```
pip install --upgrade pip
```

```
Collecting pip
  Downloading https://files.pythonhosted.org/packages/ac/cf/0cc542fc93de2f3b9b53cb979c7d1118cffb93204afb46299a9f858e
     |████████████████████████████████| 1.6MB 13.6MB/s
Installing collected packages: pip
  Found existing installation: pip 19.3.1
    Uninstalling pip-19.3.1:
      Successfully uninstalled pip-19.3.1
Successfully installed pip-21.1
```

```
pip install hdbscan
```

```
    prefix = None
    Installing build dependencies ... done
    Getting requirements to build wheel ... done
      Preparing wheel metadata ... done
Requirement already satisfied: cython>=0.27 in /usr/local/lib/python3.7/dist-packages (from hdbscan) (0.29.22)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from hdbscan) (1.15.0)
Requirement already satisfied: scikit-learn>=0.20 in /usr/local/lib/python3.7/dist-packages (from hdbscan) (0.22.2
Requirement already satisfied: numpy>=1.16 in /usr/local/lib/python3.7/dist-packages (from hdbscan) (1.19.5)
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.7/dist-packages (from hdbscan) (1.0.1)
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.7/dist-packages (from hdbscan) (1.4.1)
Building wheels for collected packages: hdbscan
  Building wheel for hdbscan (PEP 517) ... done
  Created wheel for hdbscan: filename=hdbscan-0.8.27-cp37-cp37m-linux_x86_64.whl size=2311911 sha256=c693a7f1b7ae6
  Stored in directory: /root/.cache/pip/wheels/73/5f/2f/9a259b84003b84847c259779206acecabb25ab56f1506ee72b
Successfully built hdbscan
Installing collected packages: hdbscan
  WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617
  distutils: /usr/local/include/python3.7/hdbscan
  sysconfig: /usr/include/python3.7m/hdbscan
  WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617
```

```
  distutils: /usr/local/bin
  sysconfig: /usr/bin
  WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local
  sysconfig: /usr
  WARNING: Additional context:
  user = False
  home = None
  root = None
  prefix = None
WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/include/python3.7/UNKNOWN
sysconfig: /usr/include/python3.7m
WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/bin
sysconfig: /usr/bin
WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local
sysconfig: /usr
WARNING: Additional context:
user = False
home = None
root = None
prefix = None
Successfully installed hdbscan-0.8.27
WARNING: Running pip as root will break packages and permissions. You should install packages reliably by using ve
```

Creating a copy...

```
pip install folium
```

```
WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
distutils: /usr/local/lib/python3.7/dist-packages
sysconfig: /usr/lib/python3.7/site-packages
WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
```

```
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/include/python3.7/UNKNOWN
  sysconfig: /usr/include/python3.7m
  WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/bin
  sysconfig: /usr/bin
  WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local
  sysconfig: /usr
  WARNING: Additional context:
  user = False
  home = None
  root = None
  prefix = None
  Requirement already satisfied: folium in /usr/local/lib/python3.7/dist-packages (0.8.3)
  Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (from folium) (1.19.5)
  Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from folium) (2.11.3)
  Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from folium) (2.23.0)
  Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from folium) (1.15.0)
  Requirement already satisfied: branca>=0.3.0 in /usr/local/lib/python3.7/dist-packages (from folium) (0.4.2)
  Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from jinja2->folium) (1.1
  Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->folium) (2.10)
  Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->folium) (
  Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->folium)
  Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (fr
  WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/include/python3.7/UNKNOWN
  sysconfig: /usr/include/python3.7m
  WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/bin
  sysconfig: /usr/bin
  WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local
  sysconfig: /usr
```

```
WARNING: Additional context:
user = False
home = None
root = None
prefix = None
```



```
t will break packages and permissions. You should install packages reliably by using venv
```

```
import matplotlib
%matplotlib inline
%config InlineBackend.figure_format = 'svg'
import matplotlib.pyplot as plt
plt.style.use('ggplot')

import pandas as pd
import numpy as np

from tqdm import tqdm

from sklearn.cluster import KMeans, DBSCAN
from sklearn.metrics import silhouette_score,calinski_harabasz_score,davies_bouldin_score
from sklearn.datasets import make_blobs
from sklearn.neighbors import KNeighborsClassifier

from ipywidgets import interactive

from collections import defaultdict

import hdbscan
import folium
import re




from google.colab import drive
drive.mount("/content/drive")
```

```
Mounted at /content/drive
```

```
from tqdm import tqdm

from sklearn.cluster import KMeans, DBSCAN
from sklearn.metrics import silhouette_score
from sklearn.datasets import make_blobs
from sklearn.neighbors import KNeighborsClassifier

from ipywidgets import interactive

from collections import defaultdict

import folium
import re


cols=[ "user","check-in time", "latitude","longitude","location id"]
df = pd.read_csv(
    "/content/drive/My Drive/CNN/Brightkite_totalCheckins.txt", delimiter="\t", header=None,names=cols
)
df.head()
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4747287 entries, 0 to 4747286
Data columns (total 5 columns):
 #   Column         Dtype
---  ------         -----
```



```
 3   longitude      float64
 4   location id    object
dtypes: float64(2), int64(1), object(2)
memory usage: 181.1+ MB
```

```
df.describe()
```

```
df["user"].nunique()
```

```
51406
```

```
df.isna().values.any()
```

```
True
```

```
print(f'Before dropping NaNs \t:\tdf.shape = {df.shape}')
df.dropna(inplace=True)
print(f'After dropping NaNs \t:\tdf.shape = {df.shape}')
```



```
Before dropping NaNs 	:	df.shape = (4747287, 5)
After dropping NaNs 	:	df.shape = (4747281, 5)
```

```
X = np.array(df[["latitude","longitude"]], dtype='float64')
```

```
from sklearn.model_selection import train_test_split
X_train, X_test = train_test_split(X, test_size = 0.005, random_state = 0)
```

```
print(pd.DataFrame(X_test).info())
print(pd.DataFrame(X_test).describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23737 entries, 0 to 23736
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   0       23737 non-null  float64
 1   1       23737 non-null  float64
dtypes: float64(2)
memory usage: 371.0 KB
None
                  0             1
count  23737.000000  23737.000000
mean      34.318984    -42.857492
std       17.420033     84.029079
min      -90.000000   -166.599692
25%       33.504479    -99.164130
50%       37.783171    -79.796294
75%       41.947969      0.000000
max       70.662689    175.783333
```

```
m = folium.Map(location=[X_test[:,0].mean(), X_test[:,1].mean()], zoom_start=9
```

```
m = folium.Map(location=[X_test[:,0].mean(), X_test[:,1].mean()], zoom_start=9,
               tiles='Stamen Toner')

for i in range(0,23737):
    folium.CircleMarker(
        [X_test[i,0],X_test[i,1]],
        radius=1,
        color='#1787FE',
        fill=True,
        fill_colour='#1787FE'
    ).add_to(m)


m
```

Creating a copy...

```
wcss=[]
for i in range(1,15):
    kmeans=KMeans(n_clusters=i,init='k-means++',max_iter=300,n_init=10,random_state=0)
    kmeans.fit(X_test)
    wcss.append(kmeans.inertia_)
```

```
pip install kneed
```

```
  WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/include/python3.7/UNKNOWN
  sysconfig: /usr/include/python3.7m
  WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/bin
```

```
  sysconfig: /usr/bin
  WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local
  sysconfig: /usr
  WARNING: Additional context:
  user = False
  home = None
  root = None
  prefix = None
  Collecting kneed
    Downloading kneed-0.7.0-py2.py3-none-any.whl (9.4 kB)
  Requirement already satisfied: numpy>=1.14.2 in /usr/local/lib/python3.7/dist-packages (from kneed) (1.19.5)
  Requirement already satisfied: scipy in /usr/local/lib/python3.7/dist-packages (from kneed) (1.4.1)
  Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from kneed) (3.2.2)
  Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->kneed) (0.
  Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->kn
  Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages
  Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->kneed
  Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from cycler>=0.10->matplotlib->kneed
  Installing collected packages: kneed
    WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617
    distutils: /usr/local/lib/python3.7/dist-packages
    sysconfig: /usr/lib/python3.7/site-packages
    WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617
    distutils: /usr/local/lib/python3.7/dist-packages
    sysconfig: /usr/lib/python3.7/site-packages
    WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617
    distutils: /usr/local/include/python3.7/kneed
    sysconfig: /usr/include/python3.7m/kneed
    WARNING: Value for scheme.scripts does not match. Please report this to <https://github.com/pypa/pip/issues/9617
    distutils: /usr/local/bin
    sysconfig: /usr/bin
    WARNING: Value for scheme.data does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
    distutils: /usr/local
    sysconfig: /usr
    WARNING: Additional context:
    user = False
    home = None
    root = None
    prefix = None
  WARNING: Value for scheme.platlib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
```

```
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.purelib does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
  distutils: /usr/local/lib/python3.7/dist-packages
  sysconfig: /usr/lib/python3.7/site-packages
  WARNING: Value for scheme.headers does not match. Please report this to <https://github.com/pypa/pip/issues/9617>
```

Creating a copy...

```
plt.title('The Elbow Method')
plt.xlabel('Number od clusters')
plt.ylabel('WCSS')
plt.show()
```

```
from kneed import KneeLocator
kn = KneeLocator(range(1,15),wcss, curve='convex', direction='decreasing')
print(kn.knee)
```

```
3
```

```
k=3
andom_state=17).fit(X_test)
ict(X_test)
```



```
[0 0 0 ... 0 0 2]
```

```
print(list(class_predictions).count(0))
print(list(class_predictions).count(1))
print(list(class_predictions).count(2))
```

```
15379
3338
5020
```

```
kmeans=KMeans(n_clusters=3,init='k-means++',max_iter=300,n_init=10,random_state=0)
y_kmeans=kmeans.fit_predict(X_test)
```

```
plt.scatter(X_test[y_kmeans==0,0],X_test[y_kmeans==0,1],s=100,c='red',label='cluster 1')
plt.scatter(X_test[y_kmeans==1,0],X_test[y_kmeans==1,1],s=100,c='blue',label='cluster 2')
plt.scatter(X_test[y_kmeans==2,0],X_test[y_kmeans==2,1],s=100,c='green',label='cluster 3')
plt.scatter(kmeans.cluster_centers_[:,0],kmeans.cluster_centers_[:,1],s=300,c='yellow',label='centroids')
plt.title('Clusters of users')
plt.xlabel('Latitude')
plt.ylabel('Longitude')
plt.show()
```

Creating a copy...

```
best_silhouette, best_k = -1, 0

for k in tqdm(range(2, 10)):
    model = KMeans(n_clusters=k, random_state=1).fit(X_test)
    class_predictions = model.predict(X_test)

    curr_silhouette = silhouette_score(X_test, class_predictions)
    if curr_silhouette > best_silhouette:
        best_k = k
        best_silhouette = curr_silhouette

print(f'K={best_k}')
print(f'Silhouette Score: {best_silhouette}')
```

```
100%|██████████| 8/8 [01:05<00:00,  8.23s/it]K=3
Silhouette Score: 0.751058304924117
```

```
print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
```

```
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')
```

```
Calinski ignoring outliers: 215308.6325502918
Calinski outliers as singletons: 215308.6325502918
```



```
g outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_
```

```
no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Davies Bouldin ignoring outliers: 0.4798002561377974
Davies Bouldin as singletons: 0.4798002561377974
```

```
kmeans_dat=pd.DataFrame(X_test)
kmeans_dat['CLUSTERS_KMEANS'] = y_kmeans
kmeans_dat.head()
```

DBSCAN

```
model = DBSCAN(eps=0.01, min_samples=5).fit(X_test)
class_predictions = model.labels_
```

```
print(f'Number of clusters found: {len(np.unique(class_predictions))}')
print(f'Number of outliers found: {len(class_predictions[class_predictions==-1])}')

print(f'Silhouette ignoring outliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti
```



```
er+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Silhouette outliers as singletons: {silhouette_score(X_test, no_outliers)}')
```

```
Number of clusters found: 785
Number of outliers found: 11834
Silhouette ignoring outliers: 0.7661951430103257
Silhouette outliers as singletons: 0.2988007062481111
```

```
print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')
```

```
Calinski ignoring outliers: 577981985.2748314
Calinski outliers as singletons: 68803410.67455058
```

```
print(f'Davies Bouldin ignoring outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Davies Bouldin ignoring outliers: 0.14653707772832836
Davies Bouldin as singletons: 0.12403104775502434
```

```
dbscan_dat=pd.DataFrame(X_test)


dbscan_dat['CLUSTERS_DBSCAN'] = class_predictions
dbscan_dat.head()
```

```
plt.figure(figsize=(10, 8))
plt.scatter(X_test[:,0], X_test[:,1], c=model.labels_.astype(float))
```

HDBSCAN

```
model = hdbscan.HDBSCAN(min_cluster_size=5, min_samples=2,
                        cluster_selection_epsilon=0.01)

class_predictions = model.fit_predict(X_test)


print(f'Number of clusters found: {len(np.unique(class_predictions))-1}')
print(f'Number of outliers found: {len(class_predictions[class_predictions==-1])}')

print(f'Silhouette ignoring outliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Silhouette outliers as singletons: {silhouette_score(X_test, no_outliers)}')
```

```
Number of clusters found: 1485
Number of outliers found: 3801
Silhouette ignoring outliers: 0.6548991258646143
Silhouette outliers as singletons: 0.4596338355071577
```

```
print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
```

```
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')
```

```
Calinski ignoring outliers: 456389.3477635357
Calinski outliers as singletons: 151607.63399134102
```



```
g outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_
```

```
no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin outliers as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Davies Bouldin ignoring outliers: 0.4204648261224274
Davies Bouldin outliers as singletons: 0.3548516344098311
```

```
hdbscan_dat=pd.DataFrame(X_test)
hdbscan_dat['CLUSTERS_HDBSCAN'] = class_predictions
hdbscan_dat.head()
```

```
plt.figure(figsize=(10, 8))
plt.scatter(X_test[:,0], X_test[:,1], c=model.labels_.astype(float))
```

BIRCH

```
from sklearn.cluster import Birch

brc = Birch(n_clusters=None)
```

```
class_predictions=brc.fit_predict(X_test)
```



```
print(f'Number of clusters found: {len(np.unique(class_predictions))-1}')
nd: {len(class_predictions[class_predictions==-1])}')

print(f'Silhouette ignoring outliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Silhouette outliers as singletons: {silhouette_score(X_test, no_outliers)}')
```

```
Number of clusters found: 875
Number of outliers found: 0
Silhouette ignoring outliers: 0.6563531735169894
Silhouette outliers as singletons: 0.6563531735169894
```

```
print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')
```

```
Calinski ignoring outliers: 2869745.220027992
Calinski outliers as singletons: 2869745.220027992
```

```
print(f'Davies Bouldin ignoring outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Davies Bouldin ignoring outliers: 0.38315825945478854
Davies Bouldin as singletons: 0.38315825945478854
```

```
plt.figure(figsize=(10, 8))
plt.scatter(X_test[:,0], X_test[:,1], c=brc.labels_.astype(float))
```

Creating a copy...

OPTICS

```
from sklearn.cluster import OPTICS
```



```
class_predictions= model.fit_predict(X_test)
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/cluster/_optics.py:802: RuntimeWarning: divide by zero encountered ir
  ratio = reachability_plot[:-1] / reachability_plot[1:]
```

```
print(f'Number of clusters found: {len(np.unique(class_predictions))-1}')
print(f'Number of outliers found: {len(class_predictions[class_predictions==-1])}')

print(f'Silhouette ignoring outliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Silhouette outliers as singletons: {silhouette_score(X_test, no_outliers)}')
```

```
Number of clusters found: 5463
Number of outliers found: 3076
Silhouette ignoring outliers: 0.791322364163566
Silhouette outliers as singletons: 0.6747560455330907
```

```
print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')
```

```
Calinski ignoring outliers: 971943.1093159069
Calinski outliers as singletons: 726856.1953551086
```

```
print(f'Davies Bouldin ignoring outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
```

```
print(f'Davies Bouldin outliers as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Davies Bouldin ignoring outliers: 0.33043150147184736
Davies Bouldin outliers as singletons: 0.29516973191501505
```



```
plt.scatter(X_test[:,0], X_test[:,1], c=model.labels_.astype(float))
```

Creating a copy...

Gowalla

```
cols=[ "user","check-in time", "latitude","longitude","location id"]
dataset = pd.read_csv(
    "/content/drive/My Drive/CNN/Gowalla_totalCheckins.txt", delimiter="\t", header=None,names=cols
)
dataset.head()
```

```
X = np.array(dataset[["latitude","longitude"]], dtype='float64')
```

```
from sklearn.model_selection import train_test_split
X_train, X_test = train_test_split(X, test_size = 0.005, random_state = 0)
```

```
pd.DataFrame(X_test).info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32215 entries, 0 to 32214
```

```
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   0       32215 non-null  float64
 1   1       32215 non-null  float64
```

Creating a copy...

```
pd.DataFrame(X_test).describe()
```

OPTICS

```
model = OPTICS(min_samples=2)
```

```
class_predictions= model.fit_predict(X_test)
```

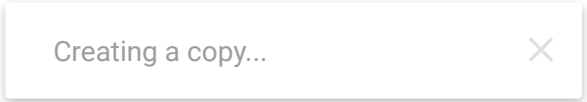
```
/usr/local/lib/python3.7/dist-packages/sklearn/cluster/_optics.py:802: RuntimeWarning: divide by zero encountered ir
  ratio = reachability_plot[:-1] / reachability_plot[1:]
```

```
print(f'Number of clusters found: {len(np.unique(class_predictions))-1}')
print(f'Number of outliers found: {len(class_predictions[class_predictions==-1])}')

tliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti

er+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Silhouette outliers as singletons: {silhouette_score(X_test, no_outliers)}')

print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')

print(f'Davies Bouldin ignoring outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```



```
Number of clusters found: 9013
Number of outliers found: 5790
Silhouette ignoring outliers: 0.6622318262693725
Silhouette outliers as singletons: 0.5177249882562976
Calinski ignoring outliers: 915006.6944089212
Calinski outliers as singletons: 688000.7410681586
Davies Bouldin ignoring outliers: 0.3736579135589328
Davies Bouldin as singletons: 0.3229703827501595
```

```
plt.figure(figsize=(10, 8))
plt.scatter(X_test[:,0], X_test[:,1], c=model.labels_.astype(float))
```



DBSCAN

```
model = DBSCAN(eps=0.01, min_samples=5).fit(X_test)
class_predictions = model.labels_
```

```
print(f'Number of clusters found: {len(np.unique(class_predictions))-1}')
print(f'Number of outliers found: {len(class_predictions[class_predictions==-1])}')

print(f'Silhouette ignoring outliers: {silhouette_score(X_test[class_predictions!=-1], class_predictions[class_predicti
```



```
er+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
 singletons: {silhouette_score(X_test, no_outliers)}')

print(f'Calinski ignoring outliers: {calinski_harabasz_score(X_test[class_predictions!=-1], class_predictions[class_pre

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Calinski outliers as singletons: {calinski_harabasz_score(X_test, no_outliers)}')

print(f'Davies Bouldin ignoring outliers: {davies_bouldin_score(X_test[class_predictions!=-1], class_predictions[class_

no_outliers = np.array([(counter+2)*x if x==-1 else x for counter, x in enumerate(class_predictions)])
print(f'Davies Bouldin as singletons: {davies_bouldin_score(X_test, no_outliers)}')
```

```
Number of clusters found: 770
Number of outliers found: 16560
Silhouette ignoring outliers: 0.5952262243383366
Silhouette outliers as singletons: 0.13146048379479877
Calinski ignoring outliers: 229765407.89558798
Calinski outliers as singletons: 22426541.654199768
Davies Bouldin ignoring outliers: 0.24912855423904723
Davies Bouldin as singletons: 0.11772929911226311
```

```
plt.figure(figsize=(10, 8))
plt.scatter(X_test[:,0], X_test[:,1], c=model.labels_.astype(float))
```

VORONOI DIAGRAM

```
from scipy.spatial import Voronoi, voronoi_plot_2d
```

```
from scipy.spatial import Voronoi, voronoi_plot_2d
vor = Voronoi(X_test)
```

```
fig = voronoi_plot_2d(vor)
fig.set_figheight(10)
fig.set_figwidth(8)
plt.show()
```

```
fig = voronoi_plot_2d(vor, show_vertices=False, line_colors='orange',
                      line_width=2, line_alpha=0.6, point_size=2)
plt.show()
```

```
vor.regions
```

```
   [[7, 4, 3, 6],
    [39, 12, 5, 7, 6, 38],
    [72, 70, 69, 71],
    [70, 39, 38, 69],
    [71, 14, -1, 3, 6, 38, 69],
    [72, 12, 39, 70],
    [83, 41, 40, 82],
    [123, 53, 117, 118, 116, 52, 51, 121],
    [135, 103, 33, 11, 29, 58, 131, 134],
    [138, 61, 133, 132, 137],
    [142, 140, 141],
    [142, 59, 62, 61, 133, 136, 104, 140],
    [148, 145, 144, 147],
    [151, 146, 60, 150],
    [165, 80, 79, 164],
    [172, 13, -1, 14, 171],
    [179, 87, 17, 19, 20, 43, 42, 91, 177, 176, 178],
    [188, 96, 98, 97, 22, 15, 16, 40, 41, 19, 20, 187],
    [198, 26, 24, 25, 197],
    [202, 114, 51, 52, 200],
    [201, 113, 49, -1, 115, 116, 52, 200],
```

```
        [206, 202, 114, 204],
        [206, 202, 200, 201, 205],
        [],
        [213, 207, 209, 208, 54, 119, 212],
        [234, 232, 231, 233],
                                  , 231],
                                  46, 232],
                                  33],
        [240, 238, 237, 239],
        [239, 235, 62, 61, 138, 139, 237],
        [240, 236, 153, 238],
        [240, 236, 235, 239],
        [242, 154, 63, 9, 67, 68, 36, 34, 35, 241],
        [249, 65, 37, 36, 34, 33, 11, 248],
        [250, 66, 65, 249],
        [263, 261, 262],
        [267, 263, 262, 162, 264],
        [272, 270, 269, 271],
        [270, 163, 164, 165, 269],
        [271, 75, 78, 167, 166, 80, 165, 269],
        [272, 77, 76, 163, 270],
        [282, 5, 7, 4, 81, 281],
        [283, 171, 14, 71, 72, 12, 5, 282],
        [287, 16, 15, 286],
        [292, 175, 84, 86, 85, 291],
        [294, 102, 103, 33, 34, 35, 293],
        [298, 296, 297],
        [298, 181, 88, 89, 296],
        [298, 181, 180, 297],
        [319, 194, 317],
        [319, 194, 108, 107, 106, 105, 318],
        [321, 318, 105, 320],
        [335, 27, 50, 203, 204, 114, 51, 121, 120, 334],
        [348, 50, 27, 48, 111, 110, 8, -1, 49, 347],
        [351, 203, 204, 206, 205, 349],
        [351, 203, 50, 348, 346, 350],
        [355, 353, 352, 354],
        [354, 118, 116, 115, 352],
```

Creating a copy...

```
vor.vertices
```

```
array([[  592.94570217,   225.91359725],
       [-1256.29761491,     2.80940194],
       [  148.91512435,   120.00058698],
       ...,
       [   30.26663329,   -97.73944915],
                          97.7396132 ],
                          97.73990186]])
```



```
vor.ridge_points
```

```
array([[20779, 15140],
       [20779, 11903],
       [20779, 19332],
       ...,
       [ 9842, 16020],
       [ 6537,  7770],
       [ 7770, 16020]], dtype=int32)
```

```
vor.ridge_vertices
```

```
[23, 197],
[26, 198],
[197, 198],
[51, 114],
[52, 200],
[114, 202],
[200, 202],
[-1, 115],
[-1, 49],
[49, 113],
[113, 201],
[115, 116],
[200, 201],
[114, 204],
[202, 206],
[204, 206],
[201, 205],
[205, 206],
[54, 208],
[54, 119],
[119, 212],
```

```
       [207, 213],
       [207, 209],
       [208, 209],
       [212, 213],
       [231, 233],
```



```
       [30, 225],
       [30, 60],
       [146, 232],
       [224, 225],
       [224, 231],
       [147, 234],
       [148, 152],
       [151, 152],
       [144, 228],
       [227, 228],
       [227, 233],
       [237, 239],
       [237, 238],
       [238, 240],
       [239, 240],
       [62, 235],
       [138, 139],
       [139, 237],
       [235, 239],
       [153, 238],
       [153, 236],
       [236, 240],
       [235, 236],
       [9, 63],
       [9, 67],
       [34, 35],
       [34, 36],
       [35, 241],
       [36, 68],
       [63, 154],
       [67, 68],
```

```
vor.furthest_site
```

```
False
```