Project Report
Prerna Singh, 2013149

Question:

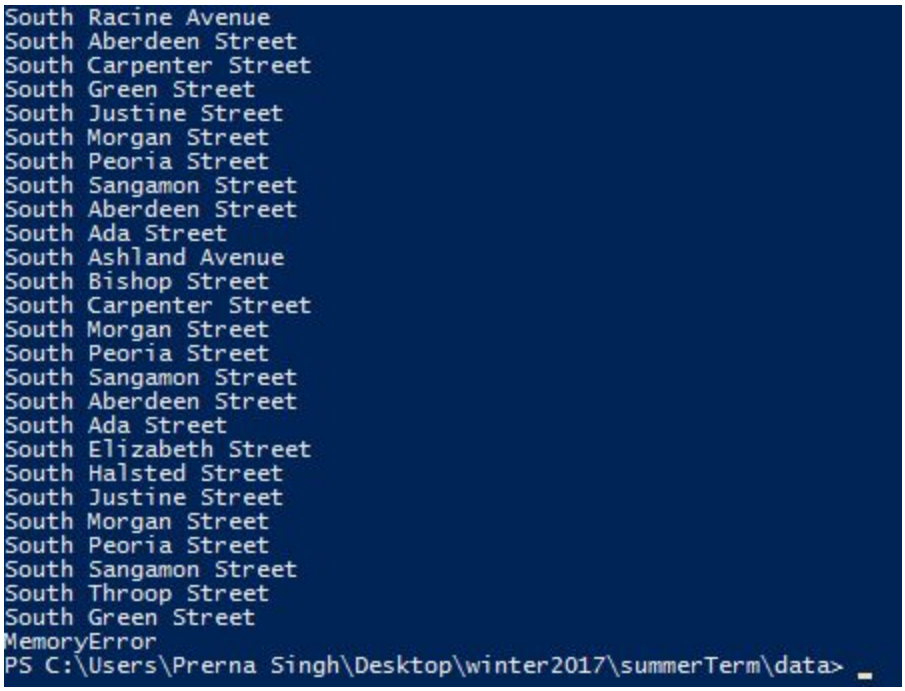
Download Chicago map from Open Street Map and arrange all the street names in the chronological order and generate an output in a JSON format. You should ensure that street names are proper. For example, “Street" shouldn’t be represented as “St.” or something else. Similarly, “Avenue” shouldn’t be represented as “Ave.” or something else.

Solution:

I followed the following steps:

i) Dataset

- Downloaded the OpenStreetMap dataset for Chicago. The dataset was about 2GB.
- **Problem Faced** - Large size of the dataset(2 GB), it gave memory error on my system
- So, I took a smaller dataset - example.osm



ii) Checked the element tag for nodes and way. Then looked for street names for auditing them.

iii) Auditing Street Names

- Made street names proper. For example - converted “N. Lincoln Ave” to “North Lincoln Ave”.
- The image below outputs some of the improved street names:

```
PS C:\Users\Prerna Singh\Desktop\winter2017\summerTerm>
PS C:\Users\Prerna Singh\Desktop\winter2017\summerTerm> cd data
PS C:\Users\Prerna Singh\Desktop\winter2017\summerTerm\data> python code.py
{'Ave': set(['N. Lincoln Ave', 'North Lincoln Ave']),
 'Rd.': set(['Baldwin Rd.']),
 'St.': set(['West Lexington St.'])}
North Lincoln Avenue
North Lincoln Avenue
Baldwin Road
West Lexington Street
PS C:\Users\Prerna Singh\Desktop\winter2017\summerTerm\data>
```

iv) Created a Json file

- The output with corrected street names were dumped in a json file.

v) Insert Data in MongoDB

- The JSON data was inserted into MongoDB and then sorted.

vi) Sorting

- Created a pipeline to sort the data.
- In the pipeline, I first checked each element for “address.street” field.
- If this field existed, the element was added to a group.
- This group was then sorted.

The output obtained after sorting is as follows:

Output

```
PS C:\Users\Prerna Singh\Desktop\winter2017\MongoDB> python main.py
chicago db is created
[{u'_id': u'Baldwin Rd.'},
 {u'_id': u'Baldwin Road'},
 {u'_id': u'N. Lincoln Ave'},
 {u'_id': u'North Lincoln Ave'},
 {u'_id': u'North Lincoln Avenue'},
 {u'_id': u'West Lexington St.'},
 {u'_id': u'West Lexington Street'}]
PS C:\Users\Prerna Singh\Desktop\winter2017\MongoDB>
```

```
PS C:\Users\Prerna Singh\Desktop\winter2017\MongoDB> python main.py
chicago db is created
{u'_id': ObjectId('596f1e70e6cade1f744e21e7'), u'type': u'node', u'id': u'261114295', u'created': {
tamp': u'2012-03-28T18:31:23Z'}}
[{u'_id': u'Baldwin Road'},
 {u'_id': u'North Lincoln Avenue'},
 {u'_id': u'West Lexington Street'}]
PS C:\Users\Prerna Singh\Desktop\winter2017\MongoDB>
```

How to run the code?

- Run main.py. This will create a .json file in the same folder.
- Comment line 100 (test.test())
- Uncomment from line 101 to 110.
- Run main.py again. We will get the sorted output.

Code

- Changes made by me - ### **changes made** ## is written above every function that I added or changed.
- Rest of the functions are taken from Udacity.