

2020

# Predicting Traffic Accident Severity



Applied Data Science  
Capstone

Rautela, Sanjay

# Coursera Applied Data Science Capstone

---

This project is for the Applied Data Science Capstone-car accident severity, this is the final project for Data Science Professional certificate by IBM.

## Introduction

---

Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product.

Analyzing a significant range of factors, including weather conditions, special events, roadworks, traffic jams among others, an accurate prediction of the severity of the accidents can be performed.

These insights, could allow law enforcement bodies to allocate their resources more effectively in advance of potential accidents, preventing when and where a severe accident can occur as well as saving both, time and money. In addition, this knowledge of a severe accident situation can be warned to drivers so that they would drive more carefully or even change their route if it is possible or to hospital which could have set everything ready for a severe intervention in advance.

Governments should be highly interested in accurate predictions of the severity of an accident, in order to reduce the time of arrival and thus save a significant amount of people each year. Others interested could be private companies investing in technologies aiming to improve road safeness.

## Key Facts

1. Approximately 1.35 million people die each year as a result of road traffic crashes.
2. The 2030 Agenda for Sustainable Development has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2020.
3. Road traffic crashes cost most countries 3% of their gross domestic product.
4. More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists.
5. 93% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately 60% of the world's vehicles.
6. Road traffic injuries are the leading cause of death for children and young adults aged 5-29 years.

## Who is at risk?

**Socioeconomic status** More than 90% of road traffic deaths occur in low- and middle-income countries. Road traffic injury death rates are highest in the African region. Even within high-income countries, people from lower socioeconomic backgrounds are more likely to be involved in road traffic crashes.

### Age

Road traffic injuries are the leading cause of death for children and young adults aged 5-29 years.

### Sex

From a young age, males are more likely to be involved in road traffic crashes than females. About three quarters (73%) of all road traffic deaths occur among young males under the age of 25 years who are almost 3 times as likely to be killed in a road traffic crash as young females.

## Risk factors

**The Safe System approach: accommodating human error** The Safe System approach to road safety aims to ensure a safe transport system for all road users. Such an approach considers people's vulnerability to serious injuries in road traffic crashes and recognizes that the system should be designed to be forgiving of human error. The cornerstones of

this approach are safe roads and roadsides, safe speeds, safe vehicles, and safe road users, all of which must be addressed in order to eliminate fatal crashes and reduce serious injuries.

## **Speeding**

- An increase in average speed is directly related both to the likelihood of a crash occurring and to the severity of the consequences of the crash.
- For example, every 1% increase in mean speed produces a 4% increase in the fatal crash risk and a 3% increase in the serious crash risk. The death risk for pedestrians hit by car fronts rises rapidly (4.5 times from 50 km/h to 65 km/h)..
- In car-to-car side impacts the fatality risk for car occupants is 85% at 65 km/h.

## **Driving under the influence of alcohol and other psychoactive substances**

- Driving under the influence of alcohol and any psychoactive substance or drug increases the risk of a crash that results in death or serious injuries.
- In the case of drink-driving, the risk of a road traffic crash starts at low levels of blood alcohol concentration (BAC) and increases significantly when the driver's BAC is  $\geq 0.04$  g/dl.
- In the case of drug-driving, the risk of incurring a road traffic crash is increased to differing degrees depending on the psychoactive drug used. For example, the risk of a fatal crash occurring among those who have used amphetamines is about 5 times the risk of someone who hasn't.

## **Nonuse of motorcycle helmets, seatbelts, and child restraints**

- Correct helmet use can lead to a 42% reduction in the risk of fatal injuries and a 69% reduction in the risk of head injuries.
- Wearing a seatbelt reduces the risk of death among drivers and front seat occupants by 45 - 50%, and the risk of death and serious injuries among rear seat occupants by 25%.
- The use of child restraints can lead to a 60% reduction in deaths.

## **Distracted driving**

There are many types of distractions that can lead to impaired driving. The distraction caused by mobile phones is a growing concern for road safety.

- Drivers using mobile phones are approximately 4 times more likely to be involved in a crash than drivers not using a mobile phone. Using a phone while driving slows reaction times (notably braking reaction time, but also reaction to traffic signals), and makes it difficult to keep in the correct lane, and to keep the correct following distances.
- Hands-free phones are not much safer than hand-held phone sets, and texting considerably increases the risk of a crash.

## **Unsafe road infrastructure**

The design of roads can have a considerable impact on their safety. Ideally, roads should be designed keeping in mind the safety of all road users. This would mean making sure that there are adequate facilities for pedestrians, cyclists, and motorcyclists. Measures such as footpaths, cycling lanes, safe crossing points, and other traffic calming measures can be critical to reducing the risk of injury among these road users.

## **Unsafe vehicles**

Safe vehicles play a critical role in averting crashes and reducing the likelihood of serious injury. There are a few UN regulations on vehicle safety that, if applied to countries' manufacturing and production standards, would potentially save many lives. These include requiring vehicle manufacturers to meet front and side impact regulations, to include electronic stability control (to prevent over-steering) and to ensure airbags and seatbelts are fitted in all vehicles. Without these basic standards the risk of traffic injuries – both to those in the vehicle and those out of it – is considerably increased.

## **Inadequate post-crash care**

Delays in detecting and providing care for those involved in a road traffic crash increase the severity of injuries. Care of injuries after a crash has occurred is extremely time-sensitive: delays of minutes can make the difference between life and death. Improving post-crash care requires ensuring access to timely prehospital care and improving the quality of both prehospital and hospital care, such as through specialist training programs.

## **Inadequate law enforcement of traffic laws**

If traffic laws on drink-driving, seatbelt wearing, speed limits, helmets, and child restraints are not enforced, they cannot bring about the expected reduction in road traffic fatalities and injuries related to specific behaviors. Thus, if traffic laws are not enforced or are perceived as not being enforced it is likely they will not be complied with and therefore will have very little chance of influencing behavior.

Effective enforcement includes establishing, regularly updating, and enforcing laws at the national, municipal, and local levels that address the above-mentioned risk factors. It includes also the definition of appropriate penalties.

## What can be done to address road traffic injuries?

### Role of Government

Road traffic injuries can be prevented. Governments need to take action to address road safety in a holistic manner. This requires involvement from multiple sectors such as transport, police, health, education, and actions that address the safety of roads, vehicles, and road users.

Effective interventions include designing safer infrastructure and incorporating road safety features into land-use and transport planning, improving the safety features of vehicles, improving post-crash care for victims of road crashes, setting and enforcing laws relating to key risks, and raising public awareness.

### Role of Data Science

Analyzing a significant range of factors, including weather conditions, special events, roadworks, traffic jams among others, an accurate prediction of the severity of the accidents can be performed.

## Data

### Exploring Vehicle Accidents in France

- This notebook represents the exploration and visualization of the dataset Accidents in France Dataset.
- This data set consists of all the recorded accidents in France since 2005 to 2016.

### Reading Data

There are 5 tables and it will be a good idea to read, explore and analyses all the tables separately. To avoid this situation, we need to merge all the tables into one data frame based on the attribute representing accident ID i.e. Num\_Acc attribute. Once the data from 5 tables are merged we can check the dimension of our dataframe.

#### Shape of Dataframe

- Rows: 3553976

- Columns: 52

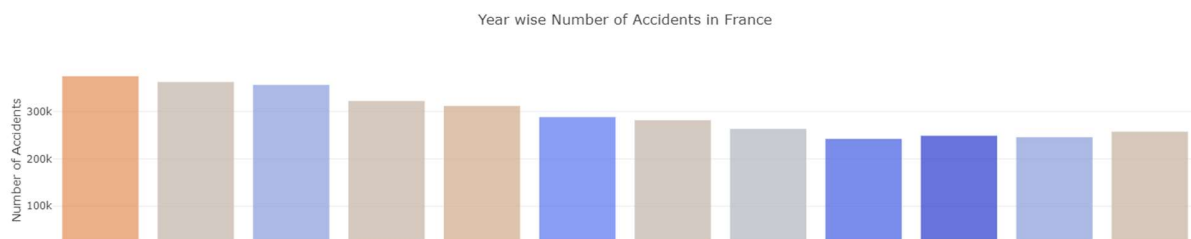
## Cleaning the Data

- We identify the percentage of NaN values in the dataset and eliminate the columns where most of the values are NaN
- Drop the column which are relevant for data exploration.
- Taking a dataset at a time, we have removed unwanted columns.
- Similarly, we have replaced the NaN with the highest value in the range.

## Exploration

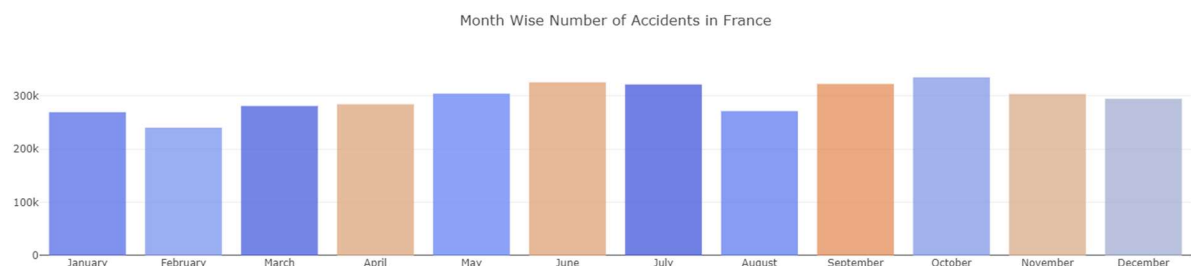
### Exploration based on date of accidents

*Trend in Accidents per year from 2005 to 2016*



- Number of accidents in France have steadily decreased from 2005 to 2013, but an increasing trend is observed from 2013 to 2016.
- In this dataset, 2005 had the largest number of accidents equal to 374561.

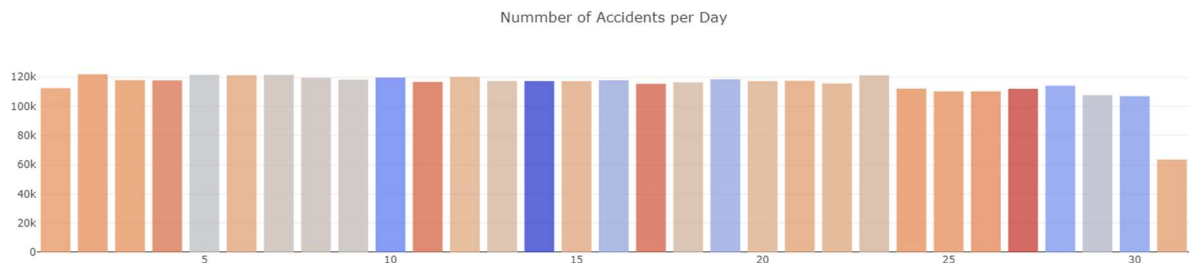
*Frequency of Accidents on monthly basis*



- June, July, September, October have the highest number of accidents, while February has the lowest.
- On an average about 296,164 accidents occur every month in France.
- October has the highest number of accidents (with about 334,884 incidents) than any other month

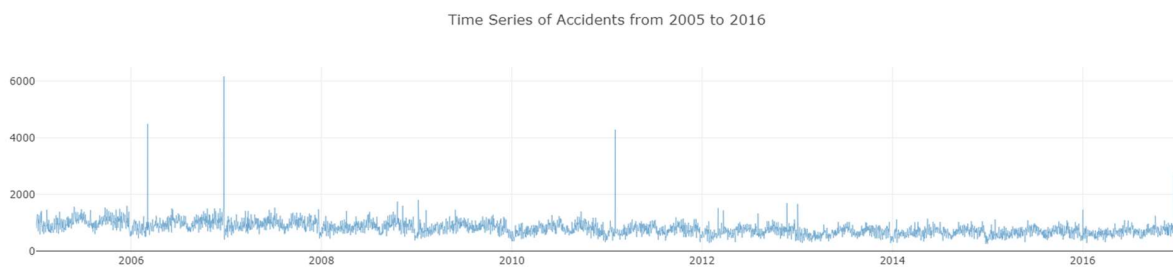
- Weather in France during September and October is cold and wet whereas, June and July form the peak tourist season

### *Safest day-of-the-month to drive*

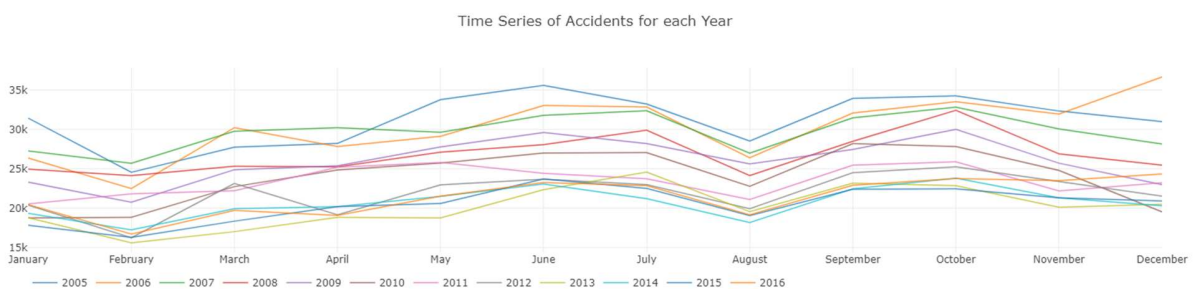


- Number of accidents per month-day is mostly uniform. 31st is lowest because only 7 months have 31 days.
- On an average about 114,644 accidents occur every day in France.

### *Time series of all accidents from 2005 to 2016*



### *Time series for all accidents in each year*

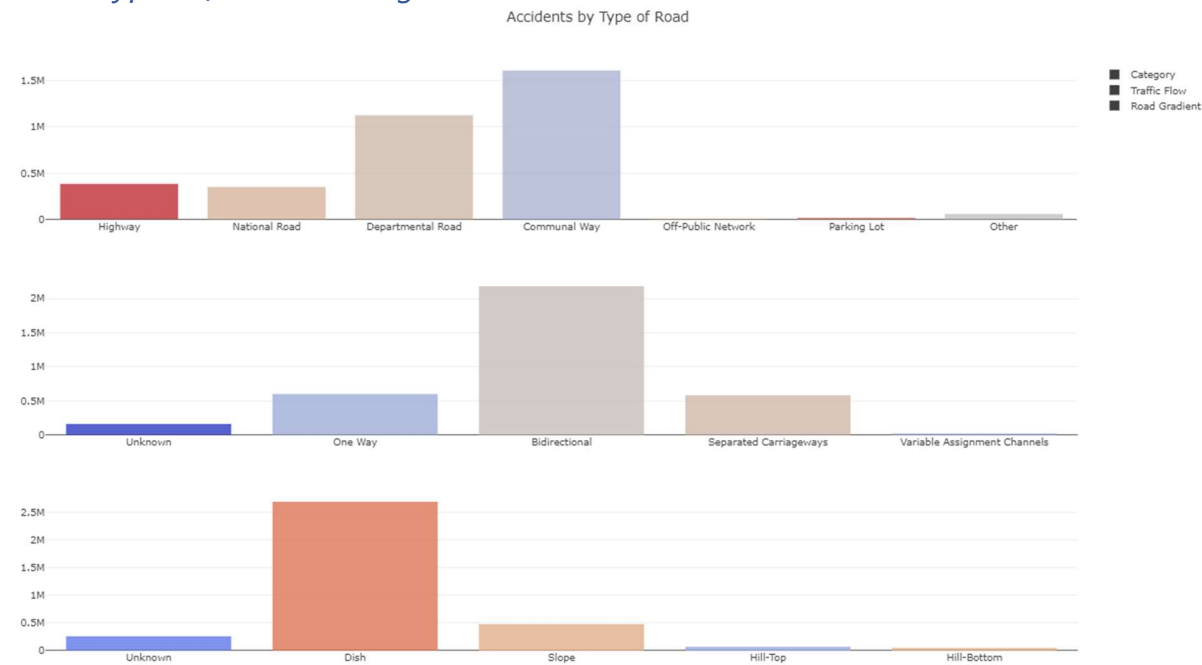


- A sharp rise is observed in the months of June, July, September and October.
- Sharp drops are observed in February and August.
- December 2006 has the highest number of accidents at 36,648.
- February 2013 has the lowest number of accidents at 15,605.

