

# Demographic Data Summarization

This repository contains an R function to summarize the distribution of demographic data by calculating counts and percentages based on specified categorical variables. The function is designed to be flexible and can handle a wide range of demographic datasets.

## Function Overview

```
{r} summarize_demographic_distribution(data, demographic_variable, status_column, group_column)
```

This function provides a summary of counts and percentages for each category within a demographic variable. It groups the data by the specified demographic variable, calculates counts for positive and negative statuses, and calculates their respective percentages for different groups.

### Arguments:

- **data** : A data frame containing the demographic dataset.
- **demographic\_variable** : A character string specifying the demographic variable to summarize (e.g., "district", "sex", "age\_group", etc.).
- **status\_column** : A character string specifying the column in the data frame that represents the status variable (e.g., "covidstatus").
- **group\_column** : A character string specifying the grouping variable for calculating counts and percentages (e.g., "village", "district", etc.).

### Returns:

- A data frame summarizing the counts and percentages for each category within the demographic variable, broken down by the status\_column and group\_column.

### Example:

```
```{r} # Example data data <- data.frame( covidstatus = c("positive", "negative", "positive", "negative"), village = c("village1", "village2", "village1", "village2"), district = c("District1", "District1", "District2", "District2") )
```

## Call the function

```
result <- summarize_demographic_distribution(data, "district", "covidstatus", "village")
```

## View the result

```
print(result)
```

### Expected Output:  
The function will return a summarized data frame that looks like this (values are just an example):

```{r}

|   | variable | district  | status   | Summary   |
|---|----------|-----------|----------|-----------|
| 1 | district | District1 | positive | 2 (66.7%) |
| 2 | district | District1 | negative | 1 (33.3%) |
| 3 | district | District2 | positive | 1 (50.0%) |
| 4 | district | District2 | negative | 1 (50.0%) |

# How It Works:

- 1. The function groups the data by the demographic variable (e.g., "district" ).
- 2. For each group, it calculates counts of positive and negative statuses within the status\_column .
- 3. Percentages are calculated based on the total count of each group.
- 4. The output includes counts along with their respective percentages.

# Installation

- 1. **Clone the repository:**

```
git clone https://github.com/your-username/your-repository.git
```
- 2. **Install necessary R packages:** The function uses the dplyr and tidyr packages. You can install them by running:

```
{r} install.packages(c("dplyr", "tidyr"))
```
- 3. **Source the function:** After cloning the repository, you can source the R script containing the function:

```
{r} source("summarize_demographic_distribution.R")
```

# How to Use:

- 1. **Prepare your data:** Make sure your data frame contains the necessary columns, such as (demographic) variable that you wish to summarize, a status column (e.g., "covidstatus" ), and a grouping column (e.g., "village" ).
- 2. **Call the function:** Use the summarize\_demographic\_distribution() function to generate the summary.

```
{r} summary <- summarize_demographic_distribution(data, "district", "pfstatus", "village") print(summary)
```
- 3. **Interpret the output:** The resulting data frame will contain the summary of counts and percentages for each category within the demographic variable.

# Example Dataset

Here's an example of how your data might look before using the function:

| covidstatus | village  | district  |
|-------------|----------|-----------|
| positive    | village1 | District1 |
| negative    | village2 | District1 |
| positive    | village1 | District2 |
| negative    | village2 | District2 |

# Contributing

Feel free to open issues, suggest features, or submit pull requests to contribute to this project! Contributions are always welcome.

# Reporting Issues:

If you find any bugs or issues with the function, please create an issue on the GitHub repository page, and we will try to address it as soon as possible.

# License

This project is licensed under the MIT License – see the LICENSE (LICENSE) file for details.