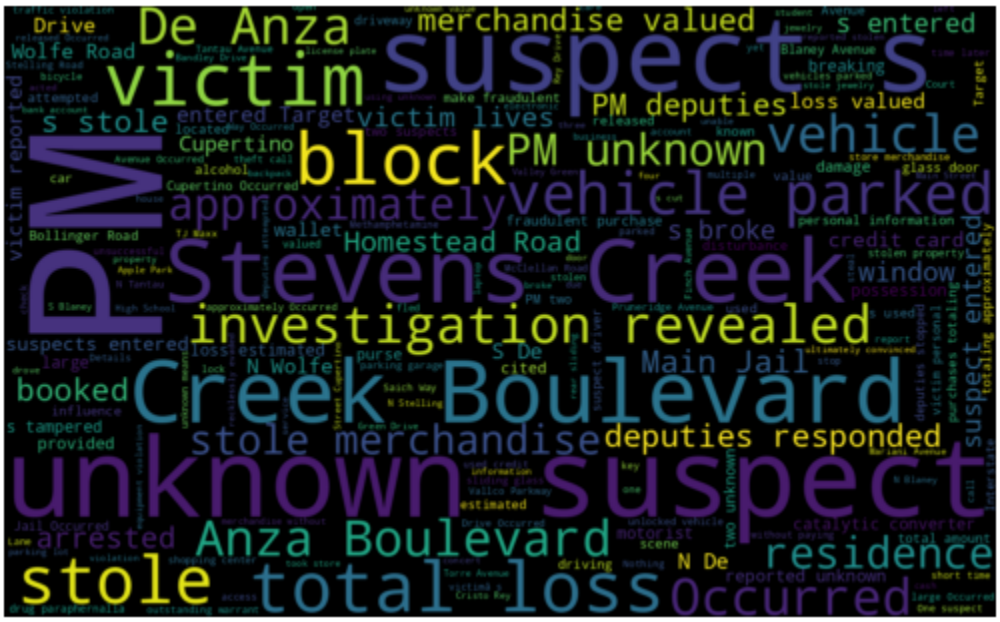
Code Sprint Participants: Arnav S, Anush V

