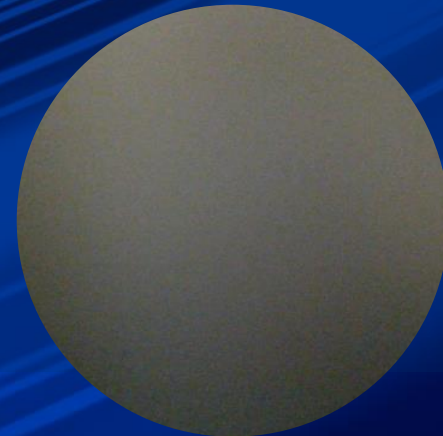


Trends of crop yield in the Netherlands with respect to emission in water from year 2000 to 2015

Bilal Ahmad Butt (23072530)

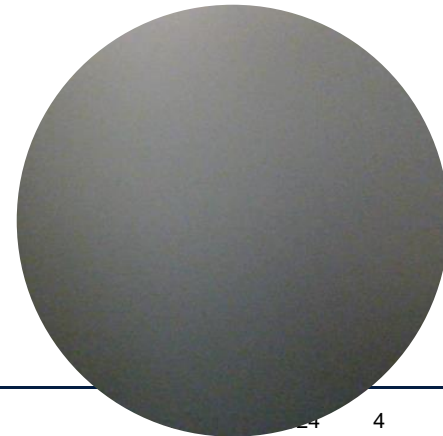


- 1 Introduction
 - 1.1 Motivation
 - 1.2 Problem in hand
 - 1.3 Hypothesis
- 2 Methods
 - 2.1 Data pipeline
 - 2.2 Data cleaning
 - 2.3 Data transformation
 - 2.4 About datasets
 - 2.5 Data source 1: Vegetables yield year 2000 to 2015
 - 2.6 Data source 2: Emissions to water year 2000 to 2015
- 3 Results
- 4 Conclusion

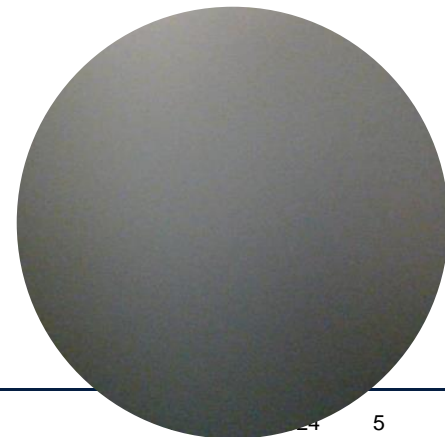


Introduction

- ☐ Reduce the water pollution.
- ☐ Protect water.
- ☐ Food security.
- ☐ Improve economy, health and reduce diseases.



- ☐ Has the vegetable yield been increased or decreased?
- ☐ Has the pollution of water increased or decreased?
- ☐ What is the relation between vegetable yield and water?
- ☐ What is the most grown vegetable in Netherlands?
- ☐ What is the most emitted element's compounds are released in water?



- ❑ With the decrease in heavy elements and nutrients; the yield of different vegetables will increase.

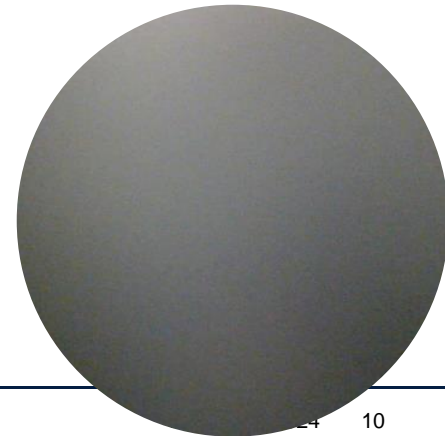
Methods



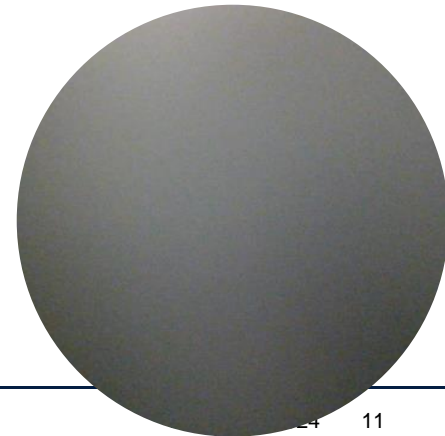
- ❑ Get the data from the sources → created a class `DataPipeline` in `/project/pipeline.py`
 - `download_data()`: to down the data.
 - `transform_data()`: to tranform the data.
 - `establish_database_connection()`: creates database and connects with it and return a cursor to perform CRUD on db.
 - `load_data()`: to transport all the cleaned and transformed into database.
 - `run_pipeline()`: a driver method to initial all the process.
 - Implemented utility methods.

- ☐ Using pandas `read_csv`, downloaded the data and cleaned it.
- ☐ Separated the data based on delimiter semi-colon(;).
- ☐ Configured that on receiving a bad line which is out of UTF8 to be skipped.
- ☐ Took out the columns (headers) out of data.
- ☐ Set the first column (header) as Index.

- ❑ Based on the data sources, different transformations were performed.
- ❑ Data source 1: Vegetable data
 - Setup the index and renamed it for easier access.
 - Correction of columns.
- ❑ Data source 2: Water data
 - Setup the index and renamed it for easier access.
 - Renamed the columns.
 - Removed the columns with NAN value.



