Getting Coordinates

In [1]:

```
import pandas as pd
```

In [2]:

```
path = 'Before_BERT(Preprocessed).csv'
df = pd.read_csv(path, index_col = 0, keep_default_na=False)
```

In [3]:

```
! pip install geopy
! pip install Nominatim
```

```
Requirement already satisfied: geopy in c:\users\hp\anaconda3\lib\site-packa
ges (2.1.0)
Requirement already satisfied: geographiclib<2,>=1.49 in c:\users\hp\anacond
a3\lib\site-packages (from geopy) (1.50)
Requirement already satisfied: Nominatim in c:\users\hp\anaconda3\lib\site-p
ackages (0.1)
```

In [4]:

```
from geopy.geocoders import Nominatim
```

In [5]:

```
df.columns
```

Out[5]:

```
Index(['Job Title', 'Company Name', 'Industry', 'Company Location',
       'New Job (90 Days)', 'Year Started', 'Profile Headline',
       'Profile Summary', 'School', 'Degree', 'Education End',
       'Field of Study', 'Domain', 'CompanyName', 'JobTitle', 'My Network',
       'Country', 'Continent', 'FieldOfStudy'],
      dtype='object')
```

In [6]:

```
x = df['Company Location'].iloc[0]
```

In [7]:

```
from geopy.geocoders import Nominatim

address=x
geolocator = Nominatim(user_agent="immansi@gmail.com")
location = geolocator.geocode(address)
print(x)
print(location.address)
print((location.latitude, location.longitude))
```

```
Dublin, Ohio, United States
Dublin, Franklin County, Ohio, United States
(40.0992294, -83.1140771)
```

In [8]:

```
df['latitude'] = 0
df['longitude'] = 0

```

In [9]:

```
import time

total_length = len(df['Company Location'])
query_start = 0
query_end = 50

latitude = []
longitude = []
exceptions = []

i=0
for idx, addr in enumerate(df['Company Location']):
    i+=1
    geolocator = Nominatim(user_agent="ram")
    location  = geolocator.geocode(addr, timeout=100, language = 'en')
    try:
        df['latitude'].iloc[idx] = location.latitude
        df['longitude'].iloc[idx] = location.longitude
    except:
        print(idx, addr)
        exceptions.append(idx)
    if i%50 == 0:
        time.sleep(1)
```

```
C:\Users\HP\anaconda3\lib\site-packages\pandas\core\indexing.py:670: Setting
WithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/
stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pand
as.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-v
ersus-a-copy)
  iloc._setitem_with_indexer(indexer, value)
```

```
44 Greater Paris Metropolitan Region
161 Brabantine City Row
231 Mumbai Suburban district, Maharashtra, India
355 Ganderbal, Jammu & Kashmir (Union Territory), India
453 Mumbai Suburban district, Maharashtra, India
723 Dallas-Fort Worth Metroplex
731 Dallas-Fort Worth Metroplex
777 Greater Paris Metropolitan Region
794 Gothenburg Metropolitan Area
1028 Mumbai Suburban district, Maharashtra, India
1094 Mumbai Suburban district, Maharashtra, India
1095 Greater Paris Metropolitan Region
1144 Mumbai Suburban district, Maharashtra, India
1357 Midnapore Sadar, West Bengal, India
1366 Greater Kassel Area
1407 Greater Barcelona Metropolitan Area
1432 Greater Barcelona Metropolitan Area
1487 Mumbai Suburban district, Maharashtra, India
1495 Mumbai Suburban district, Maharashtra, India
1540 Chatra district, Jharkhand, India
1556 Udhampur, Jammu & Kashmir (Union Territory), India
1594 Greater Toulouse Metropolitan Area
1700 Ganderbal, Jammu & Kashmir (Union Territory), India
1852 Greater Bristol Area, United Kingdom
1918 Gothenburg Metropolitan Area
```

```
2180 Greenville-Spartanburg-Anderson, South Carolina Area
2190 Jammu & Kashmir (Union Territory), India
2209 Sri Potti Sri Ramulu Nellore, Andhra Pradesh, India
2214 Greater Toulouse Metropolitan Area
2307 Mumbai Suburban district, Maharashtra, India
2373 Ganderbal, Jammu & Kashmir (Union Territory), India
2564 Greater Syracuse-Auburn Area
2582 Mumbai Suburban district, Maharashtra, India
2605 Greater Bristol Area, United Kingdom
2730 Gothenburg Metropolitan Area
2746 Greater Barcelona Metropolitan Area
2972 Mumbai Suburban district, Maharashtra, India
3031 Dallas-Fort Worth Metroplex
3117 Mumbai Suburban district, Maharashtra, India
3235 Vishakhapatnam, Andhra Pradesh, India
3268 Mumbai Suburban district, Maharashtra, India
3313 Mumbai Suburban district, Maharashtra, India
3318 Mumbai Suburban district, Maharashtra, India
3420 Mumbai Suburban district, Maharashtra, India
3569 Greater Tampa Bay Area
3641 Mumbai Suburban district, Maharashtra, India
3664 Greater Minneapolis-St. Paul Area
3696 Greater Munich Metropolitan Area
3736 Dallas-Fort Worth Metroplex
3836 Greater Paris Metropolitan Region
3856 Dallas-Fort Worth Metroplex
3894 Dallas-Fort Worth Metroplex
3952 Mumbai Suburban district, Maharashtra, India
3982 Greater Munich Metropolitan Area
4000 Greater Munich Metropolitan Area
4048 Dallas-Fort Worth Metroplex
4060 Jammu, Jammu & Kashmir (Union Territory), India
4071 Greater Edmonton Metropolitan Area
4078 Greater Munich Metropolitan Area
4117 Frankfurt Rhine-Main Metropolitan Area
4153 Greater Minneapolis-St. Paul Area
4161 Gothenburg Metropolitan Area
4298 Greater Munich Metropolitan Area
4322 Greater Bordeaux Metropolitan Area
4505 Greater Munich Metropolitan Area
4512 Greater Pittsburgh Region
4529 Greater Munich Metropolitan Area
4536 Greater Pittsburgh Region
4567 Mumbai Suburban district, Maharashtra, India
4675 Greater Minneapolis-St. Paul Area
4725 Brande, Middle Jutland, Denmark
```

In [10]:

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="my_user_agent")
for idx in exceptions:
    addr = df.iloc[idx]['Company Location'].split()
    for word in addr:
        if word == "Greater" or word == "City" or word=="Row":
            continue
        loc = geolocator.geocode(word)
        if loc is None:
            continue
        else:
            df['latitude'].iloc[idx] = location.latitude
            df['longitude'].iloc[idx] = location.longitude
            break
```

In [11]:

```
coords = df[['latitude', 'longitude']]
```

In [12]:

```
coords.to_csv('Coordinates.csv', index = False)
```