UCLA vs. USC vs. Criminals: How Crime Differs between the Two Universities in the LA Area

Tiffany Hu, Jade Liang, Bryan Mui, Clark (Xuanyao) Qian, Jeremy Reyes, Carolynn Rui, Zijia Zhang

Stats 140XP – Winter 2025

Professor Vivian Lew

Friday, March 14, 2024

Contents

1 Abstract

The primary question this report seeks to answer is if there is a significant difference in crime rates against college-aged individuals between the UCLA and USC areas. In addition, we'll be investigating if there's a difference in the significance of factors such as victim demographics, location of crime, etc. between the two universities. We analyzed a dataset of crimes during 2020 from the LAPD utilizing Chi-Square tests and found that there is a significant difference in crime rates between the schools. Furthermore, it appears that the sex and ethnicity of the victim are significantly different between them as well. However, this analysis is not exhaustive, and does not dive much deeper than checking for significant differences; there is potential for checking each individual factor, but due to time constraints, this was not fully explored.

2 Introduction

College life is often considered to be a right of passage for many young and adventurous people. The feeling of independence can open up a whole new world to the eye of the beholder. However, the unparalleled exhilaration of newfound liberty is tempered by the need to remain vigilant against potential crime. This is especially true for the students of UCLA and USC: the cross-town university rivals of LA. But while danger is a given, which one is more dangerous? Is there sufficient evidence to prove that one is more dangerous than the other? This is a question concerned parents or potential students might consider if they're worried about the dense urban environment LA has to offer, enough so that it could completely sway their college decision. Additionally, school administration should tailor their safety protocols in a way that's appropriate in order to protect students who could be more vulnerable or susceptible to becoming victims. Thus, an answer is in order, which we will provide via thorough analysis of a dataset of crime during 2020 provided by the LAPD. The need for caution may be influenced by the environment shaped by the university and its campus, making it essential to understand how crime differs between these two highly-regarded universities.

This research could be essential not only to the safety of students, but also to the future optics of each university as a whole. While there obviously won't be a shortage of students applying to each university, the perceived danger of one's campus could potentially drive away promising students. UCLA and USC will want to work hard to attract applicants and retain current students, and part of that greatly depends on what campus life has to offer; if campus and the surrounding area looks like a crime zone all the time, it's not a very good look. And, of course, the safety of students could depend on this knowledge. Knowing which crimes affect which people the most will allow universities to improve and refine their safety protocols, resulting in happier and healthier students.

3 Literature Review

Crime has a profound impact on individuals and communities, with demographics and location playing a crucial role in shaping crime patterns. Research has shown that certain groups, such as women and non-Hispanic whites, report higher rates of assault, while sexual violence leads to increased use of mental health and medical services regardless of demographic factors (Sorenson & Siegel, 1992). Additionally, crimes motivated by gender identity often involve severe violence, highlighting the heightened risks faced by transgender individuals (Stotzer, 2008). In urban areas, crime disproportionately affects marginalized communities, as seen in Detroit, where Black adolescents are significantly more likely to experience fatal violence, and previously injured youth face a high likelihood of re-victimization (Clery et al., 2020). The geographic distribution of crime is also a critical factor, as studies suggest that crime diversity is influenced by area size and proximity to key locations such as hospitals, schools, and workplaces (Curtis-Ham et al., 2022; Lentz, 2018). Given the varying crime rates and severity across different environments, understanding how location and demographics interact is essential for evaluating campus safety, particularly in areas surrounding UCLA and USC.

The methodologies in the literature provide valuable tools for analyzing crime patterns near UCLA and USC. The classification-based approach by Almanie and their colleagues (2015) utilizes Naïve Bayes and decision tree classifiers to predict crime types based on time and location, enhancing hotspot identification through dimensionality reduction and demographic analysis. Similarly, Risk Terrain Modeling (RTM) has been effective in predicting homeless-related crime by incorporating geospatial risk factors (Yoo & Wheeler, 2019). We could potentially apply RTM to campus areas to highlight environmental contributors to crime. Additionally, aggregating crime types offers a clearer measure of crime severity (Rosenfeld & Austin, 2023). Combining these methods, classification models, RTM, and crime aggregation, can provide a comprehensive framework for understanding and mitigating crime risks near college campuses.

4 Exploratory Data Analysis

The Exploratory Data Analysis for this project focused on comparing crime patterns near UCLA and USC. We started by calculating crime distances from both campuses using Euclidean distance and subsetting data to victims aged 18–24 as well as the location 3 miles within the center of UCLA or USC. Initial investigation on crime rates revealed an evident contrast: 2,622 crimes occurred within 3 miles of UCLA, while USC had 25,391, suggesting a significantly higher crime density around USC. In addition, demographic analysis showed more female than male victims at both locations, despite the statistically significant differences in crime frequency confirmed by chi-square tests (p=0.0) between the universities.

5 Research Questions

The main question we will be answering is: Is there a statistically significant difference between frequency (and type/severity) of crimes near UCLA and USC, particularly crimes where the victim age is 18-24? The UCLA/USC area is defined to be a radius of 3 miles around each campus. The center of UCLA campus is 34.0722° N, 118.4441° W, near Royce Hall. The center of USC campus is 34.0 224° N, 118.2851° W. Although our primary focus will be to analyze the difference in frequency, type and severity level of crimes near both campuses, further investigation will be directed towards a comparison of the crimes' victim demographics, both by campus and in general. Moreover, we desire to conduct an examination into the specific types of crime that are most associated with a certain campus' area.

Regarding frequency level, relevant background information and prevailing belief leads our team to believe that there is a higher occurrence of crime near USC than near UCLA. Generally, the area around the collegiate campus of USC is perceived to be less upscale and more urban than that of UCLA; these are qualities that align with the assumption of increased criminal incidence. Additionally, we predict that there is a significant difference in the types of crimes committed around UCLA and USC, although our team doesn't have a strong opinion on the specificity of how crime sort varies. Concerning severity, we predict that crimes near USC will be higher in severity on average than crimes near UCLA, a forecast that, once again, is based on the popular cultural narrative that the area surrounding USC might be "undesirable" or unsafe.

In reference to our secondary hypothesis, we predict that if a statistically significant difference between the gender of victims exists, then there is a higher occurrence of female victims than male victims, since women are disproportionately victims of crimes such as stalking, sexual assault, and domestic assault, and societally perceived as a more vulnerable group. We do not have specific predictions regarding which racial or ethnic groups are most targeted or which types of crimes are most prevalent in each area.

6 Methods

The data utilized in this study is the LAPD Crime Data 2020–present, which encompasses all crimes recorded by the Los Angeles Police Department (LAPD) within the Los Angeles area. Observations from the dataset are crimes that were committed in Los Angeles, as reported by the LAPD. The dataset was taken from original paper incident reports later transcripted to a digital data format, meaning that the reports are official, primary sources from the LAPD. For the purpose of analysis, an image of the dataset was downloaded from data.lacity.org website on February 19, 2025. The dataset includes various features for each crime case, such as victim demographics, crime type, and the time and location of the incident. To focus on crimes occurring near the University of California, Los Angeles (UCLA) and the University of Southern California

(USC), we filtered the data to include only crimes that occurred within a 3-mile radius of the geographic center of each university.

This report is an observational study - no manipulation was done in the environment while attempting to make quantifiable and qualifiable deductions from the observations in the dataset. Our experimental design involves separating the data into two different groups - crimes with locations committed near the UCLA campus, and crimes located near the USC campus. We believe that a one-sample t-test is sufficient to determine whether the crime rates differ significantly between the two groups. Additionally, we will analyze how victim demographics and crime types vary between the areas surrounding UCLA and USC. These analyses aim to provide insights into potential spatial and demographic differences in crime patterns near the two universities.

The limitations for the experiment are twofold - the inaccuracy of observational studies and potential errors in the dataset. Because we are conducting an observational analysis of the data, the findings that we deduce may not reflect reality. Real world experiments involve controlled variables, which are not present in observational studies. Secondly, the crime dataset was originally a collection of paper reports. Such paper reports were transcribed to produce the dataset in digital format, which may contain errors. Such errors should be cleaned from the data prior to analysis.

7 Primary Hypotheses – Results & Discussion

7.1 Frequency

Using a chi-square test to evaluate whether there is a statistically significant difference in the frequency of crime near UCLA and USC, we receive a test statistic of 18506.67 and a p-value of approximately 0. We can confidently conclude that there is a difference in the frequency of criminal occurrences across the areas surrounding UCLA and USC.

7.2 Crime Type

We executed a chi-square test of independence to discern whether there was a statistically significant difference in the types of crimes committed near UCLA and USC, utilizing the crime code attached to each report to categorize cases into different "types" of crimes. Our test returned a chi-square statistic of 1083.94 and a p-value = 2.63e-162, with a degrees of freedom equal to 105. The remarkably small p-value indicates that there is substantial evidence of there being a statistically significant disparity between the expected and observed crime type distributions, suggesting that there is, indeed, a difference between the types of crimes committed near UCLA and USC. Below is a table (Figure 1) displaying the theoretical counts of crimes of different codes assuming that there is no difference between the campuses. We observe that there are 106 unique crime codes recorded in the dataset, as shown by the number of rows in the table.

```
Expected Frequencies:
 location_by_university        UCLA          USC
Crm Cd
110                        5.335166    51.664834
121                       23.493449   227.506551
122                        1.123193    10.876807
210                      134.315139  1300.684861
220                       20.123871   194.876129
...                             ...          ...
940                       12.729518   123.270482
943                        0.748795     7.251205
946                       15.256702   147.743298
951                        0.093599     0.906401
956                       22.931853   222.068147

[106 rows x 2 columns]
```

Figure 1: Expected Counts of Different Crimes Codes Across Campuses

7.3 Severity Level

We classified different crime cases into one of three categories depending on their corresponding crime code: Minor (100-399), Moderate (400-699), and Severe (700+). The categories were based on the general groupings of crime codes by the LAPD, as well as a examination of the specific types correlating to different codes in the dataset. In the graph below (Figure 2), we observe the distribution of crime counts by severity and campus. Immediately, we distinguish the increased number of observations in the “USC” class. In terms of shape, there is a small, barely detectable difference between the count of minor and moderate instances for crimes near UCLA. On the other hand, for cases near USC, the frequency of different severity levels are more visually distinct from one another: the graph shows the greatest occurrence being for minor crimes, with a smaller frequency of moderate instances and the least count of severe crimes.

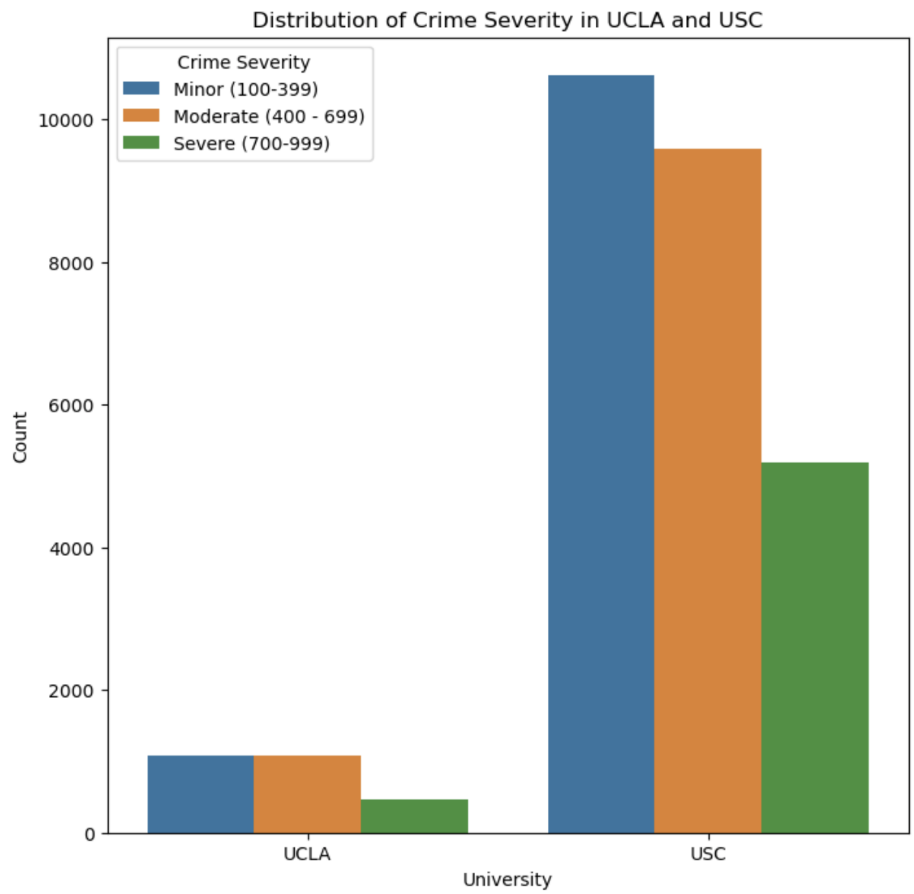

Figure 2: Distribution of Crime Severity in UCLA and USC

In order to consider the severity levels' relative frequency as compared to the total count of crimes for each campus, a graph below (Figure 3) depicts the proportion of different crime severity levels across either campus. We observe a similar pattern from the frequency graph in Figure 2. For crimes near UCLA, minor and moderate crimes occur within 1% of each other, both individually making up about 41-42% of the total number of crimes in the area. Additionally, about 17-18% of the total crimes constitute the "severe" level. We observe slightly different dynamics for crimes near USC. There is a higher proportion of crimes that are categorized as "severe," about 20%. The difference between the proportion of minor crimes (42-43%) and the percentage of moderate crimes (37-38%) is about 5-6%. This difference is much greater than that observed in the crime pool near UCLA. These graphic visualizes illuminate USC's increased experience of minor and severe crimes, as well as UCLA's increased percentage of moderate crimes.

