ABES Engineering College, Ghaziabad

B. Tech Odd Semester Make-Up Test

Printed Pages: 2
Session: 2022-2023

Course Code: KDS501 **Roll No:**
Course Name: Introduction to Data Analytics and Visualization
Date of Exam: 27-Dec-2022
Maximum Marks:100 **Time:** 1:30-4:30

Instructions:

1. **Attempt All sections.**
2. **If require any missing data, then choose suitably.**

---

1 a) Cloud computing is a big shift from the traditional way businesses thinking about IT resources. Explain Public Cloud and Private cloud with their major five characteristics.

Solution: Step-1

The two primary types of cloud environments: (1) public clouds and (2) private clouds.

Public Clouds

Public clouds have gotten the most hype and attention. With a public cloud users are basically loading their data onto a host system and they are then allocated resources as they need them to use that data. They will get charged according to their usage. There are definitely some advantages to such a setup:

- The bandwidth is as - needed and users only pay for what they use.
- It isn't necessary to buy a system sized to handle the maximum capacity ever required and then risk having half of the capacity sitting idle much of the time.
- If there are short bursts where a lot of processing is needed then it is possible to get it with no hassle. Simply pay for the extra resources.
- There's typically very fast ramp - up. Once granted access to the cloud environment, users load their data and start analyzing.
- It is easy to share data with others regardless of their location since a public cloud by definition is outside of a corporate firewall. Anyone can be given permission to log on to the environment created.

Private Clouds

▪ A private cloud has the same features of a public cloud, but it's owned exclusively by one organization and typically housed behind a corporate firewall. A private cloud is going to serve the exact same function as a public cloud, but just for the people or teams within a given organization.

▪ One huge advantage of an onsite private cloud is that the organization will have complete control over the data and system security.

▪ Data is never leaving the corporate firewall so there's absolutely no concern about where it ' s going.

Step-2

▪ **On-demand self-services:**

The Cloud computing services does not require any human administrators, user themselves are able to provision, monitor and manage computing resources as needed.

▪ **Broad network access:**

The Computing services are generally provided over standard networks and heterogeneous devices.

▪ **Rapid elasticity:**

The Computing services should have IT resources that are able to scale out and in quickly and on as needed basis. Whenever the user require services it is provided to him and it is scale out as soon as its requirement gets over.

▪ **Resource pooling:**

The IT resource (e.g., networks, servers, storage, applications, and services) present are shared across multiple applications and occupant in an uncommitted manner. Multiple clients are provided service from a same physical resource.

▪ **Measured service:**

The resource utilization is tracked for each application an d occupant, it will provideboth the user and the resource provider with an account of what has been used. This is done for various reasons like monitoring billing and effective use of resource.

1 b) Data empowers to make decision informed, justify the statement and explain the various methods of primary and secondary data collection in detail.

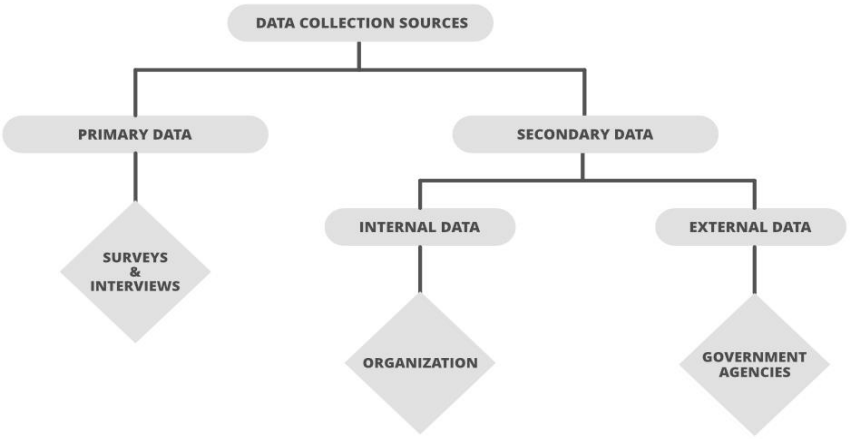
Solution: Step-1

- Data collection is the process of acquiring, collecting, extracting, and storing the voluminous amount of data which may be in the structured or unstructured form like

text, video, audio, XML files, records, or other image files used in later stages of data analysis. In the process of big data analysis, “Data collection” is the initial step before starting to analyze the patterns or useful information in data. The data which is to be analyzed must be collected from different valid sources.

- Data collection starts with asking some questions such as what type of data is to be collected and what the source of collection is. Most of the data collected are of two types known as “qualitative data“ which is a group of non-numerical data such as words, sentences mostly focus on behavior and actions of the group and another one is “quantitative data” which is in numerical forms and can be calculated using different scientific tools and sampling data.

Step-2



Primary Data Collection

Interview method:

- The data collected during this process is through interviewing the target audience by a person called interviewer and the person who answers the interview is known as the interviewee. Some basic business or product related questions are asked and noted down in the form of notes, audio, or video and this data is stored for processing. These can be both structured and unstructured like personal interviews or formal interviews through telephone, face to face, email, etc.

Survey method:

- The survey method is the process of research where a list of relevant questions are asked and answers are noted down in the form of text, audio, or video. The survey method can be obtained in both online and offline mode like through website forms and email. Then that survey answers are stored for analyzing data. Examples are online surveys or surveys through social media polls.

3. Observation method:

- The observation method is a method of data collection in which the researcher keenly observes the behavior and practices of the target audience using some data collecting tool and stores the observed data in the form of text, audio, video, or any raw formats. In this method, the data is collected directly by posting a few questions on the participants. For example, observing a group of customers and their behavior towards the products. The data obtained will be sent for processing.

Experimental method:

- The experimental method is the process of collecting data through performing experiments, research, and investigation. The most frequently used experiment methods are CRD, RBD, LSD, FD.

- **Secondary Data Collection**

1. Internal source:

- These types of data can easily be found within the organization such as market record, a sales record, transactions, customer data, accounting resources, etc. The cost and time consumption is less in obtaining internal sources.

2. External source:

The data which can't be found at internal organizations and can be gained through external third party resources is external source data. The cost and time consumption is more because this contains a huge amount of data. Examples of external sources are Government publications, news publications, Registrar General of India, planning commission, international labor bureau, syndicate services, and other non-governmental publications.

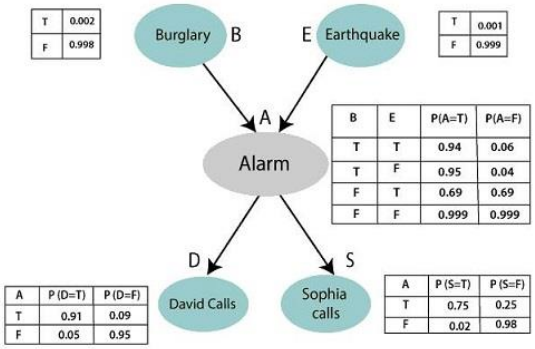
3. Other sources:

- **Sensors data:** With the advancement of IoT devices, the sensors of these devices collect data which can be used for sensor data analytics to track the performance and usage of products.

- **Satellites data:** Satellites collect a lot of images and data in terabytes on daily basis through surveillance cameras which can be used to collect useful information.

- **Web traffic:** Due to fast and cheap internet facilities many formats of data which is uploaded by users on different platforms can be predicted and collected with their permission for data analysis. The search engines also provide their data through keywords and queries searched mostly.

2. a) **Explain Baysian Belief Network. Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and David and Sophia both called the Harry. Solve this equation P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E) by probabilities given in below tables.**



T	0.002
F	0.998

T	0.001
F	0.999

B	E	P(A=T)	P(A=F)
T	T	0.94	0.06
T	F	0.95	0.04
F	T	0.69	0.69
F	F	0.999	0.999

A	P (D=T)	P (D=F)
T	0.91	0.09
F	0.05	0.95

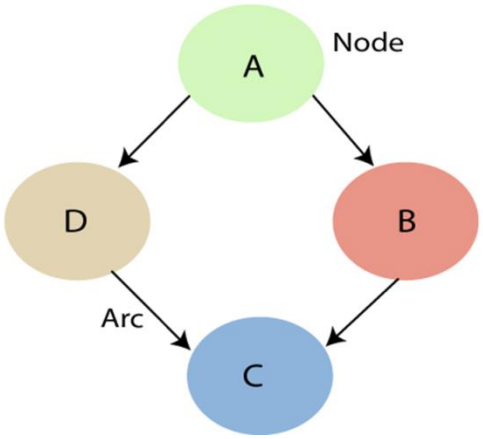
A	P (S=T)	P (S=F)
T	0.75	0.25
F	0.02	0.98

Solution: Step-1

Bayesian Belief Network in artificial intelligence. Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

"A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph." It is also called a Bayes network, belief network, decision network, or Bayesian model. Bayesian networks are probabilistic, because these networks are built from a probability distribution, and also use probability theory for prediction and anomaly detection.

Note: The Bayesian network graph does not contain any cyclic graph. Hence, it is known as a directed acyclic graph or DAG.



A directed graph G with four vertices A,B,C, and D. If p(xA,xB,xC,xD) factorizes with respect to G, then we must have p(xA,xB,xC,xD) = p(xA)p(xB|xA)p(xC|xB)p(xD|xC).

Step-2

P(S, D, A, ¬B, ¬E) = P (S|A) *P (D|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).
= 0.75* 0.91* 0.001* 0.998*0.999
= 0.00068045.

2. b) Write the steps involved in Principal Component Analysis (PCA). Consider the two dimensional patterns: (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8).

Compute the principal component using PCA Algorithm.

Solution: Step-1

Step-01: Get data.

Step-02: Compute the mean vector (μ).

Step-03: Subtract mean from the given data.

Step-04: Calculate the covariance matrix.

Step-05: Calculate the eigen vectors and eigen values of the covariance matrix.

Step-06: Choosing components and forming a feature vector.

Step-07: Deriving the new data set.

Step-2

Solution-

We use the above discussed PCA Algorithm-

Step-01:

Get data.

The given feature vectors are-

$x_1 = (2, 1)$

$x_2 = (3, 5)$

$x_3 = (4, 3)$

$x_4 = (5, 6)$

$x_5 = (6, 7)$

$x_6 = (7, 8)$

Step-02:

Calculate the mean vector (μ).

Mean vector (μ) = ((2 + 3 + 4 + 5 + 6 + 7) / 6, (1 + 5 + 3 + 6 + 7 + 8) / 6)

= (4.5, 5)

Step-03:

Subtract mean vector (μ) from the given feature vectors.

$x_1 – μ = (2 – 4.5, 1 – 5) = (-2.5, -4)$

$x_2 – μ = (3 – 4.5, 5 – 5) = (-1.5, 0)$

$x_3 – μ = (4 – 4.5, 3 – 5) = (-0.5, -2)$

$x_4 – μ = (5 – 4.5, 6 – 5) = (0.5, 1)$

$x_5 - \mu = (6 - 4.5, 7 - 5) = (1.5, 2)$

$x_6 - \mu = (7 - 4.5, 8 - 5) = (2.5, 3)$

Step-04:

Calculate the covariance matrix.

Covariance matrix is given by-

Now,

Covariance matrix

$= (m_1 + m_2 + m_3 + m_4 + m_5 + m_6) / 6$

On adding the above matrices and dividing by 6, we get-

Step-05:

Calculate the eigen values and eigen vectors of the covariance matrix.

λ is an eigen value for a matrix M if it is a solution of the characteristic equation $|M - \lambda I| = 0$.

From here,

$(2.92 - \lambda)(5.67 - \lambda) - (3.67 \times 3.67) = 0$

$16.56 - 2.92\lambda - 5.67\lambda + \lambda^2 - 13.47 = 0$

$\lambda^2 - 8.59\lambda + 3.09 = 0$

Solving this quadratic equation, we get $\lambda = 8.22, 0.38$

Thus, two eigen values are $\lambda_1 = 8.22$ and $\lambda_2 = 0.38$.

Clearly, the second eigen value is very small compared to the first eigen value.

So, the second eigen vector can be left out.

Eigen vector corresponding to the greatest eigen value is the principal component for the given data set.

So. we find the eigen vector corresponding to eigen value λ_1.We use the following equation to find the eigen vector-

