

W200 - Project 2 - March 2021

Car Crash Data Analysis in the USA in Midwest - 2014

Estrella Ndriansay, Viswanathan Thiagarajan, Kumar Narayanan

Dataset: The data set that is being analyzed has > 200K recorded cases of the Car Crashes in the midwest area with Chicago as the main city. It covers several regions around Chicago.

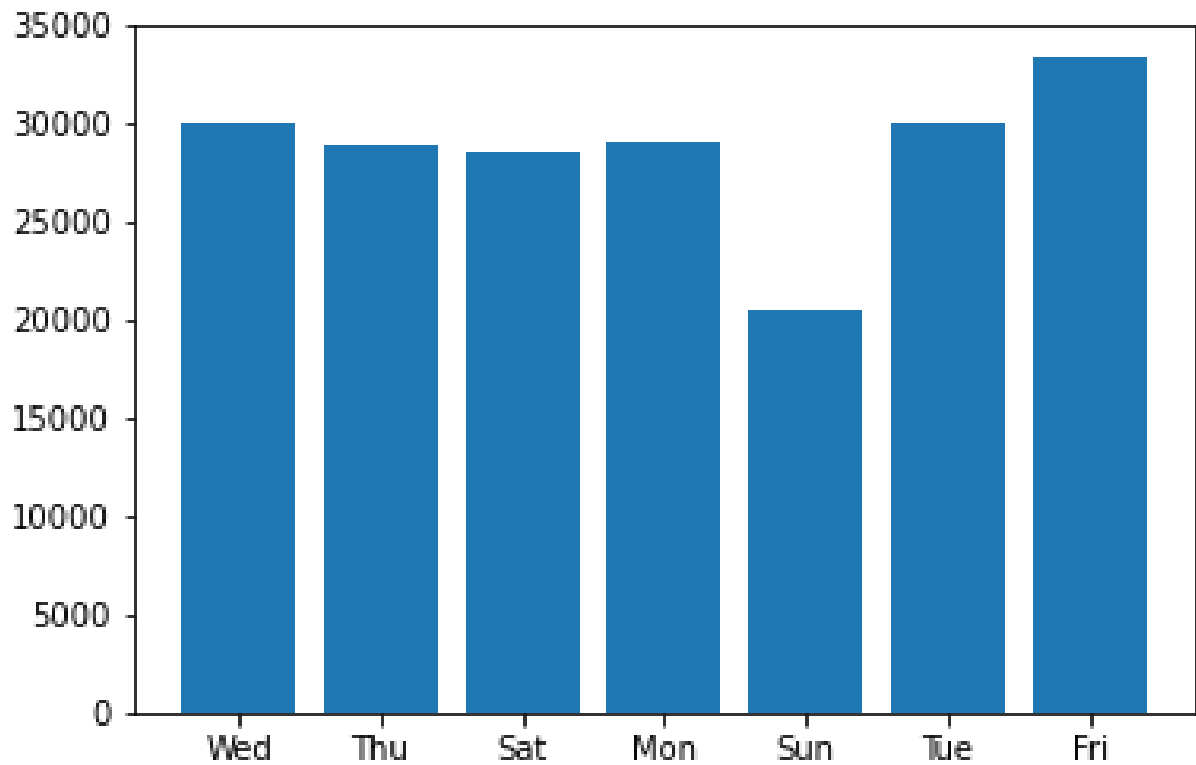
Link: <https://www.kaggle.com/qcarver/crashes-2014-csv>

The car crash data set (**CrashReport2014.csv**) has the following properties.

