generate_dummy_trackerlog

August 16, 2021

0.1 DOCUMENTATION

AUTHOR: Rebecca Wright

DATE: 6/20/21

PURPOSE

To create a dummy dataset of mock user "trackerlog" data for use in TRKR App product development. Program uses existing CSV file of Google Jobs API query results to create trackerlog entries. Program creates 50 dummy user ids, then randomly populates anywhere from 1 to 50 trackerlog entries for each user. While data points related to job posting remain static between entries, the following data points relating to the log entry are randomly generated and vary between entries: - is_active status - end_result status - date_created - date_updated

DEPENDANCIES

A CSV file must already exist containing the Google Jobs API query results. This program iteration used a file called 'job_listing_result_log_with_links.csv' containing 340 job listings.

DEVELOPMENT NOTES

This program does not protect against repeated usage of the same random job_id listing by the same user_id. Additional validation must be added to the FOR loop to protect against this.

*** #### IMPORTS

```
[103]: import datetime
       import random
       import pandas as pd
```

*** #### FUNCTIONS

```
[94]: # returns datetime.date as 2021-05-30
      def random_apply_date():
          start_date = datetime.date(2020, 6, 1)
          end_date = datetime.date(2021, 4, 1)

          time_between_dates = end_date - start_date
          days_between_dates = time_between_dates.days
          random_number_of_days = random.randrange(days_between_dates)
          random_date = start_date + datetime.timedelta(days=random_number_of_days)
          return(random_date)
```

```
[97]:  # returns datetime.date as 2021-05-30
       # produces a lapsed date from apply_date between 0 and 59 (0 representes never␣
       ↪heard back)
       def random_lapse_date(start_date):
           random_number_of_days = random.randrange(1,60)
           random_date = start_date + datetime.timedelta(days=random_number_of_days)
           return(random_date)
```

*** #### VARIABLES

Import CSV file with archived API query results

```
[43]:  jobs_df = pd.read_csv('job_listing_result_log_with_links.csv', index_col=False)
```

```
[44]:  jobs_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 340 entries, 0 to 339
Data columns (total 9 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   title                340 non-null    object
 1   company_name         340 non-null    object
 2   location             340 non-null    object
 3   via                  340 non-null    object
 4   description          340 non-null    object
 5   extensions           340 non-null    object
 6   detected_extensions  340 non-null    object
 7   job_id               340 non-null    object
 8   app_url              340 non-null    object
dtypes: object(9)
memory usage: 24.0+ KB
```

Dummy user id values

```
[98]:  user_id_vals = range(1,51)
```

Dummy job id values from row count of imported CSV file

```
[98]:  sample_jobs_id_vals = range(jobs_df.shape[0])
```

Dummy is_active values

```
[98]:  is_active_vals = [0,1]
```

Dummy range from 1 to 50 used to randomly determine number of log entries to be generated for each user

```
[98]:  num_of_apps_vals = range(1,51)
```

*** #### MAIN

Populate empty tracker_log dictionary using FOR loop

```python
[99]: tracker_log = {}
      index = 0
      for user in user_id_vals:
          for x in range(random.choice(num_of_apps_vals)):
              rand_j = random.choice(sample_jobs_id_vals)
              rand_i = random.choice(is_active_vals)
              if rand_i == 0:      #app is inactive and some resolution was reached
                  date_create = random_apply_date()
                  tracker_entry = {
                                  'user_id': user,
                                  'job_id': jobs_df.loc[rand_j]['job_id'],
                                  'company': jobs_df.loc[rand_j]['company_name'],
                                  'position': jobs_df.loc[rand_j]['title'],
                                  'location': jobs_df.loc[rand_j]['location'],
                                  'application_url': jobs_df.loc[rand_j]['app_url'],
                                  'is_active': rand_i,
                                  'end_result': random.choice([1,2,3]),   # 1 is a␣
      ↪yes, 2 is a no, 3 is a maybe
                                  'date_created': str(date_create),   #application␣
      ↪creation date
                                  'date_updated': str(random_lapse_date(date_create))
                                  }
              else:   #app is still active or pending result (is_active status is 1)
                  tracker_entry = {
                                  'user_id': user,
                                  'job_id': jobs_df.loc[rand_j]['job_id'],
                                  'company': jobs_df.loc[rand_j]['company_name'],
                                  'position': jobs_df.loc[rand_j]['title'],
                                  'location': jobs_df.loc[rand_j]['location'],
                                  'application_url': jobs_df.loc[rand_j]['app_url'],
                                  'is_active': rand_i,
                                  'end_result': 0,
                                  'date_created': str(random_apply_date()),  ␣
      ↪#application creation date
                                  'date_updated': ""
                                  }
              tracker_log[index] = tracker_entry
              index += 1
```