- ❑ Two datasets have been used.
 - Downloaded from CBS Open data StatLine.
- ❑ Data repository is Netherlands statistics database.
 - Offers data on the economy and society.
- ❑ First dataset: Vegetables yield year 2000 to 2015.
- ❑ Second dataset: Emissions to water year 2000 to 2015.
- ❑ Data engineering is done for 2000, 2005, 2010, 2015.



Methods

Data source 1: Vegetables yield year 2000 to 2015

- ❑ 126 rows, 1 index and 2 columns.
 - Index represents vegetables.
- ❑ Represents yield in million kilograms.
- ❑ `df_vegetable.index.unique()`
 - Onion/Garlic, Strawberry, Spinach, Belgian endive, Lettuce, Iceberg, Tuberous, Radishes, Carrot, Cabbage, Legume, Bean, Apple, Courgettis.
- ❑ 2000, 2005, 2010, 2015

	year	gross_yield_million_kilogram
Vegetables		
Onion/Garlic	1998	3339
Onion/Garlic	2000	3814,8
Onion/Garlic	2005	4255
Onion/Garlic	2006	4151,2
Onion/Garlic	2010	4626,8
...
Courgette	2010	18
Courgette	2015	18,4
Courgette	2020	
Courgette	2021	
Courgette	2022	

126 rows × 2 columns

- ❑ Filter data

- ❑ By year

- `fixed_year_2000 = df_vegetable.loc[df_vegetable['year'] == 2000]`

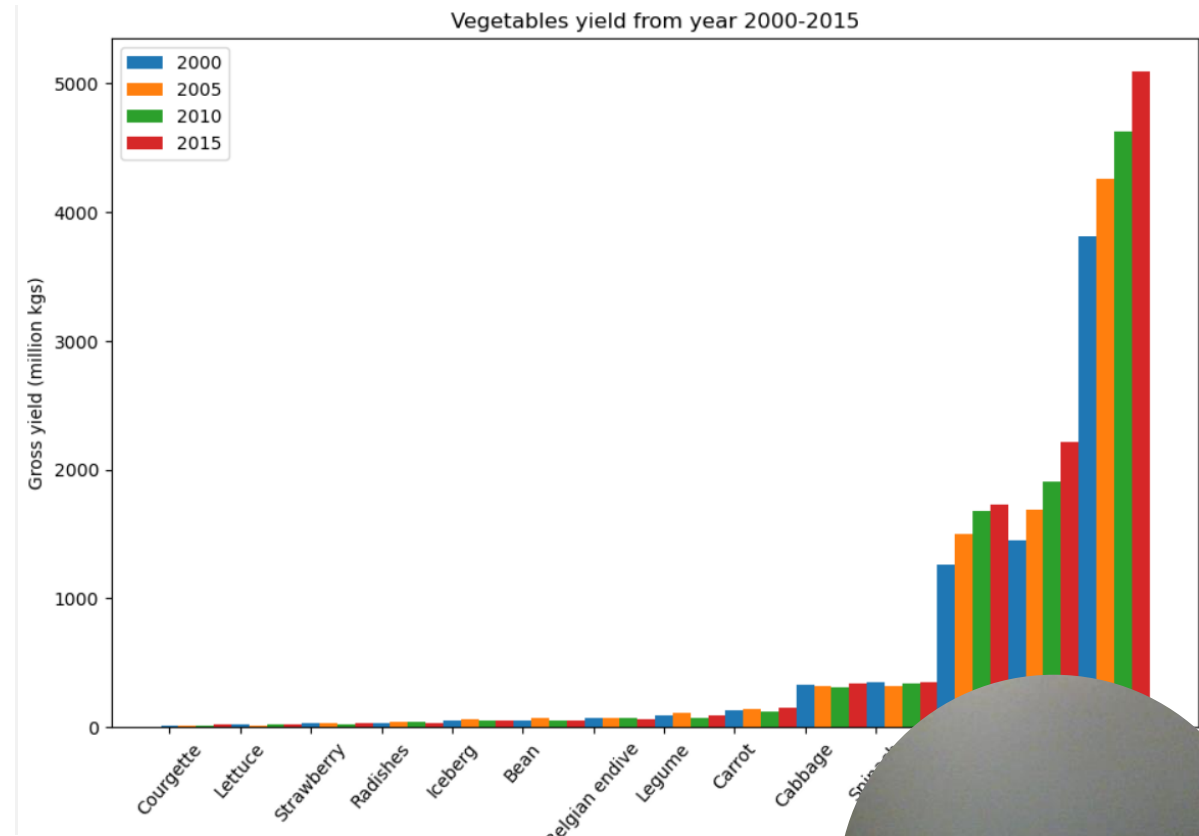
- ❑ By index

- `onion = df_vegetable.loc[df_vegetable.index == 'Onion/Garlic']`

Methods

Data source 1: Vegetables yield year 2000 to 2015

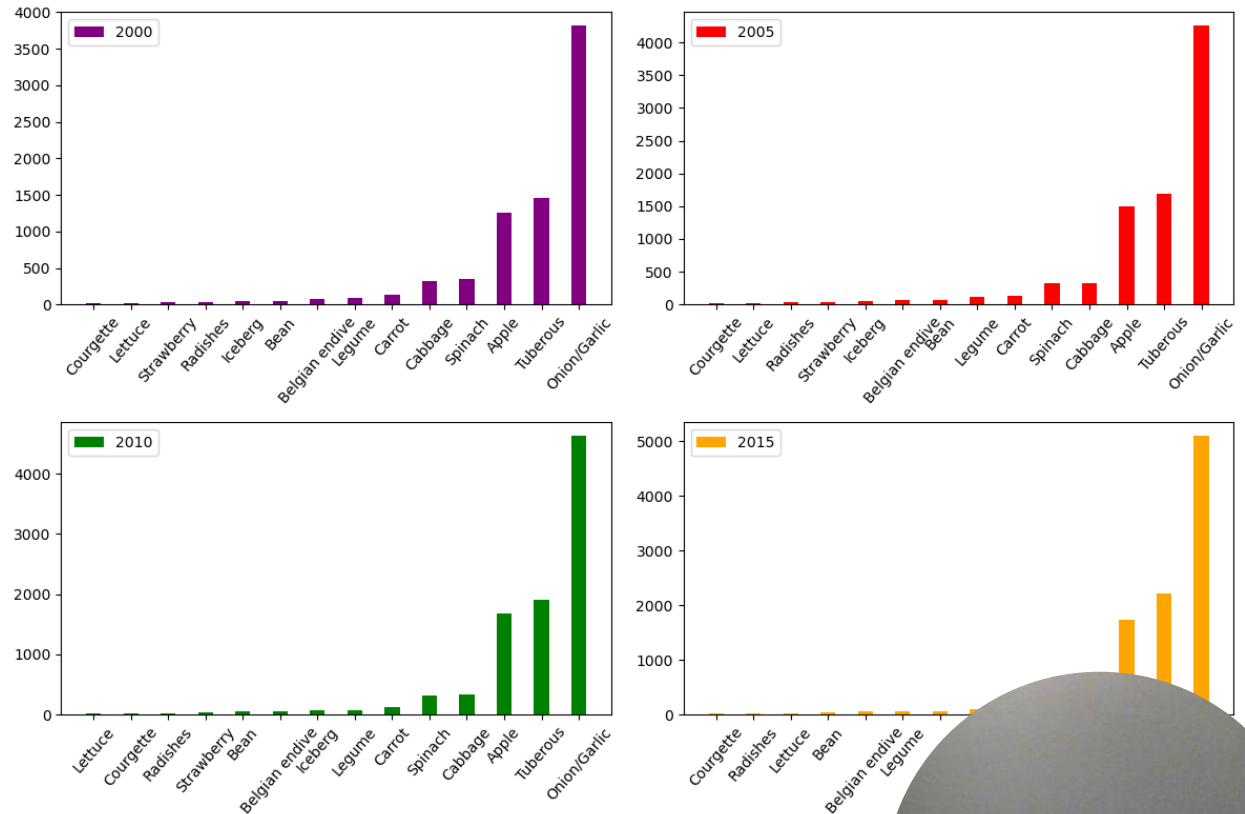
This bar chart represents quantity of the yield have as year progresses.



Methods

Data source 1: Vegetables yield year 2000 to 2015

This bar chart represents quantity of the yield have as year progresses.



Methods

Data source 2: Emissions to water year 2000 to 2015

- ❑ Multiple columns
 - Year, Heavy elements, Nutrients.
- ❑ Values represents kilograms.
- ❑ Index is the origin factor.
- ❑ 2000, 2005, 2010, 2015

	Year	Chromium	Copper	Mercury	Lead	Nickel	Zinc	OtherNutrients	Phosphorus
Origin									
Industrial	1995.0	27414.0	40266.0	408.0	13053.0	28869.0	91518.0	7433.0	5391283.0
Industrial	2000.0	10418.0	30330.0	165.0	8362.0	12573.0	63879.0	4702.0	3268213.0
Industrial	2005.0	8492.0	22300.0	86.0	3706.0	10389.0	49463.0	3334.0	1929056.0
Industrial	2010.0	2761.0	13454.0	23.0	1512.0	5493.0	25867.0	3029.0	1740726.0
