

NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI-15
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
B.Tech. DEGREE, VI SEMESTER, I CYCLE TEST, FEB-2022
CSPE64 - Data Analytics

DATE: 23 -02-2022 TIME: 10:00 a.m. -11:00 a.m. MAX.MARKS:20 Marks

Answer all Questions

5 x 4 =20 marks

1. The table below shows the demand for a particular brand of fax machine in a department store in each of the last twelve months.

| | | | | | | | | | | | | |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|
| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Demand | 12 | 15 | 19 | 23 | 27 | 30 | 32 | 33 | 37 | 41 | 49 | 58 |

- a. Calculate the four months moving average for months 4 to 12. What would be your forecast for the demand in month 13?
 - b. Apply exponential smoothing with a smoothing constant of 0.2 to derive a forecast for the demand in month 13.
 - c. Which of the two forecasts for month 13 do you prefer and why?
2. Suppose that the data for analysis includes the attribute age. The age values for the 10 data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.
- a. Give the five-number summary of the data.
 - b. Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
 - c. Show a boxplot of the data.
3. Explain the stream processor architecture for the Twitter data
4. i) Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2, 6, 5. Our hash functions will all be of the form $h(x) = ax + b \text{ mod } 32$ for some a and b . You should treat the result as a 5-bit binary integer. Determine the tail length for each stream element and the resulting estimate of the number of distinct elements if the hash function is:
- (a) $h(x) = 2x + 1 \text{ mod } 32$.
 - (b) $h(x) = 3x + 7 \text{ mod } 32$.
 - (c) $h(x) = 4x \text{ mod } 32$.
- ii) Using DGIM algorithm there are several ways that the bit-stream 1001011011101 could be partitioned into buckets. Find all of them.
5. Compute the surprise number (second moment) for the stream 3, 1, 4, 1, 3, 4, 2, 1, 2. What is the third moment of this stream? Apply the Alon-Matias-Szegedy Algorithm to estimate the surprise number. For each possible value of i , if X_i is a variable starting position i , what is the value of X_i .value?