

END SEMESTER EXAMINATION, JUNE- 2025**B. Sc. (Hons.) in Data Science Semester - IV****Course Name: Introduction to Artificial Intelligence****Course Code: BDS242****Time allotted: 03 Hours****Max. Marks: 100****Note: Read all the instructions carefully.****Q.1. Attempt all.****(10×2=20)**

		CO	PO
a.	What is a heuristic function in AI search algorithms?	1	1
b.	What is depth-first search, and when is it useful?	1	2
c.	Parsing mean.....	2	1
d.	Regular grammar is Stronger than CFG. True/False	2	1
e.	Prolog is a OOP type of language. True/False	3	2
f.	Zero order logic is also known as.....	3	1
g.	Why Fuzzy sets are required?	4	2
h.	Fuzzy sets are used to represent Boolean logic. True/False	4	2
i.	What is speech recognition?	5	1
j.	Write one application of machine learning	5	1

Q.2. Attempt any five questions. Each carries equal marks**(5X4=20)**

a.	How a 8-Puzzle problem can be solved using informed search. Show with an example.	1	1
b.	Explain Bottom-up Parsing with the help of an example	2	2
c.	What are the advantages and disadvantage of logical representation of knowledge?	3	1
d.	How union operation is performed in Fuzzy set? Show with the help of an example.	3	1
e.	What is a pattern recognition? Where it can be used?	5	2
f.	Write the differences between top-down parsing and bottom-up parsing.	2	2
g.	What are the problems associated with Hill-climbing algorithm? How they can be solved?	1	3

P.T.O/-

Q3. Attempt any three questions. Each carries equal marks.

(3X10=30)

a.	(i) Write three application of pattern recognition. (ii) Define Training and testing data? (iii) How many Phases are required in Pattern Recognition System?	5	2
b.	(i) Write the two differences between Informed and Uninformed search (ii) Explain the Heuristic function (iii) Who is considered as father of AI?	1	1
c.	(i) Why first order logic is required? (ii) Elaborate the Case grammar? (iii) How forward chaining is used?	3	2
d.	(i) What is Inference engine? (ii) Differentiate between digital controller and fuzzy controller (iii) Name any two-membership function	3	2

Q4. Attempt any two questions. Each carries equal marks.

(2X15=30)

a.	(i) Compare DFS with BFS. (ii) What is minimax algorithm? (iii) Why pruning is required in it?	1	2
b.	(i), Why Fillmore grammar is required? (ii) Explain about Transition nets in detail.	3	2
c.	(i) Write two advantages and disadvantages of Fuzzy logic. (ii) Explain the use of fuzzy logic taking the example of a real-life scenario.	4	1

END SEMESTER EXAMINATION, JUNE- 2025**B. Sc. (Hons.) in Data Science Semester - IV****Course Name: Estimation and Inferential Statistics****Course Code: BDS241**

Time allotted: 03 Hours

Max. Marks: 100

Note: Read all the instructions carefully.

Q.1. Objectives Type**(10 × 2=20)****(Very short/ Short Answer, MCQ, One-liners, Fill in the Blank, True/ False) Attempt all.**

S. No.	Question	CO	PO
a.	Alternative hypothesis	3	1
b.	Wider confidence intervals indicate more uncertainty about the estimate. (T/F)	3	1
c.	Confidence interval gives an exact range where the population parameter lies with 100% certainty. (T/F)	2	1
d.	In the chi-square test of independence if $r \times c$ table contains 4 rows and 4 columns, then find the degrees of freedom.	5	1
e.	Write down the properties of a good estimator.	1	1
f.	The degree of freedom for the independent t test is _____.	4	1
g.	Write the confidence interval for the population mean when sample size is small and variance of population is unknown.	1	1
h.	Give two real world examples where paired t test is applied.	4	2
i.	In a certain testing of hypothesis concerning differences of two means if samples are of sizes $n_1 = 14$ and $n_2 = 4$, then the test statistics to be used is _____.	4	1
j.	Confidence interval	1	1

Q.2. Short notes / Short Answer Type (Attempt any five out of Seven)**(5 × 4=20)**

a.	Define sufficiency. Let $X_1, X_2, X_3, \dots, X_n$ be a random sample from normal population with unknown mean and known population variance. Find the sufficient estimator for population mean.	1	1
b.	Explain the hypothesis testing procedure of t test of single mean.	2	2
c.	A random sample of size $n=10$ is taken from a population. Given that the sample mean is 0.5060 with a standard deviation of 0.0040 construct a 95% confidence interval for the population mean.	2	2
d.	Define Neyman Factorization theorem for sufficiency.	1	1
e.	If 12 determinations of the specific heat of iron have a standard deviation of 0.0086, test the null hypothesis that $\sigma = 0.010$ for such determinations against the alternative hypothesis $\sigma \neq 0.010$ at 0.01 level of significance.	2	2
f.	Explain the hypothesis testing procedure for Wilcoxon signed rank test.	5	1
g.	Two random samples of measurements of the heat-producing capacity of specimens of coal from two mines are taken with respective variances 15,750 and 10,920. Use 0.02 level of significance to test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_1^2 \neq \sigma_2^2$.	4	2

P.T.O./-

Q.3. Structured Question (Attempt any three out of four)

a.	Show that $\frac{\sum x_i(\sum x_i - 1)}{n(n-1)}$ is an unbiased estimator of θ^2 , for the sample $x_1, x_2, \dots, x_n$ drawn on X which takes the values 1 or 0 with respective probabilities θ and $(1 - \theta)$.	1	2
b.	(i) The following data constitute a random sample of 15 measurement of the octane rating of a certain kind of gasoline: 97.5, 95.2, 97.3, 96.0, 96.8, 100.3, 97.4, 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5, 94.9. Use signed rank test at 0.05 level of significance to test whether the mean octane rating of the given kind of gasoline is 98.5. Given critical value of statistic is 1.34. (ii) The following are the average weekly losses of worker-hours due to accidents in industrial plants before and after certain safety program was put into operation: Before: 7, 11, 14, 11, 12, 6, 9, 8, 10, 11, 13, 7, 8, After: 10, 13, 14, 15, 9, 10, 13, 11, 11, 15, 11, 10, 8 Use Wilcoxon sign rank test at 0.05 level of significance to test whether the safety program is effective. Given that critical value of a statistic is 21.	5	2
c.	The following random samples are measurements of the heat-producing capacity of specimens of coal from two mines: Mine 1: 59, 68, 44, 71, 63, 46, 69, 54, 48 Mine 2: 50, 36, 62, 52, 70, 41. Use the 0.05 level of significance to test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 \neq 0$. Given the critical value of the t statistic is 2.10.	4	2
d.	Explain the hypothesis testing procedure of F statistic for equality of two population variances.	4	2

Q.4. Long Answer / Essay type (Attempt any two out of three)

(2 X 15=30)

a. To determine whether there really is a relationship between an employee's performance in the company's training program and his or her ultimate success in the job, the company takes a sample of cases from its very extensive files and obtain the results shown in the following table:

Performance in training program

		Below average	Average	Above average
Success in job (employer's rating)	Poor	63	42	15
	Average	58	61	31
	Very good	14	47	29

Use the 0.05 level of significance to test whether performance in the training program and success in the job are associated with each other. Given that Chi square critical value is 9.48.

b. The following are the weights gains of two random samples of young turkeys fed two different diets but otherwise kept under identical conditions:
 Diet 1: 16.3, 10.1, 10.7, 13.5, 14.9, 11.8, 14.3, 10.2, 12, 14.7, 23.6, 15.1, 14.5, 18.4, 13.2, 14
 Diet 2: 21.3, 23.8, 15.4, 19.6, 12, 13.9, 18.8, 19.2, 15.3, 20.1, 14.8, 18.9, 20.7, 21.1, 15.8, 16.2.
 Use Mann Whitney U test at the 0.01 level of significance to test the null hypothesis that the two populations sampled are identical against the alternative hypothesis that on the average the second diet produces a greater gain in weight. Given that critical chi square value of U statistic is 12.

c. Define maximum likelihood estimator, and the procedure. Further, find the maximum likelihood estimator for the parameter of Poisson distribution.

5	1
5	2
3	2

END SEMESTER EXAMINATION, JUNE- 2025
B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Applied Data Science using Python

Course Code: BDS243

Time allotted: 03 Hours

Max. Marks: 100

Note: Read all the instructions carefully.

Q.1. Objectives Type

Attempt all.

(10×2=20)

a.	Create a tuple containing the names of four countries and print each country name along with its index.	CO1	PO1
b.	Differentiate set and list with example	CO1	PO1
c.	What is the output of print tuple1 + tuple2 if tuple1 = ('abcd', 1234 , 2.93, 'john', 80.2) and tuple2 = (128, 'john')?	CO2	PO1,4
d.	Perform element-wise addition of two NumPy arrays.	CO3	PO1,2
e.	Multiply each element of a NumPy array by a scalar value.	CO2	PO1,2,4
f.	Describe how to identify unique values in a pandas Series.	CO4	PO1,2,4,5
g.	How do you count the occurrences of each unique value in a pandas Series?	CO4	PO1,2,
h.	Describe common strategies for handling missing values in Python.	CO5	PO1,2
i.	How does the presence of missing values affect data analysis and modeling?	CO5	PO1
j.	Explain the concept of basic indexing and slicing in NumPy arrays	CO3	PO1

Q.2. Short notes Type (Attempt any five out of Seven)

(5×4=20)

a.	What are dictionaries in Python, and how do they differ from lists? Explain with examples how dictionaries are useful for storing and accessing data.	CO1	PO1,2
b.	Explain the concept of missing values in data analysis and discuss common strategies for handling missing data in Python. Provide examples demonstrating how missing values can be identified and handled using Pandas.	CO2	PO1,2
c.	Given a NumPy array arr = np.array([10, 20, 30, 40, 50]), calculate the cumulative sum of the elements along the array and print the result.	CO3	PO1,2,3
d.	Sort the DataFrame based on values in a specific column in ascending order. Rank the DataFrame based on values in a different column.	CO4	PO1
e.	Discuss the importance of data formatting in data analysis and machine learning tasks.	CO5	PO1,2
f.	Briefly explain map() and lambda ()	CO2	PO1,2
g.	Discuss the use of Scikit-learn	CO3	PO1,2

P.T.O./-

Q.3. Structured Question (Attempt any three out of four)

(3×10=30)

a.	Describe the following: (i) Creating the list (ii) Accessing values in the lists (iii) Updating the list (iv) Deleting the list elements	CO1	PO1,2
b.	(i) What do you mean by visualization? (ii) Develop line chart for the following data. years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3].	CO5	PO1,2,3
c.	(i) Describe how indexing, selection, and filtering work in Pandas DataFrames. (ii) Explain the concept of the Series data structure in the Pandas library. Discuss its characteristics, including index labels and data alignment, and provide examples of creating and manipulating Series objects.	CO4	PO1,2,3
d.	Write Python code to generate a random matrix of size 5x5 using NumPy. Then, calculate the sum of all elements in the matrix. Explain how NumPy simplifies such numerical operations compared to traditional Python lists.	CO3	PO1,2,3

Q.4. Long Answer type (Attempt any two out of three)

(2×15=30)

Q.3. Given following type (Attempt any two out of three)					(2×15=30)																																																								
a.	<table><thead><tr><th>ProductID</th><th>Quantity</th><th>Revenue</th><th>CustomerID</th><th>Region</th></tr></thead><tbody><tr><td>Product3</td><td>NaN</td><td>2739.948652</td><td>8504</td><td>Region3</td></tr><tr><td>Product3</td><td>NaN</td><td>1679.149019</td><td>2571</td><td>Region3</td></tr><tr><td>Product3</td><td>NaN</td><td>882.511665</td><td>3199</td><td>Region3</td></tr><tr><td>Product2</td><td>NaN</td><td>1511.998497</td><td>8698</td><td>Region1</td></tr><tr><td>Product3</td><td>47.0</td><td>4605.004648</td><td>6212</td><td>Region2</td></tr><tr><td>Product2</td><td>NaN</td><td>562.870815</td><td>1181</td><td>Region1</td></tr><tr><td>Product1</td><td>18.0</td><td>288.524524</td><td>8069</td><td>Region3</td></tr><tr><td>Product1</td><td>31.0</td><td>740.447113</td><td>5023</td><td>Region1</td></tr><tr><td>Product3</td><td>36.0</td><td>2509.334012</td><td>5387</td><td>Region2</td></tr><tr><td>Product2</td><td>44.0</td><td>3657.869004</td><td>3569</td><td>Region3</td></tr></tbody></table> i. Define Missing Values? ii. Write a code to Fill missing values by mean of Quantity Attribute. iii. Write the output of print(dataset.isnull().count()) iv. Write a code to replace Region1 with Dehradun, Region2 with Rishikesh and Region3 with Haridwar. Calculate descriptive statistics (mean, median and standard deviation) for Quantity and Revenue.				ProductID	Quantity	Revenue	CustomerID	Region	Product3	NaN	2739.948652	8504	Region3	Product3	NaN	1679.149019	2571	Region3	Product3	NaN	882.511665	3199	Region3	Product2	NaN	1511.998497	8698	Region1	Product3	47.0	4605.004648	6212	Region2	Product2	NaN	562.870815	1181	Region1	Product1	18.0	288.524524	8069	Region3	Product1	31.0	740.447113	5023	Region1	Product3	36.0	2509.334012	5387	Region2	Product2	44.0	3657.869004	3569	Region3	CO4	PO1,2
ProductID	Quantity	Revenue	CustomerID	Region																																																									
Product3	NaN	2739.948652	8504	Region3																																																									
Product3	NaN	1679.149019	2571	Region3																																																									
Product3	NaN	882.511665	3199	Region3																																																									
Product2	NaN	1511.998497	8698	Region1																																																									
Product3	47.0	4605.004648	6212	Region2																																																									
Product2	NaN	562.870815	1181	Region1																																																									
Product1	18.0	288.524524	8069	Region3																																																									
Product1	31.0	740.447113	5023	Region1																																																									
Product3	36.0	2509.334012	5387	Region2																																																									
Product2	44.0	3657.869004	3569	Region3																																																									
b.	(i) What are dictionaries in Python, and how do they differ from lists? Explain with examples how dictionaries are useful for storing and accessing data. (ii) Discuss String Manipulations with examples				CO1	PO1,2,3																																																							
c.	Write on the following: (i) Scikit (ii) Matplotlib (iii) NumPy				CO5,2	PO1,2																																																							

END SEMESTER EXAMINATION, JUNE- 2025**B. Sc. (Hons.) in Data Science Semester – IV****Course Name: Design and Analysis of Algorithms****Course Code: BDS245****Time allotted: 03 Hours****Max. Marks: 100****Note: Read all the instructions carefully.****Q.1. Attempt all.****(10×2=20)**

QN	Questions	CO	PO
a.	How an algorithm's time efficiency is measured?	1	2
b.	What is an Exhaustive Search?	2	1
c.	Define the basic principles of backtracking.	3	2
d.	How to calculate the efficiency of Dijkstra's algorithm?	2	2
e.	Write the running time of 0/1 knapsack problem.	2	1
f.	Define Travelling salesman problem.	2	2
g.	Explain Warshalls algorithm	2	1
h.	What is the time complexity of binary search in worst case?	3	2
i.	What is a pivot?	2	2
j.	Define Knapsack Problem.	2	1

Q.2. Attempt any five out of Seven.**(5×4=20)**

QN	Questions	CO	PO
a.	Illustrate Merge sort algorithm and discuss its time complexity.	1	2
b.	What are the fundamental steps involved in algorithmic problem solving?	2	1
c.	Write and explain bellman ford Generation Algorithm?	3	2
d.	Compare and contrast FIFO and LC branch – and – bound search techniques.	2	2
e.	Explain binary search with three order traversals with example	2	1
f.	Solve a travelling sales person problem using dynamic programming.	2	2
g.	Explain in detail the Graph Traversals.	2	1

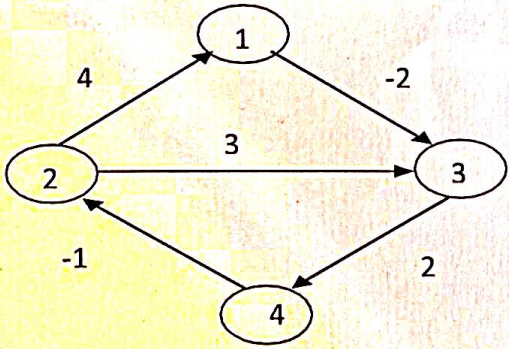
P.T.O./-

Q.3. Attempt any three out of four.

QN	Questions	CO	PO
a.	I. Write about 0/1 knapsack problem with an example. II. Explain Strassen's matrix multiplication and its time complexity	2	2
b.	I. Write a greedy algorithm for sequencing unit time jobs with dead lines and profits. II. Simulate binary search algorithm for the following apply binary search to searching for k=56 in the following array? {10,12,24,29,39,40,51,56,69}	2	1
c.	I. Describe Travelling Salesperson Problem (TSP) using Dynamic Programming. II. Distinguish between Prim's and Kruskal's spanning tree algorithm III. Explain the steps in Building a Huffman Tree. Find the codes for the alphabets given below according to the frequency. - (Space) 4 A 2 E 5 H 1 I 2 L 2 M 2 P 2 R 1 S 2 X 1	3	2
d.	I. What is meant by principle of optimality? II. Write in detail about Hamiltonian cycles. Give example to it. III. Explain the concepts of P, NP and NP completeness.	3	2

Q.4. Attempt any two out of three.

(2×15=30)

QN	Questions	CO	PO
a.	I. Explain General method of Greedy method. Find the greedy solution for following job sequencing with deadlines problem $n = 7$, $(p_1, p_2, p_3, p_4, p_5, p_6, p_7) = (3, 5, 20, 18, 1, 6, 30)$, $(d_1, d_2, d_3, d_4, \dots, d_7) = (1, 3, 4, 3, 2, 1, 2)$ II. Write an algorithm for Binary search and discuss its complexity	2	1
b.	I. Find an optimal solution to the knapsack instance $n=7$ objects and the capacity of knapsack $m=15$. The profits and weights of the objects are $(P_1, P_2, P_3, P_4, P_5, P_6, P_7) = (10, 5, 15, 7, 6, 18, 3)$ $(W_1, W_2, W_3, W_4, W_5, W_6, W_7) = (2, 3, 5, 7, 1, 4, 1)$. II. Explain and compare best and worst time complexity of Quick Sort.	3	2
c.	I. Give a suitable example and explain the breadth first search and depth first search algorithms. II. Write the Floyd algorithm to find all pairs shortest path and derive its time complexity. Solve the following using Floyd's algorithm. 	3	2

Enrol. No.

S	R	H	U																
---	---	---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Regn. No.

D	D																		
---	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

END SEMESTER EXAMINATION, JUNE- 2025
B. Sc. (Hons.) in Data Science Semester - IV

Course Name: Operating System**Course Code: BDS244****Time allotted: 03 Hours****Max. Marks: 100****Note: Read all the instructions carefully.****Q.1. Attempt All questions. Each carries equal marks.****(10x2=20)**

a.	What is a time-sharing system?	CO1	PO1
b.	Define Operating System.	CO1	PO1
c.	What is deadlock?	CO2	PO2
d.	What is semaphore?	CO2	PO1
e.	What is physical address space and physical address?	CO4	PO1
f.	Hard disk is a secondary memory. True/False	CO4	PO4
g.	List out the different file access methods	CO5	PO1
h.	Define process	CO5	PO1
i.	List out the CPU scheduling algorithms.	CO1	PO1
j.	CSCAN is disk scheduling algorithm. True/False	CO4	PO4

Q.2. Attempt any five questions. Each carries equal marks**(5x4=20)**

a.	Define operating system and its functions.	CO1	PO1
b.	What is CPU scheduling? Write its type and explain them?	CO2	PO2
c.	What is critical section problem? What are three conditions to satisfy?	CO2	PO2
d.	Explain paging and virtual memory,	CO3	PO3
e.	Compare UNIX operating system and Windows operating system.	CO2	PO3
f.	Explain attribute and operation of files.	CO5	PO4
g.	Explain disk and disk scheduling algorithms.	CO4	PO5

Q3. Attempt any three questions. Each carries equal marks.**(3x10=30)**

a.	Explain Operating system components in detail.	CO2	PO2
b.	What are the different Page replacement techniques? Explain any two.	CO3	PO2
c.	What are the CPU scheduling algorithm? Explain SRTF and RR with example.	CO3	PO1
d.	Consider a reference string: 4, 7, 6, 1, 7, 6, 1, 2, 7, 2. The number of frames in the memory is 3. Find out the number of page faults in Optimal, LRU and FIFO Page Replacement Algorithm	CO3	PO2

P.T.O./-

Q4. Attempt any two questions. Each carries equal marks. (2x15=30)

a.	Consider five processes P1, P2, P3, P4 and P5. The processes are assumed to have arrived in the order P3, P2, P1, P4, P5. Draw Gantt charts and calculate turnaround time & waiting time of each using Priority pre-emptive, non pre-emptive priority (a smaller priority number implies a higher priority), and RR (quantum = 1) scheduling. Also find which scheduling have minimal average waiting time? <table><tr><th>PROCESS</th><th>ARRIVAL TIME</th><th>BURST TIME</th><th>PRIORITY</th></tr><tr><td>P1</td><td>2</td><td>10</td><td>3</td></tr><tr><td>P2</td><td>4</td><td>1</td><td>1</td></tr><tr><td>P3</td><td>0</td><td>2</td><td>3</td></tr><tr><td>P4</td><td>1</td><td>1</td><td>4</td></tr><tr><td>P5</td><td>3</td><td>5</td><td>2</td></tr></table>	PROCESS	ARRIVAL TIME	BURST TIME	PRIORITY	P1	2	10	3	P2	4	1	1	P3	0	2	3	P4	1	1	4	P5	3	5	2	CO2	PO3																																																																		
PROCESS	ARRIVAL TIME	BURST TIME	PRIORITY																																																																																										
P1	2	10	3																																																																																										
P2	4	1	1																																																																																										
P3	0	2	3																																																																																										
P4	1	1	4																																																																																										
P5	3	5	2																																																																																										
b.	What is deadlock? What are the necessary conditions for a deadlock. Explain deadlock avoidance and recovery mechanisms.	CO2	PO4																																																																																										
c.	Using Banker's algorithm, answer the following questions. <table><tr><th rowspan="2">PROCESS</th><th colspan="4">Allocation</th><th colspan="4">Max</th><th colspan="4">Available</th></tr><tr><th>A</th><th>B</th><th>C</th><th>D</th><th>A</th><th>B</th><th>C</th><th>D</th><th>A</th><th>B</th><th>C</th><th>D</th></tr><tr><td>P0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>2</td><td>1</td><td>0</td><td>1</td><td>5</td><td>2</td><td>0</td></tr><tr><td>P1</td><td>1</td><td>2</td><td>3</td><td>1</td><td>1</td><td>6</td><td>5</td><td>2</td><td></td><td></td><td></td><td></td></tr><tr><td>P2</td><td>1</td><td>3</td><td>6</td><td>5</td><td>2</td><td>3</td><td>6</td><td>6</td><td></td><td></td><td></td><td></td></tr><tr><td>P3</td><td>0</td><td>6</td><td>3</td><td>2</td><td>0</td><td>6</td><td>5</td><td>2</td><td></td><td></td><td></td><td></td></tr><tr><td>P4</td><td>0</td><td>0</td><td>1</td><td>4</td><td>0</td><td>6</td><td>5</td><td>6</td><td></td><td></td><td></td><td></td></tr></table> a. What is the content of the matrix Need? b. Is the system in a safe state? c. If a request from process P1 arrives for (0,4,2,0), can the request be granted immediately? d. How many resources of type A, B, C, D are there?	PROCESS	Allocation				Max				Available				A	B	C	D	A	B	C	D	A	B	C	D	P0	0	1	1	0	0	2	1	0	1	5	2	0	P1	1	2	3	1	1	6	5	2					P2	1	3	6	5	2	3	6	6					P3	0	6	3	2	0	6	5	2					P4	0	0	1	4	0	6	5	6					CO3	PO5
PROCESS	Allocation				Max				Available																																																																																				
	A	B	C	D	A	B	C	D	A	B	C	D																																																																																	
P0	0	1	1	0	0	2	1	0	1	5	2	0																																																																																	
P1	1	2	3	1	1	6	5	2																																																																																					
P2	1	3	6	5	2	3	6	6																																																																																					
P3	0	6	3	2	0	6	5	2																																																																																					
P4	0	0	1	4	0	6	5	6																																																																																					
