MINOR PROJECT 2

Mid Evaluation Report - Spatial Data Mining Lab

Ambulance Allocation and Accident Data Analysis

Minor Project 2	23.03.2019
Spatial Data Mining Lab	19B16CS313

Kanishka Garg
16803012
Batch : B13

Akash Gupta
16103201
Batch : B15

Submitted To : Mr. Mahendra Gurve

INTRODUCTION

In this fast moving world where Cab, Food, Electrician all are just a click away, we tend to forget about taking use of this Technology in our Safety and Health. This project will focus on analyzing some trends in requirement of Healthcare Facility like Accident Prone areas. We will further plot the coordinates of the most appropriate place for planting an Ambulance Centre. We will be reading various Research papers for this purpose as mentioned below and will therefore try to bring up a good and unified solution.

So, our project was based on plotting the Ambulance Centres and doing Accident Data Analysis to make these Centres, more efficient.

DATASET DESCRIPTION

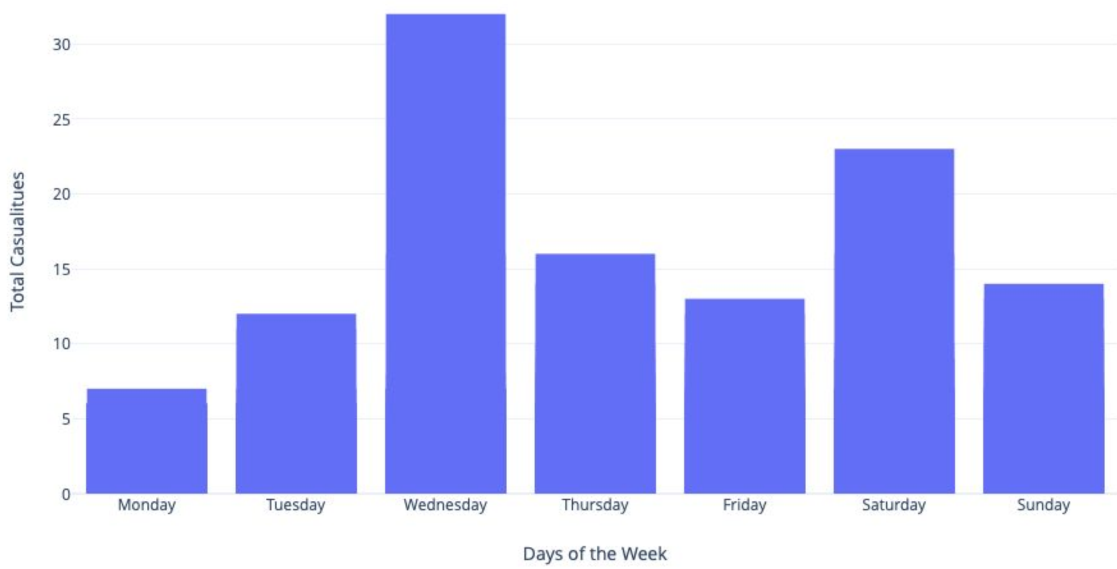
It is an Accident Dataset of Australia with the following features:

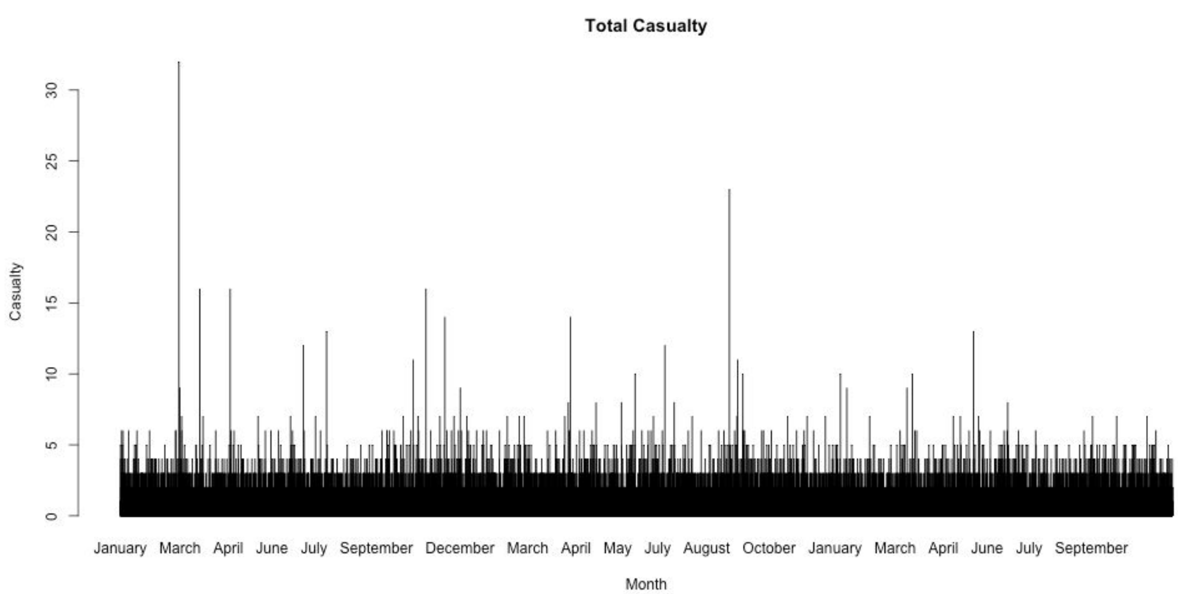
S. No.	Name of the Field
1	Crash_Ref_Number
2	Crash_Severity
3	Crash_Year
4	Crash_Month
5	Crash_Day_Of_Week
6	Crash_Hour
7	Crash_Nature
8	Crash_Type
9	Crash_Longitude_GDA94
10	Crash_Latitude_GDA94
11	Crash_Street
12	Crash_Street_Intersecting
13	Loc_Suburb
14	Loc_Local_Government_Area
15	Loc_Post_Code
16	Loc_Police_Division
17	Loc_Police_District
18	Loc_Police_Region
19	Loc_Queensland_Transport_Region
20	Loc_Main_Roads_Region
21	Loc_ABS_Statistical_Area_2
22	Loc_ABS_Statistical_Area_3
23	Loc_ABS_Statistical_Area_4
24	Loc_ABS_Remoteness
25	Loc_State_Electorate
26	Loc_Federal_Electorate
27	Crash_Controlling_Authority
28	Crash_Roadway_Feature
29	Crash_Traffic_Control
30	Crash_Speed_Limit
31	Crash_Road_Surface_Condition
32	Crash_Atmospheric_Condition
33	Crash_Lighting_Condition
34	Crash_Road_Horiz_Align
35	Crash_Road_Vert_Align
36	Crash_DCA_Code
37	Crash_DCA_Description
38	Crash_DCA_Group_Description
39	Count_Casualty_Fatality
40	Count_Casualty_Hospitalised
41	Count_Casualty_MedicallyTreated
42	Count_Casualty_MinorInjury
43	Count_Casualty_Total
44	Count_Unit_Car
45	Count_Unit_Motorcycle_Moped
46	Count_Unit_Truck
47	Count_Unit_Bus
48	Count_Unit_Bicycle
49	Count_Unit_Pedestrian
50	Count_Unit_Other

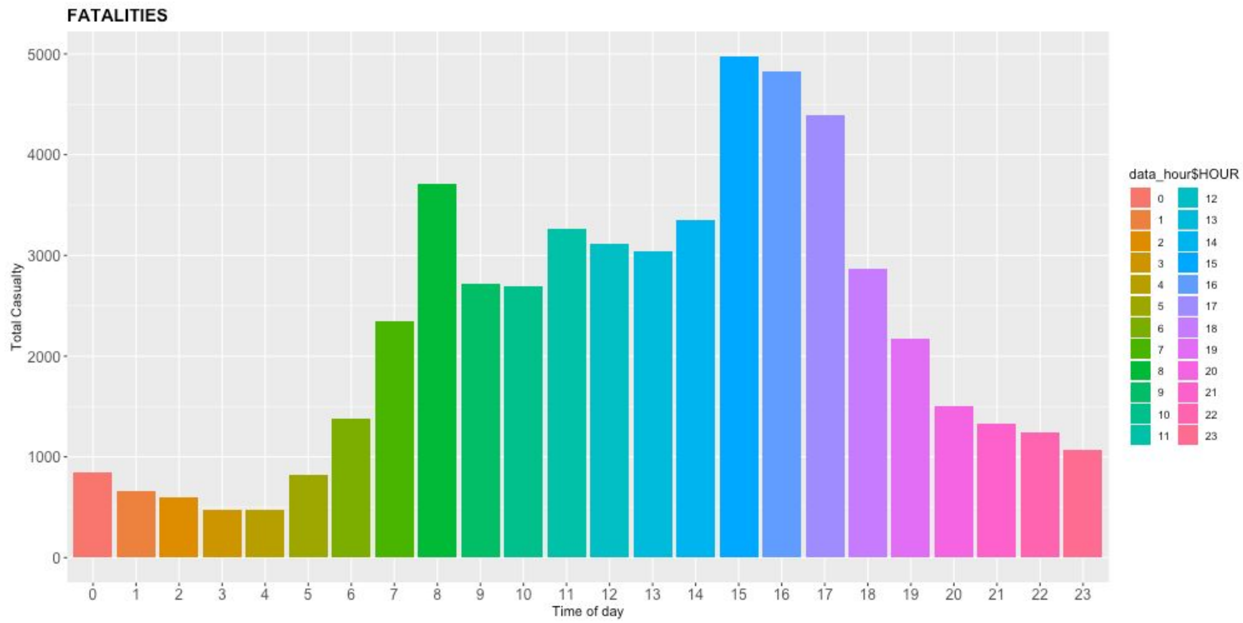
WORKING

Phase 1:

Visualizing of Dataset using R:







R Code:

```
data<-read.csv("locations_data.csv", header=T)
names(data)
##creating subset of data
accident <- data
names(accident)
#month
accident$Crash_Month <- factor(accident$Crash_Month
                          , levels=c(1,2,3,4,5,6,7,8,9,10,11,12)
                          , labels=c( "January","Feburary","March","April"
                                      ,"May","June","July","August"
                                      ,"September","October","November","December"))
#severity
accident$Crash_Severity <- factor(accident$Crash_Severity
                             , levels=c(1,2,3,4,5)
                             , labels=c( "Property damage only","Minor injury","Medical treatment","Hospitalisation"
                                         ,"Fatal"))
#Lighting Condition
accident$Crash_Lighting_Condition <- factor(accident$Crash_Lighting_Condition
                                   , levels=c(1,2,3,4,5)
                                   , labels=c( "Unknown","Daylight","Dawn/Dusk","Darkness - Lighted"
                                               ,"Darkness - Not lighted"))
#Lighting Condition
accident$Crash_Speed_Limit <- factor(accident$Crash_Speed_Limit
                                      , levels=c(1,2,3,4,5)
                                      , labels=c( "0 - 50 km/h","60 km/h","70 km/h","80 - 90 km/h"
                                                  ,"100 - 110 km/h"))
#Removing null values
accident <- accident[rowSums(is.na(accident)) == 0,]
png(file = "barchart_stacked.png")
barplot(accident$Count_Casualty_Total, main = "Total Casualty", names.arg = accident$Crash_Month, xlab = "Month", ylab = "Casualty", col="blue")
dev.off()
#####  Fatalities by hour  ##############
library(ggplot2)
library(data.table)
hour<-as.factor(accident$Crash_Hour)
data_hour<-data.table(hour,accident$Count_Casualty_Total)
data_hour<-data_hour[,sum(V2),by=hour]
names(data_hour)<-c("HOUR","fatals")
png(filename="hour.png" , width=1000, height=500)
ggplot(data=data_hour, aes(x=data_hour$HOUR, y=data_hour$fatals, fill=data_hour$HOUR)) +
  geom_bar(stat="identity")+xlab("Time of day") + ylab("Total Casualty") +
  ggtitle("FATALITIES")+theme( plot.title = element_text(size=14, face="bold") ,
                               axis.text.x= element_text(angle=0, size=12),
                               axis.text.y= element_text(angle=0, size=12)
                               ,legend.position="right")
```

Phase 2:

