

MSc. I(DS)/02.22.004 Reg. No

--	--	--	--	--	--	--	--



M.Sc. COMPUTER SCIENCE WITH SPECIALISATION IN DATA SCIENCE
FIRST SEMESTER EXAMINATION, FEBRUARY 2022
20-359-0104 PYTHON FOR DATA ANALYTICS
(Regular)

Time : 3 Hours

Maximum Marks:50

(Answer ANY FIVE questions)
 Each question carries EQUAL Marks

QUESTIONS		MARKS																																																																																																																					
I.	1. Explain the pipeline of Data Analysis Process 2. Write short note on a) Quantitative and Qualitative Data Analysis b) Scipy	6 4																																																																																																																					
II	Write short notes on with necessary examples a) Pandas Index. b) Difference between pandas Series and DataFrame.	5 5																																																																																																																					
III	With examples, describe the data preparation phase of data manipulation.	10																																																																																																																					
IV	<p>From the sample dataset given below:</p> <table><tr><th>month_number</th><th>facecream</th><th>facewash</th><th>toothpaste</th><th>bathingsoap</th><th>shampoo</th><th>moisturizer</th><th>total_units</th><th>total_profit</th></tr><tr><td>1</td><td>2500</td><td>1500</td><td>5200</td><td>9200</td><td>1200</td><td>1500</td><td>21100</td><td>211000</td></tr><tr><td>2</td><td>2630</td><td>1200</td><td>5100</td><td>6100</td><td>2100</td><td>1200</td><td>18330</td><td>183300</td></tr><tr><td>3</td><td>2140</td><td>1340</td><td>4550</td><td>9550</td><td>3550</td><td>1340</td><td>22470</td><td>224700</td></tr><tr><td>4</td><td>3400</td><td>1130</td><td>5870</td><td>8870</td><td>1870</td><td>1130</td><td>22270</td><td>222700</td></tr><tr><td>5</td><td>3600</td><td>1740</td><td>4560</td><td>7760</td><td>1560</td><td>1740</td><td>20960</td><td>209600</td></tr><tr><td>6</td><td>2760</td><td>1555</td><td>4890</td><td>7490</td><td>1890</td><td>1555</td><td>20140</td><td>201400</td></tr><tr><td>7</td><td>2980</td><td>1120</td><td>4780</td><td>8980</td><td>1780</td><td>1120</td><td>29550</td><td>295500</td></tr><tr><td>8</td><td>3700</td><td>1400</td><td>5860</td><td>9960</td><td>2860</td><td>1400</td><td>36140</td><td>361400</td></tr><tr><td>9</td><td>3540</td><td>1780</td><td>6100</td><td>8100</td><td>2100</td><td>1780</td><td>23400</td><td>234000</td></tr><tr><td>10</td><td>1990</td><td>1890</td><td>8300</td><td>10300</td><td>2300</td><td>1890</td><td>26670</td><td>266700</td></tr><tr><td>11</td><td>2340</td><td>2100</td><td>7300</td><td>13300</td><td>2400</td><td>2100</td><td>41280</td><td>412800</td></tr><tr><td>12</td><td>2900</td><td>1760</td><td>7400</td><td>14400</td><td>1800</td><td>1760</td><td>30020</td><td>300200</td></tr></table> <p>a) Write a program to read all product sales data and show it using a multiline plot? Ensure all the necessary attributes like title, labels, and legend include.</p>	month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit	1	2500	1500	5200	9200	1200	1500	21100	211000	2	2630	1200	5100	6100	2100	1200	18330	183300	3	2140	1340	4550	9550	3550	1340	22470	224700	4	3400	1130	5870	8870	1870	1130	22270	222700	5	3600	1740	4560	7760	1560	1740	20960	209600	6	2760	1555	4890	7490	1890	1555	20140	201400	7	2980	1120	4780	8980	1780	1120	29550	295500	8	3700	1400	5860	9960	2860	1400	36140	361400	9	3540	1780	6100	8100	2100	1780	23400	234000	10	1990	1890	8300	10300	2300	1890	26670	266700	11	2340	2100	7300	13300	2400	2100	41280	412800	12	2900	1760	7400	14400	1800	1760	30020	300200	5
month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit																																																																																																															
1	2500	1500	5200	9200	1200	1500	21100	211000																																																																																																															
2	2630	1200	5100	6100	2100	1200	18330	183300																																																																																																															
3	2140	1340	4550	9550	3550	1340	22470	224700																																																																																																															
4	3400	1130	5870	8870	1870	1130	22270	222700																																																																																																															
5	3600	1740	4560	7760	1560	1740	20960	209600																																																																																																															
6	2760	1555	4890	7490	1890	1555	20140	201400																																																																																																															
7	2980	1120	4780	8980	1780	1120	29550	295500																																																																																																															
8	3700	1400	5860	9960	2860	1400	36140	361400																																																																																																															
9	3540	1780	6100	8100	2100	1780	23400	234000																																																																																																															
10	1990	1890	8300	10300	2300	1890	26670	266700																																																																																																															
11	2340	2100	7300	13300	2400	2100	41280	412800																																																																																																															
12	2900	1760	7400	14400	1800	1760	30020	300200																																																																																																															

