VA Data: Data Report:

Current struggle is that they utilize a number of calculated fields and I/we do not have their feature vector table.

On the surface they do not appear to list the logic. So without the scripts, logic and table it will take a bit of time to produce and the same attributes with a similar accuracy.

Here is what I have done so far:

-Dropped in indexing on fields identified in the Engineering Notebook

-Turned on Oracle Text and configured that for one of the wildcard searches (CTXCAT index for short text field)

-Begin a query to do their strict and variable filtering...just now opening it up a bit based on ear.

The DOB and DOD columns in the person table are showing it as Varchar. The data in the table is showing 4-digit Year or null. It will be good, if we declare it as number for rating calculations.

We need to change it to DATE, if we need to maintain complete DOB or DOD.

 CREATE TABLE "DEVELOPER"."AH4929\_PERSON"   
   (    "PTCPNT\_VET\_ID" NUMBER(15,0),   
    "GENDER" VARCHAR2(9 BYTE),   
    "DOB" VARCHAR2(4 BYTE),   
    "DOD" VARCHAR2(4 BYTE),   
    "STATE\_CODE" VARCHAR2(2 BYTE)

So, a lot of numerical fields end up as character fields...why that is a loss. Sometimes I see they put in things like "Unknown" rather than leaving it as null which complicates things. I want to leave the source alone but as we build out our learning/training tables or analysis tables, where it makes sense, we can reset the data type and transform.

select

CASE WHEN rd.diagnosis\_code = '6100' THEN 1 ELSE 0 END AS "A6100",

CASE WHEN rcc.cntntn\_clsfcn\_id = '3140' THEN 1 ELSE 0 END AS "C3140",

CASE WHEN rd.diagnosis\_code = '6204' THEN 1 ELSE 0 END AS "A6204",

CASE WHEN rcc.rcc.cntntn\_clsfcn\_id = '6850' THEN 1 ELSE 0 END AS "C6850",

CASE WHEN rd.diagnosis\_code = '6260' THEN 1 ELSE 0 END AS "A6260",

CASE WHEN rd.diagnosis\_code = '6200' THEN 1 ELSE 0 END AS "A6200",

p.dob "DOB",

rcc.date\_of\_claim "MAXDATEO",

rd.prfil\_dt "ADATEPR",

rcc.claim\_ro\_number "RONUMBER"

from ah4929\_person p

left join ah4929\_rating\_corp\_claim  rcc on rcc.ptcpnt\_vet\_id = p.ptcpnt\_vet\_id

left join ah4929\_rating\_decision rd on rd.ptcpnt\_vet\_id = rcc.ptcpnt\_vet\_id and rd.prfil\_dt = rcc.prfil\_dt

where 1=1

and rcc.prfil\_dt >= rcc.date\_of\_claim --strict filter on claim

and CTXSYS.CATSEARCH(rcc.bnft\_claim\_type\_cd, '020\*', '') > 0 --variable filter on claim

and rd.system\_type\_cd='C' --strict filter on diagnosis

and NOT (rd.begin\_dt IS NULL) --strict filter on diagnosis

and rd.begin\_dt < rd.prfil\_dt --strict filter on diagnosis

and (rd.end\_dt IS NULL) OR (rd.end\_dt >= rd.prfil\_dt) --strict filter on diagnosis

and rd.dsblty\_decn\_type\_cd = 'SVCCONNCTED'

and rcc.cntntn\_clsfcn\_id IN (2200,2210,2220,3140,3150,4130,4210,4700,4920,6850) --variable for ear

and rd.diagnosis\_code IN ('6100','6200','6201','6202','6204','6205','6207','6209','6210','6211','6260') --Variable for ear

;

-Added additional indexes to the identified columns used in the MITRE Engineering Notebooks

-Added support for Oracle Text indexing in an attempt to increase the speed at which text searches (heavily used) execute

-Multiple discussions regarding the current state of the data. There are a number of aggregated fields that caused some issues.

\*We mitigated this risk by taking the approach of aggregating within the two primary tables, we will join the aggregated data.

Had an overview meeting with team to review the data and looking to pick back up Tuesday to breakdown the tasking to make us more efficient in producing the aggregate data.