

Los Angeles Thefts 2022

“Theft of Identity”

Contents

[1. Abstract](#)

[2. Introduction](#)

[3. Crimes](#)

[3.1. Victims](#)

[3.2. Time](#)

[3.3. Location](#)

[4. Logistic Regression](#)

[5. Conclusion](#)

[6. Recommendations](#)

[7. References](#)

1. Abstract

Report analyses Los Angeles (L.A.) crime numbers provided by the Los Angeles police department and contains a lot of terminology, analysis goal is to represent data in easy-to-approach and easy-to-understand data. The report shows the distribution of crimes based on victims' gender, descent, and age. The crime increased by 12% and one crime in particular had a 106% jump from 2021. Due to this increase report will focus on the *Theft of Identity*. Furthermore, the report will analyse a time trend of crime occurrence and spatial data display the city's crime heatmap. Spatial data and statistical analysis will identify the most problematic locations linked to *Theft of Identity* and possible crimes committed with a stolen identity. The motivation behind the analysis is to help uncover and streamline the most prominent theft and its trends to establish a focus on specific locations, to help allocate police resources better.

2. Introduction

The report is part of a Uniform Crime Reporting (UCR) Program to help agency staff formulate policies and make strategic and operational decisions. The report focuses on the *Theft of Identity* crime and uses spatial data, aggregated data, cohort analysis, linear and logistic regression to identify trends and correlations between *Theft of Identity* and location to pinpoint where police should focus first.

3. Crimes

In 2022 L.A.'s Police Department reported 234 921 crimes. Which is a 12.86% increase from the last year. (Table 1.)

| Year | Crimes reported |
|------|-----------------|
| 2021 | 208 190 |
| 2022 | 234 921 |

Table 1. L.A. Crimes 2021 and 2022

While for the most part individual crimes for 2021 and 2022 are close in their amount there is a notable increase in *Theft Of Identity*. (Table 2.)

| Year | Crime | Crimes reported |
|------|-------------------|-----------------|
| 2021 | Theft Of Identity | 10 788 |
| 2022 | Theft Of Identity | 22 287 |

Table 2. Theft of Identity increase

The increase of 106.34% is very concerning. *Theft of Identity* is a unique crime in that committing Identity Theft is linked with other crimes, meaning that this crime could be considered a method to commit another crime.

The reasoning behind the focus on *Theft of Identity* is that looking back at overall data for 2022, there were 133 unique crimes reported which are grouped based on their nature with a *Theft or Attempt of it* making up 58.60% of all crimes. (Figure 1.)

Theft makes up more than half of reported crimes. *Thef of Identity* will be a specific crime we will focus on due to its increase in 2022 and nature to be used as a tool to commit other crimes.

| | |
|-------------------------------|--------|
| Any Kind Of Theft Or Attempt | 58.60% |
| Crimes Against Animals | 0.03% |
| Crimes against children | 0.95% |
| Damage To Property | 8.17% |
| Other Miscallaneous Crime | 1.00% |
| Physical Damage Or Attempt | 23.56% |
| Scams, Fraud, Bribes, Forgery | 0.84% |
| Stalking | 1.19% |
| Violation | 3.85% |
| Weapons | 1.80% |

Figure 1. Crime distribution

3.1. Victims

Given that L.A. is a multicultural city and reported victims consist of a variety of ethnic groups. In 2022 *Theft of Identity* was reported 22,287 times. (Figure 2.)

The Hispanic/Latin/Mexican ethnic group being at the top of victims makes sense since this ethnic group makes up 46.9%¹ of the whole L.A. metro city’s population. Another thing to keep in mind is that L.A. is a multicultural, multi-ethnic city and assimilation between different groups

occurs and defining ethnicity, and race becomes way harder therefore descent definitions could be revised and reformed.

