Module 2 - Session 1 - Data exploration

Working effectively with data

CivicDataLab

2021/08/03 (updated: 2021-08-04)

Past sessions

Session	Topic	Date	Slide Deck	Session Recording	Notes
1	Data Collection	16/07/21			
2	Web scraping & Data Standards	26/07/21			

Thanks for sharing your feedback and suggestions with us. If you haven't got a chance to do it yet, the form is still open here -> https://forms.gle/6pzx8XKBZ7YUGV336

Data exploration Use-Cases

Exploring state level mortality data

Dataset - Link

Source: Devdatalab

Objective:

- Import data in google sheets
- Compare mortality data across states from 2019 - 2021
- Compare mortality data for Karnataka across years, as shown here

Tags

google-sheets *data-viz* *pivot* *data-structure*

Most cited IPC acts/sections across courts

Dataset - Link

Source: Contributed by IndianKanoon on the Justice Hub

Objective:

- Import data in google sheets
- Filter data for IPC sections
- Find how many sections contribute to 50% of the cases
- Add a column for the section title

Tags

google-sheets *functions* *lookup* *scraping* *XML* *JSoN*

Working with databases

- Dealing with large datasets
- Platform agnostic
- Programming language agnostic
- Easy to share and maintain as compared to multiple data files

Exploring data from eCourts

Dataset - Link - *The database contains 81.2 million cases*

Source: Devdatalab

Objective:

- Understand how the data is structured
- Import the data in a database
- Explore the sample datasets
- Find out the total cases present for each district for the year 2018

Tags

database *large-datasets* *sqlite* *eCourts*

Tools

1. Sublime Text - *For text processing*
2. SQLite Browser - *For working with database*
3. JSon Editor - *For editing/viewing JSoN files*
4. XML to JSON - *For converting XML to JSON*
5. jq - *For working with JSON files*
6. Agenty - *Chrome extension for scraping data*
7. CSVLint - *For validating CSV files*

Do try this at home

civic data lab

Exercise - 1

- Link to NALSA dashboard
- Create a CSV file with variables available under the Victim Compensation Schemes table for these states:
  - Delhi
  - Maharashtra
  - Karnataka
  - West Bengal
  - Uttar Pradesh
- Create a chart to compare the yearly compensation numbers between these states
- Create a folder here and upload the dataset (including the chart)

Exercise - 2

- Install SQLite DB Browser
- Create a new database
- Load the judges_clean dataset in the DB
- Find the distribution of male/female judges in **Bengaluru** district court where judge position is *chief metropolitan magistrate*
- Save the file, as CSV, in the drive

Other Resources

- Working with SQLite
- SQL Tutorial
- Databases and SQL practice