# Analysis on Crash Reports on Motor Carriers - May 2018 to April 2020

## I. Introduction

When an accident occurs, each Motor Carrier is first classifying it as Department of Transportation (DOT) recordable or as none DOT recordable accident. If an accident is a DOT recordable accident then it could be found on the Federal Motor Carrier Association Safety Measurement System (FMCSA SMS)[1]. There are several categories by which each accident is classified: towed away from the scene, hazmat release, traffic way description, access control description, road surface condition, weather condition description, light condition description, citation issued description. Another data that is recorded for each DOT recordable accident is number of injuries, fatalities, state, severity weight and time weight. The data association rule mining will be applied on this dataset and we will see which are the most frequent itemsets and the association rules with the highest confidence and lift.

In the following sections the data used for analysis will be described, the methods used for analysis, results and conclusion. Section II contains a description of the dataset used for this analysis. The methodology for analysis is presented in section III. In section IV, I report and discuss the results. Finally, section V provides conclusions.

## II. Data Description

As shown in table 1 there are several categories by which each accident is classified: towed away from the scene (yes or no), hazmat release (yes or no), traffic way description (Two-Way Trafficway Not Divided, Two-Way Trafficway Divided Positive Barrier, One-Way Trafficway Not Divided, Not Reported, Two-Way Trafficway Divided Unprotected Median or Unknown), access control description (partial, full no control), road surface condition (Dry, Sand Mud Dirt Oil Or Gravel, Water(Standing Moving), Wet, Ice, Snow, Other, Unknown, Slush), weather condition description (No Adverse Conditions, Fog, Rain, Other, Snow, Blowing Sand Soil Dirt Or Snow, Severe Crosswinds, Unknown, Sleet Hail), light condition description (daylight, dusk, dark-lighted, dark-not lighted), citation issued description (yes, no or unknown). Another data, also shown in table 1, that is recorded for each DOT recordable accident is number of injuries which maximum is 74 per accident, fatalities which maximum is 20 per accident, state, severity weight (1, 2 or 3) and time weight (1, 2 or 3). Severity weight is measured based on severity of the accident. If there is a towing the weight is 1, if there is towing and injuries the severity weight is 2 and if there is fatality the severity weight is 3. As shown on table 2, 62.63% of the accidents are with severity 1, 36.75% are with severity 2 and 0.38% are with severity 3. Which means that the majority of the accidents are having only towing. On table 3 is shown the frequency and percentage for the time weight. In time weight, if it is 1, which is 53.47% of all accidents, the accident is older than a year. If it is 2 which is 25.16% of all accidents, is older than 6 months or if it is 3, which is 21.04% of all accidents, than the accident occurred between 0 and 6 months. The majority of the accidents occurred more than a year ago.

TABLE I. List of attributes

| Attribute | Type | Example Value | Description |
|---|---|---|---|
| REPORT_STATE | Nominal (string) | IA | State where accident occurred |
| FATALITIES | Numeric (integer) | 1 | Number of fatalities |
| INJURIES | Numeric (integer) | 5 | Number of injured |
| TOW_AWAY | Ordinal (yes or no) | Y | If a vehicle was towed away from the scene |
| HAZMAT_RELEASED | Ordinal (yes or no) | N | If hazmat was released |
| TRAFFICWAY_DESC | Nominal (string) | Two-Way Trafficway Not Divided | Trafficway description |

| Attribute | Type | Example Value | Description |
|---|---|---|---|
| ACCESS_CONTROL_DESC | Nominal (string) | Full Control | Access control description |
| ROAD_SURFACE_CONDITION_DESC | Nominal (string) | Dry | Road surface condition description |
| WEATHER_CONDITION_DESC | Nominal (string) | Snow | Weather condition description |
| LIGHT_CONDITION_DESC | Nominal (string) | Daylight | Light condition description |
| SEVERITY_WEIGHT | Numeric (integer) | 2 | Severity weight of the accident |
| TIME_WEIGHT | Numeric (integer) | 1 | Time weight of the accident |
| CITATION_ISSUED_DESC | Nominal (string) | Yes | If citation was issued to the driver |

TABLE 2. FREQUENCY AND PERCENTAGE ON SEVERITY WEIGHT

| Severity weight | Frequency | Percentage |
|---|---|---|
| Not listed | 897 | 0.33% |
| 1 | 170371 | 62.63% |
| 2 | 99968 | 36.75% |
| 3 | 775 | 0.28% |
| **Total** | **272011** | **100.00%** |

TABLE 3. FREQUENCY AND PERCENTAGE ON TIME WEIGHT

| Time weight | Frequency | Percentage |
|---|---|---|
| Not listed | 897 | 0.33% |
| 1 | 145454 | 53.47% |
| 2 | 68411 | 25.16% |
| 3 | 57219 | 21.04% |
| **Total** | **272011** | **100.00%** |

In Fig 1 is shown the frequency distribution of states where accident occurred. It is noticeable that the top three states with the highest number of accidents in the period of May 2018 to April 2020 are Texas with 27767 accidents, California with 22262 accidents and Florida with 14615 accidents. The top three states with the least accidents are Allaska with 42, Hawaii with 115 and Vermont with 312 accidents.

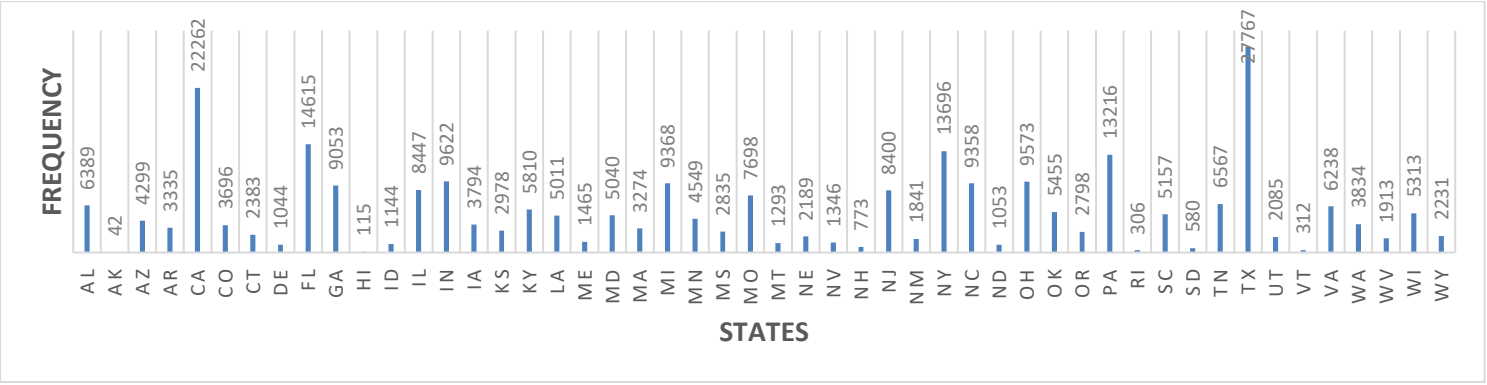


Fig 1. Frequency distribution of states where accidents occurred

In Fig 2 is shown the ratio of vehicles license state vs. state where accident occurred. Top three states with the highest ratio of occurred accidents and vehicle license plate from the same state are Wyoming, West Virginia and NM. Top three states with the lowest ratio of vehicles license state vs. state where accident occurred are Indiana, Illinois and Oklahoma.

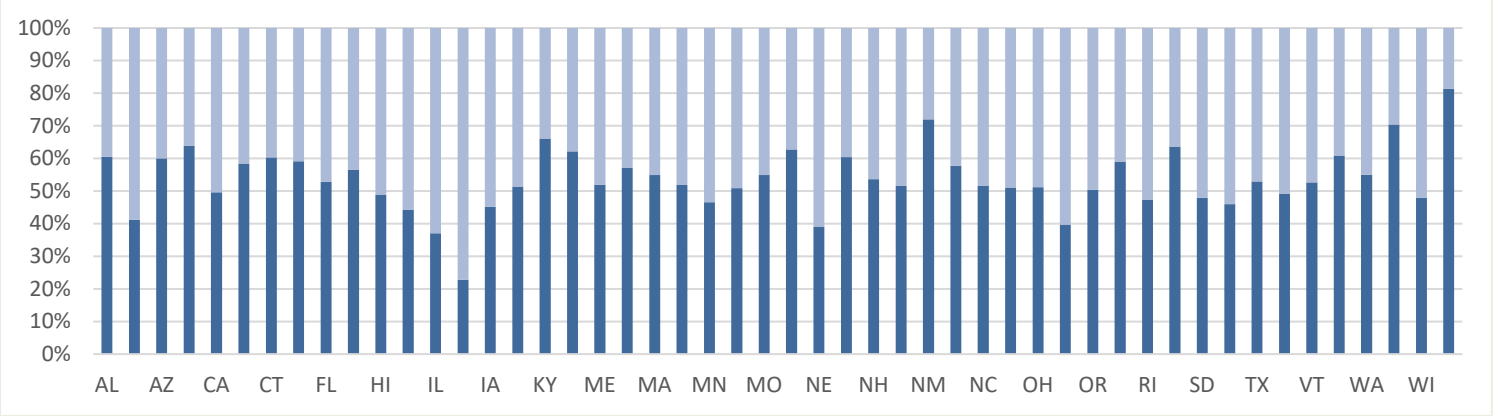


Fig 2. Ratio of vehicle license state vs. state of where accident occurred

In Fig 3 is shown the frequency distribution of trafficway. Top three trafficways are two-way trafficway not divided 86,737, two way trafficway divided positive barrier 86,405 and two-trafficway divided unprotected median 57,611.

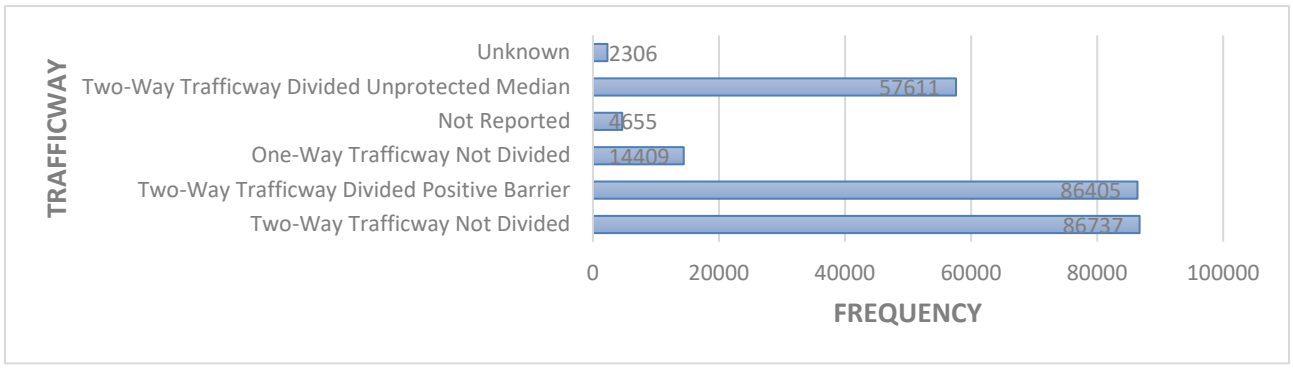


Fig 3. Frequency distribution of trafficway

In Fig 4 is shown frequency distribution of road surface condition. The majority of the accidents 205,867 out of 272,011 occurred on a dry road surface.

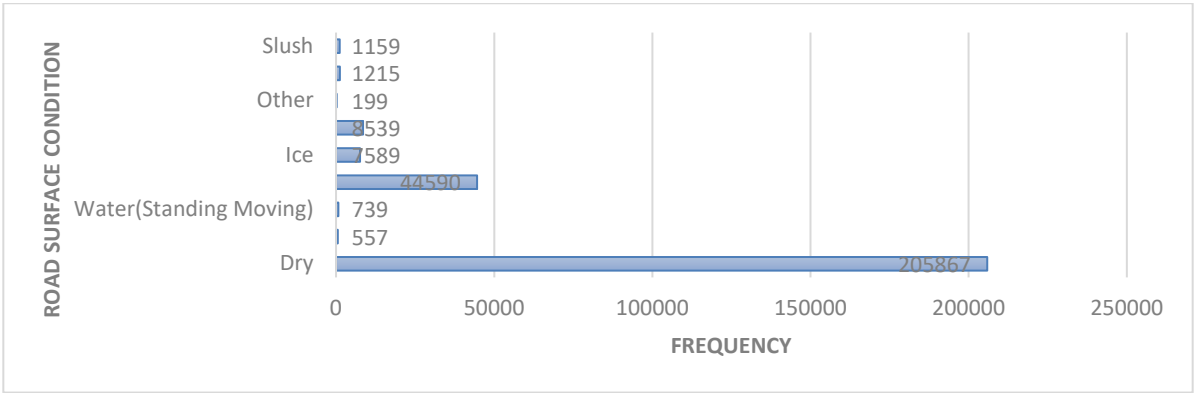


Fig 4. Frequency distribution of road surface condition

In Fig 5 is ahown the frquency distribution of weather condition, where it can be noticed that 209,239 out of 272,011 accidents occurred on no adverse weather conditions.

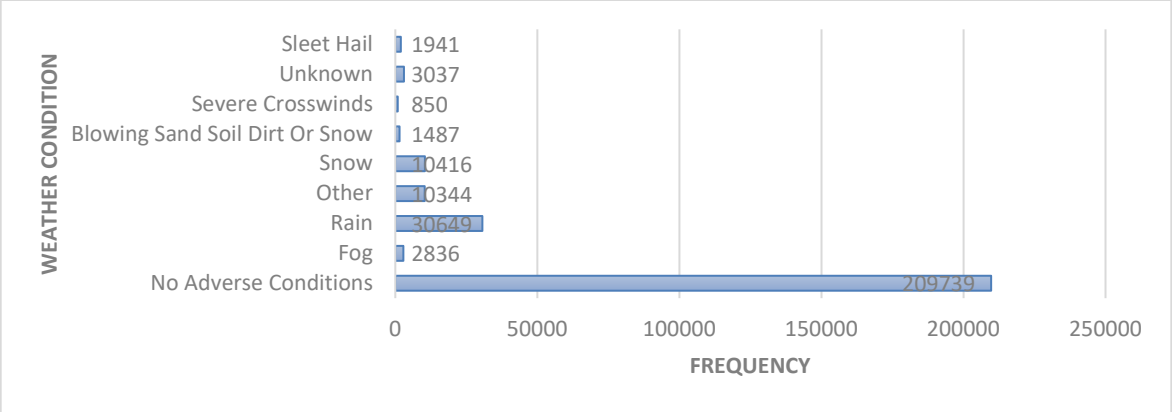


Fig 5. Frquency distribution of weather condition

## III. Methodology

Orange [2] was used to process the data and to retrieve useful analysis. In Fig. 6 it is shown the steps that were taken to preprocess the data and find association rules and frequent itemsets. After preprocessing the data was sent to do association rule mining [3] and find frequent itemsets.

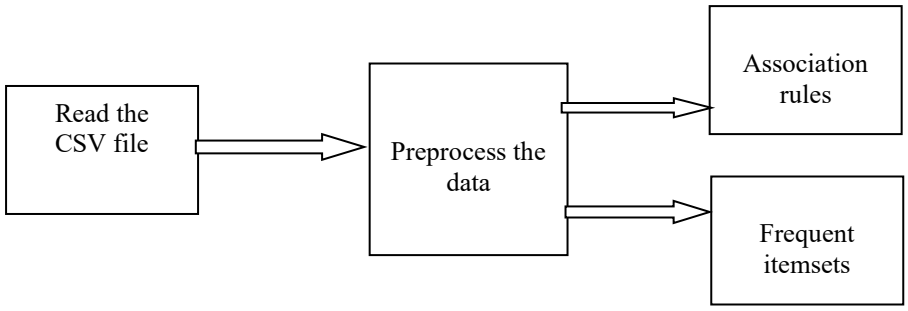


Fig 6. Taken steps to perform analysis on the data.

After reading the dataset from the CSV file some of the columns that would not give any relevant results were ignored. Then the dataset is sent to preprocessing where with equal frequency discretization was selected because it gave the best results. And after that the data was sent to association rules mining and frequent datasets. At association rules minimal support was set at 0.01%, minimal confidence was set at 90% and maximum number of rules was set on 10,000. The top 3 results with highest lift are shown in Table 4 in Section 4. At frequent datasets the minimum support was set on 1% and the maximum number of itemsets was set on 10,000. Top 3 results are shown in Table 5 Section 4.

## IV. Results and Discussion

As shown on Table 4 accidents with fatalities <= 0.5, injuries <= 0.5, no hazmat release, two way trafficway not divided, dry road surface, daylight, time weight <= 1.5 and no citations issued have smaller chances to have no adverse condition and severity weight less than 1.5. Accidents with fatalities <= 0.5, vehicle towed away, no hazmat released, two-way trafficway not divided, dry road surface, daylight, time weight <= 1.5 and no citation issued have smaller chances to have no adverse weather condtions and severity weight 1.5. Accidents with fatalities <= 0.5, injuries <= 0.5, no hazmat release, two way trafficway not divided, dry road surface, daylight, time weight <= 1.5, no citations issued have smaller chances to have a tow away, no adverse weather conditions and severity weight <= 1.5.

TABLE 4. TOP 3 ASSOCIATION RULES RESULTS

| Support | Confidence | Lift | Antecedent | Consequent |
|---|---|---|---|---|
| 0.018 | 0.96 | 2.029 | FATALITIES=< 0.5, INJURIES=< 0.5, HAZMAT_RELEASED=N, TRAFFICWAY_DESC=Two-Way Trafficway Not Divided, □¨ ROAD_SURFACE_CONDITION_DESC=Dry, LIGHT_CONDITION_DESC=Daylight, TIME_WEIGHT=< 1.5, CITATION_ISSUED_DESC=NO | WEATHER_CONDITION_DESC=No Adverse Conditions, SEVERITY_WEIGHT=< 1.5 |
| 0.018 | 0.96 | 2.029 | FATALITIES=< 0.5, INJURIES=< 0.5, TOW_AWAY=Y, HAZMAT_RELEASED=N, TRAFFICWAY_DESC=Two-Way Trafficway Not Divided, □¨ ROAD_SURFACE_CONDITION_DESC=Dry, LIGHT_CONDITION_DESC=Daylight, TIME_WEIGHT=< 1.5, CITATION_ISSUED_DESC=NO | WEATHER_CONDITION_DESC=No Adverse Conditions, SEVERITY_WEIGHT=< 1.5 |
| 0.018 | 0.96 | 2.029 | FATALITIES=< 0.5, INJURIES=< 0.5, HAZMAT_RELEASED=N, TRAFFICWAY_DESC=Two-Way Trafficway = Not Divided, □¨ ROAD_SURFACE_CONDITION_DESC=Dry, LIGHT_CONDITION_DESC=Daylight, TIME_WEIGHT=< 1.5, CITATION_ISSUED_DESC=NO | TOW_AWAY=Y, WEATHER_CONDITION_DESC=No Adverse Conditions, SEVERITY_WEIGHT=< 1.5 |

On Table 5 is shown the top 3 results of the frequent itemset results. Accidents with fatalities =< 0.5 92.3% of the cases have tow away, 70.8% of the cases occurred in no adverse conditions and 51.84% of them occurred in a daylight. Accidents with tow away, 72.80% of them occurred in no adverse conditions, 71.55% on a dry road surface and 66.97% of them occurred in a daylight. Accidents with no adverse weather conditions, 56.55% of them occurred on a daylight, 47.31% are with severity weight =< 1.5 and 40.93% have time weight less or equal to 1.5.

TABLE 5. TOP 3 FREQUENT ITEMSETS RESULTS

| | **Tow away** | **No adverse condition** | **Daylight** |
|---|---|---|---|
| **Fatalities =< 0.5** | 92.30% | 70.80% | 51.84% |
| | | | |
| | **No adverse weather condition** | **Dry road surface** | **Daylight** |
| **Tow away** | 72.80% | 71.55% | 66.97% |
| | | | |
| | **Daylight** | **Severity weight =< 1.5** | **Time weight =< 1.5** |
| **No adverse weather conditions** | 56.55% | 47.31% | 40.93% |

## V. CONCLUSIONS

The data association rule mining was applied on the FMCSA Crash report (May 2018 to April 2020) dataset and we saw which are the most frequent itemsets and the association rules with the highest confidence and lift.

Top three states with the highest number of accidents in the period of May 2018 to April 2020 are Texas with 27767 accidents, California with 22262 accidents and Florida with 14615 accidents. The top three states with the least accidents are Allaska with 42, Hawaii with 115 and Vermont with 312 accidents. Top three states with the highest ratio of occurred accidents and vehicle license plate from the same state are Wyoming, West Virginia and NM. Top three states with the lowest ratio of vehicles license state vs. state where accident occurred are Indiana, Illinois and Oklahoma. As shown on Table 4 accidents with fatalities <= 0.5, injuries <= 0.5, no hazmat release, two way trafficway not divided, dry road surface, daylight, time weight <= 1.5 and no citations issued have smaller chances to have no adverse condition and severity weight less than 1.5. Accidents with fatalities <= 0.5, vehicle towed away, no hazmat released, two-way trafficway not divided, dry road surface, daylight, time weight <= 1.5 and no citation issued have smaller chances to have no adverse weather condtions and severity weight 1.5. Accidents with fatalities <= 0.5, injuries <= 0.5, no hazmat release, two way trafficway not divided, dry road surface, daylight, time weight <= 1.5, no citations issued have smaller chances to have a tow away, no adverse weather conditions and severity weight <= 1.5. On Table 5 is shown the top 3 results of the frequent itemset results. Accidents with fatalities =< 0.5 92.3% of the cases have tow away, 70.8% of the cases occurred in no adverse conditions and 51.84% of them occurred in a daylight. Accidents with tow away, 72.80% of them occurred in no adverse conditions, 71.55% on a dry road surface and 66.97% of them occurred in a daylight. Accidents with no adverse weather conditions, 56.55% of them occurred on a daylight, 47.31% are with severity weight =< 1.5 and 40.93% have time weight less or equal to 1.5.

# REFFERENCE

[1] FMCSA SMS data set - https://ai.fmcsa.dot.gov/SMS/Tools/Downloads.aspx

[2] Orange (Software) – Developer: University of Ljubljana

[3] Association Rule Mining - Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207. CiteSeerX 10.1.1.40.6984. doi:10.1145/170035.170072. ISBN 978-0897915922.