Entfernungen zwischen den Capital Bikeshare Fahrradstationen

- Phase: Data Understanding -

Die Ziele dieses Notebooks sind die Analyse der Entfernungen zwischen den Fahrradstationen, welche durch eine Luftlinienberechnung der Start- und Endstationen durchgeführt wurden sowie die Darstellung der Sehenswürdigkeiten und Fahrradstationen auf einer Karte.

Dieses Notebook nutzt die folgenden Dateien: trips_raw.pkl, Station_Data.csv, trips_clean.pkl, sightseeing_coordinates.pkl

Folgende Dateien werden durch dieses Notebook erzeugt: station_location.pkl, airdistance_tripdata_coordinates.pkl

In [1]:
```
import datetime
import pandas as pd
import matplotlib.pyplot as plt
import geopy.distance
from datetime import date
from workalendar.core import Calendar
```

In [2]:
```
RAWDATA_PATH = '../data/raw'
DATA_PATH = '../data/'
TRIPS_FILE = 'trips_raw.pkl'
STATION_LOCATION = RAWDATA_PATH+'/Station_Data.csv'
```

In [3]:
```
df_station_data = pd.read_csv(STATION_LOCATION)
```

In [4]:
```
df_station_data.to_pickle(DATA_PATH+'station_location.pkl')
```

In [5]:
```
df_trips=pd.read_pickle(DATA_PATH+'trips_clean.pkl')
```

In [6]:
```
df_station_data = df_station_data[['TERMINAL_NUMBER', 'LATITUDE', 'LONGITUDE']]
```

In [7]:
```
df_station_data.head()
```

Out[7]:

	TERMINAL_NUMBER	LATITUDE	LONGITUDE
0	31612	38.894758	-76.997114
1	31226	38.916442	-77.068200
2	31227	38.900283	-77.029822
3	31228	38.899700	-77.023086
4	31504	38.932514	-76.992889

In [8]:
```
# Wie viele Werte sind pro Spalte vorhanden und nicht null?
# Ergebnis: 596 Stationen
df_station_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 596 entries, 0 to 595
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   TERMINAL_NUMBER  596 non-null    int64
 1   LATITUDE         596 non-null    float64
 2   LONGITUDE        596 non-null    float64
dtypes: float64(2), int64(1)
memory usage: 14.1 KB
```

In [9]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import math
```

In [10]:
```
from math import radians, cos, sin, asin, sqrt

def haversine(lon1, lat1, lon2, lat2):
    # Dezimal Zahlen  umrechnen
    lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])

    # Haversine Formel
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat/2)**2 + cos(lat1) * cos(lat2) * sin(dlon/2)**2
    c = 2 * asin(sqrt(a))
    # Erdradius:
    erdradius = 6371 * c
    return erdradius * 1000
```

In [11]:
```
def listenMethode(longst1, latst1, longst2, latst2):
    ergebnislist = [];
    x = 0;
    while x < len(longst1):
        t = haversine(longst1[x], latst1[x],longst2[x],latst2[x])
        ergebnislist.append(t)
        x = x+1
    return ergebnislist
```

In [12]:
```
df_complete =pd.merge(df_trips, df_station_data, right_on='TERMINAL_NUMBER', left_on='start_station_id').drop('TERM
INAL_NUMBER', axis=1).rename(columns={'LATITUDE':'Latitude_start_station', 'LONGITUDE':'Longitude_start_station'})
```

In [13]:
```
df_complete =pd.merge(df_complete, df_station_data, right_on='TERMINAL_NUMBER', left_on='end_station_id').drop('TER
MINAL_NUMBER', axis=1).rename(columns={'LATITUDE':'Latitude_end_station', 'LONGITUDE':'Longitude_end_station'})
```

In [14]:
```
# Einführung eines neuen Dataframes zum Speichern der Entfernungen
df_station_data_test = df_complete
```

In [15]:
```
# Hier soll für jede Zeile im Dataframe der  Wert für Entfernung gefüllt werden
# Pseudocode
# For each Zeile  do
# haversine(Latitude_start_station,Longitude_start_station,Latitude_end_station,Longitude_end_station)
# Schreibe das Ergebnis in das Feld AirDistance der Zeile
# Wiederhole es für jede Zeile
```

In [16]:
```
df_station_data_test
```

Out[16]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Lati
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	
...	...	...	...	...	...	...	...	...	...	...	
10232877	2017-10-14 12:46:54	2017-10-14 13:14:53	32211	32211	W23202	Casual	2017-10-14	12	2017-10-14	13	
10232878	2017-06-18 12:34:17	2017-06-18 16:42:29	32211	32211	W23247	Casual	2017-06-18	12	2017-06-18	16	
10232879	2017-08-06 18:42:16	2017-08-06 19:12:20	32211	32211	W23247	Casual	2017-08-06	18	2017-08-06	19	
10232880	2016-11-05 19:12:56	2016-11-05 20:11:14	32211	32211	W23261	Casual	2016-11-05	19	2016-11-05	20	
10232881	2016-11-06 15:11:39	2016-11-06 15:39:41	32211	32211	W23261	Casual	2016-11-06	15	2016-11-06	15	

10232882 rows × 14 columns

In [17]:
```
df_station_data_test
```

Out[17]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Lati
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	
...	...	...	...	...	...	...	...	...	...	...	...
10232877	2017-10-14 12:46:54	2017-10-14 13:14:53	32211	32211	W23202	Casual	2017-10-14	12	2017-10-14	13	
10232878	2017-06-18 12:34:17	2017-06-18 16:42:29	32211	32211	W23247	Casual	2017-06-18	12	2017-06-18	16	
10232879	2017-08-06 18:42:16	2017-08-06 19:12:20	32211	32211	W23247	Casual	2017-08-06	18	2017-08-06	19	
10232880	2016-11-05 19:12:56	2016-11-05 20:11:14	32211	32211	W23261	Casual	2016-11-05	19	2016-11-05	20	
10232881	2016-11-06 15:11:39	2016-11-06 15:39:41	32211	32211	W23261	Casual	2016-11-06	15	2016-11-06	15	

10232882 rows × 14 columns

In [18]:
```
df_station_data_test.head()
```

Out[18]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Latitude_sta
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	

In [19]:
```
Lst_long_start_station = list(df_station_data_test["Longitude_start_station"])
len(Lst_long_start_station)
```

Out[19]: 10232882

In [20]:
```
Lst_long_end_station = list(df_station_data_test["Longitude_end_station"])
len(Lst_long_end_station)
```

Out[20]: 10232882

In [21]:
```
Lst_lat_end_station = list(df_station_data_test["Latitude_end_station"])
len(Lst_lat_end_station)
```

Out[21]: 10232882

In [22]:
```
Lst_lat_start_station = list(df_station_data_test["Latitude_start_station"])
len(Lst_lat_start_station)
```

Out[22]:
```
10232882
```

In [23]:
```
# Reihenfolge der Parameter: longst1, latst1, longst2, latst2):
ergebnislist = listenMethode(Lst_long_start_station,Lst_lat_start_station,Lst_long_end_station,Lst_lat_end_station)
```

In [24]:
```
len(ergebnislist)
```

Out[24]:
```
10232882
```

In [25]:
```
ergebnislist
```

```
Out[25]: [2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 2862.058435528746,
 4470.925786658126,
 4470.925786658126,
 4470.925786658126,
```

4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
4470.925786658126,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,

