worksheet_03

February 6, 2024

1 Worksheet 03

Name: Daniyal Ahmed UID: U11469883

1.0.1 Topics

- Intro to DS

1.0.2 Linear Algebra Review

If you need a linear algebra review, please read through the following pdf before next class

1.0.3 Intro to Data Science

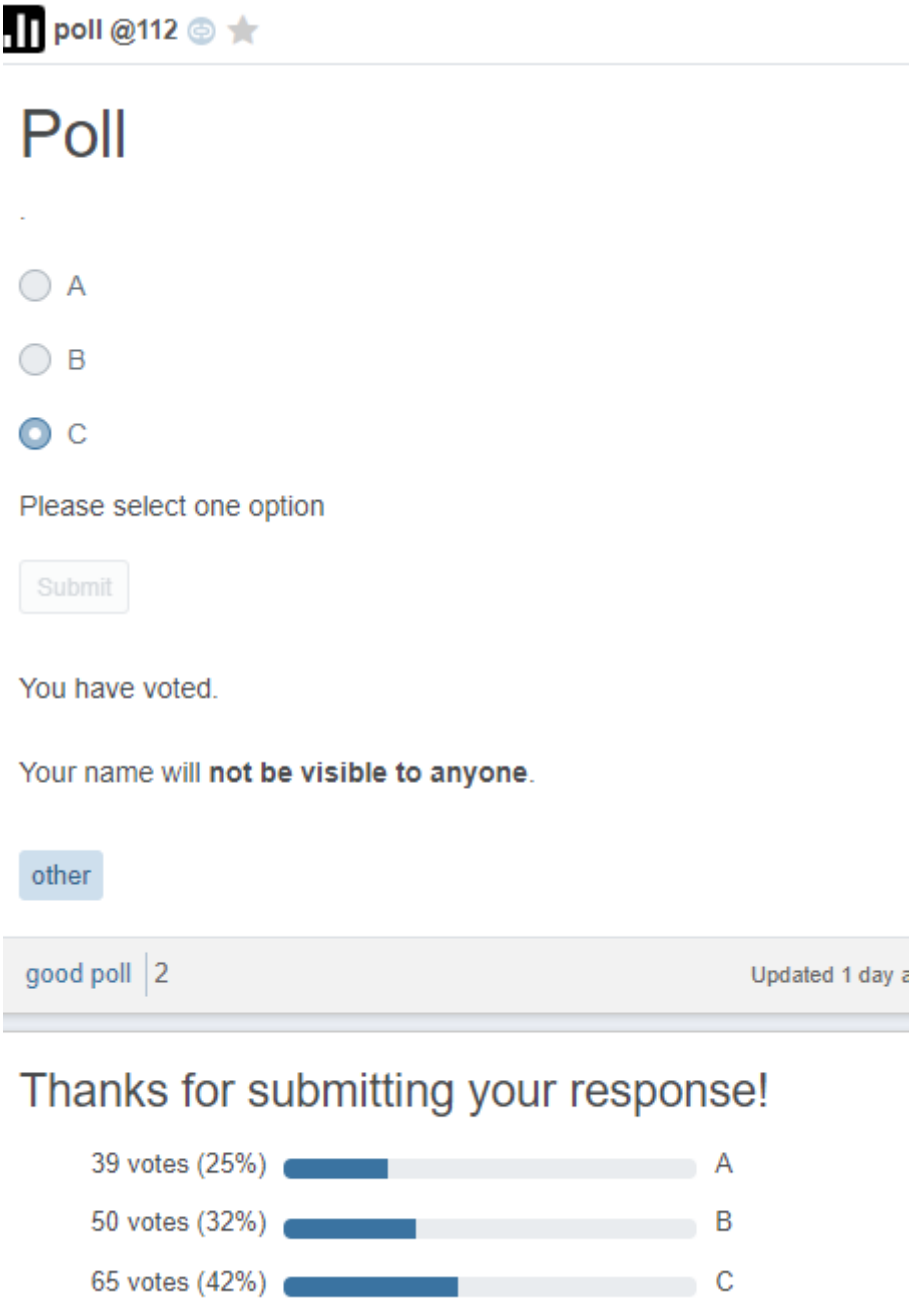
a) what property must a hypothesis have?

A good Hypothesis must be falsifiable

b) what examples would you have wanted to try?

I would have wanted to try (100,900,100), and (100,200,300)

c) Poll 1

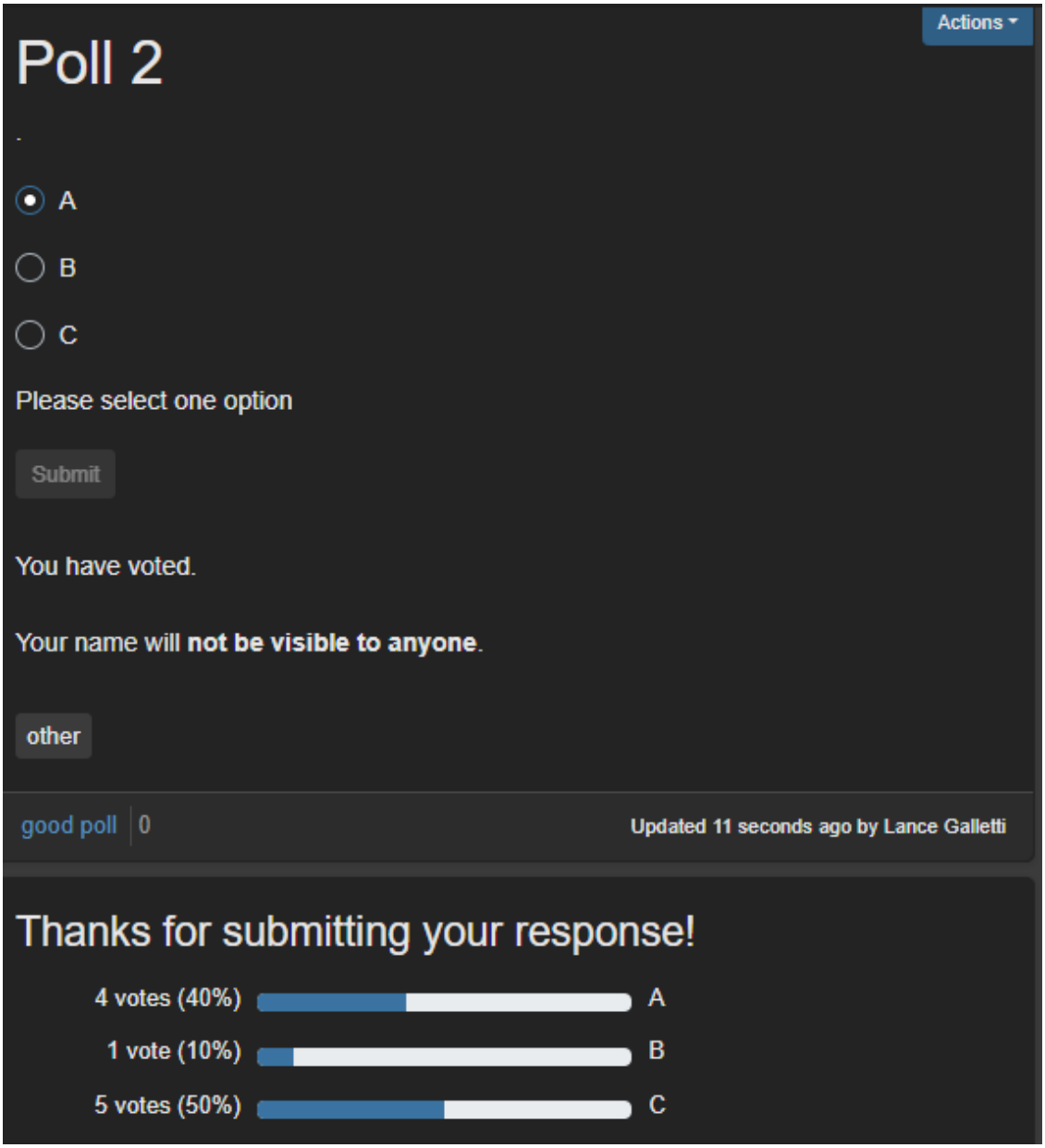


d) Given the hypothesis (x, 2x, 3x), for each of the following, determine whether they are positive or negative examples:

- (2, 4, 6)
- (6, 8 , 10)
- (1, 3, 5)

(2,4,6) are positive examples while (1,3,5) and (6,8,10) are negative examples

e) Poll 2



f) Describe steps of a Data Science Workflow

- Process Data: First we process the data and put in a format we want to work with
- Explore Data: Explore how different which data types relate to the target
- Extract Features: You might be presented with a very relevant dataframe, but you might want to combine attributes like age or income or income+ assets etc.
- Create a Model: This stage requires you to have done all the previous steps correctly because data science is all about the data and how you treat the data

g) Give a real world example for each of the following data types:

- record
- graph
- image

- text
- Records: There are vectors, we can use these to see how closely + related people's ages and balances are using other Data Science tools i.e clustering
- Graphs: Social Networks like Facebook, or looking at how busy traffic maybe at certain intersections and roads with wieghted graphs. In this case the nodes would be the intersections and the edges would be the roads
- Images: we represent as matrix of pixels, a real world example is astronomy since astronomers look at images and use math and measure individual pixels to calculate how far or close a planet or any other object in space might be

h) Give a real world example of unsupervised learning

Unsupervised learning is better used to find more information about the data, for example a form of unsupervised learning having a dataset of a collection of articles and looking at the data to answer the question do the articles cover the same task?

i) Give a real world example of supervised learning

Supervised learning is better used to find correlation and such for example, Lets say we want to map out how temprature effected cricket chirps, that would mean we plot and measure the number of cricket chirps per minute and the temprature. Then we can change the temprature around and see how the cricket chirps are affected