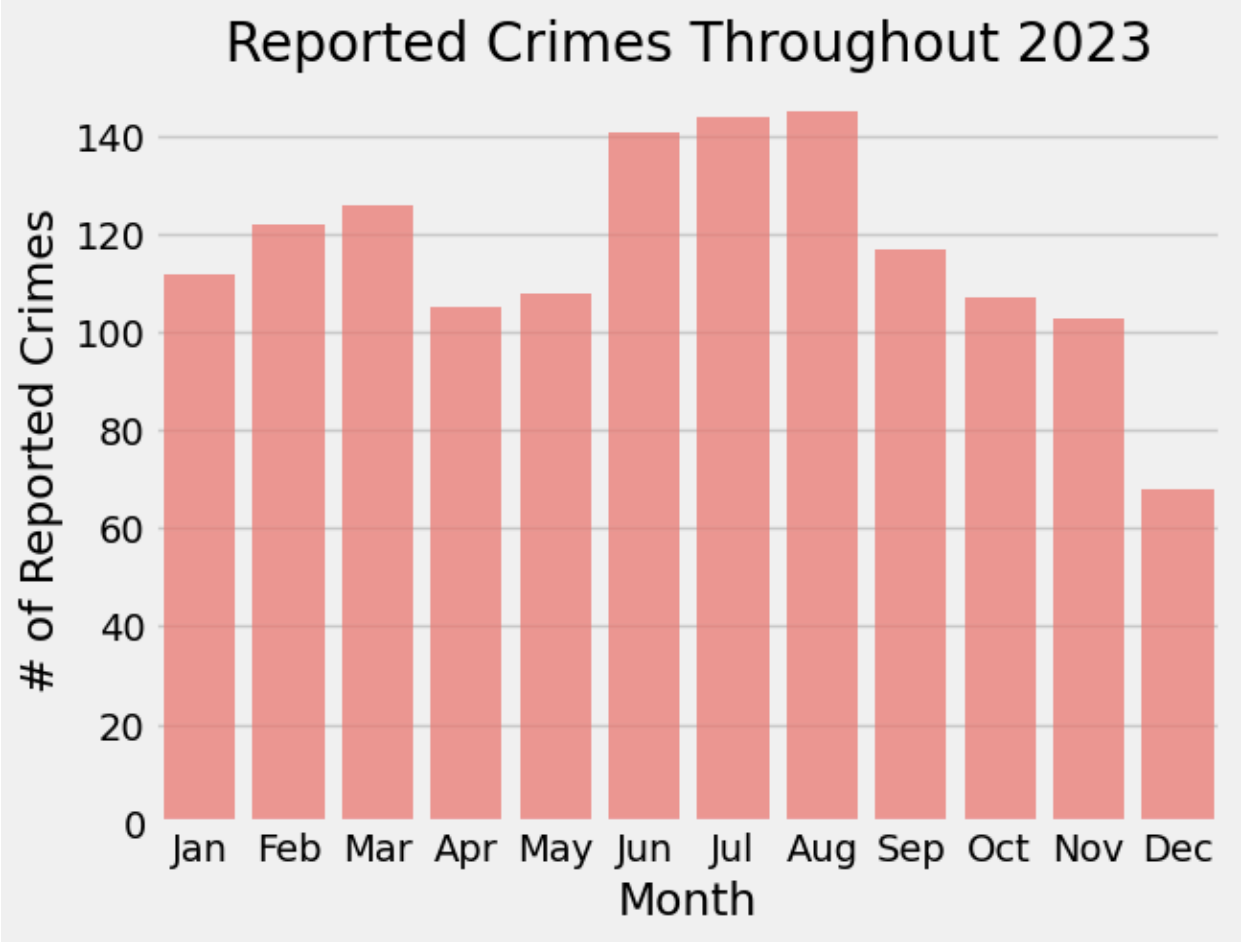
Summary: Every week in 2023, the Cupertino Police posted .doc files of weekly crime reports, including the location, date, and summary of each report in that week. This project involves using Python to convert these .doc files into a single pandas dataframe, constructing new features using existing ones, creating diagrams from these features, and analyzing them. Note that in the police department, Cupertino is divided into 9 sections as shown in this map.

