SELECT date FROM Avacado\_dataset.avocado;

ALTER TABLE avocado ADD COLUMN Date\_new DATE;

SET SQL\_SAFE\_UPDATES = 0;

UPDATE avocado SET Date\_new = STR\_TO\_DATE (Date, '%y-%m-%d'); -- change Date column with string format

ALTER TABLE avocado DROP column Date; ALTER TABLE avocado CHANGE COLUMN Date\_new Date DATE;

select distinct date from Avacado\_dataset.avocado;

UPDATE avocado SET Date\_new = STR\_TO\_DATE (Date, '%Y-%m-%d') WHERE STR\_TO\_DATE (Date, '%Y-%m-%d') IS NOT NULL;

UPDATE avocado SET Date = TRIM(Date);

SELECT Date, Date\_new FROM Avacado\_dataset.avocado;

SELECT \* FROM Avacado\_dataset.avocado;

-- removing unnecessary column ALTER TABLE avocado DROP COLUMN MyUnknownColumn;

select \* from Avacado\_dataset.avocado;

-- removing duplicates

ALTER TABLE avocado ADD COLUMN id INT AUTO\_INCREMENT PRIMARY KEY;

With CTE as (select \*, row\_number () over (partition by region,year,type order by id) as rn\_1 from Avacado\_dataset.avocado)

DELETE from avocado where id IN (select id from CTE where rn\_1 >1);

select id from Avacado\_dataset.avocado; select \* from Avacado\_dataset.avocado;

-- Check for Negative or Unusual Values in Numeric Columns

select \* from Avacado\_dataset.avocado where 'AveragePrice' < 0 OR 'Total Volume' < 1 OR '4046' < 0 OR '4770' < 0 OR 'Small Bags' < 0 OR 'Large Bags' < 0 OR 'XLarge Bags' < 0;

-- check for spacing, dimensions verification

select type, TRIM(TYPE) AS New\_type, TRIM (region) as new\_region from Avacado\_dataset.avocado;

update avocado SET TYPE = TRIM (type); update avocado SET region = TRIM (region);

UPDATE avocado SET TYPE = LOWER (type);

-- adding total\_num\_of\_bags

select \* from Avacado\_dataset.avocado where 'Total Bags' = 'Small Bags' + 'Large Bags' + 'XLarge Bags' ; -- which regions have highest avocado prices select region, ROUND(AVG(AveragePrice), 2) as avg\_price FROM Avacado\_dataset.avocado group by region order by avg\_price DESC;

ALTER TABLE avocado ADD COLUMN avg\_price DECIMAL (5,2);

Update avocado a JOIN (select region, ROUND(AVG(AveragePrice), 2) as avg\_price FROM Avacado\_dataset.avocado group by region) as avg\_table ON a.region = avg\_table.region set a.avg\_price = avg\_table.avg\_price ;

select \* from Avacado\_dataset.avocado;