Analysis of iRacing Race Summary and Lap Data

Wereratgames.com, August 2018

Executive Summary

This document presents initial, exploratory analysis of three seasons of Pro Mazda official racing in order to ascertain causes of incidents (principally off-track driving, collisions with objects, and collisions with other iRacers) that influence drivers' Safety Rating (SR). The analysis comprises: 3 seasons, 36 weeks, 2,849 sessions with 3,418 discrete races (sub-sessions), 4,181 iRacers, 43.68K entries, and just over 749,000 individual laps. While this sample is for a single series (the C-license, open-wheel Pro Mazda series) and therefore not necessarily representative of other types of road racing nor to oval racing, the series' availability to most iRacers makes it a useful object of study.

What wasn't studied:

The sample does include driver identifiers, these were not used to form conclusions based on identity. Driver identifiers were used in examining whether, in an aggregate sense, patterns of abnormal incident count or type were present, and in forming other medians or aggregates.

Additionally, a small number of unofficial sessions are present in the sample data; these were removed to focus on official races only.. In subsequent machine learning stages, records for unofficial sessions will also be removed (for instance, it would not be helpful to learn that a good way of retaining SR is to participate in unofficial races.

What was studied:

The sample includes these noteworthy items:

For the sessions themselves:

- The race season
- The race week
- The course and its identifier
- Session and subsession IDs
- The race date and time

For the entries:

- Car numbers (which represent at-start ranking of drivers iRating)
- Customer IDs
- Start and Finish Position
- Drivers' license levels (identified by color)
- Drivers' total incident points
- Drivers' fastest lap number and time, and average lap time
- Race season points earned; aggregated across a session, these correlate to the strength of field (SOF)

For the laps:

- Session and customer information key to correlating to entry and session data

- Lap number
- Lap time
- An indicator field which was divided into Black Flag, Car Contact, Other Contact, Lost Control, Off Track, Pitted, Reset, and Towed indicators; others exist but were not considered for this analysis.

The data was encapsulated in JSON format. This was parsed to reveal the above information. Additionally, information about the difference between a driver's at-start ranking and their qualifying (start) position, and between their start and finish position, proved informative. Lastly, aggregation was performed to derive counts of each type of incident by driver/race combination.

After this initial work, the author performed a variety of exploratory analyses and visualization which led to the following conclusions:

- **Race Week** – The author had anecdotally noted that incident counts appeared to decrease with race week (as drivers gain proficiency with a car, and familiarity with fellow drivers, racing becomes 'cleaner.') The data in this sample do not support this hypothesis. Again, samples from other series (notably multi-class series) may differ.

- **Course** – The designs of some courses naturally provide racing opportunities which provide more likelihood for incidents. In the sample, the *seven* courses with the highest occurrence of car contact employ long straights followed by heavy braking zones as a feature.

- **Race Lap** – Perhaps unsurprisingly, incidents decline as a race proceeds. Car contacts fall off dramatically, while off-tracks and spins trail off. This information might be used to 'tune' the way incidents are recorded.

- **Qualifying** – This was not a surprise feature of the data. Those drivers who qualified among the top five in a race enjoyed a significantly lower likelihood of a car contact in their race relative to other drivers.

- **Qualifying** – Conversely, qualifying position did produce a likely predictor of incidents. Drivers who qualified at or near their expected position (their car number was close to their starting position) had an average likelihood of car contact, but drivers who qualified five or more places better than expected had a significantly higher likelihood of car contact.

- **Atypical Drivers** – Analysis of off track and car contact totals by driver/race reveals that the vast majority of drivers have a low frequency of car contact (most commonly zero) and off track incidents (most commonly two or less), but a minority of drivers repeatedly are involved in one or both of these incidents.

Analysis of iRacing Race Summary and Lap Data

Wereratgames.com, August 2018

Initial Data Exploration

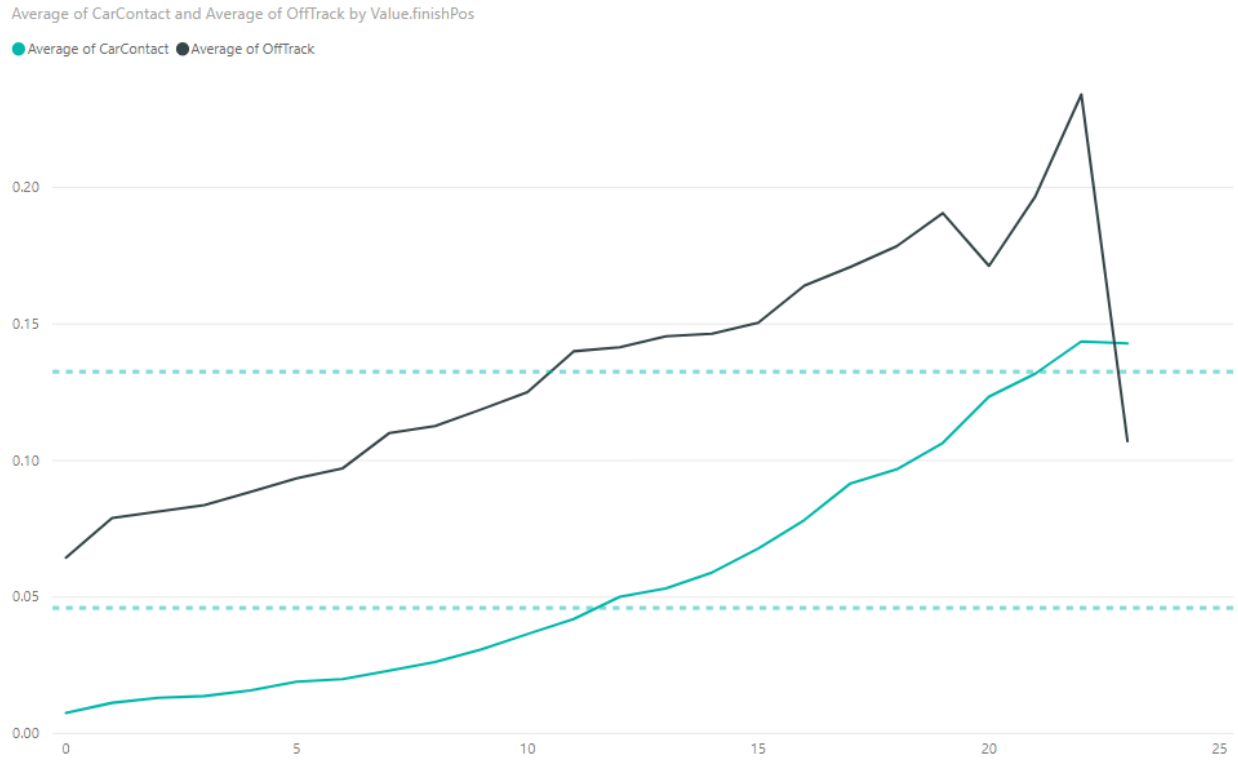
As previously noted, the data contained information about unofficial races (those races without a sufficient driver group). Because these results don't count for SR or iRating, drivers may have behaved differently, if they participated at all, than they would have for official races. Accordingly, this data was eliminated from the analysis.

Secondly, session, grid, and lap information was joined by the subsession ID (each individual race) and driver information was linked to laps by driver (customer) ID. The laps information contains a single flag that indicates whether the driver pitted, went off track, incurred a black flag penalty, reset, contacted an object, contacted another car, lost control, incurred a connection discontinuity, an interpolated crossing, or a clock smash, or towed. These were separated sans the connection items, and aggregation performed by each for each driver/subsession to get a count of off-tracks, collisions, etc. per race.

Note that the flags indicate whether the driver incurred each incident in a lap; they do not total how many. If a driver incurred four off-tracks in one lap, the off-track flag would be set, but only once. Thus, the total incident count for a driver/race is independently of interest as it does include these totals.

Sanity Checks and Basic Statistics

One would naturally expect that drivers who avoid incidents would finish better than those involved in off-track or collisions – 'To finish first, first one must finish.' Accordingly, this visualization shows these maxims to be true; if the data suggested otherwise, one might suspect the data of being poorly formed!

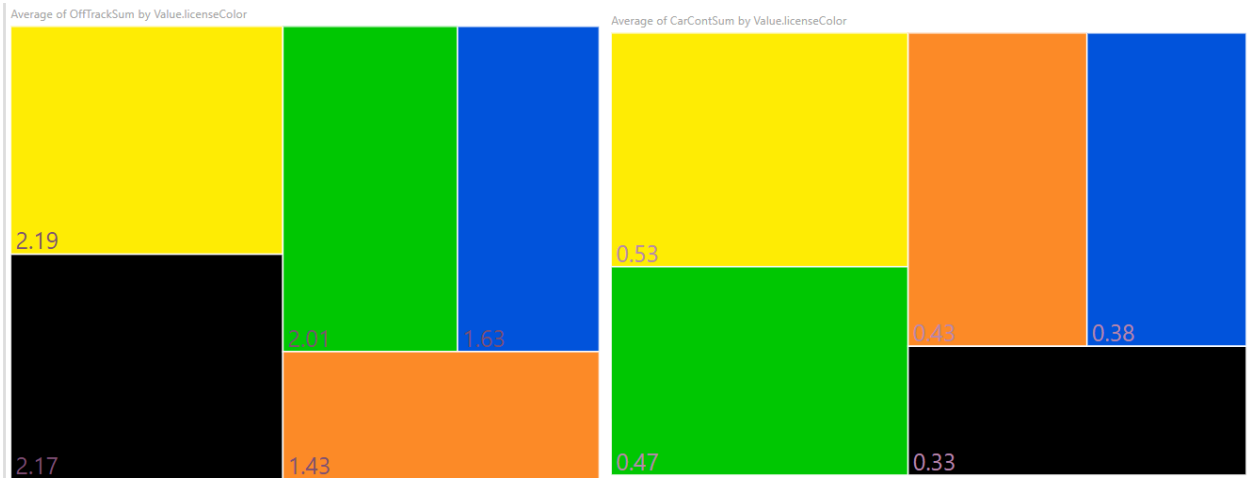


Ex. 1: No surprise – doing well requires not crashing!

Note: the '0' position is first in the data set. The dashed lines represent median results. The sharp decrease in likelihood of an offtrack at the further edge appears to be an anomaly of the low volume of races with 24 racers.

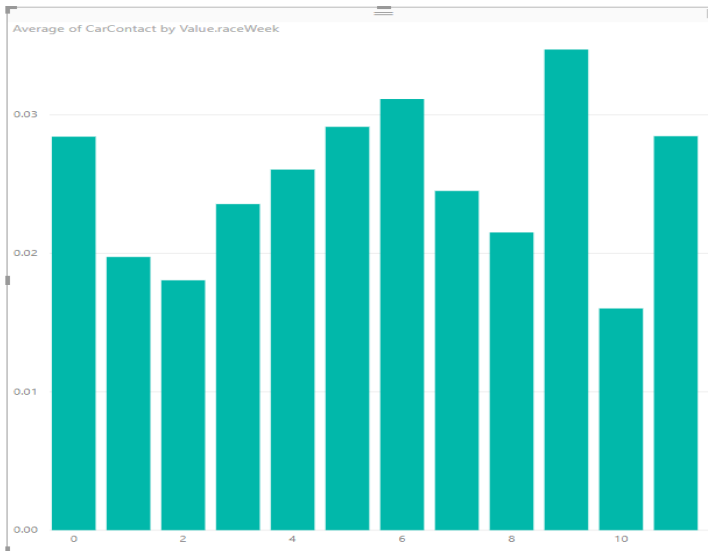
Similarly, since the whole point of a Safety Rating and license color is to represent continued avoidance of incidents and safe driving, one would expect "A" drivers to be safer than "B" and so forth.

This appears to be partly true; Pro, "A," and "B" drivers were less likely to have car contact, but "C" and "Pro" drivers *more* chance of an off-track incident per race, although neither was particularly significant in the statistical sense. "D" drivers in this series were fewer in number and less prone to off-tracks, possibly because their participation in the series was the means to advancing to a "C" license or beyond.



Ex. 2-3: Incidence of car contact and off-track incidents met expectations.

Lastly, as noted in the summary, the author anticipated a negative correlation between race week and incidents. As it turned out, the track was a better predictor of trouble than iRacers' experience in the season.



Ex. 4: Car contacts varied throughout the weeks in the three sample seasons.

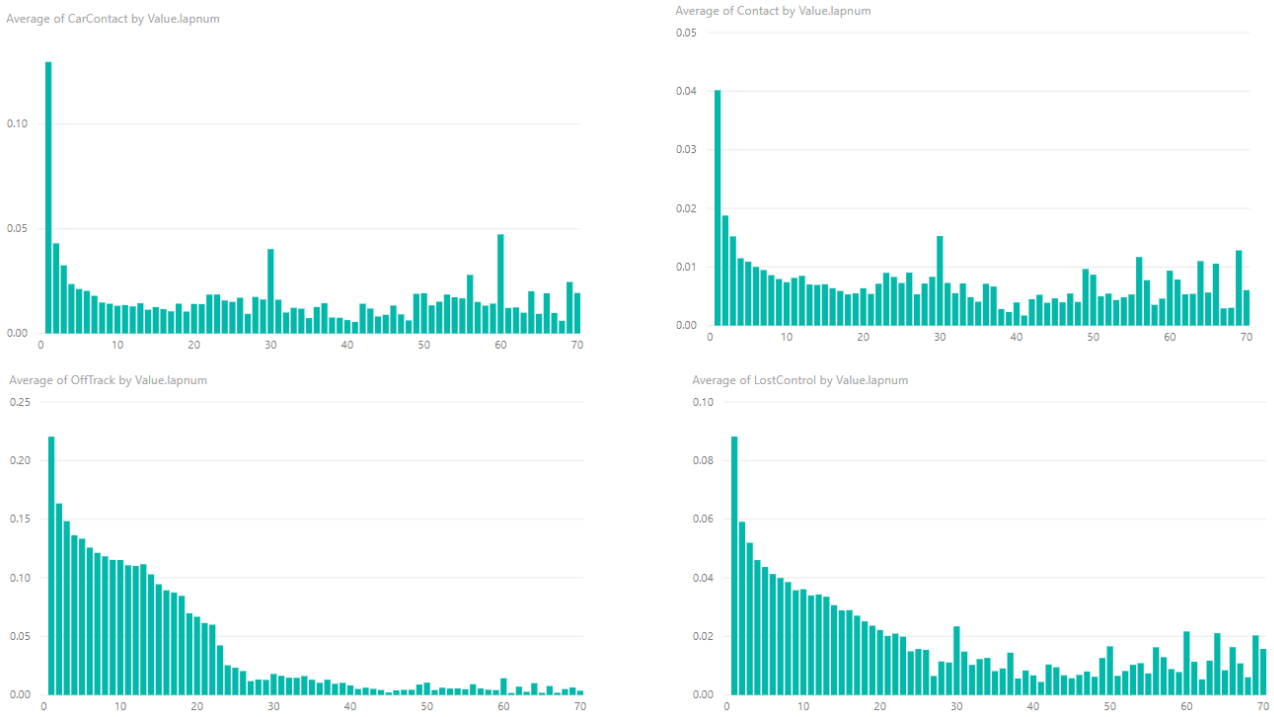
Ex. 5-6: Track geometry makes a difference – the "Green Hell" can be!

Potentially Valuable Predictors

None of the above should particularly surprise iRacers, although the author had anticipated some improvement in incident frequency as a season progressed. Neither did some of the meaningful findings, although seeing them backed by three-quarters of a million laps lends some surety.

Argy-Bargy

The first 'obvious' finding is that the first laps are more dangerous. What's useful is the rate at which that risk decays.

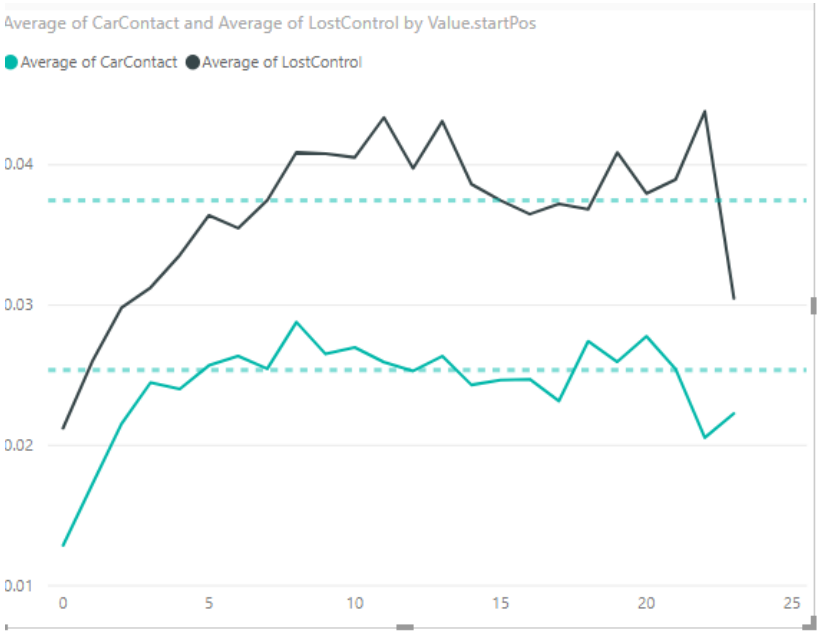


Ex. 7: Different types of incidents become less likely at different rates.

The above clearly indicate that after lap 1, the likelihood of a car contact declines almost immediately to a norm. Drivers 'push' hard and may shunt throughout early laps and, as seen by various spikes, at the last lap of a race. Off-tracks and spins take significant time to reach a low steady state at which point racing has settled into circulating the track and waiting for others' mistakes.

Clean Air

Another no-surprise is that iRacers who start in the front are less likely to have a spin or car collision; they have less traffic and so fewer unexpected events to react to. Drivers are especially wary of turning the polesitter, or of being the polesitter and losing the race through error.

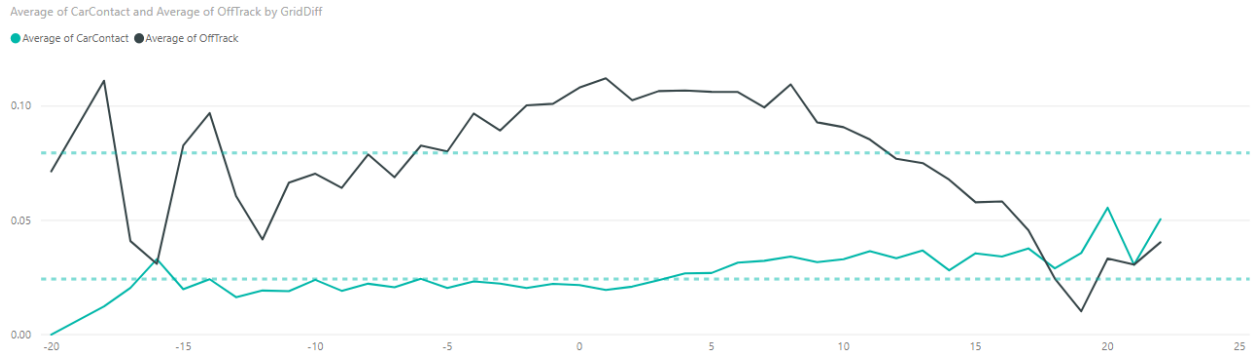


Ex. 8: Qualifying matters

As in prior examples, the dashed line represents a median. The variation on the right side is, again, due to fewer samples.

Hard-Chargers cause Over-Qualifiers

In order to arrive at one of the more potentially meaningful predictors of trouble, it was necessary to create features based on the difference between an iRacer's car number and start position. The car number is the ranking of that driver's iRating at the time of entering the race. In a static world, car #1 would qualify first, car #2, second, and so on. When a qualifying error annuls an iRacer's lap, that driver may start significantly down the order, elevating other drivers in the grid and creating at best a dynamic race, and at worst a host of 'divebombs' and disaster. Interestingly, the correlation is not between the lower-placed drivers, but with those that qualified five or more places ahead of their iRating-based car number:

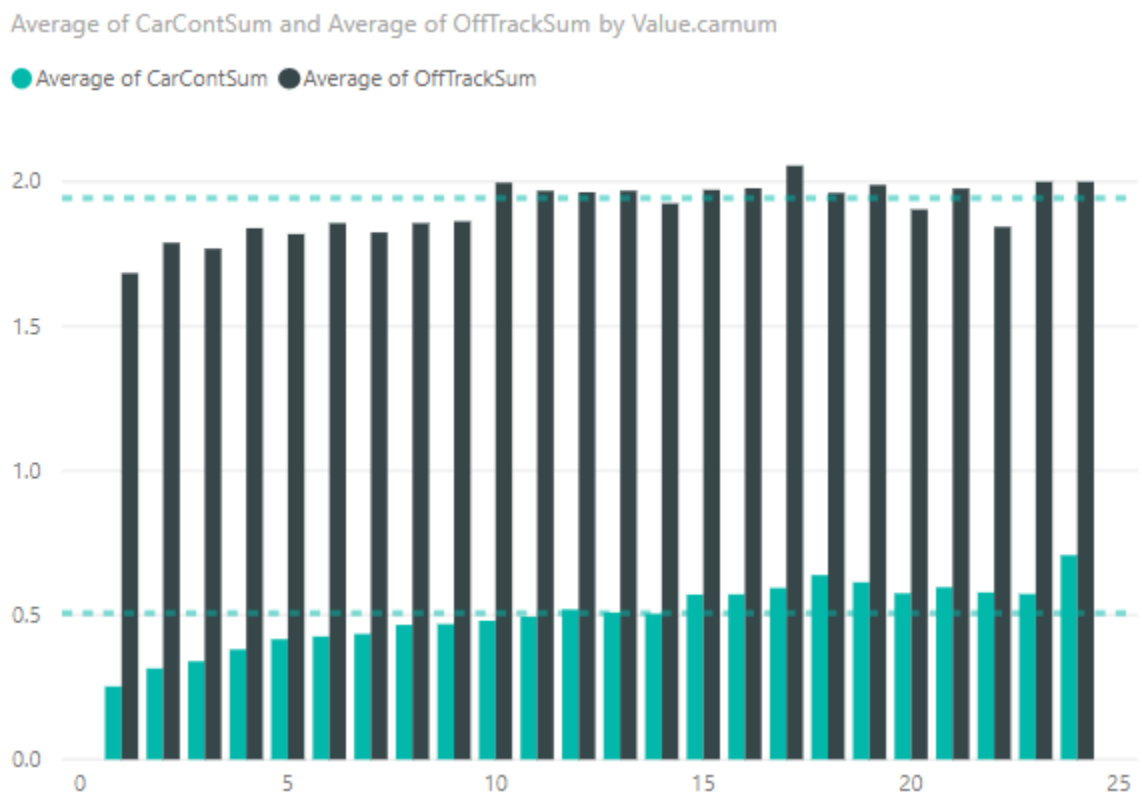


Ex. 9: Off-tracks correlate with competition, but being further in front causes crashes

Further investigation may be warranted to decipher this finding and determine if the risky qualifying position is caused by a stellar "hot-lap" or more frequently by fast drivers making qualifying mistakes who then put pressure on those ahead during the race.

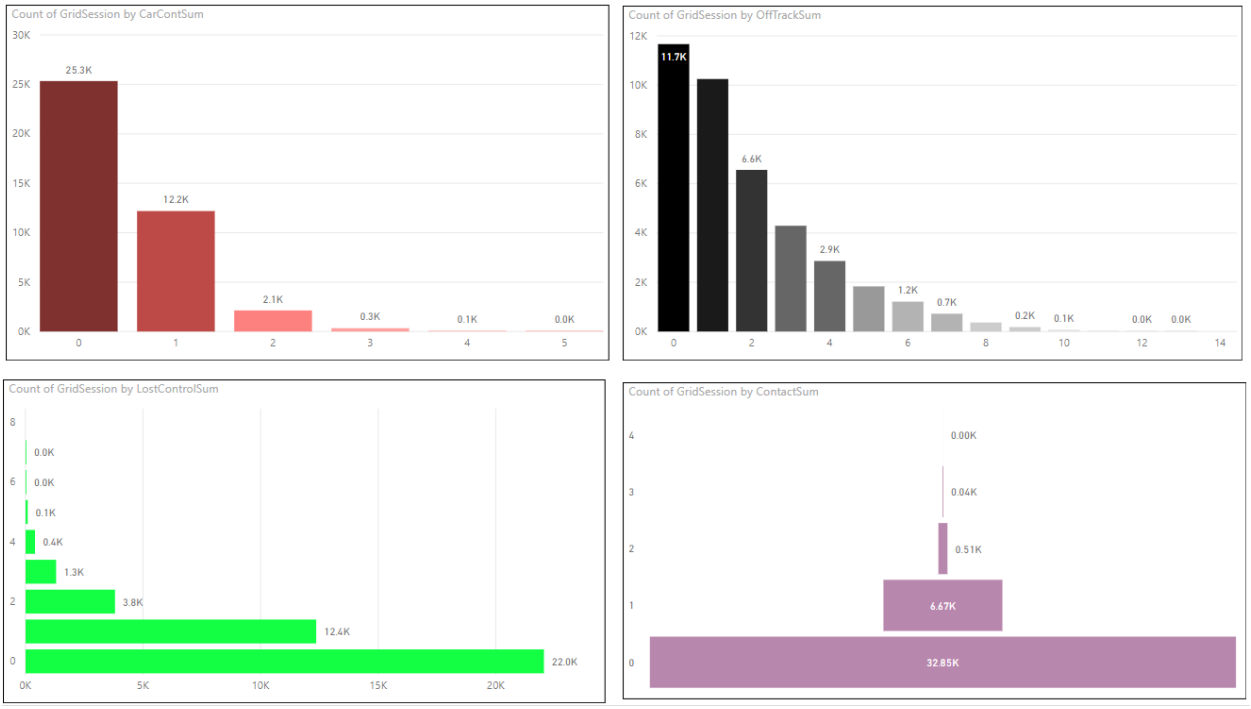
Petition for Redress

In the various conversations about Safety Rating, it's frequently assumed that a wide class of drivers cause trouble – whether in a certain iRacing 'club,' those who joined more recently, poorly-iRated drivers, or any number of other theories. None of that appears to be true, thankfully. As example, iRating appears to predict trouble less reliably than simply starting near the front:



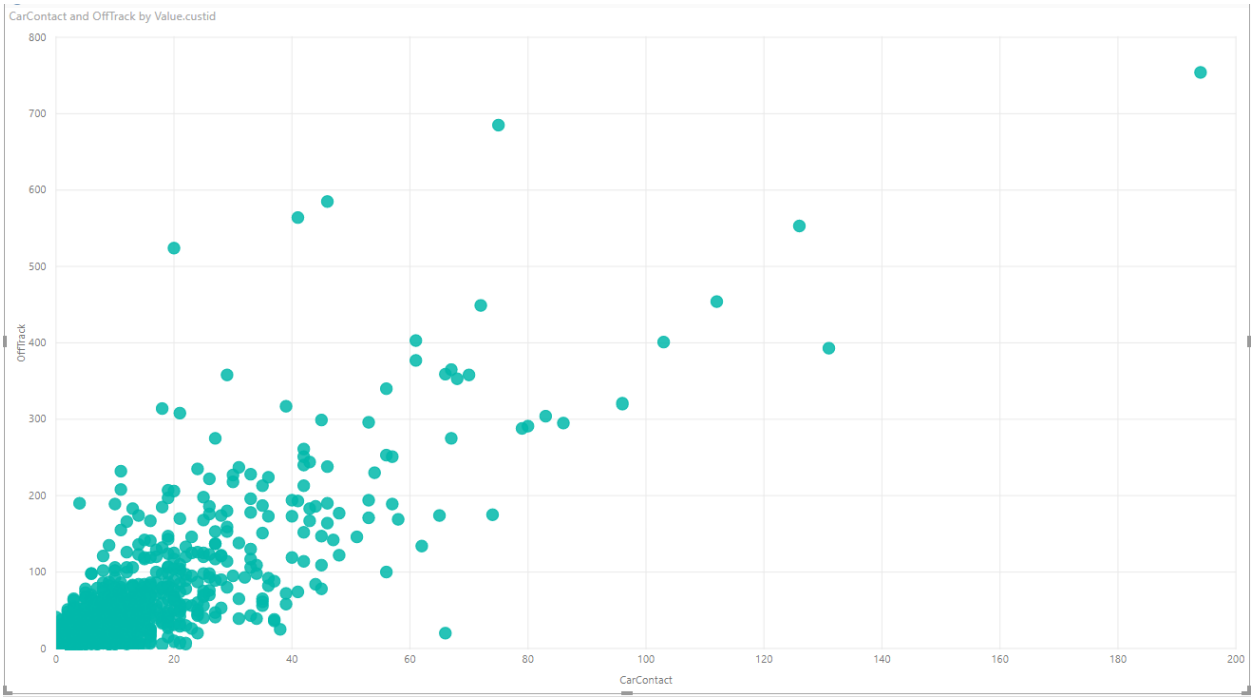
Ex. 10: Weak correlation between iRating (Car #) and incidents

Instead, when aggregating all the sessions' off-tracks and car contacts, one finds that the vast majority of iRacers fall into a common standard; but outliers who regularly incur accidents and errors are definitely present.



Ex. 11: Most iRacers drive cleanly – mostly

The above are counts of laps with the various types of incidents, per racer, per race.



Ex. 12: There are a few wild drivers out of the thousands who iRace

The above are totals of laps with car contact and with off-tracks among iRacers who drive at least 6 races (to eliminate one-off data points). Drivers are not identified – both because doing so is a violation of iRacing policy and because this data set is a sampling of three seasons of one series in iRacing's 10-year history, not an attempt to find 'bad apples.'

Nonetheless, this data clearly indicates that there are a relative few who abuse track limits and other drivers in their iRacing. While a more thorough analysis is pending involving machine learning methods against this data and a similar set, the author is prepared to make, if not strident demands, initial recommendations.

Conclusion and Recommendations

This preliminary analysis has significantly reassured the author that the core of the iRacing subscriber base is aligned toward good driving; and that, despite individual and group frustration at times, the system of Safety Rating and race petitioning has not failed to provide a reasonable simulation of race courtesy.

Nonetheless, parts of this information lead the author to some suggestions that might be tried, or potentially trialed in a voluntary or beta environment, toward the refinement of this system; and at least one that iRacers can put into practice today.

1. **Use the protest system** – The data suggests that the individuals who cause car collisions and otherwise drive unsafely are uncommon. Use of the protest system with the intent of allowing iRacing staff to coach/guide drivers and act more strongly in recalcitrant examples, seems warranted.
2. **Reconsider qualifying** – Because there's a correlation between unusually strong starting position and car collisions, the current qualifying system may require some further analysis and adjustment. In some sessions, qualifying may be done once for a week; but more commonly it occurs by race, often with the effect that a single off-track (1x) may grid a good driver in the back. The data suggests that, when this happens, someone, usually the qualified driver, may end up causing an incident.
3. **Consider progressive 4x penalties** – The incidence of multiple car contact events for a driver in a race is rare. Further, when it does occur, a large number of other drivers may be impacted by that one driver. The author suggests that, on a trial basis if possible or in a test environment, the penalties per race for suggestive car collisions be progressively worse (4x for the first, 8x for the second, 12x for the third). Some tuning of the current 17x limit for disqualification may need to accompany this, and the suggestion is intended for the road racing environment, but the author is confident that live marshals would think poorly of drivers who engaged in three (or more!) car collisions in a single sprint-length race.