Capital Bikeshare Analyseprojekt

- Phase: Data Understanding -

Dieses Notebook stellt auf Basis der Empfehlungen von Touristen, die 15 beliebtesten Sehenswürdigkeiten der Stadt Washington DC für die folgenden Notebooks bereit.

Die Basis bildet eine KML-Datei, KML steht für Keyhole Markup Language. KML-Dateien beschreiben Geodaten und sind ähnlich aufgebaut wie eine XML Datei. Für die Erstellung der KML Datei wurde auf folgende Touristenempfehlungsseiten zugegriffen:

    https://washington.sehenswuerdigkeiten-online.de/sehenswuerdigkeiten/sights_washington.html

    https://globusliebe.com/washington-dc-sehenswuerdigkeiten/

Um die Geodaten in Form einer KML-Datei zu erhalten wurde diese mit Google Earth (Erreichbar unter: https://earth.google.com/web/ (https://earth.google.com/web/)) erstellt.

Die für dieses Notebook benötigte Rohdatei heißt: sights_washington.kml

Dieses Notebook erstellt die Datei sightseeing_coordinates.pkl.

```
In [1]: RAWDATA_PATH = '../data/raw'
        DATA_PATH = '../data/'
```

```
In [2]: from pykml import parser
        import pandas as pd
```

```
In [3]: filename = DATA_PATH + 'sights_washington.kml'
```

```
In [4]: ## Quelle: https://gist.github.com/nishadhka/3ba801ca980da5b76004631c1935f604
        # Angepasst an die eigenen Bedürfnisse.
        with open(filename):
            folder = parser.parse(filename).getroot().Document

        plnm=[]
        cordi=[]
        for pm in folder.Placemark:
            plnm1=pm.name
            plcs1=pm.Point.coordinates
            plnm.append(plnm1.text)
            cordi.append(plcs1.text)

        db=pd.DataFrame()
        db['place_name']=plnm
        db['cordinates']=cordi

        def dump(obj):
            for attr in dir(obj):
                if hasattr( obj, attr ):
                    print( "obj.%s = %s" % (attr, getattr(obj, attr)))

        db['Longitude'], db['Latitude'],db['value'] = zip(*db['cordinates'].apply(lambda x: x.split(',', 2)))

        db
```

Out[4]:

	place_name	cordinates	Longitude	Latitude	value
0	Union Station	-77.0057621,38.8977859,13.56601484618045	-77.0057621	38.8977859	13.56601484618045
1	Kapitol	-77.0090505,38.8899389,0	-77.0090505	38.8899389	0
2	Library of Congress	-77.00471899999999,38.888684,0	-77.00471899999999	38.888684	0
3	United States Botanic Garden	-77.0128833,38.8881451,0	-77.0128833	38.8881451	0
4	Smithsonian Institution Building	-77.0260229,38.8887861,0	-77.0260229	38.8887861	0
5	National Museum of Natural History	-77.02606539999999,38.8912662,0	-77.02606539999999	38.8912662	0
6	Newseum Residences	-77.0196499,38.8932873,0	-77.0196499	38.8932873	0
7	Weißes Haus	-77.0365298,38.8976763,0	-77.0365298	38.8976763	0
8	Washington Monument	-77.0352791,38.8894838,0	-77.0352791	38.8894838	0
9	Lincoln Memorial	-77.05017599999999,38.8892686,0	-77.05017599999999	38.8892686	0
10	Vietnam Veterans Memorial	-77.04771319999999,38.8912933,0	-77.04771319999999	38.8912933	0
11	Jefferson Memorial	-77.0364536,38.88138060000001,0	-77.0364536	38.88138060000001	0
12	National Air and Space Museum	-77.01986789999999,38.88816010000001,0	-77.01986789999999	38.88816010000001	0
13	National Mall	-77.0229772,38.8896198,0	-77.0229772	38.8896198	0
14	Smithsonian National Zoological Park	-77.04978439999999,38.9296156,0	-77.04978439999999	38.9296156	0

```
In [5]: db_test = db
```

```
In [6]: # In der Spalte value sind nur Nullen erlaubt.
        db_test['value'] =0
```

```
In [7]: longToList = db_test['Longitude'].apply(lambda x: x[:11]).tolist()
```

```
In [8]: latToList = db_test['Latitude'].apply(lambda x: x[:10]).tolist()
```

```
In [9]: latToList
```

```
Out[9]: ['38.8977859',
         '38.8899389',
         '38.888684',
         '38.8881451',
         '38.8887861',
         '38.8912662',
         '38.8932873',
         '38.8976763',
         '38.8894838',
         '38.8892686',
         '38.8912933',
         '38.8813806',
         '38.8881601',
         '38.8896198',
         '38.9296156']
```

```
In [10]: db_test['Latitude'] = pd.DataFrame({'Latitude':latToList})
         db_test['Longitude'] = pd.DataFrame({'Longitude':longToList})
         db_test
```