	b) Write a program to read face cream and facewash product sales data and show it using the horizontal bar chart? Ensure all the necessary attributes like title, labels, and legend include.	5																																								
V	<p>Reduce the Iris flower dataset from 4 dimension to 3 dimension or 2 dimension using Principal Component Analysis (PCA) and plot the result in a 3D/2D scatterplot.</p> <table><thead><tr><th></th><th>sepal length</th><th>sepal width</th><th>petal length</th><th>petal width</th></tr></thead><tbody><tr><td>0</td><td>5.1</td><td>3.5</td><td>1.4</td><td>0.2</td></tr><tr><td>1</td><td>4.9</td><td>3.0</td><td>1.4</td><td>0.2</td></tr><tr><td>2</td><td>4.7</td><td>3.2</td><td>1.3</td><td>0.2</td></tr><tr><td>3</td><td>4.6</td><td>3.1</td><td>1.5</td><td>0.2</td></tr><tr><td>..</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>148</td><td>6.2</td><td>3.4</td><td>5.4</td><td>2.3</td></tr><tr><td>149</td><td>5.9</td><td>3.0</td><td>5.1</td><td>1.8</td></tr></tbody></table> <p>[150 rows x 4 columns]</p>		sepal length	sepal width	petal length	petal width	0	5.1	3.5	1.4	0.2	1	4.9	3.0	1.4	0.2	2	4.7	3.2	1.3	0.2	3	4.6	3.1	1.5	0.2	148	6.2	3.4	5.4	2.3	149	5.9	3.0	5.1	1.8	10
	sepal length	sepal width	petal length	petal width																																						
0	5.1	3.5	1.4	0.2																																						
1	4.9	3.0	1.4	0.2																																						
2	4.7	3.2	1.3	0.2																																						
3	4.6	3.1	1.5	0.2																																						
..																																						
148	6.2	3.4	5.4	2.3																																						
149	5.9	3.0	5.1	1.8																																						
VI	Implement K-Nearest Neighbor Classifier without using <code>train_test_split()</code> function for splitting the dataset.	10																																								
VII	<p>a) From the sample data given below, draw a pie chart with a slice extracted (Transport) from the pie and include the necessary function to find the percentage represented by each slice.</p> <table><thead><tr><th>Expenses</th><th>Amount</th></tr></thead><tbody><tr><td>Rent</td><td>7000</td></tr><tr><td>Grocery</td><td>3000</td></tr><tr><td>Transport</td><td>800</td></tr><tr><td>Current</td><td>300</td></tr><tr><td>School fee</td><td>2000</td></tr><tr><td>Savings</td><td>1900</td></tr></tbody></table> <p>b) Give an example of plotting pie Charts with a Pandas Data Frame.</p>	Expenses	Amount	Rent	7000	Grocery	3000	Transport	800	Current	300	School fee	2000	Savings	1900	5																										
Expenses	Amount																																									
Rent	7000																																									
Grocery	3000																																									
Transport	800																																									
Current	300																																									
School fee	2000																																									
Savings	1900																																									