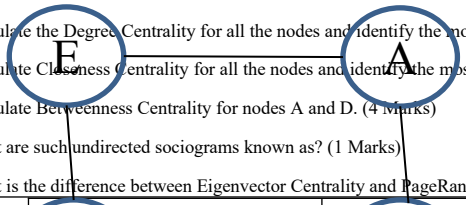
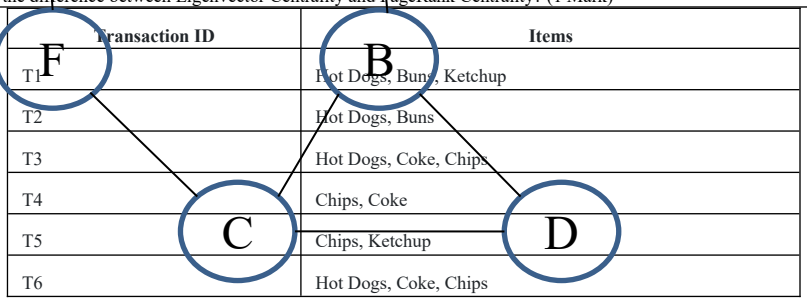
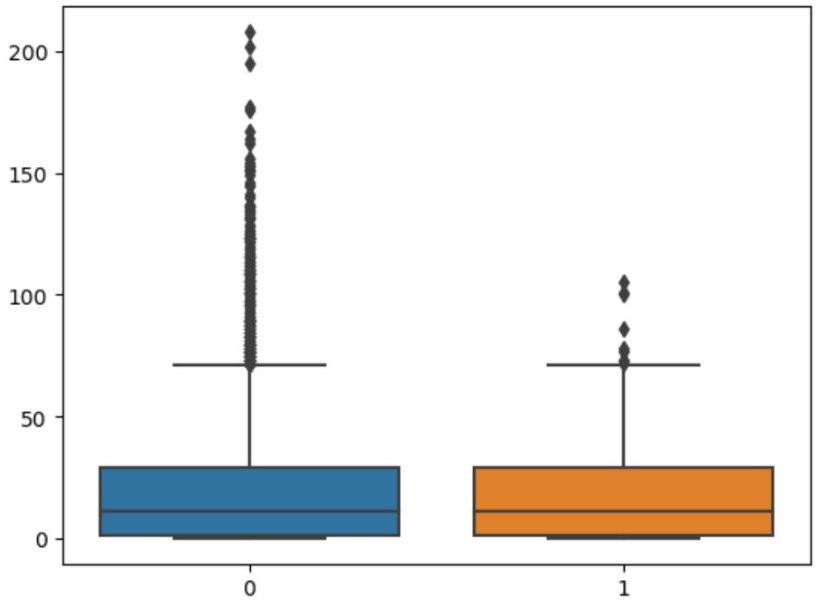
SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: Jan 24 to Apr 24 **Maximum Marks: 50 Examination: ESE Exam Date: 1st April, 2024 Duration: 3 Hrs**		
Programme code: 01 **Programme: MBA BA Minor**	**Class:** SY	**Semester/Trimester: VI**
College: K. J. Somaiya Institute of Management	**Name of the department/Section/Center: Business Analytics**	
Course Code: 217P01M623	**Name of the Course: Application Based Analytics**	

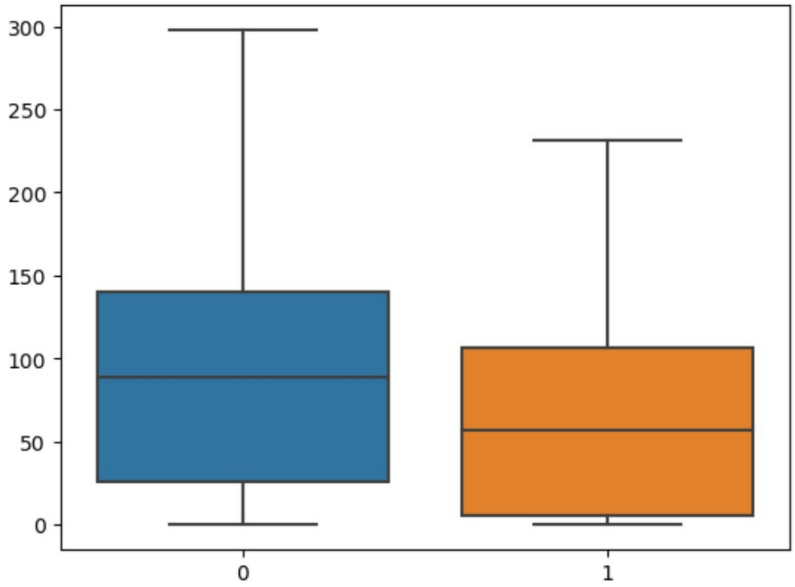
Instructions:

1. **Attempt any five questions.**
2. **All questions carry equal marks.**
3. **Make suitable assumptions if required and state them.**
4. **Use of a calculator is permitted.**

Question No.		Max. Marks
1	Based on the following sociogram:  a. Calculate the Degree Centrality for all the nodes and identify the most influential node(s). (2 Marks) b. Calculate Closeness Centrality for all the nodes and identify the most significant node(s). (2 Marks) c. Calculate Betweenness Centrality for nodes A and D. (4 Marks) d. What are such undirected sociograms known as? (1 Marks) e. What is the difference between Eigenvector Centrality and PageRank Centrality? (1 Mark)	10
2	 **Transaction ID** / **Items** T1 — Hot Dogs, Buns, Ketchup T2 — Hot Dogs, Buns T3 — Hot Dogs, Coke, Chips T4 — Chips, Coke T5 — Chips, Ketchup T6 — Hot Dogs, Coke, Chips a. Use an appropriate algorithm to find frequent items bought together by using minimum support of 2 and 3. Show step-by-step iteration. (8 Marks) b. Which algorithm have you used and why? (2 Marks)	10
3	A. Discuss any five pillars of People Analytics. B. Which of the following (a or b) will you choose as a feature to detect the anomaly and why?	5 + 5



a



b

4 Explain the following Python codes and identify the business problem(s) where the codes are used: 10

	Code	Explanation	Business Problem
a	from networkx.algorithms.community.modularity_max import greedy_modularity_communities c = list(greedy_modularity_communities(G_symmetric)) print(len(c))		
b	Q1 = df.Number_of_cards.quantile(0.25) Q3 = df.Number_of_cards.quantile(0.75) df_box = df[(df.Number_of_cards <Q1 - 1.5 * (Q3-Q1))| (df. Number_of_cards > Q3 + 1.5 * (Q3-Q1))] df_box		
c	from sklearn import preprocessing le = preprocessing.LabelEncoder() for i in categorical: df[i]=le.fit_transform(df[i])		
d	rules = association_rules(frequent_items, metric = 'lift', min_threshold = 1)		

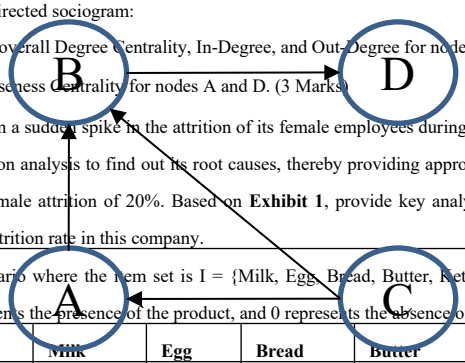
		rules.sort_values('confidence', ascending = False, inplace = True) rules			
	e	import nltk nltk.downloader.download('vader_lexicon') from nltk.sentiment import SentimentIntensityAnalyzer			

5

A. For the following directed sociogram:

a. Calculate the overall Degree Centrality, In-Degree, and Out-Degree for node B. (2 Marks)

b. Calculate Closeness Centrality for nodes A and D. (3 Marks)



B. A Company has seen a sudden spike in the attrition of its female employees during the last year, and the company's CEO wants HR to perform an attrition analysis to find out its root causes, thereby providing appropriate solutions. The company decided to have a stringent goal of female attrition of 20%. Based on **Exhibit 1**, provide key analysis and insights and a few recommendations to reduce the female attrition rate in this company.

5 + 5

6

Consider a Big Bazar scenario where the item set is I = {Milk, Egg, Bread, Butter, Ketchup, Cookies}. The database comprises twelve transactions where 1 represents the presence of the product, and 0 represents the absence of the product.

Transaction ID	Milk	Egg	Bread	Butter	Ketchup	Cookies
t1	1	1	1	1	0	0
t2	1	1	0	1	1	0
t3	0	0	1	1	1	0
t4	1	0	1	1	0	0
t5	0	0	1	1	0	1
t6	1	0	1	1	0	1
t7	1	0	0	0	0	1
t8	1	0	1	1	0	0
t9	0	1	1	1	0	1
t10	1	0	1	1	0	0
t11	1	0	1	1	0	0
t12	1	0	1	0	1	1

Considering the following association rules:

Association Rules	
Antecedent	**Consequent**
{Milk, Bread}	{Butter}
{Cookies}	{Milk, Bread}
{Egg}	{Bread}

a. Compute support, confidence, and lift for the association rule Antecedent → Consequent. (8 Marks)

b. Identify the frequent item set from the set of association rules and justify your selection. (2 Marks)

10

Exhibit 1

Dept		
Finance	HR	Marketing
Operations	Sales	Technology

Gender	
Female	Male

Title	
Manager	Team Leader
Team Member	VP

Sum of Employees	Column Labels												
Row Labels	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Grand Total
Headcount	8	4	10	14	15	15	14	14	15	15	15	15	154
Terminations	3	2											5
Grand Total	**11**	**6**	**10**	**14**	**15**	**15**	**14**	**14**	**15**	**15**	**15**	**15**	**159**
Avg HCT	8	6	7	9	10	11	11	12	12	12	13	13	13
YTD Terminations	3	5	5	5	5	5	5	5	5	5	5	5	5

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	YTD
YTD Attrition %	38%	83%	68%	56%	49%	45%	44%	43%	41%	40%	40%	39%	39%
Goal 20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%