Module 1 - Task 1

- Preparing Employee Dataset
- Import Pandas and alias it as 'pd'.
- Read the CSV file movies Uncleaned_employees_final_dataset.csv into a Pandas DataFrame named 'df'.
- To import the 'Uncleaned_employees_final_dataset' file, which is located in the root path of your project, you should use the following path: './Uncleaned_employees_final_dataset'.
- Create a copy of the DataFrame.
- Remove duplicate rows using data.drop_duplicates().
- Eliminate rows with missing values using data.dropna().
- Save the cleaned data to a new CSV file named 'employee.csv' using to_csv with the index set to False and encoding specified as 'utf-8'.
- The DataFrame 'data' now reflects the changes and has been saved to the CSV file 'employee.csv'.

Module 2 - Task 1

- Data Download, Import, and Database Connection.
- Download the dataset employee.csv which is exported in Module 1 - Task 1.
- Create the table on MYSQL using your credentials provided here
- Use the provided login information to access the database by clicking the ""localhost"" link located on the Database Info tab. Once there, you need to upload the required datasets in the specific database mentioned in the database info tab. Rename the table to 'employee' using the Operations tab within the database interface and then click on ""Run test"" to complete the task.
- Use the %load_ext sql command to load the SQL extension in your Jupyter Notebook environment. This extension allows you to run SQL commands directly within your notebook.
- Use the %sql magic command to specify the connection string for your MySQL database. Replace <user>, <password>, and <db_name> with your actual database credentials and details.

Module 2 - Task 2

- Find the average age of employees in each department and gender group.
- Use the SQL SELECT statement to retrieve data from the database.
- The COUNT(*) function is used to count the number of rows that match the specified condition.
- The FROM clause specifies the database table, in this case, 'ecommerce'.
- The WHERE clause is used to filter the rows. It specifies that only rows where the 'reviews_text' column contains the term 'Alexa' should be counted. The % symbols before and after 'Alexa' serve as wildcard characters, allowing for matching any text that includes 'Alexa' within it.

Module 2 - Task 3

- List the top 3 departments with the highest average training scores.
- Employ the SQL SELECT statement.
- Calculate the average training scores by using the AVG() function.
- Round the average training scores to two decimal places with the ROUND() function.
- Group the data by department utilizing the GROUP BY clause.
- Arrange the results in descending order based on the average training scores.
- Display only the top 3 departments with the highest average training scores.

Module 2 - Task 4

- Find the percentage of employees who have won awards in each region
- Use the SQL SELECT statement.
- Calculate the percentage of employees who have won awards in each region.
- Count the number of employees who have won awards by filtering for awards_won = 1.
- Compute the percentage of employees who received awards by dividing the count of award-winning employees by the total count of employees with awards won.
- Utilize the ROUND() function to round the resulting percentages to two decimal places.
- Group the data by the region to display the award percentages per region.

Module 2 - Task 5

- Show the number of employees who have met more than 80% of KPIs for each recruitment channel and education level.
- Utilize the SQL SELECT statement.
- Calculate the count of employees who have met more than 80% of KPIs.
- Filter employees by the condition where KPIs_met_more_than_80 = 1.
- Group the data by both the recruitment channel and education level.
- Display the count of employees who have met more than 80% of KPIs for each recruitment channel and education level.

Module 2 - Task 6

- Find the average length of service for employees in each department, considering only employees with previous year ratings greater than or equal to 4
- Use the SQL SELECT statement.
- Calculate the average length of service for employees in each department.
- Consider only employees with a previous year rating greater than or equal to 4 .
- Group the data by department.
- Display the rounded average length of service for each department.

Module 2 - Task 7

- List the top 5 regions with the highest average previous year ratings.
- Utilize the SQL SELECT statement.
- Calculate the average previous year ratings for each region.
- Group the data by region.
- Arrange the regions in descending order based on the average previous year ratings.
- Display only the top 5 regions with the highest average previous year ratings.

Module 2 - Task 8

- List the departments with more than 100 employees having a length of service greater than 5 years.
- Utilize the SQL SELECT statement.
- Filter and count the number of employees with a length of service greater than 5 years.
- Group the employees by department.
- Apply the HAVING clause to select only departments where the employee count is more than 100.

Module 2 - Task 9

- Show the average length of service for employees who have attended more than 3 trainings, grouped by department and gender.
- Utilize the SQL SELECT statement.
- Calculate the average length of service for employees who attended more than 3 trainings.
- Group the employees by both department and gender.
- Round the average length of service to two decimal places.

Module 2 - Task 10

- Find the percentage of female employees who have won awards, per department. Also show the number of female employees who won awards and total female employees.
- Use the SQL SELECT statement to extract the necessary data.
- Calculate the percentage of female employees who have won awards in each department.
- Employ a CASE statement to identify female employees who have won awards (awards_won = 1) and count them for the percentage calculation.
- Calculate the count of female employees who won awards (Females_won_award) and the total count of female employees (Total_females) for each department.
- Utilize the gender condition to filter only female employees (gender = 'f').
- Group the data by department to display the results per department.

Module 2 - Task 11

- Calculate the percentage of employees per department who have a length of service between 5 and 10 years
- Utilize the SQL SELECT statement.
- Calculate the percentage of employees in each department with a length of service between 5 and 10 years.
- Use a CASE statement to identify employees with a length of service falling between 5 and 10 years.
- Count the number of employees falling within the specified length of service range for the percentage calculation.
- Calculate the total count of employees in each department to compute the percentage.
- Employ the GROUP BY clause to organize the results by department.

Module 2 - Task 12

- Find the top 3 regions with the highest number of employees who have met more than 80% of their KPIs and received at least one award, grouped by department and region.
- Utilize the SQL SELECT statement to extract the required data.
- Filter employees who have met more than 80% of their KPIs and received at least one award.
- Group the employees by department and region.
- Count the number of employees meeting the criteria for each department and region.
- Arrange the results in descending order based on the total count of employees.
- Limit the output to the top 3 regions with the highest count of qualified employees.

Module 2 - Task 13

- Calculate the average length of service for employees per education level and gender, considering only those employees who have completed more than 2 trainings and have an average training score greater than 75
- Use the SQL SELECT statement to extract the required data.
- Filter employees who have completed more than 2 trainings and have an average training score greater than 75.
- Group the employees by education level and gender.
- Calculate the average length of service for each education level and gender.
- Round the average length of service to two decimal places for better readability.

Module 2 - Task 14

- For each department and recruitment channel, find the total number of employees who have met more than 80% of their KPIs, have a previous_year_rating of 5, and have a length of service greater than 10 years.
- Utilize the SQL SELECT statement to retrieve the required data.
- Filter employees meeting specific criteria: those who have met more than 80% of their KPIs, have a previous_year_rating of 5, and have a length of service greater than 10 years.
- Group the employees by department and recruitment channel.
- Calculate the total number of employees satisfying the specified conditions for each department and recruitment channel.

Module 2 - Task 15

- Calculate the percentage of employees in each department who have received awards, have a previous_year_rating of 4 or 5, and an average training score above 70, grouped by department and gender
- Use the SQL SELECT statement to gather the necessary data.
- Filter employees based on certain criteria: those who have received awards, possess a previous_year_rating of 4 or 5, and have an average training score above 70.
- Group the employees by department and gender.
- Calculate the percentage of employees who meet the specified conditions for each department and gender.

Module 2 - Task 16

- List the top 5 recruitment channels with the highest average length of service for employees who have met more than 80% of their KPIs, have a previous_year_rating of 5, and an age between 25 and 45 years, grouped by department and recruitment channel.
- Utilize the SQL SELECT statement to acquire the necessary data.
- Filter employees who have met more than 80% of their KPIs, possess a previous_year_rating of 5, and fall within the age range of 25 to 45 years.
- Group the employees by department and recruitment channel.
- Calculate the average length of service for these employees.
- Sort the results in descending order based on the average service length.
- Limit the output to display the top 5 recruitment channels with the highest average service length.