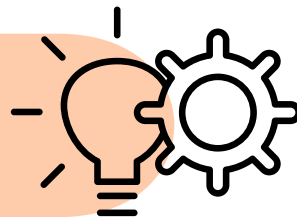


A DATASET GUIDE TO INFORMED DECISION-MAKING IN THE BAY AREA

BY DONA INAYYAH, CARLOS DE LEON, KARTIK PATEL, AKEMI SMART

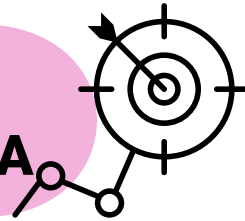
I. INTRODUCTION



II. UNDERSTANDING DATA



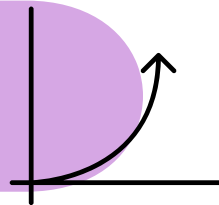
III. IDENTIFYING LOCAL ISSUES IN THE BAY AREA



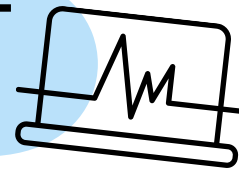
IV. EXPLORING RESEARCH QUESTIONS



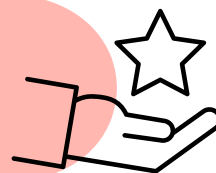
V. FINDING DATA



VI. CHALLENGES YOU MIGHT FACE



VII. TIPS & ADVICE



INTRODUCTION

Goal: To encourage you to **leverage data for informed decision-making**, and contribute valuable insights to address local issues. This guide is specifically focused on issues relevant to the Bay Area in California, but it is extendable and adaptable to any community of your choice.

UNDERSTANDING DATA

1.1. The Essence of Data:

- **Clearer understanding of their environment**, enabling easier navigation of challenges.
- Powerful tool for **monitoring the consequences of decisions**.
- **Empowers forecasting of future events** with confidence by examining past data and present trends.

1.2. Overview of Types of Data:

- **Quantitative Data:** Represented by statistics, counts, and measures, quantitative data is **objective and measurable**.
- **Qualitative Data:** Revolves around observations, open-ended questionnaires, interviews, and **descriptive information**.

IDENTIFYING LOCAL ISSUES IN THE BAY AREA

The San Francisco Bay Area, renowned for its diverse communities, thriving tech industry, and picturesque landscapes, boasts one of the world's strongest economies. However, this prosperity has brought forth a myriad of pressing challenges that significantly impact the community's well-being and overall quality of life. **Here are some key issues:**

2.1. Housing Crisis and Homelessness

- Homeownership is virtually unattainable, and SF remains the third most expensive city for renting in the US.
- Individuals are compelled into substandard living conditions, lengthy commutes, and homelessness.



2.2. Transportation

- Unprecedented levels of freeway congestion and historic overcrowding on transit systems like BART, Caltrain, and Muni.
- Transit systems operate at or beyond their capacities, leading to environmental impacts, lost time, and heightened costs.

2.3. Crime Rates

- Surge in crime, such as car break-ins, robberies, homicides & the Fentanyl epidemic.
- Safety and security of residents are impacted.
- Major retailers like Target and Walgreens have reported closures and reduced hours.

References:

- [Has the cost of living in the Bay Area always been this high?](#)
- [The ongoing crisis of homelessness in the Bay Area: What's working, what's not](#)
- [Plan Bay Area 2040-Final Plan.](#)

EXPLORING RESEARCH QUESTIONS

Here are some **sample research questions** to consider, and where to find them:

3.1. Reliable & Reputable Sources

- DataSF: <https://datasf.org/opendata/>
- The U.S. Census: <https://data.census.gov/>
- World Health Organization (WHO): <https://www.who.int/data>
- UNICEF: <https://data.unicef.org/resources/resource-type/datasets/>

3.2.1. Housing Crisis and Homelessness:

- What is the correlation between population density and homelessness rates?
- How does the GDP of a region correlate with its homelessness statistics?
- Is there a relationship between median salary levels and homelessness rates?

3.2.2. Transportation:

- How has commute time changed over the years due to economic and housing factors?
- What are the peak hours and peak destinations for transit systems like BART, Caltrain, and Muni?
- Is there a correlation between housing affordability and transportation challenges?

3.2.3. Crime Rates:

- What are the crime rate variations between different neighborhoods in the Bay Area?
- How do socioeconomic factors influence crime rates in the region?
- Is there a correlation between certain types of crime and specific demographic factors?

FINDING DATA

Here is a walkthrough on finding relevant data:

Example Research Question:

Is there a relationship between median salary levels and homelessness rates?

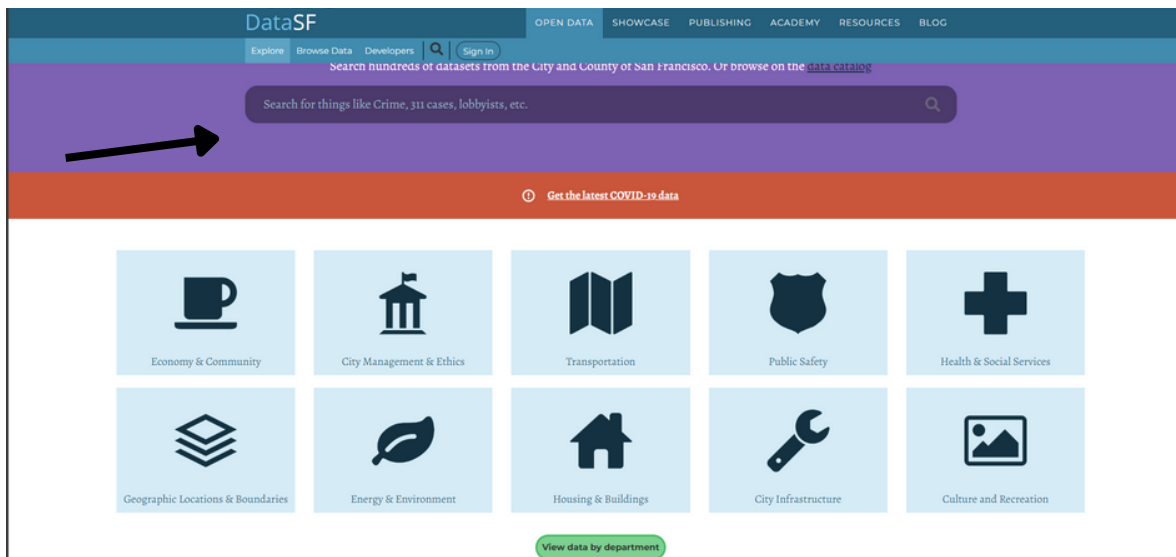
4.1. Identify Keywords:

"median salary"
"homelessness rates"

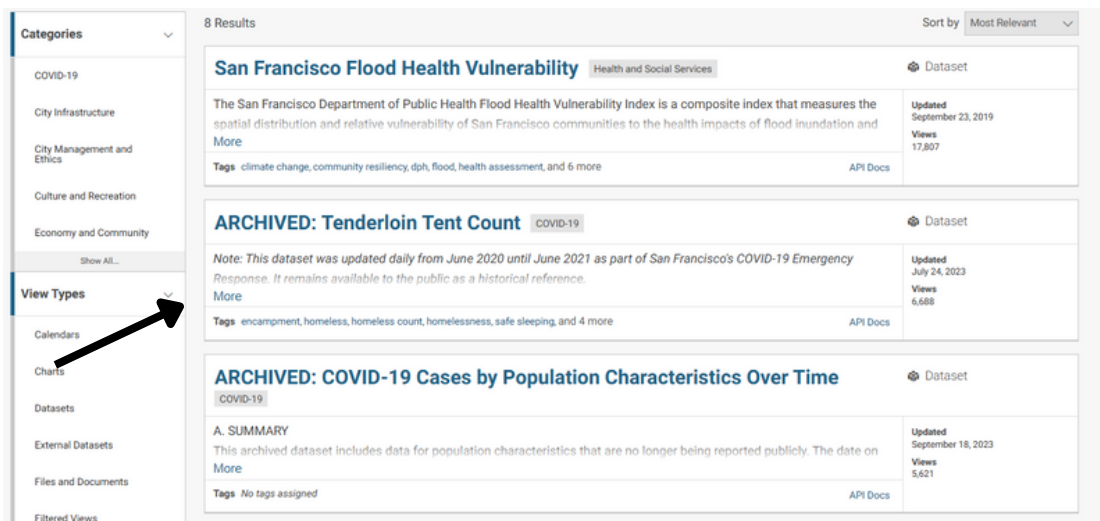


To find the dataset for homelessness in the Tenderloin district of San Francisco, we will be using data SF.

