knowledgehut
upGrad

Advanced Data Analysis in SQL – Part II

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of this Session, You Will:

- Use Rank functions to assign sequential numbering to rows in a result set.
- Capture trends in a time series using moving averages.
- Perform advance pattern matching using regular expressions.
- Improve the efficiency and performance of queries.
- Learn the importance of temporary tables and views.
- Understand the different types of Indices in MySQL and their importance.

What Have We Learned So Far?

- Miscellaneous functions for manipulating text and date columns.
- Primary Keys are used to uniquely identify each row in a table.
- Foreign Keys are used to establish a relationship between two tables.
- Entity Relationship Diagrams are used to create a blueprint of an entire database.
- JOINS in MySQL is used to analyze data from multiple tables simultaneously.
- UNION clause is used to concatenate multiple result-sets of SELECT query.
- CASE WHEN statements allow us to insert conditional logic within a query.
- Window functions are used to perform calculations across a set of rows within a defined window.

Pop Quiz

Q. Which of the following statements demonstrates the correct usage of the PARTITION BY clause in a window function?

a. **SUM(sales) OVER (PARTITION BY product_id)**

b. AVG(quantity) PARTITION BY category_name OVER ()

c. COUNT(*) OVER (PARTITION BY order_date DESC)

d. length(order_id) OVER (PARTITION BY customer_id ORDER BY purchase_date)

Pop Quiz

Q. Which of the following statements demonstrates the correct usage of the PARTITION BY clause in a window function?

a. SUM(sales) OVER (PARTITION BY product_id)

b. AVG(quantity) PARTITION BY category_name OVER ()

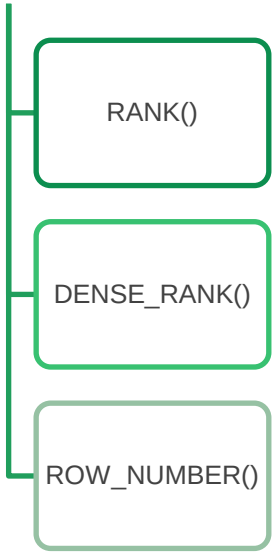
c. COUNT(*) OVER (PARTITION BY order_date DESC)

d. length(order_id) OVER (PARTITION BY customer_id ORDER BY purchase_date)

RANK Function in SQL

Purpose and Syntax of RANK Function

The purpose of rank functions in MySQL is to assign a rank or position to each row within a specified order or grouping. These functions allow you to determine the relative position of each row based on specific criteria.



Syntax of Rank Functions

```
rank_function() OVER (
                [PARTITION BY column1, column2, ...]
                ORDER BY column_name [ASC|DESC]
)
```

Demo – RANK Function in SQL

Poll Time

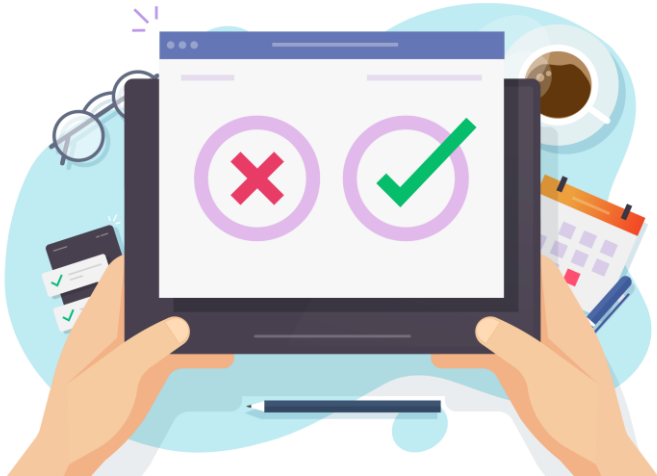
Q. Which of the following rank functions in MySQL handles ties by assigning the same rank to rows with the same values?

a. ROW_NUMBER()

b. DENSE_RANK()

c. COUNT()

d. RANK()



Poll Time

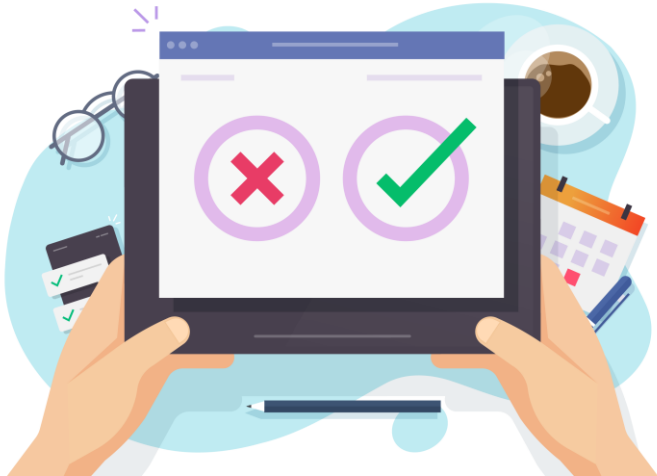
Q. Which of the following rank functions in MySQL handles ties by assigning the same rank to rows with the same values?

a. ROW_NUMBER()

b. DENSE_RANK()

c. COUNT()

d. NTILE()



Moving Average in MySQL

Purpose and Syntax of Moving Average

The purpose of calculating a moving average is to smooth out fluctuations and identify trends or patterns within a dataset over time.

Syntax of Rank Functions

```
SELECT
    column_name,
    AVG(column_name) OVER (
        ORDER BY date_column
        ROWS BETWEEN <n> PRECEDING AND CURRENT ROW
    ) AS moving_average
FROM table_name;
```

Demo – Moving Average

Pop Quiz

Q. Which of the following statements about moving averages is true?

a. Moving averages are used to calculate the total sum of a dataset

b. Moving averages are only applicable to time-series data

c. Moving averages help in identifying short-term fluctuations in data

d. Moving averages smooth out data by calculating the average of a subset of adjacent values

Pop Quiz

Q. Which of the following statements about moving averages is true?

a. Moving averages are used to calculate the total sum of a dataset

b. Moving averages are only applicable to time-series data

c. Moving averages help in identifying short-term fluctuations in data

d. Moving averages smooth out data by calculating the average of a subset of adjacent values

Regular Expressions in SQL

Purpose and Syntax of Regular Expressions in SQL

The purpose of regular expressions (regex) is to search, match, and manipulate text patterns in a flexible and powerful way.

Syntax of REGEX

```
SELECT column_name
FROM table_name
WHERE column_name REGEXP 'pattern';
```

Demo – Regular Expressions in SQL

Poll Time

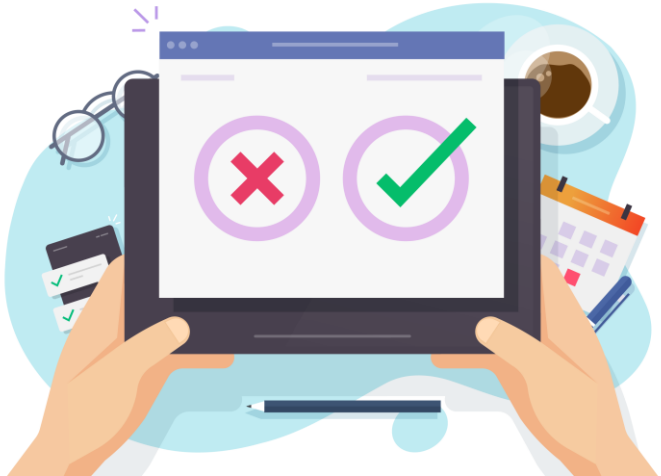
Q. Which of the following operators is used for regular expression matching in MySQL?

a. LIKE

b. =

c. REGEXP

d. BETWEEN



Poll Time

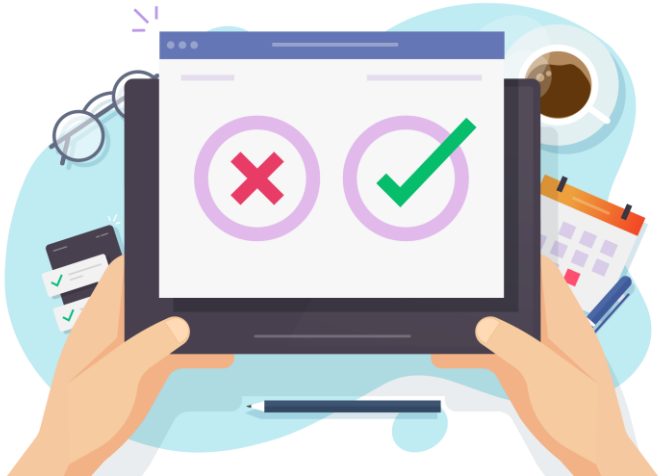
Q. Which of the following operators is used for regular expression matching in MySQL?

