love, love, love (In love)', 'Look up in the mirror like "Damn, she the one"', '[Verse 2]', 'And she never tell me to exercise', 'We always get extra fries', 'And you know the sex is fire', 'And I gotta testify', 'I get flowers every Sunday', '[Pre-Chorus 2]', "True love ain't something you can buy yourself", 'True love finally happens when you by yourself', 'So if you by yourself, then go and buy yourself', '[Chorus]', "'Cause I'm my own soulmate (Yeah, yeah)", 'I know how to love me (Love me)', "I know that I'm always gonna hold me down", '(Always gonna hold me down)', "Yeah, I'm my own soulmate (Yeah, yeah)", "No, I'm never lonely (Lonely)", "(No, I'm never lonely, no)", 'Look up in the mirror like "Damn, she the one"', '[Post-Chorus]', 'One, one, one, one, one, one (The one)', '(She the one, she the one)', 'Like, "Damn, she the one"', '(She the one, she the one, she the one)', 'One, one, one (The one)', 'Bad bitch in the mirror like, "Yeah, I\'m in love"', "(I'm in love, I'm in love, I'm in love)", 'Love, love, love, love, love, love (In love)', "(I'm in love, I'm in love)", 'Like "Yeah, I\'m in love"', "(I'm in love, yeah, I'm in love)", 'Love, love, love (In love)', 'Look up in the mirror like "Damn, she the one"', '[Bridge]', "I'm the one (Ooh)", "I'm the one (You're the one)", "I'm the one (Oh, you're the one)", 'Look up in the mirror like "Damn, she the one"', 'One, one, one, one, one, one', 'Like "Damn, she the one"', 'One, one, one, one', 'Bad bitch in the mirror like "Yeah, I\'m in love"', 'Love, love, love, love, love, love (In love)', "(I'm in love, yeah, I'm in love)", 'Love, love, love (In love)', 'Look up in the mirror (Ow)', '[Chorus]', "'Cause I'm my own soulmate (Yeah, yeah)", 'I know how to love me (Love me)', "I know that I'm always gonna hold me down", 'Look up in the mirror like "Damn, she the one"', "Yeah, I'm my own soulmate (Yeah, yeah)", "No, I'm never lonely (Lonely)", 'Look up in the mirror like "Damn, she the one"'], 6: ['[Intro: Lizzo]', "I've been waitin' for this one", 'Turn it up', '[Chorus: Lizzo]', 'Slow songs, they for skinny hoes', "Can't move all of this here to one of those", "I'm a thick bitch, I need tempo (", ')', 'Fuck it up to the tempo', '[Verse 1: Lizzo]', 'Pitty-pat, pitty-pat, pitty-pitty-pat (', ')', "Look at my ass, it's fitty-fitty-fat (", ')', 'Kitty cat, kitty cat, kitty-kitty cat (', ')', 'Prrr me a glass, boy, I like my water wet (', ')', ' (', ')', 'Catch that (', ')', ' (', ')', 'Baby, baby (', ')', ' (', ')', 'He look like he could gain a little weight (', ')', 'Lick the icing off, put the rest in your face (', ')', '[Chorus: Lizzo]', 'Slow songs, they for skinny hoes', "Can't move all of this here to one of those (", ')', "I'm a thick bitch, I need tempo (", ')', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Slow songs, they for skinny hoes', 'Fuck it up to the tempo (', ')', '[Verse 2: Lizzo]', 'Fuck it up, fuck it up', "Boyfriend watchin', oh now he wanna knuckle up", "Get on this ride, baby, you gon' have to buckle up", ', ', 'All means necessary (Ayy)', '[Chorus: Lizzo]', 'Slow songs, they for skinny hoes', "Can't move all of this here to one of those (", ')', "I'm a thick bitch, I need tempo (", ')', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Slow songs, they for skinny hoes', 'Fuck it up to the tempo (Missy)', '[Verse 3: Missy Elliott]', "Let's go (Let's go), let's go (Let's

go)", 'All the fly girls head to the dance floor', 'All the thick girls down on the flrrr (On the floor)', 'Ice on my neck like brrr (Like brrr)', 'Thick girl (Thick girl), split (Split)', 'Get a bag (Get a bag), get rich (Get rich)', 'Hair done (Hair done), fixed (Fixed)', 'If you see a hater, tell him quit (Stop)', 'Get your own dough (Own dough)', 'Get your bread, own dough (Own dough)', 'Go on, ladies, head to the floor (Floor)', 'Fuck up the tempo, thick girls get low (Woo)', '[Chorus: Lizzo]', 'Slow songs, they for skinny hoes', "Can't move all of this here to one of those (", ')', "I'm a thick bitch, I need tempo (", ')', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Fuck it up to the tempo', 'Slow songs, they for skinny hoes', 'Fuck it up to the tempo', '[Outro: Lizzo]'], 7: ['[Chorus: Lizzo]', "That's exactly how I feel", "That's exactly how I feel", "That's exactly how I feel", '[Verse 1: Lizzo]', 'I woke up this morning (God)', 'Wrong side of the bed', "I don't have to 'splain it, oh (How I feel)", 'I might be a bitch (Bitch)', 'I might make a friend (Friend)', "Ain't I so amazing? (Uh-huh)", '[Pre-Chorus: Lizzo]', 'Love me or hate me', "Ooh, I ain't changing", "And I don't give a fuck, no", '[Chorus: Lizzo]', "That's exactly how I feel (Uh)", "That's exactly how I feel (Say it again)", "That's exactly how I feel (Woo)", '[Verse 2: Lizzo & ', ']', "Don't hide no emotions (Emotions)", "Wear 'em on my sleeve (On my sleeve)", 'All my feelings Gucci (', ', haha)', "Can't hold back my tears", 'That would be a crime', "'Cause I look pretty cryin'", "Oh, they ain't tell you?", '[Pre-Chorus: Lizzo]', 'Love me or hate me', "Ooh, I ain't changing", "And I don't give a fuck, no", '[Chorus: Lizzo & ', ']', "That's exactly how I feel (How I feel, ", ')', "That's exactly how I feel (How I feel, ", ')', "That's exactly how I feel", '[Verse 3: Gucci Mane]', "I feel like hurtin' they feelings (Yeah)", "I don't get mad, I get millions (Yeah)", "My heart ain't got no feelings", "My cars ain't got no ceiling", "I ain't with that acting all friendly", 'Please step away from the Bentley', "You're either not with me or with me", "The fye close by like it's Whitney", "Stay in your lane, don't tempt me", "Mr. Bust-A-Whole-Clip-'Til-It's-Empty", 'Fire chains on like burr (Burr)', "Feelin' like rockin' my fur (Fur)", 'Rolls Royce with the chauffeur (Huh?)', 'How your toes feel on a mink fur?', "It's Gucci", '[Pre-Chorus: Lizzo]', "Cry 'cause I want to", 'Smile if I want to, huh', 'Yes, I can', "Love 'cause I want to", 'Get so mad I could scream (I could scream)', 'I could scream (I could scream)', 'I could scream (Sing it)', 'Yeah, yeah, yeah, yeah', '[Chorus: Lizzo]', "That's exactly how I feel (How I feel)", "That's exactly how I feel (How I feel)", "That's exactly how I feel"], 8: ['[Verse 1]', 'Black, white, ebony all sound good to me', 'Two tone recipe, got good chemistry', 'J. F. Kennedys, kiss hood celebrities', "Don't matter to me 'cause I like everything", '[Chorus]', '(Yeah, yeah, yeah, yeah, yeah, ', ')', '[Verse 2]', "Big dick energy (Tell 'em)", "Tastes like collard greens (Tell 'em)", "Big ass patty cake (Tell 'em)", "Hey, happy days (Tell 'em)", "Bitch, don't label me (Tell 'em)", "I'm me, famously (Tell 'em)", "Nobody's safe, nobody", 'You and you and you and', '[Chorus]', '(Yeah, yeah, yeah, yeah, yeah)', '[Post-Chorus]', 'Color', '[Verse 3]', "Sick six piston (Tell 'em)", "Tiptoe kissin' (Tell 'em)", "Catcall kittens (Tell 'em)", "Whole milk lickin' (Tell 'em)", "See my vision (Tell 'em)", "Rainbow smitten (Tell 'em)", '[Chorus]', '(Yeah, yeah, yeah, yeah, yeah)', '[Post-Chorus]', 'Color', "My whole world o' color, babe", 'Oh, if I try to tell you

then you try to hurt me', "I'ma tell your momma", '[Bridge]', 'Color me, color
me, color me, color me', 'Color me, color me, color me, color me', 'Color me,
color me, color me, color me', 'Color me, color me, color me, color me',
'[Chorus]', '(Yeah, yeah, yeah, yeah, yeah)', "It's dark under the covers"], 9:
['[Chorus]', 'Heaven, help me', "If love ain't dead, I'ma kill it 'cause it's
killin' me", "Cold world, they'll be livin' in a fantasy", 'Heaven, help me',