$$MX = \lambda X$$

where-

M = Covariance Matrix

X = Eigen vector

λ = Eigen value

Solving these, we get-

$2.92X_1 + 3.67X_2 = 8.22X_1$

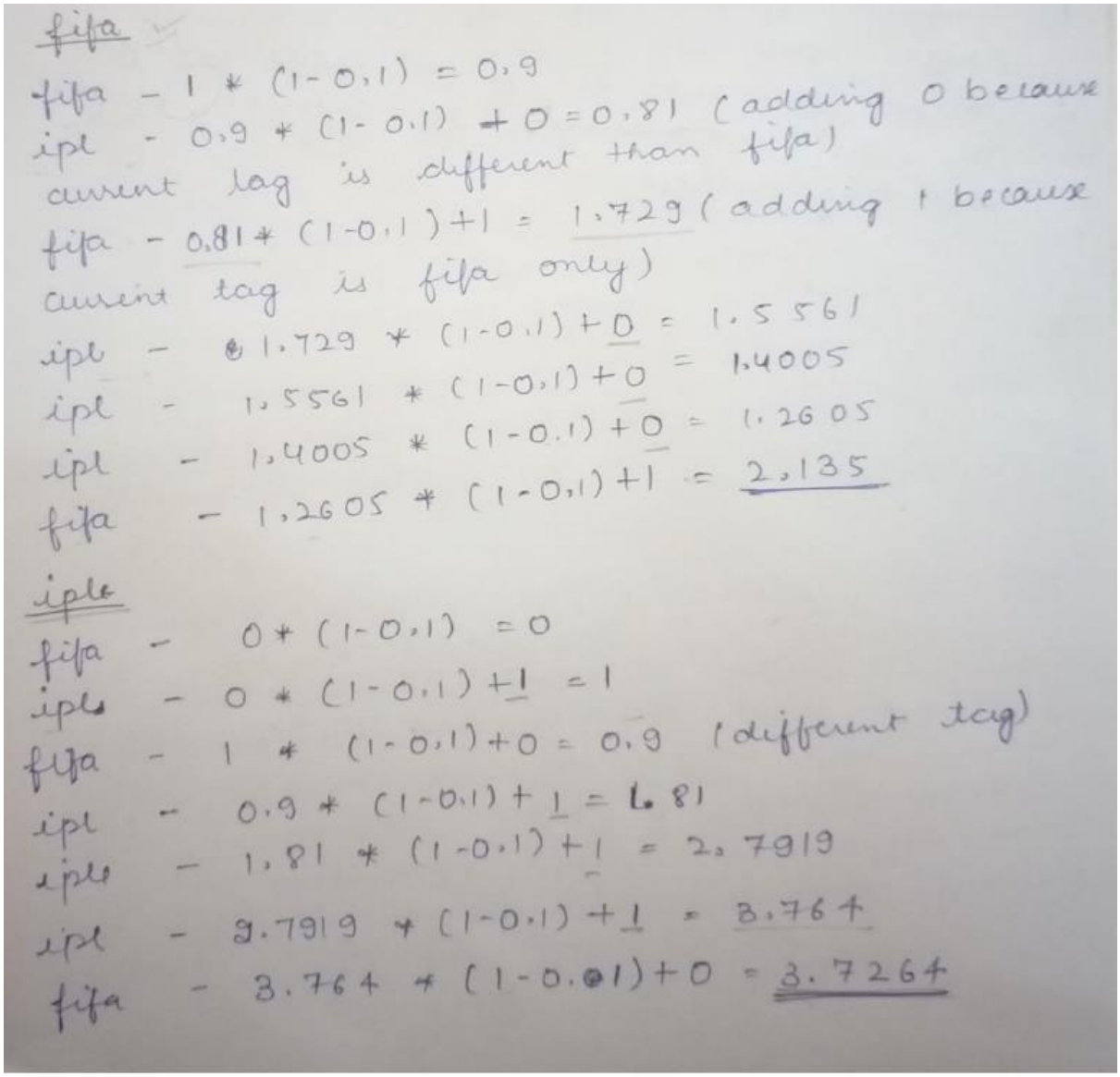
$3.67X_1 + 5.67X_2 = 8.22X_2$

On simplification, we get-

$5.3X_1 = 3.67X_2$ ……(1)
$3.67X_1 = 2.55X_2$ ……(2)

3 a) Elaborate the Decaying Window algorithm with graphical representation. Consider a sequence of Twitter tags below: {FIFA, IPL, FIFA, IPL, IPL, IPL, FIFA}
Let each element weight is 1, and constant c is 0.1. Find the most trending element in the given stream.

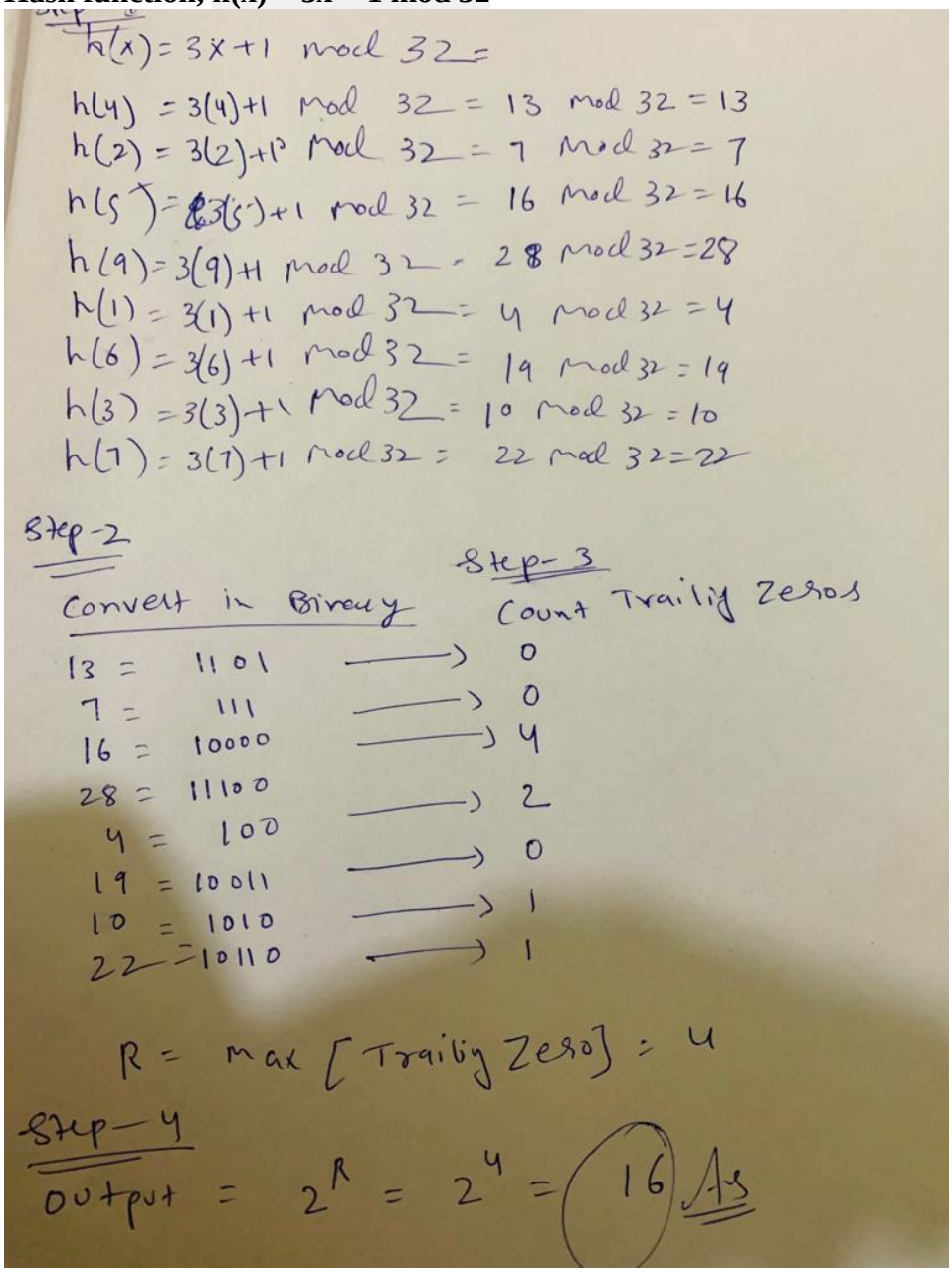
Solution:



In the end of the sequence, we can see the score of fifa is 2.135 but ipl is 3.7264.

So, ipl is more trending than fifa. Even though both of them occurred same number of times in input their score is still different.

3 b) Determine the distinct element in the stream using the Flajolet Martin algorithm.
Stream: 4, 2, 5 ,9, 1, 6, 3, 7
Hash function, h(x) = 3x + 1 mod 32



4 a) For the following given transaction data set, generate rules using the Apriori algorithm. Consider the values as Support=22% and Confidence= 70%.

Transaction-ID	Item Purchased
1	I1, I2, I5
2	I2, I4
3	I2, I3
4	I1, I2, I4
5	I1, I3
6	I2, I3
7	I1, I3
8	I1, I2, I3, I5
9	I1, I2, I3

Solution:

Given minimum support=22% and confidence=70%

Step 1) Find Frequent Item Set and their support

Item	Frequency	Support (in %)
I1	6	6/9=66%
I2	7	7/9=80%
I3	6	6/9=66%
I4	2	2/9=22.2%
I5	2	2/9=22.2%

Support (item) = Frequency of item/Number of transactions

Step 2) Remove all the items whose support is below given minimum support and form the two items candidate set and write their frequencies.

Item	Frequency	Support (in %)
I1,I2	4	4/9=44.4%
I1,I3	4	4/9=44.4%
I1,I4	1	1/9=11.1%
I1,I5	2	2/9=22.2%
I2,I3	4	4/9=44.4%
I2,I4	2	2/9=22.2%
I2,I5	2	2/9=22.2%
I3,I4	0	0/9=0%
I3,I5	1	1/9=11.1%
I4,I5	0	0/9=0%

4 b) Cluster the following eight points (with (x, y) representing locations) into three clusters:

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)
Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).
The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as-
P(a, b) = |x2 – x1| + |y2 – y1|
Use K-Means Algorithm to find the three cluster centers after the second iteration.

Solution: Part1

Iteration-01:

- We calculate the distance of each point from each of the center of the three clusters.
- The distance is calculated by using the given distance function.

The following illustration shows the calculation of distance between point A1(2, 10) and each of the center of the three clusters-

Calculating Distance Between A1(2, 10) and C1(2, 10)-

P(A1, C1)

= |x2 – x1| + |y2 – y1|

= |2 – 2| + |10 – 10|

= 0

Calculating Distance Between A1(2, 10) and C2(5, 8)-

P(A1, C2)

= |x2 – x1| + |y2 – y1|

= |5 – 2| + |8 – 10|

= 3 + 2

= 5

Calculating Distance Between A1(2, 10) and C3(1, 2)-

P(A1, C3)

= |x2 – x1| + |y2 – y1|

= |1 – 2| + |2 – 10|

= 1 + 8

= 9

In the similar manner, we calculate the distance of other points from each of the center of the three clusters.

Next,

- We draw a table showing all the results.
- Using the table, we decide which point belongs to which cluster.
- The given point belongs to that cluster whose center is nearest to it.

Given Points	**Distance from center (2, 10) of Cluster-01**	**Distance from center (5, 8) of Cluster-02**	**Distance from center (1, 2) of Cluster-03**	**Point belongs to Cluster**
A1(2, 10)	0	5	9	C1
A2(2, 5)	5	6	4	C3
A3(8, 4)	12	7	9	C2
A4(5, 8)	5	0	10	C2
A5(7, 5)	10	5	9	C2

A6(6, 4)	10	5	7	C2
A7(1, 2)	9	10	0	C3
A8(4, 9)	3	2	10	C2

From here, New clusters are-

Cluster-01:

First cluster contains points-

- A1(2, 10)

Cluster-02:

Second cluster contains points-

- A3(8, 4)
- A4(5, 8)
- A5(7, 5)
- A6(6, 4)
- A8(4, 9)

Cluster-03:

Third cluster contains points-

- A2(2, 5)
- A7(1, 2)

Now,

- We re-compute the new cluster clusters.
- The new cluster center is computed by taking mean of all the points contained in that cluster.

For Cluster-01:

We have only one point A1(2, 10) in Cluster-01.

- So, cluster center remains the same.

For Cluster-02:

Center of Cluster-02

= ((8 + 5 + 7 + 6 + 4)/5, (4 + 8 + 5 + 4 + 9)/5)

= (6, 6)

For Cluster-03:

Center of Cluster-03

= ((2 + 1)/2, (5 + 2)/2)

= (1.5, 3.5)

This is completion of Iteration-01.

Iteration-02:

We calculate the distance of each point from each of the center of the three clusters.

- The distance is calculated by using the given distance function.

The following illustration shows the calculation of distance between point A1(2, 10) and each of the center of the three clusters-

Calculating Distance Between A1(2, 10) and C1(2, 10)-

P(A1, C1)

= |x2 – x1| + |y2 – y1|

= |2 – 2| + |10 – 10|

= 0

Calculating Distance Between A1(2, 10) and C2(6, 6)-

P(A1, C2)

= |x2 – x1| + |y2 – y1|

= |6 – 2| + |6 – 10|

= 4 + 4

= 8

Calculating Distance Between A1(2, 10) and C3(1.5, 3.5)-

P(A1, C3)

= |x2 – x1| + |y2 – y1|

= |1.5 – 2| + |3.5 – 10|

= 0.5 + 6.5

= 7

In the similar manner, we calculate the distance of other points from each of the center of the three clusters.

Next,

- We draw a table showing all the results.
- Using the table, we decide which point belongs to which cluster.
- The given point belongs to that cluster whose center is nearest to it.

Given Points	Distance from center (2, 10) of Cluster-01	Distance from center (6, 6) of Cluster-02	Distance from center (1.5, 3.5) of Cluster-03	Point belongs to Cluster
A1(2, 10)	0	8	7	C1
A2(2, 5)	5	5	2	C3
A3(8, 4)	12	4	7	C2
A4(5, 8)	5	3	8	C2

A5(7, 5)	10	2	7	C2
A6(6, 4)	10	2	5	C2
A7(1, 2)	9	9	2	C3
A8(4, 9)	3	5	8	C1

From here, New clusters are-

Cluster-01:

First cluster contains points-

- A1(2, 10)
- A8(4, 9)

Cluster-02:

Second cluster contains points-

- A3(8, 4)
- A4(5, 8)
- A5(7, 5)
- A6(6, 4)

Cluster-03:

Third cluster contains points-

- A2(2, 5)
- A7(1, 2)

Now,

- We re-compute the new cluster clusters.
- The new cluster center is computed by taking mean of all the points contained in that cluster.

For Cluster-01:

Center of Cluster-01

= ((2 + 4)/2, (10 + 9)/2)

= (3, 9.5)

For Cluster-02:

Center of Cluster-02

= ((8 + 5 + 7 + 6)/4, (4 + 8 + 5 + 4)/4)

= (6.5, 5.25)

For Cluster-03:

Center of Cluster-03

= ((2 + 1)/2, (5 + 2)/2)

= (1.5, 3.5)

This is completion of Iteration-02.

After second iteration, the center of the three clusters are-

- C1(3, 9.5)
- C2(6.5, 5.25)
- C3(1.5, 3.5)

5 a) Visualization is needed to present the facts available in the unstructured datasets. Explain, in brief, the most challenging issues that occur during effective data visualization in real life.

Solution:

1. Usability

- The usability issue is critical to everyone, especially in light of successful commercialization stories Although the overall growth of information visualization is accelerating, the growth of usability studies and empirical evaluations has been relatively slow. Furthermore, usability issues still tend to be addressed in an ad hoc manner and limited to the particular systems at hand.

2. Understanding elementary perceptual–cognitive tasks

- Understanding elementary and secondary perceptual– cognitive tasks is a fundamental step toward engineering information visualization systems. The general understanding of elementary perceptual–cognitive tasks must be substantially revised and updated in the context of information visualization.

3. Prior knowledge

- This seemingly philosophical problem has many practical implications. As a vehicle for communicating abstract information, information visualization and its users must have a common ground. This is consistent with the user-centered design tradition in human–computer interaction (HCI).

4. Education and training

- The education problem is the fourth user-centered challenge. We are facing the challenge internally and externally. The internal aspect of the challenge refers to the need for researchers and practitioners within the field of information visualization to learn and share various principles and skills of visual communication and semiotics.

5. Intrinsic quality measures

- It's vital for the information visualization field to establish intrinsic quality metrics. Until recently, the lack of quantifiable quality measures has not been much of a concern. In part, this is because of the traditional priority of original and innovative work in this community. The lack of quantifiable measures of quality and benchmarks, however, will undermine information visualization advances, especially their evaluation and selection.

6. Scalability

- The scalability problem is a long-lasting challenge for information visualization. Unlike the field of scientific visualization, supercomputers have not been the primary source of data suppliers for information visualization. Parallel computing and other high-performance computing techniques have not been used in the field of information visualization as much as in scientific visualization and a few other fields. In addition to the traditional approach of developing increasingly clever ways to scale up sequential computing algorithms, the scalability issue should be studied at different levels—such as the hardware and the high-performance computing levels— as well as that of individual users.

7. Aesthetics

The purpose of information visualization is the insights into data that it provides, not just pretty pictures. But what makes a picture pretty? What can we learn from making a pretty picture and enhancing the representation of insights? It's important, therefore, to understand how insights and aesthetics interact, and how these two

5 b) Explain the various graphical techniques for visualizing the data facts and effective decision making.

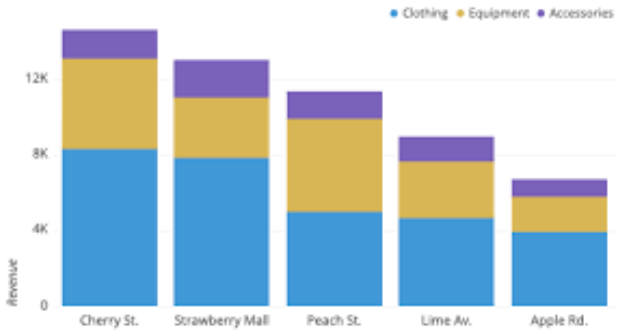
Solution:

1. Pie Chart

- Pie charts are one of the most common and basic data visualization techniques, used across a wide range of applications. Pie charts are ideal for illustrating proportions, or part-to-whole comparisons.
- Pie Chart Example
- Because pie charts are relatively simple and easy to read, they're best suited for audiences who might be unfamiliar with the information or are only interested in the key takeaways. For viewers who require a more thorough explanation of the data, pie charts fall short in their ability to display complex information.

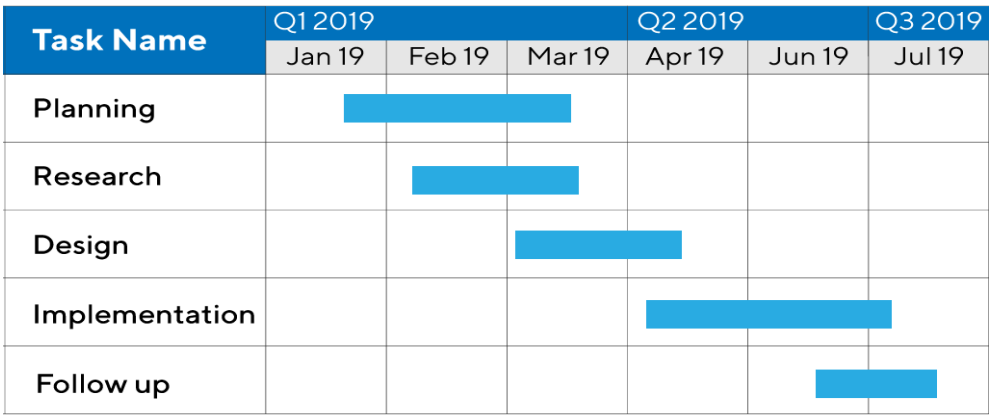
2. Bar Chart

- The classic bar chart, or bar graph, is another common and easy-to-use method of data visualization. In this type of visualization, one axis of the chart shows the categories being compared, and the other, a measured value. The length of the bar indicates how each group measures according to the value.
- One drawback is that labeling and clarity can become problematic when there are too many categories included. Like pie charts, they can also be too simple for more complex data sets.



3. **Gantt charts** are particularly common in project management, as they're useful in illustrating a project timeline or progression of tasks. In this type of chart, tasks to be performed are listed on the vertical axis and time intervals on the horizontal axis. Horizontal bars in the body of the chart represent the duration of each activity.Utilizing Gantt charts to display timelines can be incredibly helpful, and enable team members to keep track of every aspect of a project. Even if you're not a project management professional, familiarizing yourself with Gantt charts can help you stay organized.

Gantt Chart

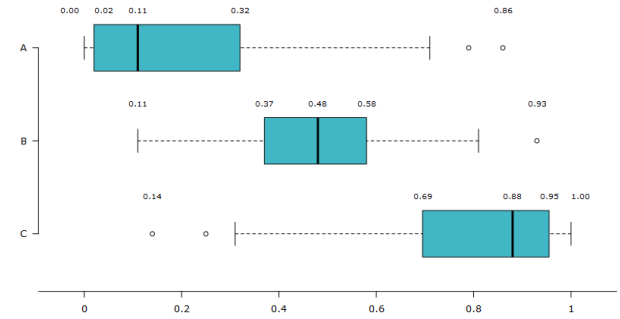


4. **Heat Map** is a type of visualization used to show differences in data through variations in color. These charts use color to communicate values in a way that makes it easy for the viewer to quickly identify trends. Having a clear legend is necessary in order for a user to successfully read and interpret a heatmap. There are many possible applications of heat maps. For example, if you want to analyze which time of day a retail store makes the most sales, you can use a heat map that shows the day of the week on the vertical axis and time of day on the horizontal axis. Then, by shading in the matrix with colors that correspond to the number of sales at each time of day, you can identify trends in the data that allow you to determine the exact times your store experiences the most sales.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Average Monthly Temperatures at Central Park, New York													
2		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
3	2009	27.9	36.7	42.4	54.5	62.5	67.5	72.7	75.7	66.3	55.0	51.2	35.9	
4	2010	32.5	33.1	48.2	57.9	65.3	74.7	81.3	77.4	71.1	58.1	47.9	32.8	
5	2011	29.7	36.0	42.3	54.3	64.5	72.3	80.2	75.3	70.0	57.1	51.9	43.3	
6	2012	37.3	40.9	50.9	54.8	65.1	71.0	78.8	76.7	68.8	58.0	43.9	41.5	
7	2013	35.1	33.9	40.1	53.0	62.8	72.7	79.8	74.6	67.9	60.2	45.3	38.5	
8	2014	28.6	31.6	37.7	52.3	64.0	72.5	76.1	74.5	69.7	59.6	45.3	40.5	
9	2015	29.9	23.9	38.1	54.3	68.5	71.2	78.8	79.0	74.5	58.0	52.8	50.8	
10	2016	34.5	37.7	48.9	53.3	62.8	72.3	78.7	79.2	71.8	58.8	49.8	38.3	
11	2017	38.0	41.6	39.2	57.2	61.1	72.0	76.8	74.0	70.5	64.1	46.6	33.4	
12														

5. **Box and Whisker Plot**, or box plot, provides a visual summary of data through its quartiles. First, a box is drawn from the first quartile to the third of the data set. A line within the box represents the median. “Whiskers,” or lines, are then drawn extending from the box to the minimum (lower extreme) and maximum (upper extreme). Outliers are represented by individual points that are in-line with the whiskers.
This type of chart is helpful in quickly identifying whether or not the data is symmetrical or skewed, as well as providing a visual summary of the data set that can be easily interpreted.

Chart 4.5.2.1
Box and whisker plots and five-number summaries of distributions A, B and C



6. **Scatter Plot:** Another technique commonly used to display data is a scatter plot. A scatter plot displays data for two variables as represented by points plotted against the horizontal and vertical axis. This type of data visualization is useful in illustrating the relationships that exist between variables and can be used to identify trends or correlations in data. Scatter plots are most effective for fairly large data sets, since it's often easier to identify trends when there are more data points present. Additionally, the closer the data points are grouped together, the stronger the correlation or trend tends to be.