a. LIKE

b. =

c. REGEXP

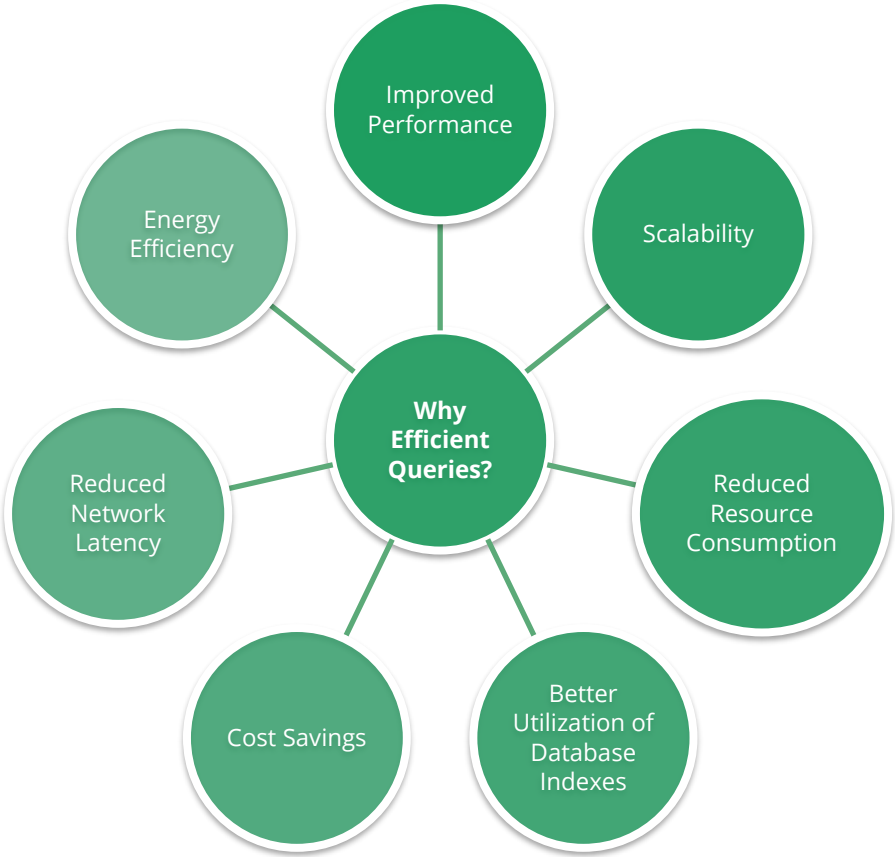
d. BETWEEN




Q&A

Making Efficient Analysis

Significance of Making Efficient Analysis



Introduction to Temporary Table in SQL

What is a Temporary Table?

In MySQL, a temporary table is a special type of table that is created and used for a specific session or connection. Temporary tables are stored in the temporary directory of the MySQL server and are automatically dropped or deleted when the session ends or when explicitly deleted by the user.

Demo – Temporary Table in SQL

Pop Quiz

Q. Which of the following statements about temporary tables in MySQL is true?

a. Temporary tables are stored in the same location as permanent tables

b. Temporary tables are visible and accessible to all database sessions

c. Temporary tables are automatically dropped at the end of the session that created them

d. Temporary tables participate in database backups and replication

Pop Quiz

Q. Which of the following statements about temporary tables in MySQL is true?

a. Temporary tables are stored in the same location as permanent tables

b. Temporary tables are visible and accessible to all database sessions

c. Temporary tables are automatically dropped at the end of the session that created them

d. Temporary tables participate in database backups and replication

Introduction to View in SQL

Purpose and Syntax of a View in SQL

In MySQL, a view is a virtual table that is derived from the result of a query. It does not store any data of its own but instead retrieves and displays data from the underlying tables based on the defined query.

