analysis

September 27, 2024

```
[1]: import polars as pl
```

```
[2]: df = pl.read_csv('data/
     ↪Motor_Vehicle_Crashes_-_Vehicle_Information__Three_Year_Window_20240925.csv')
     df.head()
```

[2]: shape: (5, 19)

Year Event Type i64	Case Partial Vehicle ID VIN i64	Vehicle Body Type str	Registrati on Class str	… ---	Contributi ng Factor 2 --- str	Contribut ing Factor 2 Descript… --- str	--- str
2018 3N6CMOKN2 Entered	15232120 FK713460	VAN TRUCK	AGRICULTUR AL TRUCK	…	HUMAN	Not Entered	Not Entered
2018 5J6TF2H55 Entered	15232535 FL001556	SUBURBAN	PASSENGER OR SUBURBAN	…	HUMAN	Unknown	Not Entered
2018 null Entered	15232536	SEDAN	Not Entered	…	HUMAN	Unknown	Not Entered
2018	15232557	4 DOOR	PASSENGER	…	VEHICLE	Other*	Not

```
               1N4AL3AP8
                              SEDAN       OR
Entered        DC229156
                                          SUBURBAN

   2018   15232556    SUBURBAN   PASSENGER   …   HUMAN        Not Appli   Not
   3GYFK2220
                                 OR                          cable
Entered        9G213337
                                 SUBURBAN
```

[4]: `df.shape`

[4]: (3659841, 19)

[5]: `df.tail()`

[5]: shape: (5, 19)

Year Event Type	Case Partial Vehicle ID VIN	Vehicle Body Type	Registration Class	…	Contributing Factor 2	Contributing Factor 2 Descript…	…
--- i64	--- i64	--- str	--- str		--- str	--- str	--- str
2022 Tree, Collision With Fixed Obj…	20508555 JF1VA1L63 J8831039	4 DOOR SEDAN	PASSENGER OR SUBURBAN	…	ENVMT	Pavement Slippery	

Year	Case Vehicle ID	Vehicle Body Type	Registration Class	…			
2022 null Entered	20508572	UNKNOWN VEHICLE	Not Entered	…	HUMAN	Passing or Lane Usage Improper	Not Entered
2022 1N4AA6AP8 Entered	20508559 HC426518	4 DOOR SEDAN	PASSENGER OR SUBURBAN	…	HUMAN	Not Appli cable	Not
2022 Appli cable	20508562 5TDBZRFH2 KS736051	SUBURBAN	PASSENGER OR SUBURBAN	…	HUMAN	Not Appli cable	Not
2022 Appli cable	20508561 1GNERGKW4 NJ192103	SUBURBAN	PASSENGER OR SUBURBAN	…	HUMAN	Unsafe Lane Changing	Not

```
[6]: df.collect_schema()
```

```
[6]: Schema([('Year', Int64),
         ('Case Vehicle ID', Int64),
         ('Vehicle Body Type', String),
         ('Registration Class', String),
         ('Action Prior to Accident', String),
         ('Type / Axles of Truck or Bus', String),
         ('Direction of Travel', String),
         ('Fuel Type', String),
         ('Vehicle Year', Int64),
         ('State of Registration', String),
         ('Number of Occupants', Int64),
         ('Engine Cylinders', Int64),
         ('Vehicle Make', String),
         ('Contributing Factor 1', String),
         ('Contributing Factor 1 Description', String),
         ('Contributing Factor 2', String),
```

```
        ('Contributing Factor 2 Description', String),
        ('Event Type', String),
        ('Partial VIN', String)])
```

[8]: `df.describe()`

[8]: shape: (9, 20)

statistic	Year	Case Vehicle ID	Vehicle Body Type	…	Contributing Factor 2	Contributing Factor 2 Descript…	Event Type	Partial VIN
str	f64	f64	str	---	str	str	str	str
count	3.659841e6	3.659841e6	3659841	…	3659841	3659841	3588870	2936903
null_count	0.0	0.0	0	…	0	0	70971	722938
mean	2019.910059	1.7671e7	null	…	null	null	null	null
std	1.444028	1.3545e6	null	…	null	null	null	null
min	2018.0	1.4378647e7	2 DOOR SEDAN	…	ENVMT	Accelerator Defective	Animal, Collision With	0
25%	2019.0	1.6506852e7	null	…	null	null	null	null

50% null	2020.0 null	1.7682515 e7	null	…	null	null
75% null	2021.0 null	1.8855223 e7	null	…	null	null
max Unknown 81054962	2022.0 yv1ah99x	2.0508759 e7	WELL SERVICING RIG	…	VEHICLE	Windshiel d Inadequ ate

[ ]: