**SQL scripts** **used in the project**

#1) Create a new database table

CREATE TABLE food\_waste(

code INT,

country VARCHAR(100),

region VARCHAR(100),

cpc\_code INT,

commodity VARCHAR(50),

year VARCHAR(5),

loss\_percentage DECIMAL (10,2),

loss\_percentage\_original VARCHAR(10),

loss\_quantity VARCHAR(100),

activity VARCHAR(1000),

supply\_stage VARCHAR(1000),

treatment VARCHAR(2000),

cause\_of\_loss VARCHAR(2000),

sample\_size VARCHAR(2000),

datacollection\_method VARCHAR(1000),

reference VARCHAR(1000),

url VARCHAR(1000),

notes VARCHAR(2000)

)

#2) Copy the CSV data to the database table

COPY food\_waste

FROM 'C:\Users\Public\SQL datasets\Food loss and waste database.csv'

DELIMITER ','

CSV HEADER;

#3) In order to copy the data correctly, some data types or data sizes had to be changed, for e.g.

ALTER TABLE food\_waste

ALTER COLUMN region TYPE VARCHAR(1000);

#4) Let’s look at the whole data

SELECT \*

FROM food\_waste;

#5) The dataset was exported with data from 1965 (i.e. The maximum period possible); to check all the years included in this range, I ran the following query, where I can see that the first year included is actually 1966 and it goes until 2021, with no data only for 1967 in between

SELECT DISTINCT(year)

FROM food\_waste

ORDER BY year DESC;

#6) I am also curious to see whether all the countries in the dataset have data points for this entire 55 years period; I query this by using the count function, requesting to count the distinct number of years that each country shows

SELECT COUNT(DISTINCT year) AS years\_of\_data,

country

FROM food\_waste

GROUP BY country

ORDER BY years\_of\_data DESC;

#7) I also want to know about how many distinct commodities we have waste data available and also for each country, on how many distinct commodities we have data available about

SELECT COUNT(DISTINCT commodity)

FROM food\_waste;

SELECT COUNT(DISTINCT commodity) AS distinct\_commodities,

country

FROM food\_waste

GROUP BY country

ORDER BY distinct\_commodities DESC;

#8) Top and bottom countries by food waste

SELECT AVG(loss\_percentage) AS loss\_percentage\_avg,

country

FROM food\_waste

GROUP BY country

ORDER BY loss\_percentage\_avg DESC

LIMIT 10;

SELECT AVG(loss\_percentage) AS loss\_percentage\_avg,

country

FROM food\_waste

GROUP BY country

ORDER BY loss\_percentage\_avg ASC

LIMIT 10;

#9) top 10 commodities and loss stages with most waste

SELECT AVG(loss\_percentage) AS loss\_percentage\_avg,

commodity

FROM food\_waste

GROUP BY commodity

ORDER BY loss\_percentage\_avg DESC

LIMIT 10;

SELECT AVG(loss\_percentage) AS loss\_percentage\_avg,

supply\_stage

FROM food\_waste

GROUP BY supply\_stage

ORDER BY loss\_percentage\_avg DESC

LIMIT 10;

#10) instance of looking more in detail into some of the data queried above

SELECT \*

FROM food\_waste

WHERE country = 'Haiti'  
  
#11) we now ran the query for a restricted time period

WITH last\_20\_years AS (

SELECT \* FROM food\_waste

WHERE year > 2000 AND year < 2022

)

SELECT AVG(loss\_percentage),

country

FROM last\_20\_years

GROUP BY country

ORDER BY avg DESC;

#12) let's now observe the waste trend by comparing rolling averages of the last 3 years down until the first year of available data

WITH yearly\_avg AS(

SELECT AVG(loss\_percentage) AS average,

year

FROM food\_waste

GROUP BY year

)

SELECT \*,

AVG(average) OVER(

ORDER BY year DESC

ROWS BETWEEN 2 PRECEDING AND CURRENT ROW

) AS rolling\_average

FROM yearly\_avg