Somerville Happiness Survey Analysis

By Anant Kumar

Table of Contents

Executive Summary

The City of Somerville in Massachusetts has conducted a unique Happiness Survey with its residents every two years since 2011. As rates of anxiety, depression and other mental health issues increase nationwide [1], Somerville's rich data set presents an opportunity to reveal what factors enable people to report higher levels of happiness, wellbeing, and satisfaction from life in general. According to Merriam Webster [2], happiness is "a state of well-being and contentment". If happiness is indeed a state, then there must be factors that play a role in attaining that state.

Instead of creating complicated models, the scope of the current analysis is simply to understand how the lives and personal dynamics of people in Somerville relate to their happiness. I want to explore the connections between demographics, family situations, money, and happiness to gain practical insights that could improve policies, resource allocations, and community spaces contributing to the well-being of all the communities and social groups.

Major Findings

Trends across six years (2011-2021)

- **Peak Happiness:** Residents reported the highest overall happiness in 2015 in the six years being considered.

- **Fluctuations:** Satisfaction with life, happiness, and satisfaction with Somerville as a place generally moved in tandem, peaking in 2015 and dipping in 2017, 2019, and 2021. 2011 stands as an exception.

Demographic Differences

- Ethnicity and Happiness: Among respondents, the Hispanic or Latino population consistently reported the highest happiness and life satisfaction, while those of two or more races expressed the lowest.

- Marital Happiness: Married respondents reported the highest happiness levels. In contrast, single individuals (never married) and those living with partners without

marriage reported the lowest happiness and life satisfaction. Additionally, individuals living with children reported higher levels of happiness and satisfaction.

- Gender and Happiness: Gender differences showed that females reported higher happiness and overall life satisfaction compared to males. Nonbinary individuals reported significantly lower levels of happiness and satisfaction.

- Age and Happiness: Retirees and those over 75 experienced the highest happiness and life satisfaction, while young people (17 and under) and students reported the lowest levels.

Socioeconomic Factors

- Income and Happiness: Income-wise, individuals earning above $200k reported the highest happiness and life satisfaction, whereas those earning less than $10k reported the lowest levels. However, there is no linear relationship between higher income and increased happiness.

- Housing and Happiness: Homeownership significantly correlated with higher happiness and life satisfaction compared to renting.

Main Conclusions

The Somerville Happiness Survey reveals interesting patterns of wellbeing, happiness and satisfaction levels being tied to several demographic factors. Most notably, Hispanic/Latino and married participants consistently report greater happiness than multi-racial and single groups respectively. This could be due to stronger cultural communities and family support systems. Residents with children living at home demonstrate exceptional happiness and satisfaction, potentially due to the enhanced sense of community and belonging children bring. Conversely, lower happiness levels among single and cohabiting individuals without marriage may stem from a lack of extended social and emotional support structures beyond romantic relationships.

Meanwhile, younger respondents such as students report lower life satisfaction. This could be attributed to financial instability and the challenges of transitioning from adolescence to adulthood. This could also be linked to the struggle among young people to find a supportive community and a sense of belonging. In conclusion, the presence of a support system and a

community emerges as a significant factor in overall happiness and life satisfaction, whether it be in the form of extended family, a spouse, or children.

Exploring socioeconomic factors also reveals interesting patterns, with higher income correlating with increased happiness and lower income correlating to decreased happiness but lacking a straightforward linear relationship. The association between homeownership and higher happiness might be due to the sense of stability and ownership which a house brings.

The analysis also raises some questions when you see the differences in happiness across genders. Women seem to be happier overall, which raises questions like - is it about how society expects us to act? Are there different gender roles we play that affect our feelings? And then there's the fact that non-binary people report feeling less happy than others. It requires deeper analysis to understand what unique challenges they have, and how societal norms might be impacting their well-being.

The Dataset

The Somerville Happiness Survey is an ongoing initiative by the city government of Somerville, Massachusetts to understand the wellbeing overall happiness of its population. The survey, called Happiness Survey, is done every two years providing a valuable source of data to inform policymaking and governance decisions at the local level. The main goal of this initiative is to get input to help leaders make decisions that support happiness and wellbeing across the community. Hearing from a random sample of residents gives a broad range of perspectives beyond just the people who show up at public forums or meetings.

Somerville has been doing this since 2011 when they became the first US city to survey happiness. The idea came from seeing some countries (notably Bhutan who started measuring GNH- Gross National Happiness from 2008) start to track overall wellbeing beyond basic economic numbers [3]. The city planning team put the survey together with help from academics and researchers. Past survey results have led to new investments - more recycling pickup spots, more trees planted, upgrading parks, etc. Illustrating that survey has resulted in tangible benefits to the community.

Data Collection Process

Every two years, residents randomly get survey via mail, email, phone or Facebook. It is usually also translated into the most common local languages like Spanish. The survey asks simple questions like:

- How happy do you feel today?
- How satisfied with Somerville overall?

It also includes topics like:

- Your neighborhood
- Safety
- Outdoor green spaces
- Getting around the city
- Housing costs
- City services

Research Process

This happiness analysis was conducted based on dataset published on Data.Gov website based on a biennial survey in the city of Somerville in Massachusetts, USA.

The DCOVA framework explained in Levine, Stephan, and Szabat (2016) is used to inspect the data:

Define

The Somerville happiness survey responses dataset was used to do the entire analysis. The survey had 10 questions related to happiness and satisfaction as well as demographic questions like age range, marital status, housing status, income, employment status etc. Not all the data collected was used for this analysis, below is a data dictionary of the data being used in the current analysis.

Survey Question	Data Type	Comment
Combined_ID	Integer	A unique ID for the row
Survey Year	Integer	The year the survey was taken
How.happy.do.you.feel.right.now	Integer	Possibble Values: Integer from 1 to 10 Write-in half values (e.g. 5.5) have been rounded to the next integer (5.5 => 6)
How.satisfied.are.you.with.your.life.in.general	Integer	Possibble Values: Integer from 1 to 10 Write-in half values (e.g. 5.5) have been rounded to the next integer (5.5 => 6)
How.satisfied.are.you.with.Somerville.as.a.place.to.live	Integer	Possibble Values: Integer from 1 to 10 Write-in half values (e.g. 5.5) have been rounded to the next integer (5.5 => 6)
What.is.your.gender	Text	•People who specified a gender such as "genderqueer" were coded as "nonbinary." •People who indicated a binary gender and were trans (e.g. "trans man") were coded with the appropriate binary gender ("trans man" => "male"). •People who checked off both male and female (and did not indicate a nonbinary gender) and/or wrote in something like "A couple filled this out" were coded as "multiple people."
Age	Text	Possible Values: Most years (2011 phone or online, 2013, 2015, 2017, 2019, 2021): • 18-24 • 25-34 • 35-44 • 45-54 • 55-64 • 65-74 • 75+ 2011 (Mail): • 18-21 • 22-25 • 26-30 • 31-40 • 41-50 • 51-60 • 61+
Marital.status	Text	Possible Values:

		• Single (never married) • Divorced/ Separated • Living with a partner, but not married • Married • Widowed
What.is.your.race.or.ethnicity	Text	Possible Values: • White • Black or African American • Hispanic or Latino • Asian or Pacific Islander • American Indian or Alaska Native • Two or more races • Another race
Do.you.have.children.age.18.or.younger.who.live.with.you	Text	Possible Values: • yes • no
Describe.your.housing.status.in.Somerville	Text	Possible Values: • Own • Rent • Other
What.is.your.annual.household.income	Text	Possible Values: Most Years (2015, 2017, 2019, 2021): • Less than $10,000 • $10,000 to $24,999 • $25,000 to $49,999 • $50,000 to $74,999 • $75,000 to $99,999 • $100,000 to $149,999 • $150,000 to $199,999 • $200,000 or more 2013: Same as most years, but the maximum bucket is "$150,000 or more" 2011 (phone): • Less than $20,000 • $20,000 to $39,999 • $40,000 to $59,999 • $60,000 to $79,999 • $80,000 to $99,999 • $100,000 to $119,999 • $120,000 to $139,999 • $140,000 to $159,999 • $160,000 to $179,999 • $180,000 to $199,999 • $200,000 or more 2011 (mail): • Less than $10,000 • $10,000 to $19,999 • $20,000 to $39,999 • $40,000 to $59,999

		• $60,000 to $79,999 • $80,000 to $99,999 • $100,000 or more
Precinct	Text	Precinct
survey_method	Text	Possible Values: • Mail • Phone • Email • Facebook 18-24 year olds
employment_status	Text	Possible Values: • Employed • Unemployed • Student • Homemaker • Retired • Self-employed • Disabled

Figure 1: Survey Questionnaire

Dependent Variables

Three dependent variables (given below) indicating happiness were considered in this analysis. Each question is on a 10-point scale.

- How happy are you right now?
- How satisfied are you with your life in general?
- How satisfied are you with Somerville as a place to live?

Independent Variables

Demographic and socioeconomic factors, including age, gender, marital status, ethnicity, housing status, income, presence of children, precinct, employment status, and survey method, were employed as variables for the analysis of dependent factors. Additionally, the survey year was considered to analyze patterns over the years.

- Survey Year
- What.is.your.gender
- Age
- Marital.status
- What.is.your.race.or.ethnicity
- Do.you.have.children.age.18.or.younger.who.live.with.you
- Describe.your.housing.status.in.Somerville
- What.is.your.annual.household.income
- Precinct
- survey_method
- employment_status

Collect

The Somerville happiness survey responses dataset was downloaded from the data.gov in CSV format, it contained 10,744 records. This dataset was fed into an SQL Server database named

'Somerville'. A staging table named 'SomervilleStagingTable' was created and populated with the happiness survey responses data.

Data Quality: Data quality check for all the dependent and independent variables was done. Null values were also checked.

- **Survey Year**- No Null values and no discrepancies
- **How happy do you feel right now**- 222 Null values which were deleted and there is no other discrepancy.
- **How satisfied are you with your life in general** - 228 Null values which were deleted and there is no other discrepancy.
- **How satisfied are you with Somerville as a place to live** - 6245 Null values which were kept as this is a significant percentage of the total dataset and there is no other discrepancy.
- **What is your gender**- 453 Null values which were deleted and there is no other discrepancy.
- **Age** - 456 Null values which were deleted and there is no other discrepancy. Apart from 2011, the age ranges are in sync. Keeping 2011 data for other analysis and we can filter out 2011 data for specific income and age analysis as and when needed.
- **Marital status** - 4343 Null values which were replaced with 'Unknown' as the number of nulls is a significant percentage of the total dataset. 'Single (never married)' was changed to 'Single, Never Married' and 'Divorced/ Separated' was changed to 'Divorced'.
- **What is your race or ethnicity**- 639 Null values which were deleted and there is no other discrepancy.
- **Do you have children age 18 or younger who live with you** - 6269 Null values which were replaced with 'Unknown' as the number of nulls is a significant percentage of the total dataset.

- **Describe your housing status in Somerville** - 6511 Null values which were replaced with 'Unknown' as the number of nulls is a significant percentage of the total dataset.

- **What is your annual household income** - 1056 Null values which were deleted and there is no other discrepancy. Apart from 2011, the income ranges are in sync. Keeping 2011 data for other analysis and we can filter out 2011 data for specific income and age analysis as and when needed.

- **Precinct -** 9207 Null values which were replaced with 'Unknown' as the number of nulls is a significant percentage of the total dataset.

- **Survey method** - No Null values and no discrepancies

- **Employment status-** 10218 Null values which were replaced with 'Unknown' as the number of nulls is a significant percentage of the total dataset.

Final Count**:** After removing all the NULL values and replacing some nulls with 'Unknown' the dataset has 9053 records.

Exploratory Data Analysis

How.happy.do.you.feel.right.now		*How.satisfied.are.you.with.your.life.in.general*		*How.satisfied.are.you.with.Somerville.as.a.place.to.live*	
Mean	7.5	Mean	7.7	Mean	7.8
Standard Error	0.0	Standard Error	0.0	Standard Error	0.0
Median	8.0	Median	8.0	Median	8.0
Mode	8.0	Mode	8.0	Mode	8.0
Standard Deviation	1.9	Standard Deviation	1.8	Standard Deviation	2.0
Sample Variance	3.6	Sample Variance	3.3	Sample Variance	3.8
Kurtosis	1.4	Kurtosis	1.6	Kurtosis	2.1
Skewness	-1.1	Skewness	-1.1	Skewness	-1.4
Range	9.0	Range	9.0	Range	9.0
Minimum	1.0	Minimum	1.0	Minimum	1.0
Maximum	10.0	Maximum	10.0	Maximum	10.0
Sum	67686.0	Sum	69560.0	Sum	29127.0
Count	9053.0	Count	9053.0	Count	3741.0

Figure 2: Exploratory Data Analysis of Dependent Variables

Organize

After data is inserted into the staging table, it's organized into 11-dimension tables and 1 fact table.

- dimYear- contains the survey year

- dimGender - contains the gender of the respondents

- dimAge - contains the age range of the respondents

- dimMaritalStatus - contains the marital status of the respondents

- dimRace - contains the race of the respondents

- dimChildren - contains the presence of children in the respondents' home

- dimHousingStatus - contains the housing status of the respondents

- dimIncome - contains the income range of the respondents

- dimPrecinct - contains the precinct of the respondents

- dimsurveymethod - contains the survey method used for the survey

- dimEmploymentStatus - contains the employment status of the respondents

- FactSomerville- contains the primary key of all the dimension tables and the three dependent variables being used in the analysis.

The Somerville database was linked to Excel file "Somerville Analysis.xlsm". These tables were then loaded into the PowerPivot model for reporting and visualization. The Somerville database was also linked to Tableau file "Somerville Tableau Analysis". The Entity Relationship Diagram is given below in figure 3.
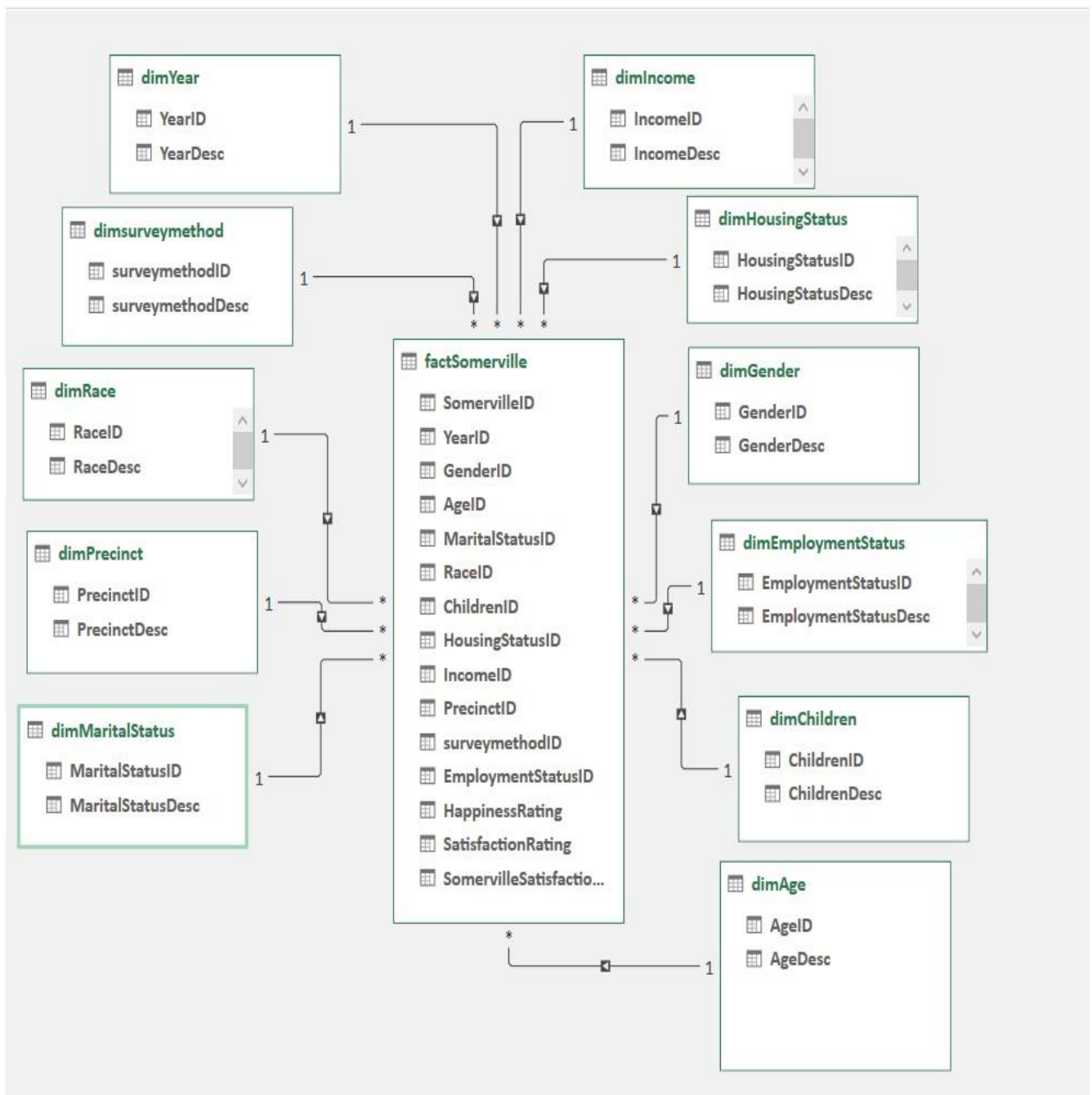
Figure 3: Somerville Entity Relationship Diagram

Visualize

Within the Somerville Analysis.xlsm file, visualizations were created. Two dashboards- one for happiness and other for satisfaction were created with slicers see figure 3 & 4.

Happiness Dashboard (Figure 4)

