Cleaned Data of All Students

Description

A dataset that takes schools from every Washington state school districts and reports statistics like test scores, student growth, geographic location, and student median income.

Keywords

Washington State   Education   Inequality   socioeconomic

public schools

How to use it?

👍 Intended Use

- **Intended Domain.** NA
- **Intended Use.** To analyze educational inequality in Washington state
- **Other Responsible Uses.** No

⚠ Restrictions on Use

- no

☑ Known Uses

- **Analyze inequality from county to county in Washington state.**
  https://data.wa.gov/Education/Report-Card-Growth-from-2014-15-to-Current-Year/ufi5-ki2f

ⓘ Do Not Use

- **Domain.** NA
- **Do not use to analyze individuals or schools outside of Washington state.** because data for those tasks are not represented

About the dataset

People

Owned by	Maxx Ferrian, Arnav Khare ⌄
Created by	Maxx Ferrian, Arnav Khare & 2 others ⌄
Maintained by	Arnav Khare, Maxx Ferrian ⌄
Funding	Washington State ⌄
Management	Washington State ⌄

Useful links

Dataset access point ⌃
https://github.com/ChillarneyMan/INFO-201-Final/blob/main/CleanedDataOfAllStudents.csv

A GitHub repository that has the data ⌃
https://github.com/ChillarneyMan/INFO-201-Final/blob/main/CleanedDataOfAllStudents.csv

Technical information

Publish Date	2023-11-23
Format	csv
Instances	2654
Version	Original
License	NA
Collection timeframe	2015-2019
Collection process	⌃

Merging three datasets and cleaning them by removing unnecessary values, missing values, and repetition. Also created filtered columns based on our given data needs.

Inference risks

At a Glance ⌄

	About humans	Upstream sources	Technical review	Ethical review	Update frequency
	No	Yes	Unsure	Unsure	Unsure

▦ Data values ⌄
What values are in each column?

Collection and Labeling Protocols	Data Imputation Protocols
Data Manipulation Protocols	Missing Data
Raw Data	

Number of issues

- Risky — 0
- Safe — 2
- Unknown — 3

▥ Feature selection ⌄
Which columns were chosen and why?

Cultural or Domain Assumptions	Proxy Characteristics
Planning Representation	Domain Knowledge

Number of issues

- Risky — 1
- Safe — 1
- Unknown — 2

⚇ Representation ⌄
Which rows were included and why?

Subpopulation Information	Representation
Individual Inferences	Individual Inferences - Mitigation
Collection Representation	Other Representation Issues

Number of issues

- Risky — 1
- Safe — 1
- Unknown — 4

⏰ General risks ⌄
Any additional risks?

Individual Information	Consent
Generalized Inferences	Generalized Inferences - Mitigation
Sensitive Content	Documented Known Issues
Other Known Issues	

Number of issues

- Risky — 0
- Safe — 1
- Unknown — 6