I will use R and Python to analyse the dataset.

For the intelligent data analysis ,as there are two situations possible in which intelligent data analysis can help us to find solutions to certain problems or provide

answers to questions that arise.

In the first case the problem at hand is by no means new ,but it is already solved as a matter of routine.

In second case a certain question arises for the first time and only little experience is available or the experience is not directly avialable to this new question.In

such cases,it is supposed that data from related situations may be helpful to generalize the new problem or that unknown relationship can be discovered from the data to

gain insights into this unfamiliar area.Here is the second situation situation for my project.

In this project(Aim is to find a pattern so that maximum passengers might be kept alive),we are going to follow the CRISP-DM proposal which is

about what the intelligent data analysis process should look like.

CRISP-DM consist of six phases.The phase are in order as: project understanding then data understanding then data preparation then modeling then evaluation then

deployment. We have graphical representation of phases in word file.

1.Project Understanding: This is often called as bussiness understanding.This event happened in 1816 a.d. which was quite threatening and happened while travelling

in Titanic ship.There were many reasons behind this unwanted event.I am interested to know that what might be pertaining situations at that time which caused the death

of passengers.All told I am interested in finding a typical pattern which could lead for passengers in order to keep them alive.

So from where should I get such promising pattern? Should I stick to pattern if I get that?

can I generalize that pattern over other passengers in order to get that how many of them could have been survived and how many of them could not?

Ultimately,because of these sudden arised questions,I can state following:

(a)Project Objective is to find a typical pattern in the attributes of passengers so that it can be generalized over other passengers in order to predict that how many

of them could have been survived.

(b)Other objective should be that which attribute is best describing the death or alive status of passengers.

2.Data Understanding: I will approach the data from a neutral point of view as one should do.Since we must not trust any data as long as we have not carried out some

simple plausibility checks.So we shall do some plausibility checks later in this phase using particular methods.Since this phase requires taking a closer look at the

data that is why we follow:

(a) Attribute Understanding : So for we have considered the aim of project.But in order to proceed further we need data.Fortunately we have data.

\*\*Talking about data: I have two datasets. One the train data and other is test data. I will emphasise on train data only.So what about train data?

> dim(titanic)

[1] 891 13 So our dataset contains 891 rows and 13 columns.It's mean that there are 12 features and obivously 1 target variable.The variables are:

"PassengerId" "Survived" "Pclass" "Name" "Sex" "Age" "SibSp" "Parch" "Ticket" "Fare" "Cabin"

"Embarked" "Ag\_grp". The target variable being "Survived".

So now we come to main topic of attribute understaing.