Investigation of eCollisionAnalytics, done on 2024-02-02

#####

Objective - To count how many duplicated CASE\_NBR existed

SQL -

>>>

select

[CASE\_NBR],

count(\*) as DuplicateCount

from [eCollisionAnalytics].[ECRDBA].[COLLISIONS]

group by [CASE\_NBR]

having count(\*) > 1

>>>

Result – 158092 rows

CASE\_NBR DuplicateCount

0000510C 2

0000641C 2

0001143C 2

0001150C 2

0001393C 2

0001771C 2

0002222C 2

0003104C 2

0003264C 2

0003597C 2

0003990C 2

0004620C 2

0004637C 2

0004685C 2

0005614C 2

0005683C 2

#####

Objective – Further drilling down to get the exact number of duplicates by year. For example, to select year “2017”, I can get all the numbers of CASE\_NBR that appears on CASE\_YEAR = 2017, which has at least 1 duplicate in other CASE\_YEAR

SQL -

>>>

with CaseNbrByYear as (

select case\_nbr

from [eCollisionAnalytics].[ECRDBA].[COLLISIONS]

where case\_year = 2015

),

Duplicate as (

select

[CASE\_NBR],

count(\*) as DuplicateCount

from [eCollisionAnalytics].[ECRDBA].[COLLISIONS]

group by [CASE\_NBR]

having count(\*) > 1

)

SELECT

CNBY.case\_nbr

FROM

CaseNbrByYear CNBY

INNER JOIN

Duplicate DUP ON CNBY.case\_nbr = DUP.[CASE\_NBR]

>>>

Result –

2013 -> 1175 rows

2014 -> 1127 rows

2015 -> 967 rows

2016 -> 47015 rows

2017 -> 23982 rows

2018 -> 16698 rows

2019 -> 51660 rows

2020 -> 10301 rows

2021 -> 1904 rows

2022 -> 0 rows

2023 -> 0 rows