

Project: HIVE

Domain: Healthcare

Context:

This data is about the cause of death from 2001 to 2008 in France, this will help the France government to identify the reasons which are the reasons for deaths across different genders and take necessary steps.

Data:

We have a predefined **dataset (CauseofDeath.csv)** having 4 columns. The dataset has different attributes like.

1. Time(year)
2. SEX (male or female)
- 3.Cause
- 4.Value (count of deaths)

Objective:

Show case your hive skills.

Learning Outcomes:

Once you successfully complete this exercise, you will be able to work on dynamic partitioned tables, windows functions in Hive, using SERDE for building the hive tables.

Grading Criteria: 50 marks

1. Create the table using the above data set, also create the partition based on the time? After creating the partitioned tables, you need to display the partitions. (15 marks)
2. Total data points collected for each year? (5 marks)
3. What are different types of cause of death? (5 marks)
4. Total number of deaths, find out the year in which the deaths are more? (12.5 marks)
5. Top 5 causes of death and Top 5 causes of death across different sex? (12.5 marks)