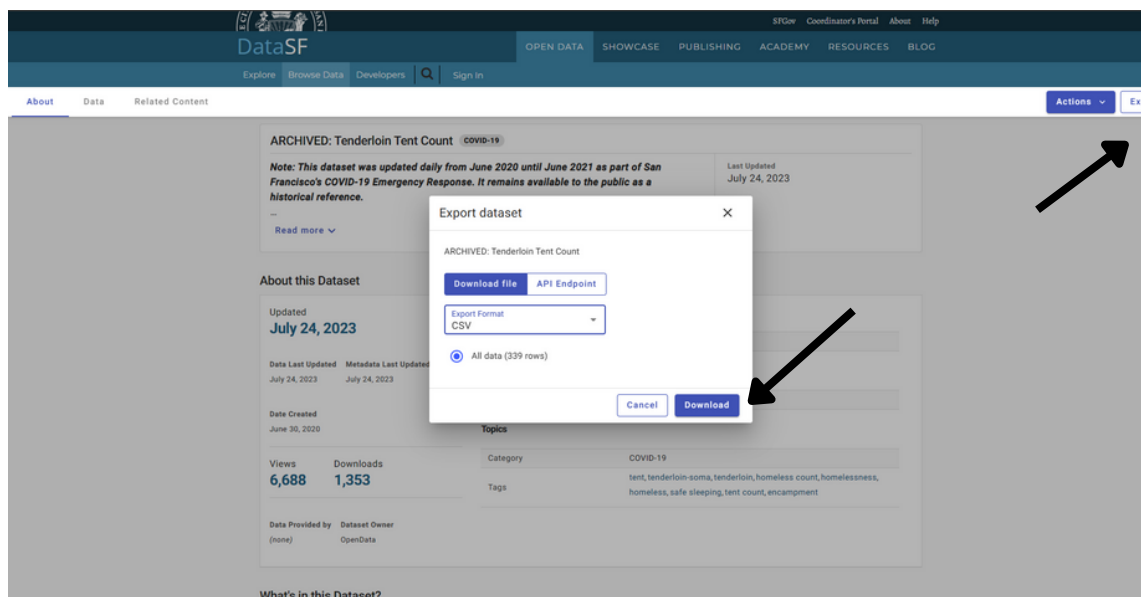
Step 1: Go to DataSF and search for the Keywords of your chosen topic.



Step 2: Find the dataset from the results that will help in your study. You can also rely on the tags that DataSF provides in order to better understand the data that is in the dataset.



Step 3: Download the Data set if it fits your criteria. There are also multiple options for the file choice such as CSV, XML, RSS, RDF



CHALLENGES YOU MAY FACE

Here are some **things to consider** when working with the key issues mentioned above:

- Defining and measuring homelessness is challenging since factors such as temporary housing solutions and unreported cases contribute to the complexity.
- Obtaining accurate and real-time transportation data can be challenging, considering the multitude of variables influencing commute patterns and transit usage.
- We have to consider reporting biases, changing legal definitions, and evolving societal perceptions of criminal activities when working with crime-related data.

TIPS & ADVICE

- **Be intrigued and follow your interests**, for a more engaging analysis.
- **Stay Persistent.** Locating relevant data can be challenging, but don't get disheartened. Alternative sources may hold valuable information.
- **Stay Open-Minded.** Your findings are dynamic, but not set in stone. Always explore alternative explanations (just like what we did above).
- **Think double E's.** Ethical and Empathetic considerations should be the compass of your data analysis journey.
- **Correlation does not imply Causation.** Just take a look at this...
 1. Sour Cream Consumption and Motorcycle Deaths
 2. Sour Cream Consumption and Motorcycle DeathsA correlation means causation argument might be more believable if the title was "attempting to consume sour cream *while* riding a motorcycle."