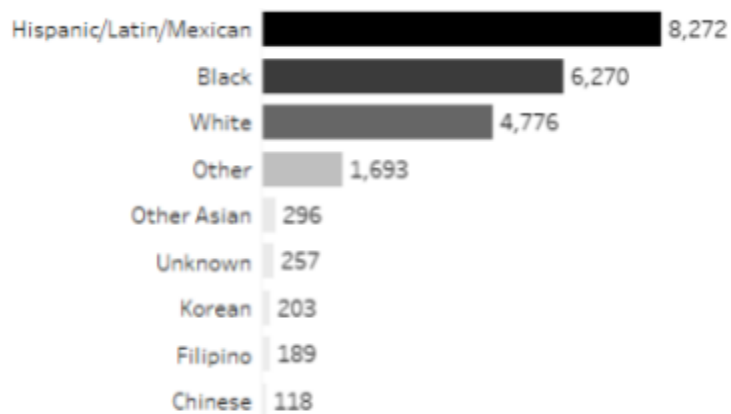


Figure 2. Crime distribution across ethnic groups

The gender distribution suggests that Females could be a target of this crime since the L.A. Gender ratio is almost equal². (Figure3.)

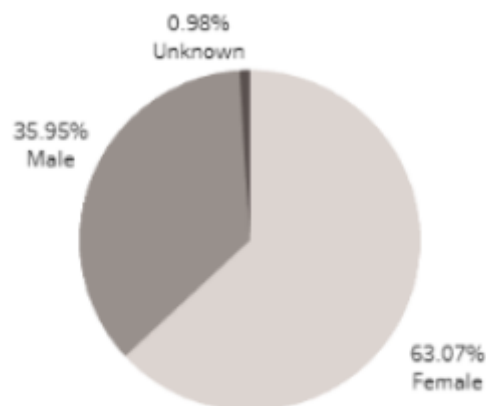


Figure 3. Crime distribution genders

¹ Demographics of Los Angeles https://en.wikipedia.org/wiki/Demographics_of_Los_Angeles

² L.A. Population by Gender <https://www.neilsberg.com/insights/los-angeles-ca-population-by-gender/>

Looking at the age distribution, victims of identity theft for the most part are adults with age between 21 and 40. (Figure 4.)

A quick overview of victims is a good way to understand data distribution between specific groups. Another thing to mention simple aggregation can provide insight into data quality, especially when working with this type of data. Dealing with categories such as *Unknown or Undisclosed* can be very tricky for pattern, insight, and trend searching since to perform analysis data must be well-defined. Talking about this crime specifically the data is well defined and collected because the percentages of *Undisclosed/Unknown* are very low, but if these categories become very large the analysis would suffer in its quality since undefined data could belong in any other well-defined

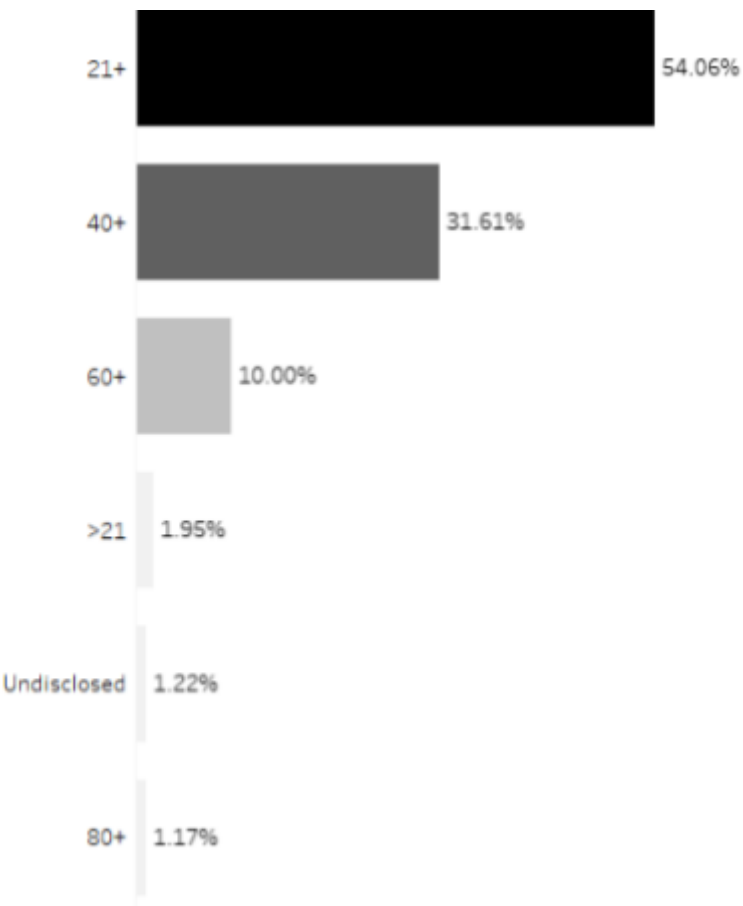


Figure 4. Crime distribution age

groups and can cause problems where a crime can become biased to an undefined group. While aggregated numbers are a good overall approach to display and check the quality of the data, for easier trend searching Cohort Analysis within a context of all thefts will help to establish the reasoning why we are focusing on this specific crime.

