SQL query

1. Get distance function

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: UserDefinedFunction [dbo].[getDistanceSTDistanctMile] Script Date: 8/12/2019 4:00:19 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE function [dbo].[getDistanceSTDistanceMile] (@lat1 as real,

@lon1 as real,

@lat2 as real,

@lon2 as real)

returns real as

begin

declare @d as real

declare @NWI geography, @EDI geography

SET @NWI = geography::Point( @lat1,@lon1, 4326)

SET @EDI = geography::Point( @lat2,@lon2, 4326)

SELECT @d = @NWI.STDistance(@EDI) / 1609.344

return @d

end;

GO

1. CompiledDataVersion1 view

This view gather all chosen fields of property into one place

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[CompiledDataVersion1] Script Date: 8/12/2019 4:09:24 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

create view [dbo].[CompiledDataVersion1]

as

select

a.ParcelNumber,

a.AppraisalAccountType,

a.Buildings,

a.LandGrossSquareFeet,

a.LandNetSquareFeet,

a.AppraisalDate,

a.Latitude,

a.Longtitude,

b.BuildingID,

b.PropertyType,

b.SquareFeet,

b.PercentComplete,

b.Condition,

b.Quality,

c.PhysicalAge,

c.YearBuilt,

c.YearRemodeled,

d.AccountType,

d.TaxableValuePriorYear,

d.TaxableValueCurrentYear,

e.Zipcode

from dbo.AppraisalAccount a

left join dbo.Improvement b ON b.ParcelNumber = a.ParcelNumber

left join dbo.ImprovementBuiltas c ON c.ParcelNumber = b.ParcelNumber AND c.BuildingID = b.BuildingID

left join dbo.TaxAccount d ON d.ParcelNumber = c.ParcelNumber

left join dbo.AddressPoint e ON e.ParcelNumber = d.ParcelNumber

WHERE a.AppraisalAccountType = 'Residential' AND b.PropertyType = 'Residential' AND d.AccountType = 'REAL'

AND e.Zipcode in ('98328','98338','98375','98387','98385','98448','98374','98373','98446','98445','98444','98467','98464','98371','98466','98391','98372','98335','98335','98394','98332','98329')

GO

1. Calculate distance between filtered Property and 10% sample of Crime data

insert into [dbo].[LandCrimeDistanceSample2]

select ta.ParcelNumber, d.OBJECTID,

ta.Latitude,ta.Longtitude, d.lattitude, d.longtitude,

[dbo].[getDistanceSTDistanceMile](ta.Latitude,ta.Longtitude, d.lattitude, d.longtitude) as Distance

from

(select a.ParcelNumber, a.Latitude as Latitude, a.Longtitude as Longtitude, ap.ZipCode

from AppraisalAccount a inner join Sale s on a.ParcelNumber= s.ParcelNumber

left join AddressPoint ap on a.ParcelNumber= ap.ParcelNumber

where s.SaleDate >='2018-01-01' and

ap.ZipCode in ('98328', '98338','98375','98387',

'98385','98448','98374','98373','98446',

'98445','98444','98467','98464','98371',

'98466','98391','98372','98335','98335',

'98394','98332','98329')

) ta ,

(select \* from CrimeData TABLESAMPLE (10 PERCENT)) d

1. Groupedsample View

This view integrate house attribute data from CompiledDataVersion1 view, table LandCrimeDistanceSample2 and table Crime Data. Then It aggregated and grouped crime by new 5 categories as mentioned in the paper.

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[groupedsample] Script Date: 8/12/2019 4:17:27 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE view [dbo].[groupedsample]

as

select d.\*, e.CrimeCategory, e.NoofCrime

from CompiledDataVersion1 d right join

(select tb1.ParcelNumber, tb1.CrimeCategory, count( tb1.CrimeCategory) as NoofCrime

from

(select a.ParcelNumber, c.Public\_Nam,case

when Public\_Nam in ('Arson - Non-residential',

'Arson - Residential',

'Burglary - Non-residential',

'Burglary - Residential','Fraud or Forgery','Motor Vehicle Theft','Possession of Stolen Property',

'Robbery - Business','Robbery - Residential','Robbery - Street','Robbery - Other',

'Theft - Gas Station Runout','Theft - Mail','Theft - Other','Theft - Vehicle Prowl','Theft -Shoplifing','Trafficking in Stolen Property'

) then 'PropertyCrime'

when Public\_Nam like '%Drug%' then 'DrugCrime'

when Public\_Nam = 'Homicide' then 'Homicide'

when Public\_Nam in ('Assault - Aggravated','Assault - Simple','Intimidation','Telephone Harassment','Vandalism - Non-residential','Vandalism - Residential') then 'PersonalCrime'

else 'OtherCrime'

end as CrimeCategory

from LandCrimeDistanceSample2 a left join CrimeData c on a.ObjectID = c.OBJECTID

where a.Distance <=1) tb1

group by tb1.ParcelNumber, tb1.CrimeCategory) e on d.ParcelNumber = e.ParcelNumber

GO

1. Groupeddatafull view

This view spread crime data in groupedsample view to column. This action is to tidy data table by using pivot function, make every single row is correspond with only one property.

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[groupeddatatull] Script Date: 8/12/2019 4:23:55 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE view [dbo].[groupeddatafull]

as

select a.ParcelNumber, a.AppraisalAccountType,

a.Buildings, a.LandGrossSquareFeet,

a.LandNetSquareFeet, a.AppraisalDate,

a.Latitude, a.Longtitude, a.BuildingID, a.PropertyType,

a.SquareFeet, z.PercentComplete, a.Condition, a.Quality, a.PhysicalAge,

a.YearBuilt, a.YearRemodeled,

a.AccountType, a.TaxableValuePriorYear,

a.TaxableValueCurrentYear, a.Zipcode, b. [DrugCrime], b.[Homicide], b.[OtherCrime], b.[PersonalCrime], b.[PropertyCrime]

from [dbo].[CompiledDataVersion1] a left join

(select ParcelNumber,[DrugCrime],[Homicide],[OtherCrime],[PersonalCrime],[PropertyCrime]

from

(select ParcelNumber, CrimeCategory, NoofCrime

from [dbo].[groupedsample] ) as a

pivot (max(a.NoofCrime) for a.CrimeCategory in ([DrugCrime],[Homicide],[OtherCrime],[PersonalCrime],[PropertyCrime]))as p ) b

on a.ParcelNumber = b.ParcelNumber

WHERE

--Zipcode in ('98371', '98372', '98373','98374','98375') and

a.ParcelNumber in( select l.ParcelNumber from LandCrimeDistanceSample2 l)

GO

1. GroupedDataWithSale view

This view is the data in groupeddatafull integrated with sale data. We only select those most recent sale data from 01-01-2018 until now.

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[GroupedDatawithSale] Script Date: 8/12/2019 4:32:09 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

/\*\*\*\*\*\* Script for SelectTopNRows command from SSMS \*\*\*\*\*\*/

CREATE view [dbo].[GroupedDatawithSale]

as

select a.ParcelNumber,

a.Buildings,

a.LandNetSquareFeet,

a.SquareFeet,

a.Condition,

a.Quality,

a.PhysicalAge,

a.YearBuilt,

a.YearRemodeled,

a.TaxableValueCurrentYear,

a.TaxableValuePriorYear,

isnull(a.DrugCrime,0) as DrugCrime,

isnull(a.Homicide,0) as Homicide,

isnull(a.PropertyCrime,0)as PropertyCrime,

isnull(a.PersonalCrime,0) as PersonalCrime,

isnull(a.OtherCrime,0) as OtherCrime, s.SalePrice, s.SaleDate

from groupeddatatull a inner join

(select ParcelNumber, SalePrice, SaleDate

from(

select Row\_number()

OVER (

partition BY (Sale.ParcelNumber)

ORDER BY Sale.SaleDate Desc) SaleOrder, ParcelNumber, SaleDate,SalePrice

from Sale) so

where SaleOrder=1) s on a.ParcelNumber = s.ParcelNumber

where s.SaleDate >='2018-01-01'

GO

After having a view with final data, we insert those data into table ***FinalData0809***. To keep this dataset constant

1. FinalNormalizeData view

This view is a normalized version of final dataset. We used this view for clustering section.

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[FinalNormalizedData] Script Date: 8/12/2019 4:34:52 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

CREATE view [dbo].[FinalNormalizedData]

as

select ParcelNumber, (cast(f.Buildings as float) - 0)/(10-0) as Buildings

, (f.LandNetSquareFeet - 1500)/(4316796-1500) as LandNetSquareFeet

, (f.SquareFeet - 192)/(7230-192) as SquareFeet

, (cast(f.PhysicalAge as float) - 0)/(119-0) as PhysicalAge

, (cast(f.YearBuilt as float) - 1890)/(2019-1890) as YearBuilt

, (cast(f.YearRemodeled as float)- 0)/(2019-0) as YearRemodeled

, (f.TaxableValueCurrentYear - 0)/(1979600-0) as TaxableValueCurrentYear

, (f.TaxableValuePriorYear - 0)/(1882900-0) as TaxableValuePriorYear

, (cast(f.DrugCrime as float)- 0)/(51-0) as DrugCrime

, (cast(f.Homicide as float) - 0)/(2-0) as Homicide

, (cast(f.PropertyCrime as float)- 0)/(459-0) as PropertyCrime

, (cast(f.PersonalCrime as float)- 0)/(108-0) as PersonalCrime

, (cast(f.OtherCrime as float)- 0)/(393-0) as OtherCrime

, (f.SalePrice - 500)/(21700000-500) as SalePrice

, f.SaleDate, f.Condition, f.Quality

from FinalData0809 f

GO

Finally, We used visual studio SQL data tool to cluster final dataset, then create data with cluster column.

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[FinalDataClusterWithCrime] Script Date: 8/13/2019 3:06:41 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

create view [dbo].[FinalDataClusterWithCrime]

as

select f.\*, c.[$CLUSTER] as Cluster

from FinalData0809 f inner join ClusteredDataWithCrime c on f.ParcelNumber = c.ParcelNumber

GO

USE [19su5510\_lehuy]

GO

/\*\*\*\*\*\* Object: View [dbo].[FinalDataClusterWithoutCrime] Script Date: 8/13/2019 3:10:30 PM \*\*\*\*\*\*/

SET ANSI\_NULLS ON

GO

SET QUOTED\_IDENTIFIER ON

GO

create view [dbo].[FinalDataClusterWithoutCrime]

as

select f.\*, c.[$CLUSTER] as Cluster

from FinalData0809 f inner join ClusteredDataWithoutCrime c on f.ParcelNumber = c.ParcelNumber

GO