

## Project on Similarity Search

Ankit Dhall

The query was pre-defined into variables. The categorical variables of this query were then encoded into numeric values to enable distance calculation.

This encoding was done on the basis of the encoding used in the MIF generation code.

Reference points were also defined on the basis of the reference points used in the MIF generation code.

After running the search algorithm, the nearest/most similar data objects from the input file were retrieved and stored in a csv file.

### Query 3:

```
1st_Road_Class = "Unclassified"
1st_Road_Number = "0"
Accident_Severity = "Slight"
Day_of_Week = "Friday"
Did_Police_Officer_Attend_Scene_of_Accident = "1"
Junction_Control = "Give way or uncontrolled"
Junction_Detail = "T or staggered junction"
Light_Conditions = "Daylight"
Local_Authority_(District) = "Sefton"
Number_of_Casualties = "1"
Number_of_Vehicles = "1"
Pedestrian_Crossing-Human_Control = "0"
Pedestrian_Crossing-Physical_Facilities = "0"
Police_Force = "Merseyside"
Road_Surface_Conditions = "Wet or damp"
Road_Type = "Single carriageway"
Speed_limit = "30"
Urban_or_Rural_Area = "Urban"
Weather_Conditions = "Fine no high winds"
Year = "2005"
```

Execution of the code was done in a Jupyter Notebook and the following query and results were obtained.

localhost:8888/notebooks/similarity\_search.ipynb

jupyter similarity\_search Last Checkpoint: a minute ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [56]: query
Out[56]:
```

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_Condition
0	Unclassified	0	Slight	Friday	1	Give way or uncontrolled	T or staggered junction	