3.1.a Descent Cohort

In simple terms Cohort is a group of people who share the same characteristics, traditionally this type of analysis is performed to define and explain user behaviour in the e-commerce field, but in our case cohort analysis can streamline trends and explain data quality.

First, let's look at the Descent cohort and thefts (Figure 5. Thefts > 1000).

| | Undisclo.. | Black | Hispanic.. | Other | Unknown | White |
|---|------------|--------|------------|--------|---------|--------|
| VEHICLE - STOLEN | 84.44% | | | | | |
| THEFT-GRAND (\$950.01 & OVER)EXCPT,GUN.. | | | 6.18% | 13.74% | 11.57% | 10.05% |
| THEFT PLAIN - PETTY (\$950 & UNDER) | | 12.99% | 10.34% | 15.81% | 10.98% | 13.86% |
| THEFT FROM MOTOR VEHICLE - GRAND (\$95.. | | | 12.22% | | | 14.25% |
| SHOPLIFTING - PETTY THEFT (\$950 & UNDER) | | | | | 29.62% | |
| ROBBERY | | 9.82% | 12.53% | | 13.86% | 4.18% |
| BURGLARY | | 9.51% | 7.69% | 27.95% | 33.96% | 16.85% |
| THEFT OF IDENTITY | | 50.12% | 31.12% | 22.59% | | 18.82% |
| THEFT FROM MOTOR VEHICLE - PETTY (\$950.. | 15.56% | | 4.33% | | | 4.02% |
| BURGLARY FROM VEHICLE | | 17.56% | 15.60% | 19.91% | | 17.98% |

Figure 5. (Thefts > 1000)

The Descent Cohort displays the percentage of victims who suffered a specific theft. The key takeaway from this is that *Theft of Identity* is always reported and does not have an undisclosed or unknown descent description and in context to all thefts covers all descent cohorts. Furthermore, identity theft half of the time is being reported by people of black descent.

3.1.b Gender Cohort

Gender Cohort same as descent shows our focus crime to be a well-defined, gender split within the cohort with Female victims being two times larger than the Male cohort for identity theft. (Figure 6.) In context with other crimes identity theft takes up 17.55% of all crimes against a specific gender and Female victims that report any kind of theft more than a third of the time report *Theft of Identity* crime.

| | Female | Male | Unknown | Total |
|--|--------|--------|---------|--------|
| VEHICLE - STOLEN | | | 62.85% | 19.60% |
| THEFT-GRAND (\$950.01 & OVER)EXCPT,GUN.. | 7.88% | 8.15% | 3.05% | 6.48% |
| THEFT PLAIN - PETTY (\$950 & UNDER) | 12.92% | 11.01% | 2.87% | 9.06% |
| THEFT OF IDENTITY | 36.39% | 16.70% | | 17.55% |
| THEFT FROM MOTOR VEHICLE - PETTY (\$950 .. | 3.45% | 3.70% | 11.87% | 6.17% |
| THEFT FROM MOTOR VEHICLE - GRAND (\$95.. | 9.61% | 12.34% | | 7.66% |
| SHOPLIFTING - PETTY THEFT (\$950 & UNDER) | | 4.48% | 7.15% | 3.94% |
| ROBBERY | 4.80% | 9.83% | 3.54% | 6.33% |
| BURGLARY FROM VEHICLE | 15.10% | 16.87% | | 11.07% |
| BURGLARY | 9.84% | 14.41% | 8.56% | 11.19% |
| BIKE - STOLEN | | 2.52% | | 0.96% |

Figure 6. (Thefts > 1000)

3.1.c Age Cohort

Age Cohort reinforces the previous cohort's insights that the *Theft of Identity* data is very well documented because this crime has no undisclosed/unknown data. (Figure 7.).

| | 21+ | 40+ | 60+ | Undisclo.. | Total |
|--|---------------|---------------|---------------|------------|---------------|
| VEHICLE - STOLEN | | | | 58.67% | 20.78% |
| THEFT-GRAND (\$950.01 & OVER)EXCPT,GUN.. | 7.98% | 7.64% | | 3.82% | 5.90% |
| THEFT PLAIN - PETTY (\$950 & UNDER) | 12.40% | 11.25% | 13.99% | 3.42% | 9.08% |
| THEFT OF IDENTITY | 27.90% | 27.54% | 28.18% | | 17.96% |
| THEFT FROM MOTOR VEHICLE - PETTY (\$950 .. | 3.51% | 3.96% | | 11.12% | 6.07% |
| THEFT FROM MOTOR VEHICLE - GRAND (\$95.. | 9.59% | 13.87% | 19.79% | | 7.79% |
| SHOPLIFTING - PETTY THEFT (\$950 & UNDER) | 3.77% | | | 7.75% | 4.12% |
| ROBBERY | 7.30% | 6.60% | | 4.05% | 5.51% |
| BURGLARY FROM VEHICLE | 19.55% | 15.47% | 13.16% | | 11.32% |
| BURGLARY | 8.01% | 13.67% | 24.88% | 11.18% | 11.47% |

Figure 7 (Thefts > 1000).

All age groups that report any kind of theft report identity theft close to third. And this crime makes up 17.96% of all crimes between all adult groups.

Looking at victims' cohort we see that *Theft of Identity* is well-reported. Well-defined data is important for future analysis especially in building classification models and/or searching for correlation³. Having looked at victims' data next step is to isolate a time when *Theft of Identity* happens the most.

³ Side Note: Looking at the Cohort analysis of *Vehicle-Stolen* crime can be problematic since all three cohorts for this crime are only reported as undisclosed/unknown, this flaw could be a sign of biased data for this crime and drawing insights about victims for this crime should be approached carefully and in general might be an area for separated investigation on how data was collected and/or processed.

3.2. Time

Looking at time data for 2022 we can see that reported *Theft of Identity* is consistent every month where the difference between the lowest and highest point is around 1200. The average for the year 2022 is 1857 thefts. (Figure 8.)

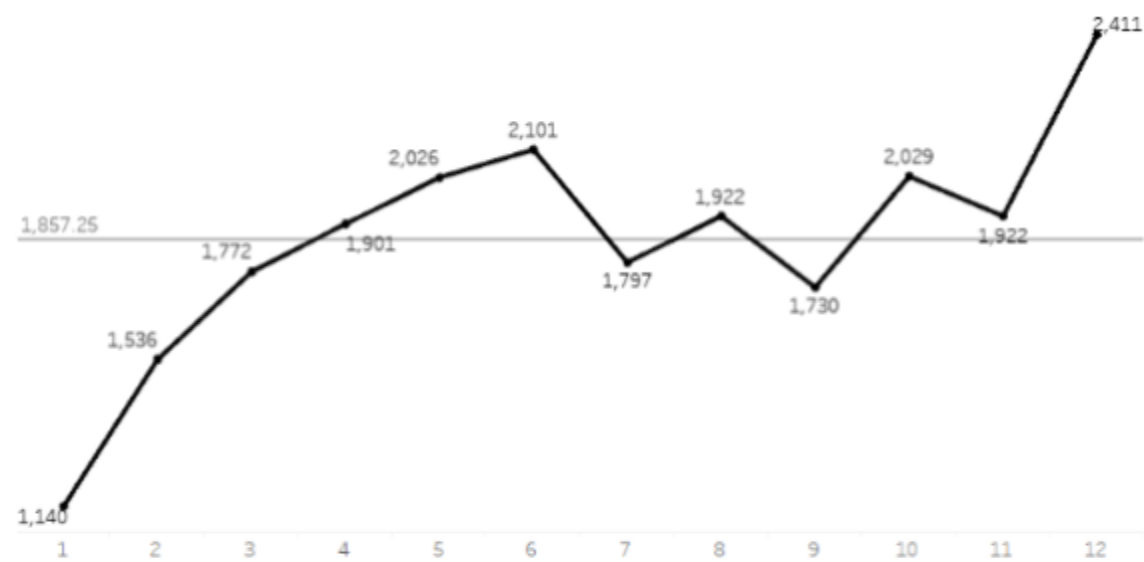


Figure 8.(Theft of Identity 2022)

Compared with the previous year the average crime occurrence grew by 106%. (Figure 9.)

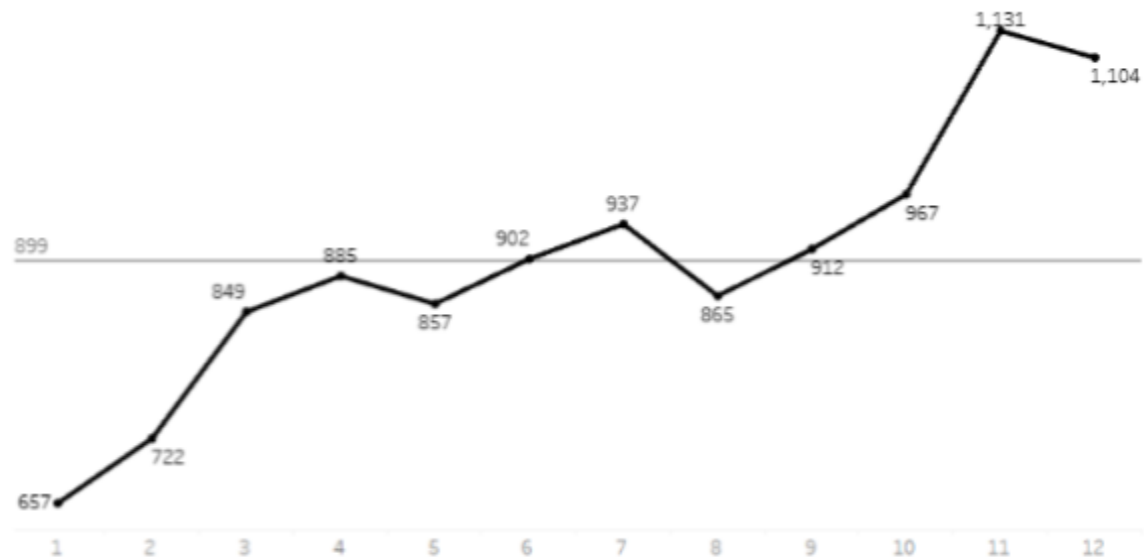


Figure 9. (Theft of Identity 2021)

It is hard to distinguish if there is seasonality, but it is clear that first 6 months for both years identity theft increases and dips at middle-to-end of summer and continues to trend upwards. Overall, we can say that the year 2022 is continuing a 2021 upward trend.

Looking at monthly cohorts to find time patterns we can see that *Theft of Identity* is one of the consistent crimes because it is being reported every month. (Figure 10.) Also, identity theft makes up close to one-third of all thefts reported in any given month.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|----------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| VEHICLE - STOLEN | 36.21% | 33.63% | 29.42% | 33.82% | 32.64% | 30.73% | 34.35% | 34.21% | 34.16% | 28.36% | 26.05% | 19.85% | 30.56% |
| THEFT PLAIN - PETTY (\$950 & U.. | | | 14.27% | | | | | | | | | 12.72% | 2.72% |
| THEFT OF IDENTITY | 19.48% | 27.06% | 24.26% | 28.99% | 30.03% | 31.96% | 28.94% | 30.05% | 28.94% | 27.78% | 26.58% | 26.32% | 27.52% |
| THEFT FROM MOTOR VEHICLE.. | | | | | | | | | | 13.78% | 15.54% | 13.96% | 4.21% |
| BURGLARY FROM VEHICLE | 23.75% | 20.38% | 16.28% | 20.22% | 18.08% | 19.08% | 17.33% | 17.81% | 18.20% | 15.35% | 14.55% | 13.65% | 17.61% |
| BURGLARY | 20.56% | 18.94% | 15.77% | 16.97% | 19.25% | 18.23% | 19.39% | 17.93% | 18.69% | 14.73% | 17.28% | 13.50% | 17.37% |

Figure 10. (Thefts > 1000)

For any given weekday data, we can see a downward trend with a weekend being the least theft-intensive. (Figure 11.)



Figure 11.

As for the weekday cohort, we can identify that four types of theft are persistent throughout a single day. *Theft of Identity* continues to be consistent as it was in the monthly cohort and makes up close to 20% of all thefts that are being reported any given day. (Figure 12.)

| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday | Total |
|-----------------------------|--------|---------|-----------|----------|--------|----------|--------|--------|
| VEHICLE - STOLEN | 34.21% | 19.51% | 30.73% | 17.72% | 21.64% | 21.11% | 32.64% | 22.12% |
| THEFT-GRAND (\$950.01 &.. | | 6.38% | | 6.29% | 6.74% | 7.06% | | 5.42% |
| THEFT PLAIN - PETTY (\$95.. | | 9.22% | | 9.59% | 9.24% | 9.21% | | 7.67% |
| THEFT OF IDENTITY | 30.05% | 17.17% | 31.96% | 19.00% | 18.39% | 15.96% | 30.03% | 19.91% |
| THEFT FROM MOTOR VEH.. | | 8.56% | | 10.09% | 6.92% | 8.59% | | 7.06% |
| SHOPLIFTING - PETTY THE.. | | 3.82% | | 5.31% | | | | 2.07% |
| ROBBERY | | 6.39% | | 5.42% | 6.88% | 6.72% | | 5.22% |
| BURGLARY FROM VEHICLE | 17.81% | 11.16% | 19.08% | 10.73% | 11.95% | 12.65% | 18.08% | 12.74% |
| BURGLARY | 17.93% | 11.42% | 18.23% | 10.80% | 11.52% | 11.48% | 19.25% | 12.57% |
| THEFT FROM MOTOR VEH.. | | 6.38% | | 5.06% | 6.71% | 7.22% | | 5.21% |

Figure 12. (Thefts > 1000)

A strange insight could be drawn when talking about what hour of the day most *Theft of Identity* occurs at first glance. (Figure 13.)

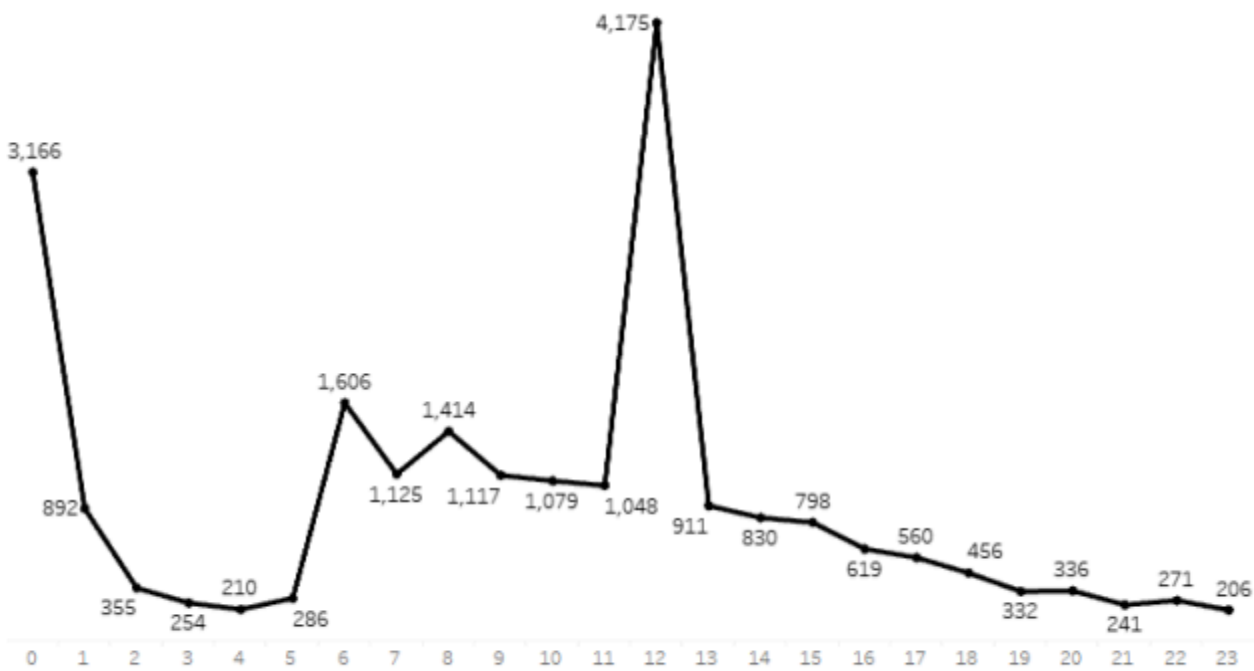


Figure 13.

