Python Applications

Tan Sein Jone

University of British Columbia

July 26, 2024

Table of contents

1. Data Preview
2. Data Wrangler
3. Data Manipulation
4. Cleaning
5. Visualization
6. Task

Table of Contents

1. Data Preview

2. Data Wrangler

3. Data Manipulation

4. Cleaning

5. Visualization

6. Task

Pollution Data

- Synthetic data on pollution levels in a province by industry and year
- Data is in a csv file
- Relatively long dataset

Table of Contents

1. Data Preview
2. Data Wrangler
3. Data Manipulation
4. Cleaning
5. Visualization
6. Task

Data Wrangler

- Data Wrangler is an extension in VS Code that allows you to preview data in a more readable format.
- It is useful for viewing data in a tabular format.
- It provides important summary statistics for the data.

Data Wrangler

- You can access it in a couple of ways in VS Code.
- Open a csv file and in the top right corner, click on the icon that looks like a table.
- You can also right-click on a csv file and select "Show in Data Wrangler".

- When you load in data to the python kernel in VS Code, you can also view the data in Data Wrangler.
- In the bottom tool bar, you should see a tab for Jupyter.
- You should be able to see a button to the left of each variable that allows you to view the data in Data Wrangler.

Table of Contents

1. Data Preview

2. Data Wrangler

3. Data Manipulation

4. Cleaning

5. Visualization

6. Task

Data Manipulation

- Select columns and rows
- Get pollution numbers for BC

Data Manipulation

- Filter data
- Filter out pollution data post 2000
- Name the highest polluting sector for every province in every year

Data Manipulation

- Filter and Group data
- Total pollution for every province sector pair
- Get total pollution by province
- Group pollution into pre 2000 and post 2000

Data Manipulation

- Merge data
- Map full names to province abbreviations

Data Manipulation

- Applying functions
- Assume you want to change pollution values to logs for all columns except 'Year' and 'Province'
- Getting Total pollution for every sector
- Get a summary of pollution for every province

Table of Contents

1. Data Preview
2. Data Wrangler
3. Data Manipulation
4. Cleaning
5. Visualization
6. Task

Cleaning

- Handling duplicates
- Handling missing values

Table of Contents

1. Data Preview

2. Data Wrangler

3. Data Manipulation

4. Cleaning

5. Visualization

6. Task

Visualization

- Using matplotlib
- Plot a line graph of pollution by sector for British Columbia over time

Visualization

- Using seaborn
- Total emissions over time by province

Visualization

- Heatmap

Table of Contents

1. Data Preview

2. Data Wrangler

3. Data Manipulation

4. Cleaning

5. Visualization

6. Task

Task

- Plot emissions over time by sector
- Track the total pollution of the top 5 polluters in the year 2000