**Team Mars Crime Data Analysis Project**

**Team Members**  
Marzia Fahimi  
Allison Ramirez  
Anna Amosi  
Sara Hail

**Course**  
Advanced-Data Analytics  
Prof. Soumyakant Padhee

**Project Task Plan**

**1. Description of the Data**

The dataset includes data on violent crime rates in California from 2000 to 2013, sourced from the FBI and other law enforcement agencies. It tracks the number of violent crimes per 1,000 people across various racial/ethnic groups and geographic regions, including county and MPO (Metropolitan Planning Organization) regions.

**Key Variables:**

* **report year**: Year of the recorded data.
* **race\_eth\_code / race\_eth\_name**: Codes and names representing racial/ethnic groups.
* **Geo type, geo type value, geo name**: Geographic identifiers, such as county names and MPO regions.
* **rate**: Crime rate per 1,000 population.
* **dof\_population**: Population count used for rate calculation.

**Limitations**: The data is limited to California, restricting generalizability to other states. Some data may be missing, and certain variables are available only for specific years. Additionally, external socio-economic factors influencing crime trends aren’t included, limiting deeper causal analysis.

**2. Research Questions**

Our analysis will address the following main research questions:

**Q1**: How have violent crime rates in California evolved across different racial/ethnic groups and regions from 2000 to 2013?

**Q2**: How do regional factors and the type of crime influence variations in violent crime rates across different areas?

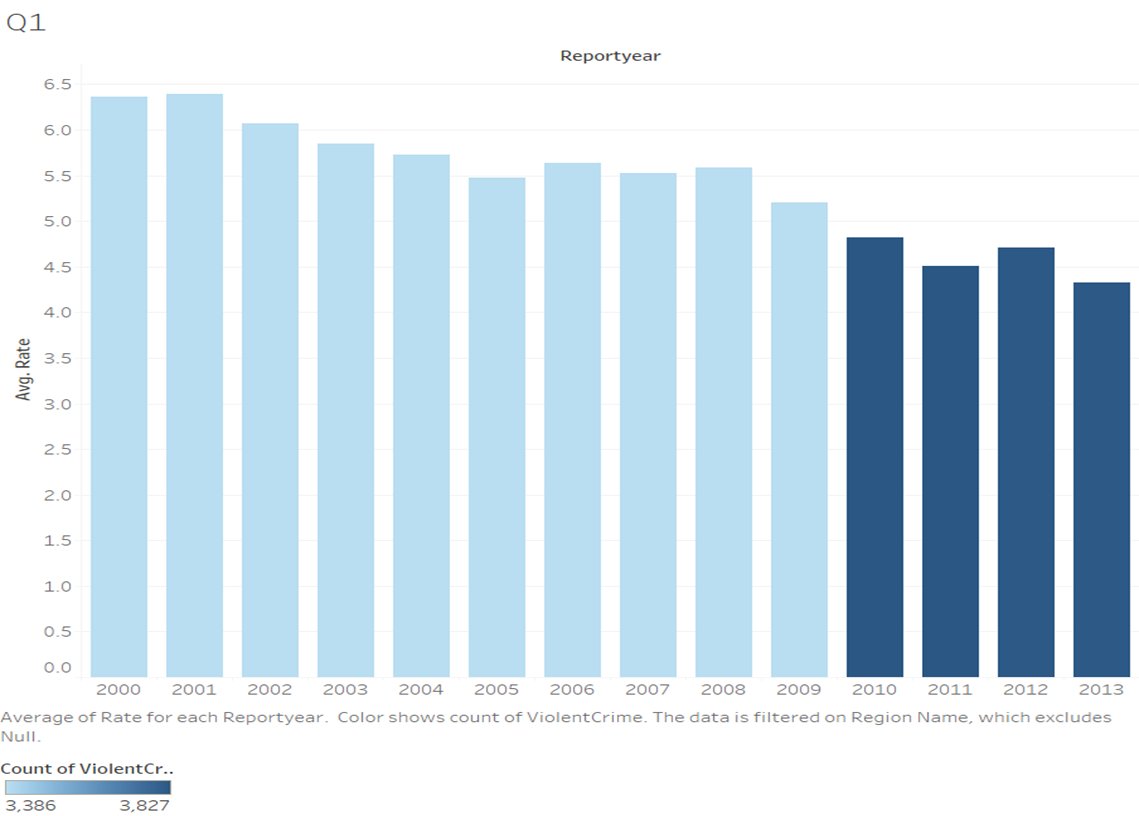
**Q3**: Are there specific MPO regions or racial/ethnic groups that experience consistently higher or lower crime rates?

**3. Pathways and Sub-Questions**

To explore these questions comprehensively, we’ve broken them down into targeted sub-questions, each with specific analytical pathways.

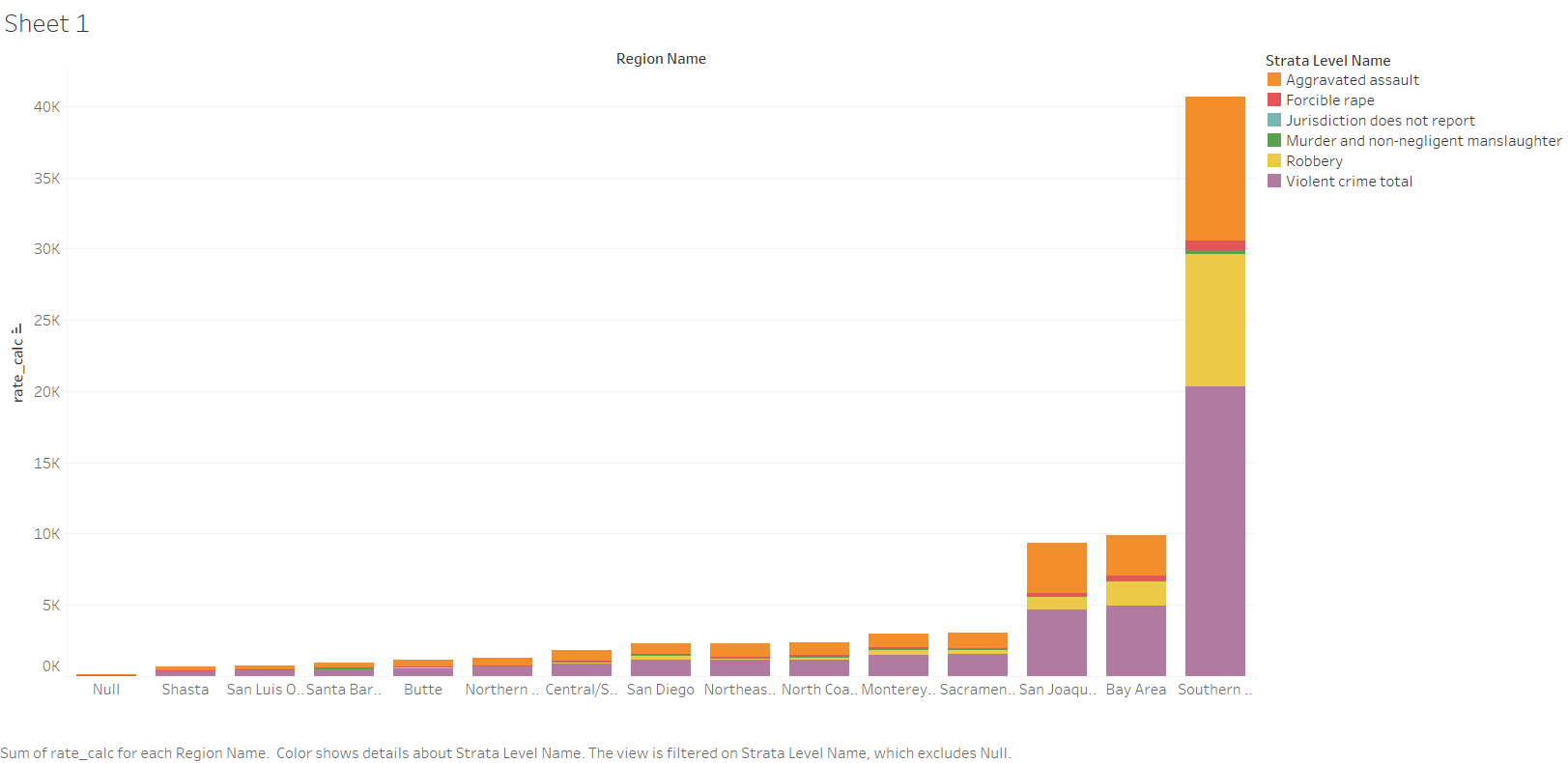
**Q1**: *How have violent crime rates in California evolved across different racial/ethnic groups and regions from 2000 to 2013?*

