Project 1: NoSQL

CSE 511: Data Processing at Scale

Lucy Zhang
Ira A. Fulton Schools of Engineering, Arizona State University
lzhan370@asu.edu

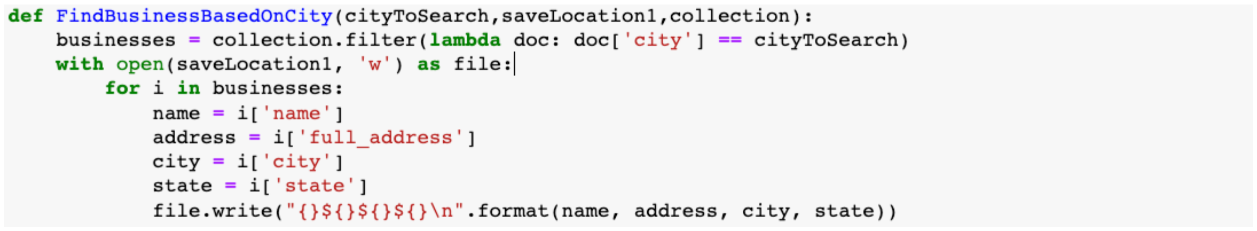
Reflection:

The goal of this project is to implement two functions to find businesses from a NoSQL database (UnQLite) based on two different criteria: city and geographical proximity to a specific location, and a distance calculation function. The data set includes fields such as business name, full address, city, state, and geographical coordinates.

Firstly, I implemented a function named ‘distanceFunction’ that calculates the geographical distance between two coordinates in miles, based on the formula given in the project instructions. This function accepts two pairs of latitude and longitude and returns the distance between them.

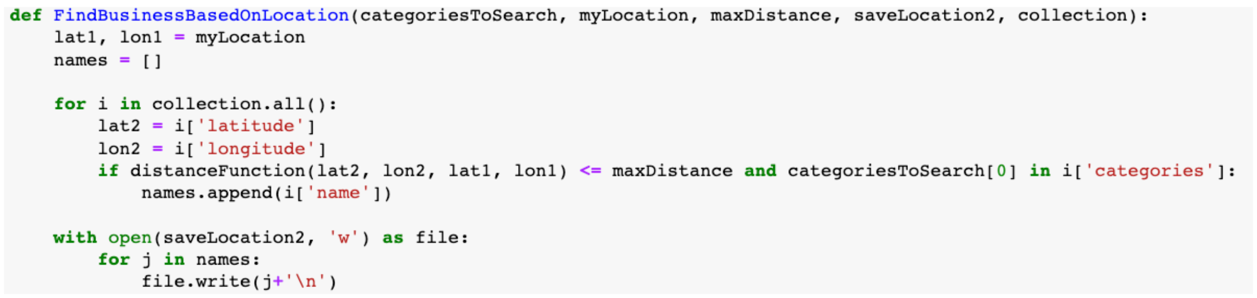
```
import math
def distanceFunction(lat2, lon2, lat1, lon1):
    R = 3959
    p1 = math.radians(lat1)
    p2 = math.radians(lat2)
    d1 = math.radians(lat2 - lat1)
    d2 = math.radians(lon2 - lon1)

    a = math.sin(d1/2) * math.sin(d1/2) + math.cos(p1) * math.cos(p2) * math.sin(d2/2) * math.sin(d2/2)
    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1 - a))
    d = R * c
    return d
```

The ‘FindBusinessBasedOnCity’ function was created to retrieve businesses based on the city they're located in. I used the filter function to iterate through the entire database collection and return businesses located in the city specified. The results are then written to a file in the required format.



```python
def FindBusinessBasedOnCity(cityToSearch,saveLocation1,collection):
    businesses = collection.filter(lambda doc: doc['city'] == cityToSearch)
    with open(saveLocation1, 'w') as file:
        for i in businesses:
            name = i['name']
            address = i['full_address']
            city = i['city']
            state = i['state']
            file.write("{}${}${}${}\n".format(name, address, city, state))
```

