CAS data exploration

Outline:

1. Dataset inspection
2. National trend
3. Regional variation
4. Roadside features and vehicle types
5. Work in progress...

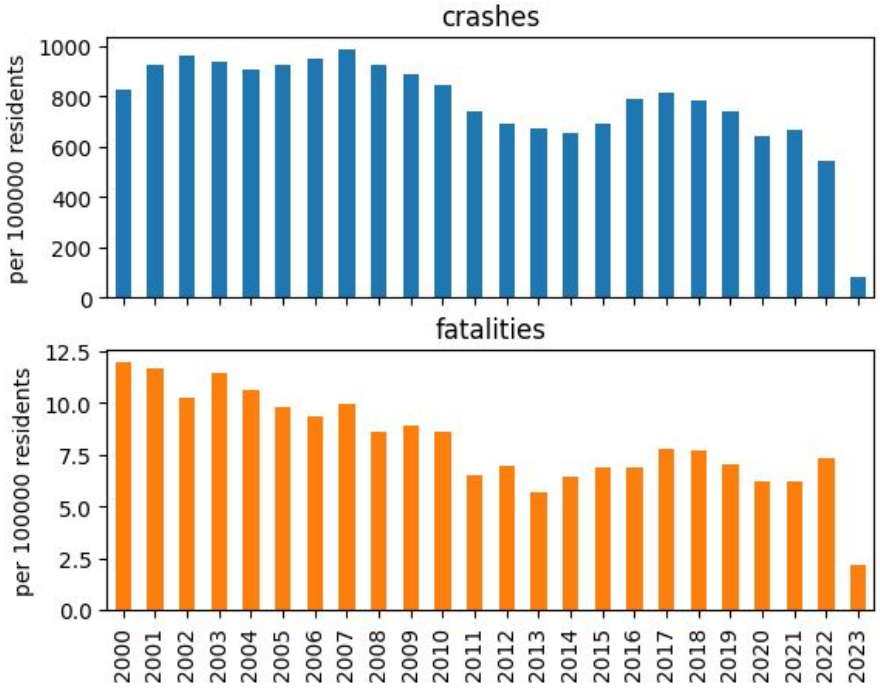
Data inspection

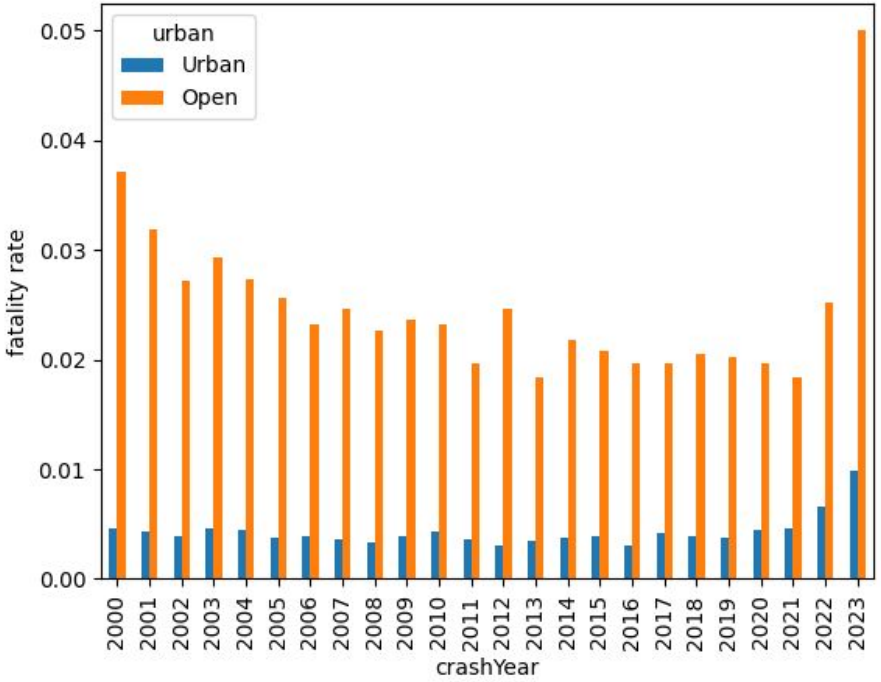
Out[2]:

	X	Y	OBJECTID	advisorySpeed	areaUnitID	bicycle	bridge	bus	carStationWagon	cliffBank	...	tre
0	2.037858e+06	5.707835e+06	1	NaN	544801.0	0.0	NaN	0.0	2.0	NaN	...	Nal
1	1.799424e+06	5.815528e+06	2	NaN	528900.0	0.0	NaN	0.0	2.0	NaN	...	Nal
2	1.756461e+06	5.936053e+06	3	NaN	507000.0	0.0	NaN	0.0	0.0	NaN	...	Nal
3	1.551129e+06	5.171320e+06	4	NaN	597513.0	0.0	0.0	0.0	2.0	0.0	...	0.
4	1.245391e+06	4.849172e+06	5	NaN	611500.0	0.0	NaN	0.0	1.0	NaN	...	Nal
...	...	...	...	...	...	...	...	...	...	...	...	.
821739	1.757918e+06	5.914599e+06	1318959	NaN	518600.0	0.0	NaN	0.0	2.0	NaN	...	Nal
821740	1.833766e+06	5.638669e+06	1318960	NaN	554900.0	0.0	0.0	0.0	0.0	0.0	...	1.
821741	1.758255e+06	5.918060e+06	1318961	NaN	517400.0	0.0	0.0	0.0	1.0	0.0	...	0.
821742	1.773738e+06	5.888266e+06	1318962	NaN	NaN	0.0	0.0	0.0	0.0	0.0	...	0.
821743	1.754396e+06	5.875116e+06	1318963	NaN	526104.0	0.0	0.0	0.0	1.0	0.0	...	0.

