

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	Implement Decision Tree Model using Iris dataset using Python/R and interpret decision rules of classification.	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	Load Iris Dataset. Apply K-means Algorithm using Python/R to group similar data points into clusters. Determine optimal number of clusters using Elbow Method . Visualize clustering results and analyze cluster characteristics.	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	Load Iris Dataset. Apply K-means Algorithm using Python/R to group similar data points into clusters. Determine optimal number of clusters using Silhouette analysis . Visualize clustering results and analyze cluster characteristics.	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Consider a scenario where you have test scores from a sample of students and you want to compare the mean of these scores with hypothesized population mean.</p> <p>Student Score = [72, 88, 64, 74, 67, 79, 85, 75, 89,77]</p> <p>Apply One Sampled T-Test using Python/R for above problem. Assume hypothesized mean as 70. Formulate Null and Alternative Hypothesis for a given problem. Interpret the results and draw the conclusion. [20]</p> <p>B. Apply Feature Scaling technique like standardization and normalization using Python/R to Boston Housing dataset. [20]</p>	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. The employee's aptitude and job proficiency score is as follows.</p> <table><tr><td>aptitude</td><td>85</td><td>65</td><td>50</td><td>68</td><td>87</td><td>74</td><td>65</td><td>96</td><td>68</td><td>94</td><td>73</td><td>84</td><td>85</td><td>87</td><td>91</td></tr><tr><td>jobprof</td><td>70</td><td>90</td><td>80</td><td>89</td><td>88</td><td>86</td><td>78</td><td>67</td><td>86</td><td>90</td><td>92</td><td>94</td><td>99</td><td>93</td><td>87</td></tr></table> <p>Perform Chi-Square Test to study correlation between aptitude and job proficiency of employee. Formulate Null and Alternative Hypothesis for a given problem. Interpret results and draw conclusions based on test outcome. [20]</p> <p>B. Perform Logistic Regression on the Iris dataset using Python/R to predict binary outcome. Evaluate model's performance using classification metrics. [20]</p>	aptitude	85	65	50	68	87	74	65	96	68	94	73	84	85	87	91	jobprof	70	90	80	89	88	86	78	67	86	90	92	94	99	93	87	40
aptitude	85	65	50	68	87	74	65	96	68	94	73	84	85	87	91																			
jobprof	70	90	80	89	88	86	78	67	86	90	92	94	99	93	87																			
2.	Viva	05																																
3.	Journal	05																																

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.

A. Create CSV file from given data. Read the data from CSV files into a data frame. Perform **Data pre-processing** tasks such as handling missing values and outliers using Python/R [20]

Country	Age	Salary	Purchased
France	44	72000	No
Spain	27	48000	Yes
Germany	30	54000	No
Spain	38	61000	No
Germany	40		Yes
France	35	58000	Yes
Spain		52000	No
France	48	79000	Yes
Germany	50	83000	No
France	37	67000	Yes

B. Implement **Multiple Linear Regression** using Python/R on the below housing dataset to predict price of a house. Evaluate model's performance using classification metrics. [20]

Bedrooms	Bathrooms	Sqft_living	Floors	Grade	Sqft_above	Sqft_basement	Price
3	1	1180	1	7	1180	0	221900
3	2.25	2570	2	7	2170	400	538000
2	1	770	1	6	770	0	180000
4	3	1960	1	7	1050	910	604000
3	2	1680	1	8	1680	0	510000
4	4.5	5420	1	11	3890	1530	267800
3	2.25	1715	2	7	1715	0	257500
3	1.5	1060	1	7	1060	0	291850
3	1	1780	1	7	1050	730	229500
3	2.5	1890	2	7	1890	0	323000
3	2.5	3560	1	8	1860	1700	662500

2.

Viva

05

3.

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Apply Feature Scaling technique like standardization and normalization using Python/R to numerical features of below dataset. [20]</p> <table><tr><th>Make</th><th>Model</th><th>Color</th><th>Mileage</th><th>Sell Price</th></tr><tr><td>Honda</td><td>Accord</td><td>Red</td><td>63,512</td><td>4000</td></tr><tr><td>Honda</td><td>Accord</td><td>Blue</td><td>95,135</td><td>2500</td></tr><tr><td>Toyota</td><td>Camry</td><td>Black</td><td>75,006</td><td>45000</td></tr><tr><td>Nissan</td><td>Altima</td><td>Green</td><td>69,847</td><td>3826</td></tr><tr><td>Toyota</td><td>Corolla</td><td>Black</td><td>87,278</td><td>2224</td></tr><tr><td>Honda</td><td>Civic</td><td>White</td><td>1,38,789</td><td>2723</td></tr><tr><td>Ford</td><td>F-150</td><td>Black</td><td>89,073</td><td>3950</td></tr><tr><td>Chevrolet</td><td>Silverado</td><td>Green</td><td>1,09,231</td><td>4959</td></tr><tr><td>Chevrolet</td><td>Impala</td><td>Silver</td><td>87,675</td><td>3791</td></tr><tr><td>Dodge</td><td>Charger</td><td>Silver</td><td>34,853</td><td>4349</td></tr><tr><td>Dodge</td><td>Charger</td><td>Silver</td><td>58,173</td><td>4252</td></tr></table> <p>B. Implement Multiple Linear Regression on the “Pima Indian Diabetes dataset” using Python/R. [20]</p>	Make	Model	Color	Mileage	Sell Price	Honda	Accord	Red	63,512	4000	Honda	Accord	Blue	95,135	2500	Toyota	Camry	Black	75,006	45000	Nissan	Altima	Green	69,847	3826	Toyota	Corolla	Black	87,278	2224	Honda	Civic	White	1,38,789	2723	Ford	F-150	Black	89,073	3950	Chevrolet	Silverado	Green	1,09,231	4959	Chevrolet	Impala	Silver	87,675	3791	Dodge	Charger	Silver	34,853	4349	Dodge	Charger	Silver	58,173	4252	40
Make	Model	Color	Mileage	Sell Price																																																										
Honda	Accord	Red	63,512	4000																																																										
Honda	Accord	Blue	95,135	2500																																																										
Toyota	Camry	Black	75,006	45000																																																										
Nissan	Altima	Green	69,847	3826																																																										
Toyota	Corolla	Black	87,278	2224																																																										
Honda	Civic	White	1,38,789	2723																																																										
Ford	F-150	Black	89,073	3950																																																										
Chevrolet	Silverado	Green	1,09,231	4959																																																										
Chevrolet	Impala	Silver	87,675	3791																																																										
Dodge	Charger	Silver	34,853	4349																																																										
Dodge	Charger	Silver	58,173	4252																																																										
2.	Viva	05																																																												
3.	Journal	05																																																												

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.

A. Consider the dataset given below.

40

Make	Model	Color	Mileage	Sell Price	Buy Price
Honda	Accord	Red	63,512	4000	3000
Honda	Accord	Blue	95,135	2500	2000
Toyota	Camry	Black	75,006	45000	44000
Nissan	Altima	Green	69,847	3826	3000
Toyota	Corolla	Black	87,278	2224	2100
Honda	Civic	White	1,38,789	2723	1900
Ford	F-150	Black	89,073	3950	3000
Chevrolet	Silverado	Green	1,09,231	4959	4500
Chevrolet	Impala	Silver	87,675	3791	3500
Dodge	Charger	Silver	34,853	4349	3500
Dodge	Charger	Silver	58,173	4252	4000

Create Pivot Table in Excel for following analysis and visualize the data using PivotChart: [20]

- How many cars do you have by make and model and by color?
- Find out profit margin of all different Make of cars.
- Find out average cost of vehicles.
- Find out percentage of cars of each color.

B. Perform **Logistic Regression** on the Iris dataset using Python/R to predict binary outcome. Evaluate model's performance using classification metrics. [20]

2.

Viva

05

3.

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Apply Feature Scaling technique like standardization and normalization using Python/R to numerical features of below dataset. [20]</p> <div><div><div></div></div><table><tr><th>Country</th><th>Age</th><th>Salary</th><th>Purchased</th></tr><tr><td>France</td><td>44</td><td>72000</td><td>No</td></tr><tr><td>Spain</td><td>27</td><td>48000</td><td>Yes</td></tr><tr><td>Germany</td><td>30</td><td>54000</td><td>No</td></tr><tr><td>Spain</td><td>38</td><td>61000</td><td>No</td></tr><tr><td>Germany</td><td>40</td><td>85000</td><td>Yes</td></tr><tr><td>France</td><td>35</td><td>58000</td><td>Yes</td></tr><tr><td>Spain</td><td>31</td><td>52000</td><td>No</td></tr><tr><td>France</td><td>48</td><td>79000</td><td>Yes</td></tr><tr><td>Germany</td><td>50</td><td>83000</td><td>No</td></tr><tr><td>France</td><td>37</td><td>67000</td><td>Yes</td></tr></table></div> <p>B. Load the Iris dataset. Perform Principal component Analysis (PCA) using Python/R on a dataset to reduce dimensionality. Select appropriate number of principle components. Visualize the data in the reduced-dimensional space. [20]</p>	Country	Age	Salary	Purchased	France	44	72000	No	Spain	27	48000	Yes	Germany	30	54000	No	Spain	38	61000	No	Germany	40	85000	Yes	France	35	58000	Yes	Spain	31	52000	No	France	48	79000	Yes	Germany	50	83000	No	France	37	67000	Yes	40
Country	Age	Salary	Purchased																																											
France	44	72000	No																																											
Spain	27	48000	Yes																																											
Germany	30	54000	No																																											
Spain	38	61000	No																																											
Germany	40	85000	Yes																																											
France	35	58000	Yes																																											
Spain	31	52000	No																																											
France	48	79000	Yes																																											
Germany	50	83000	No																																											
France	37	67000	Yes																																											
2.	Viva	05																																												
3.	Journal	05																																												

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<div>A. Convert Categorical Column to numerical representation (Feature Dummification) of below dataset using Python/R. [20]</div> <div><div></div><table><tr><th>Country</th><th>Age</th><th>Salary</th><th>Purchased</th></tr><tr><td>France</td><td>44</td><td>72000</td><td>No</td></tr><tr><td>Spain</td><td>27</td><td>48000</td><td>Yes</td></tr><tr><td>Germany</td><td>30</td><td>54000</td><td>No</td></tr><tr><td>Spain</td><td>38</td><td>61000</td><td>No</td></tr><tr><td>Germany</td><td>40</td><td>85000</td><td>Yes</td></tr><tr><td>France</td><td>35</td><td>58000</td><td>Yes</td></tr><tr><td>Spain</td><td>31</td><td>52000</td><td>No</td></tr><tr><td>France</td><td>48</td><td>79000</td><td>Yes</td></tr><tr><td>Germany</td><td>50</td><td>83000</td><td>No</td></tr><tr><td>France</td><td>37</td><td>67000</td><td>Yes</td></tr></table></div> <div>B. Perform Logistic Regression on the Iris dataset using Python/R to predict binary outcome. Evaluate model's performance using classification metrics. [20]</div>	Country	Age	Salary	Purchased	France	44	72000	No	Spain	27	48000	Yes	Germany	30	54000	No	Spain	38	61000	No	Germany	40	85000	Yes	France	35	58000	Yes	Spain	31	52000	No	France	48	79000	Yes	Germany	50	83000	No	France	37	67000	Yes	40
Country	Age	Salary	Purchased																																											
France	44	72000	No																																											
Spain	27	48000	Yes																																											
Germany	30	54000	No																																											
Spain	38	61000	No																																											
Germany	40	85000	Yes																																											
France	35	58000	Yes																																											
Spain	31	52000	No																																											
France	48	79000	Yes																																											
Germany	50	83000	No																																											
France	37	67000	Yes																																											
2.	Viva	05																																												
3.	Journal	05																																												

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Create CSV file from given data. Read the data from CSV files into a data frame.</p> <table><tr><th>Make</th><th>Model</th><th>Color</th><th>Mileage</th><th>Sell Price</th><th>Buy Price</th></tr><tr><td>Honda</td><td>Accord</td><td>Red</td><td>63,512</td><td>4000</td><td>3000</td></tr><tr><td>Honda</td><td>Accord</td><td>Blue</td><td>95,135</td><td>2500</td><td>2000</td></tr><tr><td>Toyota</td><td>Camry</td><td>Black</td><td>75,006</td><td>45000</td><td>44000</td></tr><tr><td>Nissan</td><td>Altima</td><td>Green</td><td>69,847</td><td>3826</td><td>3000</td></tr><tr><td>Toyota</td><td>Corolla</td><td>Black</td><td>87,278</td><td>2224</td><td>2100</td></tr><tr><td>Honda</td><td>Civic</td><td>White</td><td>1,38,789</td><td>2723</td><td>1900</td></tr><tr><td>Ford</td><td>F-150</td><td>Black</td><td>89,073</td><td>3950</td><td>3000</td></tr><tr><td>Chevrolet</td><td>Silverado</td><td>Green</td><td>1,09,231</td><td>4959</td><td>4500</td></tr><tr><td>Chevrolet</td><td>Impala</td><td>Silver</td><td>87,675</td><td>3791</td><td>3500</td></tr><tr><td>Dodge</td><td>Charger</td><td>Silver</td><td>34,853</td><td>4349</td><td>3500</td></tr><tr><td>Dodge</td><td>Charger</td><td>Silver</td><td>58,173</td><td>4252</td><td>4000</td></tr></table> <p>Perform transformation functions on given data using Python/R. [20]</p> <ul style="list-style-type: none">• Display records of the car having Sell Price greater than 4000.• Sort the car data in ascending order.• Group the data according the “Make” of car. <p>B. Load the Iris dataset. Perform Principal component Analysis (PCA) using Python/R on a dataset to reduce dimensionality. Select appropriate number of principle components. Visualize the data in the reduced-dimensional space. [20]</p>	Make	Model	Color	Mileage	Sell Price	Buy Price	Honda	Accord	Red	63,512	4000	3000	Honda	Accord	Blue	95,135	2500	2000	Toyota	Camry	Black	75,006	45000	44000	Nissan	Altima	Green	69,847	3826	3000	Toyota	Corolla	Black	87,278	2224	2100	Honda	Civic	White	1,38,789	2723	1900	Ford	F-150	Black	89,073	3950	3000	Chevrolet	Silverado	Green	1,09,231	4959	4500	Chevrolet	Impala	Silver	87,675	3791	3500	Dodge	Charger	Silver	34,853	4349	3500	Dodge	Charger	Silver	58,173	4252	4000	40
Make	Model	Color	Mileage	Sell Price	Buy Price																																																																					
Honda	Accord	Red	63,512	4000	3000																																																																					
Honda	Accord	Blue	95,135	2500	2000																																																																					
Toyota	Camry	Black	75,006	45000	44000																																																																					
Nissan	Altima	Green	69,847	3826	3000																																																																					
Toyota	Corolla	Black	87,278	2224	2100																																																																					
Honda	Civic	White	1,38,789	2723	1900																																																																					
Ford	F-150	Black	89,073	3950	3000																																																																					
Chevrolet	Silverado	Green	1,09,231	4959	4500																																																																					
Chevrolet	Impala	Silver	87,675	3791	3500																																																																					
Dodge	Charger	Silver	34,853	4349	3500																																																																					
Dodge	Charger	Silver	58,173	4252	4000																																																																					
2.	Viva	05																																																																								
3.	Journal	05																																																																								

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Consider you have a dataset that contains the exam scores of students from three different classes: A, B, and C.</p> <p>Class A = [85, 90, 88, 82, 87]</p> <p>Class B = [76, 78, 80, 81, 75]</p> <p>Class C = [92, 88, 94, 89, 90]</p> <p>Perform One Way ANOVA Test using Python/R to determine if there is a significant difference in the mean exam scores among these classes. Formulate Null and Alternative Hypothesis for a given problem. Interpret results and draw conclusions based on test outcome. [20]</p> <p>B. Load the Wine Quality dataset. Perform Principal component Analysis (PCA) using Python/R on a dataset to reduce dimensionality. Select appropriate number of principle components. Visualize the data in the reduced-dimensional space. [20]</p>	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. The company is accessing the difference in time to complete the task between two groups of employees. State the hypothesis and do the Two Sampled T-Test using Python/R for the given dataset. [20]</p> <p>Group-I: Experience(0-1 year)</p> <p>Group-II: Experience(1-2 year)</p> <p>Time taken by Group1: 85,95,100,80,90,97,104,95,88,92,94,99</p> <p>Time taken by Group2: 83,85,96,92,100,104,94,95,88,90,93,94</p> <p>B. Perform Multiple Linear Regression on the “Pima Indian Diabetes dataset” using Python/R. [2-]</p>	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1

Construct a **Decision Tree** using Python/R to classify whether a person can play tennis or not. Test the model and check prediction of the model is correct or not. The following data set recorded whether or not a person played tennis depending on the outlook and wind conditions.

- Each instance (example) is represented by the three attributes.
 - o Outlook: a value of {Sunny, Overcast, Rain}.
 - o Wind: a value of {Weak, Strong}.
 - o PlayTennis: the classification attribute (i.e., Yes- the person plays tennis; No - the person does not play tennis).

Date	Outlook	Wind	PlayTennis
1	Sunny	Weak	No
2	Sunny	Strong	No
3	Overcast	Weak	Yes
4	Rain	Weak	Yes
5	Rain	Weak	Yes
6	Rain	Strong	No
7	Overcast	Strong	Yes
8	Sunny	Weak	No
9	Sunny	Weak	Yes
10	Rain	Weak	Yes

Date	Outlook	Wind	PlayTennis
11	Sunny	Strong	Yes
12	Overcast	Strong	Yes
13	Overcast	Weak	Yes
14	Rain	Strong	No
15	Sunny	Strong	Yes
16	Overcast	Strong	No
17	Overcast	Weak	Yes
18	Rain	Weak	No
19	Sunny	Weak	No
20	Rain	Strong	Yes

2

Viva

05

3

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.

A. Implement **Decision Tree Model** on Titanic dataset using Python/R and interpret decision rules of classification. Perform **Linear Regression** on the following dataset in Python/R for predicting the weight of the person depending on height. [20]

Height	151	174	138	186	128	136	179	163	152
weight	63	81	56	91	47	57	76	72	62

B. Create CSV file from given data. Read the data from CSV files into a data frame.

Make	Model	Color	Mileage	Sell Price	Buy Price
Honda	Accord	Red	63,512	4000	3000
Honda	Accord	Blue	95,135	2500	2000
Toyota	Camry	Black	75,006	45000	44000
Nissan	Altima	Green	69,847	3826	3000
Toyota	Corolla	Black	87,278	2224	2100
Honda	Civic	White	1,38,789	2723	1900
Ford	F-150	Black	89,073	3950	3000
Chevrolet	Silverado	Green	1,09,231	4959	4500
Chevrolet	Impala	Silver	87,675	3791	3500
Dodge	Charger	Silver	34,853	4349	3500
Dodge	Charger	Silver	58,173	4252	4000

Perform transformation function on given data using Python/R. [20]

- Display records of the car having Buy Price greater than equal to 3000.
- Sort the car data in ascending order.
- Group the data according to the “Model” of car.

40

2.

Viva

05

3.

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.

A. Perform **Linear Regression** on the following dataset in Python/R for predicting the salary of the person depending on his/her years of experience. [20]

Years of Experience	Salary
2	30000
10	95000
4	45000
20	178000
8	84000
12	120000
22	200000

B. Consider dataset given below.

Supplier ID	Part Number	Part Name	Part Price	Status
SP301	A001	water	6800	In
SP302	A002	altenator	3800	In
SP303	A003	air filter	4500	In
		wheel		In
SP304	A004	bearing	3582	stock
SP305	A005	muffler	1600	In
				Out of
SP306	A006	oil pan	1005	stock
SP307	A007	brake pads	6500	In
		brake		Out of
SP308	A008	rotors	8549	stock
SP309	A009	headlight	6500	In
SP310	A010	brake	1500	In
SP311	A011	Strut	4500	In
SP312	A012	Deive	1580	In
		CV Boot		In
SP313	A013	Kit	2650	stock
SP314	A014	Oil Pump	4660	In
SP315	A015	Oil Filter	4350	In
SP316	A016	Fuel Filter	1280	In
		Tie Rod		In
SP317	A017	End	1800	stock
SP318	A018	Ball Joint	2500	In
		Steering		Out of
SP319	A019	Rack	2700	stock
				Out of
SP320	A020	Piston	4500	stock

Apply VLOOKUP function to retrieve information for the following queries. Also write down the steps for the same. [20]

- Find the Part Name for Part Number “A002”.
- Find the Supplier ID for the Part Name “Ball Joint”.
- Find the Part Price for Part Name “muffler”.
- Find the Status of Part Number “A008”.

2.

Viva

05

3.

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Perform Linear Regression on the Iris dataset of R/Python for predicting petal.width on petal.length. [20]</p> <p>B. A student is enrolled in the class. His /Her current grade is 65 which is average of all type of work during the semester and he/she needs atleast 72 to pass this class. Final exam is not yet completed. Apply what-if analysis using Goal seek to determine the marks need to obtained in Final Exam to complete the goal. [20]</p> <table><tr><th>Semester Work</th><th>Grade</th></tr><tr><td>Paper Presentation</td><td>58</td></tr><tr><td>Case Study</td><td>70</td></tr><tr><td>Assignment</td><td>72</td></tr><tr><td>Practical</td><td>60</td></tr><tr><td>Final Exam</td><td>60</td></tr><tr><td>Final Grade</td><td>65</td></tr></table>	Semester Work	Grade	Paper Presentation	58	Case Study	70	Assignment	72	Practical	60	Final Exam	60	Final Grade	65	40
Semester Work	Grade															
Paper Presentation	58															
Case Study	70															
Assignment	72															
Practical	60															
Final Exam	60															
Final Grade	65															
2.	Viva	05														
3.	Journal	05														

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.

A. Perform Linear Regression on the Iris dataset of R/Python for predicting petal.width on petal.length. [20]

B. Consider the dataset given below.

Supplier ID	Part Number	Part Name	Part Price	Status
SP301	A001	water	6800	In
SP302	A002	altenator	3800	In
SP303	A003	air filter	4500	In
SP304	A004	wheel bearing	3582	In stock
SP305	A005	muffler	1600	In
SP306	A006	oil pan	1005	Out of stock
SP307	A007	brake pads	6500	In
SP308	A008	brake rotors	8549	Out of stock
SP309	A009	headlight	6500	In
SP310	A010	brake	1500	In
SP311	A011	Strut	4500	In
SP312	A012	Deive	1580	In
SP313	A013	CV Boot Kit	2650	In stock
SP314	A014	Oil Pump	4660	In
SP315	A015	Oil Filter	4350	In
SP316	A016	Fuel Filter	1280	In
SP317	A017	Tie Rod End	1800	In stock
SP318	A018	Ball Joint	2500	In
SP319	A019	Steering Rack	2700	Out of stock
SP320	A020	Piston	4500	Out of stock

Apply VLOOKUP function to retrieve information for the following queries. Also write down the steps for the same. [20]

1. Find the Part Name for Part Number “A016” .

2. Find the Supplier ID for the Part Name “Oil Pump”.

3. Find the Part Price for Part Name “Brake”.

4. Find the Status of Part Number “A020”.

40

2.

Viva

05

3.