Industrial	2013.0	2347.0	12164.0	27.0	1588.0	4723.0	26503.0	2380.0	1361662.0
Industrial	2014.0	2215.0	12362.0	12.0	1296.0	4468.0	20166.0	2383.0	1308156.0
Chemical Manufacturing	1995.0	7001.0	8386.0	216.0	5331.0	6131.0	27685.0	3592.0	3140946.0
Chemical Manufacturing	2000.0	2495.0	4780.0	25.0	1622.0	3328.0	26936.0	1684.0	1416894.0
Chemical Manufacturing	2005.0	3147.0	3761.0	15.0	1538.0	1959.0	21810.0	504.0	192690.0
Chemical Manufacturing	2010.0	555.0	891.0	3.0	150.0	1080.0	8305.0	346.0	158679.0
Chemical Manufacturing	2013.0	396.0	585.0	2.0	292.0	1052.0	9371.0	320.0	109549.0
Chemical Manufacturing	2014.0	543.0	564.0	3.0	118.0	807.0	6330.0	284.0	102991.0
Pharmaceuticals Manufacturing	1995.0	1.0	164.0	0.0	271.0	163.0	665.0	76.0	50402.0
Pharmaceuticals Manufacturing	2000.0	20.0	162.0	0.0	10.0	235.0	910.0	99.0	52101.0
Pharmaceuticals Manufacturing	2005.0	12.0	149.0	0.0	4.0	52.0	471.0	91.0	50765.0
Pharmaceuticals Manufacturing	2010.0	45.0	136.0	0.0	0.0	16.0	573.0	49.0	24224.0
Pharmaceuticals Manufacturing	2013.0	12.0	8.0	0.0	0.0	11.0	316.0	57.0	40978.0
Pharmaceuticals Manufacturing	2014.0	18.0	122.0	0.0	0.0	52.0	793.0	108.0	81313.0

- ❑ Filter data

- ❑ By year

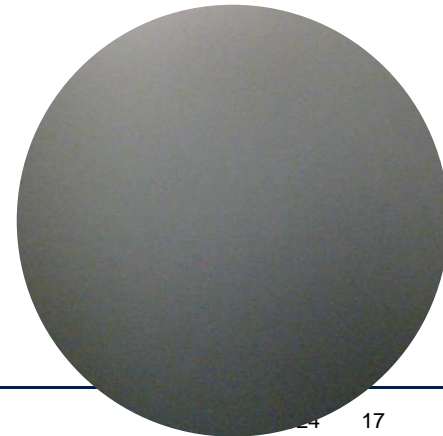
- `water_year_2000 = df_water[df_water['Year'] == 2000]`

- ❑ By index

- `Industrial = df_water.loc[df_water.index == 'Industrial']`

- ❑ By column

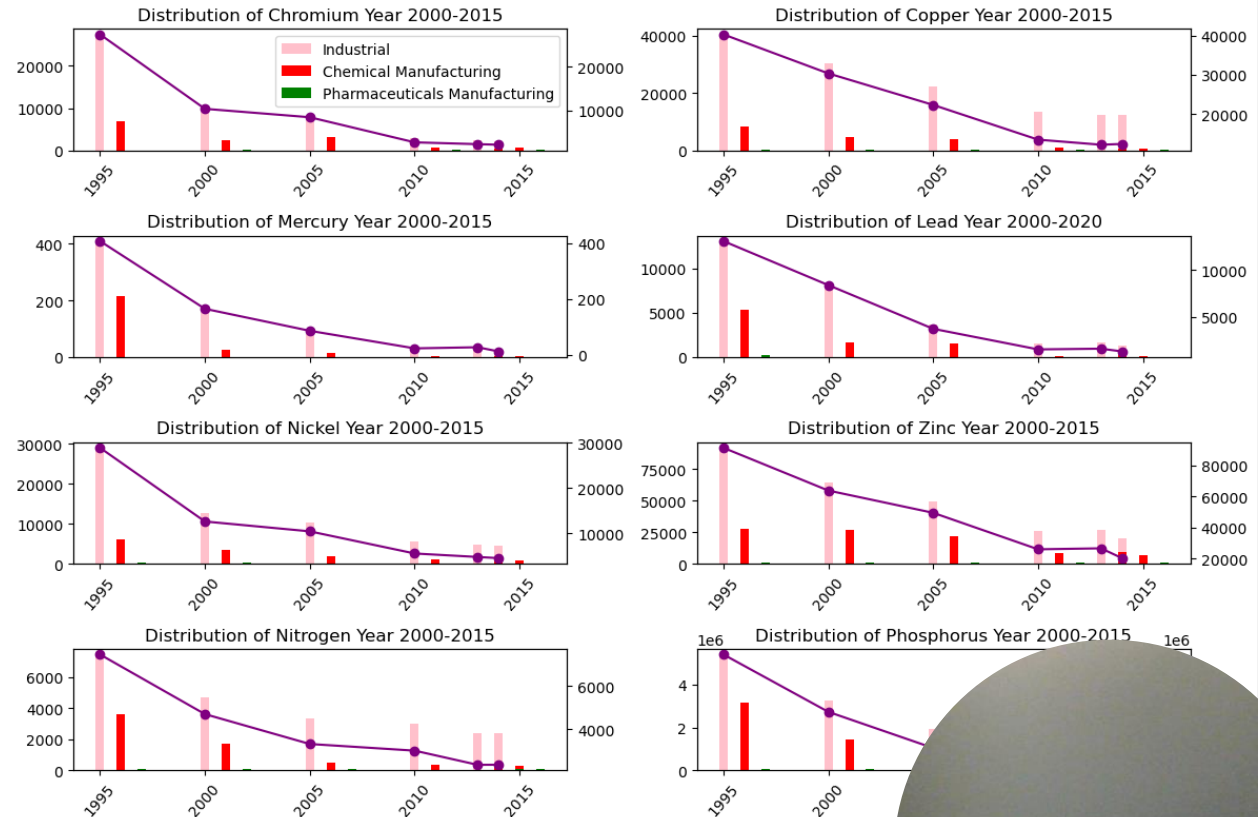
- `Chromium = df_water[['Chromium']]`



Methods

Data source 2: Emissions to water year 2000 to 2015

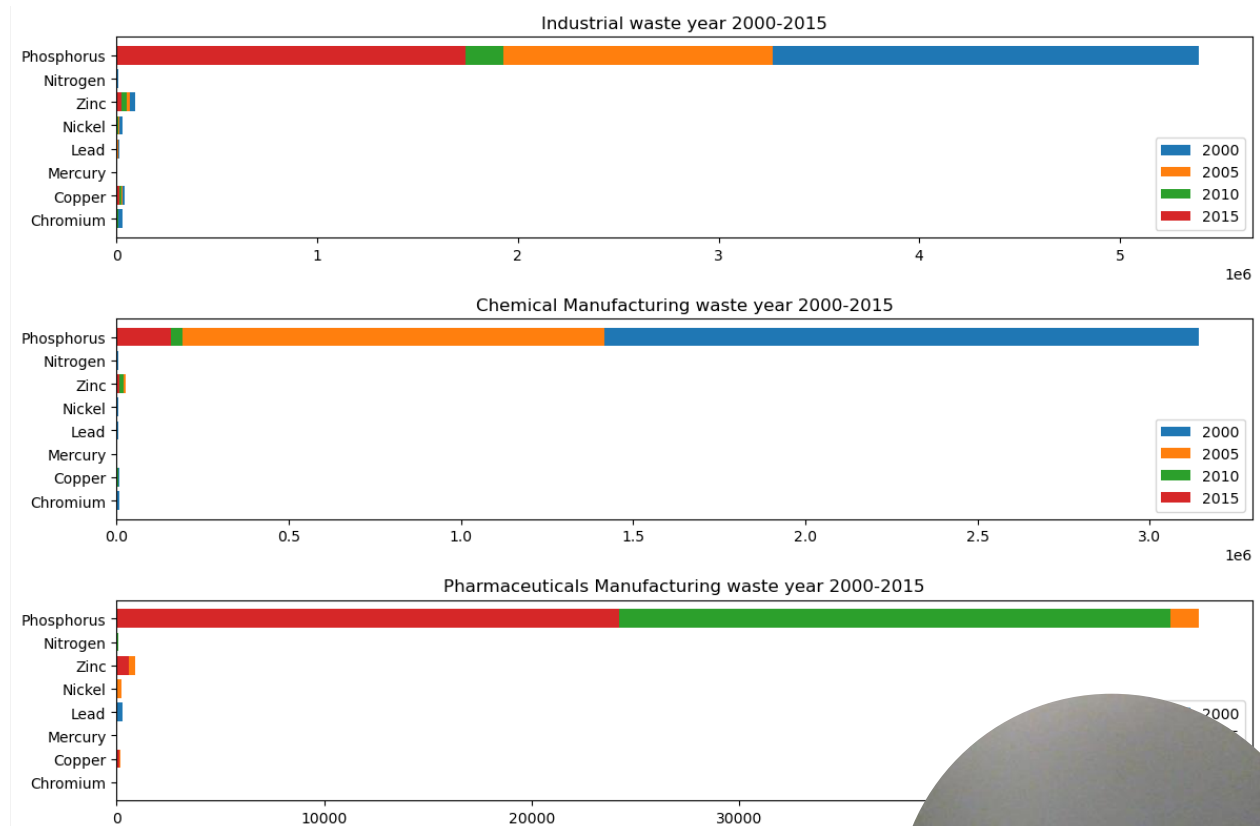
Combination of curve and bar chart represents emission of compounds compose of different heavy elements and nutrients into water.



Methods

Data source 2: Emissions to water year 2000 to 2015

Horizontal bar graphs represents emission of compounds compose of different heavy elements and nutrients into water.



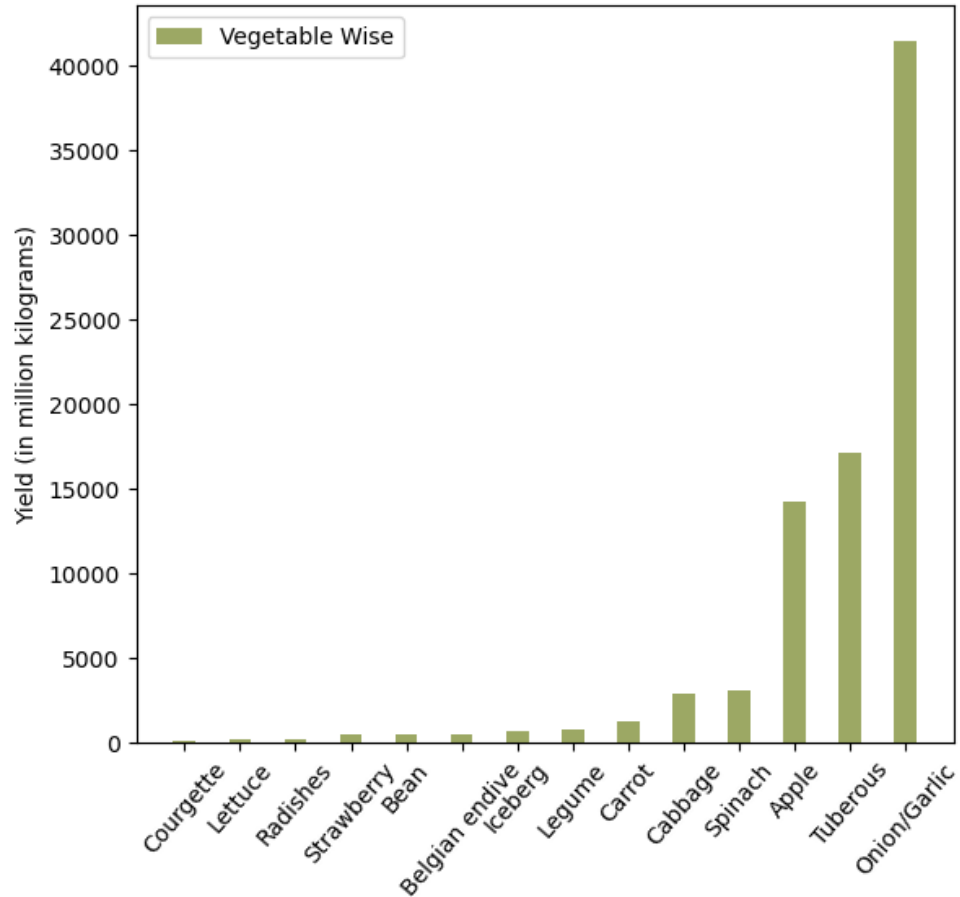
Results



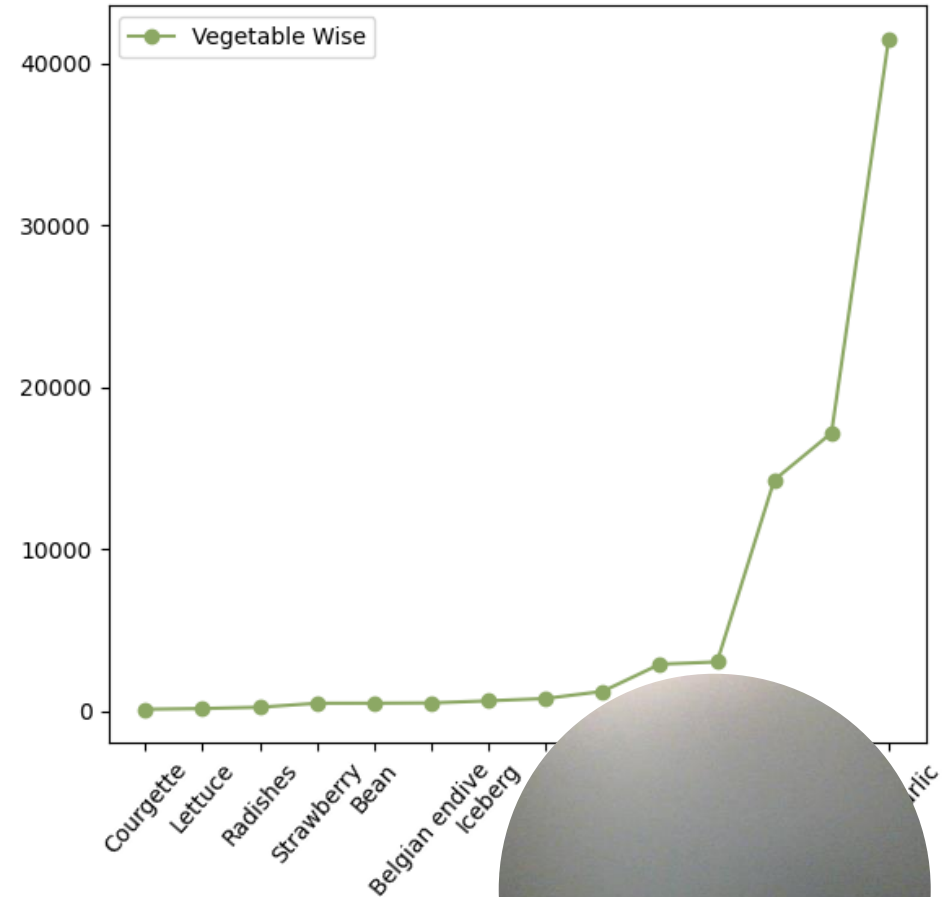
Results

Data source 1: Vegetables yield year 2000 to 2015

Vegetable yield over the years 1995-2020



Vegetable yield over the years 1995-2020

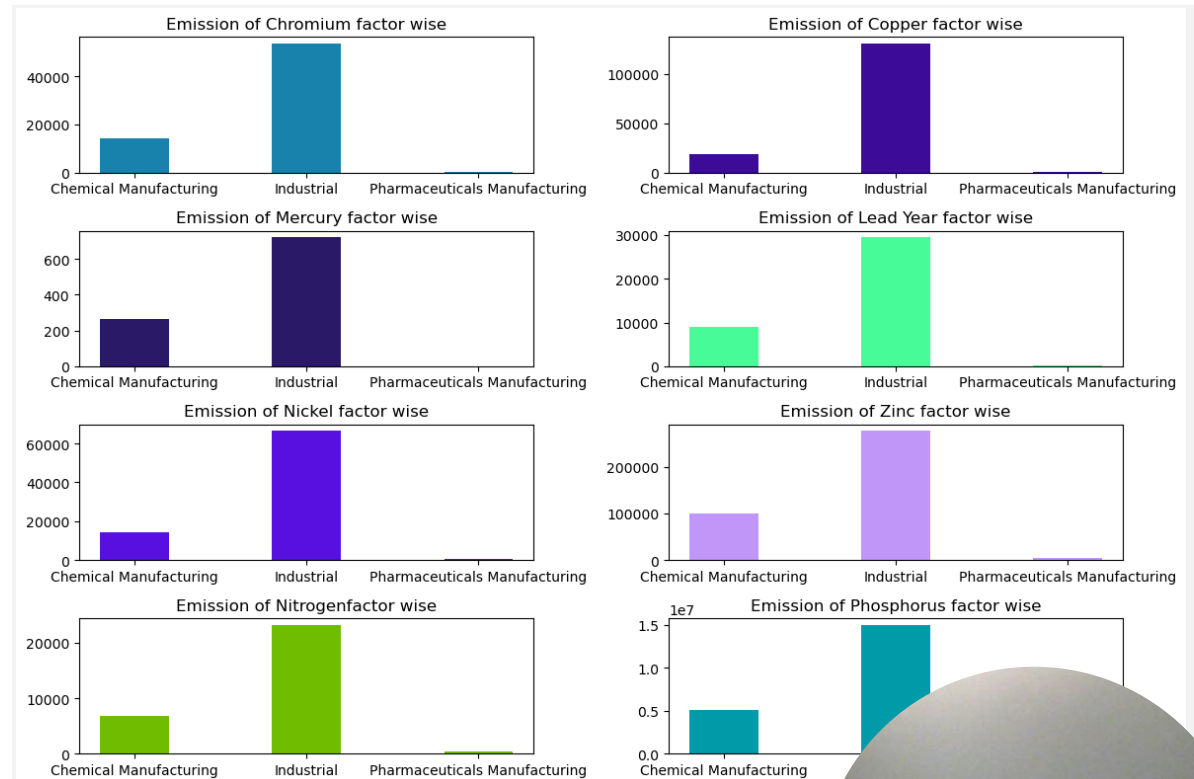


❑ Most grown vegetables in Netherlands are onion and garlicks.

Results

Data source 2: Emissions to water year 2000 to 2015

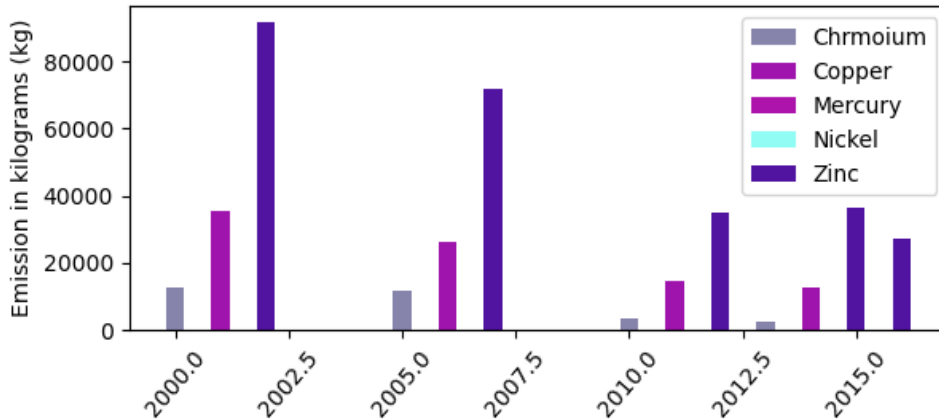
- ❑ Y-axis are adjusted according to each element automatically by matplotlib.
- ❑ Phosphorous was the most emitted element in the waters of Netherland by Industries.



Results

Analysis Illustration

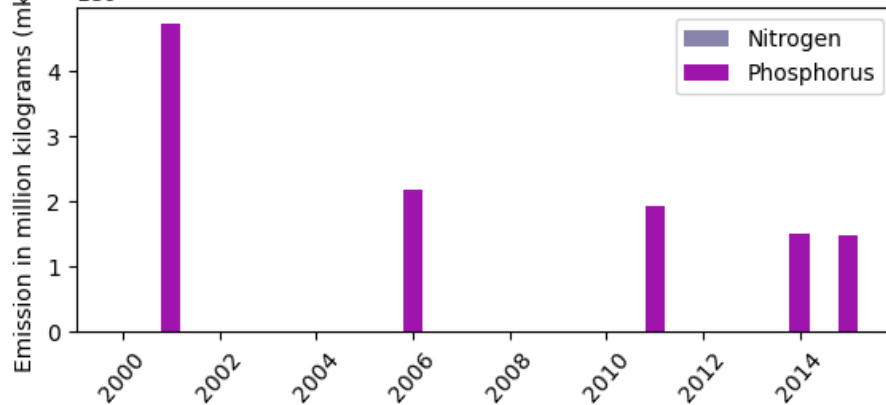
Heavy elements emission in water over the years 2000-2015



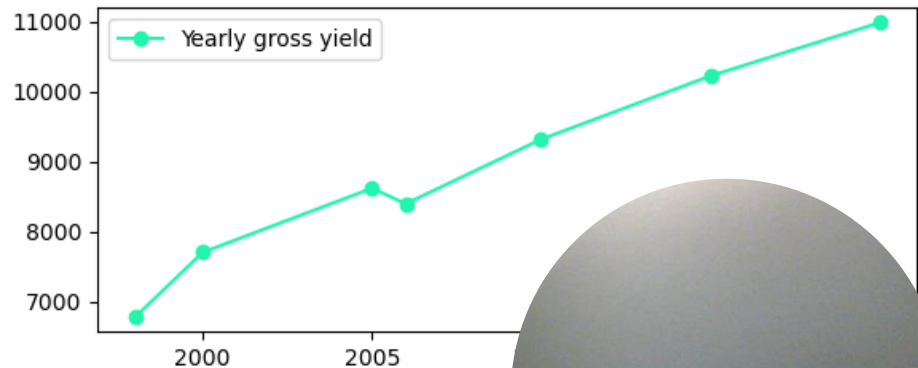
Vegetable yield (in million kilograms)



1e6 Nutrients emission in water over the years 2000-2015



Vegetable yield (in million kilograms)

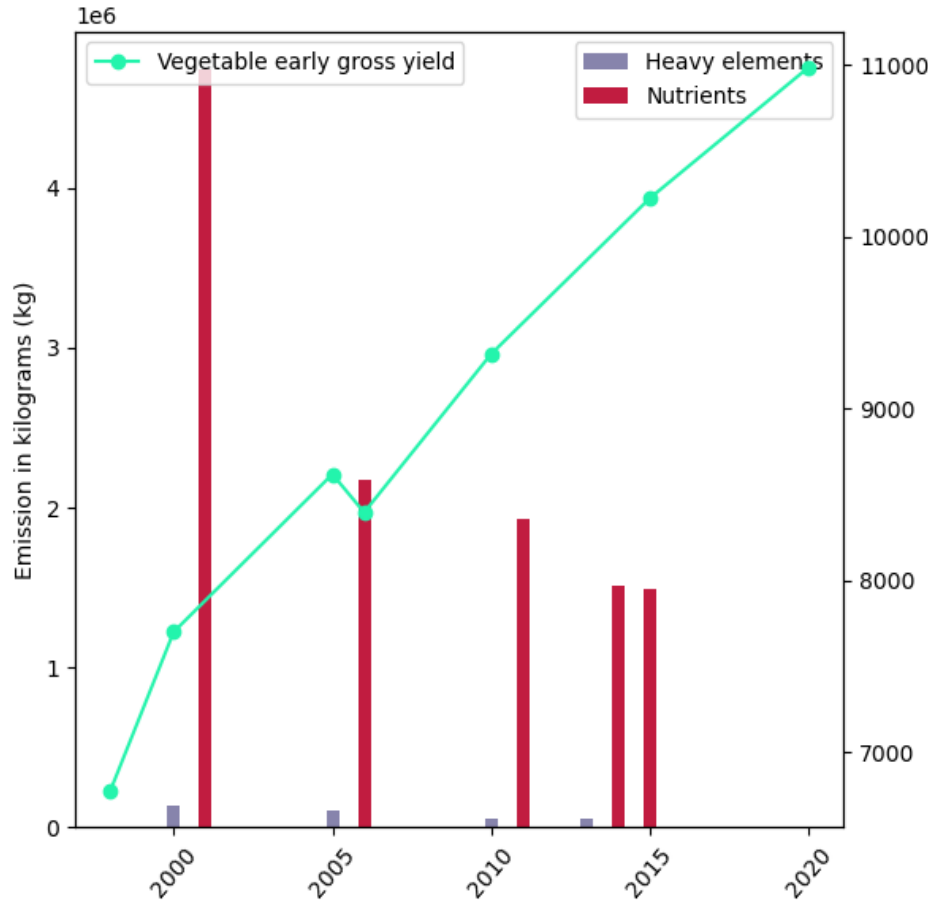


Reduction in compounds of heavy elements. Nutrients emission has also reduced

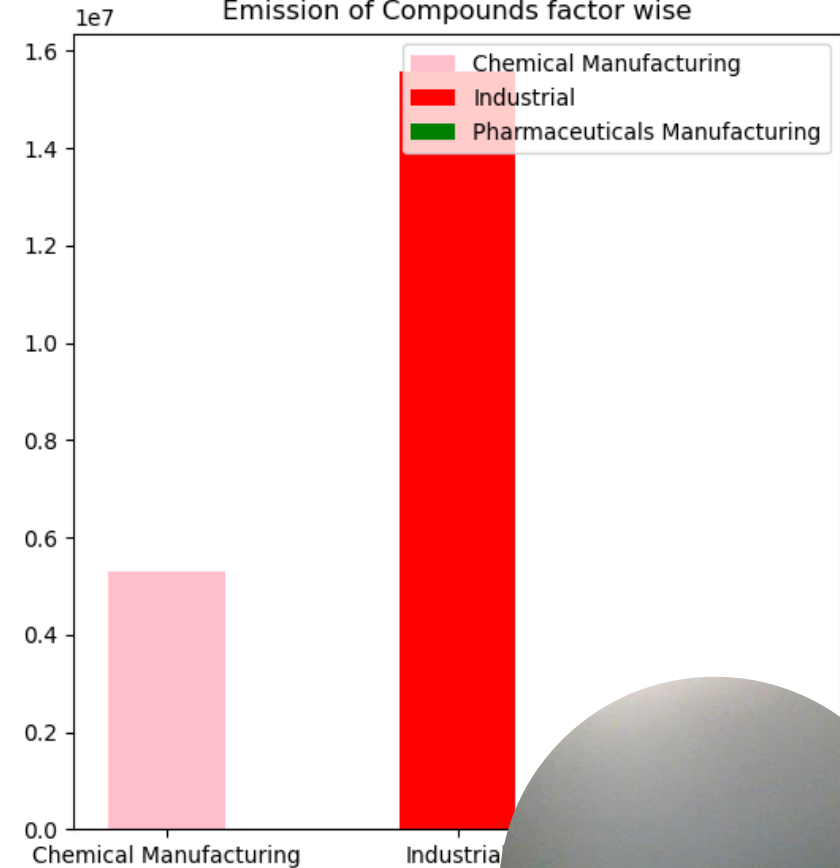
Results

Analysis Illustration

Heavy elements emission in water over the years 2000-2015



Emission of Compounds factor wise

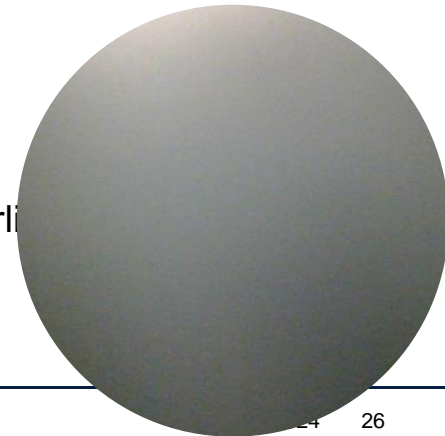


- ❑ Heavy elements emission into water has been reduced and thus yield of vegetable increased significantly.

Conclusion



- ❑ Our hypothesis was supported.
- ❑ Too much nitrogen and phosphorous can be harmful.
- ❑ From year 2000 to 2015: Phosphorus had a 300% decreases.
- ❑ Nitrogen was reduced to 2775 kilograms till year 2015 and it is 237% decrease
- ❑ Mostly Industries are emitter of heavy elements waste into water.
- ❑ Mercury was never emitted from Pharmaceuticals Manufacturing companies.
- ❑ Onion/Garlic. Tuberous and Apple → the most produced in Netherland.
- ❑ Nitrogen, Chromium based compounds has a huge correlation with Onion/Garlic



Thank you for your attention.