Syntax of VIEWS

```
CREATE VIEW view_name AS
SELECT column1, column2, ...
FROM table_name
WHERE condition;
```

Demo – View in SQL

Poll Time

Q. Which of the following statements about views in MySQL is true?

a. Views store their own copy of data separate from the underlying tables

b. Views can be modified directly, allowing insert, update, and delete operations

c. Views provide an additional layer of security by allowing direct access to the underlying tables

d. Views are created using the CREATE TABLE statement in MySQL

Poll Time

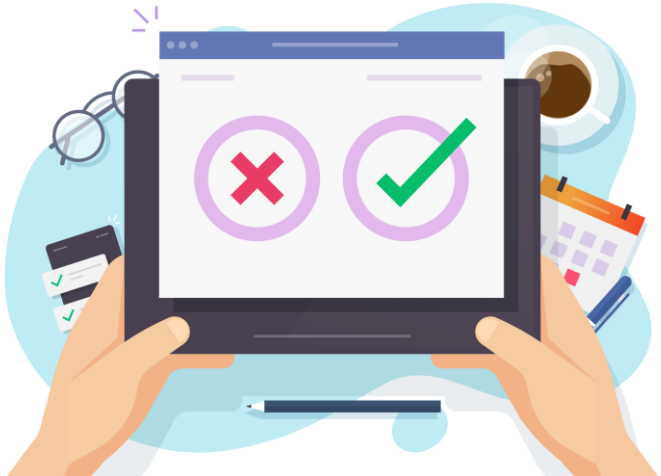
Q. Which of the following statements about views in MySQL is true?

a. Views store their own copy of data separate from the underlying tables

b. Views can be modified directly, allowing insert, update, and delete operations

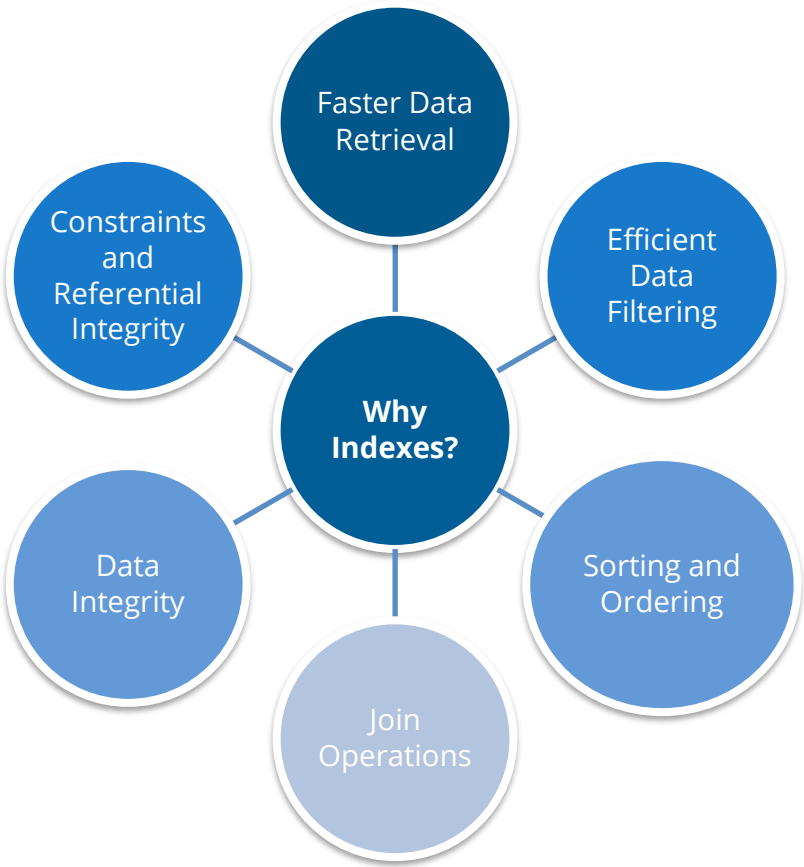
c. Views provide an additional layer of security by allowing direct access to the underlying tables

d. Views are created using the CREATE TABLE statement in MySQL



Introduction to SQL Indexes

Purpose of SQL Indexes



Types of Indexes

Primary Key Index

- Primary key indexes provide fast data retrieval for primary key lookups and enforce data integrity.

Unique Index

- A unique index ensures that the values in the indexed column(s) are unique, meaning no duplicate values are allowed.

Full-Text Index

- Full-Text indexes are used for efficient searching of text-based data, such as documents or textual content stored in MySQL.

Spatial Index

- Spatial indexes are used to optimize queries involving spatial data, such as geographical coordinates or spatial objects.

Regular Index

- For regular index, the specified column values do not require to be unique and can be NULL.

Demo - Indexes


Q&A

Summary

- Rank functions assign sequential numbering to rows in a result set.
- Moving averages are used to analyze the trend in a numerical column.
- Regular expressions are used to perform advance string manipulations using patterns.
- Temporary tables are used to store intermediate results for complex queries.
- Views are virtual tables that are created based on a SELECT statement.

Activity 1

Pre-requisites:

- MySQL Workbench
- Company_db Database

Scenario:

Perform the below operations on the tables present in the companydb database.

- Assign **rank** to each **employee** in their respective **department** based on their **total hours of work**.
- **Filter** all the **employees** who have worked more than the **average hours** spent by the employees of the respective **department**. Please note to solve the above problem by using **temporary tables**.
- Create an **Index** on the **employee** table using the **super_ssn** column.

Next Session:
SQL for Data Science - Case Study

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad

SQL for Data Science – Case Study

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of this Session, You Will:

- Solve a case study pertaining to agent commission analysis.
- Create a database to store the commission data.
- Run multiple SELECT queries to analyze the underlying data.
- Increase the efficiency of the queries by using indexes, views, and temporary tables.

Recap

Pop Quiz

Q. Which of the following statements best describes the difference between views and temporary tables in MySQL?

a. Temporary tables store their own copy of data, while views do not store data

b. Temporary tables are accessible to all database sessions, while views are session-specific

c. Temporary tables are used for complex queries, while views are used for simple queries

d. Temporary tables can be modified directly, while views are read-only by default

Pop Quiz

Q. Which of the following statements best describes the difference between views and temporary tables in MySQL?

a. **Temporary tables store their own copy of data, while views do not store data**

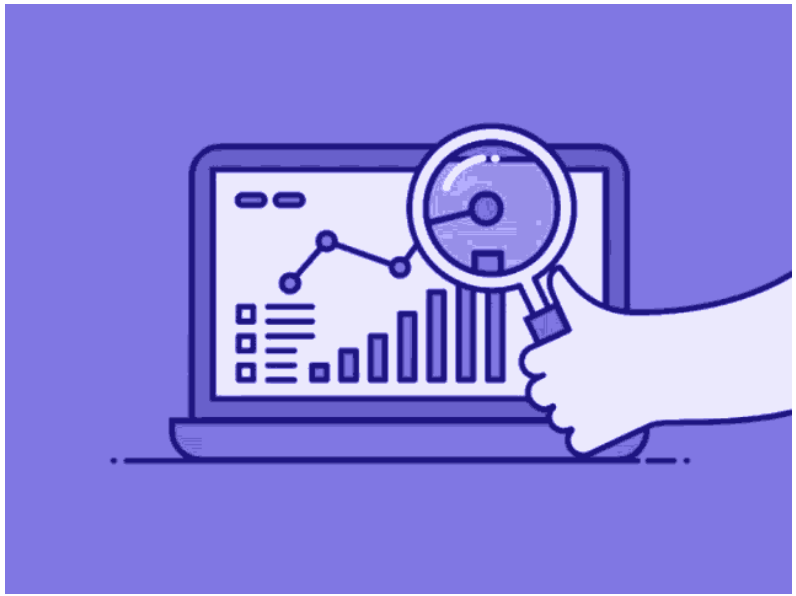
b. Temporary tables are accessible to all database sessions, while views are session-specific

c. Temporary tables are used for complex queries, while views are used for simple queries

d. Temporary tables can be modified directly, while views are read-only by default