```
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
1671.2640458023147,
...]
```

In [26]:
```
# Die Erzeugte Liste als Distanz an den Dataframe anhängen
df_station_data_test["distance_Float"] = pd.Series(ergebnislist)
```

In [27]:
```
df_station_data_test.head()
```

Out[27]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Latitude_sta
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	

In [28]:
```
df_station_data_test.distance_Float.value_counts()
```

Out[28]:
```
0.000000       383964
1471.487553     38819
1400.095493     33014
881.757354      25509
1805.147812     21935
                ...  
8952.927355         1
6356.254968         1
7670.931918         1
7226.973312         1
8535.910021         1
Name: distance_Float, Length: 49254, dtype: int64
```

Erkenntnis: Es gibt 383964 Fahrten bei denen die Start und Endstation identisch ist.

In [29]:
```
df_station_data_test
```

Out[29]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Lati
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	
...	...	...	...	...	...	...	...	...	...	...	
10232877	2017-10-14 12:46:54	2017-10-14 13:14:53	32211	32211	W23202	Casual	2017-10-14	12	2017-10-14	13	
10232878	2017-06-18 12:34:17	2017-06-18 16:42:29	32211	32211	W23247	Casual	2017-06-18	12	2017-06-18	16	
10232879	2017-08-06 18:42:16	2017-08-06 19:12:20	32211	32211	W23247	Casual	2017-08-06	18	2017-08-06	19	
10232880	2016-11-05 19:12:56	2016-11-05 20:11:14	32211	32211	W23261	Casual	2016-11-05	19	2016-11-05	20	
10232881	2016-11-06 15:11:39	2016-11-06 15:39:41	32211	32211	W23261	Casual	2016-11-06	15	2016-11-06	15	

10232882 rows × 15 columns

In [30]:
```
df_station_data_test.groupby(['distance_Float','start_station_id', 'end_station_id'])
#.value_counts()
```

Out[30]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001BB47B95708>
```

In [31]:
```
df_station_data_test
```

Out[31]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Lati
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	
...	...	...	...	...	...	...	...	...	...	...	
10232877	2017-10-14 12:46:54	2017-10-14 13:14:53	32211	32211	W23202	Casual	2017-10-14	12	2017-10-14	13	
10232878	2017-06-18 12:34:17	2017-06-18 16:42:29	32211	32211	W23247	Casual	2017-06-18	12	2017-06-18	16	
10232879	2017-08-06 18:42:16	2017-08-06 19:12:20	32211	32211	W23247	Casual	2017-08-06	18	2017-08-06	19	
10232880	2016-11-05 19:12:56	2016-11-05 20:11:14	32211	32211	W23261	Casual	2016-11-05	19	2016-11-05	20	
10232881	2016-11-06 15:11:39	2016-11-06 15:39:41	32211	32211	W23261	Casual	2016-11-06	15	2016-11-06	15	

10232882 rows × 15 columns

In [32]:
```
df_station_data_test.start_station_id.value_counts()
```

Out[32]:
```
31623    208889
31258    195279
31247    172899
31200    155122
31201    126116
          ...
32072        23
31817        20
31819        11
31715        10
31815         8
Name: start_station_id, Length: 483, dtype: int64
```

In [33]:
```
df_station_data_test.end_station_id.value_counts()
```

Out[33]:
```
31623    218024
31258    195938
31247    178054
31200    173026
31201    135701
          ...
31817        29
32072        16
31715        13
31815        13
31819         4
Name: end_station_id, Length: 483, dtype: int64
```

In [34]:
```
df_station_data_test['combined']=df_station_data_test.apply(lambda x:'%s_%s' % (x['start_station_id'],x['end_station_id']),axis=1)
```

In [35]:
```
anzahlIdentischerRouten = df_station_data_test.combined.value_counts()
```

In [36]:
```
anzahlIdentischerRouten
```

Out[36]:
```
31247_31247    22481
31258_31249    21301
31247_31258    20824
31258_31247    17995
31258_31258    16277
               ...
32034_31202        1
31294_31506        1
31622_31106        1
31093_31206        1
31045_31215        1
Name: combined, Length: 86256, dtype: int64
```

Die beliebteste Route führt vom Lincoln Memorial (31258) zum Jefferson Memorial (31249).

In [37]:
```
# Temporäre Spalte wieder löschen
df_station_data_test.drop('combined', axis=1)
```

Out[37]:

	start_ts	end_ts	start_station_id	end_station_id	bike_number	Member type	start_date	start_hour	end_date	end_hour	Lati
0	2015-10-15 10:58:35	2015-10-15 14:57:10	31219	31634	? (0x0000000074BEBCE4)	Member	2015-10-15	10	2015-10-15	14	
1	2016-07-08 20:44:21	2016-07-08 21:11:54	31219	31634	W00099	Casual	2016-07-08	20	2016-07-08	21	
2	2017-10-14 19:47:46	2017-10-14 20:02:32	31219	31634	W00139	Member	2017-10-14	19	2017-10-14	20	
3	2017-05-14 15:50:53	2017-05-14 16:33:31	31219	31634	W00242	Casual	2017-05-14	15	2017-05-14	16	
4	2016-06-22 19:03:37	2016-06-22 19:20:51	31219	31634	W00277	Member	2016-06-22	19	2016-06-22	19	
...	...	...	...	...	...	...	...	...	...	...	
10232877	2017-10-14 12:46:54	2017-10-14 13:14:53	32211	32211	W23202	Casual	2017-10-14	12	2017-10-14	13	
10232878	2017-06-18 12:34:17	2017-06-18 16:42:29	32211	32211	W23247	Casual	2017-06-18	12	2017-06-18	16	
10232879	2017-08-06 18:42:16	2017-08-06 19:12:20	32211	32211	W23247	Casual	2017-08-06	18	2017-08-06	19	
10232880	2016-11-05 19:12:56	2016-11-05 20:11:14	32211	32211	W23261	Casual	2016-11-05	19	2016-11-05	20	
10232881	2016-11-06 15:11:39	2016-11-06 15:39:41	32211	32211	W23261	Casual	2016-11-06	15	2016-11-06	15	

10232882 rows × 15 columns

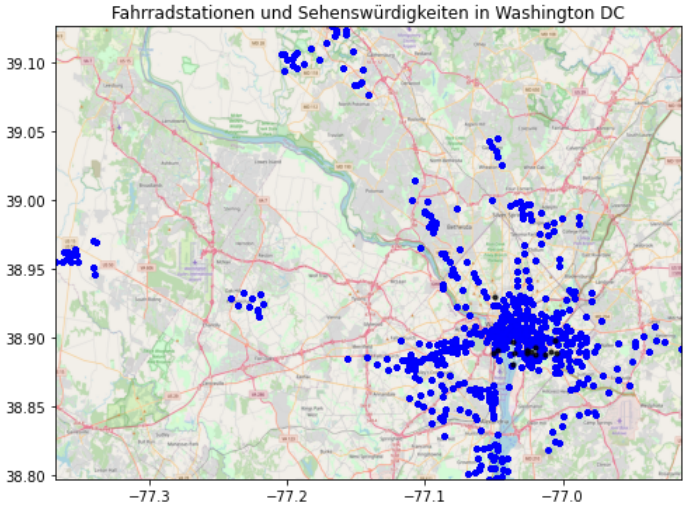
Darstellung der Sehenswürdigkeiten und Stationen in Washington DC

In [38]:
```
df_sights = pd.read_pickle(DATA_PATH + 'sightseeing_coordinates.pkl')
```

In [39]:
```
BBox = ((df_station_data_test.Longitude_start_station.min(),   df_station_data_test.Longitude_start_station.max(),
         df_station_data_test.Latitude_start_station.min(), df_station_data_test.Latitude_start_station.max()))

map_w= plt.imread('../images/map.png')

# Sehenswürdigkeiten und Fahrradstationen in einer Karte darstellen
fig, ax = plt.subplots(figsize = (8,7))
ax.scatter(df_station_data_test.Longitude_start_station, df_station_data_test.Latitude_start_station, zorder=1, alp
ha= 0.5,c='blue', s=10)
ax.scatter(df_sights.Longitude, df_sights.Latitude, zorder=1, alpha= 0.8, c='black', s=10)
ax.set_title('Fahrradstationen und Sehenswürdigkeiten in Washington DC')
ax.set_xlim(BBox[0],BBox[1])
ax.set_ylim(BBox[2],BBox[3])
ax.imshow(map_w, zorder=0, extent = BBox, aspect= 'equal');
```



In [40]:
```
# Das Ergebnis für die Weiterverwendung im Data Understanding Notebook speichern
df_station_data_test.to_pickle(DATA_PATH + 'airdistance_tripdata_coordinates.pkl')
```

In [ ]: