# Predictive Analytics Telematics Exercise

**INSTRUCTIONS:** Please review the following document in its entirety. Please use any statistical program of your choice to address the questions below. Additionally, prepare a presentation that addresses each of the questions below. PowerPoint, R Markdown (presentation or report format), Shiny, and Beamer (LaTeX) are all acceptable presentation formats. All code and related output will be reviewed by the interview panel. Be prepared to justify any assumptions that you make and present all supporting evidence of your analysis (e.g. tables, charts).

## Part 1 (GPS Data)

**Description of Data Set**

You have been provided a data set with GPS data (`sample_trips.csv`). The data set contains 9687 rows and 4 variables, described as follows.

| Variable | Description |
|---|---|
| `trip_id` | Trip identifier |
| `local_dtm` | Timestamp |
| `latitude` | latitude |
| `longitude` | longitude |

**Tasks to Be Completed**

**Analysis:** Answer the following research questions to the best of your ability. Be prepared to walk the interview panel through your code and your thought process.

(1) What steps did you take to clean the data?
(2) Set a threshold for hard events (braking, accelerating, etc) based on the data. How did you set the thresholds?
(3) Give a summary of Hard Events, Idling and distance travelled for each trip.

## Part 2 (Modeling)

**Description of Data Set**

You have been provided with a hypothetical data set (`simulated_summary_total.csv`) of 30,000 vehicles on which 1 Hz. telematics data has been collected. The data has been summarized to one row per vehicle. The dataset contains the following columns: .

| Variable | Description |
|---|---|
| `Vehicle` | An integer value used to delineate each vehicle. |
| `Days` | The number of days for which data was collected. |
| `Distance` | The number of miles a vehicle was driven during the experience period. |
| `HardBrakes` | The number of hard braking events detected in the 1 Hz. data. |
| `HardAccelerations` | The number of hard acceleration events detected in the 1 Hz. data. |
| `NightTime_Pct` | The percentage of total miles driven at night. |
| `VehicleType` | The type of vehicle. |
| `Loss_In` | An indicator that the vehicle has been in a collision. (1 indicates a collision, 0 otherwise) |

# Predictive Analytics Telematics Exercise

**Tasks to Be Completed**

**Analysis:** Answer the following research questions to the best of your ability. Be prepared to walk the interview panel through your code and your thought process.

(4) Is there a statistically significant difference between vehicle types?
(5) Are hard brakes and hard accelerations equally important in predicting risk?

**Model Building:** As an insurer, it is important to know how to segment your business based on the likelihood of having a collision. Utilizing Loss_In as the response variable, build a model that segments the vehicles based on the likelihood that they will have a collision. (Note: across an entire insurer's book of business a collision is a rare occurrence.) Use any model specification and related data transformations you deem to be most appropriate for this analysis. Prepare a clear and easily understandable interpretation of the results.

**Data Set Enhancement:** Suppose you've been asked to enhance the existing data set to enable additional rounds of analysis and facilitate additional model development.

(1) State at least three research questions you would like to address and describe your thought process behind how you formulated these research questions.
(2) Prepare a list of 5-7 additional attributes you would like to add to the data set. Prepare a brief explanation for each attribute.
(3) Estimate and justify the appropriate sample size (and sampling technique, if desired) that would be required to address the research questions you defined.



Nationwide
is on your side