

**PROJECT TITLE: ANALYSIS OF ACCIDENTS**

**Submitted By:**

M N CHARITH REDDI

N AKSHAY REDDY

S SIVAKUMAR REDDY

P NITHEESH

B KIRAN REDDY

***CERTIFICATE***

This is to certify that this project based Class report entitled **“ANALYSIS OF ACCIDENTS ”** is a bonafide work done byM N Charith Reddi, N.Akshay Reddy, S.Sivakumar Reddy, P.Nitheesh, B.Kiran Reddy in partial fulfillment of the requirement for the award of **internship conducted by smart bridge** during the summer 2019.

***DECLARATION***

We hereby declare that this project based lab report entitled“ **ANALYSIS OF ACCIDENTS”** has been prepared by us in partial fulfillment of the requirement for the award of**internship conducted by smart bridge** during the summer 2019.

We also declare that this project based class report is of our own effort and it has not been submitted to any other university for the award of any degree.

**Date: 22-06-2019**

**Place: Hyderabad**

**Submitted By:**

M N CHARITH REDDI

N AKSHAY REDDY

S SIVAKUMAR REDDY

P NITHEESH

B KIRAN REDDY

**TABLE OF CONTENTS**

1.Abstract…………………………………..

2.Introduction………………………………

3.Importing Packages………………………

4.Exploratory Data Analysis……………….

5.Plots and Models………………………….

6.Algorithms………………………………..

7.Result…………………………………….

8.Conclusion……………………………….

**INTRODUCTION**

This project aims, through machine learning techniques, at creation a model for road traffic accidents classifications. Dataset contains information on accidents across Leeds. Data includes the District Name where accident occurred, Neighborhood Name, Street, Month, Day, Time, Part of the day to know whether the accident occurred in morning, afternoon, evening or in the night, injuries(mild or serious), victims, Number of vehicles involved, Longitude and the Latitude . It is important that measures be based on scientific and objective surveys of the causes of accidents and severity of injuries. One of the key objectives in accident data analysis is to identify the main factors associated with a road and traffic accident. accidents. In this Project, we are applying machine learning algorithms to modeling the severity of injury that occurred during traffic and the number of victims who are involved in road accidents.

**OBJECTIVES OF RESEARCH**

In recent years , there is a increase in the researches attention to determine the significantly affect the severity of the drivers injuries which is caused due to the road accidents. Accurate and comprehensive accident records are the basis of accident analysis. The effective use of accident records depends on some factors, like the accuracy of the data, record retention, and data anlaysis. There is many approaches applied to this scenario to study this problem.

**PROBLEM STATEMENT**

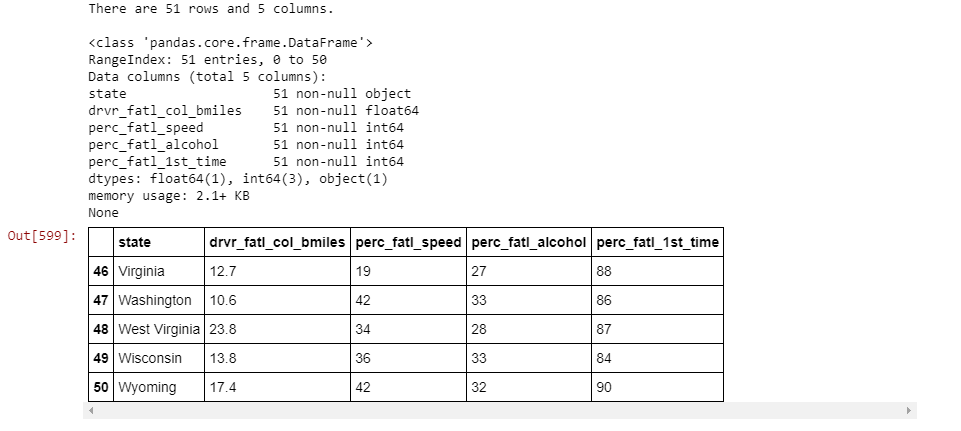
There is many approaches applied to this scenario to study this problem.The costs of fatalities and injuries due to traffic accidents have a great impact on the society. In recent years, researchers have paid increasing attention to determining factors that significantly affect severity of driver injuries caused by traffic accident.There are many inventories in in automobile industries to design and build safety measures for automobiles, but traffic accidents are unavoidable. There is a huge number of accidents prevailing in all urban and rural areas. Patterns involved with different circumstances can be detected by developing an accurate prediction models which will be capable of automatic separation of various accidental scenarios. These cluster will be useful to prevent accidents and develop safety measures. We believe to acquire maximum possibilities of accident reduction using low budget resources by using some scientific measures.

**3.Importing packages**

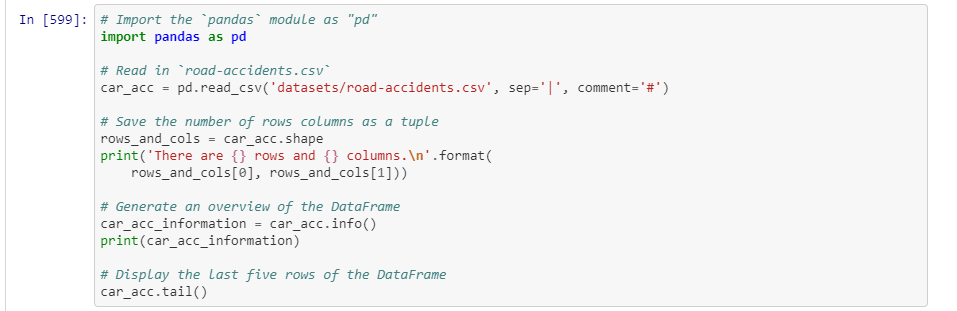
* Numpy
* Pandas
* Imputer
* Seaborn
* Matplotlib
* Sklearn(Logistic Regression, SVC, Decision Tree Classifier,

Random Forest Classifier)

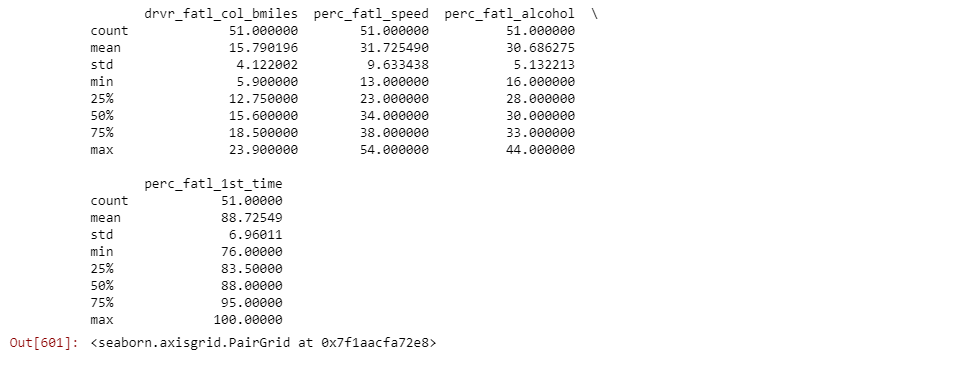
Exploratory Data Analysis

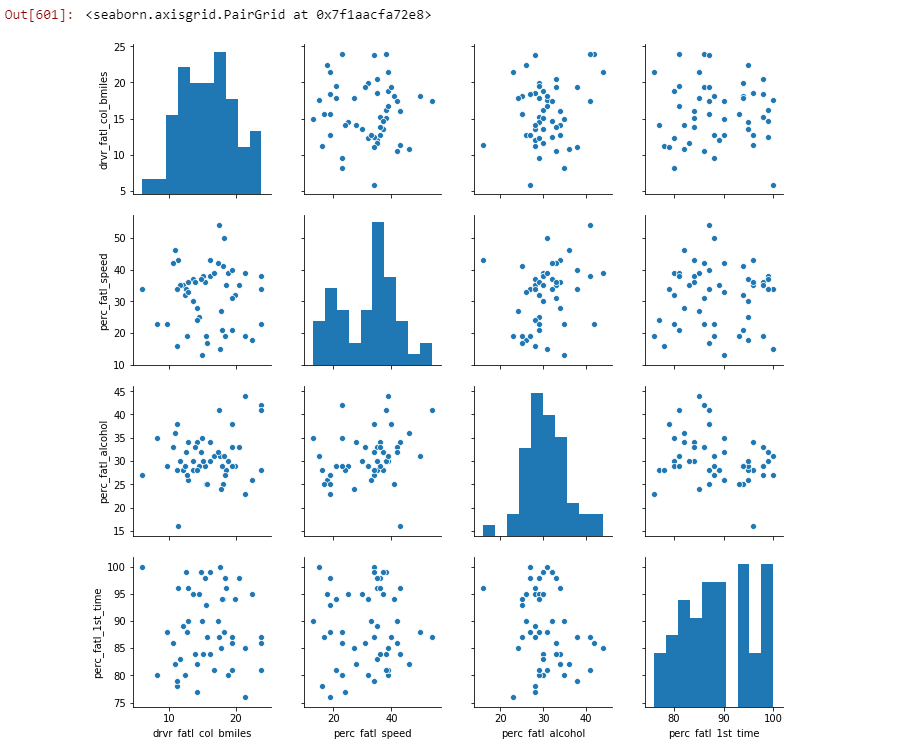
****

Reading the csv file

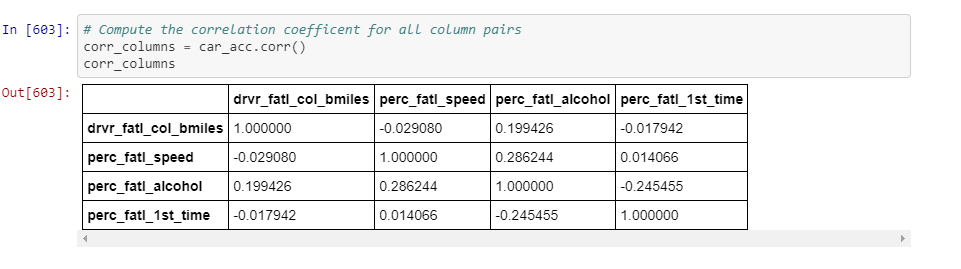


Textual and graphical analysis





Correlation coefficient



