FinalProject_MS4_KanaparthiVenkata

March 3, 2022

[10]:
```
import requests
from requests.exceptions import HTTPError
import pandas as pd
import numpy as np
```

[4]:
```
url ='https://api.openbrewerydb.org/breweries'
error = "N"
```

0.0.1 Reading the Data from API

[78]:
```
try:
    r = requests.get(url)
    r.raise_for_status()
except HTTPError as http_err:
    error = "Y"
except Exception as err:
    error = "O"
    errorReason = 'Connection Error Occured'
if error == 'O':
    #errorReason = err
    print(errorReason)
else:
    events = r.json()
    #print(type(events))
    df = pd.DataFrame.from_dict(events)
    #print(df)

df.head(5)
```

[78]:
```
     id                          obdb_id                      name  \
0  9094                bnaf-llc-austin                 Bnaf, LLC
1  9180           boulder-beer-co-boulder           Boulder Beer Co
2  9754  clermont-brewing-company-clermont  Clermont Brewing Company
3 10186     dimensional-brewing-co-dubuque   Dimensional Brewing Co.
4 10217   dixie-brewing-co-inc-new-orleans     Dixie Brewing Co Inc.

  brewery_type                      street address_2 address_3         city  \
0     planning                        None      None      None       Austin
```

```
1     regional            2880 Wilderness Pl      None      None      Boulder
2     planning                           None      None      None     Clermont
3     planning                           None      None      None      Dubuque
4     contract  6221 S Claiborne Ave Ste 101      None      None  New Orleans

       state county_province postal_code        country     longitude  \
0      Texas            None  78727-7602  United States          None
1   Colorado            None  80301-5401  United States  -105.2480158
2    Florida            None  34711-2108  United States          None
3       Iowa            None       52001  United States          None
4  Louisiana            None  70125-4191  United States          None

    latitude       phone                        website_url  \
0       None        None                               None
1  40.026439        None                               None
2       None        None                               None
3       None        None  http://www.dimensionalbrewing.com
4       None  5048228711                               None

                 updated_at                 created_at
0  2018-07-24T00:00:00.000Z  2018-07-24T00:00:00.000Z
1  2018-08-24T00:00:00.000Z  2018-07-24T00:00:00.000Z
2  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
3  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
4  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
```

0.0.2 1. Replace the Headers

[79]:
```
print(df.columns)
df=df.rename(columns={"name": "brewery_name", "street": "street_name",␣
 ↪"address_2": "addressline_2", "address_3": "addressline_3", "city":␣
 ↪"city_name", "state": "state_name"})
print(df.columns)
df.head(5)
```

```
Index(['id', 'obdb_id', 'name', 'brewery_type', 'street', 'address_2',
       'address_3', 'city', 'state', 'county_province', 'postal_code',
       'country', 'longitude', 'latitude', 'phone', 'website_url',
       'updated_at', 'created_at'],
      dtype='object')
Index(['id', 'obdb_id', 'brewery_name', 'brewery_type', 'street_name',
       'addressline_2', 'addressline_3', 'city_name', 'state_name',
       'county_province', 'postal_code', 'country', 'longitude', 'latitude',
       'phone', 'website_url', 'updated_at', 'created_at'],
      dtype='object')
```

[79]:
```
     id                          obdb_id              brewery_name  \
0  9094                bnaf-llc-austin                 Bnaf, LLC
1  9180          boulder-beer-co-boulder           Boulder Beer Co
2  9754  clermont-brewing-company-clermont  Clermont Brewing Company
3 10186     dimensional-brewing-co-dubuque   Dimensional Brewing Co.
4 10217   dixie-brewing-co-inc-new-orleans     Dixie Brewing Co Inc.

  brewery_type                  street_name addressline_2 addressline_3  \
0     planning                         None          None          None
1     regional           2880 Wilderness Pl          None          None
2     planning                         None          None          None
3     planning                         None          None          None
4     contract  6221 S Claiborne Ave Ste 101          None          None

     city_name state_name county_province postal_code        country  \
0       Austin      Texas            None  78727-7602  United States
1      Boulder   Colorado            None  80301-5401  United States
2     Clermont    Florida            None  34711-2108  United States
3      Dubuque       Iowa            None       52001  United States
4  New Orleans  Louisiana            None  70125-4191  United States

      longitude   latitude       phone                        website_url  \
0          None       None        None                               None
1  -105.2480158  40.026439        None                               None
2          None       None        None                               None
3          None       None        None  http://www.dimensionalbrewing.com
4          None       None  5048228711                               None

                 updated_at                created_at
0  2018-07-24T00:00:00.000Z  2018-07-24T00:00:00.000Z
1  2018-08-24T00:00:00.000Z  2018-07-24T00:00:00.000Z
2  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
3  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
4  2018-08-11T00:00:00.000Z  2018-07-24T00:00:00.000Z
```

0.0.3 2. Create a data set with required columns

[80]:
```
df1=df[['brewery_name','brewery_type','street_name','addressline_2','addressline_3','city_name
df1.head(5)
```

[80]:
```
              brewery_name brewery_type                  street_name  \
0                Bnaf, LLC     planning                         None
1          Boulder Beer Co     regional           2880 Wilderness Pl
2  Clermont Brewing Company     planning                         None
3   Dimensional Brewing Co.     planning                         None
4     Dixie Brewing Co Inc.     contract  6221 S Claiborne Ave Ste 101
```

```
  addressline_2 addressline_3    city_name state_name        country  \
0          None          None       Austin      Texas  United States
1          None          None      Boulder   Colorado  United States
2          None          None     Clermont    Florida  United States
3          None          None      Dubuque       Iowa  United States
4          None          None  New Orleans  Louisiana  United States

  postal_code
0  78727-7602
1  80301-5401
2  34711-2108
3       52001
4  70125-4191
```

0.0.4 3. Find duplicates

[81]:
```
print("Postal Code is duplictaed - {}".format(any(df1.postal_code.
 ↪duplicated())))
```

```
Postal Code is duplictaed - False
```

0.0.5 4. Find Null values

[82]:
```
print("The column street name contains NaN - %r " % df1.street_name.isnull().
 ↪values.any())
print("The column longitude contains NaN - %r " % df.longitude.isnull().values.
 ↪any())
```

```
The column street name contains NaN - True
The column longitude contains NaN - True
```

0.0.6 5. Identify outliers and bad data

[84]:
```
size_prev = df.shape
df['longitude'] = df['longitude'].astype(float, errors = 'raise')
df = df[np.isfinite(df['longitude'])]
size_after = df.shape
print("The size of previous data was - {prev[0]} rows and the size of the new␣
 ↪one is - {after[0]} rows".
      format(prev=size_prev, after=size_after))
```

```
The size of previous data was - 20 rows and the size of the new one is - 3 rows
```