Course Name- SQL and Data Visualization

Course Code- INT 350

Continuous Assessment-II

Important Guidelines:

1. All questions in this Academic Task are compulsory.

2. It is mandatory to attempt all questions of the assignment in your own handwriting on A4 size sheets/pages with a blue color ink pen. Any other mode of attempt (typed or printed codes or table) except handwritten/drawn will not be accepted/considered as valid submission(s) under any circumstances.

3. Every attempted sheet/page should carry clear details of student such as Name, Registration number, Roll number, Question number and Page number. The page numbers should be written clearly on the bottom of every attempted sheet in a prescribed format as: for page 1; Page 1 of 4, for page 2; Page 2 of 4, for page 3; Page 3 of 4 and for page 4; Page 4 of 4, in case your assignment/document is of 4 pages.

4. After attempting the answer(s) single pdf format document (can be done with many free online available converters).

5. This PDF file should be uploaded onto the UMS interface on or before the last date of the submission.

6. Refrain from indulging into plagiarism as copy cases will be marked zero.

7. This Document contains multiple sets of papers. The allocation sheet is also attached in the CA file. All the students are advised to attempt the Set allocated to him/her.

8. If any student found indulge in malpractices like plagiarism from internet or classmates, attempting wrong set of question paper or any other, will be awarded with zero (0) marks in CA.

SQL AND DATA VISUALISATION (INT-351) CA-2
Set-1

1.List the Pros and Cons of Joins and Subquery?

[5 Marks]

2.Create the table given below

STUDENTNAME	CLASSNAME	STUDENTMARKS
Anni	Class 7	92
Kaley	Class 9	80
Tom	Class 10	85
Johny	Class 4	76
Jimmy	Class 3	95

Rank the StudentName, based on StudentMarks as the rank would be stored in a new column StudentRank.

[5 Marks]

3. Using your own dataset perform the below followings using python

a.Histogram
b.Barchart
c.Heatmap

[5 Marks]

4.Using your own dataset perform data cleaning and data handling?

[5 Marks]

5.Perform boxplot for detecting outlier in python using your own dataset?

[5 Marks]

6.Explain the necessity of data visualization?

[5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2

Set-2

1.Create the table given below

STUDENTNAME	CLASSNAME	STUDENTMARKS
AnnI	Class 7	92
Kaley	Class 9	80
Tom	Class 10	85
Johny	Class 4	76
Jimmy	Class 3	95
Peter	Class 3	90
Bob	Class 9	80
Kat	Class 3	93

In the table above, apply the RANK() function and add 3 more students, Peter, Bob and Kim.

[5 Marks]

2.Perform scatterplot and pairplot using your own dataset in python?

[5 Marks]

3.Distinguish clustering and Non Non clustering indexing?

[5 Marks]

4.Perform Bar graph and line chart using your own dataset in python for data visualization.

[5 Marks]

5.Analyse the outliers in your dataset using Boxplot.

[5 Marks]

6.Perform bar chart and stacked bar chart using your own dataset.

[5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2
SET - 3

1. Define Rank function with the SQL queries.

   Create a table RANK (first_name, last_name, city) with the values

first_name	last_name	city
Luisa	Evans	Texas
Paul	Ward	Alaska
Peter	Bennett	California
Carlos	Patterson	New York
Rose	Huges	Florida
Marielia	Simmons	Texas
Antonio	Butler	New York
Diego	Cox	California

   a) Using Rank function arrange each rows in descending order
   b) Assign the Rank_No for each rows in the table
   c) Order the columns by first name and partition by using the city column
   d) Update the first name as Diego where the city is California [5 marks]

2. Define the LAG and LEAD functions with the SQL queries and give example.

   [5 marks]

3. Explain in detail with the SQL queries for the following:
   a) Partitioning
   b) Frames [5 marks]

4. With the user defined dataset perform the outlier analysis using boxplot and give the interpretation of the graph. [5 marks]

5. By using the user defined dataset, perform the following visualization:
   a) Bar chart
   b) Scatter plot
   c) Pair plot
   d) Line chart [5 marks]

6. Explain in detail about the bar and stacked bar charts with the example.

   [5 marks]

SQL AND DATA VISUALISATION (INT-351) CA-2
SET - 4

1. Using the Rank Function

   Create a ExamResultTable and insert the following values

	StudentName	Subject	Marks
1	Lily	Maths	65
2	Lily	Science	80
3	Lily	english	70
4	Isabella	Maths	50
5	Isabella	Science	70
6	Isabella	english	90
7	Olivia	Maths	55
8	Olivia	Science	60
9	Olivia	english	89

   a) Using Rank Function arrange each rows in descending order
   b) Assign the Rank_No for each rows in the table
   c) Order the columns by Student name and partition by using the Subject column
   d) Update the Student name as Isabella where the Marks are 65, 80 and 70.

[5 marks]

2. Perform the following:
   a) Define clustering and non-clustering index.
   b) Create a Student_info and insert 5 rows into the table with the column names of ROLL_NO, NAME and DEPARTMENT.
   c) Perform Clustering and Non-Clustering index Function in the above table.

[5 marks]

3. Perform the joining and nested function for the following table

   Create a table Student and Subject and insert the 5 values

Roll no.	Name	Class	Subjects
4	Aman	10	Maths
7	Raghav	11	Science
8	Sameer	12	Biology

a. Join the two tables using INNER JOINS

ID	Name	Age	Address	Salary
1	Arun	34	Kanpur	30,000
2	Kamal	23	Lucknow	34,000
3	Ajay	32	Mumbai	25,000
4	Shubham	28	Delhi	26,000
5	Anurag	26	Bangalore	24,000
6	Shivam	27	Hyderabad	23,000
7	karan	24	Noida	32,000
8	Himanshu	33	Chennai	20,000

Create a above table as Employee and perform the following

a) Select the Employee details whose salary is greater than 24,000

b) Select the names of the Employee whose age is less than32.

[5 marks]

4. Choose the dataset of your choice and perform the following visualization

    a) Box plot

    b) Scatter plot

    c) Heat map

    d) Line plot

[5 marks]

5. Using Pair plot and heat maps perform the multivariate and bivariate analysis respectively.

[5 marks]

6. Describe in detail

    a) Distribution plot

    b) Differentiate Bar and Stacked bar

[5 marks]

SQL AND DATA VISUALISATION(INT-351) CA-2

Set-5

1. Suppose you're given the following tables called 'orders' and 'order_info'. The table 'orders' shows revenue values for unique orders along with the associated channel ('online' or 'in_store') while the table 'order_info' shows the order's ID along with its location.

Table: Orders

order_id	channel	date	month	revenue
1	online	01-09-2020 00:00	9	100
2	online	03-09-2020 00:00	9	125
3	in_store	11-10-2020 00:00	10	208
4	in_store	21-08-2020 00:00	8	80
5	online	13-08-2020 00:00	8	200
6	online	16-08-2020 00:00	8	210
7	in_store	16-08-2020 00:00	8	205
8	online	11-10-2020 00:00	10	215
9	online	16-08-2020 00:00	8	203
10	in_store	01-09-2020 00:00	9	400
11	online	01-08-2020 00:00	8	107

Table: Order_info

order_id	location
1	NYC
2	NYC
3	LAX
4	LAX
5	SEA
6	AUS
7	LON
8	LAX
9	BLD
10	SEA
11	AUS

Using these tables, write a SQL query to return the top 3 'online' orders and their associated locations based on revenue generated. [5 marks]

2. Consider the following table, annual_sale, shown below:

year	total_sale
2015	23000
2016	25000
2017	34000
2018	32000
2019	33000

Use lag() and lead() function to compare annual sale amounts across years. [5 Marks]

3. What is the difference between Stored Procedures and UDFs. [5 Marks]

4. Take any dataset of your choice and perform outlier analysis using boxplots. Write the interpretation of the graph. [5 Marks]

5. Using any dataset of your choice perform bivariate analysis and interpret each graph –
   a) Line Charts
   b) Bar Graph
   c) Box plots

[5 Marks]

