**Santa Clara County Suicide Data Analysis Project 2018-2025**

**Project Overview**

This project analyzes suicide death data from Santa Clara County, California, spanning from January 1, 2018 to present. The analysis aims to identify patterns and trends that could inform suicide prevention efforts and mental health interventions in the community.

**Data Source:** [Santa Clara County Medical Examiner-Coroner Suicide Deaths Dataset](https://data.sccgov.org/Health/Medical-Examiner-Coroner-Suicide-Deaths-dataset/aw9t-m5w4/about_data)

**Project Motivation**

This analysis was conducted in response to potential federal budget cuts (HR1) that could impact crisis mobile services in Santa Clara County. The findings were shared with Supervisor Ellenberg's office following her community address on October 6th, 2025.

**Technology Stack**

* **Python 3.11.9**
* **Polars 1.32.2** - High-performance DataFrame library for data analysis
* **Jupyter Notebook** - Interactive development environment

**Key Findings**

**Demographics**

* **Total Records Analyzed:** 1,223 suicide deaths (2018-2025)
* **Average Age:** 48 years old
* **Most Affected Demographics:**
  + White individuals: 596 cases (48.73%), average age 54.4
  + Asian individuals: 297 cases (24.28%), average age 45.6
  + Hispanic/Latino individuals: 235 cases (19.22%), average age 38.4

**Geographic Distribution**

* **San Jose** leads with the highest number of cases (577 total)
* 82 cases had no known resident zip code (listed as N/A)
* **Surprising finding:** Only 2 of 82 cases with N/A residence listed "Depression" as a contributing factor

**Contributing Factors**

* **68% (829 cases) had NO listed contributing factors** in the medical examiner report
* Only 30% (394 cases) had documented "Other Significant Conditions"
* When conditions were documented, common themes included:
  + Depression and mood disorders (most prevalent)
  + Bipolar disorder
  + Anxiety disorders
  + Schizophrenia and schizoaffective disorder
  + Substance use disorders
  + Chronic physical health conditions
  + PTSD

**Temporal Trends**

* 2018: 150 deaths
* 2019: 169 deaths
* 2020: 166 deaths
* 2021: 154 deaths
* 2022: 181 deaths (peak year)
* 2023: 138 deaths
* 2024: 168 deaths
* 2025: 97 deaths (partial year)

**Analysis Scripts**

The Jupyter notebook (SC\_Coroner\_Analysis.ipynb) contains multiple analysis cells that generate:

1. **Resident Zip Code Analysis** - Identifies geographic distribution and residence status
2. **Race Statistics Report** - Demographic breakdown with percentages and average ages
3. **Temporal Analysis** - Year-by-year death counts
4. **Significant Conditions Analysis** - Frequency analysis of documented contributing factors
5. **Depression-Specific Analysis** - Filtered datasets examining depression as a factor

**Generated Reports**

The analysis produces several text and CSV output files:

* resident\_zip\_counts.txt - Zip code distribution with city names
* race\_statistics.txt - Detailed demographic statistics
* year\_death\_counts.txt - Annual suicide counts with totals
* significant\_conditions\_report.txt - Contributing factors with percentages
* condition\_counts.csv - Structured data on all documented conditions
* resident\_city.csv - Cases with listed residence and depression
* no\_depression\_resident\_city.csv - Cases with residence but no depression listed
* no\_resident\_city.csv - Cases without listed residence

**Critical Insights**

**The 68% Mystery**

The most significant finding is that **68% of suicide victims had no documented contributing factors**. This raises important questions:

* Are people hiding depression and not seeking treatment?
* Are we missing opportunities for early intervention?
* Could home-based outreach be more effective?

**Intergenerational Risk**

Research indicates that children who lose a parent to suicide are:

* **3x more likely** to die by suicide
* **2x more likely** to attempt suicide (Source cited: York University trauma research blog)

**Community Impact Projection**

Using the Center for Suicide Prevention's estimate that each suicide affects 135 people:

* If current trends continue over the next decade: ~2,400 suicides expected
* **324,000 county residents would be affected** (approximately 1 in 6 people)

**Recommendations**

1. **Enhanced Data Collection** - Track family history of suicide for better pattern identification
2. **Community Outreach** - Coordinate programs across cities with proven messaging strategies
3. **Early Intervention** - Develop home-based support systems alongside 988 crisis line
4. **Public Awareness** - Use visible messaging ("You Matter" campaigns, 988 number displays)
5. **QPR Training** - Expand Question, Persuade, Refer suicide prevention training
6. **Protect Funding** - Maintain mental health crisis services despite federal budget pressures

**Usage**

To run the analysis:

import polars as pl

# Load the data

df = pl.read\_csv("coroner\_data.csv")

# Run analysis cells in the Jupyter notebook

# Each cell generates specific reports and visualizations

**Data Privacy & Ethics**

This analysis uses publicly available de-identified data from the Santa Clara County Medical Examiner-Coroner's office. No personally identifying information is included in this repository.

**Contact & Collaboration**

This project was developed to inform policy decisions and community mental health initiatives in Santa Clara County. For questions or collaboration opportunities, please reach out through official county channels.

**Note:** This analysis represents a critical public health issue. If you or someone you know is experiencing suicidal thoughts, please call or text 988 for the Suicide & Crisis Lifeline.