SPAD SEVERITY VS DRIVERS

MGMT321: Business Analytics Project

Quang Minh Ngo

Aashish Agrawal

Muhammad Umer

NOVEMBER 7, 2019

MACQUARIE UNIVERSITY & PACIFIC NATIONAL

Executive Summary

The report looks at different aspects of SPAD such as potential severity, distant past signal and driver's experience as well as the number of drivers. Most of SPAD has a moderate of actual severity however there are 10% potentially became major. Whenever SPAD has more than 100 meters of distance past stop sign, there is a significantly higher probability of it become a major SPAD. There could be a relationship between the number of drivers and SPAD and drivers with more than 30 years of conducting involved in more SPAD than others.

pacificnational

Table of Contents

I. Insights

A. Number of SPAD overtime

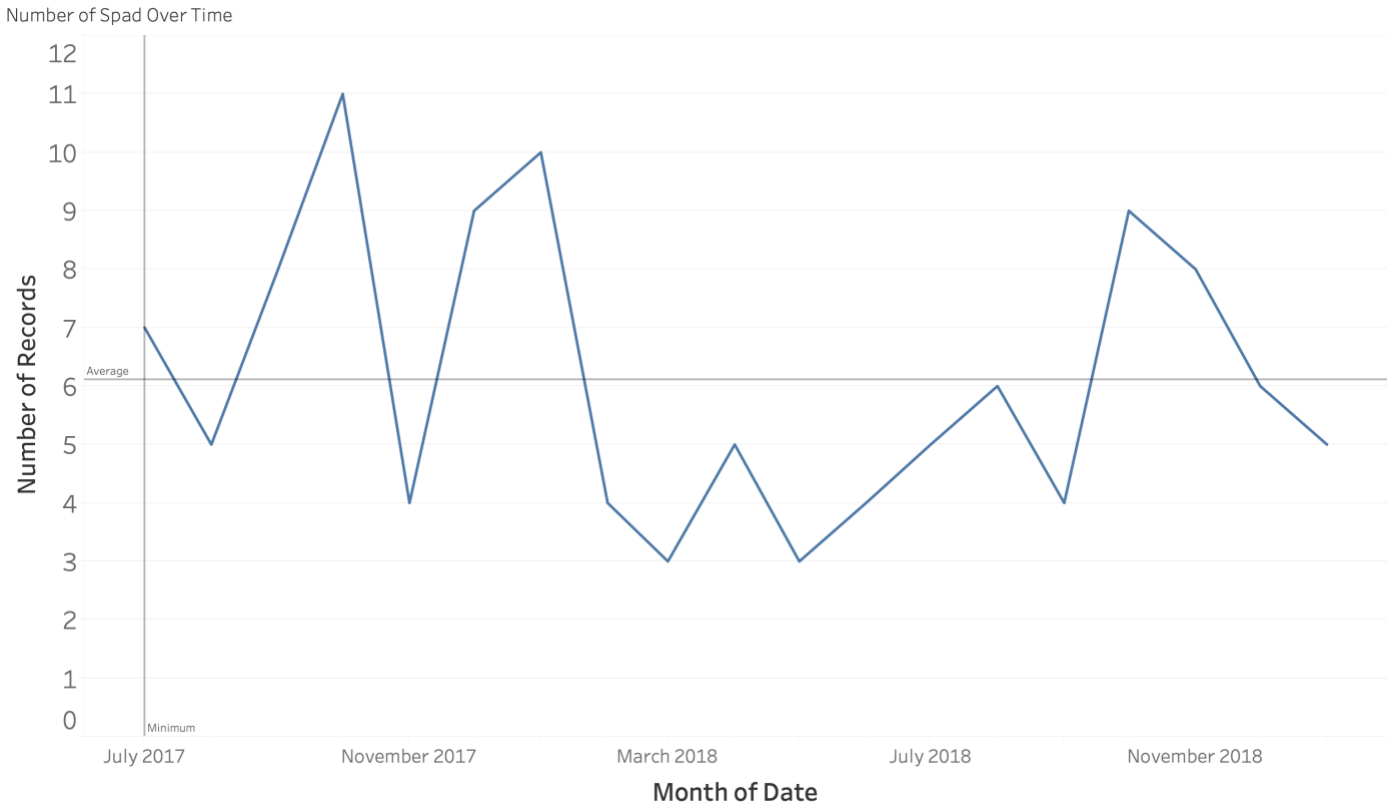


Figure 1: Number of SPAD overtime

There is not an obvious trend in the time series of all SPAD. On average, Pacific National should expect 6 SPAD every month.

B. Severity

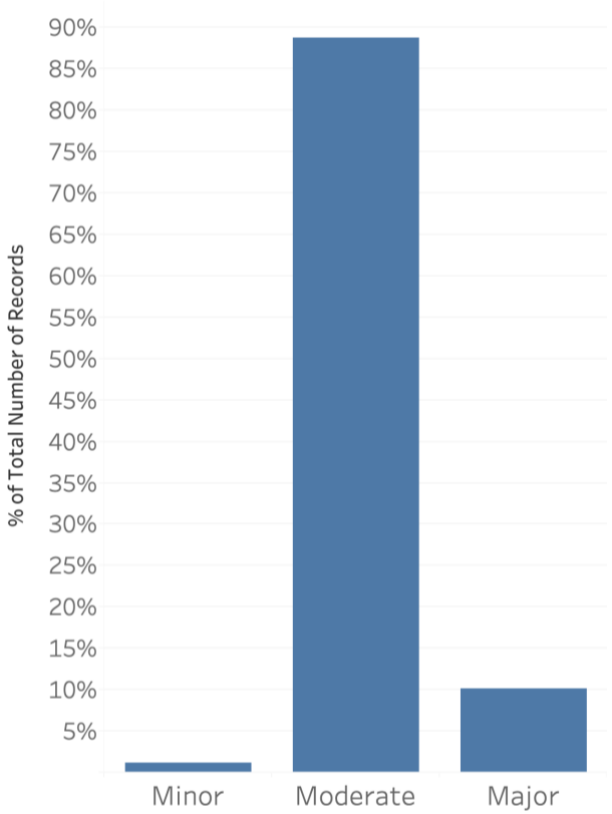


Figure 2: Potential Severity

Although nearly 100% of SPAD have moderate actual severity. There are roughly 10% of cases could have been worse and ended up as major accidents.

C. Distance Past Signal

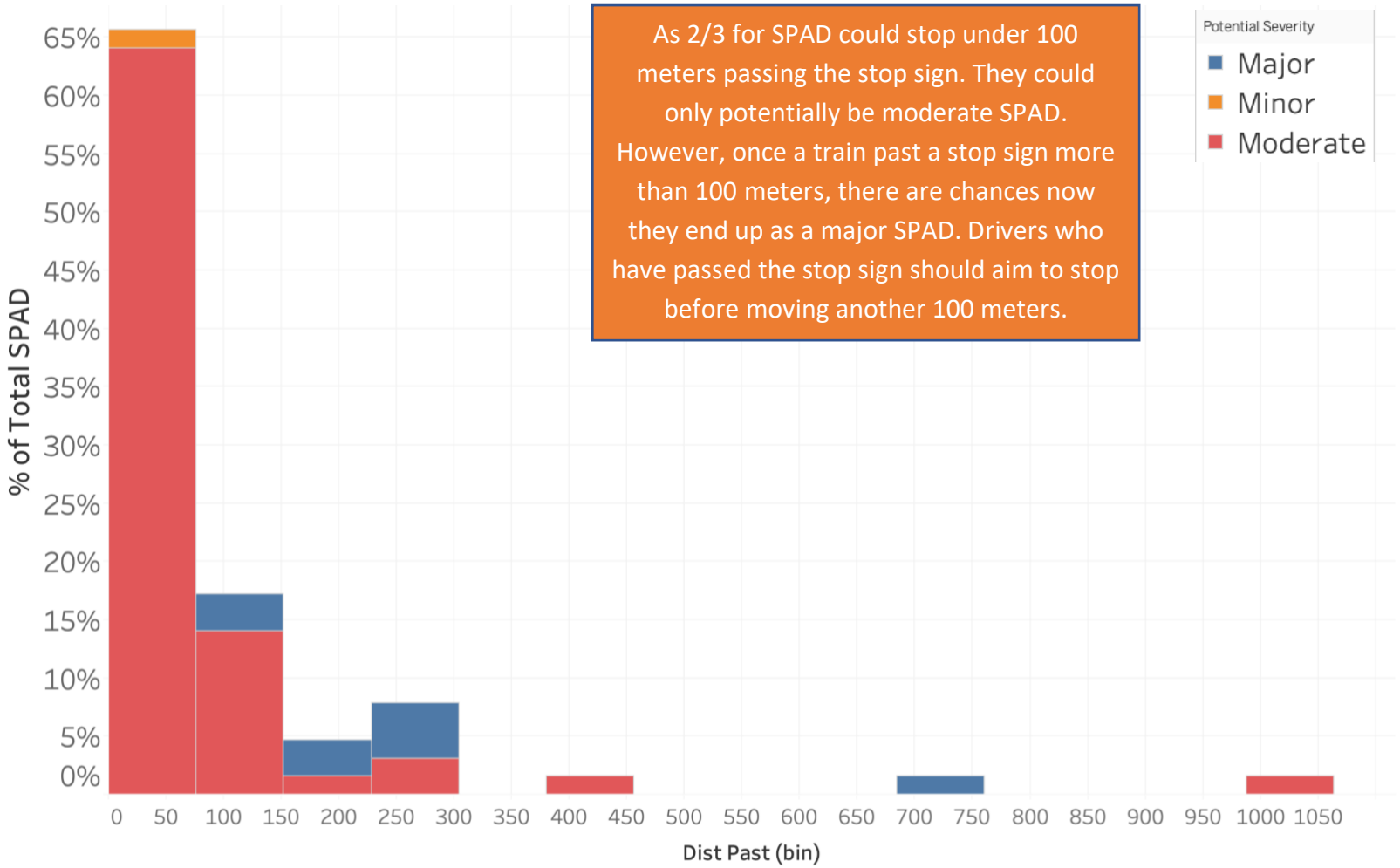
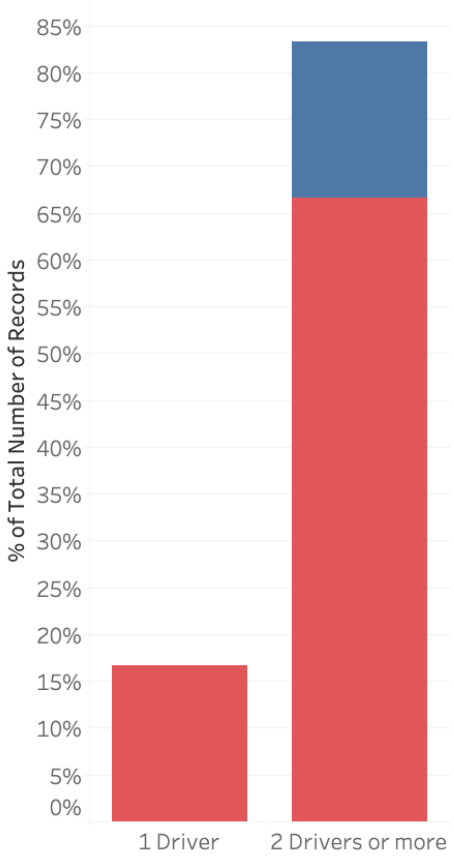


Figure 3: SPAD vs Distance Past Stop Sign

D. Number of Drivers



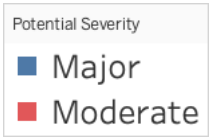
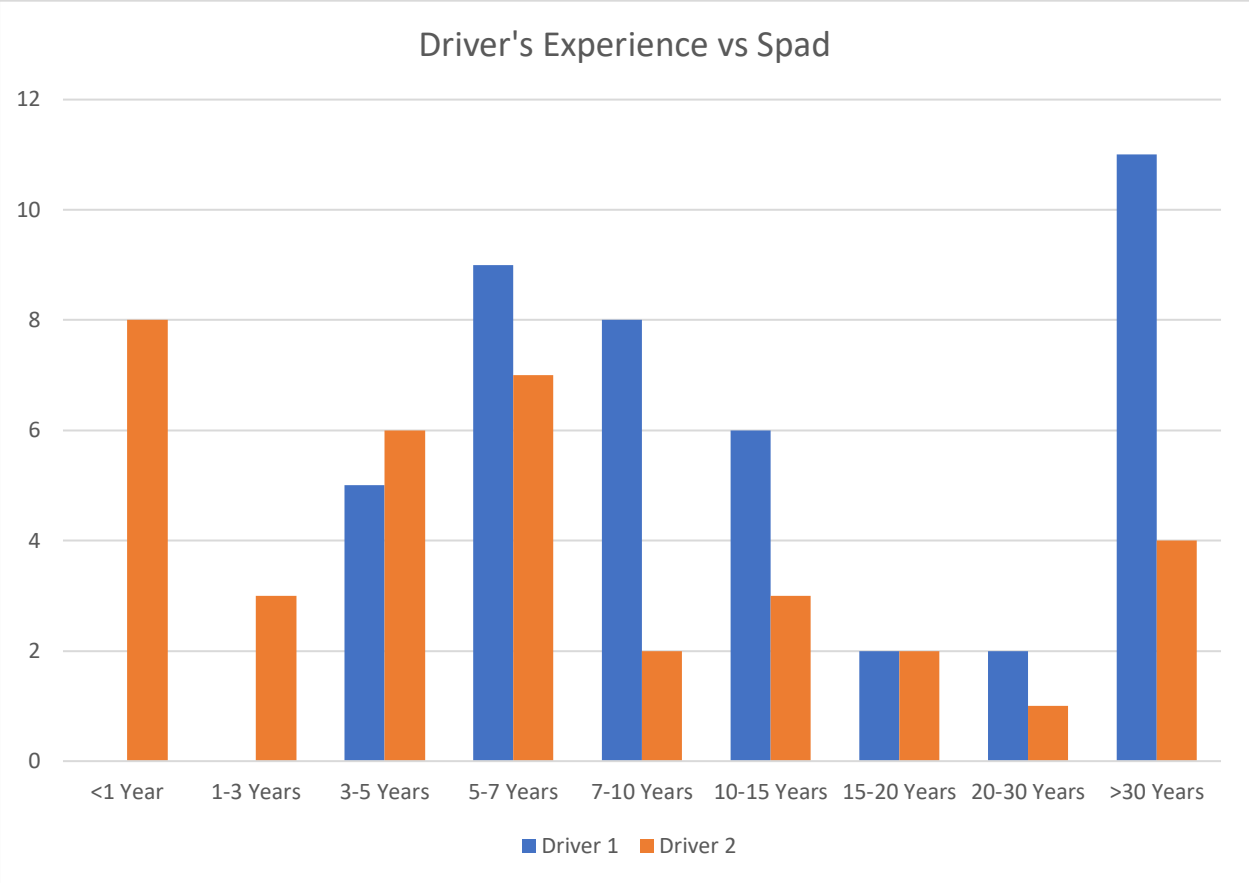


Figure 3: Number of Driver in each SPAD

More than 80% of SPAD involving 2 or more drivers and all of potential severity SPAD had 2 drivers or more on the train. It probably due to the proportion of trains with 2 or more drivers or the characteristics of those trips. However, this is still worth looking into.

E. Driver's experience



There could be a negative relationship between the driver's experience and the number of SPAD. As the more experienced drivers have, the less SPAD they have. However, drivers with more than 30 years of conducting train seem to have a high number of SPAD. It could just be of their ages. Hence, Pacific National should implement some types of examinations for drivers over a certain age.

pacificnational

II. Potential for Predictive Modelling

Our team has tried predicting the potential severity of each SPAD base on most of the logical variables. The result came back with 80% accuracy for SVM and Decision tree. Though this is a good sign for further modeling, the given dataset has many quality issues, therefore, it would not be sensible to pursuit predictive modeling at this stage. However, our team could see the potential for it and in the future, if Pacific National could obtain a better dataset with a new structure of variables, less missing value and more accuracy input, it is possible to predict potential severity of each SPAD.