* **Sub-question 1.1**: What is the overall trend in violent crime rates over the dataset’s time period?
  + **Pathway**: Use time-series analysis to plot yearly crime rates.
* **Sub-question 1.2**: Are there significant differences in crime rates across racial/ethnic groups?
  + **Pathway**: Compare rates across racial/ethnic groups, using bar charts and trend lines to visualize disparities.
* **Sub-question 1.3**: Do particular regions show distinct trends?
  + **Pathway**: Aggregate data by MPO regions and visualize regional patterns.



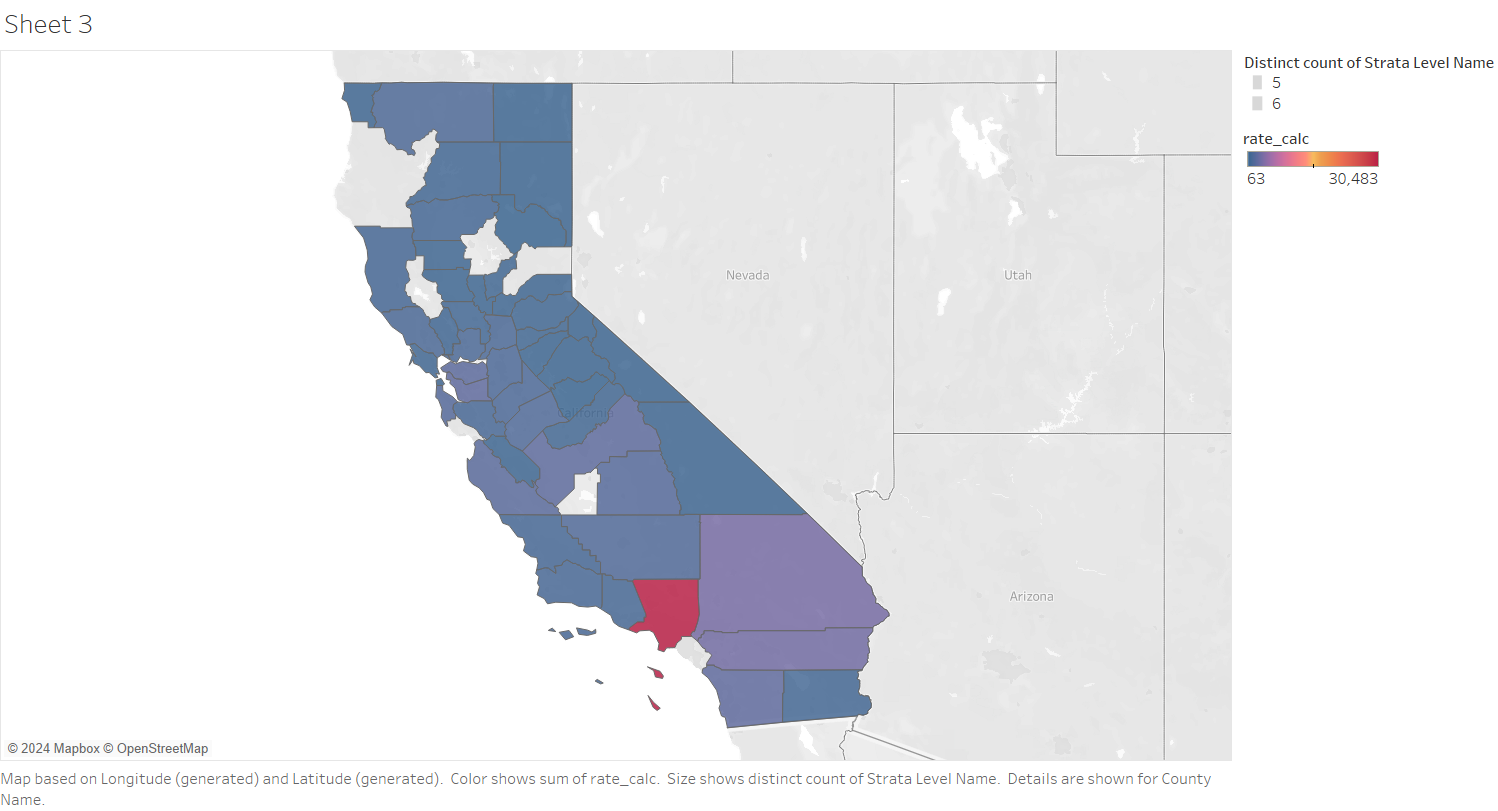
**Q2**: *What regional factors and demographic trends contribute to variations in violent crime rates?*

* **Sub-question 2.1**: How does population size impact crime rates across regions?
  + **Pathway**: Perform correlation analysis between population and crime rates.
* **Sub-question 2.2**: Do higher-density counties have consistently higher crime rates?
  + **Pathway**: Analyze crime rates by county and population density.
* **Sub-question 2.3**: What role does racial/ethnic composition play in regional crime patterns?
  + **Pathway**: Use regression analysis to examine racial demographics against crime rates.



**Q3**: *Are there specific MPO regions or racial/ethnic groups that experience consistently higher or lower crime rates?*

* **Sub-question 3.1**: Which regions have the highest average crime rates?
  + **Pathway**: Calculate average crime rates by MPO region and map these averages.
* **Sub-question 3.2**: How do crime rates differ within racial/ethnic groups across regions?
  + **Pathway**: Compare crime rates for racial/ethnic groups across MPO regions.
* **Sub-question 3.3**: Are there regions showing consistent improvement or worsening trends?
  + **Pathway**: Identify regions with notable trends over time using a time-series analysis.



**4. Timeline and Milestones**

**Week 1: Data Exploration**

* Load and clean the dataset.
* Conduct initial descriptive statistics.
* Filter data based on relevance for each research question.

**Week 2: Refinement of Research Questions**

* Finalize research questions and sub-questions.
* Structure data for time-series and regression analyses.

**Week 3: Analysis for Q1 (Trends in Crime Rates)**

* Perform time-series analysis and plot crime rate trends.
* Compare racial/ethnic group trends across the years.

**Week 4: Analysis for Q2 (Regional and Demographic Factors)**

* Conduct correlation and regression analyses.
* Map crime rates to visualize regional disparities.

**Week 5: Analysis for Q3 (Regional Consistency in Crime Rates)**

* Calculate regional averages and trends.
* Map trends to identify consistent patterns in crime rates.

**Week 6: Drafting the Final Report**

* Integrate findings, visualizations, and statistical insights into a cohesive report.

**Week 7: Presentation Preparation**

* Develop presentation slides summarizing key insights and visualizations.

**5. Deliverables**

* **Final Report**: A comprehensive analysis addressing all three research questions, with visualizations and data-driven conclusions.
* **Presentation**: Slide deck highlighting key insights, visualizations, and actionable recommendations for future policy planning.

Notes:

For each graph(question), make a story. For question 1 say why the crime rate went up/down through the year 200-2013- you can google it, make a story, find other resources such as news, articles, etc.

For question 2, check the type of crime that has a higher rate among others based in counties and tell why the rate differs in different counties. Again, build a story, based on other resources.

For question 3, tell why the northern part of California has a lower crime rate compared to the southern part ( Los Angeles = dark red color).

Dark red = extreme rate

Purple= moderate

Navy blue= the lowest