

CT-1 Section-B CSE 315
Time: 30 Mins Total Marks-15
Fall 2024

1. You are part of a research team studying the behavior and experiences of freelance workers in the tech industry who primarily take on short-term contracts. Many of these workers are difficult to locate as they do not belong to formal organizations, and some work entirely remotely. The team initially knows a few freelancers who are willing to participate, but you need to expand the sample size and reach more individuals in this hidden population. [4]
- Would **snowball sampling** be an appropriate method for collecting data on this population? If so, explain why you chose this sampling method. How does this technique help in accessing hard-to-reach groups like freelance workers?

2. What do you mean by primary and secondary sources of data? Briefly explain with an example. [3]
3. The CSV file (car.csv) contains the following features: [8]

	Car_Name	Year	Selling_Price	Present_Price	Driven_kms	Fuel_Type	Selling_type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0
...
296	city	2016	9.50	11.60	33988	Diesel	Dealer	Manual	0

- (a) Read the data set from the (car.csv) file and print 5 values of car name 'city'.
- (b) The feature set has a target variable, "Selling_price," which needs to be differentiated from input features. Write code on how you differentiate it.
- (c) Identify how much missing information is present in your dataset.
- (d) The first feature represents the Car_Name, which needs to be dropped from the dataset. Write code how to drop it.