**BUSINESS PROBLEM FRAMING**

**Objective 1: For the open data source selected, define the problem to be addressed**

The problem area is Crime Investigation. Basically, we will try to determine which factors affect the crime happenings and which geographical areas of the US are safer.

**Objective 2: Identify the stakeholders**

In this data mining project, we will be responsible for the entire project and FBI web-site will support us the crime rate data set, US Census official web-site will support us the census data set and Bureau of Economic Analysis web-site will support us the personal income data set.

**Objective 3: Determine whether the problem is suitable for an analytics solution**

I think that this problem is suitable for an analytics solution because it has historic data and by using some data mining algorithms with the help of analyzing the historic data with some variables like gender, age, sex and personal income, we can reach some good predictions of the safest places and the factors affecting the crime happenings.

**Objective 4: Refine the statement of the problem with any constraints**

“How to prevent a crime?” is a good question for this data mining process. Before coming up with a solution of how to prevent it, factors that trigger a crime should be extracted firstly. In this project, we try to come up with a solution of finding the factors that make a crime occurred. We also try to find the safest places in the US according to the historic data. The variables that we will be using during data analysis are age, sex, gender and personal income. This process would be limited because we have the data sets that are recorded in the years of 2010-2013 only. So the constraints would be time period, the statistics’ accuracy and limited data sets. This project may be extended by having a longer time-series data sets, more accurate data sets taken from the governmental departments by requesting them by a petition or a protocol and more detailed information like education, social and economic situation factors etc. that we do not use in our project.

**Objective 5: Define an initial set of business benefits**

We will use three data sets which are crime rate data set taken from FBI web-site, census data set taken from the US census official web-site, personal income data set taken from the Bureau of Economic Analysis. We will use Rapid Miner tool as a data-mining tool. We will implement Naive-Bayes or K- nearest neighbor data-mining algorithms to the training data sets. On the evaluation stage, we will get the accuracy and use the Laplace estimator. During all those stages and getting reports, we will use Rapid Miner data-mining tool.

The Project plan will include;

1. Get the open data sources.
2. Prepare the data understanding document.
3. Prepare the data sets. Cleaning and organizing data is included in this phase.
4. Modeling. Selecting models and implementing data-mining algorithms in Rapid Miner tool to the cleaned data sets are included in this stage.
5. Evaluation. Evaluating the results of accuracy is included in this stage.
6. Deployment and getting reports.