Correlation of Weather with the number of Accidents that take place in New Jersey.

Author Name: Ashwin Thobbi
Dept. name of Organization: NYU
City, Country: New York City, New York

Author Name: Ritu Sirkanungo
Dept. name of Organization: NYU
City, Country: New York City, New York

Author Name: Ashish Yadav
Dept. name of Organization: NYU
City, Country: New York City, New York

***Abstract*—The impact of weather on accidents that happen on road is examined in detail. Both the weather data and accident data are analyzed thoroughly to check the degree of causal relationship that might exist. This would be advantageous in informing travelers about making decisions to travel on certain days where there might be higher risk of accidents. Also, the Highway Agencies might be more alert during days of bad weather and regulate traffic in a more efficient manner. Hadoop and Map Reduce framework is used in conjunction with HiveQL to analyze the large data sets.**

Keywords:Hadoop,MapReduce,HiveQL,NCDC,Big Data,New Jersey State Police,Weather,Accidents

I. INTRODUCTION

Weather affects many aspects of transportation. Inclement weather affects traffic demand, traffic safety, and traffic flow relationships. Understanding these relationships will help highway agencies select better management strategies and create more efficient operating policies. In this project, we are leveraging the power of Big Data Analytics to measure the correlation between number of accidents and different seasons during which they might increase, decrease or remain constant. This would help to mitigate the risk of traveling by informing travelers to be cautious while traveling during certain days where there might be higher probability of accidents as compared to others. Given the heightened risk of drivers' involvement in a crash, highway agencies might wish to manage better and restrict use of highways during times of extreme weather, to reduce safety costs and costs associated with rescuing stranded and injured motorists in the worst weather conditions. The first step in managing the transportation systems to minimize the weather impact is to quantify its impact on traffic which this project will enable to do. In addition, this might also help other organizations like insurance companies charge the drivers in a different way. They can charge more during inclement seasons as compared to other seasons where the drivers might be less prone to accidents.

II. MOTIVATION

Bad weather can cause multifarious problems. Road accidents are one of them. This project helps to find the extent to which bad weather can result in the increase in the number of accidents. This would help to reduce the number of accidents that take place during inclement weather by increasing security, avoiding unnecessary travel, traffic monitoring etc. Highway agencies can pump up the number of personnel required to manage security. They can also restrict use of certain highways. People with visibility issues can be warned not to drive during fog rain as it might increase the risk of accidents.

The analytic can also be used by insurance companies to check if weather was a significant factor or not in the accident. Let’s take a couple of examples. Let’s say that it is raining heavily. Heavy rain affects visibility and the road conditions. It’s harder for a driver to see, and the road is more slippery. Most drivers would go a little more slowly in heavy rain than they would if it were not raining. Thus, the hypothetical “reasonable person” would also go more slowly due to the conditions. When an insurance adjuster evaluates a car accident claim that arose in bad weather, the adjuster determines if the drivers were driving reasonably for the weather conditions. If the adjuster decides that its insured was driving too fast for the conditions, the adjuster will pay on the claim, even if the defendant was driving below the speed limit. This is because the key in bad weather claims is whether the drivers were driving reasonably for the conditions. On the other hand, when a driver is driving reasonably for the conditions, then an insurer might fight liability. Let’s say that it was snowing, and the insured (i.e., the defendant) was driving slowly and cautiously (well below the speed limit), but still crashes into you. In that case, the defendant’s insurer might well deny liability on the grounds that its insured was not negligent. The project will have a metric that will help the insurance companies in deciding the weather condition.

III. RELATED WORK

There are plenty of crashes that occur on highways in U.S. each day. Most of these crashes occur in the presence of rain, sleet, snow, fog, wet pavement, slushy pavement or icy pavement. Twenty four percent of all crashes are weather related. Each year, nearly 7,400 people are killed and over 673,000 people are injured in these crashes. This paper examines driver behavior and crash risk in diverse weather; Addresses weather exposure and severity; And presents statistics on weather related crashes, injuries, and fatalities on U.S. roads from 1995 to 2005. The paper also discusses regional variance in weather related crash types and rates, fatal weather related crashes involving commercial vehicles, and the economic impact of weather related crashes. Bad weather conditions affect driver capabilities, vehicle performance, pavement friction and sometimes roadway infrastructure and therefore risk of accidents increases. Brakes of the vehicles are not as efficient on wet roads when it becomes slippery. In one of the reports, investigators concluded that drivers did not recognize that pavement friction is lower on wet pavement as compared with dry pavement. In another study of crashes during rain events in Canada, investigators concluded that crash risk during rainfall was 70 percent higher than crash risk under clear, dry conditions. In one other study done in California, it was found out that on very wet days crash frequency was twice the rate on dry days. There are many more related studies in this area. One thing to notice in these studies is that wet pavement in the absence of rainfall is not viewed as hazardous, so speed reductions are minimal and many drivers do not realize that pavement friction is significantly reduced under these conditions, leading to greater stopping distances. While during the rainfall, drivers become cautious and hence reduce speed, but they generally do not tend to cancel their trips due to rain. The combination of high volume of traffic and relatively high speed gives rise to large number of accidents during rainfall and on wet roads. Not only this, there are economic impacts of weather related crashes also. The average cost including travel delay, emergency services, property damage, medical costs, etc. per crash in 2000 was estimated to be $14,100. On average, there are 1,561,430 police –reported weather related crashes. The annual cost of these crashes is estimated to be $22 billion, which is a huge amount of cost. Weather has significant impact on road safety. One quarter of all crashes on U.S. public roads are weather related. In terms of crash frequency, rate, and severity; wet weather is far more dangerous than winter weather. Most weather related crashes happen during rainfall and on wet pavement. This may be explained by exposure and driver behavior. Driver travel much slower when faced with winter weather and many cancel trips resulting in lower traffic volumes. Weather related crashes have significant impacts on roadway mobility and the economy. It was estimated that weather related crashes causes between 94 million and 272 million hours of delay each year. Weather related accidents can be minimized by doing more research on weather related crashes. It will be really helpful to better understand about the weather related accidents and to minimize their effect. We must aware more and more drivers about the risk of increasing speeds on wet roads, and education programs can be developed to train them about the better way to drive according to various weather conditions.[1]

It has been known that accident risk increases with an increase in adverse weather conditions, especially precipitation. Huge public expenditures on road design and surfacing, as well as on snow control, are incurred on this premise. However, little empirical information exists to provide details on the overall magnitude and minimum thresholds of elevated risk, the impacts of different rains and their intensities and also how different drivers respond to these conditions. This type of information is really critical if we want to understand the nature of impacts of climate conditions on road safety. This paper analyzed past relationships between weather and road safety and investigates the usefulness of data for studying these relationships. There can be two approaches which can be taken in studying relationships between weather and road safety. The first is to identify and quantify changes in the highway transportation system as they occur in response to different weather scenarios, in order to predict the effect of weather on accident rates. The second approach is to depend on observational accident data and to document changes in accident rates as they vary with weather. Both the approaches have their good in some sense and not good in some other. In the first approach the emphasis is on process while in the second, the highway transportation system is treated more as a black box. To investigate the effect of weather conditions on road accidents, there are three key aspects which we need to consider, the driver reaction, the vehicle and the driving environment. These components are linked together by informational and mechanical interactions. When precipitation occurs, the road surface becomes either wet or icy/snowy, with resultant changes in surface friction. Considerable experimental research on the changes brought about in tire pavement friction by precipitation has been conducted. Most research has focused on wet road conditions that result from rainfall. Other parts of the highway transportation system are also affected by changes in weather. Driver visibility is affected by atmospheric conditions and it is probable also that most drivers consider weather in their decision making. Some trips are cancelled or rerouted, and driver attentiveness and control behavior are altered as a function of weather. However, empirical information in these areas is very limited. Preliminary results indicate that driver visibility decreases with increased rainfall intensity, mainly due to the film of water on the windshield. There are various results from different studies which indicate that precipitation is associated with a significant increase in traffic accidents, however, it is difficult to generalize the estimates from these results because of the diverse methodologies and settings on which the results are based. There is an opportunity to improve our knowledge in this area by analyzing the information of accidents with that collected at various different weather stations. Existing data can be used and relation between weather conditions and traffic accidents can be examined which can then be used for projecting the impact of climatic change on road safety.[2]

The association between weather and traffic variables/crash events appears to be dependent on motorists. But finding out about the effects that weather, specifically rain, has on driver reflexes in travel speeds, different traffic conditions, and on chances of accident occurrence is necessary to evaluate practical areas of traffic operation and safety measures. Research concentrates on the effects of rain on freeways in the Southeast region of the U.S. Conducted to clarify whether the results from earlier studies are location specific or not. The effect of rain on per hour mean speeds and traffic volumes were analyzed for freeway segments in Jacksonville, Florida. Results show a measurable decrease in both traffic aspects with increase in rain quantity. Crash data analyzed along those particular freeway sections depicted that hourly crash risks and crash rates /100 million vehicle miles of travel, based on rain occurrence hours, increased with increasing rain intensity. However, hour of the day and season had little effect on hourly crash occurrence. Rain intensity also dramatically increased the proportion of injury accidents in the variety of traffic conditions. Findings from this study will provide information on travelling speeds, traffic flow, and crash occurrence during rainy conditions for various rainfall levels on high speed freeways in the Southeast region of the United States Such information is necessary to understand the effects of weather on traffic in a region. This mapping will help in the engineering of real time prediction models and safety planning methods. It is evident that inclement weather effects driver behavior on freeway facilities by reducing travel speeds and increasing time head ways. Results from the above study show that during rain, average hourly crashes relative to the no. of rain exposure hours increase along Florida freeways by an amount equivalent to as much as 2.7 times greater as compared to dry weather conditions (that is for precipitation values <0.001) Also, at a very low level of significance, there was insufficient evidence to say that the hour has a measurable effect on hourly crash occurrence based on rain occurrence.[3]

The study depicts about the effects of weather conditions on automobile accidents in Montreal, Canada. It analyzed rain, mean temperature and snow and derived that all the three parameters have huge impact on road. Out of those three parameters, it was concluded that snowfall is the most prominent cause for increase in number of accidents. Dealing with daily data was one of the issues that obstructed the analysis as it is uncertain that precipitation and accidents will happen simultaneously. In the previous study it has been found that impact of temperature on road accidents tend to vary to a large extent. Many researchers analyzed the effects of rain on different areas. One of the research shows that the effect of rain were adverse in urban areas and on the contrary fatal accidents occur in scarcely populated areas. Furthermore, effects of snowfall were also recorded and the conclusion was the same. Data set for the years 1990-1992 were extracted and analyzed. However those data set failed to provide the information regarding the cost and amount of damage occurred in the accident. Data set of 1992 was also insufficient as it only provided information about the traffic density and one route site and number of accidents that occurred in a month. Several other factors like speed, size of vehicle were also required in order to estimate the number of accidents. Estimation also involved calculation on average number of accidents per day and use of correlation and regression equations that calculated the estimate for entire three years. Graphs for number of accidents were plotted against the independent variables rain, temperature and snowfall. These graphs were then put to observation and conclusions were derived from it. Annual data depicted that both snow and rain showed a proportional relationship with the number of accidents per year, however effect of temperature over accidents turned out to be varying as seen from winter and summer data. Thus, it became significantly clear the snow and rain are the important causes of accidents. It was suggested that roads should have pouring property so that the drainage problem due to rain and snow deposition over the roads could be avoided. After further studying these causes, it was observed that reduction in number of accidents was still insignificant. Hence additional studies are still required to affirm that the effects of the parameters over the number of accidents and their applicability across the country.[4]

This paper helps us to find a detailed analysis of the impact of rainfall on road traffic accidents. The authors argue about a new method for measuring rainfall (weather radar approach) and its effectiveness in correlating rainfall with weather accidents as compared to the traditional approach of using a representative surface meteorological station to represent an urban area. Rainfall is often cited as one of the major weather factors causing accidents. Rain can cause accidents in a number of ways such as loss of friction between the tyre and road and impaired visibility through rain on the windshield and spray from other vehicles. The effect of rainfall is usually expressed through relative accidents rates (RARs), the ratio of accidents recorded during a precipitation event to those during normal conditions. The matched pairs approach is commonly employed to determine RARs, and works on the basis that within a given area, the accidents observed (usually through police reports) during a period experiencing rain can be compared with a corresponding dry period. This is usually achieved by comparing a period exactly a week preceding or following the event, with the assumption being that other factors such as volume of traffic, driver demographic and light conditions will be broadly similar. It has also been shown that greater rainfall intensities lead to higher RARs and injury rates. However, there are several longstanding methodological difficulties that call into question the validity of RARs obtained in urban areas and make direct comparison between studies extremely difficult. One of the biggest problems in urban studies is the lack of representative meteorological station observations at suitable spatial and temporal resolutions. The paper attempts to address the two main methodological issues highlighted in the literature review; the spatial and temporal representativeness of meteorological data in city based accident analysis. Matched pairs analysis is performed on two large UK cities UK Meteorological (Met) Office NIMROD weather radar images directly over the urban areas to give a more representative measure of rainfall than station based approaches. Although this study considers UK cities, discussion is made on how the approach and methods can be applied to any urban area with available weather radar data, with suggestions on how other unconventional data could be used in areas without radar coverage. In this study, 3 hour period is considered a rain event if in excess of 70% of grid squares receives between 0.5 and 4 mm/hour for all 3 hours. Dry events are classified as 3 hour periods where no precipitation is captured by the radar images. Matched pairs were selected if a suitable dry day was available exactly one week before or after a rain event. Matched pairs including a public holiday or weekend were unusable, as traffic volume and road user demographic were likely to differ considerably. Additionally the ability to change travel plans in the presence of weather is likely to be greater on these days. The matched pairs were selected from weekdays between the hours of 7:00 and 20:00. The timings were chosen to Reduce the influence of low traffic volumes later at night, while still providing a large number of potential match pairs. However, this period will still have temporal variations in the ability to reduce exposure. The number of accidents that occurred during the 3 hour period for both the rain and no rain events of the matched pair were ascertained using ArcGIS. By finding an accident total for both rain and no rain events over the study period for each of the cities it was possible to derive RARs. This was repeated using station data, with the same criteria for rainfall amount. In both cases only one rain event was selected per day that being the first 3-h rain event that occurred. In this paper, Manchester and London are the two cities that are considered. The RAR for these two cities using the rainfall radar and station approaches are found. The RAR is greater than unity in both cities using the conventional station based approach suggesting that rainfall consistently represents a driving hazard. For Manchester, the relative accident rate of 1.5 implies that proportionally 1.5 accidents occur during rain events for every 1 accident during an equivalent dry spell. London has a lower RAR of 1.18. By using these RARs the number of accidents over the 3 year period attributed to sustained rained events can be estimated. In comparison with the results from the traditional station based approach, the RARs obtained from the weather radar approach diverge considerably. It must be noted that for both cities the number of matched pairs reduces when using the weather radar data. The number of viable matched pairs reduces from 45 to 32 in Manchester and from 26 to 15 in Greater London. This is to be expected as the criteria for rainfall amount must be met for the majority of grid squares in the cities over the 3-h period, as opposed to a single site in the weather station approach. The spatial and temporal criteria set for a rain event in this study limited the number of matched pairs compared to a station approach. As a result all accident types (property damage only, minor injury, serious injury) were included to obtain as large an accident count as possible. It also dictated the use of a single category for rainfall amount. Although the aim of this study was to compare a conventional and alternative approach under identical criteria rather than investigate the impact of rainfall intensities on RARs or injury rates, this would be possible using the weather radar approach. However, this would likely require a longer period of study than the 3 years which is used in the analysis of the paper.[5]

This paper explains the coupling between weather conditions and road accident risk at an aggregate level and on a monthly basis. Time series analysis models with explanatory variables that measure the weather quantitatively are used and applied to aggregate datasets of injury accidents for France, the Netherlands and the Athens region, over periods of more than twenty years. The results reveal a high degree of correlation between weather and number of injury accidents but the magnitude of the effects vary depending on whether we are considering rural roads or urban roads. This research is based on time series accident analyses conducted in Work Package 7 on "Data analysis and synthesis" of the EU-FP6 project "Safety Net - Building the European Road Safety Observatory and which illustrate the use of weather variables for analyzing changes in the number of road injury accidents Time series analysis techniques are widely used for analyzing changes in road safety trends as observed at a national level, with the use of meteorological variables to capture short term changes in road safety indicators. Fixing the time scale for the analysis is very important, as the effects of weather conditions on safety indicators may be expected to differ according to the time scale, for example depending on whether the day or the month is considered. As daily monitoring requires a large number of daily accident data counts, it is rarely used at the national level. Instead, monthly monitoring is preferred as it facilitates international benchmarking. The number of injury accidents that are recorded during inclement weather has led rain to be considered as the major meteorological explanatory factor for road accident risk. In France for instance, for the period 1990-2000, 14 % of all injury accidents took place during rainy weather and 1%, at the most during fog, frost or snow/hail. In this research, the focus is on three climatic situations: rainfall (as a risk factor), fair weather (as a mobility factor) and frost (which plays a role in both risk and mobility).These meteorological factors were measured on a daily basis by three variables: the rainfall, measured in mm, the maximum temperature during the day, measured in °C, and the occurrence of frost by means of a dummy variable (1/0). The average and extreme values of these three weather variables were both obtained. For France, the daily climatic variables were calculated by averaging the daily variables measured at the hundred meteorological observation points that are spread throughout the country. For the Netherlands, the daily variables were measured at a single centrally located observation point, the “de Bilt” station, and for the region of Athens the daily variables were measured at a single observation point in the old airport of Hellenikon, in the southern suburbs of Athens. These daily variables were then aggregated or averaged in order to construct monthly variables. The results from the model suggest that low temperatures are negatively correlated with the number of injury accidents (which is very similar to the negative correlation between the occurrence of frost and the number of injury accidents in France) and that unusually high precipitation (for Greece) is negatively correlated with the number of injury accidents. Accidents increase in June, October, November and December- but these increases do not appear to be significant at the monthly level at the usual confidence level - and decrease significantly in August: these results are no doubt linked to simultaneous changes in exposure. The authors have highlighted significant correlations between average weather variables and aggregate numbers of injury accidents for several regions of Europe: France, the Netherlands and the Athens region, on a monthly time scale over a long period. Similarities were found between France and the Netherlands but the results for rainfall effects differed for the Athens region. However, the magnitude and even the sign of the correlations vary according to the type of network considered. In addition to the results obtained with average weather variables, the effects of extremely low temperatures and extremely high precipitation could be highlighted for the Athens region. These results are consistent with previous results obtained for France with the same method and with an ARIMA model, with the exception of the negative correlation between precipitation and the number of injury accidents for the Athens region. In addition to the discussion of the results, the main points of discussion relate to the choice of the method, in particular the appropriate time series technique to use, the comparability of the results between European countries or regions, and the possibility of extending the model to other European countries or regions. The choice of the type of model used should not impact these results as, for instance for France, it has been demonstrated that different ARIMA models used with the same weather variables for fitting the aggregate number of injury accidents provided consistent results. Concerning the comparability of the results, the similarity of the results for France and the Netherlands could be extended to other European countries or regions with similar climate. Furthermore, controlling for exposure by including the number of vehicle kilometers driven in the models improves the interpretability of the model. In the case of the interurban network in France, for instance, the rainfall effect appears to be mainly direct on motorways – as exposure is unchanged, and to be partly indirect on main roads – as a result of changes in exposure. The current limitation is that the effects of extreme weather (heavy rain, low and high temperatures) are not yet modeled in a systematic manner, and this should be tested on groups of countries with similar climates. The possibility of including all the average data and extreme value data recorded in a month in a single injury accident data model has been investigated, and research in this direction should be continued as it may significantly improve the model’s explanatory fit and our understanding of how the weather affects road accident risk at a monthly scale. Although the weather itself cannot be changed, understanding how it affects road safety is useful for both analysis and prevention. Prevention is obvious an important application, and takes different forms: information campaigns, road improvements, local warning systems. [6]

IV. DESIGN

Work Flow Architecture

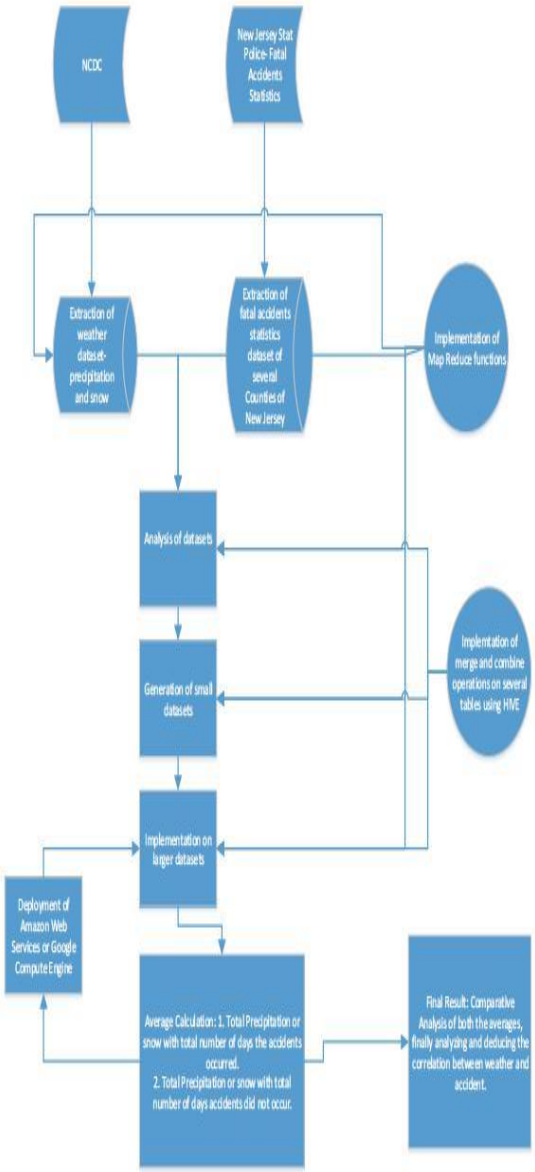

Fig 1. Work Flow Diagram (Flow Chart)

Flow Chart Description:

Aim of our project is to analyze and deduce a correlation between weather and accidents occurring in a region. We are analyzing New Jersey state dataset and are using different counties of New Jersey to support our analysis.

Step 1: Extraction and Profiling of data sets

NCDC provides us with weather data set and New Jersey State Police have provided us the fatal accidents statistics. Using the two huge data sets we have first extracted the following details:

NCDC dataset: Extraction of weather for each day across the year, mainly rain and snow values and profiling it for different counties of New Jersey

New Jersey State Police: Extraction of accidents data across the year and profiling the data on the basis counties where they have occurred.

These extraction are done by application of Map Reduce implementations.

Step 1 is repeated on several year datasets: mainly year 2008, 2009, 2010.

Step 2: Second step is to analyze the data and generate small dataset in order to make the understand implementation clearly.

Step 3: Tables, from the weather dataset and accident dataset, generated are then grouped together using HIVE

Step 4: The above implementations are then again applied for large dataset so as to support the final result

Step 5: Now, Average Calculation is done. Two final average values are derived:

a) Average: (total of precipitation or snow) by (total number of days the accidents have occurred).
b) Average: (total of precipitation or snow) by (total number of days the accidents have not occurred).

Step 6: Lastly, A Comparative study is applied on both the averages and final deduction and analysis is being stated that supports the correlation of weather and accidents.

High Level Component Architecture

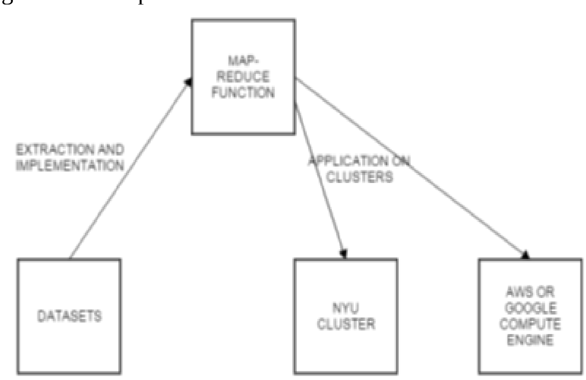

Fig 2. High Level Component Diagram

V. RESULTS

Year	Avg PREC ON DATES WITH NO ACC	Avg PREC DATES WITH ACC	TOTAL NUM OF ACC
2008	0.3037797	0.1566666	30
2009	0.2545871	0.9663157	39
2010	0.2045058	0.0609523	21
2011	0.2366564	0.0416216	40
2012	0.1404216	0.0786206	30

The above results are for Atlantic County.

VI. FUTURE WORK

The following analytic could be applied to all the counties of New Jersey State and further to all the states of United States. Using the Amazon Web Services cluster, the analysis can be done on large data and which may result in deriving a resultant equation that conforms to the analytic values. Furthermore, using the resultant equation along with the analytics of precipitation, analysis can further be refined for different type weather such as snow, drizzle, fog etc.

VII. CONCLUSION

It has been observed from the table generated that average precipitation on dates with accidents is not in direct correlation with the total number of accidents that have occurred in a county. Thus, the analysis disproves the general belief that implies a commensurate relationship between weather and accident rates This might be due to the fact that few people venture out on road during bad weather as compared to times of good weather resulting in lesser number of accidents during bad weather. Hence, we believe that improved safety measures be taken not only in times of inclement weather but during times of good weather as well. Lot of precautionary measures and warnings are given out during bad weather making drivers more cautious while driving in bad weather conditions as compared to when the weather is good. Drivers need to be made cognizant on how to drive safely in heavy traffic which is usually the case when the weather is good. To support this argument, Missouri state government website is being referred that aids the analysis by stating the fact of 2008 that provides the higher percentage of accident rates in clear weather in comparison to rainy season. [9]. This paper also supports the contrarian results that we got from our analytic.

VIII. ACKNOWLEDGMENT

We would like to thank National Climatic Data Center (NCDC) and New Jersey State Police for providing us with relevant resources for the project. We would like to thank Professor Suzanne K. McIntosh for her constant guidance throughout the project and her expeditious replies to all our emails. Lastly, we would also like to thank NYU, Amazon Web Services and Cloudera for providing us with the cluster and environment to analyze our dataset.

IX. REFERENCES

1.Pisano,L.Goodwin and M.Rossetti.Highway Crashes in Adverse Road Weather Conditional,1998

2.J. Andrey and R. Olley. Relationship between Weather and Road Safety:Past and Future Research Directions January 1990

3.M.Angel. The Effects of Rain on Traffic Safety and Operations on Florida Freeways,2014

4.M.Andreescu and D.Frost. Weather and traffic accidents in Montreal,Canada,1998

5.D.Jaroszweski, T.McNamara The influence of rainfall on road accidents in urban areas: A weather radar approach. January 2014

6.R.B. Hayat,M.Debbarh,C.Antoniu,G.Yannis Explaining the road accident risk: Weather effects.Elsevier December 2013

7. T. White. Hadoop: The Definitive Guide. O'Reilly Media Inc., Sebastopol, CA, May 2012.

8 .S. Ghemawat, H. Gobioff, S. T. Leung. The Google File System. In Proceedings of the nineteenth ACM Symposium on Operating Systems Principles – SO

9. Jackson County Missouri State: http://www.jacksongov.org/content/4847/5765/5807/