```
In [57]: search_result.head(10)
Out[57]:
```

	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_Condition
31187	Unclassified	0.0	Slight	Sunday		1.0	Data missing or out of range	Not at junction or within 20 metres
33120	Unclassified	0.0	Slight	Thursday		1.0	Give way or uncontrolled	Other junction
35329	Unclassified	0.0	Slight	Saturday		1.0	Give way or uncontrolled	T or staggered junction
35453	Unclassified	0.0	Slight	Friday		1.0	Give way or uncontrolled	T or staggered junction
35512	Unclassified	0.0	Slight	Friday		1.0	Give way or uncontrolled	T or staggered junction
35737	C	0.0	Slight	Saturday		1.0	Give way or uncontrolled	T or staggered junction
35848	C	0.0	Slight	Monday		2.0	Give way or uncontrolled	Private drive or entrance
35979	Unclassified	0.0	Slight	Friday		1.0	Give way or uncontrolled	Roundabout
36905	C	0.0	Slight	Thursday		2.0	Give way or uncontrolled	Roundabout
36937	Unclassified	0.0	Slight	Monday		1.0	Give way or uncontrolled	T or staggered junction

These search results were stored in a CSV file along with the query

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
	1st_Road_Class	1st_Road_Number	Accident_Severity	Day_of_Week	Did_Police_Officer_Attend_Scene_of_Accident	Junction_Control	Junction_Detail	Light_Condition	Local_Authority	Number_of_Pedestrians	Number_of_Pedestrians	Police_Force	Road_Type	Road_Type	Speed_Limit	Urban_or_Rural	Weather_Condition	Year				
1	Query	Unclassified	0	Slight	Friday	1	Give way or uncontrolled	T or staggered	Daylight	Sefton	1	1	0	0	Merseyside		Wet	Single carriageway	30	Urban	Fine no high	2005
2																						
3																						
4	31187	Unclassified	0	Slight	Sunday	1	Data missing or out of range	Not at junction	Daylight	South Ribble	1	1	0	0	Lancashire		Dry	Unknown	30	Urban	Fine no high	2005
5	33120	Unclassified	0	Slight	Thursday	1	Give way or uncontrolled	Other junction	Darkness	Rossendale	2	3	0	0	Lancashire		Wet	Single carriageway	30	Urban	Raining no h	2005
6	35329	Unclassified	0	Slight	Saturday	1	Give way or uncontrolled	T or staggered	Daylight	Sefton	1	2	0	0	Merseyside		Wet	Single carriageway	30	Urban	Fine no high	2005
7	35453	Unclassified	0	Slight	Friday	1	Give way or uncontrolled	T or staggered	Daylight	Sefton	2	2	0	0	Merseyside		Wet	Single carriageway	30	Urban	Fine no high	2005
8	35512	Unclassified	0	Slight	Friday	1	Give way or uncontrolled	T or staggered	Darkness	Sefton	1	1	0	1	Merseyside		Wet	Single carriageway	30	Urban	Fine no high	2005
9	35737	C	0	Slight	Saturday	1	Give way or uncontrolled	T or staggered	Daylight	Sefton	3	2	0	0	Merseyside		Wet	Single carriageway	60	Rural	Fine no high	2005
10	35848	C	0	Slight	Monday	2	Give way or uncontrolled	Private drive or	Daylight	Sefton	1	1	0	0	Merseyside		Wet	Single carriageway	30	Urban	Fine no high	2005
11	35979	Unclassified	0	Slight	Friday	1	Give way or uncontrolled	Roundabout	Daylight	Sefton	1	2	0	0	Merseyside		Wet	Roundabout	30	Urban	Fine no high	2005
12	36905	C	0	Slight	Thursday	2	Give way or uncontrolled	Roundabout	Daylight	St. Helens	1	2	0	0	Merseyside		Dry	Roundabout	30	Urban	Fine no high	2005
13	36937	Unclassified	0	Slight	Monday	1	Give way or uncontrolled	T or staggered	Daylight	St. Helens	1	1	0	0	Merseyside		Dry	Single carriageway	30	Urban	Fine no high	2005
14	36988	Unclassified	0	Slight	Monday	1	Give way or uncontrolled	T or staggered	Daylight	St. Helens	1	1	0	0	Merseyside		Dry	Single carriageway	30	Urban	Fine no high	2005
15	41400	A	6	Slight	Saturday	1	Auto traffic signal	Crossroads	Daylight	Salford	1	1	0	5	Greater Manchester		Wet	Single carriageway	30	Urban	Raining no h	2005
16	41413	Unclassified	0	Slight	Wednesday	1	Give way or uncontrolled	Crossroads	Daylight	Salford	1	2	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
17	41429	A	6	Slight	Tuesday	1	Data missing or out of range	Not at junction	Daylight	Salford	1	2	0	4	Greater Manchester		Wet	Single carriageway	30	Urban	Raining no h	2005
18	41432	Unclassified	0	Serious	Wednesday	1	Data missing or out of range	Not at junction	Daylight	Salford	1	1	0	0	Greater Manchester		Dry	One way street	30	Urban	Fine no high	2005
19	41433	Unclassified	0	Slight	Friday	2	Give way or uncontrolled	T or staggered	Daylight	Salford	1	2	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
20	41436	Unclassified	0	Slight	Wednesday	1	Give way or uncontrolled	Crossroads	Daylight	Salford	1	2	1	1	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
21	41453	Unclassified	0	Slight	Wednesday	1	Give way or uncontrolled	Crossroads	Daylight	Salford	2	3	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
22	41471	Unclassified	0	Slight	Friday	2	Give way or uncontrolled	Private drive or	Daylight	Salford	1	2	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine + high v	2005
23	41489	Unclassified	0	Serious	Tuesday	1	Data missing or out of range	Not at junction	Daylight	Salford	1	1	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
24	41490	Unclassified	0	Serious	Tuesday	1	Give way or uncontrolled	Crossroads	Daylight	Salford	2	2	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
25	41518	A	6	Slight	Tuesday	1	Give way or uncontrolled	Crossroads	Daylight	Salford	2	2	0	4	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
26	41630	Unclassified	0	Slight	Saturday	1	Give way or uncontrolled	T or staggered	Daylight	Salford	1	2	0	0	Greater Manchester		Wet	Single carriageway	30	Urban	Raining no h	2005
27	41642	Unclassified	0	Slight	Tuesday	1	Data missing or out of range	Not at junction	Daylight	Salford	1	1	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
28	41651	A	6	Slight	Friday	1	Give way or uncontrolled	Other junction	Daylight	Salford	1	1	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
29	41661	Unclassified	0	Slight	Tuesday	1	Data missing or out of range	Not at junction	Darkness	Salford	1	1	0	0	Greater Manchester		Dry	Single carriageway	30	Urban	Fine no high	2005
	querv3 result																					

In the above screenshot, the query has been highlighted in orange. The indices of the data objects (from the original data file) retrieved in the search results are highlighted in grey.