"If love ain't dead, I'ma kill it 'cause it's killin' me", '[Verse 1]', "Y'all,
who you think you sassing? (Sassing)", "Say whoa, baby, I'm a classic
(Classic)", "Oh no, don't you get your shit rocked", 'Say whoa, ', 'Say whoa, ',
'Say whoa, what you say, what you say, sucker?', '(What you say? What you
say?)', 'Say whoa', '[Pre-Chorus]', 'If you think you got me dickmatized', '(Oh
yeah, oh yeah, oh yeah, oh yeah)', 'I need to get you out of my life', 'Can I
get an amen?', '[Chorus]', 'Heaven, help me', "If love ain't dead, I'ma kill it
'cause it's killin' me", "Cold world, they'll be livin' in a fantasy", "Got me,
it's the only thing I'll ever need", 'Heaven, help me', "If love ain't dead,
I'ma kill it 'cause it's killin' me", '[Verse 2]', "Hit 'em with the bye-bye
(Bye-bye), say whoa", "Hit 'em with the drive-by (Brrr), say whoa", "Hit 'em
with the dry cry (Dry cry), say whoa", 'Then you get a new guy, hello', '[Pre-
Chorus]', 'Even if you are the love of my life', '(Oh yeah, oh yeah, oh yeah, oh
yeah)', 'Sorry, ', '[Chorus]', 'Heaven, help me', "If love ain't dead, I'ma kill
it 'cause it's killin' me", "Cold world, they'll be livin' in a fantasy", "Got
me, it's the only thing I'll ever need", 'Heaven, help me', "If love ain't dead,
I'ma kill it 'cause it's killin' me", '[Post-Chorus]', 'Heaven, help me', "If
love ain't dead, I'ma kill it 'cause it's killin' me", 'Ooh, can I get an amen?
(Amen)', '[Outro]', "Been thinkin' about you, thinkin' about you", "Thinkin'
about you, baby", "If love didn't drive me crazy, maybe I'd be your baby",
"Trust, baby, you can't make me beg or try to change, hey"], 12: ['[Chorus]', 'I
am free, yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh no',
'I am free, yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh
no', '[Verse 1]', 'Woo', "It's hot, I think I wanna kick off both of my socks",
"I can feel it boilin' up in this pot", "A closed mouth don't get fed, is you
hungry or not?", 'Like brrt-brrt, operator', 'Caboose like, choo-choo, bless my
trainer', "(Thank God, thank God, thank God) I'm gettin' thicker", "And I got
'em sweatin' in the middle of the winter", '[Pre-Chorus]', "(Get up) I don't get
dehydrated, I moisturize it daily", 'I am my inspiration, I am my inspiration',
'(Get up) But I love you, no limit, I need to know you in it', "If you got time,
let's spend it", '[Chorus]', 'I am free, yeah, yeah', "Love you so, but if you
don't", 'I have to leave, oh no', 'I am free, yeah, yeah', "Love you so, but if
you don't", 'I have to leave, oh no', '[Verse 2]', 'Woo', "I think I need a
washrag, perspiratin'", "Gimme that sweet tea, you think I'm playin'", 'I need
it all night, no sleep', "You can get it in the kitchen if we sweatin' out the
sheets", 'Like brrt-brrt, operator', "(Thank God, thank God, thank God) I'm
gettin' thicker", 'And I can make you boil up, baby, let it simmer', '[Pre-
Chorus]', "(Get up) I don't get dehydrated, I moisturize it daily", 'I am my
inspiration, I am my inspiration', '(Get up) But I love you, no limit, I need to
know you in it', "If you got time, let's spend it", '[Chorus]', 'I am free,
yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh no', 'I am
free, yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh no',

