Report on Auto Theft Data Analysis

Objective

The objective of this analysis is to guide the Toronto Police Services on how best to allocate resources to manage auto theft in the Toronto region. This involves examining the dataset for biases, potential errors, and patterns that can inform resource allocation strategies.

Data Cleaning and Preparation

1. **Handling Old Data**
    - **Deleted Rows**: Rows where `OCC_DATE` is 2000 and 2001 were removed from the dataset. These dates were considered too old and likely less relevant for current analysis. Additionally, specific rows (1994, 3081, 17362, 56052) were deleted due to the same reason. These rows were identified as potentially erroneous or outdated data.
2. **Handling Missing Geographic Data**
    - **Rows with Zero Coordinates**: A total of 717 rows contained zeros in both `LONG_WGS84` and `LAT_WGS84` columns. These values are geographically nonsensical for Toronto. Given the high number of such rows, they have been retained in the dataset but flagged as containing missing or potentially erroneous geographic data. They have not been deleted to preserve the dataset's completeness. It is noted that these rows show "NSA" (Not Specified/Available) in other columns, which reinforces their questionable reliability.

Data Analysis

1. **Temporal Analysis**
    - **Day of the Week (REPORT_DOW)**:
        - A pivot table was created with `REPORT_DOW` in Rows and `EVENT_UNIQUE_ID` in Values. The analysis revealed that the number of reports is significantly lower on Sundays compared to other days of the week. This suggests that auto theft incidents might be less reported or occur less frequently on Sundays.
    - **Time of Day (REPORT_HOUR)**:
        - Another pivot table was created with `REPORT_HOUR` in Rows and `EVENT_UNIQUE_ID` in Values. The analysis indicated that the number of reports peaks in the early morning hours, specifically at 6, 7, and 8 AM. This pattern could suggest that auto theft incidents are more frequently reported or occur during these early morning hours.
2. **Geographic Analysis**
    - **Neighborhood Analysis (NEIGHBOURHOOD_158)**:
        - A pivot table was created to examine the number of thefts across different neighborhoods. It was observed that the neighborhood of West Humber-Clairville has a disproportionately high number of reports compared to

other neighborhoods. This indicates a higher incidence of auto theft in this area, which may require increased resources or targeted intervention.

Conclusion and Recommendations

- **Data Integrity**: The dataset has been cleaned to remove outdated data and to flag rows with missing geographic information. These rows have not been deleted to avoid loss of valuable data but should be considered carefully in any further analysis.
- **Resource Allocation**: Based on the analysis:
    - **Temporal Resource Allocation**: Consider focusing resources on early morning hours (6-8 AM) and reviewing reporting patterns on Sundays.
    - **Geographic Resource Allocation**: Allocate additional resources to the West Humber-Clairville neighborhood due to its high number of reported auto theft incidents.

Work to do

- **Further Investigation**: Review flagged rows with zero geographic coordinates for potential data improvements or follow-up.
- **Ongoing Monitoring**: Continuously monitor and update resource allocation strategies based on emerging patterns and updated data.