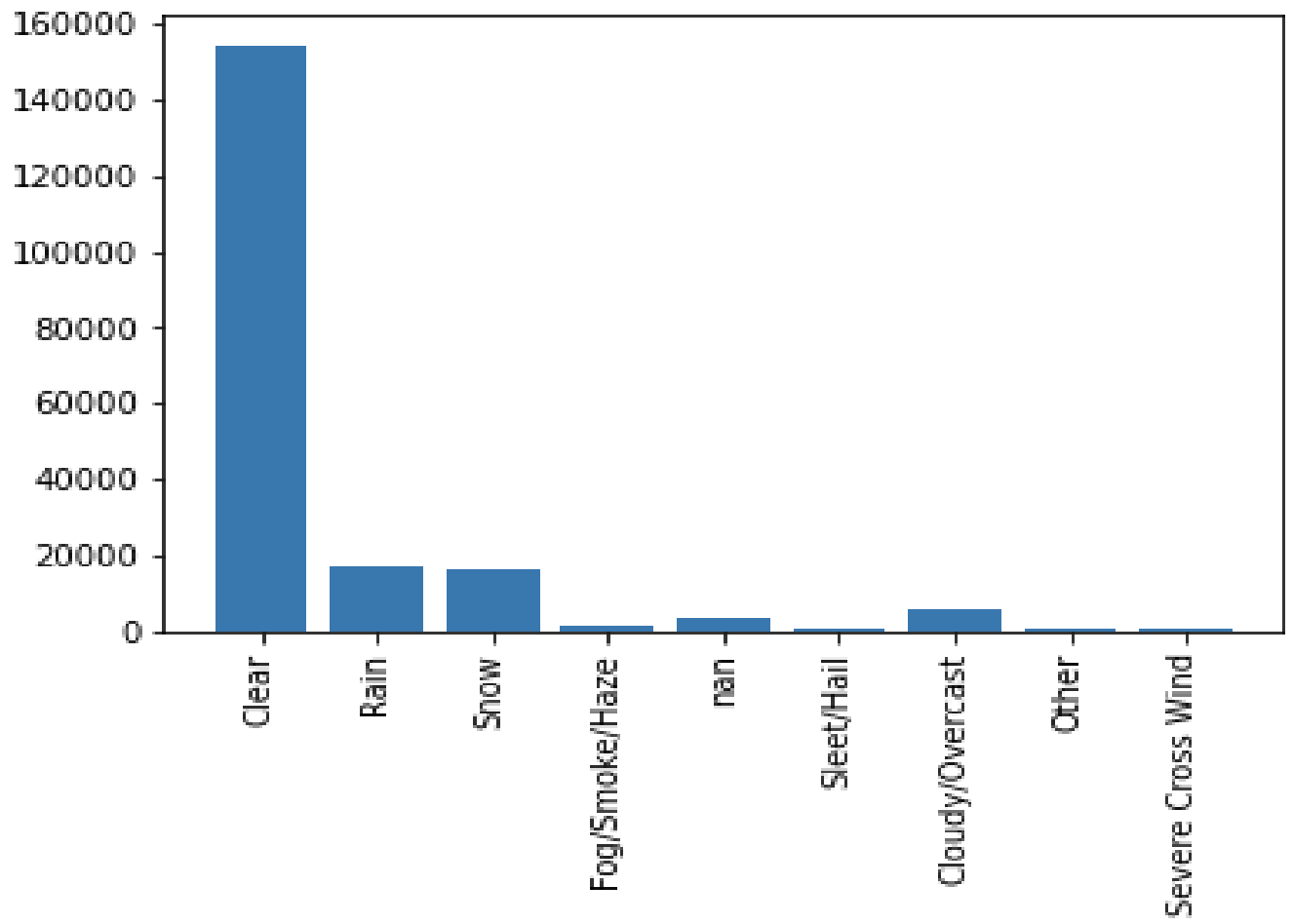
- Size: 128.61MiB (MiB is used to indicate that 1KiB = 1024B, and not 1000B, and subsequently 1MiB is 1024x1024B and not 1000x1000B).
- 292019 rows and 80 columns, of which 1st 200K rows are considered for analysis.
- Out of the 80 columns the interesting ones are:
 - ROUTE: The Highway or the road where accident happened
 - YEAR, MONTH, DAY, HOUR, and DAY_O_WEEK
 - NUM_VEH, INJURIES, FATALITIES, COLL_TYPE
 - WEATHER, LIGHTING
 - SURF_COND, RD_DEFECT, RD_FEATURE, TRAF_CNTRL
 - DRIVER_1, VEH1_TYPE, VEH1_MANUV, VEH1_EVENT1, VEH1_LOC1
 - XCOORD, YCOORD, INTERSEC, WorkZone, WorkZoneTy, WorkersPre, ExceedSpee, CellPhoneU

Initial Exploration:

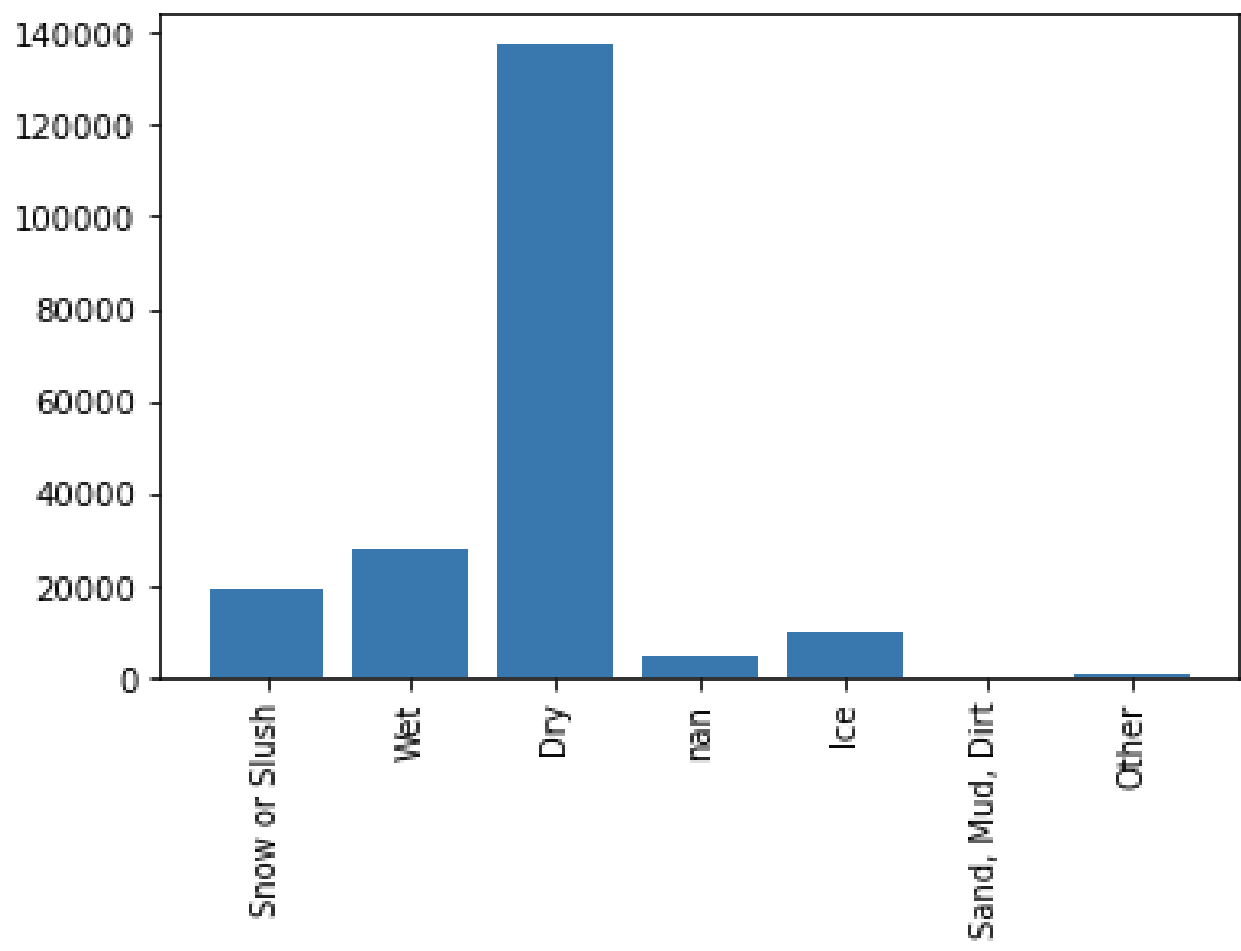
1. Accidents and crashes by the day of the week:



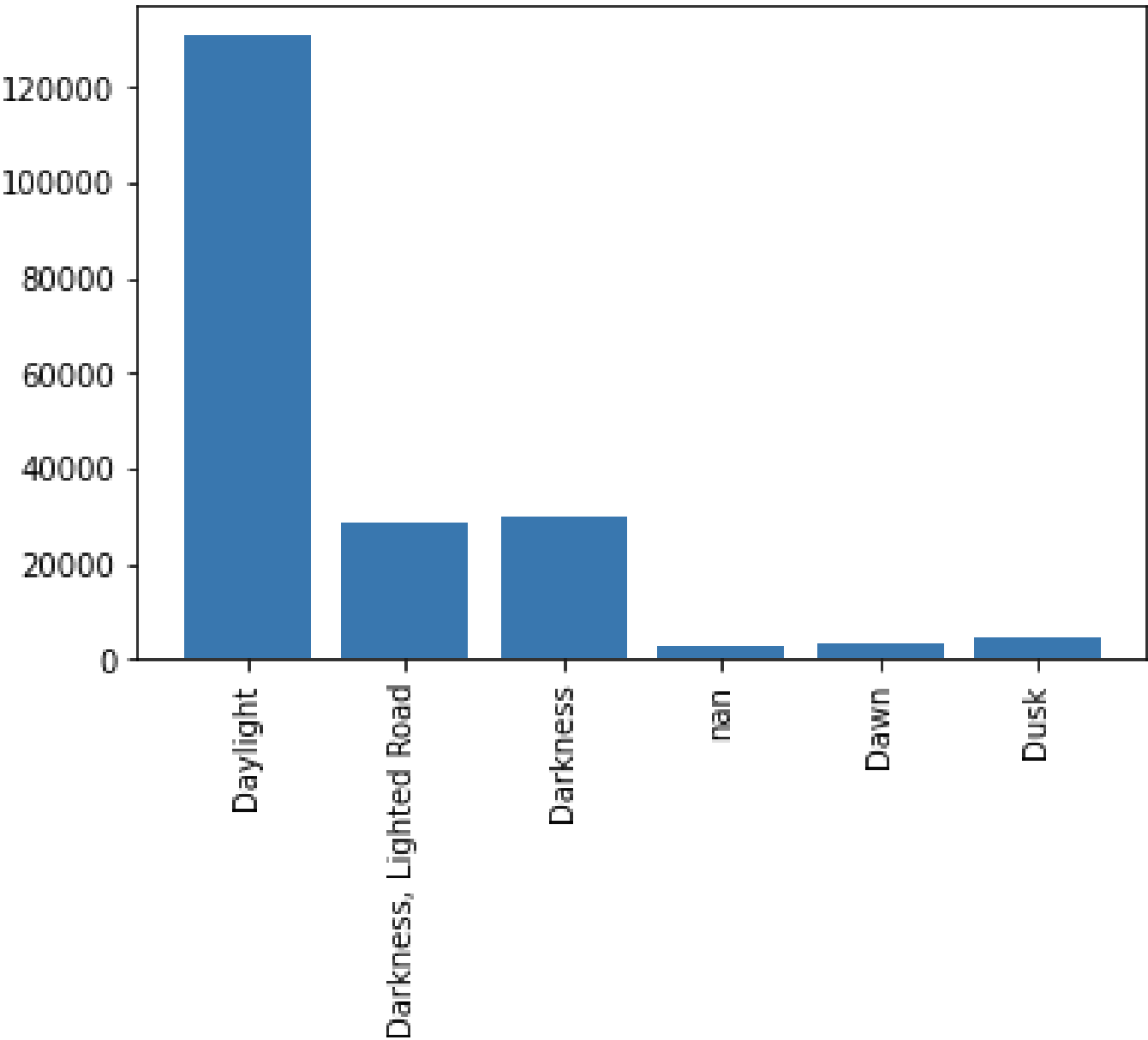
2. Accidents due to weather conditions:



3. Accidents due to Surface Conditions:



4. Accidents due to Light conditions:



Questions and analysis we plan for the final report:

- What combination of natural conditions causes the most accidents - weather, lighting, day of the week?
- Is there a particular day of the months that we see more accidents occurring? Is there any kind of clustering of the accidents?
- What are the different types of accidents? What percentage of accidents is only with the single vehicle, Vs accidents involving multiple vehicles?
- What weather condition caused most accidents?
- Is there a correlation between location (X/Y Coordinates), WorkZone, Speed limit exceeded, Cell Phone Usage and accidents?
- What type of vehicle causes most accidents?
- Are there correlations between factors such as presence of Intersection, WorkZone etc. to accidents?
- Is there a particular Highway (or route, in general) that is contributing excessively? If so, are there additional factors such as surface conditions, weather etc.?
- What about fatality rate with respect to single car Vs multiple cars, weather conditions, road conditions etc.?

We expect additional questions to pop up as and when we do the data analysis, and will add details if these correlations reveal something interesting.