**WORLD LIFE EXPECTANCY DATA CLEANING SQL COMMANDS**

1. Find out duplicates

use world;

select \* from worldlifeexpectancy;

select country, year, concat(country,year), count(concat(country,year))

from worldlifeexpectancy

group by country, year, concat(country,year)

having count(concat(country,year)) >1

;

1. **Remove duplicates**

select \* from (

Select Row\_ID,

concat(country,year),

Row\_number() over (partition by concat(country,year) order by concat(country,year)) as row\_num from worldlifeexpectancy) as row\_table

where row\_num > 1

;

DELETE from worldlifeexpectancy

where

Row\_ID IN ( select Row\_ID from (

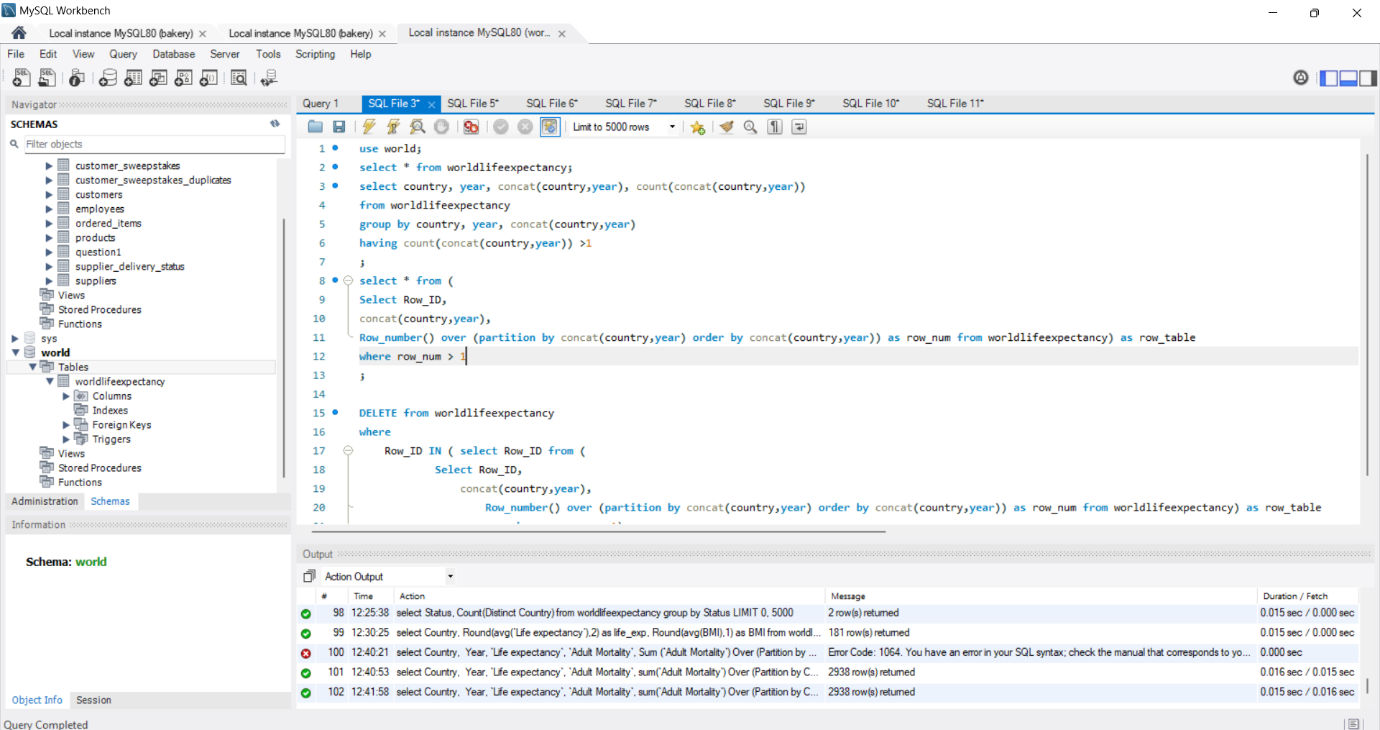
Select Row\_ID,

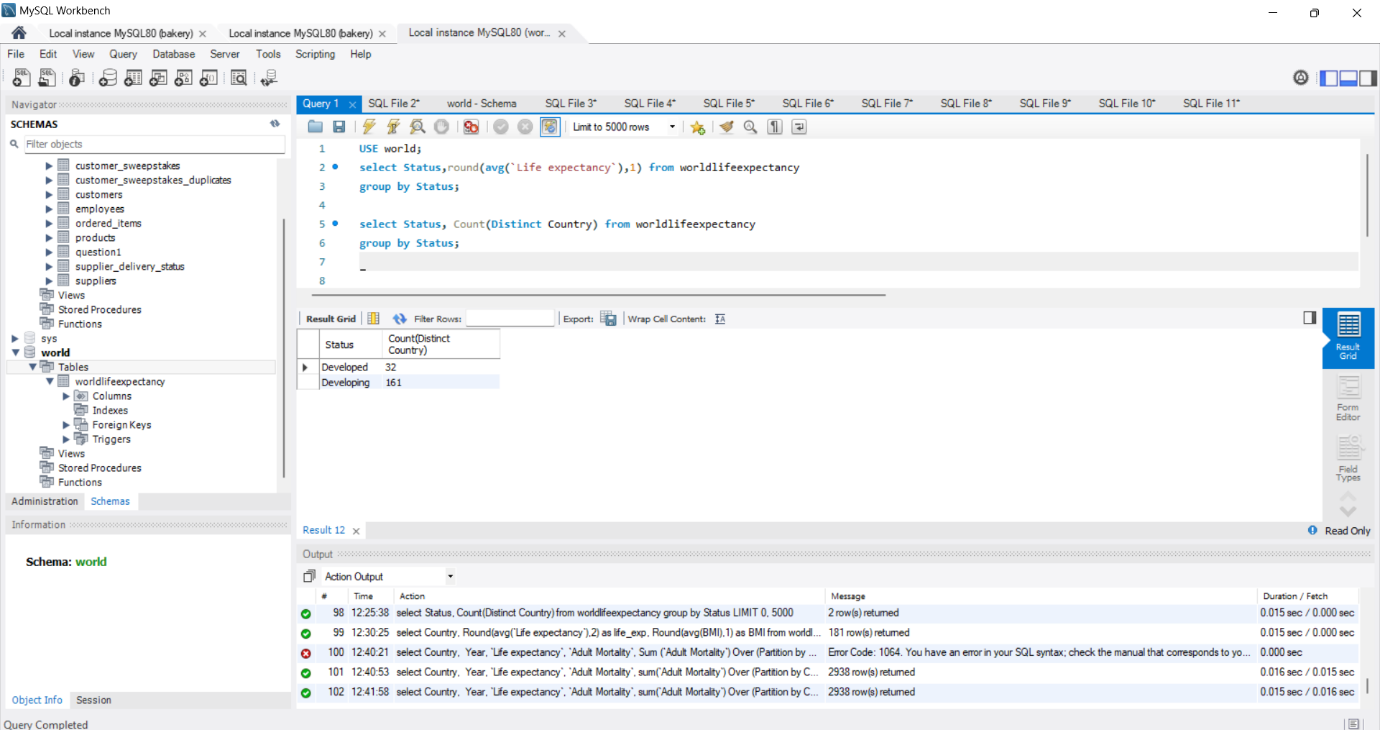
concat(country,year),

Row\_number() over (partition by concat(country,year) order by concat(country,year)) as row\_num from worldlifeexpectancy) as row\_table

where row\_num > 1)

;





**FILLING IN THE BLANKS IN THE STATUS COLUMN**

Update worldlifeexpectancy t1

Join worldlifeexpectancy t2

on t1.country = t2.country

Set t1.status = 'Developing'

where t1.status = ''

and t2.status <> ''

and t2.status = 'Developing'

;

FOR DEVELOPED COUNTRIES

Update worldlifeexpectancy t1

Join worldlifeexpectancy t2

on t1.country = t2.country

Set t1.status = 'Developed'

where t1.status = ''

and t2.status <> ''

and t2.status = 'Developed'

;

**LIFE EXPECTANCY COLUMN BLANKS FILL**

select t1.country, t1.Year, t1.`Life expectancy`,

t2.country, t2.Year, t2.`Life expectancy`,

t3.country, t3.Year, t3.`Life expectancy`,

Round((t2.`Life expectancy` + t3.`Life expectancy`)/ 2,1)

from worldlifeexpectancy t1

join worldlifeexpectancy t2

on t1.country = t2.country

and t1. year = t2.year - 1

join worldlifeexpectancy t3

on t1.country = t3.country

and t1. year = t3.year + 1

where t1.`Life expectancy` = ''

;

Update worldlifeexpectancy t1

join worldlifeexpectancy t2

on t1.country = t2.country

and t1. year = t2.year - 1

join worldlifeexpectancy t3

on t1.country = t3.country

and t1. year = t3.year + 1

set t1.`Life expectancy`= Round((t2.`Life expectancy` + t3.`Life expectancy`)/ 2,1)

where t1.`Life expectancy` = ''

;

**World Life Expectancy Exploratory Data Analysis**

**Differences in life expectancy in 15 years**

select Country, Min(`Life expectancy`), Max(`Life expectancy`),

Round (Max(`Life expectancy`) - Min(`Life expectancy`),1) as life\_expectancy\_difference

from worldlifeexpectancy

Group by Country

Having Min(`Life expectancy`) <> 0

and Max(`Life expectancy`) <> 0

order by life\_expectancy\_difference desc

;

**Average life expectancy all over the world**

select Year, Round(avg(`Life expectancy`),2)

from worldlifeexpectancy

where `Life expectancy` <> 0

group by Year

order by Year;

**Relationship between GDP and Life Expectancy**

select Country, Round(avg(`Life expectancy`),2) as life\_exp, Round(avg(GDP),1) as gdp

from worldlifeexpectancy

group by Country

Having life\_exp > 0

And gdp > 0

order by gdp asc

;

**Average life expectancy in developed and developing countries**

USE world;

select Status,round(avg(`Life expectancy`),1) from worldlifeexpectancy

group by Status;

select Status, Count(Distinct Country) from worldlifeexpectancy

group by Status;

**Relationship between BMI and Life Expectancy**

select Country, Round(avg(`Life expectancy`),2) as life\_exp, Round(avg(BMI),1) as BMI

from worldlifeexpectancy

group by Country

Having life\_exp > 0

And BMI > 1

order by BMI asc

;

**Relationship between Life Expectancy and Adult Mortality**

select Country,

Year,

`Life expectancy`,

`Adult Mortality`,

sum(`Adult Mortality`) Over (Partition by Country Order by Year) as rolling\_total

from worldlifeexpectancy;