Data Management Plan for Research Students

1. Overview

Researcher: Mihir Gujarathi, Sofia Sole Ferrarese
Project title: Distributed Data Analysis Project on a Spotify dataset
Project duration: 2 months
Project context: Streams prediction based on a Spotify dataset using ML and HCPI methods.

2. Defining your data/research sources

2.1 Where will your data/research sources come from? The dataset has been retrieved from Kaggle
2.2 How often will you get new data? We don't expect to get new data
2.3 How much data/information will you generate? Less than 50MB
2.4 What file formats will you use? The user is expected to use any software to unzip the file uploaded in WiseFlow. Then, to execute to analysis, the user should have RStudio or any other software which can run R scripts to run our code.

3. Organising your data

3.1 How will you structure and name your folders and files?

All the files are inside a unique folder, comprising the dataset and the analysis

3.2 What additional information is required to understand each data file?

The user is expected to find only two files in the folder, the RMarkdown file, and the dataset.

3.3 What different versions of each data file or source will your create?

There will be two versions. They will be equal up to EDA section.

4. Looking after your data

4.1 Where will you store your data?

In a folder inside my laptop, as well as in the cloud

4.2 How will your data be backed up?

Every time a section of the project is concluded the Rmd file will be shared by email between the two researchers. In this way we avoid any big loss of analysis execution.

4.3 How will you test whether you can restore from your backups?

We will download every time the updated file from the attachments in the emails and start working on that file for the consequent analysis steps.

5. Sharing your data

5.1 Who owns the data you generate?

By both the group members

5.2 Who else has a right to see or use this data?

Everyone can access the data from Kaggle and use it accordingly

5.3 Who else should reasonably have access to this data when you share it?

University

5.4 What should/shouldn't be shared and why?

There are no ethical restrictions as the dataset is publicly available on Kaggle

6. Archiving your data

6.1 What should be archived beyond the end of your project?

We are not sharing our files publicly, but for an assessment
6.2 For how long should it be stored?
6.3 When will files be moved into the data archive/repository?
6.4 Where will the data be stored?
6.5 Who is responsible for moving data to the data archive and maintaining it? The University
6.6 Who should have access and under what conditions? Staff Members at Brunel University to evaluate the Assessment

7. Executing your plan

7.1 Who is responsible for making sure this plan is followed? Group members
7.2 How often will this plan be reviewed and updated? In weekly meetings
7.3 What actions have you identified from the rest of this plan?
7.4 What further information do you need to carry out these actions?

Notes on completing this form

- o Type as much (or as little) as you feel you need to into each box: it will expand to accommodate what you write;
- o You can leave or remove the prompts in grey once you're done;
- o For help with completing this DMP, please contact researchdata@brunel.ac.uk