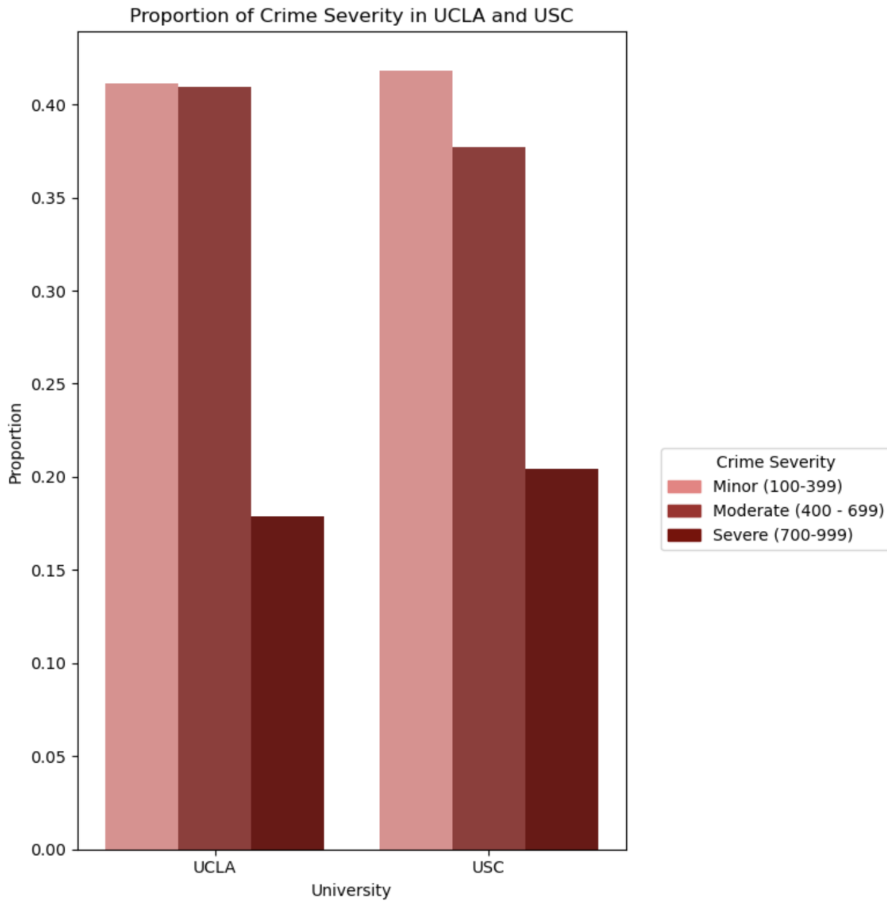

Figure 3: Proportion of Crime Severity in UCLA and USC

8 Predicting Campus Location Using Models of Crime Features – Results & Discussion

Different classifying models were implemented using features of the reported crimes in order to predict where the crime most likely took place: near UCLA or near USC. Prior to training the models, the code was cleaned in a variety of manners. In addition to NA values being dropped, the hour and minute of each crime occurrence was extracted to create a new categorical value entitled "Day_or_Night," which assigns a value of "Day" to crimes taking place between 6 AM and 6 PM and a value of "Night" to crimes taking place at any other time. Furthermore, cyclic encoding of the hour was applied using sine and cosine transformations to prevent discontinuity issues between the beginning and end of cycles in the model. In the column "Weapon Used Cd," all NA values were changed to 0. Several other variables, such as "Vict Sex," "Weapon Used Cd," "Vict Descent," and "Premis Cd," were changed to categorical types for more efficient processing. Information about the street a crime occurred on was also extracted to exclude the building number and encoded as a new variable "Street_Name."

8.1 Random Forest

The random forest model was created using the features "Hour_sin," "Hour_cos," "Weapon Used Cd," "Crm Cd," "Vict Age," "Vict Sex," "Vict Descent," "Street_Name," and "Premis Cd." Using these predictors, the model was able to obtain a 90.07% testing accuracy. The classification report of the random forest model is shown below in Figure 4.

Classification Report:

	precision	recall	f1-score	support
UCLA	0.44	0.13	0.20	807
USC	0.91	0.98	0.95	7594
accuracy			0.90	8401
macro avg	0.68	0.56	0.57	8401
weighted avg	0.87	0.90	0.88	8401

Figure 4: Classification Report of Random Forest Model

The model does a relatively poor job at accurately identifying whether a crime takes place near UCLA, with a precision rate of 44%. The model also generates a low recall rate of 0.13, meaning that it correctly identifies only 13% of the actual "UCLA" cases. The F1-score (0.20) conveys similar inefficiency, as its low value suggests a significant imbalance between the precision and recall. Overall, the model's performance for this class is weak, as it was mostly unable to accurately predict if a crime was committed near UCLA. On the other hand, the precision rate is much higher for USC (0.91). The recall (0.98) and F1-score (0.95) are also much higher than those for the "UCLA" cases, indicating that model performs significantly better at accurately predicting and identifying "USC" crimes. When comparing the support, we remark on the large disparity between the number of cases between the classes. Consequently, we can assume that the model's improved performance on USC cases can be largely attributed to the fact that it is the majority class. Overall, while the model's accuracy of 90% may seem high, it is heavily influenced by the strong performance on "USC" data (the majority class). The model has difficulty identifying "UCLA" instances, which are likely being overshadowed by the more frequent "USC" cases.

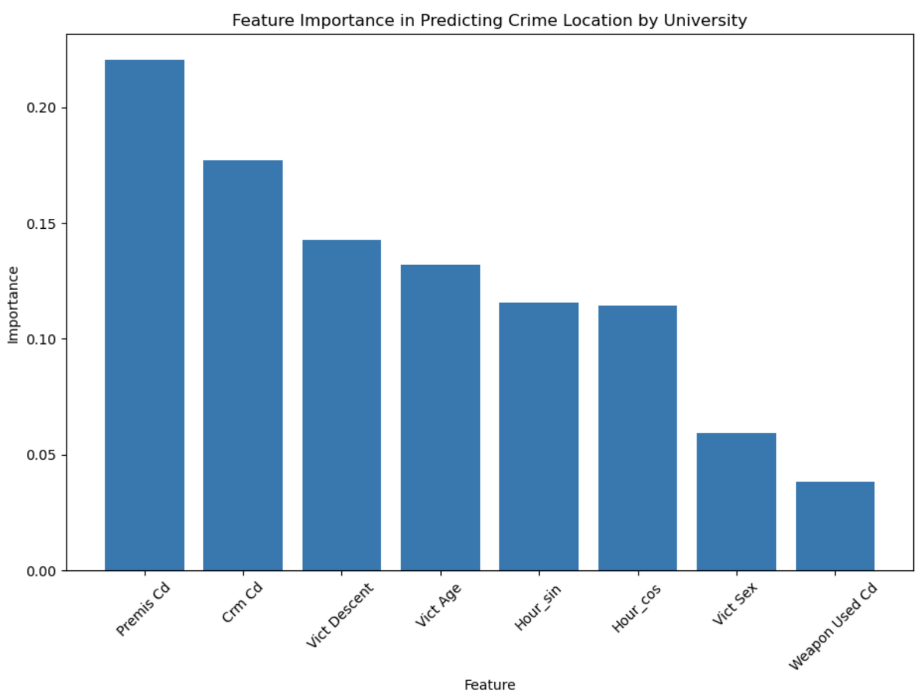

Figure 5: Feature Importance of Random Forest Model

In Figure 5, we see a graph displaying the different features utilized in the random forest model, ranked by their importance. In this particular case, "Premis Cd" plays the largest role in aiding the model with making accurate predictions, and "Weapon Used Cd" is least significant in determining which campus the crime took place near.

