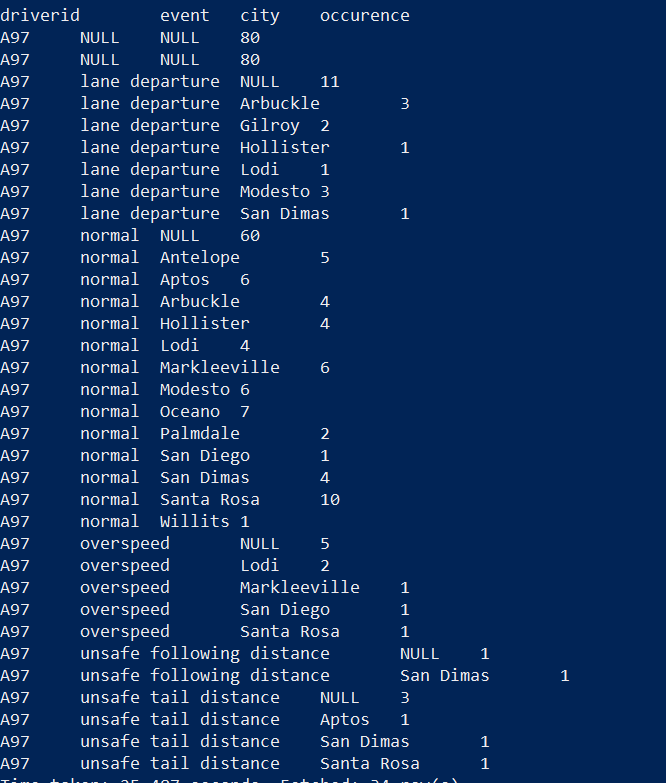
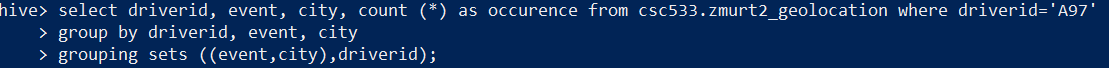
Used HiveQL in Apache Hive to analyze driver-accident risk factors using using Data Warehousing/OLAP functions present in Hive, enhanced aggregation with Grouping Sets, Cube and Rollup. Dimensional aggregation on geolocation data deduced metrics to reduce fuel costs and improve driver safety.

Q1) 1)

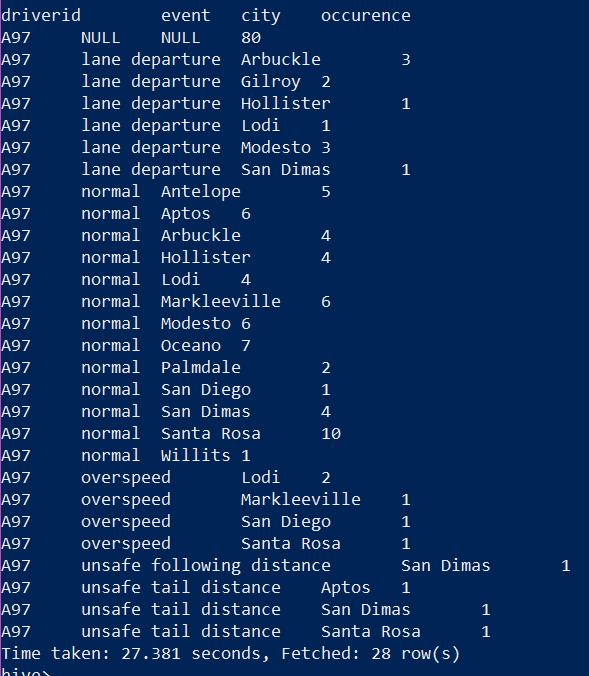


Output:



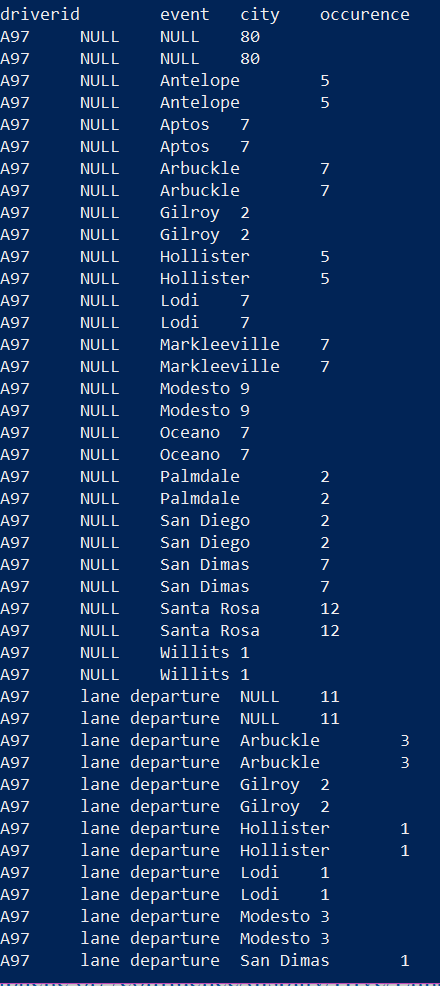
2) 

Same result as Rollup:



Q2) 1) 

Output:



Here, we see an aggregation of all possible combinations.

**Differences between Grouping Sets, Rollup and Cube:**

**Grouping Sets** clause in the Group By clause allows specification of more than one Group By option in the same record set.

**Cube** is claused with Group By and creates a subtotal of all possible combinations of the set of column in its argument. Aftesr using Cube on dimensions, we can compute all possible aggregation questions on those dimensions.

**Rollup**, also claused with Group By, is used to compute the aggregate at the hierarchy levels of a dimension.

GROUP BY a, b, c with ROLLUP assumes that the hierarchy is "a" drilling down to "b" drilling down to "c".