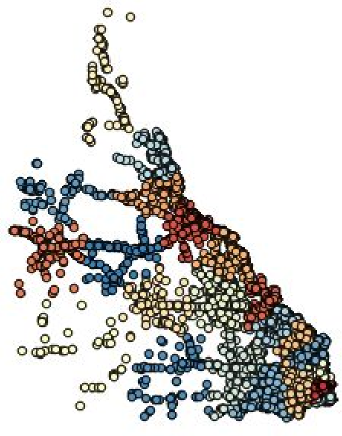
Applying DBSCAN and K- Mean Clustering to our data.

```
#Importing Libraries

#DBSCAN

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import scipy
from sklearn.cluster import DBSCAN
from sklearn import metrics

 #Importing Dataset

dataset = pd.read_csv('locations_data.csv')
X = dataset.iloc[:,[8,9]].values

db = DBSCAN(eps=0.3, min_samples=10).fit(X)

core_samples_mask = np.zeros_like(db.labels_, dtype=bool)

core_samples_mask[db.core_sample_indices_] = True
labels = db.labels_

n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)

print('Estimated number of clusters: %d' % n_clusters_)



#PLOT

import matplotlib.pyplot as plt

  # Black removed and is used for noise instead.
unique_labels = set(labels)
colors = plt.cm.Spectral(np.linspace(0, 1, len(unique_labels)))
for k, col in zip(unique_labels, colors):
    if k == -1:
        col = 'k'
    class_member_mask = (labels == k)
    xy = X[class_member_mask & core_samples_mask]
    plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=col,markeredgecolor='k', markersize=14)
    xy = X[class_member_mask & ~core_samples_mask]
    plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=col,markeredgecolor='k', markersize=6)
plt.title('Estimated number of clusters: %d' % n_clusters_)
plt.show()
```

```
#Importing Libraries

#KMEANS

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from collections import Counter

 #Importing Dataset

dataset = pd.read_csv('locations_data.csv')
X = dataset.loc[:,['Crash_Longitude_GDA94','Crash_Latitude_GDA94']]

id_n=23
kmeans = KMeans(n_clusters=id_n, random_state=0).fit(X)
id_label=kmeans.labels_

#plot result
ptsymb = np.array(['b.','r.','m.','g.','c.','k.','b*','r*','m*','r^']);
plt.figure(figsize=(12,12))
plt.ylabel('Longitude', fontsize=12)
plt.xlabel('Latitude', fontsize=12)
for i in range(id_n):
    cluster=np.where(id_label==i)[0]
    plt.plot(X.Crash_Longitude_GDA94[cluster].values,X.Crash_Latitude_GDA94[cluster].values,ptsymb[i])
plt.show()

#Adding to the main dataset

import csv
rows=[]
fields=[]
with open('locations_data.csv','r') as csv_input:
    csvreader= csv.reader(csv_input)
    fields=next(csvreader)
    for row in csvreader:
        rows.append(row)
fields.append("CLUSTERS_with_KMEAN")
i=0
for row in rows:
    row.append(id_label[i])
    i+=1

with open('locations_data.csv','w') as csvfile:
    csvwriter=csv.writer(csvfile)
    csvwriter.writerow(fields)
    csvwriter.writerows(rows)
```

Phase 3:

Visualizing The Results of our Clusters in QGIS.



Phase 4:

Read some Research Papers and Working on Possibilities of:

1. Spatial Decision Tree (ID3 or CART)
2. SPODT: Spatial Oblique Decision Tree in R
3. Hidden Markov Model on Accident Data Analysis

REFERENCES (RESEARCH WORK)

1. Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques

This paper follows the Approach of:

Statistical Analysis + Associative Rule Learning + Classification + Clustering on Accident Data.

2. Geographical Information System for Mapping Accident-Prone Roads and Development of New Road Using Multi-Attribute Utility Method

This paper follows the Approach of:

Multi Attribute Utility Theory with Spatial

3. Analyzing the Road Traffic and Accidents with Classification Techniques

This paper follows the Approach of:

Comparing many Classifiers like:

Naive Bayesian Classifier, Decision Tree Classifier, AdaBoostM1 Classifier

4. A Decision Tree for Multi-Layered Spatial Data

This paper offers an Algorithm for Spatial Decision Tree with CART (SCART)

5. A Spatial Entropy-Based Decision Tree for Classification of Geographical Information

This paper offered the approach of Spatial Decision Tree in 2006.

6. Spatial Decision Tree for Accident Data Analysis

This paper by IITK students applied the approach of above paper on Accident Data.

7. SPODT: An R Package to Perform Spatial Partitioning

This paper introduces a built-in R package for Spatial Oblique Decision Trees.

8. Crash Detection System Using Hidden Markov Models

This paper offers Approach of discrete Hidden Markov Models on real time accident Data using Jack-Knifing Cross validation.

9. Using Hidden Markov Models in Vehicular Crash Detection

This paper has presented a methodology for building a crash detection system using continuous-mode HMMs and has established the proposed strategy to be a robust methodology for early detection of automotive crashes.

10. Traffic Incident Prediction on Intersections Based on HMM

This paper uses HMM on input from Video Cameras at traffic Intersection to improve the detection rate of accidents, ease traffic jams, and reduce traffic accidents