Finally, the ‘FindBusinessBasedOnLocation’ function was written to find businesses within a certain radius from a specified location, and which belong to a specific category. I used the ‘distanceFunction’ to determine if the business location is within the given maximum distance. If a business meets these criteria, its name is added to the list. The names are then written to an output file.



```python
def FindBusinessBasedOnLocation(categoriesToSearch, myLocation, maxDistance, saveLocation2, collection):
    lat1, lon1 = myLocation
    names = []

    for i in collection.all():
        lat2 = i['latitude']
        lon2 = i['longitude']
        if distanceFunction(lat2, lon2, lat1, lon1) <= maxDistance and categoriesToSearch[0] in i['categories']:
            names.append(i['name'])

    with open(saveLocation2, 'w') as file:
        for j in names:
            file.write(j+'\n')
```

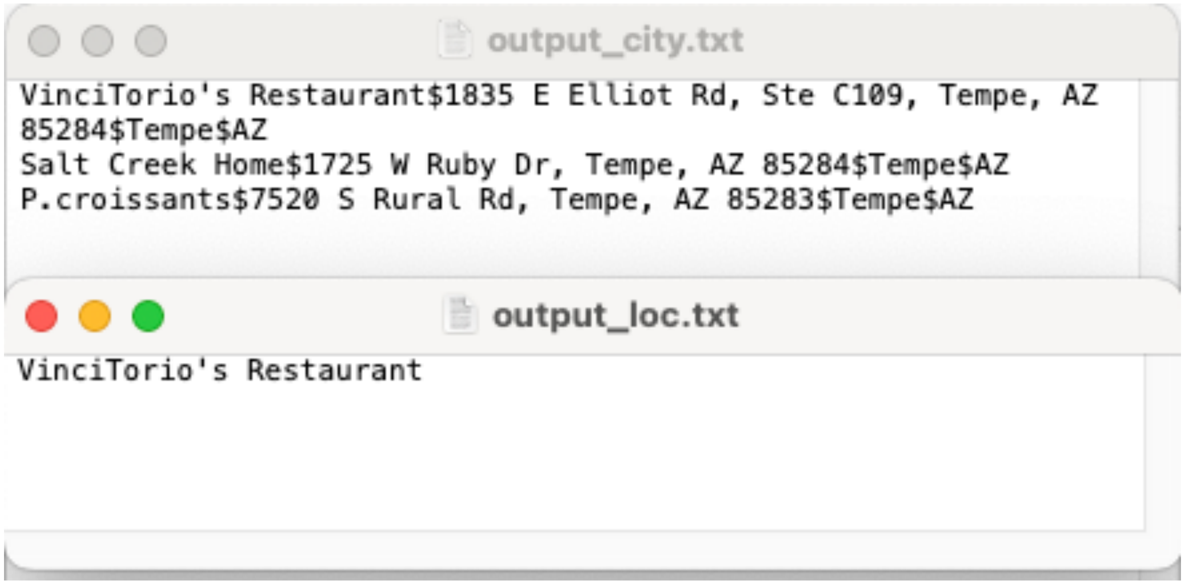
Lessons Learned:

Through this project, I gained valuable insights into working with NoSQL databases and performing data operations using Python. The project helped me understand the benefits of NoSQL databases, such as flexibility in data storage and high performance in handling large quantities of data.

I learned how to implement practical data retrieval functions using Python and write the output to files. Additionally, I discovered how to calculate geographical distances using latitude and longitude coordinates, an important concept for location-based data manipulation.

The project provided hands-on experience in data manipulation, database interaction, file I/O operations, and applying mathematical formulas in real-world scenarios. I believed this knowledge and the skills developed during this project will be extremely beneficial for similar tasks and projects in the future.

Output:



```
VinciTorio's Restaurant$1835 E Elliot Rd, Ste C109, Tempe, AZ
85284$Tempe$AZ
Salt Creek Home$1725 W Ruby Dr, Tempe, AZ 85284$Tempe$AZ
P.croissants$7520 S Rural Rd, Tempe, AZ 85283$Tempe$AZ
```

```
VinciTorio's Restaurant
```

Result:

```
Correct! Your FindBusinessBasedOnLocation function passes these test c
ases. This does not cover all possible edge cases, so make sure your f
unction does before submitting.
```

```
Correct! You FindBusinessByCity function passes these test cases. This
does not cover all possible test edge cases, however, so make sure tha
t your function covers them before submitting!
```