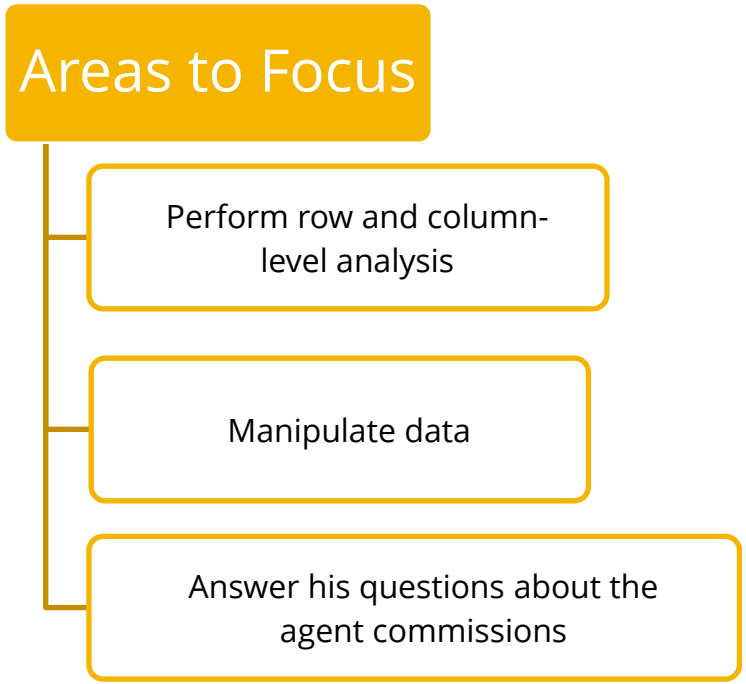
Case Study – Problem Statement

Problem Statement



Ankush is the head of the agent association. He wants to analyze agent commission data using SQL to get insights around top-performing agents and compare their performances.

Areas to Focus



Pop Quiz

Q. Which of the following is a primary purpose of MySQL in data analysis?

a. Storing and managing data

b. Generating reports and visualizations

c. Performing complex calculations and aggregations

d. Data visualization and exploration

Pop Quiz

Q. Which of the following is a primary purpose of MySQL in data analysis?

a. Storing and managing data

b. Generating reports and visualizations

c. Performing complex calculations and aggregations

d. Data visualization and exploration

Understanding the Data

Sneak Peak into the Data

Agent Table

agent_code	F_name	L_name	city	commission	mobile_no
1	Manthan	Koli	Delhi	12000	9898454545
2	Virat	Dixit	Jaipur	12150	9696123240
3	Akshay	Deshmukh	Mumbai	1500	8787987810
4	Sanket	Kumar	Chandigarh	120	8789878980
5	Gaurav	Desai	Chennai	13020	9698989560
6	Sai	Jain	Shimla	65650	7875747971

Customer Table

customer_id	c_name	c_city	agent_code	order_date
101	Nick Rimando	New York	2	2009-01-13
102	Brad Davis	New York	1	2010-03-05
103	Graham Zusi	California	5	2010-07-20
104	Julian Green	London	8	2010-04-10
105	Fabian Johnson	Paris	9	2010-08-28
106	Geoff Cameron	Berlin	5	2010-05-15

Pop Quiz

Q. Which of the following best describes a table schema in MySQL?

a. The physical structure of a table on disk

b. The set of columns and data types in a table

c. The primary key constraint applied to a table

d. The collection of rows and data stored in a table

Pop Quiz

Q. Which of the following best describes a table schema in MySQL?

a. The physical structure of a table on disk

b. The set of columns and data types in a table

c. The primary key constraint applied to a table

d. The collection of rows and data stored in a table


Q&A

Hands-on: Case Study Questions

Pop Quiz

Q. Which of the following statements best describes the usage of stored procedures in MySQL?

a. Stored procedures are used for defining complex data models and table relationships

b. Stored procedures provide a way to perform ad-hoc queries and retrieve data from the database

c. Stored procedures are primarily used for securing and encrypting sensitive data in the database

d. Stored procedures allow for encapsulating and executing a sequence of SQL statements as a single unit

Pop Quiz

Q. Which of the following statements best describes the usage of stored procedures in MySQL?

a. Stored procedures are used for defining complex data models and table relationships

b. Stored procedures provide a way to perform ad-hoc queries and retrieve data from the database

c. Stored procedures are primarily used for securing and encrypting sensitive data in the database

d. Stored procedures allow for encapsulating and executing a sequence of SQL statements as a single unit


Q&A

Summary

- **newdb** database contains the tables for performing commission analysis.
- MySQL can be used to answer a lot of critical question around the agent commission.
- You can modify the database objects and table contents.
- Using SELECT clause in conjunction with other important clauses, we can analyze the data from multiple angles.

Activity 1

Pre-requisites:

- MySQL Workbench
- Newdb Database

Scenario:

Perform the below operations on the tables present in the **newdb** database.

- Determine the **city** with the **most number** of **customers**.
- Find out which **agent** has served the **highest number** of **customers**.
- Calculate the **total number** of available **agents** in **Mumbai**.
- Calculate the **average commission** earned by **agents** in different **cities**.

Session Feedback

Next Session:
Introduction to Python for Data Science

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad