APPENDIX

use project

------# data pre processing-----

-------1. Data Exploration and Cleaning

-- Countries in 'incidence-of-tuberculosis-sdgs' not in 'tuberculosis-deaths-by-age'

SELECT [Entity] AS CountryName FROM [dbo].[ incidence-of-tuberculosis-sdgs]

EXCEPT

SELECT [Entity] FROM [dbo].[tuberculosis-deaths-by-age];

-- Countries in 'tuberculosis-deaths-by-age' not in 'incidence-of-tuberculosis-sdgs'

SELECT [Entity] AS CountryName FROM [dbo].[tuberculosis-deaths-by-age]

EXCEPT

SELECT [Entity] FROM [dbo].[ incidence-of-tuberculosis-sdgs]

SELECT [Entity] AS CountryName FROM [dbo].[tuberculosis-case-detection-rate]

EXCEPT

SELECT [Entity] FROM [dbo].[4- tuberculosis-treatment-success-rate-by-type]

SELECT [Entity] AS CountryName FROM [dbo].[4- tuberculosis-treatment-success-rate-by-type]

EXCEPT

SELECT [Entity] FROM [dbo].[tuberculosis-case-detection-rate]

------------- joining two tables-------

SELECT a.[Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis],b.\* into #newtable

FROM [dbo].[ incidence-of-tuberculosis-sdgs] AS a

JOIN [dbo].[tuberculosis-deaths-by-age] AS b

ON a.[Entity] = b.[Entity]

ORDER BY a.[Entity]

-------------203 same countries-----

------139380-----

SELECT a.\*,b.[Indicator\_Treatment\_success\_rate\_new\_TB\_cases],b.[Indicator\_Treatment\_success\_rate\_for\_patients\_treated\_for\_MDR\_TB],b.[Indicator\_Treatment\_success\_rate\_XDR\_TB\_cases]

into #secondtable

FROM #newtable AS a

INNER JOIN [dbo].[4- tuberculosis-treatment-success-rate-by-type] AS b

ON a.[Entity] = b.[Entity]

ORDER BY a.[Entity]

----------------2911890-------------

select \* from #secondtable

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

SELECT a.\*,b.[Case\_detection\_rate\_all\_forms]

into dbo.finalTB

FROM #secondtable AS a

INNER JOIN [dbo].[tuberculosis-case-detection-rate] AS b

ON a.[Entity] = b.[Entity]

ORDER BY a.[Entity]

-------65251170------

delete from [dbo].[finalTB] where [Indicator\_Treatment\_success\_rate\_new\_TB\_cases] is Null

delete from [dbo].[finalTB] where [Indicator\_Treatment\_success\_rate\_for\_patients\_treated\_for\_MDR\_TB] is Null

delete from [dbo].[finalTB] where [Indicator\_Treatment\_success\_rate\_XDR\_TB\_cases] is Null

delete from [dbo].[finalTB] where [Case\_detection\_rate\_all\_forms] is Null

select distinct \* into dbo.TB from [dbo].[finalTB]

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

--------1.2 Descriptive Statistics and Initial Observations-----------

-- Summary statistics for tuberculosis incidence

SELECT

MIN([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS min\_incidence,

MAX([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS max\_incidence,

AVG([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS avg\_incidence,

STDEV([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS stddev\_incidence

FROM [dbo].[TB]

--------min\_incidence : 2.1 ,max\_incidence : 1590 , avg\_incidence : 263.0318866162682 , stddev\_incidence : 273.00749104379094

--------Analyze the trends of TB incidence rates over the years to understand how TB is evolving globally or in specific countries

SELECT [Year], [Entity], AVG([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS avg\_incidence\_rate

FROM [dbo].[finalTB]

GROUP BY [Year],[Entity]

ORDER BY year, avg\_incidence\_rate DESC

SELECT [Year], [Entity], AVG([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) AS avg\_incidence\_rate

FROM [dbo].[finalTB]

GROUP BY [Year],[Entity]

ORDER BY year, avg\_incidence\_rate

--------Eswatini -- 971.5652173913044

-------- Age-wise TB Mortality Analysis

SELECT

year,

CONVERT(FLOAT, Deaths\_Age\_70\_plus) / CONVERT(FLOAT, Total\_Deaths) \* 100 AS Pct\_Deaths\_Age\_70\_plus,

CONVERT(FLOAT, Deaths\_Age\_50\_to\_69) / CONVERT(FLOAT, Total\_Deaths) \* 100 AS Pct\_Deaths\_Age\_50\_to\_69,

CONVERT(FLOAT, Deaths\_Age\_15\_to\_49) / CONVERT(FLOAT, Total\_Deaths) \* 100 AS Pct\_Deaths\_Age\_15\_to\_49,

CONVERT(FLOAT, Deaths\_Age\_5\_to\_14) / CONVERT(FLOAT, Total\_Deaths) \* 100 AS Pct\_Deaths\_Age\_5\_to\_14,

CONVERT(FLOAT, Deaths\_Age\_Under\_5) / CONVERT(FLOAT, Total\_Deaths) \* 100 AS Pct\_Deaths\_Age\_Under\_5

FROM

(SELECT

year,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_70\_years\_Number])) AS Deaths\_Age\_70\_plus,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_50\_69\_years\_Number])) AS Deaths\_Age\_50\_to\_69,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_15\_49\_years\_Number])) AS Deaths\_Age\_15\_to\_49,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_5\_14\_years\_Number])) AS Deaths\_Age\_5\_to\_14,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_Under\_5\_Number])) AS Deaths\_Age\_Under\_5,

SUM(CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_70\_years\_Number])

+ CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_50\_69\_years\_Number])

+ CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_15\_49\_years\_Number])

+ CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_5\_14\_years\_Number])

+ CONVERT(BIGINT, [Deaths\_Tuberculosis\_Sex\_Both\_Age\_Under\_5\_Number])) AS Total\_Deaths

FROM

[dbo].[TB]

GROUP BY

year

) AS SubQuery

ORDER BY

year DESC

------------ Avg\_success rate------

SELECT [Entity], AVG([Indicator\_Treatment\_success\_rate\_for\_patients\_treated\_for\_MDR\_TB]) AS avg\_success\_rate\_mdr, AVG([Indicator\_Treatment\_success\_rate\_XDR\_TB\_cases]) AS avg\_success\_rate\_xdr, AVG([Indicator\_Treatment\_success\_rate\_new\_TB\_cases])

AS avg\_success\_rate\_newcases FROM [dbo].[TB]

GROUP BY [Entity]

ORDER BY avg\_success\_rate\_mdr DESC, avg\_success\_rate\_xdr DESC,avg\_success\_rate\_newcases DESC ;

----3. Comparative and Correlational Analysis------

-----3.1 Correlation Between HIV Prevalence and Tuberculosis Incidence-----

--------- Correlation between HIV prevalence and tuberculosis incidence

SELECT distinct

t1.[Entity],

t1.[Year],

t1.[Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis],

[Estimated\_HIV\_in\_incident\_tuberculosis],

([Estimated\_HIV\_in\_incident\_tuberculosis] \* [Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) / (AVG([Estimated\_HIV\_in\_incident\_tuberculosis]) OVER() \* AVG([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) OVER()) AS correlation\_factor

FROM [dbo].[TB] t1

JOIN [dbo].[5- tuberculosis-patients-with-hiv-share] t2 ON t1.[Entity] = t2.[Entity] AND t1.year = t2.year

ORDER BY correlation\_factor DESC;

------2. Predictive Model for TB Incidence Rates

-----Develop a predictive model using historical data to forecast future TB incidence rates. This could involve:

-------Creating a linear regression model in SQL to predict next year's incidence rates based on past data.

-------Using rolling averages

SELECT year, [Entity],

AVG([Estimated\_incidence\_of\_all\_forms\_of\_tuberculosis]) OVER (PARTITION BY [Entity] ORDER BY year ROWS BETWEEN 2 PRECEDING AND CURRENT ROW) AS moving\_avg\_incidence

FROM [dbo].[TB]

ORDER BY[Entity], year

SELECT TOP 1000\*

FROM [dbo].[TB]

select \* from [dbo].[TB]