'[Bridge]', "I'm so pretty and so ditty", 'When these boys be after kitty', 'I just snap and pivot, uh-huh, uh-huh', "I'm so pretty and so ditty", 'When these boys be after kitty', 'I just snap and pivot, uh-huh, uh-huh', '[Pre-Chorus]', "(Get up) I don't get dehydrated, I moisturize it daily", 'I am my inspiration, I am my inspiration', '(Get up) But I love you, no limit, I need to know you in it', "If you got time, let's spend it", '[Chorus]', 'I am free, yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh no', 'I am free, yeah, yeah', "Love you so, but if you don't", 'I have to leave, oh no'], 14: ['[Chorus: Lizzo]', 'I do my hair toss, check my nails', "Baby, how you feelin'? (Feelin' good as hell)", 'Hair\u2005toss,\u2005check my nails', "Baby,\u2005how you feelin'? (Feelin' good as\u2005hell)", '[Verse 1: Lizzo]', 'Woo, child, tired of the bullshit', 'Go on, dust your shoulders off, keep it moving', 'Yes, Lord, tryna get some new shit', 'In there, swimwear, going-to-the-pool shit', 'Come now, come dry your eyes', 'You know you a star, you can touch the sky', "I know that it's hard but you have to try", 'If you need advice, let me simplify', '[Pre-Chorus: Lizzo]', "If he don't love you anymore", 'Just walk your fine ass out the door', '[Chorus: Lizzo]', 'I do my hair toss, check my nails', "Baby, how you feelin'? (Feelin' good as hell)", 'Hair toss, check my nails', "Baby, how you feelin'? (Feelin' good as hell)", '(Feeling good as hell)', "Baby, how you feelin'? (Feelin' good as hell)", '[Verse 2: Lizzo & ', ']', 'Woo, girl, need to kick off your shoes', 'Gotta take a deep breath, time to focus on you', 'All the big fights, long nights that you been through', 'I got a bottle of tequila I been saving for you (', ')', '[Pre-Chorus: Ariana Grande, ', ' & ', ']', "So girl, if he don't love you anymore", ' ', '[Chorus: Lizzo, ', ' & ', ']', ' toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell)", 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (Show it off tonight)", 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (And we gon' be alright, ", ')', 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (", ')', '[Pre-Chorus: Lizzo & ', ']', 'Ooh, yeah (', '), alright (', ')', 'Listen, ', 'Then walk your fine ass out the door (', ')', '[Chorus: Lizzo, ', ' & ', ']', 'And do your hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (Woo-hoo)", 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (Show it off tonight)", 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (And we gon' make it right)", 'Hair toss, check my nails (', ')', "Baby, how you feelin'? (Feelin' good as hell) (", ')', "(Feelin' good as hell)", "Baby, how you feelin'? (", ')', '(', ', ', ', ')']}

7 5

```
[84]:  base = "https://maps.spirithalloween.com"
       endpoint = "/api/getAsyncLocations"

       params = {"template":"domain",
                 "level":"domain"}
```

```
headers = make_headers()


spooky = requests.get(base+endpoint, params=params, headers=headers)
spooky_json = json.loads(spooky.text)


list_of_locations = []
for loc in spooky_json['markers']:
    gumbo = BeautifulSoup(loc['info'])
    tag = gumbo.find_all("div", attrs={"class": "address-two"})
    list_of_locations += tag[0].contents


final_list = []
for loc in list_of_locations:
    if "former" in loc.lower().strip():
        final_list.append(loc)

final_list
pd.Series(final_list).value_counts()
```

[84]:
```
Former Sears                          3
Former Party City                     2
Former Rue 21                         2
Former CVS                            2
Former Party Depot                    2
Former Charming Charlies              1
Former DSW Shoes                      1
Former Advance Auto                   1
Upper Floor of Former Lord and Taylor 1
Former Tuesday Morning                1
Former Rite Aid                       1
Former Danker Furniture               1
Former Joann                          1
Former Big Lots                       1
Former Annas Linens                   1
Former NAPA                           1
Former Sam Ash                        1
Former Dots                           1
Former Hair Salon                     1
Former Orvis Clothing                 1
Former Ace Hardware                   1
Former Kmart                          1
Name: count, dtype: int64
```

[ ]: