

Cleansing : Validate Datatype Value

```
setdata = spark.read.format('csv').option("header",
"true").option("inferSchema","true").load('FileStore/tables/crime/')
print("total baris sumber data: "+ str(setdata.count()))

cleaningset = setdata.withColumn("Latitude", setdata["Latitude"].cast("double")).na.drop(subset=
["Latitude"])
cleaningset2a = cleaningset.withColumn("Longitude",
cleaningset["Longitude"].cast("double")).na.drop(subset=["Longitude"])
cleaningset2b = cleaningset2a.withColumn("Month",
cleaningset2a["Month"].cast("integer")).na.drop(subset=["Month"])

cleaningset2 = cleaningset2a.na.drop(subset=[' Crime ID','Month','Year', 'Reported by', 'Falls
within', 'Longitude', 'Latitude', 'Location', 'LSOA code', 'LSOA name', 'Crime type', 'Last outcome
category'])

valid = setdata.where(setdata["Latitude"].cast("double").isNotNull())
invalid = setdata.where(setdata["Latitude"].cast("double").isNull())

print("total baris aftercleansing: "+ str(cleaningset2.count()))

total baris sumber data: 18102559
total baris aftercleansing: 12750402

# melakukan cleansing, yaitu dengan pemilihan data yang memiliki value berdasarkan value crime type
yang benar sesuai deskripsi pada data kepolisian UK
cleaningset2 = cleaningset2.where(
    (cleaningset2["Crime type"] == "Criminal damage and arson") |
    (cleaningset2["Crime type"] == "Theft from the person") |
    (cleaningset2["Crime type"] == "Public order") |
    (cleaningset2["Crime type"] == "Vehicle crime") |
    (cleaningset2["Crime type"] == "Violence and sexual offences") |
    (cleaningset2["Crime type"] == "Shoplifting") |
    (cleaningset2["Crime type"] == "Burglary") |
    (cleaningset2["Crime type"] == "Other theft") |
    (cleaningset2["Crime type"] == "Bicycle theft") |
    (cleaningset2["Crime type"] == "Robbery") |
    (cleaningset2["Crime type"] == "Other crime") |
    (cleaningset2["Crime type"] == "Drugs") |
    (cleaningset2["Crime type"] == "Possession of weapons") |
    (cleaningset2["Crime type"] == "social behaviour")
).drop()

display(cleaningset2.orderBy("Crime type").groupBy("Crime type").count())
```

Crime type
Bicycle theft
Burglary
Criminal damage and arson
Drugs
Other crime

Other theft
Possession of weapons
Public order
Robbery
Shoplifting
Theft from the person
Vehicle crime
Violence and sexual offences
social behaviour


```
# total baris setelah cleansing final
print("total baris after conditioning: "+ str(cleaningset2.count()))
```

total baris after conditioning: 11949434

Start Clustering with KMeans

```
from pyspark.sql import SparkSession
from pyspark.ml.clustering import KMeans
from pyspark.ml.feature import VectorAssembler

# inisiasi appname
spark = SparkSession.builder.appName('cluster').getOrCreate()

# mengambil data csv yang merupakan dataset berupa lat, lng, membuat jadi bertipe data double
dataclean = cleaningset2.select("Longitude","Latitude")

# set up menjadi 9 cluster dengan kmeans
kmeans = KMeans().setK(9).setSeed(1)

# inisiasi menambahkan atribut features yang dimana atribut features yaitu memiliki value array of lat,lng
assembler = VectorAssembler(inputCols=dataclean.columns, outputCol = 'features')

# jadi punya kolom features
final_data_clean = assembler.transform(dataclean)

# inisiasi kmeans dengan mengeksekusi pada kolom bernama features
kmeans = KMeans(featuresCol='features',k=9)

model = kmeans.fit(final_data_clean)

display(model.transform(final_data_clean))
```

Longitude	Latitude	features
-2.511571	51.414895	▶ [1,2,[],[-2.511571,51.414895]]
-2.515816	51.408717	▶ [1,2,[],[-2.515816,51.408717]]
-2.515816	51.408717	▶ [1,2,[],[-2.515816,51.408717]]
-2.494715	51.419948	▶ [1,2,[],[-2.494715,51.419948]]
-2.49793	51.417966	▶ [1,2,[],[-2.49793,51.417966]]

-2.497767	51.420232	▶[1,2,[],[-2.497767,51.420232]]
-2.502805	51.414033	▶[1,2,[],[-2.502805,51.414033]]
-2.500214	51.416481	▶[1,2,[],[-2.500214,51.416481]]
-2.501425	51.416692	▶[1,2,[],[-2.501425,51.416692]]
-2.501425	51.416692	▶[1,2,[],[-2.501425,51.416692]]
-2.497799	51.415233	▶[1,2,[],[-2.497799,51.415233]]
-2.49854	51.414618	▶[1,2,[],[-2.49854,51.414618]]
-2.497799	51.415233	▶[1,2,[],[-2.497799,51.415233]]
-2.49854	51.414618	▶[1,2,[],[-2.49854,51.414618]]
-2.498613	51.416002	▶[1,2,[],[-2.498613,51.416002]]
-2.498613	51.416002	▶[1,2,[],[-2.498613,51.416002]]
-2.498613	51.416002	▶[1,2,[],[-2.498613,51.416002]]
-2.501425	51.416692	▶[1,2,[],[-2.501425,51.416692]]
-2.501425	51.416692	▶[1,2,[],[-2.501425,51.416692]]
-2.501319	51.417573	▶[1,2,[],[-2.501319,51.417573]]
-2.494715	51.419948	▶[1,2,[],[-2.494715,51.419948]]
-2.504309	51.418801	▶[1,2,[],[-2.504309,51.418801]]
-2.494715	51.419948	▶[1,2,[],[-2.494715,51.419948]]
-2.499922	51.417373	▶[1,2,[],[-2.499922,51.417373]]
-2.498613	51.416002	▶[1,2,[],[-2.498613,51.416002]]
-2.501806	51.417391	▶[1,2,[],[-2.501806,51.417391]]
-2.503079	51.416757	▶[1,2,[],[-2.503079,51.416757]]
-2.503079	51.416757	▶[1,2,[],[-2.503079,51.416757]]
-2.505264	51.411433	▶[1,2,[],[-2.505264,51.411433]]
-2.501363	51.412394	▶[1,2,[],[-2.501363,51.412394]]
-2.498086	51.413838	▶[1,2,[],[-2.498086,51.413838]]
-2.498086	51.413838	▶[1,2,[],[-2.498086,51.413838]]
-2.500575	51.411345	▶[1,2,[],[-2.500575,51.411345]]
-2.505264	51.411433	▶[1,2,[],[-2.505264,51.411433]]
-2.500575	51.411345	▶[1,2,[],[-2.500575,51.411345]]
-2.500304	51.408928	▶[1,2,[],[-2.500304,51.408928]]
-2.500575	51.411345	▶[1,2,[],[-2.500575,51.411345]]
-2.504356	51.40865	▶[1,2,[],[-2.504356,51.40865]]
-2.507	51.408557	▶[1,2,[],[-2.507,51.408557]]
-2.504356	51.40865	▶[1,2,[],[-2.504356,51.40865]]
-2.502845	51.409879	▶[1,2,[],[-2.502845,51.409879]]
-2.507	51.408557	▶[1,2,[],[-2.507,51.408557]]
-2.491971	51.412101	▶[1,2,[],[-2.491971,51.412101]]
-2.490696	51.415272	▶[1,2,[],[-2.490696,51.415272]]
-2.483216	51.416346	▶[1,2,[],[-2.483216,51.416346]]
-2.490696	51.415272	▶[1,2,[],[-2.490696,51.415272]]
-2.489013	51.408499	▶[1,2,[],[-2.489013,51.408499]]

-2.491308	51.406619	▶ [1,2,[],[-2.491308,51.406619]]
-2.48419	51.406496	▶ [1,2,[],[-2.48419,51.406496]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.474568	51.410258	▶ [1,2,[],[-2.474568,51.410258]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.48623	51.413087	▶ [1,2,[],[-2.48623,51.413087]]
-2.474568	51.410258	▶ [1,2,[],[-2.474568,51.410258]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.48557	51.406445	▶ [1,2,[],[-2.48557,51.406445]]
-2.483019	51.412705	▶ [1,2,[],[-2.483019,51.412705]]
-2.483019	51.412705	▶ [1,2,[],[-2.483019,51.412705]]
-2.483019	51.412705	▶ [1,2,[],[-2.483019,51.412705]]
-2.475536	51.412124	▶ [1,2,[],[-2.475536,51.412124]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.478811	51.413145	▶ [1,2,[],[-2.478811,51.413145]]
-2.350626	51.400417	▶ [1,2,[],[-2.350626,51.400417]]
-2.346	51.400799	▶ [1,2,[],[-2.346,51.400799]]
-2.347868	51.398735	▶ [1,2,[],[-2.347868,51.398735]]
-2.346982	51.399457	▶ [1,2,[],[-2.346982,51.399457]]
-2.350626	51.400417	▶ [1,2,[],[-2.350626,51.400417]]
-2.350626	51.400417	▶ [1,2,[],[-2.350626,51.400417]]
-2.359173	51.397837	▶ [1,2,[],[-2.359173,51.397837]]
-2.3561	51.400041	▶ [1,2,[],[-2.3561,51.400041]]
-2.351833	51.398435	▶ [1,2,[],[-2.351833,51.398435]]
-2.341424	51.400219	▶ [1,2,[],[-2.341424,51.400219]]
-2.344359	51.396704	▶ [1,2,[],[-2.344359,51.396704]]
-2.33901	51.402222	▶ [1,2,[],[-2.33901,51.402222]]
-2.344359	51.396704	▶ [1,2,[],[-2.344359,51.396704]]
-2.33901	51.402222	▶ [1,2,[],[-2.33901,51.402222]]
-2.336398	51.400818	▶ [1,2,[],[-2.336398,51.400818]]
-2.400821	51.405138	▶ [1,2,[],[-2.400821,51.405138]]
-2.389637	51.400186	▶ [1,2,[],[-2.389637,51.400186]]
-2.378413	51.395277	▶ [1,2,[],[-2.378413,51.395277]]
-2.390086	51.395508	▶ [1,2,[],[-2.390086,51.395508]]
-2.392367	51.39658	▶ [1,2,[],[-2.392367,51.39658]]

-2.389883	51.395284	▶ [1,2,[],[-2.389883,51.395284]]
-2.392587	51.3972	▶ [1,2,[],[-2.392587,51.3972]]
-2.389599	51.395672	▶ [1,2,[],[-2.389599,51.395672]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.352617	51.392067	▶ [1,2,[],[-2.352617,51.392067]]
-2.355418	51.39186	▶ [1,2,[],[-2.355418,51.39186]]

Showing the first 1000 rows.



Results of Cluster (2)

```
centers = model.clusterCenters()
size_summary = model.summary.cluster
total_data = model.summary.clusterSizes
total_cost = model.computeCost(final_data_clean)

print("center "+ str(centers) +"\n")
print("total_data "+ str(total_data) + "\n")
print("total_cost "+ str(total_cost) + "\n")
print("size_summary " + str(size_summary) + "\n")
print("total baris setelah modeling with kmeans: "+ str(final_data_clean.count()))

center [array([-0.15931191,  51.51734029]), array([-2.87259142,  53.52871918]), array([-0.72140
065,  53.25522812]), array([-3.34004466,  51.28603116]), array([-1.86993476,  53.56966014]), arr
ay([-1.47562803,  54.82014664]), array([-1.75686087,  52.47886174]), array([ 0.90768687,  51.81
232548]), array([-1.37880455,  51.12808072])]

total_data [3530076, 1097394, 890795, 1072866, 1689131, 636583, 1430228, 815513, 786848]

total_cost 3151794.96011

size_summary DataFrame[prediction: int]

total baris setelah modeling with kmeans: 11949434
```

Menambahkan Kolom ID Unik

```
# inisiasi variable dataset
dataset1 = cleaningset2
dataset2 = model.transform(final_data_clean)

# start code for dataset2
from pyspark.sql import Row
from pyspark.sql.types import StructField, StructType, LongType

row = Row("foo", "bar")
row_with_index = Row(*["id"] + dataset2.columns)

def make_row(columns):
    def _make_row(row, uid):
        row_dict = row.asDict()
        return row_with_index*[uid] + [row_dict.get(c) for c in columns]
    return _make_row

f = make_row(dataset2.columns)

df_with_pk2 = (dataset2.rdd
    .zipWithUniqueId()
    .map(lambda x: f(*x))
    .toDF(StructType([StructField("id", LongType(), False)] + dataset2.schema.fields)))

#####

# start code for dataset1
from pyspark.sql import Row
from pyspark.sql.types import StructField, StructType, LongType

row = Row("foo", "bar")
row_with_index = Row(*["id"] + dataset1.columns)

def make_row(columns):
    def _make_row(row, uid):
        row_dict = row.asDict()
        return row_with_index*[uid] + [row_dict.get(c) for c in columns]
    return _make_row

f2 = make_row(dataset1.columns)

df_with_pk1 = (dataset1.rdd
    .zipWithUniqueId()
    .map(lambda x: f2(*x))
    .toDF(StructType([StructField("id", LongType(), False)] + dataset1.schema.fields)))

# menampilkan dataset2 yang dimana merupakan data setelah kmeans
display(df_with_pk2.orderBy("prediction"))
```

id	Longitude	Latitude	features
868593	-0.592272	51.263693	▶ [1,2,[],[-0.592272,51.263693]]
898623	-0.151204	51.2637	▶ [1,2,[],[-0.151204,51.2637]]
868623	-0.591612	51.265133	▶ [1,2,[],[-0.591612,51.265133]]
838653	-0.150603	51.543452	▶ [1,2,[],[-0.150603,51.543452]]

868653	-0.592272	51.263693	► [1,2,[],[-0.592272,51.263693]]
838743	-0.15221	51.15813	► [1,2,[],[-0.15221,51.15813]]
868683	-0.588262	51.263609	► [1,2,[],[-0.588262,51.263609]]
838803	-0.196508	51.112067	► [1,2,[],[-0.196508,51.112067]]
868713	-0.593017	51.264664	► [1,2,[],[-0.593017,51.264664]]
838863	-0.066134	51.322722	► [1,2,[],[-0.066134,51.322722]]
868743	-0.585401	51.260616	► [1,2,[],[-0.585401,51.260616]]
838953	-0.372065	51.408127	► [1,2,[],[-0.372065,51.408127]]
868773	-0.591636	51.255916	► [1,2,[],[-0.591636,51.255916]]
839013	-0.368866	51.407929	► [1,2,[],[-0.368866,51.407929]]
868803	-0.587329	51.262716	► [1,2,[],[-0.587329,51.262716]]
839073	-0.377878	51.406813	► [1,2,[],[-0.377878,51.406813]]
868833	-0.590143	51.261617	► [1,2,[],[-0.590143,51.261617]]
839133	-0.361722	51.405069	► [1,2,[],[-0.361722,51.405069]]
868863	-0.591636	51.255916	► [1,2,[],[-0.591636,51.255916]]
839193	-0.368512	51.403725	► [1,2,[],[-0.368512,51.403725]]
868893	-0.590396	51.261324	► [1,2,[],[-0.590396,51.261324]]
839253	-0.36442	51.402786	► [1,2,[],[-0.36442,51.402786]]
868923	-0.591636	51.255916	► [1,2,[],[-0.591636,51.255916]]
839313	-0.362267	51.402684	► [1,2,[],[-0.362267,51.402684]]
868953	-0.586062	51.263358	► [1,2,[],[-0.586062,51.263358]]
839373	-0.36001	51.401871	► [1,2,[],[-0.36001,51.401871]]
868983	-0.588624	51.260205	► [1,2,[],[-0.588624,51.260205]]
839433	-0.362482	51.406725	► [1,2,[],[-0.362482,51.406725]]
869013	-0.581788	51.259736	► [1,2,[],[-0.581788,51.259736]]
839493	-0.36141	51.409759	► [1,2,[],[-0.36141,51.409759]]
869043	-0.581483	51.257997	► [1,2,[],[-0.581483,51.257997]]
839553	-0.362482	51.406725	► [1,2,[],[-0.362482,51.406725]]
869073	-0.573897	51.260926	► [1,2,[],[-0.573897,51.260926]]
839613	-0.364123	51.405057	► [1,2,[],[-0.364123,51.405057]]
869103	-0.573897	51.260926	► [1,2,[],[-0.573897,51.260926]]
839673	-0.344444	51.403747	► [1,2,[],[-0.344444,51.403747]]
869133	-0.573897	51.260926	► [1,2,[],[-0.573897,51.260926]]
839733	-0.34811	51.400561	► [1,2,[],[-0.34811,51.400561]]
869163	-0.577995	51.262343	► [1,2,[],[-0.577995,51.262343]]
839793	-0.34341	51.402932	► [1,2,[],[-0.34341,51.402932]]
869193	-0.573897	51.260926	► [1,2,[],[-0.573897,51.260926]]
839853	-0.34341	51.402932	► [1,2,[],[-0.34341,51.402932]]
869223	-0.575173	51.260438	► [1,2,[],[-0.575173,51.260438]]
839913	-0.345413	51.403203	► [1,2,[],[-0.345413,51.403203]]
869253	-0.568744	51.257932	► [1,2,[],[-0.568744,51.257932]]
839973	-0.358127	51.399848	► [1,2,[],[-0.358127,51.399848]]

869283	-0.578762	51.255617	► [1,2,[],[-0.578762,51.255617]]
840033	-0.355742	51.40021	► [1,2,[],[-0.355742,51.40021]]
869313	-0.58341	51.257777	► [1,2,[],[-0.58341,51.257777]]
840093	-0.356101	51.403021	► [1,2,[],[-0.356101,51.403021]]
869343	-0.58341	51.257777	► [1,2,[],[-0.58341,51.257777]]
840153	-0.37714	51.397181	► [1,2,[],[-0.37714,51.397181]]
869373	-0.578278	51.252239	► [1,2,[],[-0.578278,51.252239]]
840213	-0.38176	51.395402	► [1,2,[],[-0.38176,51.395402]]
869403	-0.578762	51.255617	► [1,2,[],[-0.578762,51.255617]]
840273	-0.378373	51.396065	► [1,2,[],[-0.378373,51.396065]]
869433	-0.577538	51.254855	► [1,2,[],[-0.577538,51.254855]]
840333	-0.37714	51.397181	► [1,2,[],[-0.37714,51.397181]]
869463	-0.576158	51.253139	► [1,2,[],[-0.576158,51.253139]]
840393	-0.378373	51.396065	► [1,2,[],[-0.378373,51.396065]]
869493	-0.578278	51.252239	► [1,2,[],[-0.578278,51.252239]]
840453	-0.362476	51.399657	► [1,2,[],[-0.362476,51.399657]]
869523	-0.572772	51.251434	► [1,2,[],[-0.572772,51.251434]]
840513	-0.368456	51.398841	► [1,2,[],[-0.368456,51.398841]]
869553	-0.570782	51.25784	► [1,2,[],[-0.570782,51.25784]]
840573	-0.374562	51.402595	► [1,2,[],[-0.374562,51.402595]]
869583	-0.572772	51.251434	► [1,2,[],[-0.572772,51.251434]]
840633	-0.379488	51.399911	► [1,2,[],[-0.379488,51.399911]]
869613	-0.571793	51.255298	► [1,2,[],[-0.571793,51.255298]]
840693	-0.378888	51.401	► [1,2,[],[-0.378888,51.401]]
869643	-0.567016	51.260977	► [1,2,[],[-0.567016,51.260977]]
840753	-0.374562	51.402595	► [1,2,[],[-0.374562,51.402595]]
869673	-0.570782	51.25784	► [1,2,[],[-0.570782,51.25784]]
840813	-0.379488	51.399911	► [1,2,[],[-0.379488,51.399911]]
869703	-0.572772	51.251434	► [1,2,[],[-0.572772,51.251434]]
840873	-0.366411	51.401384	► [1,2,[],[-0.366411,51.401384]]
869733	-0.542652	51.26457	► [1,2,[],[-0.542652,51.26457]]
840933	-0.368283	51.400466	► [1,2,[],[-0.368283,51.400466]]
869763	-0.543221	51.263345	► [1,2,[],[-0.543221,51.263345]]
840993	-0.371284	51.402999	► [1,2,[],[-0.371284,51.402999]]
869793	-0.552482	51.264188	► [1,2,[],[-0.552482,51.264188]]
841053	-0.372976	51.399902	► [1,2,[],[-0.372976,51.399902]]
869823	-0.544084	51.261	► [1,2,[],[-0.544084,51.261]]
841113	-0.409255	51.392604	► [1,2,[],[-0.409255,51.392604]]
869853	-0.544084	51.261	► [1,2,[],[-0.544084,51.261]]
841173	-0.411745	51.39708	► [1,2,[],[-0.411745,51.39708]]
869883	-0.544084	51.261	► [1,2,[],[-0.544084,51.261]]
841233	-0.407414	51.388496	► [1,2,[],[-0.407414,51.388496]]

869913	-0.544084	51.261	▸ [1,2,[],[-0.544084,51.261]]
841293	-0.409255	51.392604	▸ [1,2,[],[-0.409255,51.392604]]
869943	-0.544084	51.261	▸ [1,2,[],[-0.544084,51.261]]
841353	-0.403941	51.394959	▸ [1,2,[],[-0.403941,51.394959]]
869973	-0.544084	51.261	▸ [1,2,[],[-0.544084,51.261]]
841413	-0.409255	51.392604	▸ [1,2,[],[-0.409255,51.392604]]
870003	-0.544084	51.261	▸ [1,2,[],[-0.544084,51.261]]
841473	-0.383089	51.387471	▸ [1,2,[],[-0.383089,51.387471]]
870033	-0.545669	51.264356	▸ [1,2,[],[-0.545669,51.264356]]

Showing the first 1000 rows.



```
# menampilkan dataset1 yang dimana merupakan sumber data setelah cleaning
display(df_with_pk1)
```

id	Crime ID	Month	Year	Reported by	Falls within
0	8213c26857d308b164c483db257ea5b4096a4ed8c6bc5b822073b810d71315ec	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
30	1de75978787da838c23abf65c7fbf709a7a8a1027665e6728d0f87febf544bd4	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
60	5aa3789b3f44aea8a85805264826ee65ab2e970bfdb6cd505922102d0f6b1bae	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
90	161bc4110580834cc3f11a696b4d59fdbc24c4f84af7a2d363d603155c15933e	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
120	36f405914ea34bab24b5f540ff1a961d2dfa5e2c4937c960272a55e6c15bccd7	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
150	38161324b2591802147040dcc791c083d02a9a07d4ac30b30219708f2800212b	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

180	69570bb2eca3b2e062e0657038f24d6f60d6bd5c2b670522cde549a72f4edd49	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
210	6f4e774ef0c057319b33330bdea5ce84fb1faea3b05e6563dcf9a78c0bbd2362	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
240	a640a6b37298ad1afc29dc4daf9d73f83a5205999e47112d9732ec7a6b929b4d	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
270	08788344efa9bcb3fe53269f8bd44b839435cd13bcc8b7223e38ed1dea9dd7b1	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
300	a22480661166eafe47a258cfb6f21505fe2234e6e769be90c0d1d357da6b6a67	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
330	c9d779a7b3c271a56f07797a9d082bbaa5de46fd09678336f653f2e8fad35924	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
360	7399b8ea36cdddcde5f7f030cfa97d8eeedecfb3fde144b809a92d4836fa429	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
390	2d908be5d68d14656b332fcb71dd6c67d1841852184f833c3dc0a4dec57f39fd	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
420	4eb26a9c37830427f26a73a2e9bfd2cf958dd8576ad8f966da857467466b5526	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
450	ede6daebefcb542fd454d1555a5d36728979d9f2dc2bc12ca33ca92848e963f3	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

480	62f04c18ee6f940348733c1db8cf1700f428decead0edac479bf60606bad6969	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
510	55241de520aef65472d06762f6b65ef8ede59235b4bf3a1d735eaf07fdc0417d	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
540	86c4861edd2522db347a045ea09766430f9d57d2652d05492c8bebd3c3775e8f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
570	311877567250db247e77fc18bd3399a5a765190f182f5d203f1f73e85fad5ea	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
600	7f39e79eb2e5f77c1aac04487a491ec049c54c89aa2f3cd8696196c7bb92a9af	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
630	ae27d79b527fc07c94f61d8786dfcac45fd5b63c7aae6ee9e503103e1f516bdb	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
660	7293ed353aa2ad7fe094335914f9620edfcb54860f1a9e5e390ed848f99db01c	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
690	75373d2714776a5649d485ffa6d10f4a9963bd28ac801a4f630303fddeed73b4	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
720	8da849dc19c75fc83dd949aca962136b55f2df14a68efd118f80e5f332ca7b14	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
750	9f4e6c636caffc6f93bdb0b65f44aba902f52bdbe0272b4f404e31ec86f675ba	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

780	32d8313a14db77f8e340f2f0e17bbb72f931613f41804cbb917753ac8295a959	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
810	e5fed4f71886928cf79abd6eb7bcce8a8f75389edd58e79d36b23555f94ee3ac	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
840	f27dd602dcd43e954bf488fe6576835dac43ff5858f82cdf2afe65eee0c21b44	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
870	6347421bdf1085da56ef9cbba1e602b291a11cb3ad93a6a27751dd0eeff2f8c2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
900	be0a334561feb23718d1781da94d069c190876646516f67659b03acc0b07fea2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
930	bae3ef31bfa79dff4d1b92ae7599281f24572634be00942fbd5d3d2407ff1bb6	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
960	e3e3d9f88f01c951452641f9cc00dacd65d66ed1b49b015b21b496fdc9f5d045	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
990	a63ef766b054ebcc521fc3b8429f99b0093fc75e6e6c26093509c382b15ef662	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1020	97a6eddf80835b2af253b55f4685dd877bcd84594926c9b325edd00d1d67a5eb	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1050	ce31951675937a74be127c147f069da5411a0112c9f6e8bcc6414a9435127716	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

1080	6ce50434edf3957a979b113f51e92f9ba7a0807152846eac49974ebb82cb4d8d	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1110	d13ba87b53c5d0d793769c55e2a95dbc6af9b1d1349e6450a84cdbe45cb71361	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1140	db80195970b1966ab7165e5b0136004c2ba53636e9eb7fb3b212dd2bad8a3750	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1170	2bfc267652c597ec6aea9e3cae7655e6bd91add18c97b1db202080b7b4b3ea49	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1200	355e8a61245911c636959863bf7326f505e53b9d9ae22bef80bf1fc573fefa8d	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1230	87b1040fd4b4edc134054051ac77b5695da6fa4d604f854451558622a021c099	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1260	0ca87abaed51b362dbcc8298a6656bc28fa4dc5457f81ba45a1bf016bb415ebe	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1290	d3cf824c745fdcc501a0437b6cbb678d305743cec0a7c498f55d81958df57ca8	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1320	aef2ac01ba58ab6aaf834fa96104a33df032b345f856deb7cc9eece86f7551ec	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1350	a20299ad60679e9d3237a1b2c7e64964530af4f79b322c090c1237bdfc2b4116	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

1380	96e5802cb5fba8de80ec53e48e718f5ce9b99b53e0aae5469466dc06ab199e53	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1410	020005f2a333f91897fd237d0c03a50be37f9e25297ec023ee3f1c1ad315c7f2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1440	3cfce12071ca2f2ec142dfb17ca612a02f58a6d9f33795b2bfb615d3503a2915	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1470	d818a6d382f03cab3f737f375d697c3e9d9535162db2f1d2174125956c30c1da	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1500	e2deca2b713a7866796b6798d6de49c882f7c67133cf84cd3e1f01bf57fce142	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1530	34f9f84eeb8a6dc5654657341c99e4cbe0b509300f9f1d2e6f42bfa0e1181316	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1560	ebfea65ab45ff726840c0e8d5a3e9cc4e496c39f82c6834c9a38dc7ad8869f3f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1590	4b937e8b0ca5f87923449186be827a84ad10e8c0a495e1a068fbdba0e41cdcd2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1620	b523c3c4a287b860b4fdb917401bdc397a3499c3cd4c0e97a22879c1700960b1	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1650	b119321ff667129910098f9f7e9960af1dd082d7d2ce9d565e6b6bc67380be1f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

1680	998019783fab8fc29fdeaf066e070c5c7b548a5748afa992a7bb83a8fd144cc2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1710	facf994285bf665655a668b13b78925d5e617ef8f75ef1ba1719a3c83a942d5e	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1740	df33d3647ad1a77d50379a73427d293077332f3fc948e057e6224fcf2b1f04c	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1770	e632c83a84eb135be9eaa1176d5f936318819eb894d2962852f813cea4a5ea28	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1800	c29e078d6f9e724659d40a1f99aa83a1396d1cbb997515f6f65ce18377221b1e	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1830	78f8c8a65b95b19a290693a5cb9720d7fe172764b2c1dc766c17758c05788e04	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1860	a5a94ce223a8de8452d8b06a6f8332740812dcd73f4510cb8fb64b486c0a76f6	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1890	017898e182fa252fc18f082829d2fbeece1c597358e30fd130b056c5b96d608d	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1920	d0872c2aedd21de8b031ba21815235fba71870e3758e97aea4e06f384ff54d83	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
1950	970d7bac1737bd4c0879b9fb72ca7d9652a873e4554c42303933c6891f530304	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

1980	8f51524a52fcaf1d2459b60554ca44b6c79fcbefe289c2dd0ae619217ad68a4a	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2010	244dee38138be8110bf6cdc1e1823cbb60ce6435c4452d41a5f8d3148baf8748	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2040	25dd851f4a1ac1d9f6f73ca51fd54949db7ca83c42da83ce8d63b8c73ddbe3b1	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2070	00d697f352c5d07fed5940070307c41622043adeea2ef1454b20ccb5ecf34969	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2100	cfae6ff67ae51fb9b3d47d08f6b51d1233a5c4c9bc44ab8d4abc9c0b54e4da14	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2130	57d446bffffbd23299693805068e846b08706a00f20c04faffb6380de5f5546f7	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2160	865249be9b21b1c1ae7dce7fa72a1e70c5ac8138f7058710d2b29793314f3d20	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2190	2b537c8c51e7abd0b0f2343feb3e959bcf163e7e666cdf3736c9482b3fd3d63	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2220	726f32ac663d52e017b13719b612a5bb708a1d7f1160478bee62ddf55312a1c9	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2250	e0631e55ebeab8045af026b62d127c0ad9dd2239a3ee22b356397bf1647d939f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

2280	e37784aacc109beb3edd9a181faa361d890fa16830b8bdee8b76e5fc1839cd35	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2310	20959c2c16205ef7cd48db6e8348f850cac0578bc897919f4212abdc400b084f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2340	d8baa630c4dc9725ca038797cc4a158d9bf81a97fddb4e53a4fc143681a68871	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2370	927412fa30f35fdafeefbd7e3269e93215c02f2a6031bf8765b124c242077ade	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2400	16ce83d1148b5f139386e0af9fad2819746b4f3d82d82b597c172cf9c752f77f	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2430	9184b9a29d9452356057dda3846d8e95ede7146f89a5df65f966bfd9abb4852c	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2460	1bf60a04203be616a623436d0b6400b9d75930f60b7a77a6beac12b2258378d2	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2490	33661b5025483ddd5ffe4515233341e7a8b89b842ce48616b31d4d4fa2d7ee68	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2520	752e20d9e1d47c71e2e656c1636b4630c62e9b0209b36d47eb4257c71aad2def	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2550	e000104282d15eb39b80b1282e8918592a933f7811ad85de27b9c4046c4053a0	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

2580	ba5326f70c5ae754b3481cc9fa9a9250a509f18859cb13af09d098996e6b2d37	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2610	9d6dab2c702fe96aaa602d01fa457a33a3389d70bf036851eb004ff11576dccc	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2640	b372b586e551cab3a862e2b818e2bdb5b83aa4bdae2f7d090901cd473b134aba	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2670	593e6099a9bcbda406a83dc10c9d4a6aec88b877e4605ba2752e7d742cb11cc	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2700	4d1f65d62d1cea27ad76e712e922dc8682161731c16317b141a80bf892fd6b62	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2730	f8c9cc7e33b5029752bd28292c36b89609954eef23c5cc5b7369a8a1987bc13b	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2760	cdf1ee1a24fc1d2fe7c0d91b063ce73ac68b521e02f4df77d78969c06c9e46e4	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2790	e2b1c0e183a6a4ed45314678bc4338f58817dac4fbe8be74c2858d0c713af376	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2820	c7970a53635a634a3adb6b1323389857d9701a7aa7adef516895f0073c54467a	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2850	d92296a6b7c817d2899fb91e8785ff5ec329fe5e589d11421e098865b06b90b4	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

2880	ff08e19fbbd215af7f3d6a94cba124d8e97e444daa42a2f54385498b0674c06e	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary
2910	e6ba00d7dae2273b15a62386b9f8a9f22f48d91d28ab908a177a8f12056dba24	10	2014	Avon and Somerset Constabulary	Avon and Somerset Constabulary

Showing the first 1000 rows.



```
# mempersiapkan dengan memilih kolom yang digunakan untuk join dengan tabel sumber data
df_with_pk3 = df_with_pk2.select("id","features","prediction")
```

Join Dataset1 dan Dataset2

```
# melakukan join, estimated execution: 20 min
joined = df_with_pk1.join(df_with_pk3, ["id"])
```

```
# melakukan displaying, estimated execution: 20 min
display(joined)
```

```
# total waktu dieksekusi kedua baris itu ialah: 40 min
total_after_join = str(joined.count())
# print("total baris setelah join: "+ )
```

Cancelled

```
# print baris setelah join, hasilnya harus tetap dengan setelah ataupun sebelum
print("total baris setelah join2: "+ total_after_join)
```

total baris setelah join2: 11949434

```
# Menampilkan banyaknya per jenis kriminal pada 9 hasil cluster versi 1
display(joined.orderBy("prediction").groupBy("id","Month","Year","Reported by","Falls
within","Longitude","Latitude","Location","LSOA code","LSOA name","Crime type","Last outcome
category","features","prediction").count())
```

prediction	Bicycle theft	Burglary	Criminal damage and arson	Drugs	Other crime	Other theft	Pc
0	89114	335622	361111	145343	51583	490633	29
1	19403	112513	180501	40734	18291	117059	64
2	23146	101912	133611	27693	16753	103313	67
3	18456	90861	155619	41378	18796	124178	56
4	23643	196344	243951	38724	32371	194163	11
5	10712	57738	113930	17686	12153	75544	43
6	28607	165147	190088	39937	23184	160670	10
7	17353	75836	117885	22686	16513	88411	52

8	27556	76704	114002	26039	13531	91981	59

Only showing the first ten series.



```
# Menampilkan banyaknya per jenis kriminal pada 9 hasil cluster versi 2
display(joined.orderBy("prediction").groupBy("Crime type","prediction").count())
```

Crime type	prediction
Criminal damage and arson	0
Theft from the person	0
Public order	0
Vehicle crime	0
Violence and sexual offences	0
Burglary	0
Shoplifting	0
Other theft	0
Bicycle theft	0
Robbery	0
Other crime	0
Drugs	0
Possession of weapons	0
social behaviour	0
Shoplifting	1
Vehicle crime	1
Violence and sexual offences	1
Criminal damage and arson	1
Burglary	1
Other theft	1
Theft from the person	1
Drugs	1
Bicycle theft	1
Other crime	1
Public order	1
Possession of weapons	1
Robbery	1
Criminal damage and arson	2
Vehicle crime	2
Violence and sexual offences	2
Other theft	2
Shoplifting	2
Public order	2
Other crime	2

Drugs	2
Burglary	2
Robbery	2
Bicycle theft	2
Possession of weapons	2
Theft from the person	2
social behaviour	2
Shoplifting	3
Other crime	3
Criminal damage and arson	3
Other theft	3
Violence and sexual offences	3
Vehicle crime	3
Public order	3
Drugs	3
Burglary	3
Bicycle theft	3
Robbery	3
Theft from the person	3
Possession of weapons	3
Criminal damage and arson	4
Burglary	4
Public order	4
Violence and sexual offences	4
Other theft	4
Vehicle crime	4
Shoplifting	4
Theft from the person	4
Possession of weapons	4
Drugs	4
Bicycle theft	4
Robbery	4
Other crime	4
Violence and sexual offences	5
Criminal damage and arson	5
Other theft	5
Public order	5
Vehicle crime	5
Burglary	5
Shoplifting	5
Drugs	5
Theft from the person	5


Other crime	5
Robbery	5
Bicycle theft	5
Possession of weapons	5
Vehicle crime	6
Violence and sexual offences	6
Criminal damage and arson	6
Burglary	6
Shoplifting	6



```
# Menampilkan banyaknya kriminal pada 9 hasil cluster dengan pola per bulan  
display(joined.orderBy("prediction").groupBy("Month","prediction").count())
```

Month	prediction
05	0
04	0
11	0
01	0
12	0
10	0
06	0
02	0
07	0
09	0
08	0
03	0
07	1
03	1
04	1
12	1
06	1
01	1
10	1
11	1
09	1
02	1
05	1
08	1
10	2
03	2
08	2
05	2

07	2
12	2
06	2
02	2
04	2
09	2
01	2
11	2
10	3
04	3
12	3
06	3
03	3
09	3
05	3
11	3
01	3
08	3
02	3
07	3
08	4
07	4
01	4
10	4
03	4
12	4
04	4
06	4
05	4
09	4
02	4
11	4
05	5
04	5
01	5
12	5
06	5
10	5
09	5
02	5
07	5
11	5

03	5
08	5
03	6
08	6
01	6
07	6
11	6
10	6
04	6
	

```
# Menampilkan banyaknya kriminal pada 9 hasil cluster dengan pola per Tahun  
display(joined.orderBy("prediction").groupBy("Year","prediction").count())
```

Year	prediction
2015	0
2017	0
2014	0
2016	0
2017	1
2016	1
2014	1
2015	1
2015	2
2017	2
2016	2
2014	2
2014	3
2016	3
2017	3
2015	3
2015	4
2017	4
2016	4
2014	4
2016	5
2017	5
2014	5
2015	5
2017	6
2016	6
2015	6
2014	6

2016	7
2017	7
2015	7
2014	7
2015	8
2016	8
2014	8
2017	8



```
# Menampilkan banyaknya jenis kriminal pada 9 hasil cluster dengan pola per Tahun
display(joined.orderBy("Year").groupBy("Crime type","Year","prediction").count())
```

Crime type	Year
Shoplifting	2014
Other crime	2014
Criminal damage and arson	2014
Public order	2014
Violence and sexual offences	2014
Public order	2014
Criminal damage and arson	2014
Shoplifting	2014
Bicycle theft	2014
Other theft	2014
Drugs	2014
Violence and sexual offences	2014
Burglary	2014
Vehicle crime	2014
Burglary	2014
Other theft	2014
Theft from the person	2014
Vehicle crime	2014
Other theft	2014
Burglary	2014
Bicycle theft	2014
Other crime	2014
Violence and sexual offences	2014
Violence and sexual offences	2014
Criminal damage and arson	2014
Robbery	2014
Shoplifting	2014
Drugs	2014

Shoplifting	2014
Criminal damage and arson	2014
Violence and sexual offences	2014
Shoplifting	2014
Vehicle crime	2014
Burglary	2014
Other theft	2014
Public order	2014
Other theft	2014
Shoplifting	2014
Shoplifting	2014
Other theft	2014
Criminal damage and arson	2014
Possession of weapons	2014
Violence and sexual offences	2014
Other theft	2014
Vehicle crime	2014
Bicycle theft	2014
Burglary	2014
Drugs	2014
Theft from the person	2014
Burglary	2014
Violence and sexual offences	2014
Other theft	2014
Drugs	2014
Vehicle crime	2014
Criminal damage and arson	2014
Shoplifting	2014
Robbery	2014
Burglary	2014
Other crime	2014
Public order	2014
Bicycle theft	2014
Robbery	2014
Vehicle crime	2014
Criminal damage and arson	2014
Burglary	2014
Violence and sexual offences	2014
Drugs	2014
Possession of weapons	2014
Drugs	2014
Robbery	2014

Bicycle theft	2014
Vehicle crime	2014
Vehicle crime	2014
Robbery	2014
Public order	2014
Public order	2014
Drugs	2014
Shoplifting	2014
Violence and sexual offences	2014



Crime type
Bicycle theft
Burglary
Criminal damage and arson
Drugs
Other crime
Other theft
Possession of weapons
Public order
Robbery