6. Briefly discuss about Stacked Bar Graphs with an example [5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2

Set-6

1. SQL RANK () function illustration
    a) First, create a new table named rank_demo that has one column
    b) Insert some rows into the rank_demo table
    c) Query data from the rank_demo table
    d) use the row_number() to assign ranks to the rows in the result set of rank_demo table

    [5 Marks]
2. Consider the below given table "contest" find number of days a contest will collapse with the next contest i.e. no. of days on which both contests are held

c_id	*start_date*	*end_date*
1	01-02-2015	04-02-2015
2	02-02-2015	05-02-2015
3	03-02-2015	07-02-2015
4	04-02-2015	06-02-2015
5	06-02-2015	09-02-2015
6	08-02-2015	10-02-2015
7	10-02-2015	11-02-2015

    [5 Marks]
3. What are case statements? Explain it briefly with an example [5 Marks]
4. Take any dataset of your choice and perform univariate analysis. Write the interpretation for all the graphs mentioned below –
    a) Scatter Plots
    b) Count Plot
    c) Distribution Plot [5 Marks]
5. Perform Bivariate and Multivariate Analysis using Heatmaps and Pair plots respectively.

    [5 Marks]
6. Briefly discuss about stacked bars with an example. Explain the interpretation of the graphs.

    [5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2
SET – 7

CREATE A SEPERATE DATABASE, CREATE TABLES AND INPUT THE VALUES (SHOW OUPUT FOR EACH TABLE).

Product Table:

product_id	product_name	unit_price
1	S8	1000
2	G4	800
3	iPhone	1400

Sales Table

seller_id	product_id	buyer_id	sale_date	quantity	price
1	1	1	2019-01-21	2	2000
1	2	2	2019-02-17	1	800
2	2	3	2019-06-02	1	800
3	3	4	2019-05-13	2	2800

1.Write an SQL query that reports the best seller by total sales price, If there is a tie, report them all.

Result Table

seller_id
1
3

[5 Marks]

2.Write an SQL query that reports the buyers who have bought S8 but not iPhone. Note that S8 and iP hone are products present in the Product table.

Expected Output

buyer_id
1

[5 Marks]

3. Write an SQL query that reports the products that were only sold in spring 2019. That is, between 2019-01-01 and 2019-03-31 inclusive.

Expected Output

```
+-------------+--------------+
| product_id  | product_name |
+-------------+--------------+
| 1           | S8           |
+-------------+--------------+
```

Creating Dataframe: tablefortest

srno	col_val
1	56
2	74
3	15
4	51
5	9
6	32

[5 Marks]

4. From the following dataframe, write a SQL query to find the even or odd values. Return "Even" for even number and "Odd" for odd number.

Expected Output:

srno	col_val	Even_Odd
1	56	Even
2	74	Even
3	15	Odd
4	51	Odd
5	9	Odd
6	32	Even

Table 1: sales

sale_id	product_id	year	quantity	price
1	100	2008	10	5000
2	100	2009	12	5000
7	200	2011	15	9000

Table 2: products

product_id	product_name
100	Nokia
200	Apple
300	Samsung

.

[5 Marks]

5. Write an SQL query that returns all product names of the products in the Sales table along with their selling year and price.

Expected Output:

product_name	year	price
Nokia	2008	5000
Nokia	2009	5000
Apple	2011	9000

[5 Marks]

6. Write an SQL query that returns the total quantity sold for every product id.

Expected Output:

product_id	total_quantity
100	22
200	15

[5 Marks]

SQL AND DATA VISUALISATION(INT-351) CA-2
SET – 8

CREATE A SEPERATE DATABASE, CREATE TABLES AND INPUT THE VALUES (SHOW OUPUT FOR EACH TABLE)

Create Dataframe 1:item

item_code	item_desc	cost
101	mother board	2700
102	RAM	800
103	key board	300
104	mouse	300

Create Dataframe 2: sales_info

distributor_id	item_code	retailer_id	date_of_sell	quantity	total_cost
5001	101	1001	2020-02-12	3	8100
5001	103	1002	2020-03-15	15	4500
5002	101	1001	2019-06-24	2	5400
5001	104	1003	2019-09-11	8	2400
5003	101	1003	2020-10-21	5	13500
5003	104	1002	2020-12-27	10	3000
5002	102	1001	2019-05-18	12	9600
5002	103	1004	2020-06-17	8	2400
5003	103	1001	2020-04-12	3	900

1. From the above dataframes, write a SQL query to find those retailers who have bought 'key board' but not 'mouse'. Return retailer ID.

Expected Output:

```
retailer_id|
-----------|
        1001|
        1004|
```

[5 Marks]

2. From the following dataframe, write a SQL query to display those items that were only sold in the 2nd quarter of a year, i.e. April 1st to June end for the year 2020. Return item code and item description.

Expected Output:

item_code	item_desc
103	key board

[5 Marks]

3.

```
mysql('select * from purchase')
```

	customer_id	item_code	purch_qty
0	101	504	25
1	101	503	50
2	102	502	40
3	102	503	25
4	102	501	45
5	103	505	30
6	103	503	25
7	104	505	40
8	101	502	25
9	102	504	40
10	102	505	50
11	103	502	25
12	104	504	40
13	103	501	35

.

From the following dataframe, write a SQL query to find the highest purchase with its corresponding item for each customer. In case of a same quantity purchase find the item code which is smallest.

The output must be sorted by increasing of customer_id. Return customer ID,lowest item code and purchase quantity.

Expected Output:

```
customer_id|lowest item code|purch_qty|
-----------|----------------|---------|
        101|             503|       50|
        102|             505|       50|
        103|             501|       35|
        104|             504|       40|
```

[5 Marks]

4.

Create dataframe 1:managing_body

manager_id	manager_name	running_years
51	James	5
52	Cork	3
53	Paul	4
54	Adam	3
55	Hense	4
56	Peter	2

Create dataframe 2:scheme

scheme_code	scheme_manager_id
1001	51
1001	53
1001	54
1001	56
1002	51
1002	55
1003	51
1004	52

From the following dataframes, write a SQL query to display those managers who have average experience for each scheme.

Expected Output:

```
scheme_code|Average year of experience|
-----------|--------------------------|
       1001|                      3.50|
       1002|                      4.50|
       1003|                      5.00|
       1004|                      3.00|
```

[5 Marks]

5. From the above dataframes, write a SQL query to find those schemes which executed by minimum number of employees. Return scheme code.

Expected Output:

scheme_code
1003
1004

[5 Marks]

6. From the following dataframes, write a SQL query to find those experienced manager who execute the schemes. Return scheme code and scheme manager ID.

Expected Output:

scheme_code	scheme_manager_id
1001	51
1002	51
1003	51
1004	52

[5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2

Set-9

1. Suppose you're given the following tables called 'orders' and 'order_info'. The table 'orders' shows revenue values for unique orders along with the associated channel ('online' or 'in_store') while the table 'order_info' shows the order's ID along with its location.

Table: Orders Table: Order_info

Using these tables, write a SQL query to return the top 3 'online' orders and their associated locations based on revenue generated. [5 marks]

2. Consider the following table, annual_sale, shown below:

year	total_sale
2015	23000
2016	25000
2017	34000
2018	32000
2019	33000

Use lag() and lead() function to compare annual sale amounts across years. [5 Marks]

3. What is the difference between Stored Procedures and UDFs. [5 Marks]

4. Take any dataset of your choice and perform outlier analysis using boxplots. Write the interpretation of the graph. [5 Marks]

5. Using any dataset of your choice perform bivariate analysis and interpret each graph –
   a) Line Charts
   b) Bar Graph
   c) Box plots

[5 Marks]

6. Briefly discuss about Stacked Bar Graphs with an example [5 Marks]

7.

SQL AND DATA VISUALISATION (INT-351) CA-2
Set-10

1.Given the 'orders' and 'order_info' tables, write a SQL query to find the total revenue generated from 'in_store' orders. [5 marks]

2.Consider a table 'employee_salaries' with the following structure:

Table: employee_salaries

emp_id	emp_name	salary
1	Alice	60000
2	Bob	55000
3	Carol	62000
4	David	58000

Write a SQL query to calculate the average salary of all employees.

.Write a SQL Query to rank the employees as per salary [5 Marks]

3.Compare the Joins with Nested Query [5 Marks]
4. Explain the concept of data visualization and its role in data analysis. Provide an example of a situation where effective data visualization can lead to better insights. [5 Marks]

5.Using a dataset of your choice, create a scatter plot to visualize the relationship between two variables. Interpret the scatter plot and discuss any correlation you observe. [5 Marks]

6.Describe the key differences between a bar chart and a histogram in data visualization. Provide scenarios where each type of chart is more appropriate. [5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2
Set-11

1.Given the 'orders' and 'order_info' tables, write a SQL query to find the order with the highest revenue and its associated location, regardless of the channel. [5 marks]

2.Consider a table 'customer_feedback' with the following structure:

Table: customer_feedback

feedback_id	customer_id	rating
1	101	4
2	102	5

3 103 3
4 104 4

3.Write a SQL query to calculate the average rating given by customers
Also split the customer into 4 quartiles according to the rating . [5 Marks]

4.Discuss the concept of data preprocessing in the context of machine learning. Why is it important, and what are some common preprocessing steps? [5 Marks]

5.Use Python's Seaborn library to create a heatmap to visualize the correlation matrix of variables in a dataset. Interpret the heatmap and explain the relationships between variables. [5 Marks]

6.Compare and contrast a line chart and a scatter plot in data visualization. Provide examples of scenarios where each type of chart is more suitable. [5 Marks]

SQL AND DATA VISUALISATION (INT-351) CA-2
Set-12

1.Given the 'orders' and 'order_info' tables, write a SQL query to find the total revenue generated from 'online' orders. [5 marks]

2.Consider a table 'product_inventory' with the following structure:

Table: product_inventory

product_id	product_name	quantity
1	Laptop	20
2	Smartphone	50
3	Tablet	30
4	Printer	10

Write a SQL query to calculate the total quantity of products in the inventory.
Write a SQL Query to calculate to rank the product according to quantity in such a way that it shouldnt skip the rank [5 Marks]

3.Comment on the Order of Execution of Query v/s Syntax of Query [5 Marks]
4.Explain the concept of data normalization and denormalization in the context of databases. Provide examples of situations where each approach is appropriate. [5 Marks]

5.Using a dataset of your choice, create a bar chart to compare the sales performance of different product categories. Interpret the chart and discuss any trends or insights. [5 Marks]

6.Define and describe the use cases for box plots in data visualization. Provide an example dataset where a box plot would be particularly informative. [5 Marks]

Student List with Assigned Sets

Sr. No	Registration Number	Name of the Student	Roll Number	Set Allocation
1	12113501	Shubham Kumar	RK21UTA01	SET-1
2	12112282	Palli Sai Kiran	RK21UTA02	SET-2
3	12112093	Khurram Shahin	RK21UTA03	SET-3
4	12111724	Shahriar Mumin Khan	RK21UTA04	SET-4
5	12113102	Annamdevula Ravi	RK21UTA05	SET-5
6	12113229	Gummudu Kishore Kumar	RK21UTA06	SET-6
7	12109994	Priyanshu Singh	RK21UTA07	SET-7
8	12110145	Prathipati Venkatesh	RK21UTA08	SET-8
9	12110626	Marlakunta Kedhareswer Naidu	RK21UTA09	SET-9
10	12111396	Darsi Venkat Charan	RK21UTA10	SET-10
11	12100915	Nived Suresan A	RK21UTA11	SET-11
12	12100863	C S Charithartha Sai	RK21UTA12	SET-12
13	12109514	Nikhil Singh	RK21UTA13	SET-1
14	12109665	T Tanusree	RK21UTA14	SET-2
15	12109211	Karri John Pradeep Reddy	RK21UTA15	SET-3
16	12108024	Anushka Kashyap	RK21UTA16	SET-4
17	12108472	Gopidesi Vinod Kumar	RK21UTA17	SET-5
18	12108725	Dharani K S	RK21UTA18	SET-6
19	12106386	Pentyala Kumar Govindu	RK21UTA19	SET-7
20	12106729	Kriti Mishra	RK21UTA20	SET-8
21	12106692	Garvit Joshi	RK21UTA21	SET-9
22	12107057	Yaswanth Subrahmanyam Jonnadula	RK21UTA22	SET-10
23	12107367	Shivansh Ranjan	RK21UTA23	SET-11
24	12107544	Shaik Latheef	RK21UTA24	SET-12
25	12107776	Lakshya Sharma	RK21UTA25	SET-1
26	12107627	Medam Sai Shashank	RK21UTA26	SET-2
27	12104754	Achanagari Hanu Tejesh	RK21UTA27	SET-3
28	12104652	Alexander Peter Maliyakkal	RK21UTA28	SET-4
29	12106234	Vulli B M S Pruthvi	RK21UTA29	SET-5
30	12105798	Utkrist Ark	RK21UTA30	SET-6

31	12103929	Velagalapalli Sai Kishore Chandra	RK21UTA31	SET-7
32	12115897	Kunal Yadav	RK21UTA32	SET-8
33	12115161	Mahrishi Rathore	RK21UTA33	SET-9
34	12115398	Rohan Patel	RK21UTA34	SET-10
35	12116486	Madhan Sai Thupakula	RK21UTA35	SET-11
36	12102845	Ankur Banerjee	RK21UTB36	SET-12
37	12102585	Nikhil Pathak	RK21UTB37	SET-1
38	12102610	S Surjith Subash	RK21UTB38	SET-2
39	12101918	Indukuri Satya Sudheer Varma	RK21UTB39	SET-3
40	12101692	Gurram Karthik	RK21UTB40	SET-4
41	12104702	K Somanath Sai Teja Srinivas	RK21UTB41	SET-5
42	12104879	Jarugu Mukesh Sai	RK21UTB42	SET-6
43	12107747	Mahamad Suhail	RK21UTB43	SET-7
44	12107884	Vaspari Murari	RK21UTB44	SET-8
45	12107890	Sanjana Umrao	RK21UTB45	SET-9
46	12107896	Prabhu Varun Puppala	RK21UTB46	SET-10
47	12107901	Madireddy Bharath Kumar Reddy	RK21UTB47	SET-11
48	12107624	Kanigelupula Surya Venkata Phanindra	RK21UTB48	SET-12
49	12107183	Rahul Rajput	RK21UTB49	SET-1
50	12108436	Saksham Parasher	RK21UTB50	SET-2
51	12108310	Mohammed Aasif	RK21UTB51	SET-3
52	12107941	Peyyala Akshay Mathew	RK21UTB52	SET-4
53	12109517	Adigopula Varun Kumar	RK21UTB53	SET-5
54	12109549	Pallanti Asrith Vatsal	RK21UTB54	SET-6
55	12100859	Abhinav Kumar	RK21UTB55	SET-7
56	12100568	Mandeep Singh Gill	RK21UTB56	SET-8
57	12100583	Sunkari Vedavyas	RK21UTB57	SET-9
58	12100403	Poothi Chandrasekhar Reddy	RK21UTB58	SET-10
59	12110965	Anindita Pandit	RK21UTB59	SET-11
60	12110943	Shristi Sehwag	RK21UTB60	SET-12
61	12113036	Siddharth Prahasith Bathula	RK21UTB61	SET-1

62	12112410	Nikhil Kaundal	RK21UTB62	SET-2
63	12111711	Kunal Kumar Pandit	RK21UTB63	SET-3
64	12111702	manish choudhury	RK21UTB64	SET-4
65	12112264	Bevara Hemanth Kumar	RK21UTB65	SET-5
66	12113773	Vidhya Bhusan Rath	RK21UTB66	SET-6
67	12115210	Rohan Stanislaus R	RK21UTB67	SET-7
68	12115853	Syed Faiq Husain	RK21UTB68	SET-8
69	12114879	Debasish Chandra Dey	RK21UTB69	SET-9
70	12114325	Aman Verma	RK21UTB70	SET-10