

**Analysis Assignment**

**Data Analyst**

**Bootcamp**

# Introduction

This assignment will help you to collate the learning developed throughout the Data Analyst bootcamp programme.

You are presented with a data set from a study into wellness amongst post-graduate students at a North East university. From this you should form two sensible hypotheses and conduct the necessary analysis to determine whether to reject or accept your hypotheses. The survey focused on some factors which could affect management of workload, and the university is interested in analysis which could provide insight into how to better support students.

It is expected that you will produce at least the following:

* Two sensible hypotheses based on the survey data
* Supporting statistical values and charts
* Results of regression analysis, which should include control variables and a justification for inclusion of these
* Written explanations of what the analysis shows
* A conclusion stating whether to reject or accept the null hypothesis

You should include screenshots of the outputs of the analysis you perform, along with written explanations of what the analysis shows.

**Note – you will need to deal with missing values in the dataset before you can use R to perform the analysis. Missing value notation is shown in the variable list below, these values should be adjusted to NA in the dataset, so R treats them as missing values.**

# Dataset Details

Variables:

* **How managed workload?** 
  + 1-9 Likert scale
  + Missing value - 10
  + howManagedWorkload
* **Involvement in department**
  + 1-9 Likert scale
  + Missing value - 10
  + involvementInDepartment
* **Involvement in university**
  + 1-9 Likert scale
  + Missing value - 10
  + involvementInUniversity
* **Total income this year**
  + Missing value - 99999
  + totalIncomeThisYear
* **Job out of university**
  + 1 – yes | 2 – no
  + Missing value - 9
  + jobOutOfUniversity
* **How many hours worked?**
  + Missing value - 99
  + howManyHoursWorked
* **How much rent per week?**
  + Missing value - 999
  + howMuchRentPerWeek
* **How much spent for gas?**
  + Missing value - 999
  + howMuchSpentForGas
* **How much spent for electric?**
  + Missing value - 999
  + howMuchSpentForElectric
* **How much spent for food?**
  + Missing value - 999
  + howMuchSpentForFood
* **How much from student fund?**
  + Missing value - 999
  + howMuchFromStudentFund
* **Size of overdraft**
  + Missing value - 9999
  + sizeOfOverdraft
* **Any credit cards?**
  + 1 – yes | 2 – no
  + Missing value - 9
  + anyCreditCards
* **How much credit owed?**
  + Missing value - 9999
  + howMuchCreditOwed
* **Have you borrowed?**
  + 1 – yes | 2 – no
  + Missing value - 9
  + haveYouBorrowed
* **How much borrowed owed?**
  + Missing value - 9999
  + howMuchBorrowedOwed
* **Age**
  + No missing values in dataset
  + age
* **Sex**
  + 1 – male | 2 – female
  + Missing value - 9
  + sex
* **Location**
  + Values matching the 41 officials names for the UK regions at NUTS level 2 as supplied by the Office for National Statistics
  + Missing value - 9999
  + location

# Results & Explanations of Analysis

# Introduction

## Data Preparation

Prior to any regression analysis, the data need to be cleaned.

1. Spurious values should be checked for
2. Numbers representing missing values should be replaced with appropriate nulls
3. Outliers should be removed

Image result for city & guilds logo pngA close up of a logo

Description automatically generatedA close up of a logo

Description automatically generatedA close up of a logo

Description automatically generatedA picture containing clipart

Description automatically generatedA close up of a logo

Description automatically generated

**INTEQUAL**

**INTEQUALLTD**

**INTEQUAL**

**info@intequal.co.uk**

**0239 244 9727**

**0121 262 4117**

**Havant**

31a, Basepoint Business Centre, Harts Farm Way, Havant, Hampshire, PO9 1HS

**Birmingham**

The Colemore Building, 20 Colemore Circus

Queensway, Birmingham, B4 6AT