8.2 Logistic Regression & Support Vector Machine (SVM)

Prior to the training of the logistic regression and SVM models, SMOTE (Synthetic Minority Oversampling Technique) was utilized to deal with the class imbalance in the data. Using the same features as in the random forest model, logistic regression was able to obtain a 65.21% testing accuracy. The classification report and confusion matrix of the logistic regression model are shown below in Figures 6 and 7.

Logistic Regression Classification Report:

	precision	recall	f1-score	support
UCLA	0.15	0.57	0.24	807
USC	0.94	0.66	0.77	7594
accuracy			0.65	8401
macro avg	0.54	0.61	0.51	8401
weighted avg	0.86	0.65	0.72	8401

Figure 6: Classification Report of Logistic Regression Model

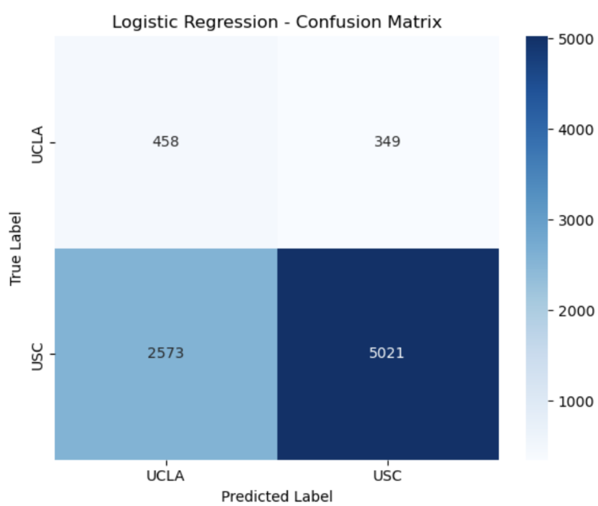

Figure 7: Confusion Matrix of Logistic Regression Model

Using the same features included in the previous models, SVM was able to obtain a 63.36% testing accuracy. The classification report and confusion matrix of the logistic regression model are shown below in Figures 8 and 9.

SVM Classification Report:

	precision	recall	f1-score	support
UCLA	0.17	0.71	0.27	807
USC	0.95	0.63	0.76	7594
accuracy			0.63	8401
macro avg	0.56	0.67	0.51	8401
weighted avg	0.88	0.63	0.71	8401

Figure 8: Classification Report of Logistic Regression Model

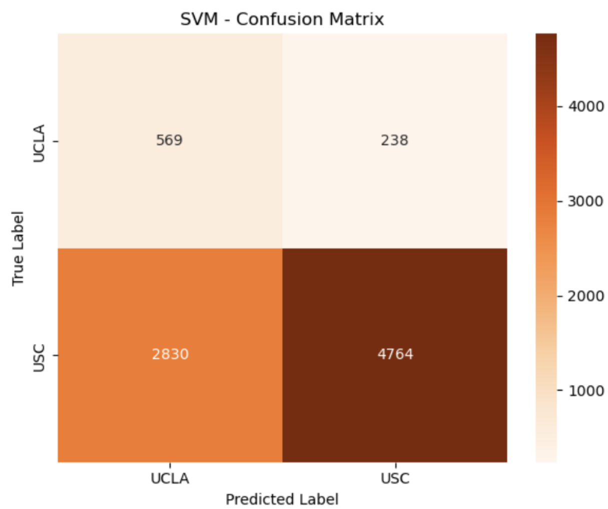

Figure 9: Confusion Matrix of Logistic Regression Model

As with the random forest model, both the logistic regression and SVM models performed poorly at accurately predicting "UCLA" cases, with respective precision rates of 15% and 17%. However, both models have improved recall rates for "UCLA" crimes, with logistic regression correctly identifying 57% and SVM correctly identifying 71% of the actual "UCLA" cases. Again, both F1-scores are low – 0.24 and 0.27 for logistic regression and SVM, respectively – indicating that the logistic regression and SVM models are performing poorly at prediction. It seems as though efforts to counteract the issues with class imbalance using SMOTE have not worked as well as hoped, as proven by the weak prediction performance of "UCLA" cases in both models. On the other hand, both models have relatively high precision, recall and F1-score values for "USC" cases and seem to be accurately predicting true "USC" cases more consistently than "UCLA" cases, mimicking an earlier pattern observed from the results of the random forest model.

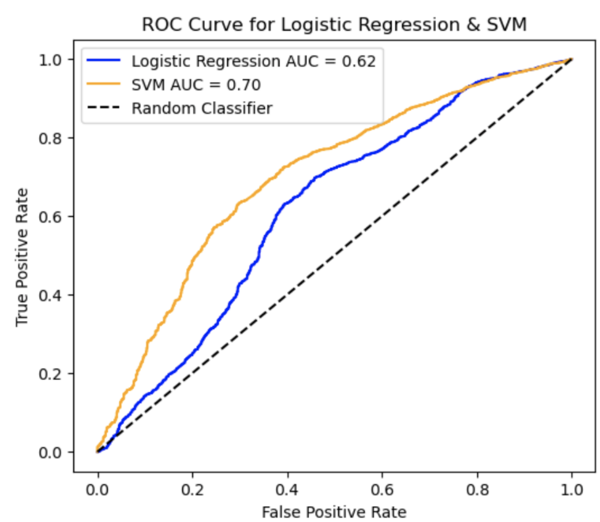

Figure 10: Receiver Operating Characteristic (ROC) Curves of Logistic Regression and SVM Models

The Receiver Operating Characteristic (ROC) curve is shown above in Figure 10. The logistic regression AUC (Area Under Curve) score is 0.60 and the SVM AUC score is 0.70. The SVM has better diagnostic accuracy, which is shown in the graph as the SVM curve exceeds the logistic regression curve. However, neither model boasts an impressive prediction rate in terms of accuracy, and both underperform compared to the previous random forest model.

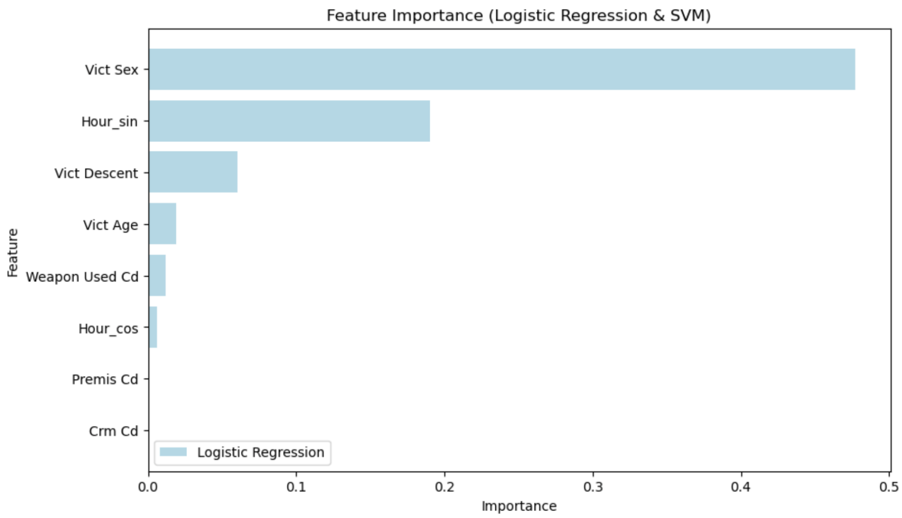

Figure 11: Feature Importance of Logistic Regression and SVM Models

In the feature importance graph depicted above in Figure 11, we observe that the logistic regression and SVM models rely most on "Vict Sex" to accurately predict which campus a crime is near. Unlike in the random forest model, "Premis Cd" is one of the least significant features by importance, tying with "Crm Cd" for the lowest importance.

8.3 XGBoost

Using the same predictors as in the previous models, an XGBoost classifier was able to obtain a 90.47% overall prediction accuracy. The classification report and confusion matrix of the XGBoost model are shown below in Figures 12 and 13.

```
XGBoost Classification Report :
              precision    recall  f1-score   support

        UCLA       0.51      0.15      0.24       807
         USC       0.92      0.98      0.95      7594

    accuracy                           0.90      8401
   macro avg       0.71      0.57      0.59      8401
weighted avg       0.88      0.90      0.88      8401
```

Figure 12: Classification Report of XGBoost Model

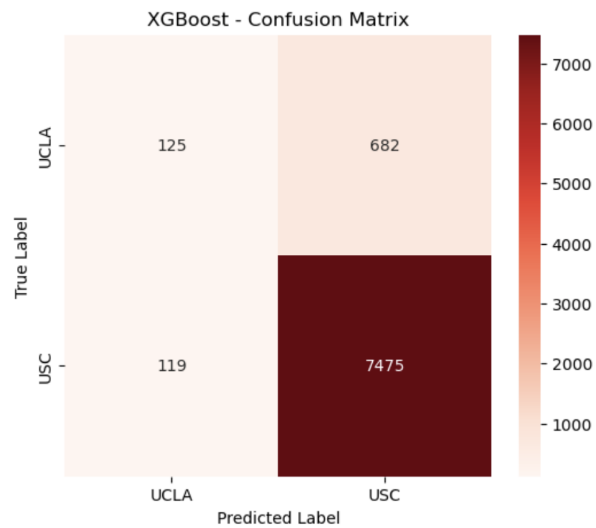

Figure 13: Confusion Matrix of XGBoost Model

The classification report illuminates the same trend we've observed in the random forest, logistic regression and SVM models, which is that the XGBoost classification method performs better on "USC" cases than "UCLA" ones. However, this model boasts the highest precision rate out of all models for crimes committed near UCLA, with a value of 51%. The recall is similarly very low (0.15) to all previous models, indicating that XGBoost remains unsuccessful in predicting true positive cases. The F1-score is also low (0.24), corresponding with the similarly low values achieved using the previous models. The XGBoost model performs much better than the SVM and logistic regression models for crimes of the "USC" class, performing similarly to the random forest model. However, because the random forest model performs slightly poorer for "UCLA" cases, the XGBoost has the highest overall accuracy rate and presents as the best performing model we've produced.

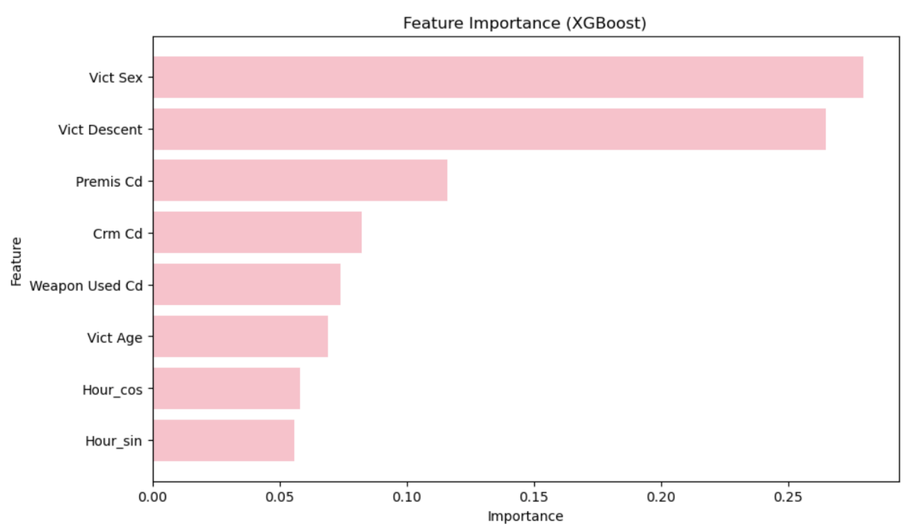

Figure 14: Feature Importance of XGBoost

In the feature importance graph above, we observe that "Vict Sex" and "Vict Descent" are the most useful predictors for the XGBoost model. On the other hand, "Hour_sin" and "Hour_cos" are the least significant in assisting the model with determining the correct campus the crime took place near.

9 Additional Hypotheses – Results & Discussion

9.1 Victim Sex

In Figure 15 (below), we see the graphed distribution of crime count by campus and victim sex, excluding the crimes where victim sex was unreported or unlisted (represented by an "X" in the original dataset). From the graph, we observe that there a barely distinguishable difference between the amount of female and male victims for crimes committed within three miles of UCLA; 1,329 and 1,280 crimes were reported to have been perpetrated against women and men, respectively. On the other hand, there is a notable disparity between the number of female and male victims for crimes committed within three miles of USC; there were 11,355 crimes where victim sex was recorded as female and 10,257 crimes where victim sex was recorded as male.

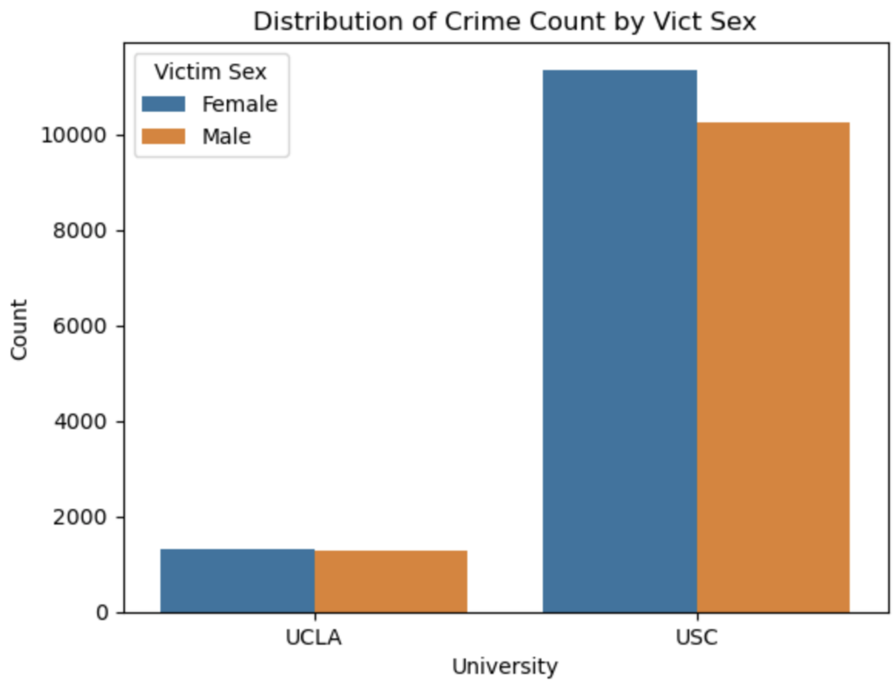

Figure 15: Distribution of Crime Count by Victim Sex

Using a chi-square of independence to analyze the difference between victim sex by campus, we receive a test statistic of 421.98 and a p-value = 3.84e-91, with a degrees of freedom equal to 3. The significantly small p-value insinuates a relationship between campus and victim sex, thus we can conclude that there is a difference in victim sex across the campuses.

Below is a mosaic plot (Figure 16) that illustrates that relationship between campus and victim sex. There is a conspicuous difference in the occurrence of crimes, as indicated by the size of the red and green areas (denoting USC and UCLA, respectively) relative to the horizontal axis. The vertical axis represents the sizes of the gender demographics within each campus' crime population, which showed an apparent contrast in the gender proportions of victims. The proportion of crimes near UCLA committed against females appears to be generally equal to the proportion of male victims, with few "X" and "H" entries, which presumably represent unknown or unrecorded inputs. On the other hand, there is a much higher instance of "unknown" gendered-victims amongst crimes near USC, and, upon closer inspection, we can deduce a greater proportion of female victims than male, which aligns with the information gleaned in Figure 15. Overall, we can conclude, both from the results of our chi-square analysis and the visualizations produced from the data, that victim sex varies systematically with the campus, with USC having a higher ratio of female victims than UCLA. In regards to the entries with an unknown victim sex ("X" and "H"), we can only speculate as to the multitude of potential reasons why victim gender would go unrecorded or unknown. However, the results and implications of our analyses remain the same.

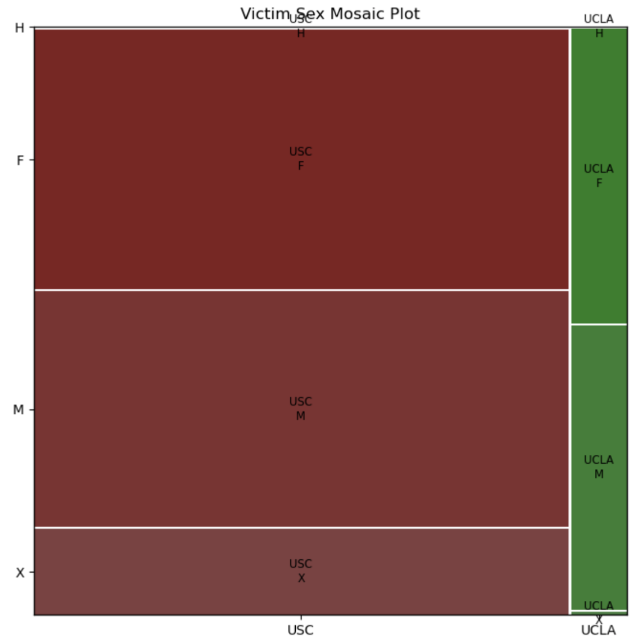

Figure 16: Mosaic Plot of Victim Sex

9.2 Victim Descent

We performed a chi-square test to analyze the differences in victim descent between crimes committed near UCLA and USC. Our results returned a chi-square statistic of 2865.71, a p-value ~ 0, and degrees of freedom equal to 18. The huge test statistic and extremely small p-value both indicate that there is a statistically significant difference in the racial/ethnic background of victims between crimes near UCLA and USC. The table below (Figure 17) showcases the theoretical counts of victims of different racial/ethnic backgrounds assuming that there is no difference between the campuses. We observe that there are 19 unique categorizations of racial/ethnic background recorded in the dataset, as shown by the number of rows in the table.

```
Expected Frequencies:
 location_by_university         UCLA          USC
Vict Descent
0                        121.816305  1179.183695
1                        407.209156  3941.790844
2                         52.996179   513.003821
3                          0.936328     9.063672
4                         11.704103   113.295897
5                          0.187266     1.812734
6                       1028.369318  9954.630682
7                         13.295861   128.704139
8                          5.337071    51.662929
...                             ...          ...
10                         0.655430     6.344570
11                       163.389280  1581.610720
12                         0.655430     6.344570
13                         0.093633     0.906367
14                         0.842695     8.157305
15                         7.771524    75.228476
16                       412.078063  3988.921937
17                       371.066886  3591.933114
18                         5.243438    50.756562

[19 rows x 2 columns]
```

Figure 17: Expected Counts of Different Racial/Ethnic Backgrounds of Victims Across Campuses

A key for the racial/ethnic backgrounds is as follows: 0 - Other Asian; 1 - Black; 2 - Chinese; 3 - Cambodian; 4 - Filipino; 5 - Guamanian; 6 - Hispanic/Latin/Mexican; 7 - American Indian/Alaskan Native; 8 - Japanese; 9 - Korean; 10 - Laotian; 11 - Other; 12 - Pacific Islander; 13 - Samoan; 14 - Hawaiian; 15 - Vietnamese; 16 - White; 17 - Unknown; 18 - Asian Indian.

When performing a similar analysis on the proportion of racial/ethnic backgrounds of victims of crimes near UCLA and USC, we receive analogously significant results that indicate the same substantial difference between racial/ethnic demographics of crime victims across campuses. The table below (Figure 18) reveals the proportion of victims of each racial/ethnic group for crimes near both campuses, and Figure 19 exhibits a graph that visualizes this information.

```
Proportions of Victim Descent by Location:
Vict Descent                   0         1         2         3         4  \
location_by_university
UCLA                    0.080473  0.085812  0.061785  0.000000  0.011442
USC                     0.042946  0.162484  0.015917  0.000394  0.003743

Vict Descent                   5         6         7         8         9  \
location_by_university
UCLA                    0.000000  0.163234  0.001526  0.004958  0.009153
USC                     0.000079  0.415862  0.005437  0.001734  0.006777

Vict Descent                  10        11        12        13        14  \
location_by_university
UCLA                    0.000000  0.128528  0.000381  0.000000  0.000763
USC                     0.000276  0.055475  0.000236  0.000039  0.000276

Vict Descent                  15        16        17        18
location_by_university
UCLA                    0.007628  0.428299  0.007246  0.008772
USC                     0.002482  0.129152  0.155392  0.001300
```

Figure 18: Proportion Table of Victim Descents by Campus

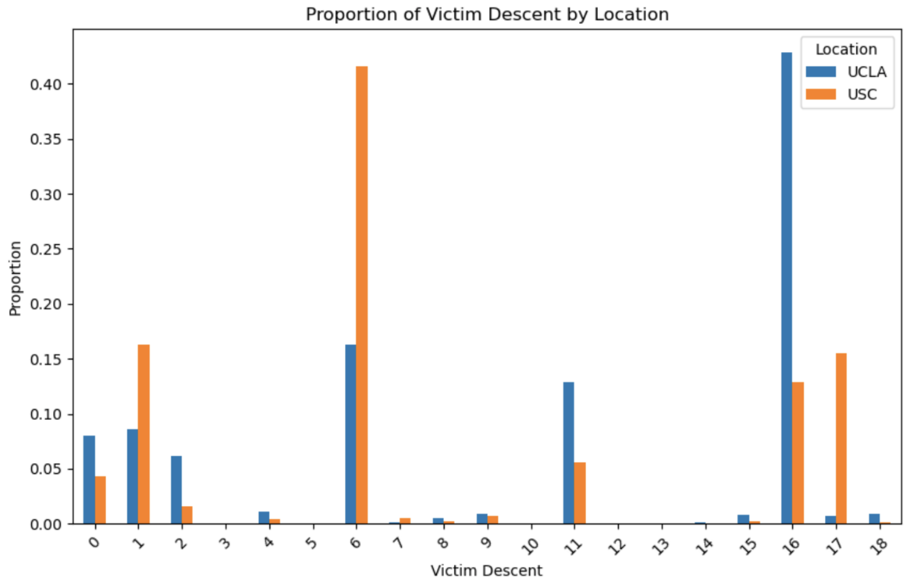

Figure 19: Proportion of Victim Descent by Location

The graph above indicates that the most significant differences across the campuses are in the proportions of victims of Other Asian (0), Black (1), Chinese (2), Hispanic/Latin/Mexican (6), Other (11), White (16), and Unknown (17) racial/ethnic descent. There is a higher percentage of victims of Black, Hispanic/Latin/Mexican and Unknown descent in the crimes committed near USC, while there is a larger proportion of Other Asian, Chinese, Other, White victims in the pool of crimes near UCLA. In the 2022-2023 academic year, 35% of students attending USC identified as Latinx and 32% identified as African American. On the other hand, in the same year, 26.6% of students attending UCLA identified as white and 26.2% identified as Asian. Thus, the specific racial/ethnic subgroups in which the crimes vary across campus correlate generally to the student demographics of each school. While USC has a higher proportion of black and hispanic crime victims, which is in accordance to the two most identified racial/ethnic groups present in their student body, there is a higher proportion of white and Asian victims in the crimes near UCLA, which is primarily made up of white- and Asian-identifying students.

In the table below (Figure 20) are the results of a two-proportion z-test conducted for each racial/ethnic group from both campuses. Using the test to determine whether there exists a statistically significant disparity in the proportions of each racial/ethnic group from 0 to 19 across crimes near UCLA and USC, we were able to attribute a “true” value for racial/ethnic groups that produced p-values < 0.05 and a “false” value for categories that had larger p-values. An assigned value of “true” suggests that there is a statistically significant difference in the proportion of that specific racial/ethnic group across crimes near the two campuses. The smallest p-values are associated with the Other Asian (0), Black (1), Chinese (2), Hispanic/Latin/Mexican (6), Other (11), Unknown (17), and Asian Indian(18) – most of these have already been discussed in our evaluation of the graph above depicting the proportion of victim descents by campus.

Significant Differences (P-value < 0.05):

	Victim Descent	P-value	Significant
0	0	3.566585e-18	True
1	1	5.751233e-25	True
2	2	7.531315e-57	True
3	3	3.093555e-01	False
4	4	1.802689e-08	True
5	5	6.494249e-01	False
6	6	2.183860e-140	True
7	7	7.261175e-03	True
8	8	4.872741e-04	True
..	...	...	...
10	10	3.950570e-01	False
11	11	3.971449e-49	True
12	12	6.547938e-01	False
13	13	7.478939e-01	False
14	14	1.853592e-01	False
15	15	3.943039e-06	True
16	16	0.000000e+00	True
17	17	2.304386e-95	True
18	18	3.541270e-16	True

[19 rows x 3 columns]

Figure 20: Z-Test Results for Each Racial/Ethnic Background Across Campuses

9.3 Map Visualizations

Figures 21 and 22 below depict heat maps that display the geographical distribution of victim descent across crimes near both campuses.

Figure 21: Heat Map of Crimes in UCLA by Descent

Figure 22: Heat Map of Crimes in USC by Descent

When we compare the two images above, we can see that USC's heat map contains a much larger volume of points. Additionally, we observe a difference in the general color patterns of the points, with USC containing a majority of darker-colored purple dots and UCLA displaying primarily yellow-colored dots, with few purple points. This information correlates to our previous analysis and discussion of the general patterns of victims' racial/ethnic backgrounds across campuses, with the bulk of USC's crimes occurring against victims of Black (1) and Hispanic/Latin/Mexican (6) racial/ethnic background and the majority of UCLA's cases involving White (16) victims. In terms of geographical distribution, we can see that crimes near USC are spread relatively equally across the entire region, with few areas being especially exempt from crime. On the other hand, crimes occurring near UCLA seem to be centered mostly in the areas southern and southwestern of the campus center. An explanation for the major inequality of crime placement across the area could involve the fact that the area north of UCLA is largely residential, while the area south of UCLA is more urban and contains a more concentrated population.

We also observe the graphical patterns conveyed in the heat maps displayed in figures 23 and 24 below, which graph victim sex across crimes near both campuses.

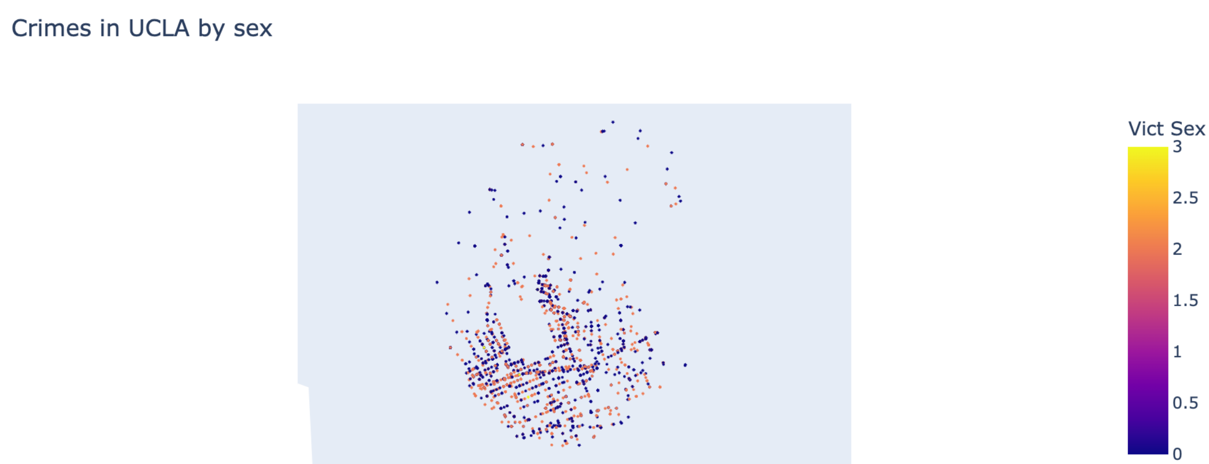

Figure 23: Heat Map of Crimes in UCLA by Sex

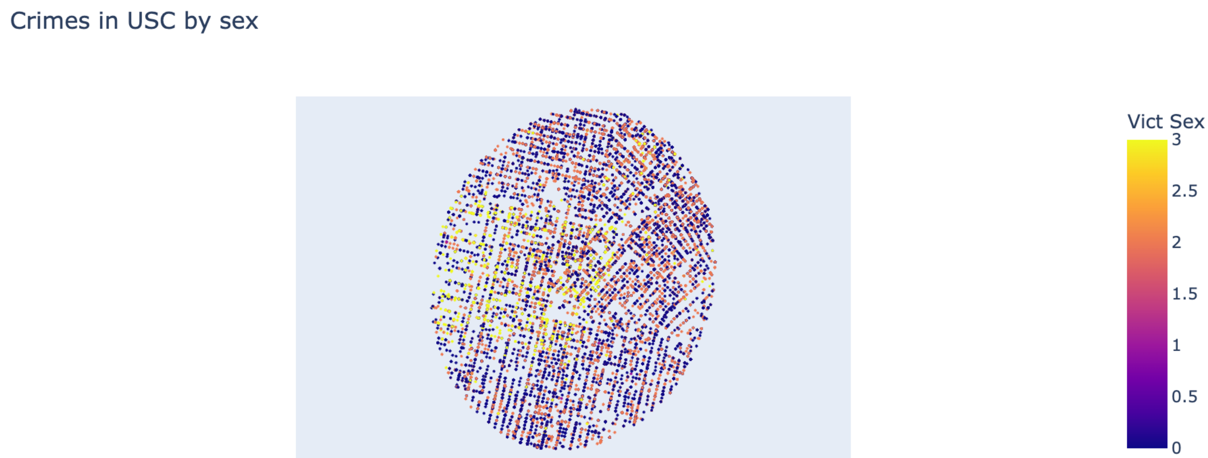

Figure 24: Heat Map of Crimes in USC by Sex

While both campuses' neighboring crimes display a relatively equal ratio of purple to orange dots, representing similar frequencies of male and female victims, USC's heat map displays a concentration of yellow-colored points in the middle and towards the western edge of the map. Our previous exploration of victim sex across universities revealed that USC had a much higher proportion of victims of unknown or unrecorded gender identity, which is referenced by this unique concentration of yellow points. Their seemingly contained location in the middle and western regions might indicate that those victims are primarily students, who, for some reason, didn't disclose information about their gender identity when reporting the crime they suffered.

10 Conclusion

Based on our analysis, it's clear that there is a statistically significant difference in crime rates between the UCLA and USC areas, the latter having not only much higher crime rates, but also a higher proportion of "Severe" crimes. Furthermore, there's a significant difference in factors of crime between UCLA and USC, namely victim sex and victim descent. Observing the differences, it appears that females are more likely to be victims of a crime in the USC area, whereas they're equally likely to be a victim of a crime as men are in the UCLA area. Compared to UCLA, Black, Hispanic, and "Unknown" people are more likely to be a victim of a crime in the USC area; "unknown" is used to indicate a collection error. On the other hand, people of various Asian ethnicities and White people are more likely to be victims of a crime around UCLA.

Considering this knowledge, it's reasonable to conclude that each university should take measures to protect its college-aged population. Universities can consider enhancing security both on-campus and in university apartments or other off-campus locations. Educating students to be aware and how to protect themselves could also be key to curtailing crime rates. Beyond the scope of UCLA and USC, the city and/or state could also consider similar methods; it's clear that safety is an issue for this age group, so it's important to protect the people who will ultimately be leaving with degrees to do greater things.

11 Limitations and Future Recommendations

However, it's important to acknowledge that this study is not definitive, and is bound by some limitations. Firstly, this is an observational study based on a publicly available dataset from the LAPD, which may introduce bias in our findings. In addition, the dataset contains many missing values and even unexplained values, including the presence of "H" as an indicator for victim sex; several data entries are missing several factors. Secondly, our analysis consists mostly of checking for significant differences between the two universities. While that can be quite telling in of itself, there is much more potential to be found in analyzing certain factors, which could potentially answer more questions about crime between UCLA and USC.

Unfortunately, this research project was mainly restricted by time, but that does not mean this topic can't be looked into further. Researchers interested in analyzing crime between UCLA and USC could consider using the same dataset as ours, but go further by analyzing other factors that have more detail. For example, "premise code" and "weapon code" could possibly reveal crime patterns for each area. Also, since our dataset doesn't tell us who's a student of UCLA or USC, researchers could use or collect data specific to students of each university in order to calculate better estimates of crimes against students. Ultimately, the scope of this study is relatively small and can be greatly improved on, but it's one that is very interesting to us. There isn't any research we found that specifically tackles this topic, so we hope our attempt can serve as a starting point to the topic.

12 References

1. Almanie, T., Mirza, R., & Lor, E. (2015). Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots. International Journal of Data Mining & Knowledge Management Process, 5(4), 01–19. https://doi.org/10.5121/ijdkp.2015.5401
2. Clery, M., Dworkis, D., Sonuyi, T., Khaldun, J., & Abir, M. (2020). Location of Violent Crime Relative to Trauma Resources in Detroit: Implications for Community Interventions. Western Journal of Emergency Medicine, 21(2), 291–294. https://doi.org/10.5811/westjem.2019.9.44264
3. Curtis-Ham, S., Bernasco, W., Medvedev, O. N., & Polaschek, D. L. L. (2022). Relationships Between Offenders' Crime Locations and Different Prior Activity Locations as Recorded in Police Data. Journal of Police and Criminal Psychology. https://doi.org/10.1007/s11896-022-09540-8
4. Lentz, T. S. (2018). Crime Diversity: Reexamining Crime Richness Across Spatial Scales. Journal of Contemporary Criminal Justice, 34(3), 312–335. https://doi.org/10.1177/1043986218770002
5. Rosenfeld, R., & Austin, J. (2023). The Future of Crime in Los Angeles and the Impact of Reducing the Prison Population on Crime Rates. CrimRxiv. https://doi.org/10.21428/cb6ab371.b2179eae
6. Sorenson, S. B., & Siegel, J. M. (1992). Gender, Ethnicity, and Sexual Assault: Findings from a Los Angeles Study. Journal of Social Issues, 48(1), 93–104. https://doi.org/10.1111/j.1540-4560.1992.tb01159.x
7. Stotzer, R. L. (2008). Gender identity and hate crimes: Violence against transgender people in Los Angeles County. Sexuality Research and Social Policy, 5(1), 43–52. https://doi.org/10.1525/srsp.2008.5.1.43
8. Student Body Demographics. USC Suzanne Dworak-Peck School of Social Work. (n.d.). https://dworakpeck.usc.edu/admissions/student-body-demographics#:~:text=Across%20all%20four%20of%20our,5%25%20of%20students%20are%20Unknown.
9. University of California-Los Angeles (110662). Data USA. (n.d.). https://datausa.io/profile/university/university-of-california-los-angeles#:~:text=9%2C293%20enrolled%20students-,The%20enrolled%20student%20population%20at%20University%20of%20California%2DLos%20Angeles,Hawaiian%20or%20Other%20Pacific%20Islanders.
10. Yoo, Y., & Wheeler, A. P. (2019). Using risk terrain modeling to predict homeless related crime in Los Angeles, California. Applied Geography, 109, 102039. https://doi.org/10.1016/j.apgeog.2019.10203