The visualization in Figure 4 illustrates that the highest happiness rating was recorded in the year 2015, followed by a consistent decline in subsequent years. While the recent decrease in rankings can be reasonably attributed to the widespread impact of Covid and its macro effects on communities and cities, a closer analysis is warranted to understand the specific factors contributing to the peak in 2015. Further investigation is needed to uncover the precise reasons behind the exceptionally high ratings during that particular year.
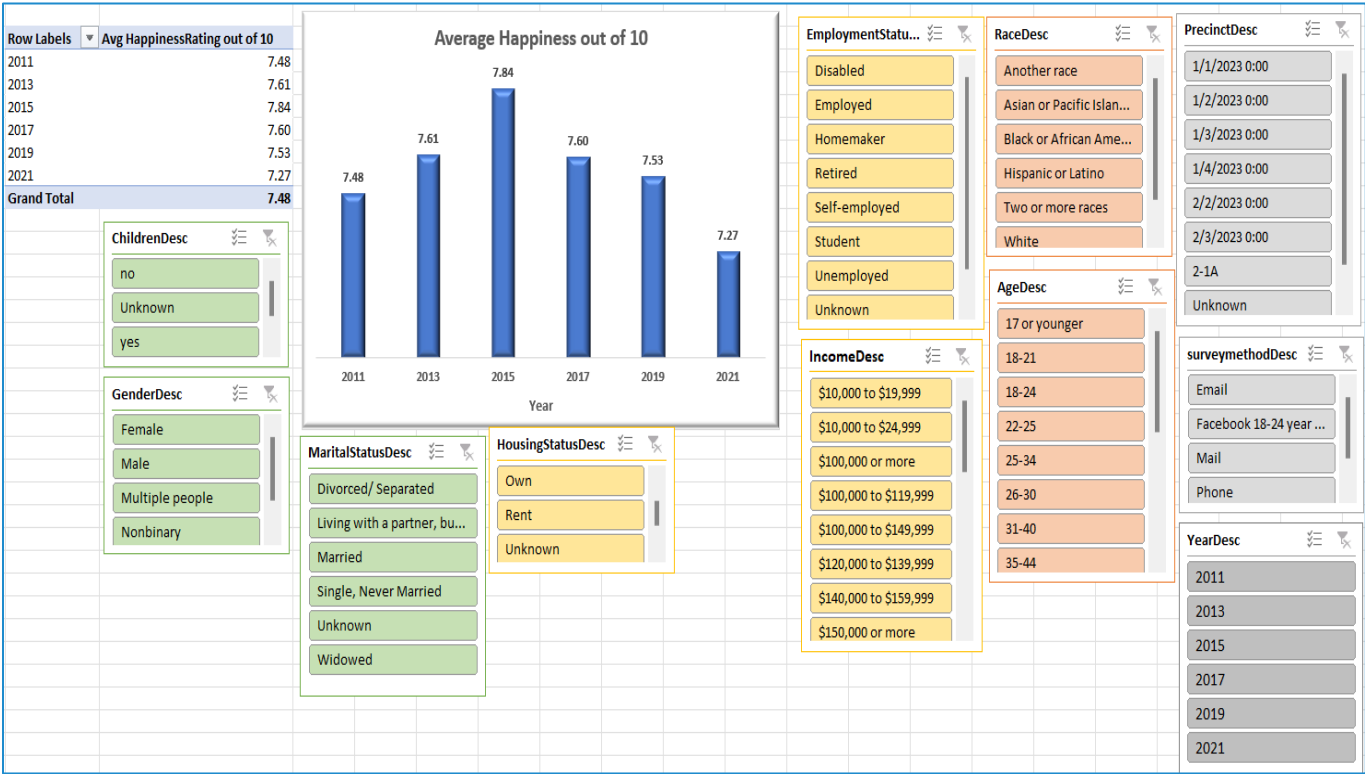
Figure 4: Somerville Happiness Dashboard

Satisfaction Dashboard (Figure 4)

The visualization in Figure 5 illustrates that the highest satisfaction rating was recorded in the year 2015, followed by a consistent decline in subsequent years exactly following the happiness trend. However, if you look closely, the satisfaction of 2017 & 2019 are still higher than the levels in 2011 and 2013 which could be attributed to the changes and improvements made by the city authorities to improve the overall living conditions in the city.
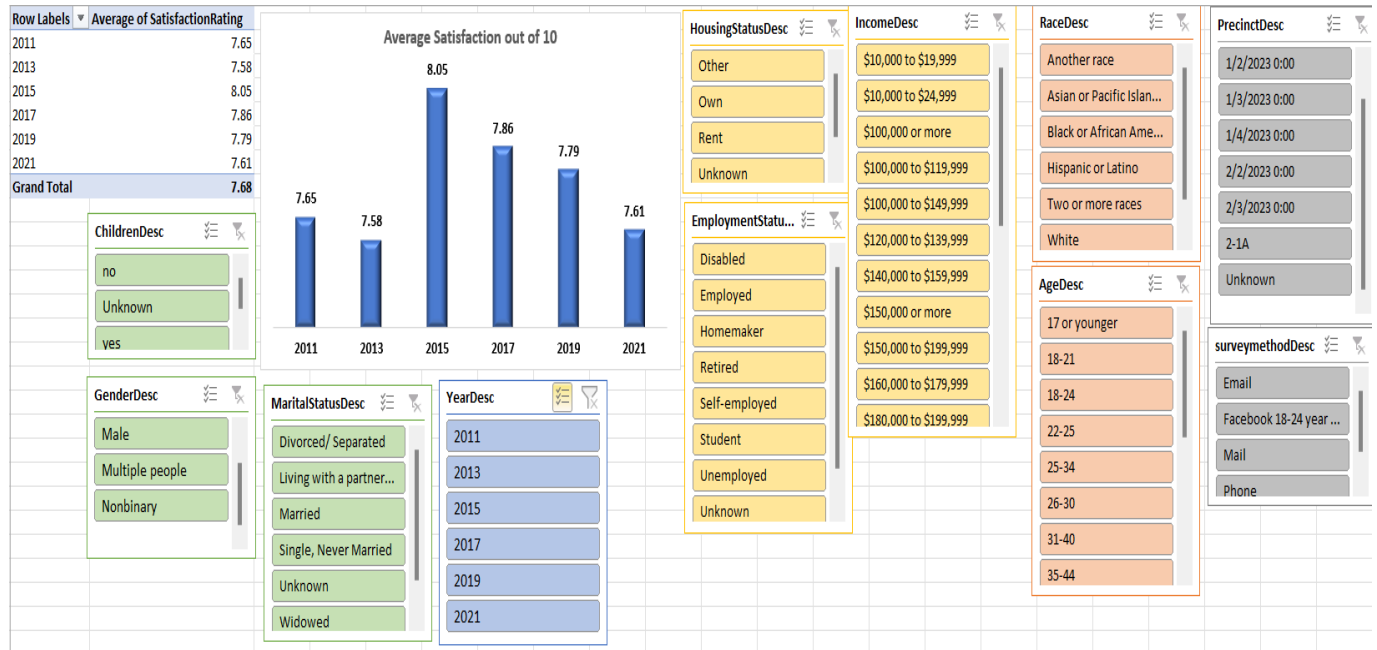
Figure 5: Somerville Satisfaction Dashboard

Somerville Happiness Charts I & II (Figure 6 & 7)

The 8 charts below reflect conclusions earlier mentioned that the Hispanic/Latino and married individuals consistently report higher happiness, possibly linked to robust cultural communities and family support. While Young respondents, particularly students, report lower life satisfaction, likely influenced by financial challenges and the struggle to find a supportive community. Women generally reported higher happiness than men, while nonbinary individuals reported lower levels. Regarding income, those earning above $200k reported the highest happiness, with no clear linear relationship. Homeownership correlated with higher happiness compared to renting.
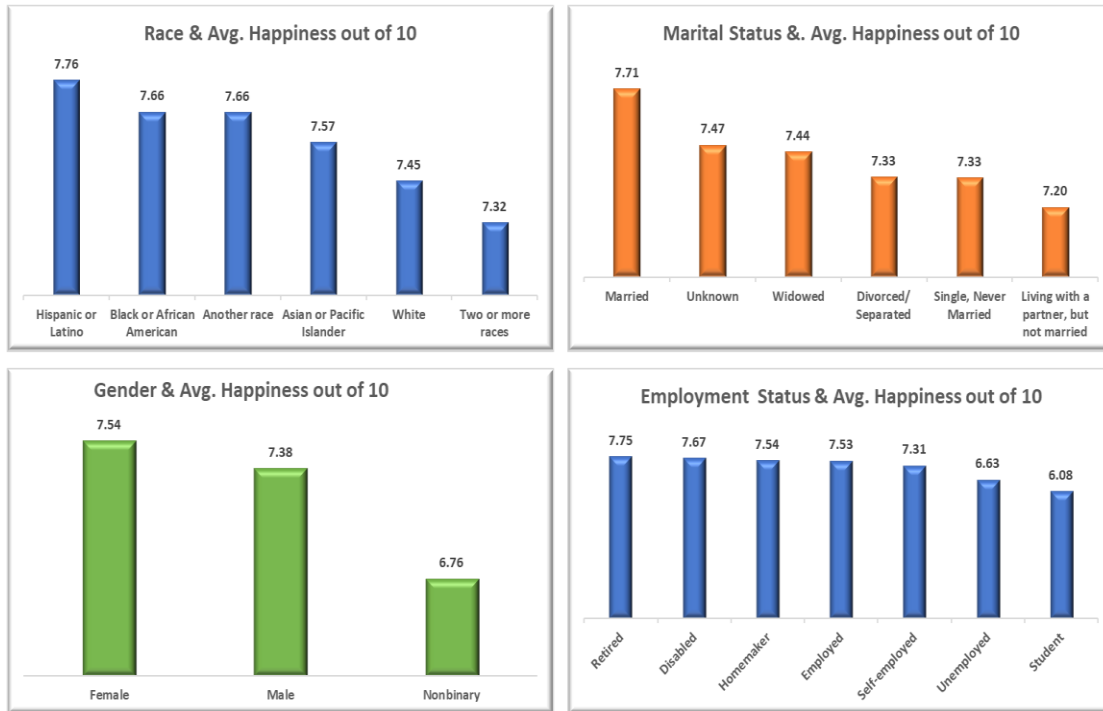
Figure 6: Somerville Happiness Charts I
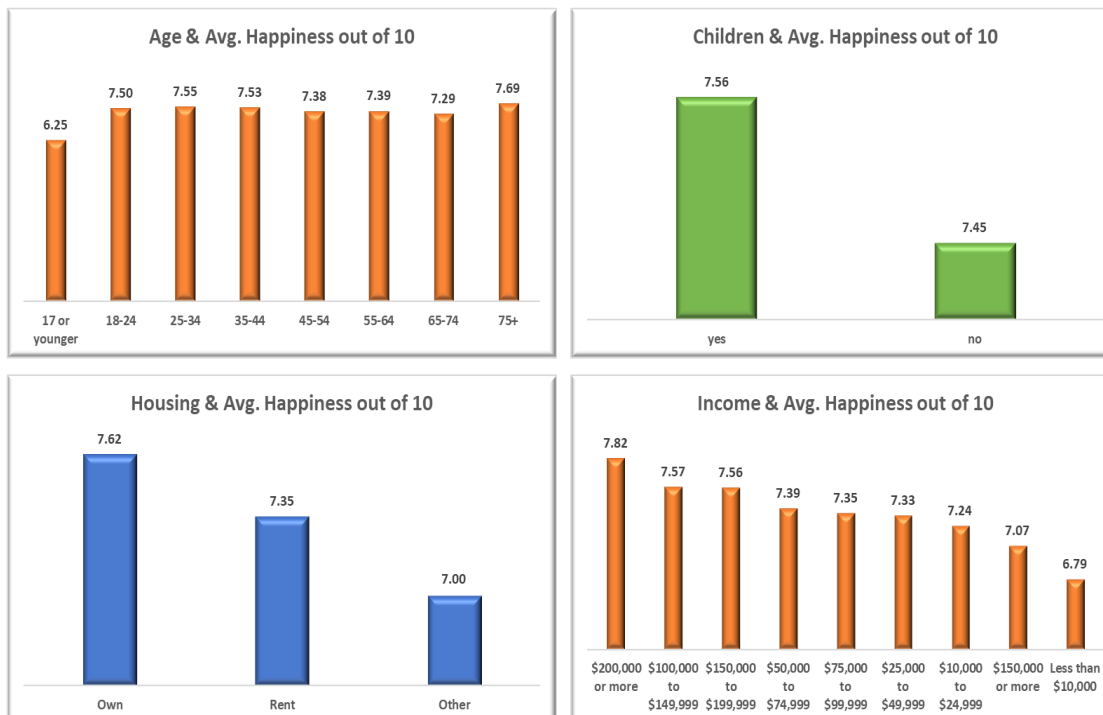
Figure 7: Somerville Happiness Charts II