821744 rows × 73 columns

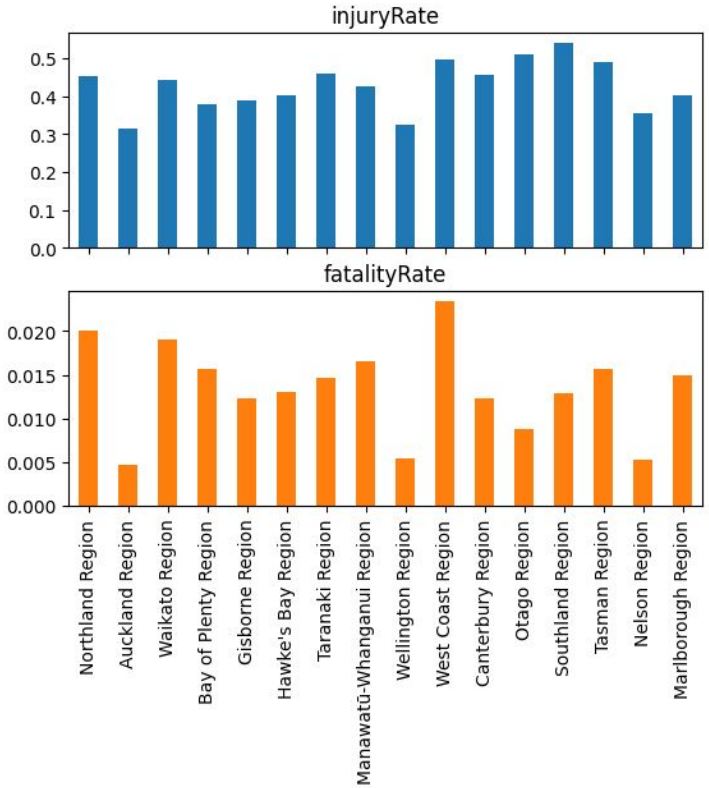
National trends

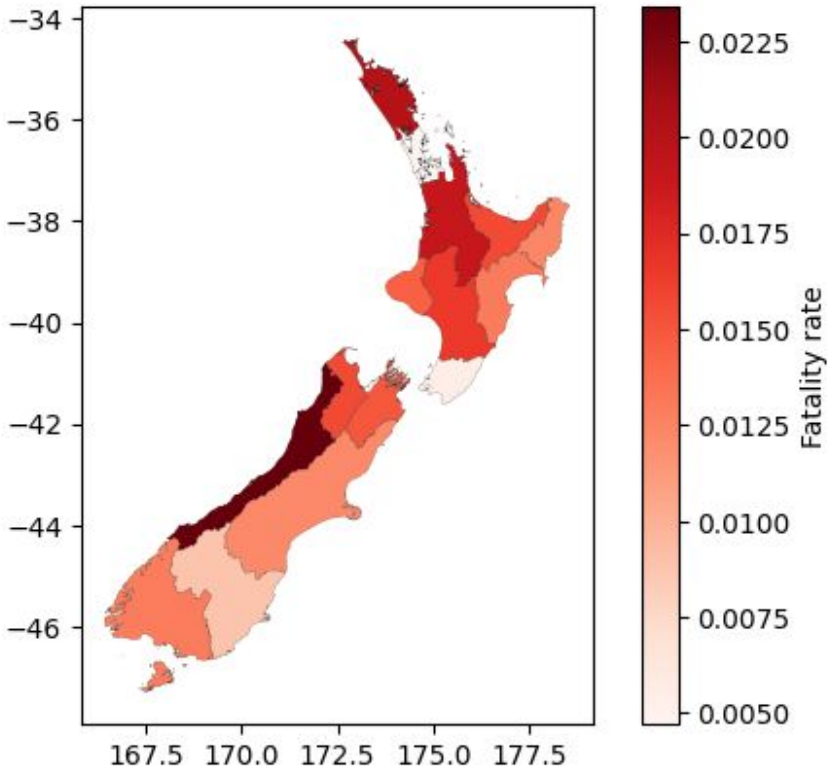




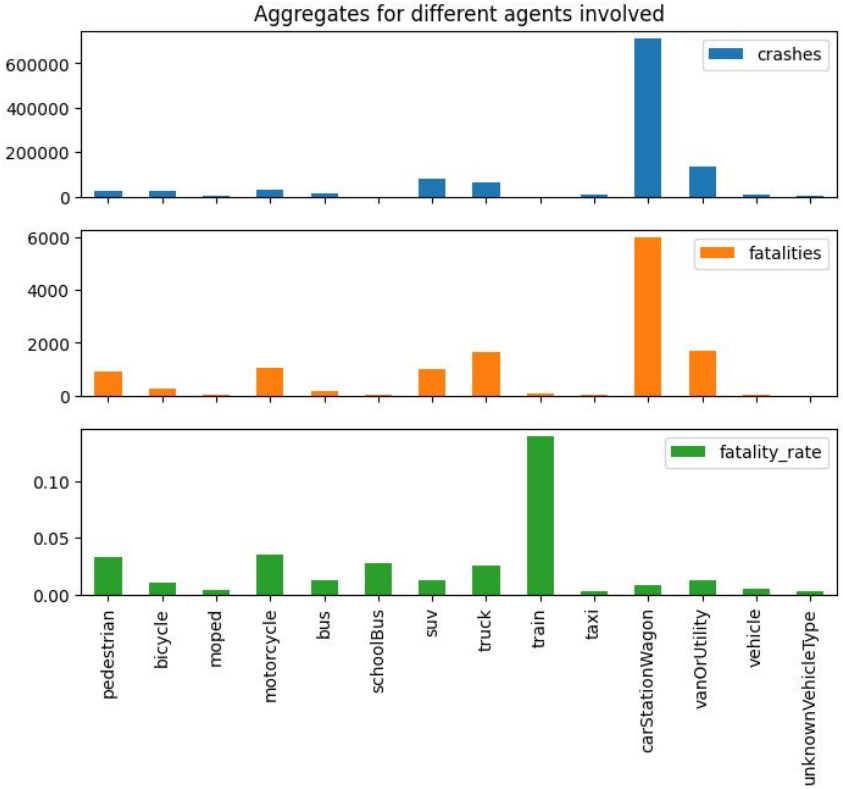
(Global annual fatalities: ~18 per 100000 pop)

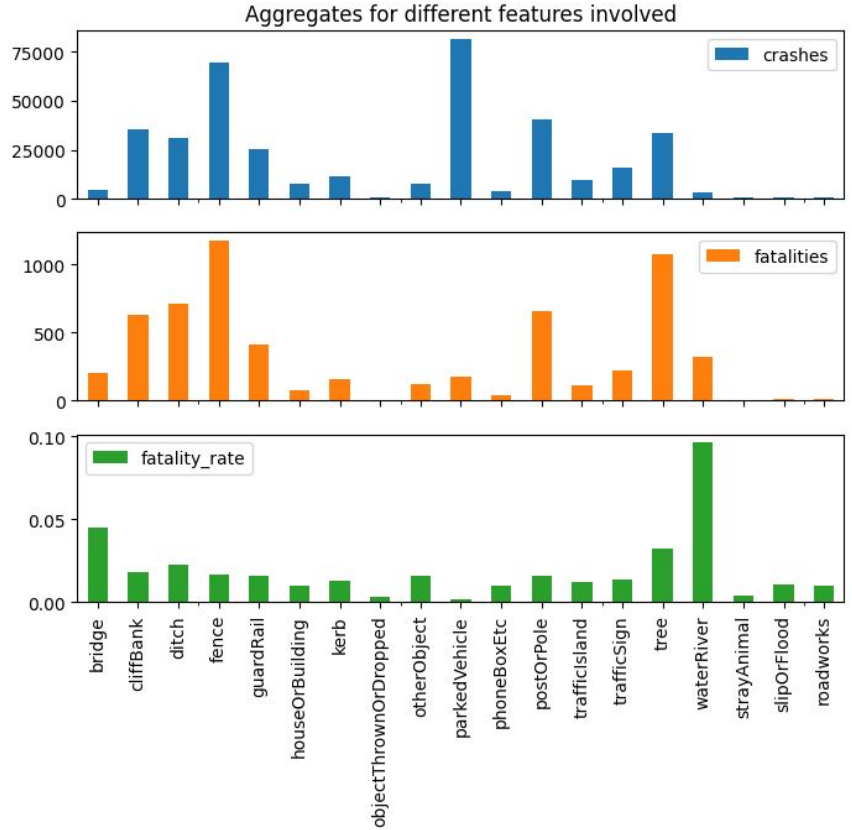
Regional variation





Lethal features and vehicle types





Work in progress...

Started looking for lethal combinations of vehicle types and particular features involved in crashes. Preliminary findings:

1) Crashes involving at least two station-wagons are the most fatal, with **1137 deaths** in 322170 crashes.
2) Crashes involving at least one truck and one station-wagon come second, featuring in 38284 crashes with **723 deaths**.

Interpretation of these partial aggregates becomes more subtle, so more care will need to be taken when communicating them.

Goal: to characterise circumstances associated with higher fatality rates.