**Incidents.csv**

For hit\_run - we have 2 missing values which can be replaced by Yes/No

For route\_type - we have 12995 missing values which can be replaced by the mode of 9 different category

For mile\_point - we have 12338 missing values can be filled with mean/median

For mile\_point\_direction - we have 12991 missing values and 1 unkown value which can be replaced by Yes/No

For lane\_direction - we have 12327 missing values and 750 unknown which can be replaced by the mode of 9 different category

For lane\_type - we have 87945 missing values can be filled or kept as is where acrs failed to track the lane type

For direction - we have 12338 missing values and 37 unknown which can be replaced by 4 other direction- mode

For distance - we have 10475 missing values which can be replaced by mean/median

For distance\_unit - we have 12328 missing values and 458 Unknown values can be filled with mode

For road\_grade - we have 12967 missing values and 152 unkown value which can be replaced by mode

For road\_name - we have 12328 missing values which can be replaced by the mode

For cross\_street\_type - we have 13016 missing values and Unknown -7276 can be filled or kept as is where acrs failed to track the lane type

cross\_street\_name - 12338 Missing - mode/drop

off\_road\_description- 85132 Missing

municipality- 86967 Missing

related\_non\_motorist- 92077 Missing

collision\_type- 13263 Other, 717 Unknown, 504 Missing

weather- 7947 Missing, 648 Unknown, other 214

surface\_condition- 14650 Missing, 407 Unknown, 99 other

light- Missing 815, UNKNOWN 698, OTHER 221

traffic\_control- Missing 16436, OTHER 1143, UNKNOWN 234

driver\_substance\_abuse- Missing 15598,UNKNOWN 5115, N/A, UNKNOWN 1154, OTHER 35 (need attention) ---Aasim's file

non\_motorist\_substance\_abuse- Missing 93195, UNKNOWN 222

first\_harmful\_event- Missing 650, OTHER 339, UNKNOWN 192

second\_harmful\_event- Missing 72948, OTHER 155, UNKNOWN 129

fixed\_object\_struck- Missing 76863, UNKNOWN 235

junction- Missing 26099, UNKNOWN 79

intersection\_type- Missing 49054, UNKNOWN 97

intersection\_area- Missing 71045, OTHER 1150, UNKNOWN 79

road\_alignment- Missing 12749, UNKNOWN 131

road\_condition- Missing 15705, OTHER 338,UNKNOWN 204

road\_division- Missing  13531,OTHER 1167,UNKNOWN 85

**Deciding on before loading and after loading transformation and analysis of Incidents.csv:**

*Before Loading transformations:*

*Mainly missing values if are less then 20K can be transformed here in any case if it's above 20K we keep values as is because it might lead to false analysis/interpretation*

*After Loading transformations:*

*import into frame and perform analysis directly*

**Importance of each parameter in taking precautions for future incidents:**

*Very High: (Mainly missing values if are less then 20K can be transformed here in any case if it's above 20K we keep values as is because it might lead to false analysis/interpretation)*

**Collision Type- Understanding the types of collisions occurring can inform safety measures and infrastructure improvements.**

**Weather- Weather conditions greatly affect road safety, influencing visibility, road surface conditions, and driver behavior.**

**Road Condition- Road condition impacts vehicle handling and braking distance, affecting the likelihood and severity of accidents.**

**Traffic Control- Traffic control measures (e.g., signals, signs) play a crucial role in regulating traffic flow and reducing accident risks.**

*High: (Mainly missing values if are less then 20K can be transformed here in any case if it's above 20K we keep values as is because it might lead to false analysis/interpretation)*

**Lighting- Adequate lighting enhances visibility and reduces the likelihood of accidents, especially at night.**

**Junction- Understanding junction characteristics can help identify high-risk areas for accidents and implement appropriate safety measures**

**Intersection Type- Different intersection types pose unique safety challenges, and understanding their characteristics is essential for targeted interventions.**

*Moderate to High: (Mainly missing values if are less then 20K can be transformed here in any case if it's above 20K we keep values as is because it might lead to false analysis/interpretation)*

**Road Grade- Road gradient affects vehicle stability and driver behavior, particularly on steep inclines or declines.**

**Direction- Knowledge of prevailing travel directions can inform signage placement and road design improvements.**

*Moderate: (Mainly missing values if are less then 20K can be transformed here in any case if it's above 20K we keep values as is because it might lead to false analysis/interpretation)*

**Mile Point- Mile point information can help identify accident-prone areas along roadways for targeted safety interventions**

**Distance- Distance-related data can provide insights into accident patterns and help assess the effectiveness of safety measures.**

**Lane Type- Understanding lane configurations aids in identifying potential hazards and optimizing road design for safety.**

**Lane Direction- Lane direction information assists in analyzing traffic flow patterns and implementing appropriate lane management strategies**

**Route Type- Route type classification helps prioritize safety improvements on different road types based on usage and risk.**

*Low:*

**Cross-Street Name**

**Latitude/Longitude/Location**

**Off-Road Description, Municipality, Related Non-Motorist, First/Second Harmful Event**

hit\_run:

Replace missing values with "Unknown" if it's possible to distinguish between unknown and Yes/No. Otherwise, replace missing values with the mode ("Yes" or "No").

route\_type:

Replace missing values with the mode of the 9 different categories.

mile\_point:

Fill missing values with the mean or median.

mile\_point\_direction:

Replace missing values and the "unknown" value with "Yes" or "No", depending on the context.

lane\_direction:

Replace missing values with the mode of the 9 different categories.

lane\_type:

Decide whether to fill missing values or keep them as is if ACRS failed to track the lane type.

direction:

Replace missing values and "unknown" values with the mode of the 4 other directions.

distance:

Fill missing values with the mean or median.

distance\_unit:

Replace missing values and "Unknown" values with the mode.

road\_grade:

Replace missing values and "unknown" values with the mode.

road\_name:

Replace missing values with the mode.

cross\_street\_type:

Decide whether to fill missing values or keep them as is where ACRS failed to track the lane type.

cross\_street\_name:

Decide whether to fill missing values with the mode or drop them.

off\_road\_description, municipality, related\_non\_motorist:

These columns have a high number of missing values. Decide whether to drop these columns or fill missing values based on further analysis.

collision\_type, weather, surface\_condition, light, traffic\_control, driver\_substance\_abuse, non\_motorist\_substance\_abuse, first\_harmful\_event, second\_harmful\_event, fixed\_object\_struck, junction, intersection\_type, intersection\_area, road\_alignment, road\_condition, road\_division:

Replace missing values and "unknown" values with the mode or decide whether to drop or fill missing values based on further analysis.