Project Plan of the Final Paper

for the CEU MSc in Business Analytics program

Gabor Horvath

2017

High level description

The goal of the project is to show - creating a risk evaulation of wildlife stikes of flights in the US - the techniques, methods, interpretations and understanding of the data analitics. The project is based on the Cross Industry Standard Process for Data Mining (CRISP-DM) process model, which is widely used worldwide for various scientific and business related data analytical projects. The use of the CRISP-DM process model will enable to for the project to cover all those areas (i.e. Business Understanding, Data understanding, Modelling, Evaulation, etc.), which are crutial of managing and delivering a successfull data analitical project.

Resource needs

Training requirements

No additonal organized / official training requirements are required above the trainings received during the courses in the program. There are tools and techiques used to fulfill the project which have not been described in the program at CEU, but there are several usefull user manuals available on the webpages of the toold creators, which would enable the use of these tools and resources for any student who have been part of the program.

Tools & resources used

Fulfilling the completition need for the project the following tools are planned to be used:

- Programming language:
    - R
- GUI for the programming language:
    - RStudio
- Documentation (including this initial project description) is created using:
    - knitr
    - ReporteRs
- Data visualisation:
    - ggplot2
- Project plan / task management:
    - Buckets
- Source code repository:
    - GitHub

Note: The list above do not contain the list of all the packages used to create the project, but the full list will be provided in the source code.

Data sources

The project will use the following data provided by multiple US government agencies:

- Federal Aviation Administration - Wildlife Strike Database
- United States Department of Transportation - Bureau of Transportation Statistics

Note: In case data enrichment would be required for the sucessfull risk modelling, additional data sources might be used as well. These possible additional data sources will be listed in the Final Paper.

Restrictions

Restrictions apply as per the restrictions set by the tools, data providers and owners of additional resources used. No addtional restrictions have been identified and set regarding the use of the results of this project.

Contributors

Student: Gabor Horvath

Mentor: Gergely Daroczi