INFM600 Information Environment

Team Exceptional

Work Plan

October 5, 2016

Overview

The Work Plan is intended to serve as a framework for our project. The purpose of the document is to demonstrate the different stages of our project, the goals and guidelines we intend to set for our team. It will help us keep a track of our progress and check if we follow our estimated deadlines for each stage.

Project Goal

The main goal of this project is to understand the end-to-end workflow while working with data; right from the seeking the data to validating it to using it to come up with some information that helps in the decision-making process for the target audience. Throughout the course of the project, we also aim to become more familiar with working in a hybrid team as well as certain tools and technologies used for information management.

For this project we will be working with raw datasets and using that we will try answering some of the related research questions. The details of the datasets, research questions and target audience are mentioned in the following sections.

Data Description

The main dataset that we would be working on for our project is the DC Metro Crime Rate [1]. This dataset contains all possible details about the crimes that have occurred in the State of Virginia. Details like when the crime started or ended (right down to the extract timestamp), where it occurred (location is described using cluster number and ward), what kind of crime it was, the weapon used (if any) for the years 2011-2013.

Research Questions

Following are a few questions that we aim to answer with the abovementioned dataset and a few other relevant datasets:

1. Does the change in seasons have an effect on the crime rate?
2. Is there a relation between the timings of bars/clubs with the rate of sexual assaults?
3. Is there any impact of the availability of guns on the crime rate?

For question one, we do not require any supporting data apart from the main dataset of DC Metro Crime. For question two, we have figured out a source to get the data regarding the location and timings of bars and clubs located in Columbia. We will be scraping this data from Yelp website [2]. For question three, we will scraping the data from FFL Gun Dealers website [3]. This website provides the address of each gun store in a particular area. But it would be a tricky task as the store details are organized by the zip-code and the crime rate by clusters.

Target Audience

 The target audience for our analysis is as follows:

* General public of the District of Columbia
* Police department of the District of Columbia
* Government

Projected Timeline

Table I has the project schedule for the coming 11 weeks, for our team. Every week, we plan to have a meeting to chalk out plans for the upcoming stages in the project. In the Gantt chart (see: Appendix), blue parts represent each phase/stage of the project, while green parts denote the specific work plan corresponding to that project phase what has already been done.

**Table I: Projected Timeline**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** | **Project Stage/Phase** | **Deliverable** | **Start Date** | **End Date** |
| 1 | Figuring out research questions and coming up with a project plan | Work Plan | 9/28/2016 | 10/4/2016 |
| 2 | Data Collection | - | 10/5/2016 | 10/8/2016 |
| 3 | Data Cleaning and Validation | Data Cleaning and Documentation | 10/9/2016 | 11/1/2016 |
| 4 | Progress Meeting | - | 10/9/2016 | 11/1/2016 |
| 5 | Analysis using R (I) | R Script (Draft) | 11/2/2016 | 11/14/2016 |
| 6 | Analysis using R (II) | R Plot (Draft) | 11/15/2016 | 11/30/2016 |
| 7 | Final Analysis and Presentation | Git Package | 12/1/2016 | 12/14/2016 |

We plan to use Google Drive to let multiple group members compile one file at the same time. Other collaborative tools that we plan to use include Slack and Github. Slack, to maintain communications among group members and Github, to store all the files of our team.

In all the timeline plans, we have set the assignment deadlines one day before the due date set by the course instructor. This way, if we encounter any problem, we have enough time to consult the instructor and make the necessary changes.

At present, we still lack the specific sub-tasks classifications for the upcoming project phases apart from the first one as we plan to analyze and discuss one phase at a time. As the weeks progress, we will keep updating the Gantt chart as the tasks for each phase becomes clearer.

Effort Allocation

Table II defines the tentative effort-allocation for the course of this project. We want to make use of the specific skills that an individual team-member has along with ensuring that the work is equally distributed amongst the group members. Table II provides a tentative effort allocation plan for our team, which may be altered as the project progresses.

**Table II: Effort Allocation Plan**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** | **Project Task/ Phase** | **Pal** | **Rahul** | **David** |
| 1 | Figuring out research questions and coming up with a project plan | Research question formulation，Work Plan first draft | Research question formulation，Work Plan final draft | Timeline plan,  Gantt chart, Research question formulation |
| 2 | Data Collection | Collect Yelp data for bars/clubs | Collect data related to the gun stores | Collect more detailed data for the DC Crime rates |
| 3 | Data Cleaning and Validation | Apply data cleaning techniques on the Yelp dataset | Apply data cleaning techniques on Gun Store dataset | Apply data cleaning techniques on DC Metro Crime rate dataset |
| 4 | Progress Meeting | Will be present for meeting | Will be present for meeting | Will be present for meeting |
| 5 | Analysis using R (I) | Conduct analysis for Question 2 using R | Conduct analysis for Question 3 using R | Conduct analysis for Question 1 using R |
| 6 | Analysis using R (II) | Same as above | Same as above | Same as above |
| 7 | Final Analysis and Presentation | TBD\* | TBD | TBD |

\*TBD- To Be Determined

**Word count: 977**

References

[1] DC Metro Crime Data | Kaggle. (n.d.). Retrieved October 3, 2016, from <https://www.kaggle.com/vinchinzu/dc-metro-crime-data>

[2] <https://www.yelp.com/search?find_desc=Nightlife&find_loc=Washington,+DC>

[3] <http://fflgundealers.net/maryland-zip-codes.html>

Appendix:

We will be using a systematic color scheme to denote the stage/task for each phase in the project.

For example：The blue parts represent each phase of the project, while green parts denote the specific tasks corresponding to that project phase.

**Figure I: Gantt chart**

****