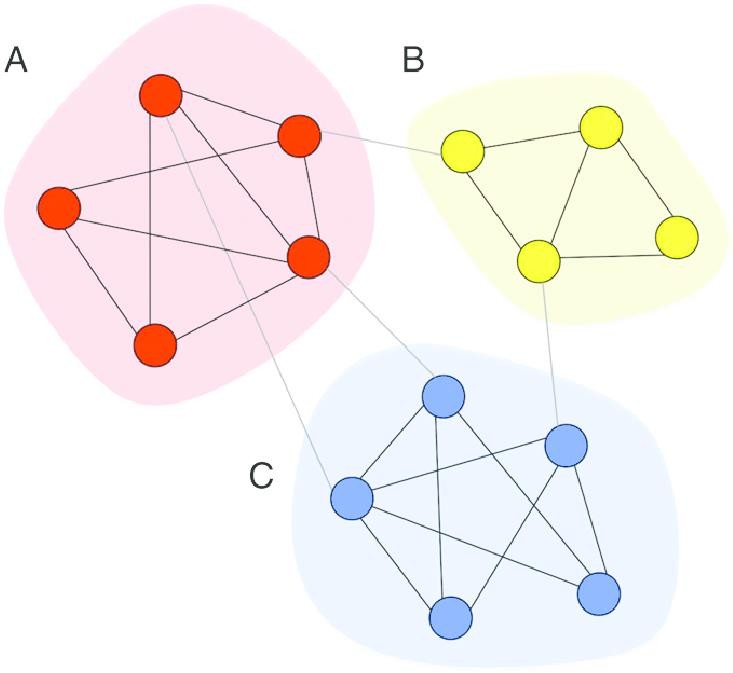
**EXPERIMENT**

###### **Objectives:**

Implement homophily for social media analysis.

###### **Background Study:**

**Homophily**

Network homophily refers to the theory in [network science](https://en.wikipedia.org/wiki/Network_science) which states that, based on node attributes, similar nodes may be more likely to attach to each other than dissimilar ones. The hypothesis is linked to the model of [preferential attachment](https://en.wikipedia.org/wiki/Preferential_attachment) and it draws from the phenomenon of homophily in social sciences and much of the scientific analysis of the creation of social ties based on similarity comes from network science. In fact, empirical research seems to indicate the frequent occurrence of homophily in real networks.

**Outcome:**

Students will be able to perform a case study on Social Network Analysis for a population.

###### **Problem Statement:**

In this case study, you will investigate homophily of several characteristics of individuals connected in social networks in rural India.

You will calculate the chance homophily for an arbitrary characteristic. Homophily is the proportion of edges in the network whose constituent nodes share that characteristic. How much homophily do we expect by chance? If characteristics are distributed completely randomly, the probability that two nodes x and y share characteristic a is the probability both nodes have characteristic a, which is the frequency of a squared. The total probability that

nodes x and y share their characteristic is therefore the sum of the frequency of each characteristic in the network.

**DATA CAMP CASE STUDY QUESTIONS AND ANSWERS**

Q1. Network homophily occurs when nodes that share an edge share a characteristic more often than nodes that do not share an edge. In this case study, we will investigate homophily of several characteristics of individuals connected in social networks in rural India.

