**Handling Null Values and Unnecessary Values**

**Null Values**

Handling null values is crucial to ensure data integrity and accuracy in the data warehouse. The following strategies can be used to handle null values:

**1- Imputation:**

* **Mean/Median Imputation**: Replace null values in numerical columns with the mean or median value of the column.
* **Mode Imputation**: Replace null values in categorical columns with the most frequent value (mode).
* **Default Values**: Assign default values to null entries based on business logic. For example, replace null values in ‘*SpeedLimit’* with a default speed limit.

**2- Flagging:**

* Add a new column to indicate whether the original value was null. This is useful for keeping track of imputed values.

**3- Removal**:

* If null values are present in non-critical fields or if the percentage of null values is very high, consider removing those rows or columns.
* Rows with null values in critical fields (e.g., primary keys or foreign keys) should be investigated and either corrected or removed.

**4- Forward or Backward Fill**:

* Use forward or backward filling techniques to propagate non-null values to neighboring null values, especially in time series data.

**Unnecessary Values**

Unnecessary values can include duplicates, outliers, and values that do not contribute to the analysis. Here's how to handle them:

**1- Duplicate Removal**:

* Remove duplicate rows based on unique identifiers (e.g., ‘*ReportNumber*’).

**2- Outlier Detection and Treatment**:

* Identify outliers using statistical methods (e.g. IQR).
* Treat outliers by capping or flooring extreme values, or by removing them if they are deemed erroneous and non-representative.

**3- Data Transformation**:

* Normalize or standardize numerical values to ensure consistent scale.
* Map categorical values to a consistent format (e.g., converting all vehicle makes to uppercase).

**4- Column Filtering:**

* Remove columns that do not provide meaningful information or are irrelevant to the analysis. For example, if ‘*OffRoadDescription*’ is always null or contains irrelevant information, it can be dropped.

**Analytics for Vehicle Crash Accidents**

**Crash Frequency Analysis**:

* Analyze the frequency of crashes over time.

**Vehicle Involvement Analysis**:

* Analyze the types of vehicles involved in crashes.

**Location-Based Crash Analysis**:

* Identify crash hotspots by location.

**Crash Severity Analysis**:

* Analyze crashes by severity and contributing factors.

**Temporal Analysis**:

* Identify patterns in crash occurrences by time of day, day of the week, and month.

**Driverless and Parked Vehicle Analysis**:

* Analyze the involvement of driverless and parked vehicles in crashes.

**Another analysis with methodology describing**

**Weather Condition Analysis:**

**Objective:** To analyze the impact of weather conditions on crash frequency.

**Methodology:**

* Aggregate crash counts by different weather conditions.
* Determine which weather conditions are associated with higher crash frequencies.

**Driver Demographics Analysis**

**Objective:** To analyze the demographics of drivers involved in crashes to identify any high-risk groups.

**Methodology:**

* Aggregate crash counts by driver age, gender, and experience level.
* Identify demographic groups with higher crash involvement.

**Road Condition Analysis**

**Objective:** To analyze the impact of road conditions on crash frequency.

**Methodology:**

* Aggregate crash counts by different road conditions.
* Determine which road conditions are associated with higher crash frequencies.

**Light Condition Analysis**

**Objective:** To analyze the impact of light conditions on crash frequency.

**Methodology:**

* Aggregate crash counts by different light conditions (daylight, dawn, dusk, night with/without streetlights).
* Determine which light conditions are associated with higher crash frequencies.

**Crash Cause Analysis**

**Objective:** To analyze the primary causes of crashes to identify common factors.

**Methodology:**

* Aggregate crash counts by primary cause (e.g., speeding, distracted driving, DUI, etc.).
* Determine the most common causes of crashes.

**Time to Incident Response Analysis**

**Objective:** To analyze the time taken for emergency services to respond to crash incidents.

**Methodology:**

* Calculate the time difference between crash occurrence and emergency response arrival.
* Aggregate response times and analyze their distribution.

**Repeat Offender Analysis**

**Objective:** To analyze patterns of repeat offenders involved in multiple crashes.

**Methodology:**

* Identify drivers involved in multiple crash incidents.
* Aggregate crash counts by driver ID or license number.

**Seatbelt Usage Analysis**

**Objective:** To analyze the impact of seatbelt usage on crash severity and outcomes.

**Methodology:**

* Aggregate crash counts by seatbelt usage status.
* Compare crash severity levels for cases with and without seatbelt usage.

**Adding Constraints and Analytics to the Business Model**

1. **Data Integrity**: Ensure the data warehouse maintains high data integrity through the implementation of primary key, foreign key, not null, unique, and check constraints.
2. **Analytical Capabilities**: Provide comprehensive analytics to uncover insights about crash frequency, vehicle involvement, location hotspots, crash severity, temporal patterns, and specific factors like driverless and parked vehicles.
3. **Reporting**: Develop reporting solutions that leverage these analytics to generate actionable insights for stakeholders, such as traffic safety agencies, policymakers, and researchers.