The line graph for any given hour has 2 points that need explaining. *Theft of Identity* crime reporting has an issue when the police have to report the exact time of this crime. While you can describe the time of an item theft way easier due to surveillance or witnesses, for identity theft it is very hard to pinpoint the exact time since for the most part a victim would suspect that his/her identity was stolen once the criminal is caught/seen after or during the other crime he/she committed. For this reason, police have 2 default times, one of them being the middle of the night and the other the middle of the day. The default times are designed to keep reporting data consistent.

If we omit default times as outliers we can identify that this crime is reported mostly in the early morning and trends down during the day. A more specific point could be made that crime starts before working hours when the city is most active. (Figure 14.)

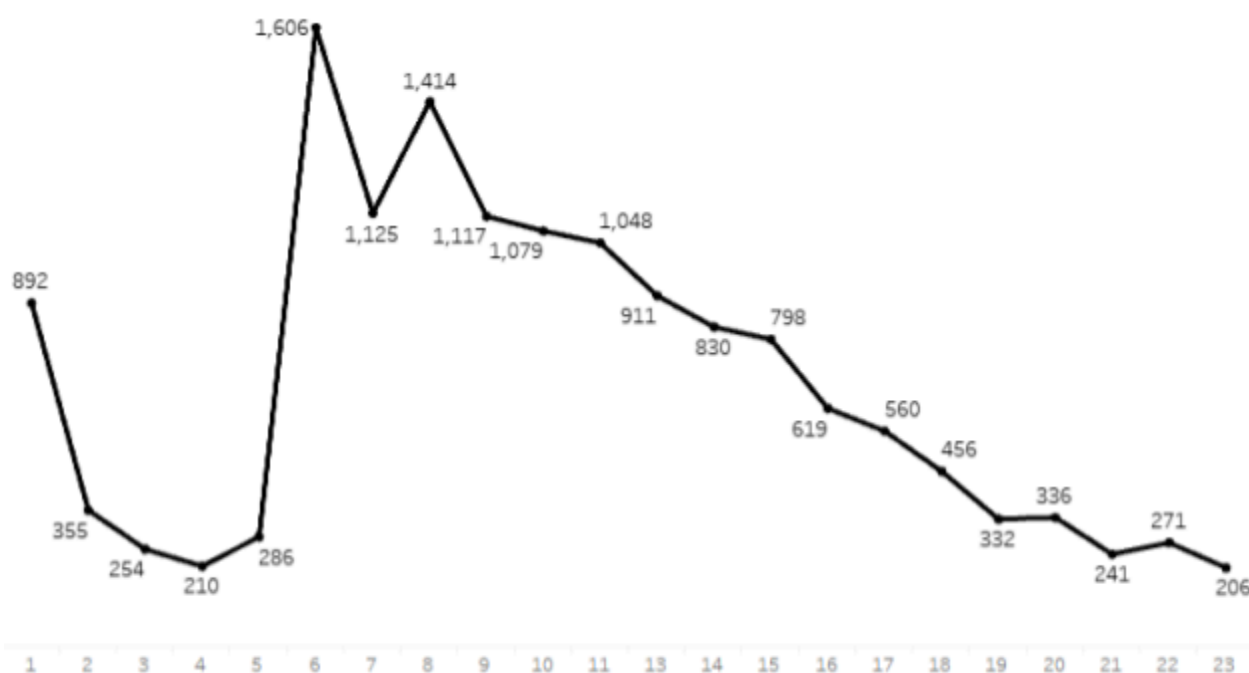


Figure 14. (Default times excluded)

In conclusion, the overall trend of time would bet that *Theft of identity* occurs every month, day and hour and for more specific insight is that data suggests that this crime occurs mostly during workdays and working hours.

