Objective: Upon completion of this session, you will be able to understand the various central measures for the given data, solve simple problems to understand the application of Probability concepts and interpretations.

Key takeaways:

- Data type classification
- Given the data, compute central measures such as mean, median, mode, quartiles.
- Properties of probability, Joint probability, conditional probability, marginal probability
- Understand the jargon and compute True positives, True negatives, False Positives and False negatives using a tree diagram or tabular format.
- Introduction to R, computing measures and solving some problems using the basic functions. Working of set.seed.

Problem 1: You plan to hire a taxi to commute. When you access the app based on the pick-up and drop point, you get an estimated charge for the travel. Name the factors that the app must be considering to arrive at an estimated cost and respective data types.
Few more examples that you come across daily:

1. Blood pressure reading
2. Number of stocks traded
3. Education background
4. Type of groceries purchased
5. Price of petrol
6. Rating a Restaurant
7. Buy a car or not
8. Lifetime of a battery

Problem 2: Here is the data of experience of a CPEE class. We have grouped individuals into 6 groups and here is the data. Compute the average and median values for each group and list your observations.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
0	3	14	9	16	13
0	12	9.5	4.5	12	10
0	3.5	4.5	9	6.5	10
0	14	7.5	8	1	1
0	3	5	4	11	3
0	2.8	2	6	5	4
1	5	4.8	5	3	6
1	9	3.6	3.5	8	3.8
1	5.5	6	2.8	3	4
16	9	8.5	12	4	8

1. What is the average in each group?
2. What is the median in each group?
3. What is the average experience across all groups?
4. What do you observe?

Problem 3: You and your friends regularly order food online and prefer door delivery services. Each one believes that their respective service providers are very prompt. To understand it better, you have started collecting the time to deliver food in 20 different occasions for all. Here is the data of delivery times.

Time taken to deliver the order in minutes				
EagleBoys	**FoodPanda**	**Swiggy**	**PiazzaHut**	**Dominos**
30	39	33	30	35
35	37	31	35	23
23	35	25	23	35
12	33	37	12	33
15	31	28	15	30
16	25	36	16	31
19	37	20	19	25
31	28	30	31	37
35	10	35	35	28
21	46	23	0	36
39	30	12	60	20
37	35	15	37	12
35	23	16	35	15
33	12	19	33	16
31	15	31	31	19
25	16	35	25	31
37	19	0	37	35
28	31	60	28	21
36	35	37	36	44
20	21	35	20	32

1. Now that you know central measures help you understand data better you go ahead with computing the central measures. (mean, median, mode, quartiles, range, inter-quartile range, standard deviation)
2. What do you observe?
3. Do you still believe that all the service providers are prompt in their services?

Few more applications
1. Performance of batsmen in a cricket team
2. Performance of stocks in market

3. Performance of machines in a manufacturing unit

Problem 4: Two people work in a factory making parts for cars. The table shows how many complete parts they make in one week.

Worker	Mon	Tue	Wed	Thu	Fri
Philip	20	21	22	20	21
Mathews	30	15	12	36	28

(a) Find the mean, median and range for Philip and Mathews.
(b) Who is more consistent?

Problem 5: Find the mode for 8,6,2,4,6,8,10,8

Problem 6: Analyze the performance of a class in a WUQ

Scores: 11, 7.5, 8.5, 10, 10, 10.5, 5.5, 10, 9, 9.5, 5.25, 8, 6.5, 10.5, 8.75, 0, 6, 6, 6.75,

8.75, 0, 9.5, 7.5, 8.5, 7

1. How is the spread of the scores? Compute range, variance & standard deviation
2. Which central tendency measure is the most representative of the performance of the class?
3. Find the 25th percentile, 50th percentile and 75th percentile for this data.

Problem 7: Temperatures in 5 cities measured on 12 days is given below. The weather department says that two cities have similar weather. Use central tendencies to identify those two cities.

City 1	29	32	36	40	43	37	36	33	32	37	31	29
City 2	20	24	31	37	40	38	37	34	34	33	28	23
City 3	23	26	32	38	41	40	35	33	35	37	30	25
City 4	20	24	29	34	37	36	32	30	33	32	27	23
City 5	19	24	29	38	43	38	33	34	36	34	29	23

Problem 8: A large retailer store regularly orders cartons of Pineapples. The average weight of the cartons is supposed to be 22 kgs. Random samples of cartons from two suppliers were weighed. The weights in kgs of the cartons were

Supplier – I	17	22	22	22	27
Supplier – II	17	19	20	27	27

1. Compute the range of carton weights from each supplier
2. Compute the mean weight of cartons from each supplier.

3. Look at the two samples again. The samples have the same range and mean. How do they differ? The retailer store uses one carton of blueberries in each blueberry muffin recipe. It is important that the cartons be of consistent weight so that the muffins turn out right.

Problem 9: What is the probability that we get a 5 Tuesday in a 30-day month?

Problem 10: Below is a table of graduates and post graduates

	Graduate	Post Graduate	Total
Male	19	41	**60**
Female	12	28	**40**
Total	**31**	**69**	**100**

1. What is the probability that a randomly selected individual is a male and a graduate? What kind of probability is it (Marginal/ Joint/Conditional)?
2. What is the probability that a randomly selected individual is a male?
3. What is the probability of a randomly selected individual being a graduate? What kind of probability is this?
4. What is the probability that a randomly selected person is a female given that the selected person is a post graduate? What kind of probability is this?

Problem 11: In a region during a 1 year period, there were 1000 deaths. It was observed that 321 people died of a renal failure and 460 people had at least one parent with renal failure. Of these 460 people, 115 died of renal failure. Calculate the probability of a person that dies of renal failure if neither of his parents had a renal failure

Problem 12: 0.5 percent of the population of an area is affected by a particular disease. A test is developed to detect the disease. This test gives a false positive 3% of the time and false negative 2% of the time.

a) Draw the tree diagram for this problem.

b) What is the probability that the test gives a positive result?

c) If a person's test turns out to be positive, what is the probability that he (actually) has the disease

Other applications:
*   What is the probability that an email is spam given a word/phrase
*   Anomaly detection. Check the references material.