

**MSG 500**

**Individual project report**

***Investigating alcohol consumption and driving behaviour and their relationship to risk taking and impulsivity***

**Fionn Delahunty**

Applied Data Science Program

[Gusdelfi@student.gu.se](mailto:Gusdelfi@student.gu.se)

09/01/2018

**Contents**

|  |  |
| --- | --- |
| **Introduction** |  |
| **Methods** |  |
| **Results** |  |
| **Discussion** |  |
| **Appendix** |  |

Note: Although this project is written in the general tense (we), I am the only contributor to this project.

**Introduction**: Main goal of the analysis, describe the data and problem

**Methods**: Traditionally, this is where you would give the theoretical descriptions of models - you don't need that here. Highlight the methods you used and why.

**Results**: Use subsections here! Examples below (but you can use others).

a) Preliminary data exploration (where you look into variable transformations, missing values, etc). Note - not pages and page of this stuff. If you had to do a lot to your data, put a summary here and the rest in an Appendix

b) Model building. Overview of the full model fit etc.

c) Model selection or model comparison (variable selection, comparing CART/Regression etc).

d) Your special topic: e.g. stability analysis, sensitivity analysis, comparing two different data, etc etc etc

**Discussion**: Summarize your main findings. Interpretation. Predictive performance. Limitations and Concerns. What would future work entail?

Appendix: Your data source URL, Your codes + other stuff (see above).

**Introduction**

The Dublin science gallery is a free public museum based in the capital of Ireland and hosted by trinity college Dublin. The purpose of the museum is to enhance publics scientific knowledge through interactive and fun exhibitions. Research teams can submit different exhibitions based on the current theme of the gallery. During 2015 the gallery hosted a risk-taking theme, as part of this exhibition participants where offered the ability to fill out two psychometric risk-taking questionnaires and discover their own risk-taking attitudes. A total of 1326 individual participated in the questionnaire over the two months it was left on display.

In total participants where asked to complete 93 questions. The first 10 of these questions related to demographic characteristics. The next thirteen questions related to participants driving behaviour while the following eight questions related to their alcohol consumption behaviour. Participants then completed the 30 item Barratt Impulsiveness Scale (BIS) in its standardised format and the revised 30-Item Domain-Specific Risk-Taking (DOSPERT) Scale in its respective standard format.

Since the data was collected no analysis has been completed on the dataset. This report is the first primary analysis of the raw dataset. We will discuss the implications of working on such a raw dataset later in the report. For the purpose of this assignment we investigated four different hypothesises related to this dataset. These hypotheses are roughly based on existing psychological research on DOSPERT and BIS scales, but are not an extension of any specific published study. Rather they are designed for the purpose of self-teaching for the MSG500 module.

The four hypotheses are follow.

1. Predicting the number of cigarettes smoked per week from a measure of risk taking in a social context (Domain-Specific Risk-Taking (DOSPERT) Scale)
2. Predicting the number of car crashes from a measure of self-control (Barratt Impulsiveness Scale (BIS))
3. Building a prediction model to predict if a participant in the dataset has smoked during their lifetime.
4. Building a prediction model to predict the age that a participant started drinking alcohol.

Following the establishment of the four hypotheses, our first step was to convert the raw dataset into a useable set of variables. Primarily this involved scoring the two psychometric questionnaires into a set of final variables. A number of questions required reversed scoring as an intermediate process. The 30 item BIS questionnaire was converted into eight final factors, while the DOSPERT questionnaire was reduced into six final factors. A brief description of these factors is provided in the appendix.

Overall this process was considerable more time consuming then initially expected. Issues arose when we discovered the data collection system has failed for one a specific question on the BIS scale ("I buy things on impulse?") meaning two of our final variables on the BIS questionnaire (BIS Sum & BIS Motor) where

##

Add in brief description of each variable