52465 Project Prep 21.11.2018

For this project you must create a data set by simulating a real-world phenomenon of your choosing.

You may pick any phenomenon you wish – you might pick one that is of interest to you in your personal or professional life.

Then, rather than collect data related to the phenomenon, you should model and synthesise such data using Python.

We suggest you use the numpy.random package for this purpose.

Speciﬁcally, in this project you should:

• Choose a real-world phenomenon that can be measured and for which you could collect at least one-hundred data points across at least four diﬀerent variables.

• Investigate the types of variables involved, their likely distributions, and their relationships with each other.

• Synthesise/simulate a data set as closely matching their properties as possible.

• Detail your research and implement the simulation in a Jupyter notebook – the data set itself can simply be displayed in an output cell within the notebook. Note that this project is about simulation – you must synthesise a data set. Some students may already have some real-world data sets in their own ﬁles. It is okay to base your synthesised data set on these should you wish (please reference it if you do), but the main task in this project is to create a synthesised data set. The next section gives an example project idea.

Ideas for a real-world phenomenon;

It must be measured, and I need to be able to collect at least one-hundred data points across four different variables.

I recently adopted a rescue puppy, and when I read the description of this project I immediately thought about different dog breeds, and their relation to being adopted from animal shelters. This seems like something there should be a lot of data on, so I decided to base my project on this topic and investigate online. The four different variables could be four different dog breeds.

‘Investigate the types of variables involved, their likely distributions, and their relationships with each other’ – how will I do this? Hopefully as I research I will stumble across an idea.

‘Synthesise/simulate a data set as closely matching their properties as possible’ – I think this should be possible once I have completed some research and established some real data on the topic.

‘Detail your research and implement the simulation in a Jupyter notebook – the data set itself can simply be displayed in an output cell within the notebook’ – having become familiar with Jupyter notebook in our last assignment for this module I feel confident that I should be able to complete this. I am hoping for some feedback on the assignment prior to handing in this project so I can implement the feedback fere.

I should research this phenomenon and see what types of real data has been collected on it, and then synthesise data for this project. I should use Python to model and synthesise such data.

1. Phenomenon = the adoption of dogs or puppies from animal shelters.
2. Variable = perhaps the breed of the dog/puppy.
3. I should try to come up with the equivalent of the examples Ian has provided in the project example (number of hours study, number of logins to Moodle, previous qualifications), perhaps this could be compared to how well trained the dog is, the amount of previous owners the dog has had and how many times it has been placed for adoption by different owners, and age of the dog/puppy.

I should investigate how these impact the likelihood of the dog/puppy being successfully adopted.

Ian’s example;

The hours and grade variables will be non-negative real number with two decimal places, logins will be a non-zero integer and qual will be a categorical variable with four possible values: none, bachelors, masters, or phd. After some online research, I ﬁnd that full-time post-graduate students study on average four hours per week with a standard deviation of a quarter of an hour and that a normal distribution is an acceptable model of such a variable. Likewise, I investigate the other four variables, and I also look at the relationships between the variables. I devise an algorithm (or method) to generate such a data set, simulating values of the four variables for two-hundred students. I detail all this work in my notebook, and then I add some code in to generate a data set with those properties.

**Marking Scheme;**

**25% Research** Investigation of the data set as demonstrated by references, background information, and approach.

**25% Development** Clear, well-written, and eﬃcient code with appropriate comments.

**25% Consistency** Good planning and pragmatic attitude to work as evidenced by commit history.

**25% Documentation** Concise descriptions and plots of variables in the data set.

**Advice for students;**

* *Your git commit history should be extensive* – I will try to ensure this.

**Etain’s research on my chosen topic:**

I adopted my puppy, Tyson, from MADRA in August of this year after our family dog, Bailey, passed away.

I wasn’t certain what type of dog I wanted to adopt. I knew I would prefer a puppy, as then I would have the experience of raising him/her from a young age. My boyfriend and I contacted MADRA, and they advised us to send in an application so they could complete a home inspection, etc. We also contacted the GSPCA who advised that they would not allow us to adopt a puppy as there was not always guaranteed to be someone at home 24/7 in our house.

We visited MADRA several times, and walked lots of their dogs who ranged in age from 1 year to approximately 10 years old, but they had no puppies on any of the occasions we visited the shelter.

We continued to contact the kind people who work at MADRA regularly so that we would be updated when any puppies arrived at the shelter, and eventually we were informed that two pitbull-cross puppies had arrived in need of a good home. We knew we could not adopt two puppies, but the staff at MADRA advised us that it was not a requirement that the puppies be rehomed together, as was the case for some of their other candidates.

I visited their website while completing this project, but could not locate any information or figures on the amount of dogs they rehome each year.

I decided to Google ‘Ireland Dog Adoption Figures’.

The results

According to the ISPCA website ‘there has been a steady increase of dog adoptions, 30 dogs are still abandoned each day in Ireland.’ I was happy to see that there is an increase in dog adoptions. I thought this may be of interest in terms of data for this project. ‘Recent statistics released by the Department of Environment, Community and Local Government have shown that there has been a steady increase in the number of dog adoptions since the Pedigree Adoption Drive first launched 6 years ago. In 2008, a startling 10,094 dogs died in Irish pounds while in 2012 4,500 dogs died in shelters nationwide.’

I learned from The Journal.ie that;

‘The latest dog control report released by the Department of Rural and Community Development, suggests that 916 dogs, not counting registered greyhounds) were put to sleep in Irish pounds in 2017, a 40% reduction from the 1,522 strays who suffered the same fate the previous year.

Meanwhile, 11,559 dogs entered Irish pounds last year, an 8% drop from the 12,549 who entered such institutions in 2016.

Dogs Trust also drew particular attention to the issues facing registered greyhounds in Ireland. Last year saw a 15% increase on the number of such greyhounds entering pounds. Of those 215, 180 were surrendered by their owners, and 80 (37%) were destroyed.’

This was interesting to me – it seems that the breed of the dog (in the case of greyhounds) has an impact on the amount which enter pounds.

I downloaded and reviewed the latest dog control report (https://drcd.gov.ie/dog-control/view-our-dog-control-statistics/). I intend to simulate figures based on the actual figures from this report for the purpose of this project.

According to an article from *The Irish Times* ‘Limerick put down three times more dogs than rest of country’. This indicates that the area/location impacts dogs in shelters/ dog welfare.

References;

* MADRA Website - <http://www.madra.ie/>
* ISPCA Website - <https://www.ispca.ie>
* Department of Environment, Community and Local Government Website –
* <https://www.thejournal.ie>
* <https://www.irishtimes.com>
* Dog control report 2017
* <https://drcd.gov.ie>