

Drop Justification

--- Missing Values in Crashes ---

	column	missing_percent	\
VEHICLE TYPE CODE 5	VEHICLE TYPE CODE 5	99.561614	
CONTRIBUTING FACTOR VEHICLE 5	CONTRIBUTING FACTOR VEHICLE 5	99.547390	
VEHICLE TYPE CODE 4	VEHICLE TYPE CODE 4	98.408010	
CONTRIBUTING FACTOR VEHICLE 4	CONTRIBUTING FACTOR VEHICLE 4	98.348052	
VEHICLE TYPE CODE 3	VEHICLE TYPE CODE 3	93.044659	
CONTRIBUTING FACTOR VEHICLE 3	CONTRIBUTING FACTOR VEHICLE 3	92.763505	
OFF STREET NAME	OFF STREET NAME	82.370849	
CROSS STREET NAME	CROSS STREET NAME	38.213525	
ZIP CODE	ZIP CODE	30.643796	
BOROUGH	BOROUGH	30.631327	
ON STREET NAME	ON STREET NAME	21.787222	
VEHICLE TYPE CODE 2	VEHICLE TYPE CODE 2	20.069510	
CONTRIBUTING FACTOR VEHICLE 2	CONTRIBUTING FACTOR VEHICLE 2	16.059128	
LATITUDE	LATITUDE	10.819789	
LONGITUDE	LONGITUDE	10.819789	
LOCATION	LOCATION	10.819789	

--- Missing Values in Persons ---

	column	missing_percent	missing_count
CONTRIBUTING_FACTOR_2	CONTRIBUTING_FACTOR_2	98.266294	5722518
CONTRIBUTING_FACTOR_1	CONTRIBUTING_FACTOR_1	98.264011	5722385
PED_ACTION	PED_ACTION	98.242048	5721106
PED_LOCATION	PED_LOCATION	98.240313	5721005
SAFETY_EQUIPMENT	SAFETY_EQUIPMENT	52.071253	3032359
EJECTION	EJECTION	48.584815	2829327
POSITION_IN_VEHICLE	POSITION_IN_VEHICLE	48.576968	2828870
EMOTIONAL_STATUS	EMOTIONAL_STATUS	46.917032	2732204
BODILY_INJURY	BODILY_INJURY	46.916294	2732161
COMPLAINT	COMPLAINT	46.916174	2732154
PERSON_AGE	PERSON_AGE	11.171550	650573
PERSON_SEX	PERSON_SEX	11.055778	643831
VEHICLE_ID	VEHICLE_ID	4.148052	241561
PED_ROLE	PED_ROLE	3.346607	194889
PERSON_ID	PERSON_ID	0.000326	19

The column was dropped because it contained over 99% missing values, providing almost no information for analysis. Retaining it would increase noise and reduce model/data quality without adding meaningful insight.