Journal

05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Create CSV file from given data. Read the data from CSV files into a data frame. Perform data pre-processing tasks such as handling missing values and outliers. [20]</p> <table><tr><th>Rollno</th><th>Name</th><th>Age</th><th>Marks</th><th>Class</th></tr><tr><td>1</td><td>Sudin</td><td>44</td><td>47</td><td>FY</td></tr><tr><td>2</td><td>Shaima</td><td>46</td><td>86</td><td>SY</td></tr><tr><td>3</td><td>Raina</td><td>27</td><td>45</td><td>TY</td></tr><tr><td>4</td><td>Paul</td><td>38</td><td></td><td>SY</td></tr><tr><td>5</td><td>Rahul</td><td>46</td><td>45</td><td>FY</td></tr><tr><td>6</td><td>Gopal</td><td></td><td>67</td><td>TY</td></tr><tr><td>7</td><td>Yatin</td><td>59</td><td>45</td><td>FY</td></tr><tr><td>8</td><td>Jim</td><td>36</td><td>34</td><td>FY</td></tr><tr><td>9</td><td>Nima</td><td>45</td><td>32</td><td>TY</td></tr></table> <p>B. Implement Linear Regression on the Iris dataset using Python/R for predicting petal.width on petal.length. [20]</p>	Rollno	Name	Age	Marks	Class	1	Sudin	44	47	FY	2	Shaima	46	86	SY	3	Raina	27	45	TY	4	Paul	38		SY	5	Rahul	46	45	FY	6	Gopal		67	TY	7	Yatin	59	45	FY	8	Jim	36	34	FY	9	Nima	45	32	TY	40
Rollno	Name	Age	Marks	Class																																																
1	Sudin	44	47	FY																																																
2	Shaima	46	86	SY																																																
3	Raina	27	45	TY																																																
4	Paul	38		SY																																																
5	Rahul	46	45	FY																																																
6	Gopal		67	TY																																																
7	Yatin	59	45	FY																																																
8	Jim	36	34	FY																																																
9	Nima	45	32	TY																																																
2.	Viva	05																																																		
3.	Journal	05																																																		

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs

Marks: 50

Date: 15 April 2025

Time: 9:00 - 11:00

Course: USCSP601 Data Science – Practical

Candidate's University Seat Number: _____

1.	<p>A. Perform Logistic Regression on the Iris dataset using Python/R to predict binary outcome. Evaluate model's performance using classification metrics. [20]</p> <p>B. Consider the dataset given below.</p> <table><tr><th>Date</th><th>Color</th><th>Region</th><th>Units</th><th>Sales</th></tr><tr><td>03-Jan-20</td><td>Red</td><td>West</td><td>1</td><td>110000</td></tr><tr><td>14-Jan-20</td><td>Blue</td><td>South</td><td>8</td><td>96000</td></tr><tr><td>21-Jan-20</td><td>Green</td><td>West</td><td>2</td><td>26000</td></tr><tr><td>30-Jan-20</td><td>Blue</td><td>North</td><td>7</td><td>84000</td></tr><tr><td>07-Feb-20</td><td>Green</td><td>North</td><td>8</td><td>25000</td></tr><tr><td>13-Feb-20</td><td>Red</td><td>South</td><td>2</td><td>60000</td></tr><tr><td>22-Feb-20</td><td>Blue</td><td>East</td><td>5</td><td>35000</td></tr><tr><td>01-Mar-20</td><td>Green</td><td>West</td><td>2</td><td>87000</td></tr><tr><td>13-Mar-20</td><td>Blue</td><td>East</td><td>8</td><td>69000</td></tr><tr><td>23-Mar-20</td><td>Blue</td><td>North</td><td>7</td><td>54000</td></tr></table> <p>Create Pivot Table in Excel for following analysis and visualize the data using PivotChart: [20]</p> <ul style="list-style-type: none">● Find out the total Sales.● Find out the Sum of Sales color-wise.● Find out the Sum of Units. <p>Find out Region-wise total sales and total units.</p>	Date	Color	Region	Units	Sales	03-Jan-20	Red	West	1	110000	14-Jan-20	Blue	South	8	96000	21-Jan-20	Green	West	2	26000	30-Jan-20	Blue	North	7	84000	07-Feb-20	Green	North	8	25000	13-Feb-20	Red	South	2	60000	22-Feb-20	Blue	East	5	35000	01-Mar-20	Green	West	2	87000	13-Mar-20	Blue	East	8	69000	23-Mar-20	Blue	North	7	54000	40
Date	Color	Region	Units	Sales																																																					
03-Jan-20	Red	West	1	110000																																																					
14-Jan-20	Blue	South	8	96000																																																					
21-Jan-20	Green	West	2	26000																																																					
30-Jan-20	Blue	North	7	84000																																																					
07-Feb-20	Green	North	8	25000																																																					
13-Feb-20	Red	South	2	60000																																																					
22-Feb-20	Blue	East	5	35000																																																					
01-Mar-20	Green	West	2	87000																																																					
13-Mar-20	Blue	East	8	69000																																																					
23-Mar-20	Blue	North	7	54000																																																					
2.	Viva	05																																																							
3.	Journal	05																																																							