Out[10]:

	place_name	cordinates	Longitude	Latitude	value
0	Union Station	-77.0057621,38.8977859,13.56601484618045	-77.0057621	38.8977859	0
1	Kapitol	-77.0090505,38.8899389,0	-77.0090505	38.8899389	0
2	Library of Congress	-77.00471899999999,38.888684,0	-77.0047189	38.888684	0
3	United States Botanic Garden	-77.0128833,38.8881451,0	-77.0128833	38.8881451	0
4	Smithsonian Institution Building	-77.0260229,38.8887861,0	-77.0260229	38.8887861	0
5	National Museum of Natural History	-77.02606539999999,38.8912662,0	-77.0260653	38.8912662	0
6	Newseum Residences	-77.0196499,38.8932873,0	-77.0196499	38.8932873	0
7	Weißes Haus	-77.0365298,38.8976763,0	-77.0365298	38.8976763	0
8	Washington Monument	-77.0352791,38.8894838,0	-77.0352791	38.8894838	0
9	Lincoln Memorial	-77.05017599999999,38.8892686,0	-77.0501759	38.8892686	0
10	Vietnam Veterans Memorial	-77.04771319999999,38.8912933,0	-77.0477131	38.8912933	0
11	Jefferson Memorial	-77.0364536,38.88138060000001,0	-77.0364536	38.8813806	0
12	National Air and Space Museum	-77.01986789999999,38.88816010000001,0	-77.0198678	38.8881601	0
13	National Mall	-77.0229772,38.8896198,0	-77.0229772	38.8896198	0
14	Smithsonian National Zoological Park	-77.04978439999999,38.9296156,0	-77.0497843	38.9296156	0

```
In [11]: sights_db = db_test
```

```
In [12]: sights_db
```

Out[12]:

	place_name	cordinates	Longitude	Latitude	value
0	Union Station	-77.0057621,38.8977859,13.56601484618045	-77.0057621	38.8977859	0
1	Kapitol	-77.0090505,38.8899389,0	-77.0090505	38.8899389	0
2	Library of Congress	-77.00471899999999,38.888684,0	-77.0047189	38.888684	0
3	United States Botanic Garden	-77.0128833,38.8881451,0	-77.0128833	38.8881451	0
4	Smithsonian Institution Building	-77.0260229,38.8887861,0	-77.0260229	38.8887861	0
5	National Museum of Natural History	-77.02606539999999,38.8912662,0	-77.0260653	38.8912662	0
6	Newseum Residences	-77.0196499,38.8932873,0	-77.0196499	38.8932873	0
7	Weißes Haus	-77.0365298,38.8976763,0	-77.0365298	38.8976763	0
8	Washington Monument	-77.0352791,38.8894838,0	-77.0352791	38.8894838	0
9	Lincoln Memorial	-77.05017599999999,38.8892686,0	-77.0501759	38.8892686	0
10	Vietnam Veterans Memorial	-77.04771319999999,38.8912933,0	-77.0477131	38.8912933	0
11	Jefferson Memorial	-77.0364536,38.88138060000001,0	-77.0364536	38.8813806	0
12	National Air and Space Museum	-77.01986789999999,38.88816010000001,0	-77.0198678	38.8881601	0
13	National Mall	-77.0229772,38.8896198,0	-77.0229772	38.8896198	0
14	Smithsonian National Zoological Park	-77.04978439999999,38.9296156,0	-77.0497843	38.9296156	0

```
In [13]: sights_db[["Longitude", "Latitude"]] = sights_db[["Longitude", "Latitude"]].apply(pd.to_numeric)
```

```
In [14]: sights_db.dtypes
```

Out[14]:
```
place_name     object
cordinates     object
Longitude      float64
Latitude       float64
value          int64
dtype: object
```

```
In [15]: sights_db.to_pickle(DATA_PATH + 'sightseeing_coordinates.pkl')
```

```
In [ ]:
```