IBM

APPLIED DATA SCIENCE CAPSTONE - DATA COLLISIONS

1)INTRODUCTION

I examined data before suggest to business problem. Fill NaN values with correct values that the way I think it is right. This infos been in own cells. Data has accident by Traffic Records. I evaluate here how accident is happened, how many people dead or get injured, how many vehicle included accident, collisions type of vehicles, as so many parameters in that data. There are some factors about driver that get drunk or alcohol. Weather, road and light condition about accident. Location of accident. All that parameters will be examined in detail. If an accident occur anywhere , data has information about there. I said "there" is that mean all info about cars, peoples,weather condition namely all factors may affect the accident.

2)BUSINESS PROBLEM

where did the accident occur ? why did the accident happen ? which factors affect the accident weather, light or road condition or all of them ? was driver drunk? was driver too speedly? I think ours target is that questions answer. I want to see factors that affect the accident. Collisions type and collisions address is interest each other. If interest , how we explain this ? which time interval is have high accident rate and why is that? then maybe find some solution about accident problem. Maybe it is road problem or psychological problem or we don’t discover yet that any new problem that we will discover with this data.

3)DATA

Data include accident that have some factors about driver, pedestrian, road cond, weather cond, road type, collisions type, accident time, location, severity, injuries, dead. How occur accident and its explanation as briefly. I did some adjustment about unknown situations. I will do some feature engineering and add some new attributes in that data. Maybe detect the location in that occur accident frequently. And below I explain data in detail and do some changing about missing values and thought to be wrong values.