## created a new table to keep the raw data

create table annual\_change\_solar(entity text,

code\_ text,

year\_collected year,

solar\_terrawatt\_hr\_yr float);

## insert data into new table from raw table

insert into my\_sql\_prol.annual\_change\_solar select \* FROM my\_sql\_prol.`annual-change-solar - copy`;

## remove the codes column in the new table

alter table annual\_change\_solar

drop column code\_

## delete rows where no photovoltaic energy was collected

delete from my\_sql\_prol.annual\_change\_solar

where solar\_terrawatt\_hr\_yr = 0

## check dublicates against number of rows

select distinct entity,year\_collected,solar\_terrawatt\_hr\_yr

from my\_sql\_prol.annual\_change\_solar

select distinct entity

from my\_sql\_prol.annual\_change\_solar

select count(distinct(entity)) as categories

from my\_sql\_prol.annual\_change\_solar

select asia,africa,america,antarctica,europe

from annual\_change\_solar

## delete continents

set sql\_safe\_updates = 0

delete from my\_sql\_prol.annual\_change\_solar

where entity regexp 'america|antarctica|europe|asia|middle|OECD|world'

set sql\_safe\_updates = 1

##check for nulls

SELECT \*

FROM my\_sql\_prol.annual\_change\_solar

where entity is null

SELECT \*

FROM my\_sql\_prol.annual\_change\_solar

where year\_collected is null

SELECT \*

FROM my\_sql\_prol.annual\_change\_solar

where solar\_terrawatt\_hr\_yr is null

## remove dublicates from all columns

SELECT entity,year\_collected,solar\_terrawatt\_hr\_yr,count(\*)

FROM my\_sql\_prol.annual\_change\_solar

group by 1,2,3

having count(\*) > 1

EXPLORING CODE

##Maximum terra Watts collected

select distinct entity, max(solar\_terrawatt\_hr\_yr) over(partition by entity) max\_growth, year\_collected

from annual\_change\_solar

order by 2 desc

limit 1

##minimum terra Watts collected

select distinct entity, min(solar\_terrawatt\_hr\_yr) over(partition by entity) max\_growth, year\_collected

from annual\_change\_solar

order by 2 desc

limit 1

## average terra Watts collected

select distinct entity, avg(solar\_terrawatt\_hr\_yr) over(partition by entity) max\_growth

from annual\_change\_solar

order by 2 desc

limit 1

select distinct entity,

min(year\_collected) over(partition by entity) ealiest\_solar\_users

from annual\_change\_solar

order by 2

limit 1

select distinct entity,

min(year\_collected) over(partition by entity) late\_solar\_users

from annual\_change\_solar

order by 2 desc

##Create a new table for prices in USA for solar installations

create table solar\_pv\_system\_costs (entity text,

code\_ text,

year\_collected year,

Total\_Installed\_Price\_Barbose\_and\_Darghouth\_twentysixten float,

Non\_Module\_Costs\_Barbose\_Darghouth\_twentysixten float,

Module\_Price\_Barbose\_Darghouth\_twentysixten float );

insert into my\_sql\_prol.solar\_pv\_system\_costs \* FROM my\_sql\_prol.`solar-pv-system-costs-raw`;

##WHY THERE WAS INCREASE IN TERRAS COLLECTED WAS BECAUSE OF DECREASE OF COSTS ALSO

SELECT costs.entity,

costs.year\_collected,

costs.Total\_Installed\_Price\_Barbose\_and\_Darghouth\_twentysixten,

chang.solar\_terrawatt\_hr\_yr

FROM my\_sql\_prol.solar\_pv\_system\_costs costs

join annual\_change\_solar chang on costs.entity = chang.entity

and costs.year\_collected = chang.year\_collected

order by 3