**Task : Extract top 3 medications for each suburb in Victoria**

**Steps (MySQL):**

(1) Create indexes for different tables based on the anticipated query fields

create index idx\_t1 on transactions(Patient\_ID,Drug\_ID);

create index idx\_t2 on ChronicIllness\_LookUp(MasterProductID,Chronicillness);

create index idx\_t3 on ATC\_LookUp(ATCLevel3Code,ATCLevel3Name,ATCLevel5Code,ATCLevel5Name);

create index idx\_t4 on Drug\_LookUp(MasterProductID,ATCLevel3Code,ATCLevel5Code);

create index idx\_t3\_5 on ATC\_LookUp(ATCLevel5Code,ATCLevel5Name);

create index idx\_t4\_5 on Drug\_LookUp(MasterProductID,ATCLevel5Code);

create index idx\_patients on patients(Patient\_ID,postcode);

(2) Now create a view out of multiple table join

CREATE VIEW ZipCodeATC5View

AS

(SELECT patients.postcode, ATCLevel5Name

FROM ATC\_LookUp FORCE INDEX (idx\_t3\_5),

transactions FORCE INDEX (idx\_t1),

ChronicIllness\_LookUp FORCE INDEX (idx\_t2),

Drug\_LookUp FORCE INDEX (idx\_t4\_5),

Patients FORCE INDEX(idx\_patients)

where (

Patients.Patient\_ID = TRANSACTIONS.Patient\_ID

and

TRANSACTIONS.Drug\_ID = Drug\_LookUp.MasterProductID

and Drug\_LookUp.ATCLevel5Code = ATC\_LookUp.ATCLevel5Code

and Drug\_LookUp.MasterProductID = ChronicIllness\_LookUp.MasterProductID)

and (ChronicIllness\_LookUp.chronicillness = "Diabetes")

and Patients.postcode between 3000 and 3139 );

This view has zip codes and respective ATCLevel5Names in it, so now it's easy to work with having just this view.

(3) Now, filter top 3 medications from each of the zip code -

select postcode, ATCLevel5Name as ATCLevel5Name, count(ATCLevel5Name) as consumption from ZipCodeATC5View

where postcode = <postcode>

group by ATCLevel5Name

order by count(ATCLevel5Name) desc limit 3;

This query will be executed from the R code. It will iterate for each zip code and produce top 3 medications per zip code. Parameter for each zip codes are given from R(in a loop).