FUNDAMENTALS OF DATA SCIENCE

UNIT 1

1) What is Data Science? Write a short note on the need of Data Science.
2) Briefly explain various data science disciplines.
   https://medium.com/@udemeudofia01/disciplines-in-data-science-a1da93306528
3) What is the difference between analysis & analytics?
   https://www.geeksforgeeks.org/difference-between-data-analytics-and-data-analysis/
4) Explain how Business Analytics, Data Analytics, Data Science related to each other with a neat diagram.
5) What are the traditional data methods and data science terms, also explain the benefits of it?
6) Briefly explain techniques for Working with Traditional Data methods with real life examples.
7) Briefly explain techniques for Working with Big Data with real life examples.
8) Briefly explain techniques for Working with Business Intelligence with real life examples.
9) Briefly explain techniques for Working with traditional methods of Data Science with real life examples.
10) Briefly explain techniques for Working with Machine Learning with real life examples.
11) Write a short note on popular Data Science tools.
12) What is Feature and Feature Engineering? Explain with its benefits
13) Explain following terms briefly
i) Feature creation   ii) Feature transformation
https://www.javatpoint.com/feature-engineering-for-machine-learning

UNIT 2

14) Define Statistics. Write a short note on probability. Also solve given problem:
Alice has 2 kids and one of them is a girl. What is the probability that the other child is also a girl? You can assume that there are an equal number of males and females in the world.

➔ Prob(two girls | at least one girl)
   = Prob(at least one girl | two girls) * Prob(two girls) / Prob(at least one girl)
   = Prob(at least one girl | two girls) * Prob(two girls) / (1 - Prob(no girls))
   = 1 * (1/4) / (1 - 1/4)
   = **1/3**

15) Amita randomly picks 4 cards from a deck of 52-cards and places them back into the deck (Any set of 4 cards is equally likely). Then, Babita randomly chooses 8 cards out of the same deck (Any set of 8 cards is equally likely). Assume that the choice of 4 cards by Amita and the choice of 8 cards by Babita are independent. What is the probability that all 4 cards chosen by Amita are in the set of 8 cards chosen by Babita?
➔

The total number of possible combination would be 52C4 (For selecting 4 cards by Anita) * 52C8 (For selecting 8 cards by Babita).

Since, the 4 cards that Anita chooses is among the 8 cards which Babita has chosen, thus the number of combinations possible is 52C4 (For selecting the 4 cards selected by Anita) * 48C4 (For selecting any other 4 cards by Babita, since the 4 cards selected by Anita are common)

16) Cross-fertilizing a red and a white flower produces red flowers 25% of the time. Now we cross-fertilize five pairs of red and white flowers and produce five offspring. What is the probability that there are no red flower plants in the five offspring?

➔

Answer

Y = # of red flowered plants in the five offspring. Here, the number of red-flowered plants has a binomial distribution with $n = 5, p = 0.25$.

$$P(Y = 0) = \frac{5!}{0!(5-0)!}p^0(1-p)^5$$
$$= 1(0.25)^0(0.75)^5$$
$$= 0.237$$

17) Ahmed is playing a lottery game where he must pick 2 numbers from 0 to 9 followed by an English alphabet (from 26-letters). He may choose the same number both times. If his ticket matches the 2 numbers and 1 letter drawn in order, he wins the grand prize and receives $10405. If just his letter matches but one or both of the numbers do not match, he wins $100. Under any other circumstance, he wins nothing. The game costs him $5 to play. Suppose he has chosen 04R to play.
What is the expected net profit from playing this ticket?
➔ **Expected value in this case**

E(X) = P(grand prize)*(10405-5)+P(small)(100-5)+P(losing)*(-5)

P(grand prize)= (1/10)*(1/10)*(1/26)

P(small) = 1/26-1/2600, the reason we need to do this is we need to exclude the case where he gets the letter right and also the numbers rights. Hence, we need to remove the scenario of getting the letter right.

P(losing) = 1-1/26-1/2600

Therefore we can fit in the values to get the expected value as $2.81

18) Write short not on Expected value. Also solve the following problem.
A local club plans to invest $10000 to host a baseball game. They expect to sell tickets worth $15000 . But if it rains on the day of the game, they won't sell any tickets and the club will lose all the money invested. If the weather forecast for the day of the game is 20% possibility of rain, is this a good investment?

Make a table of probability distribution .

Outcome	+$5000	−$10000
Probability	0.80	0.20

Use the weighted average formula.

$$\text{Expected Value} = 5000(0.8) - 10000(0.2)$$
$$= 4000 - 2000$$
$$= 2000$$

The club can expect a return of $2000 . So, it's a good investment, though a bit risky.

In other cases, we are asked to find the values of one or more variables involved in the model for which the experiment has a given expected value .

19)   Write short not on Expected value. Also solve the following problem.

A company makes electronic gadgets. One out of every 50 gadgets is faulty, but the company doesn't know which ones are faulty until a buyer complains. Suppose the company makes a $3 profit on the sale of any working gadget, but suffers a loss of $80 for every faulty gadget because they have to repair the unit. Check whether the company can expect a profit in the long term. Write the probability distribution.

➔

Write the probability distribution.

$$E(X) = \frac{49}{50} \cdot 3 + \frac{1}{50} \cdot (-80)$$
$$= \frac{147}{50} - \frac{80}{50}$$
$$= \frac{67}{50}$$
$$= 1.34$$

Since the expected value is positive, the company can expect to make a profit. On average, they make a profit of $1.34 per gadget produced.

20)   Briefly explain following terms
a)   Frequency
b)   Events and their compliments
21)   What is Permutation?
22)   Briefly explain following terms
c)   Variation with repetition
d)   Variation without repetitions
23)   Write a short note on Combination. Also solve following problem

Imagine you were asked to order a 3-tier cake for your best friend' s wedding. They asked you to get a cake with a variety of flavours, so you decide to order different batters (cake flavours) for each tier. The pastry shop you contacted offers 5 different batters, so you want to know how many distinct options you have for the cake.

24)   Write short notes on
a)   Symmetry of Combinations
b)   Combination with Separate Sample spaces
25)   Briefly explain dependence and independence of sets.
26)   What is conditional probability? Solve

Suppose you draw two cards from a deck and you win if you get a jack followed by an ace (without replacement). What is the probability of winning, given we know that you got a jack in the first turn?

➔

Let event A be getting a jack in the first turn

Let event B be getting an ace in the second turn.

We need to find

P(A) = 4/52

P(B) = 4/51 {no replacement}

P(A and B) = 4/52*4/51= 0.006

Here we are determining the probabilities when we know some conditions instead of calculating random probabilities. Here we knew that he got a jack in the first turn.

27)   What is conditional probability? Solve
Suppose you have a jar containing 6 marbles – 3 black and 3 white. What is the probability of getting a black given the first one was black too without replacement

Suppose you have a jar containing 6 marbles – 3 black and 3 white. What is the probability of getting a black given the first one was black too.

P (A) = getting a black marble in the first turn

P (B) = getting a black marble in the second turn

P (A) = 3/6

P (B) = 2/5

P (A and B) = ½*2/5 = 1/5

$$P\left(\frac{B}{A}\right) = \frac{P(A \text{ and } B)}{P(A)} = \frac{0.2}{0.5} = 0.4$$

Condition	Follow Traffic Rule	Does not follow Traffic Rule
Accident	50	500
No Accident	2000	5000

28)   A research group collected the yearly data of road accidents with respect to the conditions of following and not following the traffic rules of an accident prone area. They are interested in calculating the probability of an accident given that a person followed the traffic rules. The table of the data is given as follows:

➔

Now here our equation becomes:

P(Accident | A person follow Traffic Rule) = P(Accident and follow Traffic Rule) / P(Follow Traffic Rule)

29)   Briefly explain the Law of Total Probability.

30)   Explain Additive Law with examples.

31)   Explain multiplication Law with examples.
32)   Explain Bayes Law with examples.
33)   A bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.
34)   You are planning a picnic today, but the morning is cloudy

●   Oh no! 50% of all rainy days start off cloudy!
●   But cloudy mornings are common (about 40% of days start cloudy)
●   And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)

What is the chance of rain during the day?