COMP6235 Group Coursework – Individual Feedback Form

|  |  |  |  |
| --- | --- | --- | --- |
| **Module:** | *Foundations of Data Science* | **Lecturers:** | *ES, CP, RT, MB* |
| **Assignment:** | *Group Project* | **Weight:** | *70%* |
| **Student Name** | Am5e15 | **Student ID:** |  |

**1) Please summarise your group project (100-200 words):**

The core idea of our project was to use the crime dataset available to gain insight into crime. After cycling through some ideas we came to the concept: To relate the coordinate data of the crimes to a geographic dataset. The geographic dataset represents points of interest which we believe may be correlated to crimes and then study that correlation. To produce that into a visualisation which has an end user application we wanted to visualise the individual effects of the points of interest on the background crime rate. We chose trains, schools and banks as our points of interest. Calculated which crimes were associated with each point of interest and created various statistics of those points of interest taking into account areas, people using the point of interest (such as entry and exit figures for stations), background crime rates of the borough and applying gaussian weightings to the distance from the stations.

**2) What were the main challenges of the project and how did you and your group overcome them? (100-200 words)**

The main challenges were the idea. The data: both the quality/availability and working with big data. We cycled through many different ideas before finalise on an idea which the data allowed us to investigate. This was due to granularity issues. Once the idea was solidified I found it fascinating dealing with big data for the first time. I spent over a week trying to get my R code to process the distances of crimes from stations to a running time from 35days down to 11 hours, I overcame these obstacles using faster structure in R such as data.tables and optimising my code in apply functions but also doing these operations in memory as part of the preprocessing. (Unfortunately I don’t know fast programming languages). Finally we encountered problems with the visualisation, there were a lot of technical problems in this field I am not fully familiar with the concepts of such as how individual layers impacted our visualisation and other things.

**3) Please reflect on your contribution to the project: (100-200 words)**

My first major contribution to the group project was the idea. I felt like I had the best conceptual understanding in our group of the different ideas and the practicalities of these ideas. In the end the final idea was something I came up with and agreed on with Ramine in a group meeting.

My other contribution to the group was the data processing and modelling. We split the data by the different datasets we were planning on using (1 person doing each banks, schools and trains). I wanted to provide an example of the schema for the visualisation as early as possible. Because of this I designed scripts for data cleaning and processing onto mongodb as well as calculating the modelling statistics such as distance and crime counts / densities.

My role was primarily the ideas guy having a conceptual understanding of the project as well as frontiering the data processing and modelling so others could follow my lead. In meetings I took a back seat writing the minutes and ensure everyone was on the same page.

**4) Please reflect on the contribution of your team members. How well did you work as group? (100-200 words)**

I believe in some aspects we worked very well as a group. For 4 out of 5 members there was very high participation in very regular skype meetings (every 2 -3 days over holidays). There was good communication and clear breakdown of areas of responsibility. There was also a strong sense of “standing on the shoulders of giants” where work wasn’t repeated unnecessarily and people helped each-other on other people’s tasks or used code from other people’s tasks. Having said that there was an individual in our group who defied all of these things doing very little work, attending 11 / 18 documented meetings, agreeing to do something and then doing something else. All of this person’s work was either not able to be used or heavily altered by other members of the group.

5) On the scale of 1 -5, please indicate how you felt the group performed (1 low, 5 high)?

**3**