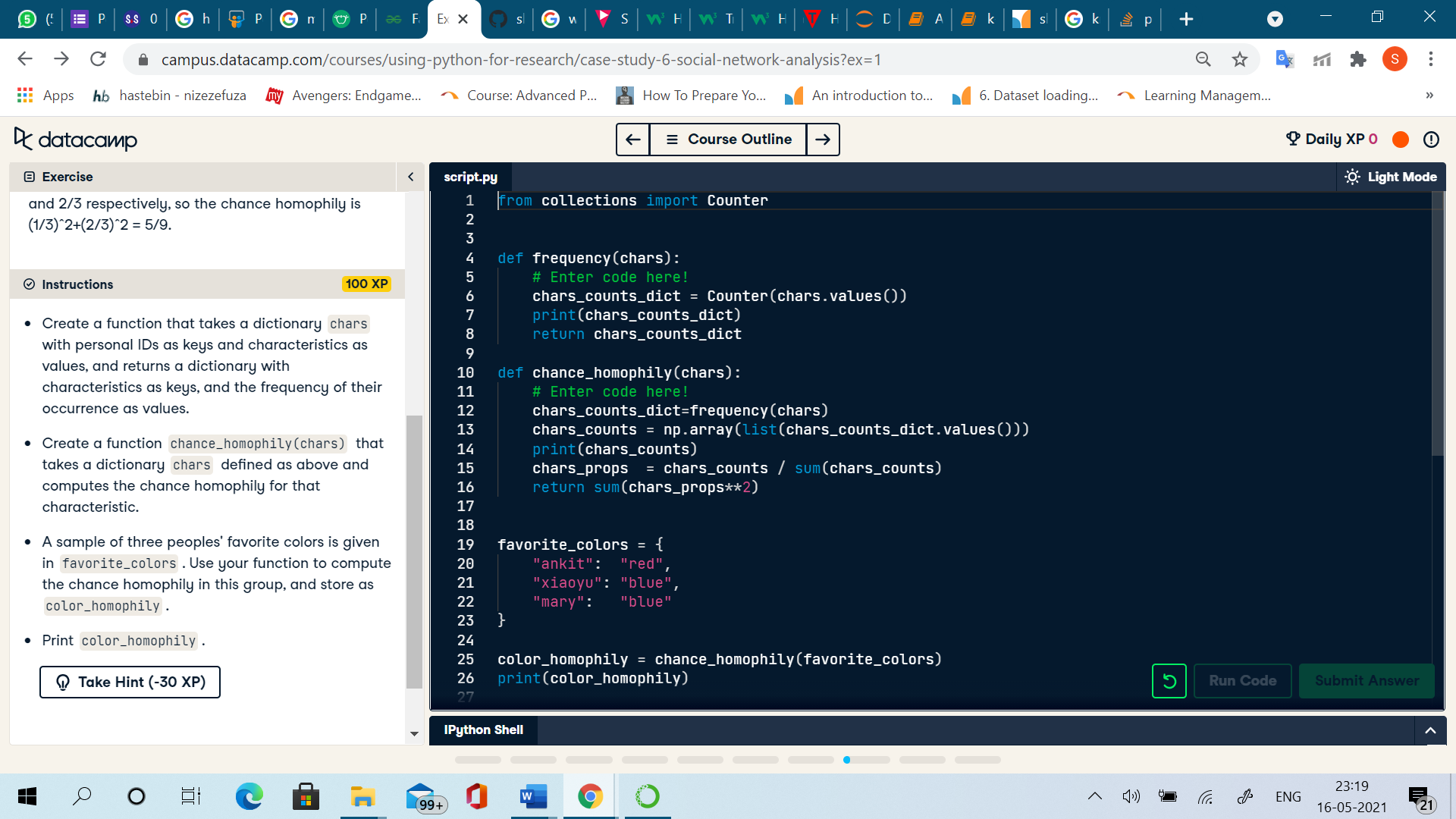
In this exercise, we will calculate the chance homophily for an arbitrary characteristic. Homophily is the proportion of edges in the network whose constituent nodes share that characteristic. How much homophily do we expect by chance? If characteristics are distributed completely randomly, the probability that two

nodes x and y share characteristic a is the probability both nodes have characteristic a, which is the frequency of a squared. The total probability that nodes x and y share their characteristic is therefore the sum of the frequency of each characteristic in the network. For example, in the dictionary favorite\_colors provided, the frequency of red and blue is 1/3 and 2/3 respectively, so the chance homophily is (1/3)^2+(2/3)^2 = 5/9.

**INSTRUCTIONS**

* Create a function that takes a dictionary chars with personal IDs as keys and characteristics as values, and returns a dictionary with characteristics as keys, and the frequency of their occurrence as values.
* Create a function chance\_homophily(chars) that takes a dictionary chars defined as above and computes the chance homophily for that characteristic.
* A sample of three peoples' favorite colors is given in favorite\_colors. Use your function to compute the chance homophily in this group, and store as color\_homophily.
* Print color\_homophily

**CODE:**



**OUTPUT:**