Convert tracker_log dictionary to dataframe

```python
[100]: final_df = pd.DataFrame.from_dict(tracker_log, orient='index')
```

Preview dataframe

```python
[101]: final_df
```

```
[101]:        user_id                                              job_id  \
        0            1   eyJqb2JfdGl0bGUiOiJTZW5pb3IgVVggRGVzaWduZXIsIE…
        1            1   eyJqb2JfdGl0bGUiOiJVWCBEZXNpZ25lciIsImNvbXBhbn…
        2            1   eyJqb2JfdGl0bGUiOiJVc2VyIEV4cGVyaWVuY2UgRGVzaW…
        3            1   eyJqb2JfdGl0bGUiOiJTZW5pb3IgVVggRGVzaWduZXIgLS…
        4            1   eyJqb2JfdGl0bGUiOiJVWCBEZXNpZ25lciAtIEFJJQSBDdX…
        …           …                                               …
        1289        50   eyJqb2JfdGl0bGUiOiJVWCBEZXNpZ25lciIsImNvbXBhbn…
        1290        50   eyJqb2JfdGl0bGUiOiJVWC9VSSBEZXNpZ25lciAtIE5ZQy…
        1291        50   eyJqb2JfdGl0bGUiOiJVWCBEZXNpZ25lciAtIElJIiwiY2…
        1292        50   eyJqb2JfdGl0bGUiOiIzRCBVSSAvIFVYIERlc2lnbmVyIi…
        1293        50   eyJqb2JfdGl0bGUiOiJKci4gVXNlciBFeHBlcmllbmNlIC…

                   company                                        position  \
        0          SiriusXM         Senior UX Designer, Automotive Programs
        1     Onward Search                                     UX Designer
        2     Finn Partners   User Experience Designer, Marketing Technology
        3      Michael Page               Senior UX Designer - Brooklyn, NY
        4     Goldman Sachs   UX Designer - RIA Custody Team - Global Markets
        …            …                                               …
        1289     Altice USA                                     UX Designer
        1290             EY             UX/UI Designer - NYC Design Studio
        1291          Eteam                               UX Designer - II
        1292        Spatial                             3D UI / UX Designer
        1293       Collabera               Jr. User Experience (UX) Designer

                        location  \
        0              New York, NY
        1     New York, NY (+1 other)
        2              New York, NY
        3              Brooklyn, NY
        4     New York, NY (+1 other)
        …                 …
        1289           New York, NY
        1290           New York, NY
        1291           New York, NY
        1292           New York, NY
        1293           New York, NY

                                    application_url  is_active  \
        0     https://www.ziprecruiter.com/c/SiriusXM/Job/Se…           0
        1     https://www.hispanicjobexchange.com/jobs/ux-de…           0
        2     https://www.linkup.com/details/9e5b89dbf0956a3…           0
        3     https://jobs.lubeninja.com/jobs/senior-ux-desi…           0
        4     https://uscareers-goldmansachs.icims.com/jobs/…           0
        …                                              …           …
        1289  https://b-jobz.com/us/web/jobposting/01701ece0…           0
```

```
1290   https://www.glassdoor.com/job-listing/ux-ui-de…          0
1291   https://www.jobilize.com/amp/job/ux-designer-i…          1
1292   https://lensa.com/3d-ui--ux-designer-jobs/new-…          1
1293   https://www.ziprecruiter.com/c/Collabera/Job/J…          0

      end_result date_created date_updated
0              2   2021-03-16   2021-03-27
1              3   2020-11-18   2020-12-25
2              1   2020-08-07   2020-09-24
3              3   2020-08-14   2020-08-30
4              1   2020-11-13   2020-11-27
…            …          …          …
1289           2   2020-10-13   2020-10-18
1290           1   2021-03-29   2021-04-04
1291           0   2020-12-22
1292           0   2020-06-05
1293           1   2020-09-30   2020-10-13

[1294 rows x 10 columns]
```

Save tracker_log dataframe dummy data to CSV file

```python
[102]: final_df.to_csv('dummy_data_tracker_log.csv', index=False)
```