# IC 252 Data Science II
## Lab 7: Exponential Distribution and Correlations

Q1. The exponential distribution is ideal to model waiting times (e.g., time to failure of sensors in a system, flood occurrence, etc.). In this assignment, we will be using the exponential distribution to model time to the arrival of the next confirmed case of Covid-19 in India. Based upon data of confirmed cases from the source www.covid19india.org, between 17th April 2020 and 23rd April 2020, there were on average 1373 confirmed cases per day, i.e., on average around 57 cases per hour.

A. Write a program to plot the probability density function of the wait time for the next Covid-19 confirmed case, where the X axis is the wait time in hours and Y axis is the probability density.
B. Write a program to find the probability of the wait time for the next Covid-19 confirmed case to be less than or equal to 1 minute (*Hint:* convert minutes into hours before using it in the cumulative density function).
C. Write a program to find the probability of the wait time for the next Covid-19 confirmed case to be between 1 minute and 2 minutes.
D. Now, write a program to find the probability of the wait time for the next Covid-19 confirmed case to be more than 2 minutes.
E. Suppose, the average number of cases per hour doubled. Write a program to find the probability of wait time for the next Covid-19 confirmed case to be between 1 minute and 2 minutes.

Q2. Consider the following dataset: Supplementary file.

(This dataset was constructed by connecting Indian Census data with data from www.covid19india.org)

Each row in data is a person and these data are of the first 2,310 Covid-19 cases in India. Column District contains the district code for the person's location. Column ConfirmedCases contains the total number of confirmed cases in the district of person's location. The columns Population, SexRatio, State, SmellTrend, and Literacy contain the population, number of males to 1000 females, state code, number of searches for the word "smell" on Google.com, and the percentage literacy in the district of person's location, respectively. The columns Age, Gender, ForeignCode, ForeignCountryCode contain the age (in years) of the person, the gender of the person (1 male and 0 female), whether person travelled to a foreign country (1 travelled and 0 not travelled), and the code of the foreign country to which the person travelled, respectively. The column Status tells whether the person was hospitalized due to Covid-19, recovered from Covid-19, or died due to Covid-19.

A. Convert the column Status such that hospitalized is coded as 1, recovered as 2, and dead as 3.
B. Find the correlation between the following random variables across all people:
    a. Status and Population
    b. Status and SexRatio
    c. Status and Literacy

d. Status and Age
e. Status and SmellTrend
f. Status and Gender.

C. What variables correlate strongly to the Status? (*Hint:* Sort the correlations from the highest to the lowest)