Project Proposal Title

(A concise and meaningful title for the application, your name, student number)

Ahmed Arif Khan

1324501

Computer Science

2XB3 L01

• Abstract:

- describes the highlights of the proposal (the problem, the solution, the dataset and experiments for verification and validation). (max: 150 words) Hint: you will write the abstract after writing your proposal!

• Objectives and scope of the project:

- concisely describe the problem that you are interested to solve. You should include a background study and the context of the application. You can use symbolic notations or figures to describe the problem. Then you will describe your tasks (what your implementation will accomplish) to address the problem. You will also include in this section your motivation for selecting this topic. You will describe why you think it is important to solve this problem. (max: 500 words)

There is a vast amount of open data publicly available on the Internet available for everyone to use except the problem is that it is inconvenient to find and work with. It takes a long amount of time and energy to find all the data that you need. After collecting a large amount of information it gets very complex to help analyze the data. This application will fix his problem by gathering all the information in one place. Then

￼￼￼

• Input/output:

- describes the type of input datasets (and the source of the dataset) that you will be using to evaluate your application and the type of output that you will be generating. (max: 200 words)

The input will be large statistical information of Canada such as crime rate, income, and health. This information will be retrieved from Canada’s open data websites. When the user goes into our application there will be two main features, explore and analyze. In analyze the user can pick any number of data sets such as hand gun violence and graph it with respect to time, age, gender, geographical location and so on. The other feature will be called explore. In this the computer will chose any topic. It will find the distribution of the topic by graphing it with respect to all the groups such as gender, age, and wealth. It will then find the correlation between that specific information and group and when the correlation becomes to high it displays it to the user in the respective graphical format. The program could graph the information in any desired chart for example pie graph, bar graph, line graph and so on. This will allow the computer to do the mindless information management so the user can go and analyze it and figure out why these trends exist and help prevent them for the future.

• Algorithmic challenges of the project:

- describes which algorithms (sorting, searching, graphs, strings, etc.) you think you will be implementing to solve this problem and what are the challenges that you should be aware of. (max: 200 words)

Visually graphing the information should be straightforward but going and finding the information that is distributed unevenly may get complex. We are going to be dealing with very large sets of data so algorithm complexity will have a big impact on its speed. First the program will need to sort all the categories of information with respect to the quantity of information available and return them in decreasing order. Then it will need to graph the data and find its correlation with respect to many different variables. --------------------------------------------

• Project timetable:

- describes the timeline of the project with a list of milestones, and deliverables. The list should not exceed 7 milestones. You should use a table or a gantt chart to describe your timetable. (max word: 150 words)

**General Layout**: Make a more defined input and output diagram. Make a sample diagram of how the view is going to look on paper. Make a general layout of the amount of classes and how they are going to interact with each other, make the test files for all the different parts of the program, build sample data to test with.

**Build The View**: Make the view, meaning the code that displays the menu the different graphs (for now it will just display sample data)

**Build The Model**: This will be making the part that graphs the data in many different ways and finds the correlation between two variables until it finds two variables that have a strong correlation.

**Make the controller**: get a large sample of information and place it into our program.

**Finish Up**: Combine all the different parts, look at the code, try it out with people, make small tweaks to make sure its working.

• References:

- use a consistent format (IEEE, APA, etc.) for your references.

4 page long

There is alot of open data online, It is available on the internet for everyone to access except the problem is, its just inconvenient to find and work with. What we are going to make is a program that lets you use and analayze all that data in a more visual formal and workable format.

- categorize the data according to age, ethinicity, gender, for things like internet time spent, disease,

- sell it to marketers

- you can predict future crime, you can help the future

- find disease groups, help prevent disease by finding where its popular,

- figure out why crime rate is higher in certain people

- if an age group have high suicide rate figure out why.

- monitor new born disease

IDEA: take the open data from the government and categorize it, After categorizing it map it with respect to age, gender, wage, location. it finds the correlation of the data with that group. If the correlation is to high it then gives you that data and graph for you to analyze. So you can figure out why and help find a solution.

For example: Lets say it takes in the crime and it graphs with respect to location, graphs it with respect to age, gender,

- There is alot of open data freely available for use but it is very overwhelming to use. To get something as simple as crime rate can be retrieved from the government websites but takes an unnecessary amount of effort, patience, and manual labor. It is a hassle to even get a simple data set let alone start to analyze it.

- We will make a website that will allow you to easily sort and easily manipulate that data. Firstly it will be able to take all the data from all the different government websites and combine it into one simple place for you to access.

- The next step will be to sort that data with respect to age, gender, wealth, ethnicity, race, culture, geographic location and so on. If it finds any two variables to be very closely correlated it will show the user the result in the appropriate graph.   
- You could also manually chose which data sets you want to compare with others

-

IDEA: