Project: Income in Urban vs Rural areas in the US

Data Source

Data Source

This is an external data source collected by a federal government agency, the US Census Bureau. This survey is called The American Community Survey (ACS). As it is a government organization that has collected this, it is a trustworthy and unbiased source.

Data Collection

This is administrative data collected as a part of the US Census. As the US Census Bureau has compiled this information, one can estimate that it is complete.

At the same time, as the data is collected through surveys, it can also be called survey data. The information is collected manually through phone and in-person interviews, online surveys, and mail responses. There's a follow-up for non-responses, so the response rate is high.

This is an ongoing, yearly survey, so the time lag is there.

Data Contents

The data contains the recorded household income of people living in different US states and counties. It also includes the total population of each county. The data is divided by state, county, and whether the area is rural or urban.

Data Relevancy

The data is trustworthy and complete. It meets the requirements for the data set.

Why this data set?

It meets the necessary criteria, is from an open, reliable source, and is fairly recent, which was quite hard to find.

I am interested in working in finance or banking, so analyzing people's incomes sounds like an appropriate choice.

Data Profile

Data cleanliness

No missing values or duplicates were found.

There were some inconsistencies in the numerical columns, specifically “Median Household Income”:

	County	State	FIPS	State FIPS Code	County FIPS Code	Total Population	Median Household Income	Urban-Rural
1753	Esmeralda County	Nevada	32009	32	9	962	-666666666	Rural
2655	Kenedy County	Texas	48261	48	261	52	-666666666	Rural

The negative values were replaced with the average median household income.

The data set also included information from Puerto Rico. These rows were eliminated for the sake of clarity.

The column types were changed to make future analysis easier.

This is the updated data profile:

Changed Records			
Did not make any changes.			
Timeliness			
This data set is from 2023. That is rather timely.			
Summary Statistics			
(Numeric Columns)			
		Total Population	Median Household income
	Minimum	43	25425
	Maximum	9848406	178707
	Mean	105721,228	66072,4418

Data Limitations

The annual reports are based on a sample of the population (around 3.5 million households). This means that if a sampling error occurs, the estimation of the yearly demographics might be not accurate.

The data is collected manually, so there is a chance of misspellings happening. However, the Bureau uses error-checking algorithms to make sure the responses are accurate.

Unregistered immigrants are not likely to report their presence in the US, so they are often not accounted for.

The missing values are usually imputed with similar households in the same geographic area. That can lead to an additional margin of error.

Because the data involves money, there might be a chance that people embellish or devaluate their true means to give a specific impression to the government.

Data Ethics

The data aims to be objective. The collection process does not aim to exclude any groups of people, it is made to be very inclusive.

Data Questions

What is the difference in household income between people living in rural and urban counties?

Where do people earn more?

In which of the urban counties do people earn the most/least? In which of the rural counties people earn the most/least?

In which states the difference is the most noticeable? In which is it the least noticeable?

How does the population size affect the average income? Is there a difference between rural and urban areas?

Why do people earn more in certain areas?