**Clarifications/Questions for moving the Solution to Discover to Define phase.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Question/Area of Clarification** | **Our Understanding** | **Specific Questions** | **Dependency** |
| **1. Precise Definition and Thresholds of "Near Miss" Events** | trigger\_source\_event types (e.g., TRIG\_AEB\_WARNING\_MSG) define near-misses. | We need the *specific quantitative thresholds* for each trigger\_source\_event type.  For example, for TRIG\_AEB\_WARNING\_MSG, is it triggered when AEB\_TIME\_TO\_COLLISION drops below a certain threshold (e.g., 2 seconds), or when AEB\_DEC\_CMD exceeds a specific negative G-force value (e.g., -0.5G)?  We see all the readings for these Signals (Does not seem to have a threshold) | Cannot start designing the Data Model until we get this info. |
| **2. Exact Event Window Duration and Trigger Point** | seq represents 100-millisecond timestamps within an event. | Please confirm the *exact, consistent duration* of the recorded window for each log\_data\_id event (e.g., always 30 seconds, or variable between 10-30 seconds).  Also, clarify if trigger\_timestamp\_IST marks the *start, middle (peak), or end* of this recorded window relative to the actual near-miss event's peak. | Needed for designing measures. Not a blocker, but will affect report accuracy |
| **3. Availability of Additional Contextual Data** | Integrating DTC, Usage Info, KYC Details will add more value | Are there other contextual data points available? Examples: road conditions (wet/dry, type of road), weather, time of day/night, driver identification (anonymized), vehicle configuration/model, or outcome of ADAS interventions (e.g., AEB activation vs. warning). | Nice to have we can consider adding this later as well. |
| **4. Implications of High Data Volume (130 Million Rows/Day)** | We observe a very high volume of data (approx. 130 million rows/day for 15,000 VINs), potentially due to congested driving conditions/behaviors in India. | Could you provide insights into the expected frequency of these events and confirm if this volume is within anticipated ranges? This is crucial for analysis expectations and potentially refining event definitions if current thresholds are too sensitive.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | day | row\_count | unique\_vin\_count | unique\_event\_count | Rows/VIN | | 11-06-2025 | 4,42,78,085.00 | 238 | 9 | 1,86,042.37 | | 12-06-2025 | 61,94,23,087.00 | 2290 | 9 | 2,70,490.43 | | 13-06-2025 | 17,07,33,107.00 | 4024 | 10 | 42,428.70 | | 14-06-2025 | 16,42,21,626.00 | 3983 | 9 | 41,230.64 | | 15-06-2025 | 13,81,35,568.00 | 3434 | 9 | 40,225.85 | | 16-06-2025 | 15,12,98,707.00 | 3838 | 9 | 39,421.24 | | 17-06-2025 | 9,23,24,237.00 | 2895 | 9 | 31,890.93 |   In addition, there are 1075 Signals for each event, which is bloating the Data load. Our sample study sees zero values for many; can we recheck this requirement. Which are the critical signals for Analysis among these.?   |  |  | | --- | --- | | trigger\_source\_event | signal\_count | | TRIG\_ELK\_SIDE\_WARNING | 1075 | | TRIG\_STS\_EMERGENCY\_BRK | 1075 | | TRIG\_ELK\_OVERTAKING\_WARNING | 1075 | | TRIG\_ELK\_ONCOMING\_WARNING | 1075 | | TRIG\_STS\_CRASH | 447 | | TRIG\_AES\_WARNING | 1075 | | TRIG\_AEB\_WARNING\_MSG | 1075 | | TRIG\_ACC\_SYS\_INFO | 1075 | | TRIG\_LCF\_SYS\_STATE | 1075 | | TRIG\_SPA\_LCF\_HANDSOFF\_WARNING | 1075 | | Critical:  We see the data only for one week starting from June 11-17 only. Latest Data is not available,  Even then the data seems too heavy. (2 Lakh/Vehicle/Day)  Need to check the Trigger Conditions.  Report Usability will be affected if every Vehicle is sending 1000s of Alerts in a day, |
| **5. Series Event Log Data** | Table used “Drivelog\_Silver”. Catalogue: prod\_msdvc\_bev\_mmc\_postsop | Reconfirm the Table. Request an Audit of this Table by MEAL/VES Team | Source Table correctness, related to point 4. |