Diagram 1: Word Cloud of all Crime Summaries



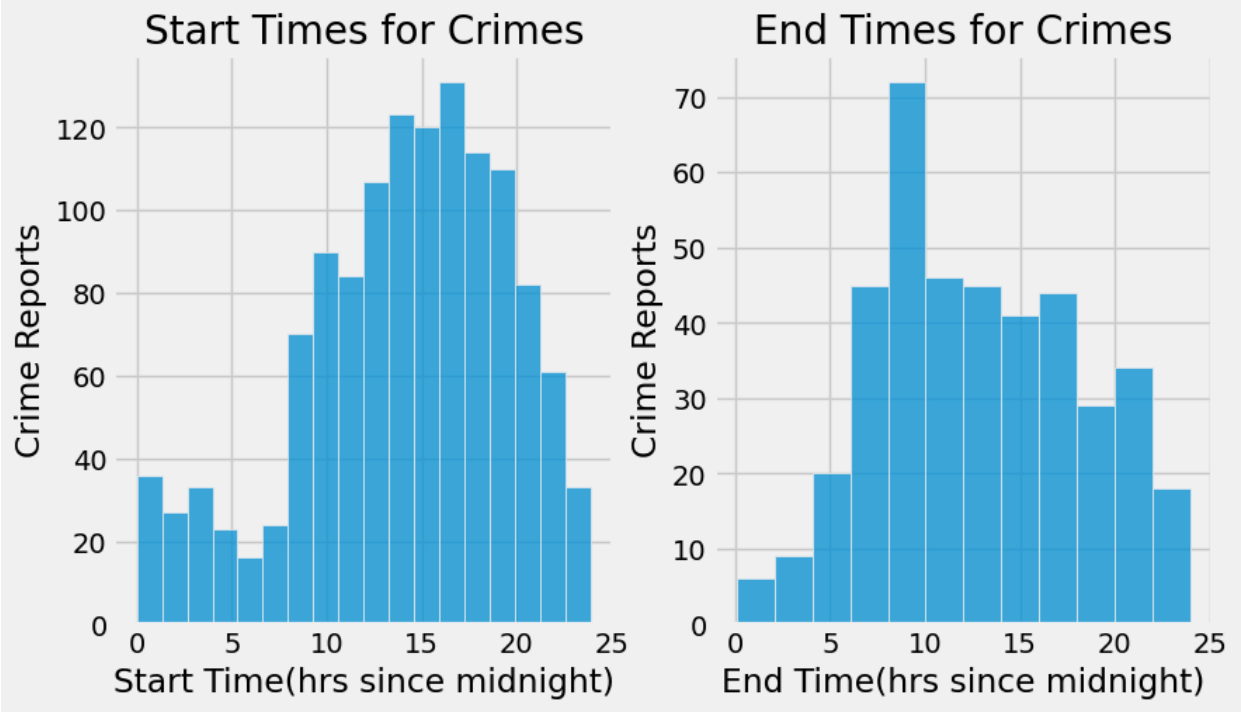
Analysis: This Word Cloud shows the most common words written in the crime report summaries throughout the entire year. With regards to location, both “Stevens Creek” and “Creek Boulevard” are highly present, showing how a lot of crimes are committed along this road. Words like “Vehicle” and “parked” are common, showing how a lot of crimes involve vehicles. Additionally, the words “stole” and “merchandise” being common show the prevalence of robbing in Cupertino.

Diagram 2: Number of Reported Crimes Every Month



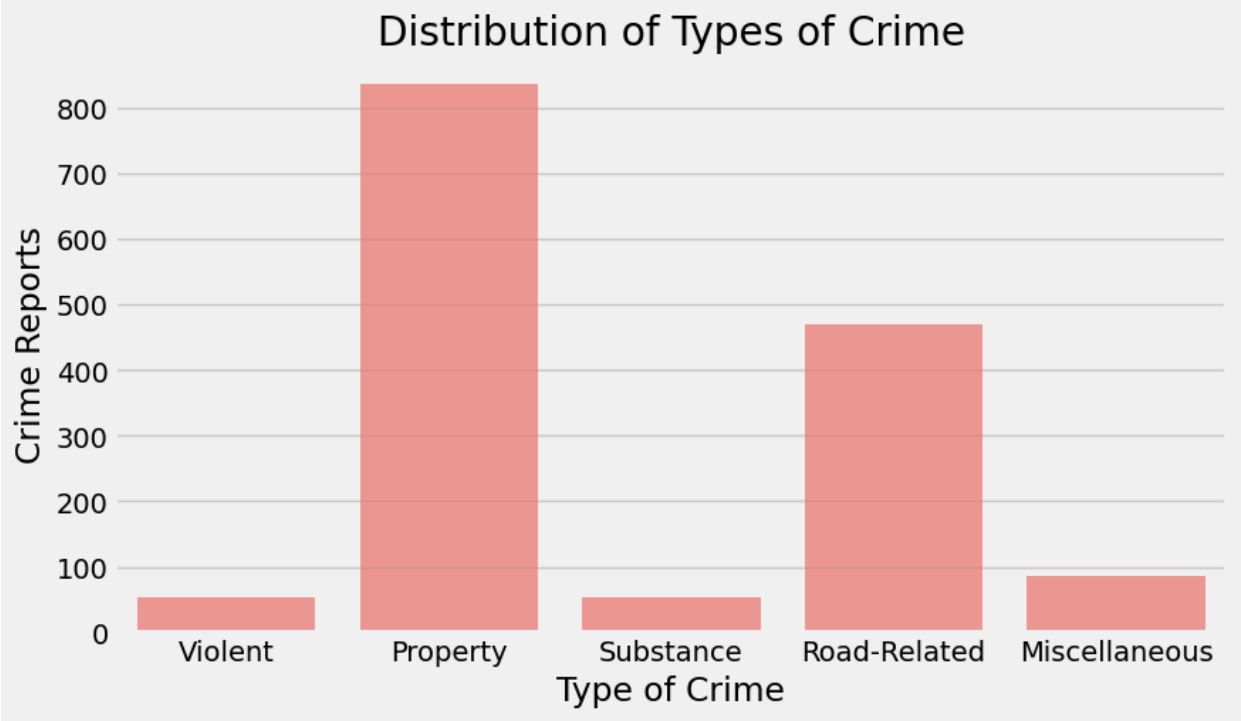
Analysis: We can see the correlation between weather conditions and crime rates, with the summer months of the year having a noticeable increase in crime over the winter months. The months of June, July and August have an average of around 145 crimes reported in contrast to the average throughout December, January and February of 100 crimes reported. This can possibly be associated with the higher temperature and harsher weather in the summer months along with the increase in tourist visits in the Summer leading to more people vulnerable to crime in this period.