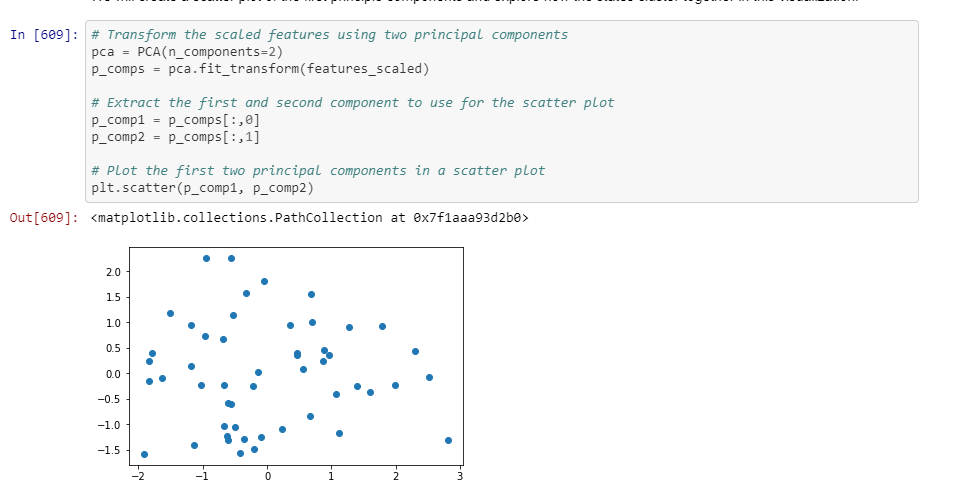
Multivariate linear regression



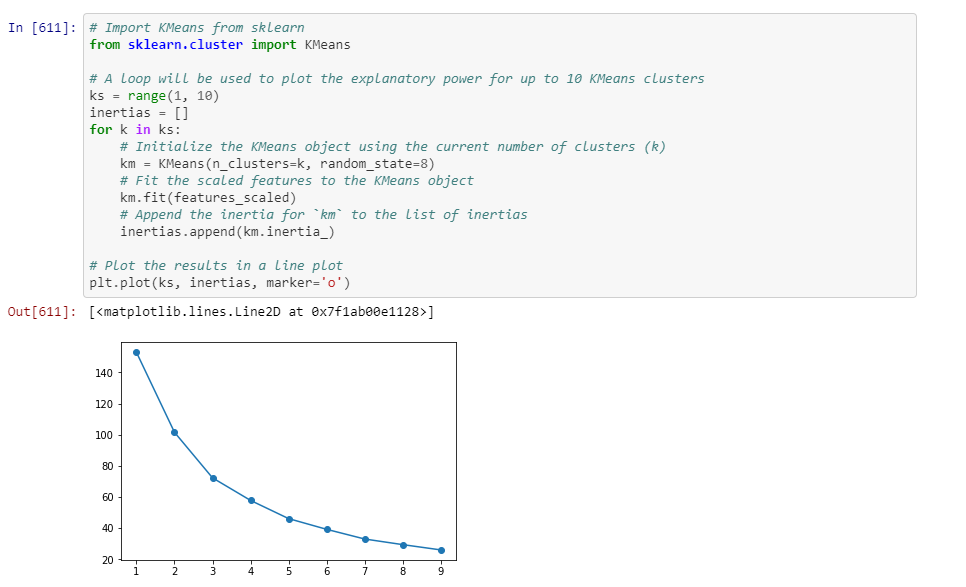
PCA on standardized data

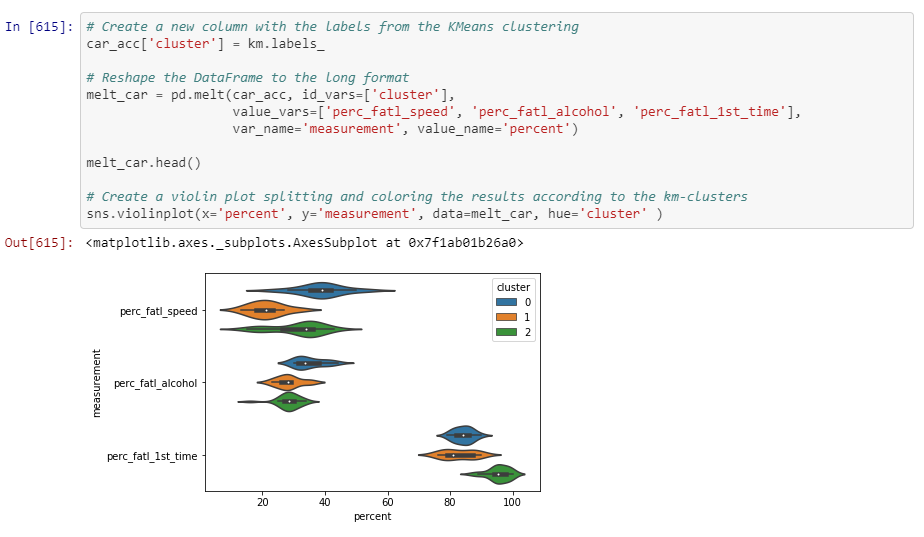


Visualisation of the first two principal components

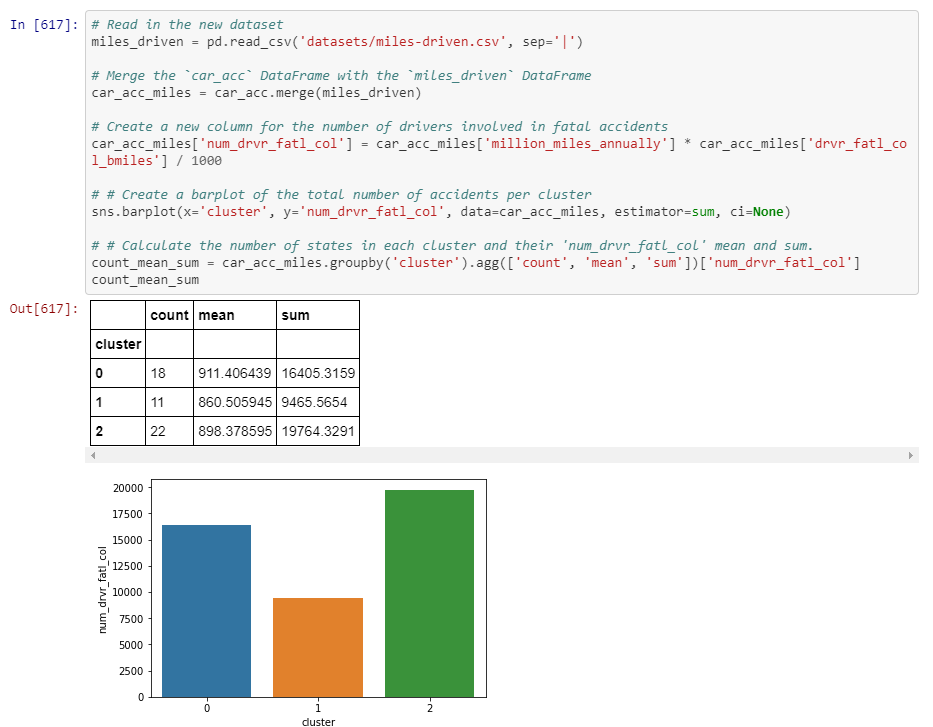


Find clusters of similar states in the data



Visualize the feature differences 

Compute the number of accidents within each cluster



**RESULT**

These are the following observations from the above models:

* The accuracy of KNN classifier on test set is:0.68
* The accuracy of Logistic regression on test set is:0.71
* The accuracy of Random Forest classifier on test set is:0.69

**CONCLUSION**

So, from the result we can say that Logistic regression, KNN classifier and Random Forest classifier the highest accuracy on acidents data set.

Overall Accuracy score is : 71.179884