Somerville Satisfaction Charts I & II (Figure 8 & 9)

The 8 charts below based on the dependent variable of satisfaction rating follow the same trend as the happiness ratings. This is also expected since the two questions in the questionnaire were similar, asking respondents to rate their happiness and their satisfaction from life. Both these are interrelated and affect each other. Happiness affects satisfaction in life and vice versa. One interesting thing to note is that the people who filled 'another race' i.e. people not falling in any of the other brackets were the most satisfied with life in general. Non-binary people, students and unemployed people gave the lowest satisfaction rating, which is expected considering the taboo in society as well as the financial hardships that come with unemployment.
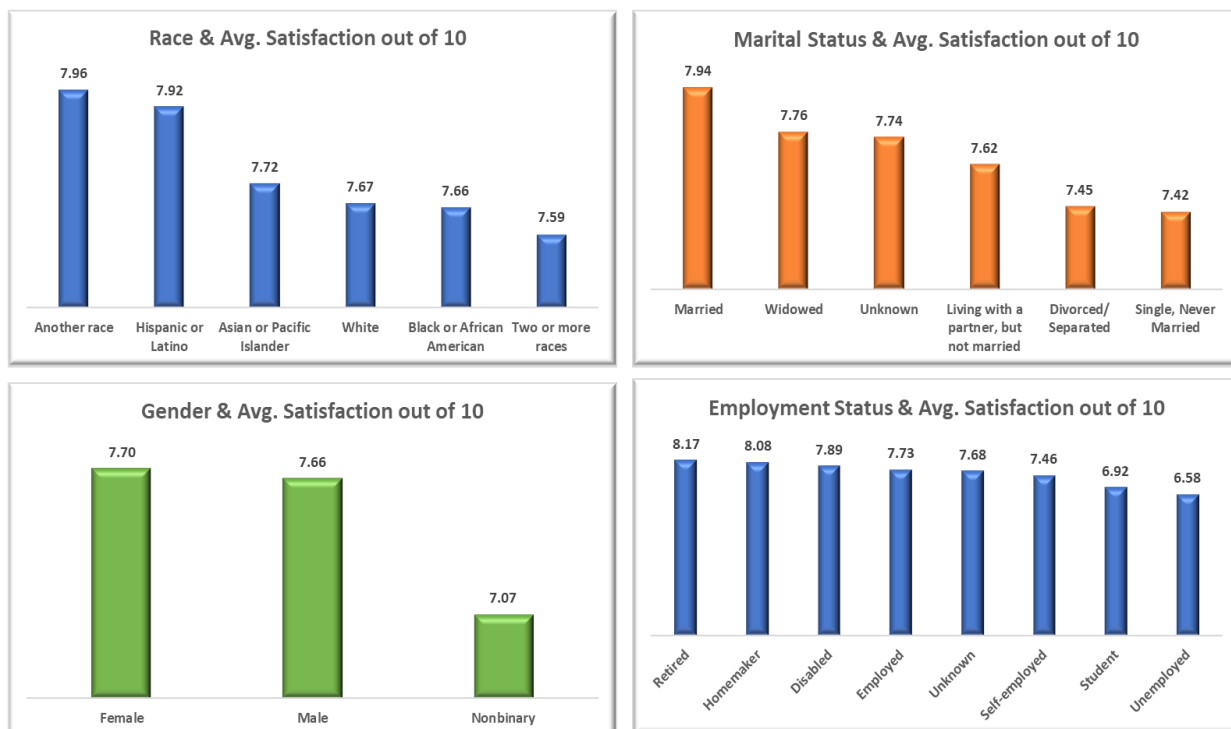
Figure 8: Somerville Satisfaction Charts 1
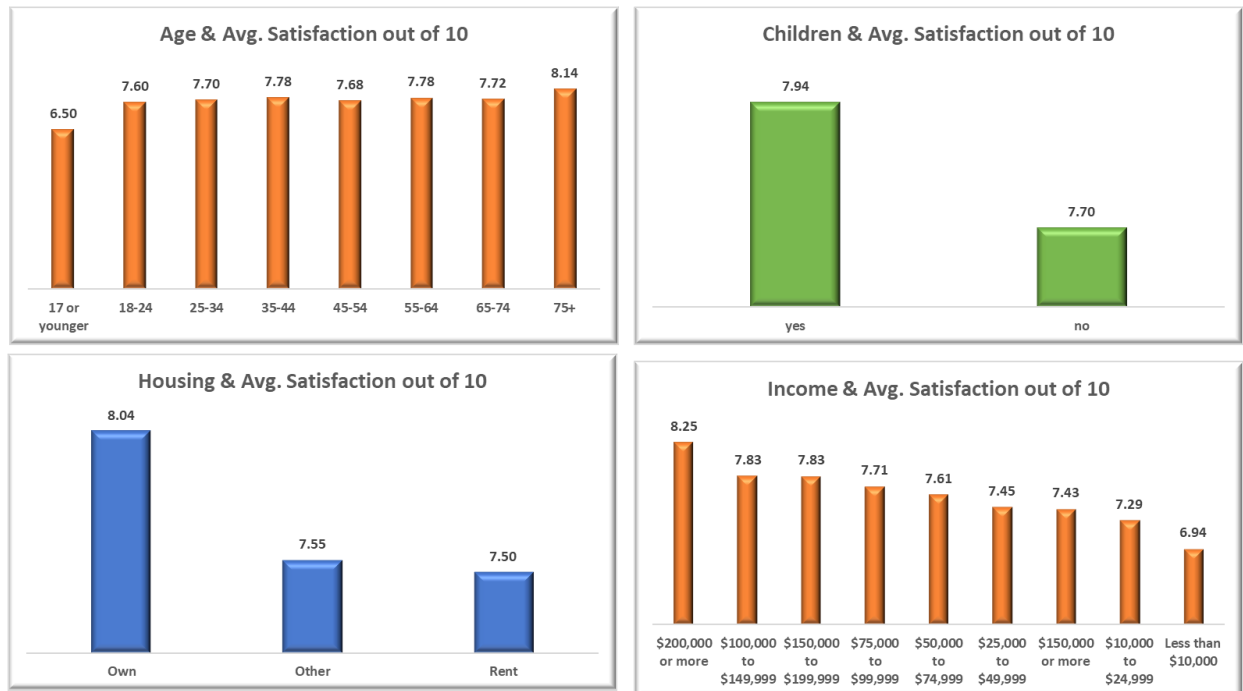
Figure 9: Somerville Satisfaction Charts 2

Line Charts (Figure 10, 11 & 12)

The line chart in Figure 10 illustrates that happiness ratings, life satisfaction ratings, and satisfaction living in Somerville ratings all exhibit a similar trend, indicating their interrelation.

The line chart in Figure 11 depicts the change in happiness levels over the years, categorized by housing status (renting or owning). This trend aligns with the overall happiness line chart, suggesting that factors other than housing status influenced happiness, as there was no significant change in happiness levels for both renters and homeowners.

The line chart in Figure 12 highlights a minor anomaly in the year 2017, where men reported higher happiness levels compared to women. This discrepancy warrants further investigation in more in-depth studies. Nevertheless, it is noteworthy that nonbinary individuals consistently reported the lowest happiness ratings across the observed years of 2017, 2019, and 2021.

Happiness, Overall Satisfaction & Satisfaction at Somerville
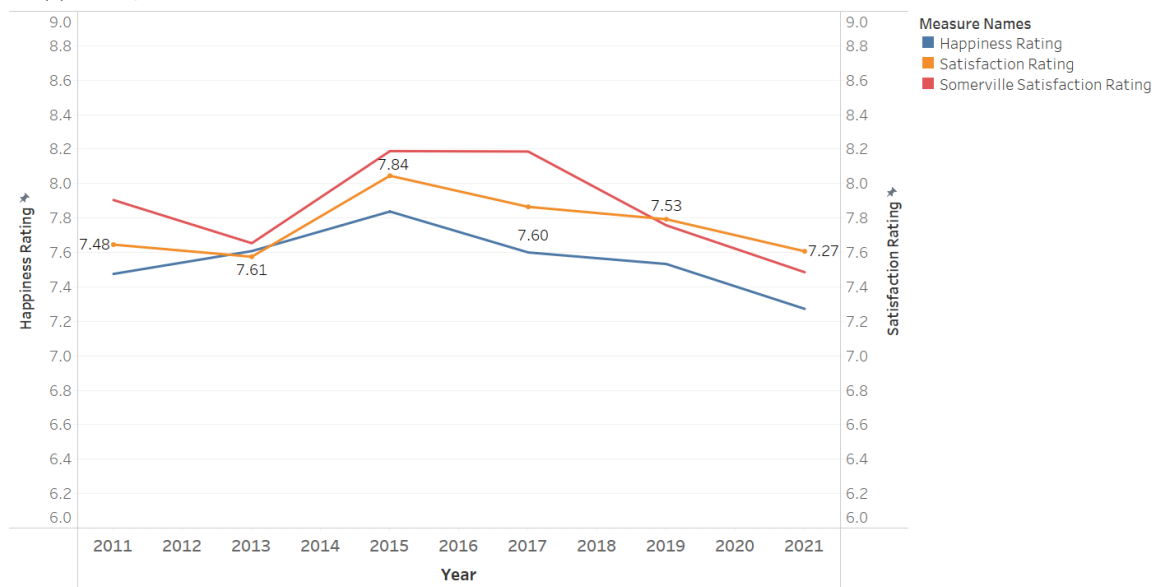
Figure 10: Somerville Happiness, Satisfaction and Satisfaction with Somerville rating with Time

Housing Trend
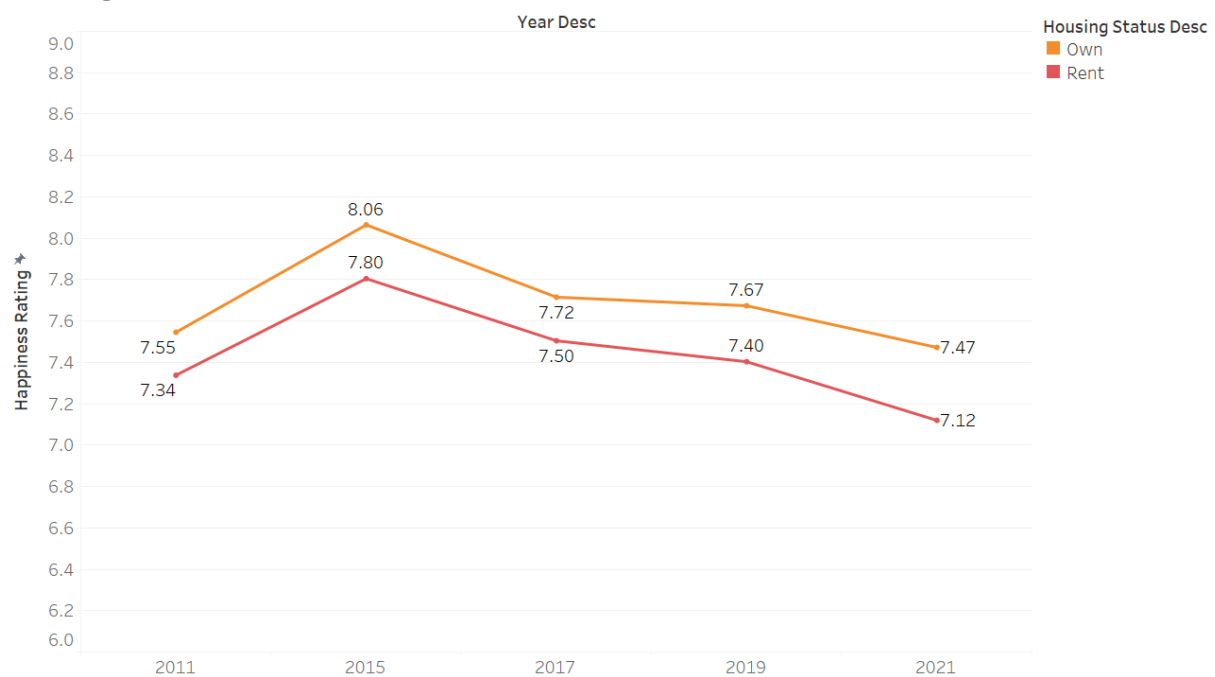
Figure 11: Year-wise Housing Trend

Line Graph-Gender
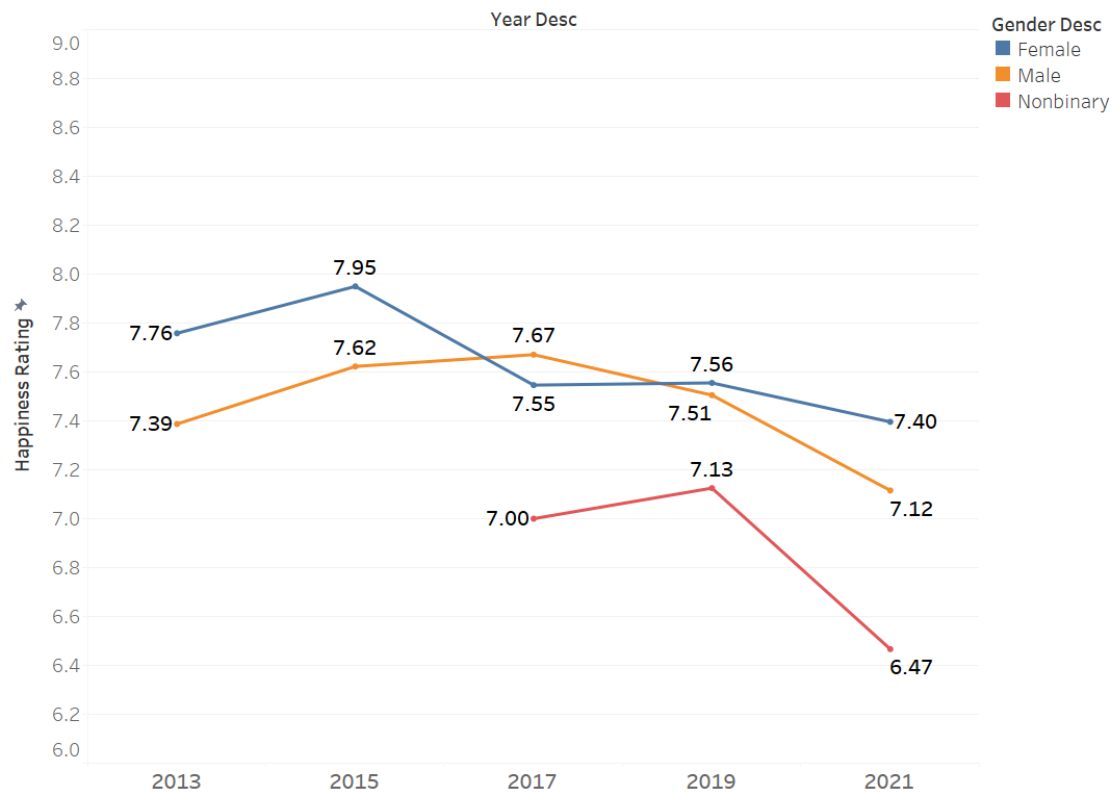
Figure 12: Year-wise Gender Trend

References

1. Dan Witters, "U.S. Depression Rates Reach New Highs"
   https://news.gallup.com/poll/505745/depression-rates-reach-new-highs.aspx
2. Merriam-Webster https://www.merriam-webster.com/dictionary/happiness
3. University of Oxford, "Bhutan's Gross National Happiness Index"
   https://ophi.org.uk/policy/bhutan-gnh-index/

APPENDIX: 1 (SQL CODE)

```
/*

Name:          Somerville DataWarehouse

Author:        Anant Kumar

Date:          12/17/2023

Version: 1.0

*/


-----------------------------------------------------------

--        Create Somerville Database Using Rightclick on Databses Tab

-----------------------------------------------------------


USE Somerville

GO


--------------------------------------------------

--        Creating & Populating SomervilleStagingTable

--------------------------------------------------


IF  EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[SomervilleStaging]') AND type in (N'U'))

DROP TABLE [dbo].[SomervilleStaging]

GO


CREATE TABLE SomervilleStaging

(

        Combined_ID INT,

        [Survey Year] VARCHAR(20),

        [How.happy.do.you.feel.right.now] INT,
```

[How.satisfied.are.you.with.your.life.in.general] INT,

[How.satisfied.are.you.with.Somerville.as.a.place.to.live] INT,

[In.general..how.similar.are.you.to.other.people.you.know] INT,

[When.making.decisions..are.you.more.likely.to.seek.advice.or.decide.for.yourself] INT,

[How.satisfied.are.you.with.your.neighborhood] INT,

[How.proud.are.you.to.be.a.Somerville.resident] INT,

[How.would.you.rate.the.following..The.availability.of.information.about.city.services] INT,

[How.would.you.rate.the.following..The.availability.of.affordable.housing] FLOAT,

[How.would.you.rate.the.following..The.cost.of.housing] INT,

[How.would.you.rate.the.following..The.overall.quality.of.public.schools] INT,

[How.would.you.rate.the.following..The.beauty.or.physical.setting.of.Somerville] INT,

[How.would.you.rate.the.following..The.effectiveness.of.the.local.police] INT,

[How.would.you.rate.the.following..Your.trust.in.the.local.police] INT,

[How.would.you.rate.the.following..The.maintenance.of.streets..sidewalks..and..squares] FLOAT,

[How.would.you.rate.the.following..The.maintenance.of.streets.and.sidewalks] INT,

[How.would.you.rate.the.following..The.availability.of.social.community.events] INT,

[How.safe.do.you.feel.walking.in.your.neighborhood.at.night] INT,

[How.satisfied.are.you.with.the.beauty.or.physical.setting.of.your.neighborhood] INT,

[How.satisfied.are.you.with.the.appearance.of.parks.in.your.neighborhood] INT,

[How.satisfied.are.you.with.the.appearance.of.parks.and.squares.in.your.neighborhood] INT,

[What.is.your.gender] VARCHAR (20),

[Age] VARCHAR (20),

[Marital.status] VARCHAR (50),

[Are.you.of.Hispanic..Latino..or.Spanish.origin] VARCHAR (10),

[What.is.your.race.or.ethnicity] VARCHAR (50),

[Do.you.have.children.age.18.or.younger.who.live.with.you] VARCHAR (10),

[Describe.your.housing.status.in.Somerville] VARCHAR (10),

[Do.you.plan.to.move.away.from.Somerville.in.the.next.two.years] VARCHAR (10),

[How.long.have.you.lived.here] VARCHAR (20),

[What.is.your.annual.household.income] VARCHAR (25),

[Are.you.a.student] VARCHAR (5),

[Ward] VARCHAR (25),

[Precinct] VARCHAR (50),

[How.anxious.did.you.feel.yesterday] INT,

[How.satisfied.are.you.with.the.quality.and.number.of.transportation.options.available.to.you] INT,

[Do.you.feel.the.City.is.headed.in.the.right.direction.or.is.it.on.the.wrong.track] VARCHAR (25),

[How.safe.do.you.feel.crossing.a.busy.street.in.Somerville] INT,

[How.convenient.is.it.for.you.to.get.where.you.want.to.go] INT,

[How.satisfied.are.you.with.the.condition.of.your.housing] INT,

[What.is.your.primary.mode.of.transportation] VARCHAR(50),

[Which.of.the.following.have.you.used.in.the.past.month.to.get.around] VARCHAR(50),

[Language] VARCHAR (25),

[survey_method] VARCHAR (50),

[language_spoken_english] INT,

[language_spoken_spanish] INT,

[language_spoken_portuguese] INT,

[language_spoken_chinese] INT,

[language_spoken_haitian_creole] INT,

[language_spoken_nepali] INT,

[language_spoken_other] INT,

[language_spoken_category] VARCHAR(100),

[race_ethnicity_asian_pacific_islander] INT,

[race_ethnicity_black] INT,

[race_ethnicity_white] INT,

[race_ethnicity_other] INT,

[race_ethnicity_prefernottosa] VARCHAR(50),

[age_mid] INT,

[tenure_mid] INT,

[household_income_mid] INT,

[somerville_median_income] INT,

[inflation_adjustment] FLOAT,

[disability_yn] VARCHAR (25),

[employment_status] VARCHAR (25),

```
        [zipcode] VARCHAR (25),

        [in_the_past_year_have_you_used_311_via_phone_online_etc]        VARCHAR (25),

        [in_the_past_year_did_you_attend_a_city_led_meeting] VARCHAR (25),

        [in_the_past_year_how_satisfied_were_you_with_your_ability_to_access_city_services] INT


)


BULK INSERT SomervilleStaging
FROM 'C:\Somerville\Somerville_Happiness_Survey_Responses_20231110.csv'
WITH    (FORMAT = 'csv',
                FIRSTROW = 2)


SELECT *
FROM SomervilleStaging



--Checking Data Quality

--Checking for NULL Values



----------SURVEY YEAR---------------------------------------------------------
            SELECT COUNT(*)
            FROM SomervilleStaging
            WHERE [Survey Year] is null


            --NO NULL VALUES IN SURVEY YEAR



----------[How.happy.do.you.feel.right.now]---------------------------------------
            SELECT COUNT(*)
            FROM SomervilleStaging
            WHERE [How.happy.do.you.feel.right.now] is null
```

```
--222 NULL VALUES IN [How.happy.do.you.feel.right.now] COLUMN

BEGIN TRAN

DELETE SomervilleStaging

WHERE [How.happy.do.you.feel.right.now] is null

COMMIT TRAN

--Nulls removed




----------[How.satisfied.are.you.with.your.life.in.general]------------------------

SELECT COUNT(*)

FROM SomervilleStaging

WHERE [How.satisfied.are.you.with.your.life.in.general] is null

--228 NULL VALUES IN [How.satisfied.are.you.with.your.life.in.general] COLUMN

BEGIN TRAN

DELETE SomervilleStaging

WHERE [How.satisfied.are.you.with.your.life.in.general] is null

COMMIT TRAN

--Nulls removed




----------[How.satisfied.are.you.with.Somerville.as.a.place.to.live]------------------

SELECT COUNT(*)

FROM SomervilleStaging

WHERE [How.satisfied.are.you.with.Somerville.as.a.place.to.live] is null

--6245 NULL VALUES in [How.satisfied.are.you.with.Somerville.as.a.place.to.live] COLUMN

--Since these are INT values andd I need to Avg them, keeping NULL values as is



----------[What.is.your.gender]---------------------------------------------------------

SELECT COUNT(*)

FROM SomervilleStaging

WHERE [What.is.your.gender] is null
```

```sql
--453 Null values in Gender Column

BEGIN TRAN

DELETE SomervilleStaging

WHERE [What.is.your.gender] is null

COMMIT TRAN

--Nulls removed
```

----------[Age]------------------------

```sql
SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Age] is null

--456 Null values in Age Column

BEGIN TRAN

DELETE SomervilleStaging

WHERE [Age] is null

COMMIT TRAN

--Nulls removed

SELECT [Survey Year]

FROM SomervilleStaging

WHERE [Age] IN ('18-21', '26-30', '22-25','31-40', '41-50', '51-60', '61+')
```
--2011 survey folowed two different age ranges however since there are 5299 such entries, keeping them as is.

----------[Marital.status]--

```sql
SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Marital.status] is null

--4343 Null values in Marriage Status column

BEGIN TRAN

UPDATE SomervilleStaging

SET [Marital.status] = 'Unknown'
```

WHERE [Marital.status] IS NULL

COMMIT TRAN

--Nulls Updated to 'Unknown'

SELECT DISTINCT [Marital.status]

FROM SomervilleStaging

BEGIN TRAN

UPDATE SomervilleStaging

SET [Marital.status] = 'Divorced/ Separated'

WHERE [Marital.status] = 'Divorced'

COMMIT TRAN

SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Marital.status] = 'Single, Never Married'

BEGIN TRAN

UPDATE SomervilleStaging

SET [Marital.status] = 'Single, Never Married'

WHERE [Marital.status] = 'Single (never married)'

COMMIT TRAN

----------[What.is.your.race.or.ethnicity]---

SELECT COUNT(*)

FROM SomervilleStaging

WHERE [What.is.your.race.or.ethnicity] is null

--639 Null values in [What.is.your.race.or.ethnicity] column

BEGIN TRAN

DELETE SomervilleStaging

WHERE [What.is.your.race.or.ethnicity] is null

COMMIT TRAN

--Nulls removed

```
SELECT DISTINCT [What.is.your.race.or.ethnicity]

FROM SomervilleStaging




----------[Do.you.have.children.age.18.or.younger.who.live.with.you]-----------------------------------------------------


SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Do.you.have.children.age.18.or.younger.who.live.with.you] IS NULL

--6269 Null values in [Do.you.have.children.age.18.or.younger.who.live.with.you] column

BEGIN TRAN

UPDATE SomervilleStaging

SET [Do.you.have.children.age.18.or.younger.who.live.with.you] = 'Unknown'

WHERE [Do.you.have.children.age.18.or.younger.who.live.with.you] IS NULL

COMMIT TRAN

--Nulls Updated to 'Unknown'




----------[Describe.your.housing.status.in.Somerville]------------------------------


SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Describe.your.housing.status.in.Somerville] is null

--6511 Null Values in [Describe.your.housing.status.in.Somerville] column

BEGIN TRAN

UPDATE SomervilleStaging

SET [Describe.your.housing.status.in.Somerville] = 'Unknown'

WHERE [Describe.your.housing.status.in.Somerville] IS NULL

COMMIT TRAN




---------------[What.is.your.annual.household.income]--------------------------------
```

```sql
SELECT COUNT(*)

FROM SomervilleStaging

WHERE [What.is.your.annual.household.income] is null

--1056 Null Values in [What.is.your.annual.household.income] column

BEGIN TRAN

DELETE SomervilleStaging

WHERE [What.is.your.annual.household.income]  is null

COMMIT TRAN

--Nulls removed

SELECT DISTINCT [What.is.your.annual.household.income]

FROM SomervilleStaging

WHERE [Survey Year] NOT IN ('2011')
```

--Apart from 2011, the income & age ranges are in sync. Keeping 2011 data for other analysis and we can filter out 2011 data for specific income and age analysis as and when needed

--------------------------[Precinct]--

```sql
SELECT COUNT(*)

FROM SomervilleStaging

WHERE [Precinct] is null

--9207 Null Values in [Precinct] Column

BEGIN TRAN

UPDATE SomervilleStaging

SET [Precinct] = 'Unknown'

WHERE [Precinct] IS NULL

COMMIT TRAN
```

--Nulls Updated to 'Unknown'

--------------------------[survey_method]--

```
            SELECT COUNT(*)

            FROM SomervilleStaging

            WHERE [survey_method] is null

            --No Null Values in [survey_method] Column



--------------------------[employment_status]--------------------------------------------------

            SELECT COUNT(*)

            FROM SomervilleStaging

            WHERE [employment_status] is null

            --10218 Null Values in [employment_status] column

            BEGIN TRAN

            UPDATE SomervilleStaging

            SET [employment_status] = 'Unknown'

            WHERE [employment_status] IS NULL

            COMMIT TRAN

            --Nulls Updated to 'Unknown'




--------------------------------------------------------------------------------------

-----------------------Creating Dimension Tables----------------------------------------

--------------------------------------------------------------------------------------


IF  EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[factSomerville]') AND type in (N'U'))

DROP TABLE [dbo].[factSomerville]


IF  EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimYear]') AND type in (N'U'))

DROP TABLE [dbo].[dimYear]


IF  EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimGender]') AND type in (N'U'))

DROP TABLE [dbo].[dimGender]


IF  EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimAge]') AND type in (N'U'))
```

DROP TABLE [dbo].[dimAge]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimMaritalStatus]') AND type in (N'U'))

DROP TABLE [dbo].[dimMaritalStatus]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimRace]') AND type in (N'U'))

DROP TABLE [dbo].[dimRace]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimChildren]') AND type in (N'U'))

DROP TABLE [dbo].[dimChildren]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimHousingStatus]') AND type in (N'U'))

DROP TABLE [dbo].[dimHousingStatus]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimIncome]') AND type in (N'U'))

DROP TABLE [dbo].[dimIncome]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimPrecinct]') AND type in (N'U'))

DROP TABLE [dbo].[dimPrecinct]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimsurveymethod]') AND type in (N'U'))

DROP TABLE [dbo].[dimsurveymethod]

IF EXISTS (SELECT * FROM sys.objects WHERE object_id = OBJECT_ID(N'[dbo].[dimEmploymentStatus]') AND type in (N'U'))

DROP TABLE [dbo].[dimEmploymentStatus]

--Creating dimYear

CREATE TABLE dimYear (

        YearID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimYear PRIMARY KEY CLUSTERED (YearID),

        YearDesc VARCHAR(10)

)

INSERT INTO dimYear

SELECT DISTINCT [Survey Year]

FROM SomervilleStaging

ORDER BY [Survey Year]

SELECT *

FROM dimYear

--Creating dimGender

CREATE TABLE dimGender (

        GenderID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimGender PRIMARY KEY CLUSTERED (GenderID),

        GenderDesc VARCHAR(20)

)

INSERT INTO dimGender

SELECT DISTINCT [What.is.your.gender]

FROM SomervilleStaging

ORDER BY [What.is.your.gender]

SELECT *

FROM dimGender

--Creating dimAge

```sql
CREATE TABLE dimAge (
        AgeID INT IDENTITY(1,1) NOT NULL
                CONSTRAINT PK_dimAge PRIMARY KEY CLUSTERED (AgeID),
        AgeDesc VARCHAR(20)
)

INSERT INTO dimAge
SELECT DISTINCT [Age]
FROM SomervilleStaging
ORDER BY [Age]

SELECT *
FROM dimAge
```

--Creating dimMaritalStatus

```sql
CREATE TABLE dimMaritalStatus (
        MaritalStatusID INT IDENTITY(1,1) NOT NULL
                CONSTRAINT PK_dimMaritalStatus PRIMARY KEY CLUSTERED (MaritalStatusID),
        MaritalStatusDesc VARCHAR(80)
)

INSERT INTO dimMaritalStatus
SELECT DISTINCT [Marital.status]
FROM SomervilleStaging
ORDER BY [Marital.status]

SELECT *
FROM dimMaritalStatus
```

--Creating dimRace

```sql
CREATE TABLE dimRace (
        RaceID INT IDENTITY(1,1) NOT NULL
                CONSTRAINT PK_dimRace PRIMARY KEY CLUSTERED (RaceID),
        RaceDesc VARCHAR(60)
)
```

```sql
INSERT INTO dimRace
SELECT DISTINCT [What.is.your.race.or.ethnicity]
FROM SomervilleStaging
ORDER BY [What.is.your.race.or.ethnicity]
```

```sql
SELECT *
FROM dimRace
```

--Creating dimChildren

```sql
CREATE TABLE dimChildren (
        ChildrenID INT IDENTITY(1,1) NOT NULL
                CONSTRAINT PK_dimChildren PRIMARY KEY CLUSTERED (ChildrenID),
        ChildrenDesc VARCHAR(10)
)
```

```sql
INSERT INTO dimChildren
SELECT DISTINCT [Do.you.have.children.age.18.or.younger.who.live.with.you]
FROM SomervilleStaging
```

```sql
SELECT *
FROM dimChildren
```

--Creating dimHousingStatus

--SELECT DISTINCT [Describe.your.housing.status.in.Somerville]

--FROM SomervilleStaging

CREATE TABLE dimHousingStatus (

        HousingStatusID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimHousingStatus PRIMARY KEY CLUSTERED (HousingStatusID),

        HousingStatusDesc VARCHAR(20)

)

INSERT INTO dimHousingStatus

SELECT DISTINCT [Describe.your.housing.status.in.Somerville]

FROM SomervilleStaging

ORDER BY [Describe.your.housing.status.in.Somerville]

SELECT *

FROM dimHousingStatus

--Creating dimIncome

SELECT DISTINCT [What.is.your.annual.household.income]

FROM SomervilleStaging

CREATE TABLE dimIncome (

        IncomeID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimIncome PRIMARY KEY CLUSTERED (IncomeID),

```
        IncomeDesc VARCHAR(20)
)


INSERT INTO dimIncome

SELECT DISTINCT [What.is.your.annual.household.income]

FROM SomervilleStaging

ORDER BY [What.is.your.annual.household.income]


SELECT *

FROM dimIncome


--Creating dimPrecinct


--SELECT DISTINCT Precinct

--FROM SomervilleStaging


CREATE TABLE dimPrecinct (

        PrecinctID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimPrecinct PRIMARY KEY CLUSTERED (PrecinctID),

        PrecinctDesc VARCHAR(20)
)


INSERT INTO dimPrecinct

SELECT DISTINCT Precinct

FROM SomervilleStaging

ORDER BY Precinct


SELECT *

FROM dimPrecinct


--Creating dimsurveymethod

--SELECT DISTINCT [survey_method]
```

```sql
--FROM SomervilleStaging


CREATE TABLE dimsurveymethod (

        surveymethodID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimsurveymethod PRIMARY KEY CLUSTERED (surveymethodID),

        surveymethodDesc VARCHAR(30)

)


INSERT INTO dimsurveymethod

SELECT DISTINCT [survey_method]

FROM SomervilleStaging

ORDER BY [survey_method]


SELECT *

FROM dimsurveymethod




--Creating dimEmploymentStatus


--SELECT DISTINCT [employment_status]

--FROM SomervilleStaging


CREATE TABLE dimEmploymentStatus (

        EmploymentStatusID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_dimEmploymentStatus PRIMARY KEY CLUSTERED (EmploymentStatusID),

        EmploymentStatusDesc VARCHAR(30)

)


INSERT INTO dimEmploymentStatus

SELECT DISTINCT [employment_status]

FROM SomervilleStaging
```

ORDER BY [employment_status]

SELECT *

FROM dimEmploymentStatus

```
------------------------------------------------------------------------------

----------------Creating factSomerville Table----------------------------------

------------------------------------------------------------------------------



CREATE TABLE factSomerville

(

        SomervilleID INT IDENTITY(1,1) NOT NULL

                CONSTRAINT PK_factSomerville PRIMARY KEY CLUSTERED (SomervilleID),


        YearID INT NOT NULL

                CONSTRAINT FK_dimYear FOREIGN KEY (YearID)

                REFERENCES dimYear(YearID),


        GenderID INT NOT NULL

                CONSTRAINT FK_dimGender FOREIGN KEY (GenderID)

                REFERENCES dimGender(GenderID),


        AgeID INT NOT NULL

                CONSTRAINT FK_dimAge FOREIGN KEY (AgeID)

                REFERENCES dimAge(AgeID),


        MaritalStatusID INT NOT NULL

                CONSTRAINT FK_dimMaritalStatus FOREIGN KEY (MaritalStatusID)

                REFERENCES dimMaritalStatus (MaritalStatusID),
```

RaceID INT NOT NULL

    CONSTRAINT FK_dimRace FOREIGN KEY (RaceID)

    REFERENCES dimRace (RaceID),

ChildrenID INT NOT NULL

    CONSTRAINT FK_dimChildren FOREIGN KEY(ChildrenID)

    REFERENCES dimChildren (ChildrenID),

HousingStatusID INT NOT NULL

    CONSTRAINT FK_dimHousingStatus FOREIGN KEY (HousingStatusID)

    REFERENCES dimHousingStatus (HousingStatusID),

IncomeID INT NOT NULL

    CONSTRAINT FK_dimIncome FOREIGN KEY (IncomeID)

    REFERENCES dimIncome (IncomeID),

PrecinctID INT NOT NULL

    CONSTRAINT FK_dimPrecinct FOREIGN KEY (PrecinctID)

    REFERENCES dimPrecinct (PrecinctID),

surveymethodID INT NOT NULL

    CONSTRAINT FK_dimsurveymethod FOREIGN KEY (surveymethodID)

    REFERENCES dimsurveymethod (surveymethodID),

EmploymentStatusID INT NOT NULL

    CONSTRAINT FK_dimEmploymentStatus FOREIGN KEY (EmploymentStatusID)

    REFERENCES dimEmploymentStatus (EmploymentStatusID),

HappinessRating FLOAT,

SatisfactionRating FLOAT,

SomervilleSatisfactionRating FLOAT

)

```
INSERT INTO factSomerville

SELECT


--SELECT 'WHEN [Survey Year] = ''' + YearDesc + ''' then ' +  CAST(YearID AS varchar(2))

--FROM dimYear


        [Survey Year] =


        CASE

                WHEN [Survey Year] = '2011' then 1

                WHEN [Survey Year] = '2013' then 2

                WHEN [Survey Year] = '2015' then 3

                WHEN [Survey Year] = '2017' then 4

                WHEN [Survey Year] = '2019' then 5

                WHEN [Survey Year] = '2021' then 6

        END,


--SELECT 'WHEN [What.is.your.gender] = ''' + GenderDesc + ''' then ' +  CAST(GenderID AS varchar(2))

--FROM dimGender


        [What.is.your.gender]=


        CASE

                WHEN [What.is.your.gender] = 'Female' then 1

                WHEN [What.is.your.gender] = 'Male' then 2

                WHEN [What.is.your.gender] = 'Multiple people' then 3

                WHEN [What.is.your.gender] = 'Nonbinary' then 4


        END,
```

```
--SELECT 'WHEN [Age] = ''' + AgeDesc + ''' then ' +  CAST(AgeID AS varchar(2))

--FROM dimAge

        [Age]=


        CASE

                WHEN [Age] = '17 or younger' then 1

                WHEN [Age] = '18-21' then 2

                WHEN [Age] = '18-24' then 3

                WHEN [Age] = '22-25' then 4

                WHEN [Age] = '25-34' then 5

                WHEN [Age] = '26-30' then 6

                WHEN [Age] = '31-40' then 7

                WHEN [Age] = '35-44' then 8

                WHEN [Age] = '41-50' then 9

                WHEN [Age] = '45-54' then 10

                WHEN [Age] = '51-60' then 11

                WHEN [Age] = '55-64' then 12

                WHEN [Age] = '61+' then 13

                WHEN [Age] = '65-74' then 14

                WHEN [Age] = '75+' then 15

        END,




--SELECT 'WHEN [Marital.status] = ''' + MaritalStatusDesc + ''' then ' +  CAST(MaritalStatusID AS varchar(2))

--FROM dimMaritalStatus


        [Marital.status]=


        CASE

                WHEN [Marital.status] = 'Divorced/ Separated' then 1

                WHEN [Marital.status] = 'Living with a partner, but not married' then 2
```

```
                WHEN [Marital.status] = 'Married' then 3

                WHEN [Marital.status] = 'Single, Never Married' then 4

                WHEN [Marital.status] = 'Unknown' then 5

                WHEN [Marital.status] = 'Widowed' then 6

        END,


--SELECT 'WHEN [What.is.your.race.or.ethnicity] = ''' + RaceDesc + ''' then ' +  CAST(RaceID AS varchar(2))

--FROM dimRace


        [What.is.your.race.or.ethnicity] =


        CASE

                WHEN [What.is.your.race.or.ethnicity] = 'Another race' then 1

                WHEN [What.is.your.race.or.ethnicity] = 'Asian or Pacific Islander' then 2

                WHEN [What.is.your.race.or.ethnicity] = 'Black or African American' then 3

                WHEN [What.is.your.race.or.ethnicity] = 'Hispanic or Latino' then 4

                WHEN [What.is.your.race.or.ethnicity] = 'Two or more races' then 5

                WHEN [What.is.your.race.or.ethnicity] = 'White' then 6

        END,


--SELECT 'WHEN [Do.you.have.children.age.18.or.younger.who.live.with.you] = ''' + ChildrenDesc + ''' then ' +
CAST(ChildrenID AS varchar(2))

--FROM dimChildren


        [Do.you.have.children.age.18.or.younger.who.live.with.you] =


        CASE

                WHEN [Do.you.have.children.age.18.or.younger.who.live.with.you] = 'yes' then 1

                WHEN [Do.you.have.children.age.18.or.younger.who.live.with.you] = 'Unknown' then 2

                WHEN [Do.you.have.children.age.18.or.younger.who.live.with.you] = 'no' then 3

        END,
```

```
--SELECT 'WHEN [Describe.your.housing.status.in.Somerville] = ''' + HousingStatusDesc + ''' then ' +
CAST(HousingStatusID AS varchar(2))
--FROM dimHousingStatus


        [Describe.your.housing.status.in.Somerville] =

        CASE

                        WHEN [Describe.your.housing.status.in.Somerville] = 'Other' then 1

                        WHEN [Describe.your.housing.status.in.Somerville] = 'Own' then 2

                        WHEN [Describe.your.housing.status.in.Somerville] = 'Rent' then 3

                        WHEN [Describe.your.housing.status.in.Somerville] = 'Unknown' then 4

        END,




--SELECT 'WHEN [What.is.your.annual.household.income] = ''' + IncomeDesc + ''' then ' +  CAST(IncomeID AS
varchar(2))
--FROM dimIncome


        [What.is.your.annual.household.income] =


        CASE

                        WHEN [What.is.your.annual.household.income] = '$10,000 to $19,999' then 1

                        WHEN [What.is.your.annual.household.income] = '$10,000 to $24,999' then 2

                        WHEN [What.is.your.annual.household.income] = '$100,000 or more' then 3

                        WHEN [What.is.your.annual.household.income] = '$100,000 to $119,999' then 4

                        WHEN [What.is.your.annual.household.income] = '$100,000 to $149,999' then 5

                        WHEN [What.is.your.annual.household.income] = '$120,000 to $139,999' then 6

                        WHEN [What.is.your.annual.household.income] = '$140,000 to $159,999' then 7

                        WHEN [What.is.your.annual.household.income] = '$150,000 or more' then 8

                        WHEN [What.is.your.annual.household.income] = '$150,000 to $199,999' then 9

                        WHEN [What.is.your.annual.household.income] = '$160,000 to $179,999' then 10

                        WHEN [What.is.your.annual.household.income] = '$180,000 to $199,999' then 11
```

```
                    WHEN [What.is.your.annual.household.income] = '$20,000 to $39,999' then 12

                    WHEN [What.is.your.annual.household.income] = '$200,000 or more' then 13

                    WHEN [What.is.your.annual.household.income] = '$25,000 to $49,999' then 14

                    WHEN [What.is.your.annual.household.income] = '$40,000 to $59,999' then 15

                    WHEN [What.is.your.annual.household.income] = '$50,000 to $74,999' then 16

                    WHEN [What.is.your.annual.household.income] = '$60,000 to $79,999' then 17

                    WHEN [What.is.your.annual.household.income] = '$75,000 to $99,999' then 18

                    WHEN [What.is.your.annual.household.income] = '$80,000 to $99,999' then 19

                    WHEN [What.is.your.annual.household.income] = 'Less than $10,000' then 20

                    WHEN [What.is.your.annual.household.income] = 'Less than $20,000' then 21

        END,



--SELECT 'WHEN [Precinct] = ''' + PrecinctDesc + ''' then ' +  CAST(PrecinctID AS varchar(2))

--FROM dimPrecinct


        [Precinct] =


        CASE


                WHEN [Precinct] = '1/1/2023 0:00' then 1

                WHEN [Precinct] = '1/2/2023 0:00' then 2

                WHEN [Precinct] = '1/3/2023 0:00' then 3

                WHEN [Precinct] = '1/4/2023 0:00' then 4

                WHEN [Precinct] = '2-1A' then 5

                WHEN [Precinct] = '2/2/2023 0:00' then 6

                WHEN [Precinct] = '2/3/2023 0:00' then 7

                WHEN [Precinct] = 'Unknown' then 8

        END,



--SELECT 'WHEN [survey_method] = ''' + surveymethodDesc + ''' then ' +  CAST(surveymethodID AS varchar(2))
```

```sql
--FROM dimsurveymethod


        [survey_method] =


        CASE

                WHEN [survey_method] = 'Email' then 1

                WHEN [survey_method] = 'Facebook 18-24 year olds' then 2

                WHEN [survey_method] = 'Mail' then 3

                WHEN [survey_method] = 'Phone' then 4

        END,


--SELECT 'WHEN [employment_status] = ''' + EmploymentStatusDesc + ''' then ' +  CAST(EmploymentStatusID AS varchar(2))
--FROM dimEmploymentStatus


        [employment_status] =

        CASE

                WHEN [employment_status] = 'Disabled' then 1

                WHEN [employment_status] = 'Employed' then 2

                WHEN [employment_status] = 'Homemaker' then 3

                WHEN [employment_status] = 'Retired' then 4

                WHEN [employment_status] = 'Self-employed' then 5

                WHEN [employment_status] = 'Student' then 6

                WHEN [employment_status] = 'Unemployed' then 7

                WHEN [employment_status] = 'Unknown' then 8

        END,

        [How.happy.do.you.feel.right.now],

        [How.satisfied.are.you.with.your.life.in.general],

        [How.satisfied.are.you.with.Somerville.as.a.place.to.live]


FROM SomervilleStaging
```

```
SELECT *

FROM factSomerville
```

--In Which Year people were the Happiest on Average?

```
SELECT Y.YearDesc, ROUND(AVG(HappinessRating),2)  AS [Average_Happiness_Rating_Outof10]

FROM  dimYear Y INNER JOIN factSomerville FS ON Y.YearID = FS.YearID

GROUP BY Y.YearDesc

ORDER BY AVG(HappinessRating) DESC
```

--are people living with children more happy?

```
SELECT C.ChildrenDesc, ROUND(AVG(HappinessRating),2)  AS [Average_Happiness_Rating_Outof10]

FROM  dimChildren C INNER JOIN factSomerville FS ON C.ChildrenID = FS.ChildrenID

GROUP BY C.ChildrenDesc

ORDER BY AVG(HappinessRating) DESC
```

--are people with their own houses generally more happy?

```
SELECT H.HousingStatusDesc, ROUND(AVG(HappinessRating),2)  AS [Average_Happiness_Rating_Outof10]

FROM  dimHousingStatus H INNER JOIN factSomerville FS ON H.HousingStatusID = FS.HousingStatusID

GROUP BY H.HousingStatusDesc

ORDER BY AVG(HappinessRating) DESC
```

--are people with jobs generally more happy?

```
SELECT E.EmploymentStatusDesc, ROUND(AVG(HappinessRating),2)  AS [Average_Happiness_Rating_Outof10]

FROM  dimEmploymentStatus E INNER JOIN factSomerville FS ON E.EmploymentStatusID = FS.EmploymentStatusID

GROUP BY E.EmploymentStatusDesc

ORDER BY AVG(HappinessRating) DESC
```

--In Which Income Group people are happiest on average?

SELECT I.IncomeDesc, ROUND(AVG(HappinessRating),2) AS [Average_Happiness_Rating_Outof10]

FROM dimIncome I INNER JOIN factSomerville FS ON I.IncomeID = FS.IncomeID

INNER JOIN dimYear Y ON Y.YearID = FS.YearID

WHERE YearDesc <> '2011'

GROUP BY I.IncomeDesc

ORDER BY AVG(HappinessRating) DESC

-----------------Creating View----------------------------------

----

--CREATE VIEW [Happiness by Age Group] AS

--SELECT A.AgeDesc, ROUND(AVG(HappinessRating),2) AS [Average_Happiness_Rating_Outof10]

--FROM dimAge A INNER JOIN factSomerville FS ON A.AgeID = FS.AgeID

--   INNER JOIN dimYear Y ON Y.YearID = FS.YearID

--WHERE YearDesc <> '2011'

--GROUP BY A.AgeDesc

--SELECT *

--FROM [Happiness by Age Group]