Name – Davender Singh

Student ID – 125700211

Course – DBW624V1A

Professor – Les King

1. **Baby Names:**

CREATE TABLE Staging\_BabyNames (

REF\_DATE VARCHAR(50) NOT NULL,

GEO VARCHAR(50) NOT NULL,

DGUID VARCHAR(50),

Sex VARCHAR(50) NOT NULL CHECK (Sex IN ('Male', 'Female')), -- Check constraint for gender

Birth\_Name VARCHAR(100) NOT NULL,

Indicator VARCHAR(50) NOT NULL,

UOM VARCHAR(50),

UOM\_ID INT,

SCALAR\_FACTOR VARCHAR(50),

SCALAR\_ID INT,

VECTOR VARCHAR(50),

COORDINATE VARCHAR(50),

VALUE VARCHAR(50) NOT NULL CHECK (ISNUMERIC(VALUE) = 1 AND CAST(VALUE AS INT) > 0), -- Check for positive numeric value

STATUS VARCHAR(50),

SYMBOL VARCHAR(50),

TERMINATED INT,

DECIMALS VARCHAR(50)

);

BULK INSERT Staging\_BabyNames

FROM 'C:\Program Files\Microsoft SQL Server\MSSQL16.SQLEXPRESS\MSSQL\BabyNames.csv'

WITH (

FIELDTERMINATOR = ',',

ROWTERMINATOR = '\n',

FIRSTROW = 2

)

DELETE FROM Staging\_BabyNames

WHERE ISNUMERIC(VALUE) = 0 OR VALUE IS NULL OR Indicator <> 'Frequency';

CREATE TABLE BabyNames (

Year INT NOT NULL,

Name NVARCHAR(255) NOT NULL,

Gender NVARCHAR(50) NOT NULL CHECK (Gender IN ('Male', 'Female')), -- Check constraint for gender

Frequency INT NOT NULL CHECK (Frequency > 0), -- Check for positive frequency

PRIMARY KEY (Year, Name, Gender)

);

--Move cleansed data from Staging\_BabyNames to BabyNames

INSERT INTO BabyNames (Year, Name, Gender, Frequency)

SELECT

CAST(LEFT(REF\_DATE, 4) AS INT), -- Extract year from REF\_DATE

Birth\_Name,

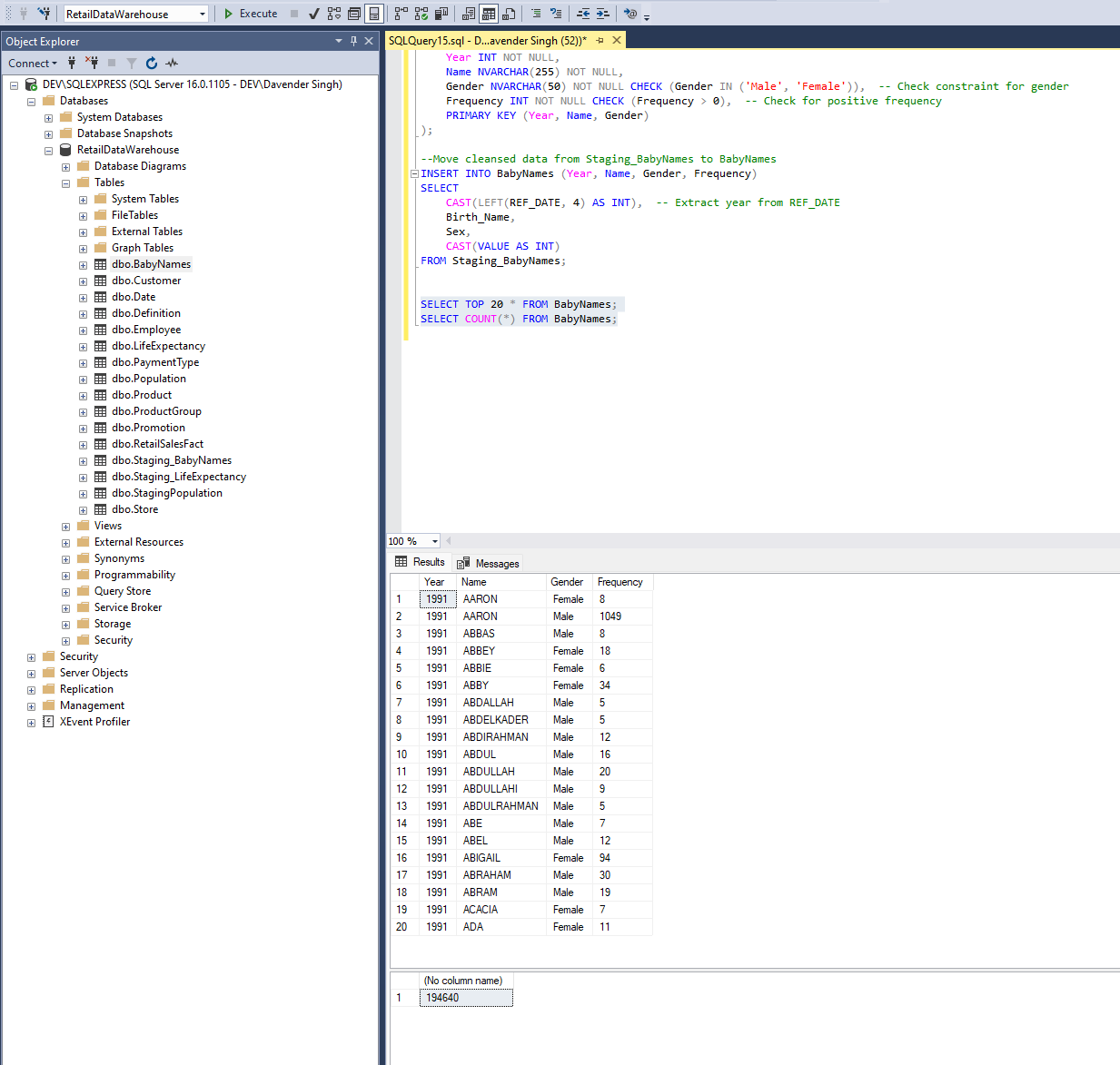
Sex,

CAST(VALUE AS INT)

FROM Staging\_BabyNames;

SELECT TOP 20 \* FROM BabyNames;

SELECT COUNT(\*) FROM BabyNames;



1. **Population:**

CREATE TABLE StagingPopulation (

REF\_DATE VARCHAR(50) NOT NULL,

GEO VARCHAR(100) NOT NULL,

DGUID VARCHAR(50) DEFAULT 'N/A',

Sex VARCHAR(50) NOT NULL CHECK (Sex IN ('Males', 'Females', 'Both sexes')),

Age\_Group VARCHAR(100) NOT NULL,

UOM VARCHAR(50) DEFAULT 'Persons' NOT NULL,

UOM\_ID INT DEFAULT 1 NOT NULL,

SCALAR\_FACTOR VARCHAR(50) DEFAULT 'units',

SCALAR\_ID INT DEFAULT 0,

VECTOR VARCHAR(50) DEFAULT 'N/A',

COORDINATE VARCHAR(50) DEFAULT 'N/A',

VALUE FLOAT,

STATUS VARCHAR(50) DEFAULT 'Active',

SYMBOL INT DEFAULT 0,

TERMINATED VARCHAR(50) DEFAULT 'No',

DECIMALS INT DEFAULT 2 NOT NULL

);

BULK INSERT StagingPopulation

FROM 'C:\Program Files\Microsoft SQL Server\MSSQL16.SQLEXPRESS\MSSQL\Population.csv' -- Replace this with the path to your file

WITH

(

FIELDTERMINATOR = ',', -- CSV field delimiter

ROWTERMINATOR = '\n',

FIRSTROW = 2 -- assuming first row is header

);

DELETE FROM StagingPopulation

WHERE

VALUE IS NULL

OR GEO = 'Canada'

OR (Sex <> 'Males' AND Sex <> 'Females')

OR Age\_Group <> 'All ages'

OR UOM <> 'Persons';

CREATE TABLE Population (

Year INT NOT NULL,

Province VARCHAR(100) NOT NULL,

Population INT NOT NULL DEFAULT 0,

Gender VARCHAR(50) NOT NULL CHECK (Gender IN ('Males', 'Females')),

CONSTRAINT PK\_Population PRIMARY KEY (Year, Province, Gender)

);

INSERT INTO Population (Year, Province, Population, Gender)

SELECT

CAST(REF\_DATE AS INT) AS Year,

GEO AS Province,

VALUE AS Population,

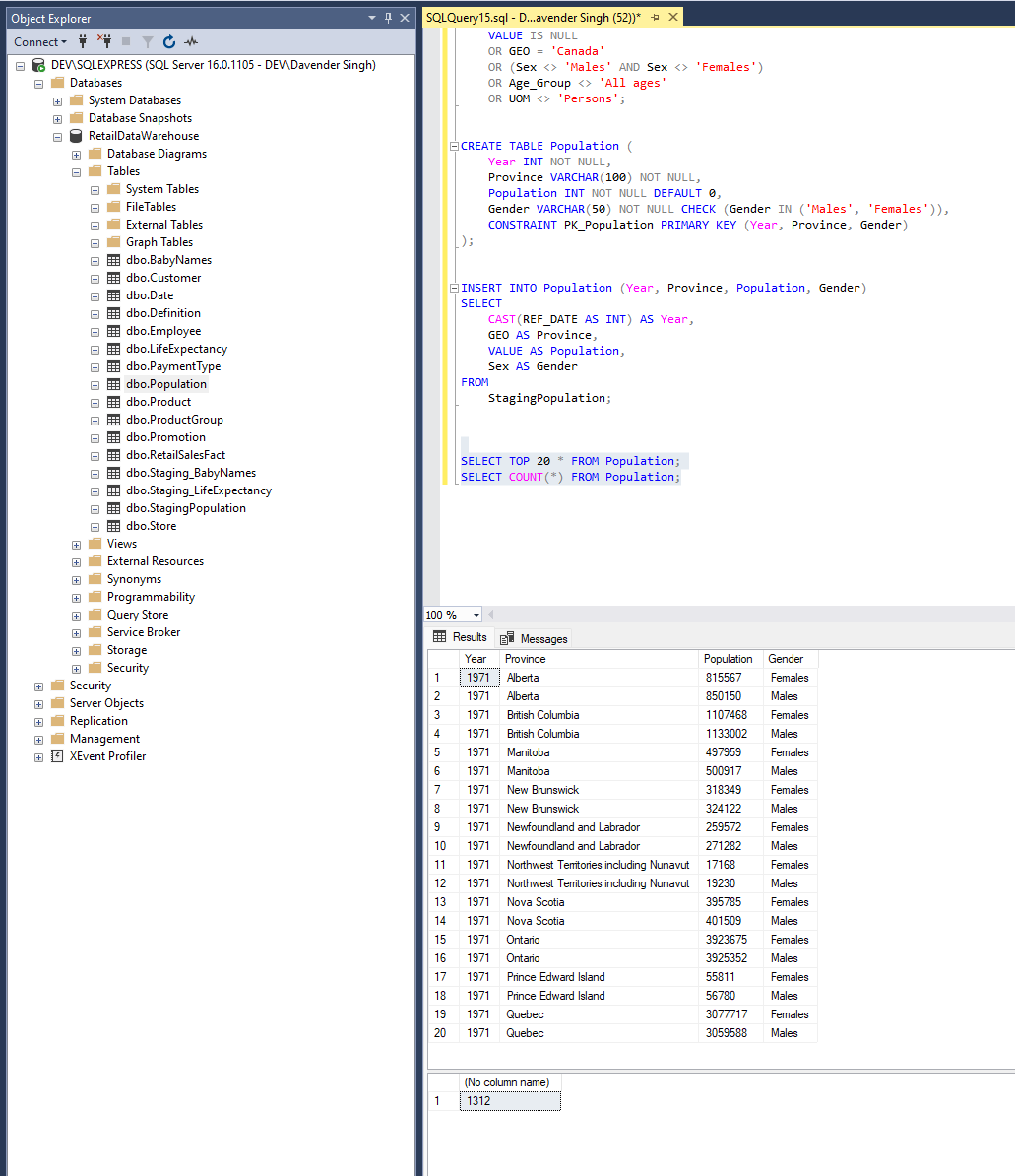
Sex AS Gender

FROM

StagingPopulation;

SELECT TOP 20 \* FROM Population;

SELECT COUNT(\*) FROM Population;



1. **Life Expectancy:**

CREATE TABLE Staging\_LifeExpectancy

(

REF\_DATE VARCHAR(255) NOT NULL,

GEO VARCHAR(255) NOT NULL,

DGUID VARCHAR(255),

Age\_Group VARCHAR(255) NOT NULL,

Sex VARCHAR(50) NOT NULL,

Element VARCHAR(255) NOT NULL,

UOM VARCHAR(255) NOT NULL,

UOM\_ID INT NOT NULL,

SCALAR\_FACTOR VARCHAR(255) NOT NULL,

SCALAR\_ID INT NOT NULL,

VECTOR VARCHAR(255) NOT NULL,

COORDINATE VARCHAR(255) NOT NULL,

VALUE FLOAT,

STATUS INT,

SYMBOL INT,

TERMINATED INT,

DECIMALS INT NOT NULL

);

BULK INSERT Staging\_LifeExpectancy

FROM 'C:\Program Files\Microsoft SQL Server\MSSQL16.SQLEXPRESS\MSSQL\LifeExpectancy.csv'

WITH (

FIELDTERMINATOR = ',',

ROWTERMINATOR = '\n',

FIRSTROW = 2

)

DELETE FROM Staging\_LifeExpectancy

WHERE

VALUE IS NULL

OR GEO = 'Canada'

OR (Sex <> 'Males' AND Sex <> 'Females')

OR Age\_Group <> '0 years'

OR Element <> 'Life expectancy (in years) at age x (ex)';

CREATE TABLE LifeExpectancy (

Birth\_Year INT NOT NULL,

Province VARCHAR(255) NOT NULL,

Gender VARCHAR(50) NOT NULL CHECK (Gender IN ('Males', 'Females')),

Life\_Expectancy\_Years FLOAT NOT NULL

);

INSERT INTO LifeExpectancy (Birth\_Year, Province, Gender, Life\_Expectancy\_Years)

SELECT

REF\_DATE AS Birth\_Year,

GEO AS Province,

Sex AS Gender,

VALUE AS Life\_Expectancy\_Years

FROM

Staging\_LifeExpectancy;

SELECT TOP 20 \* FROM LifeExpectancy;

SELECT COUNT(\*) FROM LifeExpectancy;

