/\*\*\*\*\*\* Match to Most Similar Partial & Excessive Match \*\*\*\*\*\*/

USE [Assessment]

-- Describe the patient's conditions in the table #Patients

DROP TABLE #Patients

CREATE Table #Patients (pOld Float, pMale Float, pEat Float, pTransfer Float, pGroom Float, pToilet Float, pBathe Float

, pWalk Float, pDress Float, pBowel Float, pUrine Float)

INSERT INTO #Patients (pOld, pMale, pEat, pTransfer, pGroom, pToilet, pBathe, pWalk, pDress, pBowel, pUrine)

VALUES (1.,1.,0.,0.,0.,1.,0.,1.,0.,0.,0.)

-- Labeling the variables and calculating age at time of assessment

DROP TABLE #data

SELECT [column 0] AS ID

,Cast(IIF(isnumeric([column 1])=1, [Column 1], null) AS Float)

+ CAST(IIF(isnumeric([column 5])=1, [Column 5], null) AS Float)/365. AS [Age]

,IIF ([Column 2]='M', 1, 0) AS [uMale]

,CAST(iif(isnumeric([column 5])=1, [Column 5], null) AS Float) AS [Days since first assessment]

,IIF([column 7]='1',1.,0.) AS [uEat]

,IIF([column 8]='1',1.,0.) AS [uTransfer]

,IIF([column 9]='1',1.,0.) AS [uGroom]

,IIF([column 10]='1',1.,0.) AS [uToilet]

,IIF([column 11]='1',1.,0.) AS [uBathe]

,IIF([column 12]='1',1.,0.) AS [uWalk]

,IIF([column 13]='1',1.,0.) AS [uDress]

,IIF([column 14]='1',1.,0.) AS [uBowel]

,IIF([column 15]='1',1.,0.) AS [uUrine]

,IIF([column 16]='1',1.,0.) AS [Dead]

,IIF([column 17]='1',1.,0.) AS [Assessment number]

INTO #Data

FROM [dbo].[Assessments]

-- (1306456 row(s) affected)

-- Calculating mortality within 6 months of assessment

DROP TABLE #Data2, #Dead

SELECT id, [assessment number] , Max(age) as AgeDeath INTO #Dead

FROM #data

WHERE Dead='1'

GROUP BY id, [assessment number]

DECLARE @Age as float

SET @Age=(SELECT avg(Age) FROM #Data )

SELECT #data.\*

, IIF(Agedeath is null, 0, iif(Agedeath-Age <180, 1,0)) AS Dead6M

, iif(Age>@age, 1,0) AS [uOld]

INTO #Data2

FROM #Data left join #Dead

ON #Data.ID=#Dead.ID

and #Data.[Assessment number]=#dead.[Assessment number]

-- 196,533 dead people -- merged with 1,306,456 assessments

-- Calculate probability of mortality in strata with mroe than 29 cases

DROP TABLE #strata

Select SUM(Cast(Dead6m as float))/Count(Cast(Dead6m as float)) As Prob

, Count(ID) as nCases

,[uOld],[uMale],[uEat],[uTransfer],[uGroom],[uToilet],[uBathe],[uWalk],[uDress],[uBowel],[uUrine]

INTO #Strata

FROM #Data2

GROUP BY [uOld],[uMale],[uEat],[uTransfer],[uGroom],[uToilet],[uBathe],[uWalk],[uDress],[uBowel],[uUrine]

HAVING Count(ID)>29

-- 981 strata total strata, 430 more than 29 cases

-- Calculate partial and excessive matches of strata to the patient

DROP TABLE #Matches

SELECT IIF([uWalk]=pWalk,1.,0.) + IIF([uToilet]=pToilet,1.,0.) + IIF([uOld]=pOld,1.,0.)

+ IIF([uMale]=pMale,1.,0.) + IIF([uEat]=pEat,1.,0.) + IIF([uTransfer]=pTransfer,1.,0.)

+ IIF([uGroom]=pGroom,1.,0.) + IIF([uBathe]=pBathe,1.,0.) + IIF([uDress]=pDress,1.,0.)

+ IIF([uBowel]=pBowel,1.,0.) + IIF([uUrine]=pUrine,1.,0.) AS Matches

, pWalk\*IIF([uWalk]<>pWalk,1.,0.) + pToilet\*IIF([uToilet]<>pToilet,1.,0.) + pOld\*IIF([uOld]<>pOld,1.,0.)

+ pMale\*IIF([uMale]<>pMale,1.,0.) + pEat\*IIF([uEat]<>pEat,1.,0.) + pTransfer\*IIF([uTransfer]<>pTransfer,1.,0.)

+ pGroom\*IIF([uGroom]<>pGroom,1.,0.) + pBathe\*IIF([uBathe]<>pBathe,1.,0.) + pDress\*IIF([uDress]=pDress,1.,0.)

+ pBowel\*IIF([uBowel]<>pBowel,1.,0.) + pUrine\*IIF([uUrine]<>pUrine,1.,0.) As PartialMatches

, uWalk\*IIF([uWalk]<>pWalk,1.,0.) + uToilet\*IIF([uToilet]<>pToilet,1.,0.) + uOld\*IIF([uOld]<>pOld,1.,0.)

+ uMale\*IIF([uMale]<>pMale,1.,0.) + uEat\*IIF([uEat]<>pEat,1.,0.) + uTransfer\*IIF([uTransfer]<>pTransfer,1.,0.)

+ uGroom\*IIF([uGroom]<>pGroom,1.,0.) + uBathe\*IIF([uBathe]<>pBathe,1.,0.) + uDress\*IIF([uDress]=pDress,1.,0.)

+ uBowel\*IIF([uBowel]<>pBowel,1.,0.) + uUrine\*IIF([uUrine]<>pUrine,1.,0.) As ExcessiveMatches

, \*

INTO #Matches

FROM #strata CROSS JOIN #Patients

WHERE nCases>29 order by Prob

--(430 rows affected)

--Most similar strata where a feature missed in patient is more important than a feature in the strata

SELECT TOP 1 Matches/(Matches+.2\*ExcessiveMatches+.8\*PartialMatches) AS Similarity -- Tversky's scale at .8

, \*

FROM #matches

WHERE PartialMatches+ExcessiveMatches<>0

ORDER BY Matches/(Matches+.2\*ExcessiveMatches+.8\*PartialMatches) desc

--Most similar strata where a feature missed in patient is less important than a feature in the strata

SELECT TOP 1 Matches/(Matches+.8\*ExcessiveMatches+.2\*PartialMatches) AS Similarity -- Tversky's scale at .2

, \*

FROM #matches

WHERE PartialMatches+ExcessiveMatches<>0

ORDER BY Matches/(Matches+.8\*ExcessiveMatches+.2\*PartialMatches) desc