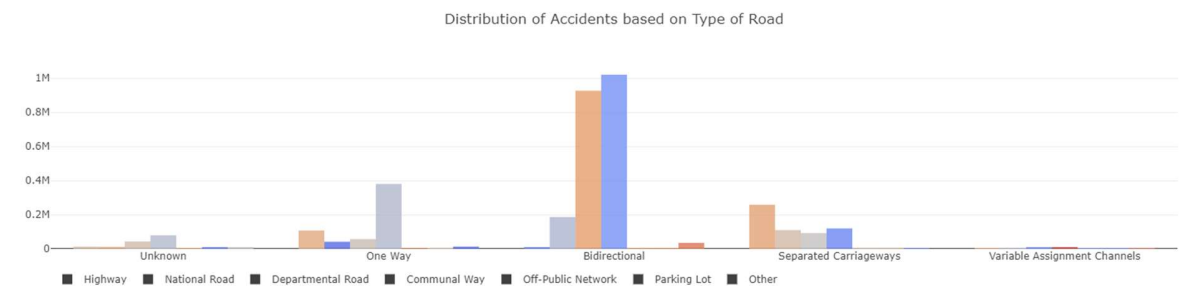


# Exploration based on roads where accidents occurred

Which types of roads are high risk?

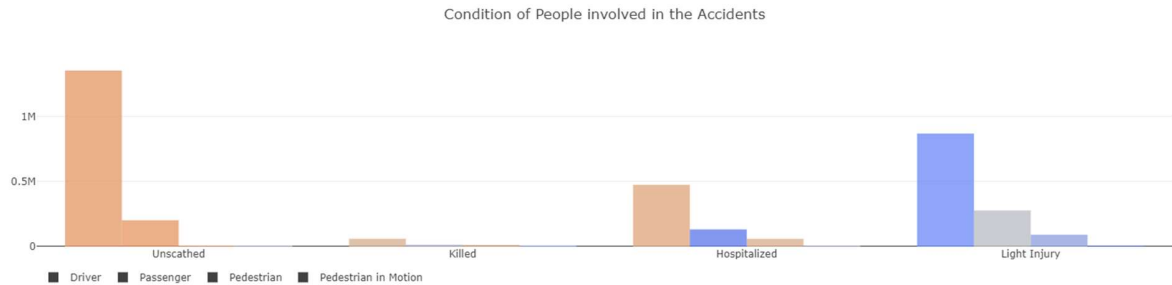


Which type of road gradient is high risk?

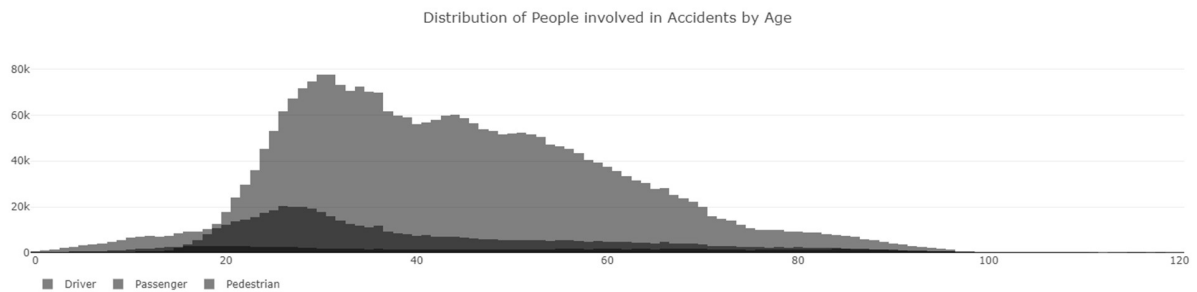


## Exploration based on people involved in the accidents

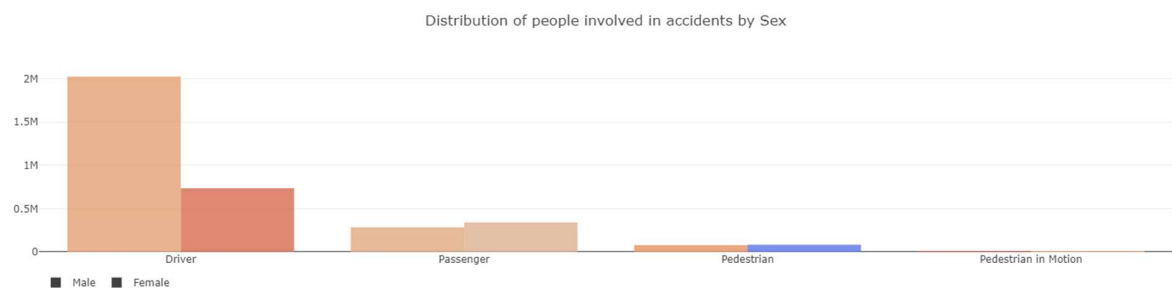
*What was the condition of the people after the accident?*



*What was the age distribution of the people involved?*

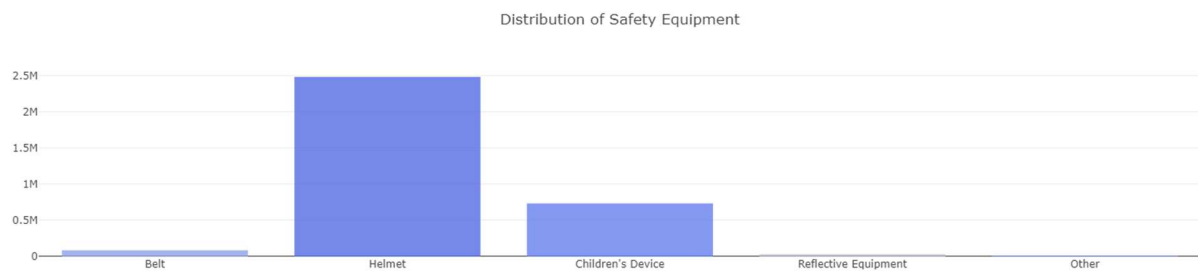


*What was the sex distribution of the people involved?*

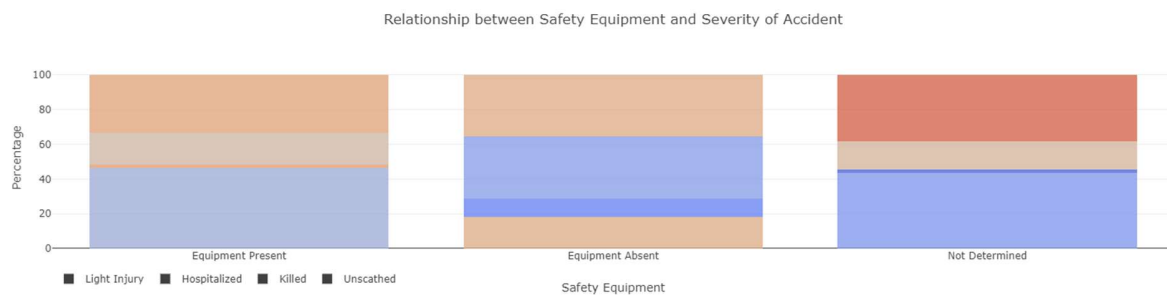


## Exploration based on use of safety equipment

*What was the distribution of Safety Equipment used?*



*Did use of Safety Equipment impact condition of people after the accident?*



- 1.9% of the people were killed and 17.8% were hospitalized if Safety Equipment was used.
- 10.7% of the people were killed and 35.8% were hospitalized if Safety Equipment was not used.

## Predictive Modeling

The regression model does not suit our prediction as our target is in binary form. Therefore, we have selected classification model, which can help us predict if the accident is sever or not. Given other information we should be able to predict the severity of the accident.

## Model

We have selected different model like KNN, Logistic Regression, Random Forest and Decision Tree. The Random Forest seems to be performing well overall.

## Solution

With the accuracy score to take into consideration, that the injuries will be only light. Similarly, knowing this information can save some resources that can be more helpful to save other people that need it.

## Performances of different models

Model	Target Value	precision	recall	f1-score	support
KNN	0	0.69	0.75	0.72	94315
	1	0.64	0.57	0.6	73682
LR	0	0.65	0.8	0.72	94315
	1	0.64	0.46	0.54	73682
DT	0	0.67	0.71	0.69	94315
	1	0.6	0.54	0.57	73682
RF	0	0.65	0.8	0.72	94315
	1	0.64	0.46	0.54	73682

## Conclusion

From the above data of different classification methods, it is evident that KNN is doing well overall.