**DATABASE NAME - EXPORT**

**TABLE NAME – EXPO, CONTINENT, ISLANDS**

**COLUMNS – EXPO(COMMODITY, COUNTRY, UNIT, QUANTITY, VALUE), CONTINENT(COUNTRY,CONTINENT), ISLANDS(ISLANDS, CONTINENT)**

QUERIES

select COMMODITY from expo

GROUP BY COMMODITY

169 ROWS

select COUNTRY from expo

GROUP BY COUNTRY

269 ROWS

select UNIT from expo OR select DISTINCT UNIT FROM EXPO

GROUP BY UNIT

239 ROWS

select QUANTITY from expo

GROUP BY QUANTITY

5780 ROWS

select VALUE from expo

GROUP BY VALUE

5337 ROWS

#CHANGING NAME OF COLUMN TO VALUE

EXEC sp\_rename 'EXPO.VALUE\_US\_million', 'Value', 'COLUMN';

#to check non numeric values

select quantity, ISNUMERIC(quantity) as result from Expo

where ISNUMERIC(quantity) = 0

UPDATE Expo

SET country = REPLACE(country, '"', '')

UPDATE Expo

SET value = REPLACE(value, 'NA', '')

UPDATE Expo

SET value = REPLACE(value, 'KGS', '')

UPDATE Expo

SET value = REPLACE(value, ',', '')

UPDATE EXPO

SET VALUE = CAST(VALUE AS FLOAT) FROM EXPO

UPDATE Expo

SET commodity = REPLACE(COMMODITY, '"', '')

UPDATE Expo

SET QUANTITY = REPLACE(quantity, 'NA', '0')

alter table expo

alter column quantity bigint

ALTER TABLE EXPO

DROP COLUMN UNIT

alter table expo

drop column quantity

select DATA\_TYPE FROM INFORMATION\_SCHEMA.COLUMNS WHERE TABLE\_NAME = 'EXPO' AND COLUMN\_NAME = 'VALUE'

this is to check the data type of a particular column of a particular table

select commodity, round(sum(value),2)as total\_value from expo

group by commodity

order by total\_value desc

select country, round(sum(value),2)as totalvalue from expo

group by COUNTRY

order by totalvalue desc

delete from expo

where quantity = 0 and value = 0

# 2222 rows were deleted left with 19977

select count(quantity) from expo

where quantity = 0

result - 10209

select count(value) from expo

where value = 0

1199

select commodity, count(distinct country) from Expo

group by COMMODITY

order by count(distinct country) desc

# to know which commodities are distributed vastly

select commodity, max(QUANTITY) as commodity\_getting\_max\_expo from expo

group by COMMODITY

order by max(QUANTITY) desc

select country, max(QUANTITY) as country\_getting\_max\_expo from expo

group by COUNTRY

order by max(QUANTITY) desc

select commodity, country, round(value, 2),

DENSE\_RANK() over(partition by commodity order by value desc) as sales\_rank

from expo

# to give rank to each country importing a particular product on the basis of value

select commodity, country, round(value, 2),

DENSE\_RANK() over(partition by country order by value desc) as sales\_rank

from expo

# to give rank to each commodity importing a particular product on the basis of value

Next I am trying to give continent name for every country

Few continents have gained a value of null because there is some ambiguity with the names of the country

To check how many and which country have got their continents as null following is the query.

with first as

(

select a.country, b.continent from expo a

left join continent b

on a.country= b.country

)

select \* from first

where first.continent is null

following is the query which shows the names of country not falling in any continent

with first as

(

select a.country, b.continent from expo a

left join continent b

on a.country= b.country

)

select first.country from first

where first.continent is null

group by first.country

# GOT 79 ROWS TO CORRECT OF COUNTRIES NOT REGISTERING UNDER CONTINENT

THERE ARE TYPOS AND MANY MORE PROBLEMS

INSERT INTO Continent(COUNTRY, CONTINENT)

VALUES ('ANTARTICA','ANTARTICA')

UPDATE Continent

SET COUNTRY = 'ANTARCTICA', CONTINENT = 'ANTARCTICA'

WHERE COUNTRY='ANTARTICA'

To solve the problem of 79 rows where countries are not matching continents there are some islands also and some typos so here new data of islands is introduced and first cleaned in excel now in sql

delete from island

where islands is null

#DATA CLEANING

SELECT

CASE

WHEN CHARINDEX('(', ISLANDS) > 0 THEN

SUBSTRING(ISLANDS, 1, CHARINDEX('(', ISLANDS) - 1)

WHEN CHARINDEX('[', ISLANDS) > 0 THEN

SUBSTRING(ISLANDS, 1, CHARINDEX('[', ISLANDS) - 1)

ELSE

ISLANDS

END AS ISLANDS

FROM

ISLAND

# CORRECTING COUNTRY NAMES

UPDATE EXPO

SET COUNTRY='BAHRAIN'

WHERE COUNTRY='BAHARAIN IS'

DELETE FROM Continent

where country like '%China'

UPDATE ISLAND

SET ISLANDS =

CASE

WHEN CHARINDEX('(', ISLANDS) > 0 THEN

SUBSTRING(ISLANDS, 1, CHARINDEX('(', ISLANDS) - 1)

WHEN CHARINDEX('[', ISLANDS) > 0 THEN

SUBSTRING(ISLANDS, 1, CHARINDEX('[', ISLANDS) - 1)

ELSE

ISLANDS

END;

#now I will join three tables and lastly merge continent and islands and put it in temp table from cte

with first as

(

select commodity,

a.country,

CASE

WHEN continent IS NOT NULL THEN continent

ELSE cont

END AS Location\_on\_earth,

a.Value

from expo a

left join continent b on a.country= b.country

left join ISLAND C on a.country= c.islands

)

SELECT \* into #data from first;

#last few left countries

UPDATE #data

SET location\_on\_earth = 'Others'

WHERE location\_on\_earth IS NULL;

#now I am moving data to power bi so I converted temp table #data to permanent table i.e exportdata and dropped temp table

CREATE TABLE dbo.ExportData (

Commodity nvarchar(50),

Country nvarchar(50),

Continent nvarchar(50),

Total\_Value float

);

INSERT INTO dbo.ExportData (Commodity ,Country, Continent, Total\_Value)

SELECT commodity, country, location\_on\_earth, value

FROM #data;

select \* from exportdata

drop table #data