Diagram 3: Start Time vs End Time for Crimes



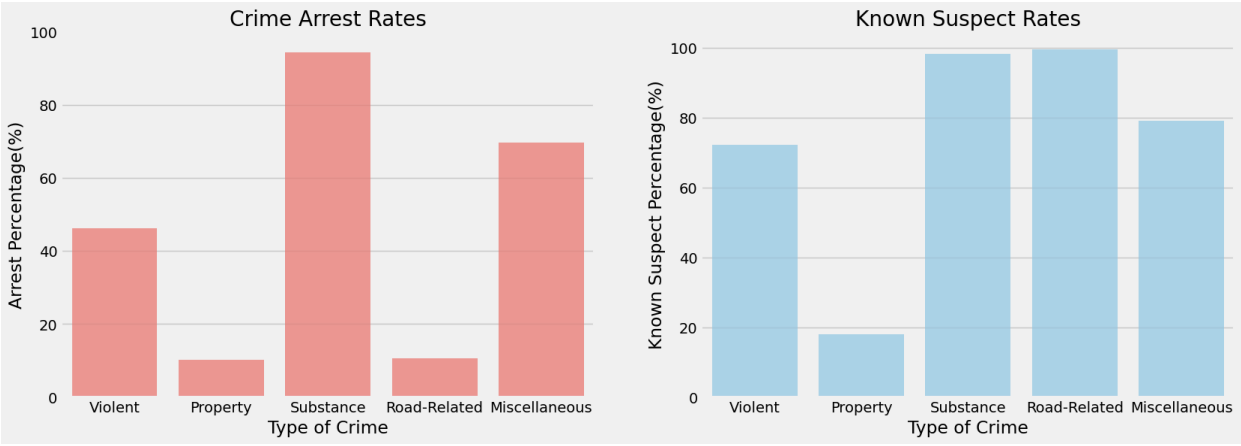
Analysis: Through these graphs we can see the use of darkness in order to conceal crimes from police: using the dark as camouflage can help people avoid the eyes of the public and get away with crimes such as murder and robbery. In addition, the peak crime ending time being at 8 hours after midnight could possibly be in association with the fact that most people come to work around 8 hours after midnight, leading to most crimes having to stop to avoid public attention and stay secretive.

Diagram 4: Prevalence of Different Crime Types



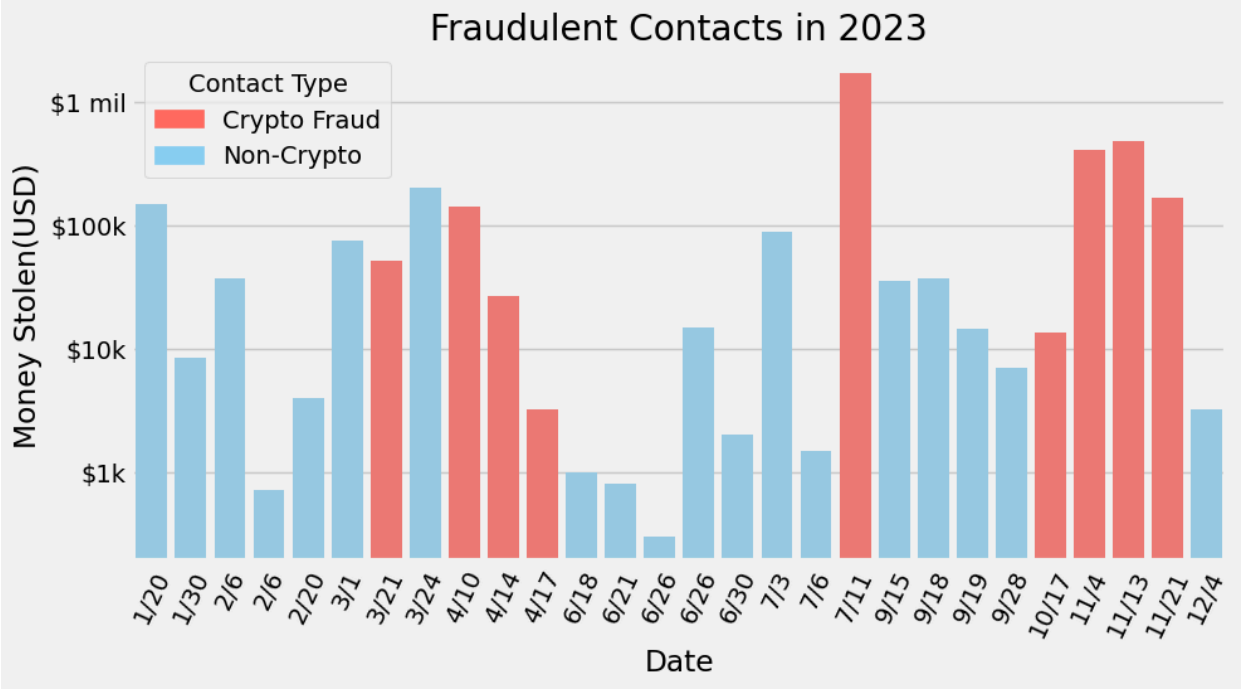
Analysis: The low amount of violent and substance crimes at 50 crimes reported each speaks to the average wealth of the city, with a lot of engineers and other high-paying jobs being employed by companies like Apple, for example. In addition, the prevalence of property related crimes come from the known high mortgage and property cost in Cupertino, with houses commonly going for millions of dollars, leading to exploitation by criminals in search of money.

Diagram 5: Rates of Arrests and Rates of Known Suspects for Crime Types



Analysis: The percentages of crimes that end up in arrests vastly depends on the type of crime involved. Notably, nearly all substance crimes end up in arrests, and around 40% of violent crimes end up in arrests, but both property and road-related crimes see very little arrest percentages. When looking at the percentage of each crime type in which the perpetrator is known, this data makes more sense. The known suspect rate for property crimes is very low, and for violent crimes it's around 70%: when accounting for this, we can see that the majority of suspects of both property and violent crimes are arrested. For substance crimes, the suspect is nearly always known, and is nearly always arrested. However, for road-related crimes, the suspect is always known, but the arrest rate is low, meaning that the end result of these crimes for the perpetrator is usually something other than arrest.

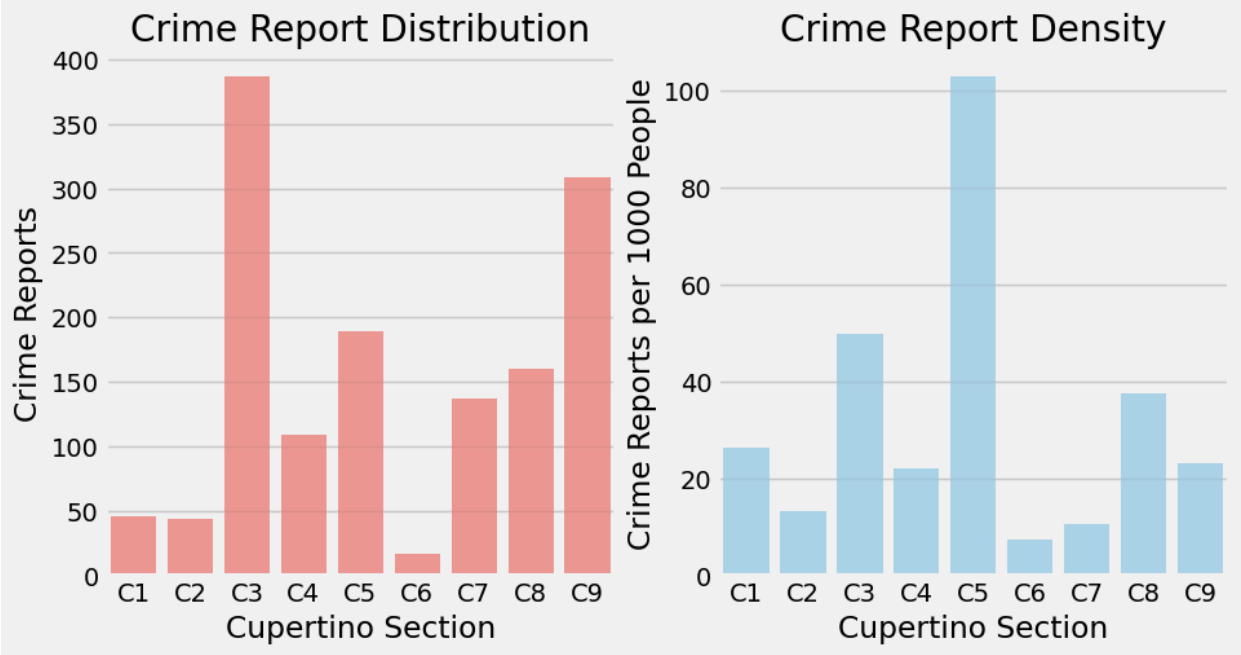
Diagram 6: All Instances of Fraudulent/Scam Contents in 2023



Analysis: Throughout 2023, there have been numerous fraudulent contacts affecting victims in Cupertino, sometimes involving massive sums of money. What's important to note in this graph is that the y-axis is log based, so what looks like a linear increase is actually an exponential increase, which shows the wide variety in money stolen with this type of crime, which can range from a few hundred dollars to over a million. There are also patterns of crypto frauds in particular in different pockets in time, such as in April

and in November, and these crypto frauds usually involve relatively high amounts of money stolen($3 million of the $3.6 million total stolen due to fraudulent contacts in Cupertino have been crypto, to be more specific).

Diagram 7: Crime Report Distribution & Density for Different Locations



Analysis: Through the graphs, we can see C3 and C9 having a much higher total crime report rate due to the high amount of population of people with around 7,700 people in section C3 and around 13,000 people for C9, whereas C7, despite having a similar population to C9, doesn't have such a high crime rate.To add on, C5 has a very high crime density, likely due to the amount of people coming there to work on Apple campus relative to the population.

Conclusion: Crime rates in Cupertino vastly vary with the type of crime, the location of the crime, the season/month of the crime, and the specific time of the crime. The analysis shown above gives more information of the specifics of these relationships, and more broadly gives information on the dynamics of crime in this city.