# Feature Selection Recommendations for Traffic Accident Prediction

This document provides recommendations for feature selection based on business logic and domain knowledge to predict traffic accidents at specific locations.

## Recommended Features to Retain

### 1. Spatial Features

- Latitude and Longitude: Geographic coordinates are critical for identifying high-risk locations (e.g., intersections, curves, or areas with heavy traffic). You can cluster these coordinates to define risk zones for better modeling.

### 2. Temporal Features

- Time: Time of day is directly correlated with accident likelihood, such as during rush hours (e.g., 7–9 AM, 5–7 PM).  
- Day of Week: Weekdays and weekends have distinct traffic patterns; weekends might involve riskier behaviors like drunk driving.  
- Month and Date in Month: Seasonal or holiday patterns can affect accident rates (e.g., icy roads in winter).

### 3. Weather and Lighting Conditions

- Weather: Conditions like rain or snow increase accident risks due to reduced visibility and slippery roads.  
- Light: Poor lighting (e.g., nighttime or dawn) is associated with higher accident rates.

### 4. Road and Traffic Control

- Road Name and Cross-Street Name: Certain roads or intersections might have inherently higher risks due to design or traffic volume.  
- Traffic Control: The presence or absence of traffic lights, stop signs or other controls can significantly impact accident likelihood.

### 5. Vehicle and Driver-Related Factors

- Driver Substance Abuse: Drunk or impaired driving is a leading cause of accidents and should be included if available.  
- Vehicle Movement: The movement type (e.g., turning, stopped in traffic) can help identify patterns in specific accident types.  
- Speed Limit: Higher speed limits are associated with more severe accidents.

### 6. Event-Related Information

- Collision Type: Understanding the type of collision (e.g., rear-end, side-impact) may help in refining predictions.

## Features to Consider Dropping

- Weekend: The weekend indicator is redundant if the Day of Week is already included.  
- Vehicle Make and Vehicle Year: These may have a limited effect in predicting accidents at a location.  
- Surface Condition: If weather information (e.g., rain, snow) is already present, this might not add much value.  
- Driver At Fault: This is a post-accident outcome and may not be useful for predictive purposes.

## Suggestions for New Features

- High-Risk Road Tags: Use historical data to tag specific roads or intersections as high-risk zones.  
- Peak Time Indicators: Create a binary feature indicating peak traffic hours.  
- Weather Impact Factors: Add flags for extreme weather events like heavy rain or snow.

## Final Recommended Feature Set

Based on the analysis and domain knowledge, the following features are recommended for retention:  
**- Latitude**  
**- Longitude**  
**- Time**  
- Day of Week  
**- Month**  
- Weather  
- Lights  
**- Road Name**  
- Traffic Control  
- Speed Limit  
- Driver Substance Abuse  
- Collision Type