Any crime has time, which we have discussed and while reporting could face some challenges, especially pinpointing the exact time of identity theft. Trends and insights about time data can help to focus resources on a specific time frame. Another attribute of any crime is a place which we are going investigate further.

3.3. Location

Having collected victims' data and investigated time trends next step is to try to isolate a crime's location. Metropolitan L.A. is divided into 21 districts. (Figure 15.)

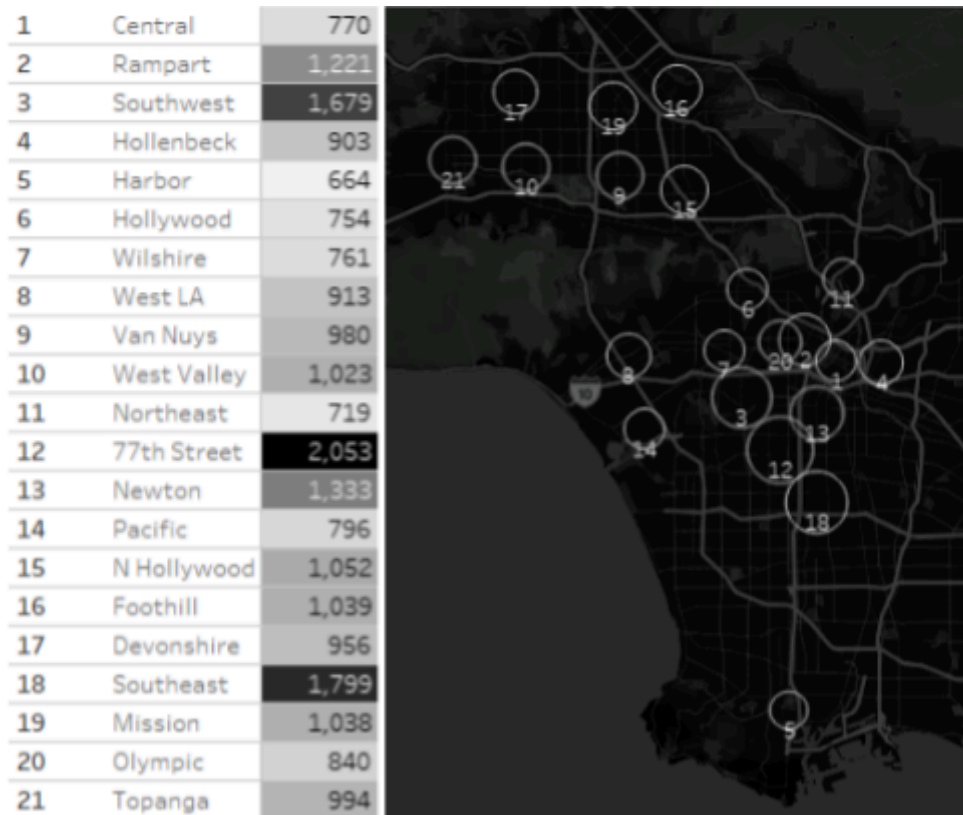


Figure 15.

Looking at the numbers we can see that *77th Street*, *Southeast*, and *Southwest* (represented by numbers 12, 18, and 3 respectively) are the top 3 districts where *Theft of Identity* occurred in 2022. To see if the numbers correlate with districts and have significance in locating the most common crime location we are going to use the results of linear regression. After running linear regression and looking at the results we can see that for the most part districts' identity theft numbers from

| | Coefficient | Std. Error | t-ratio | p-value |
|------------|-------------|------------|---------|---------|
| const | 0,0684476 | 0,00276370 | 24,77 | <0,0001 |
| Southeast | 0,0847693 | 0,00385674 | 21,98 | <0,0001 |
| 77thStreet | 0,0720246 | 0,00366721 | 19,64 | <0,0001 |
| Foothill | 0,0636903 | 0,00429394 | 14,83 | <0,0001 |

Table 3. Coefficient > 0,05

before (Figure 15.) do imply a correlation. (Table 3.)

The results based on linear regression suggest that districts *Southeast*, *77th Street*, and *Foothill* compared to other districts, while minimal, are impact zones and should help to narrow down the search.

To narrow down the location even closer we need to look at near what premises *Theft of Identity* occurred the most. Given that multiple districts share some of the streets looking at premises first is a good approach. (Figure 16.)

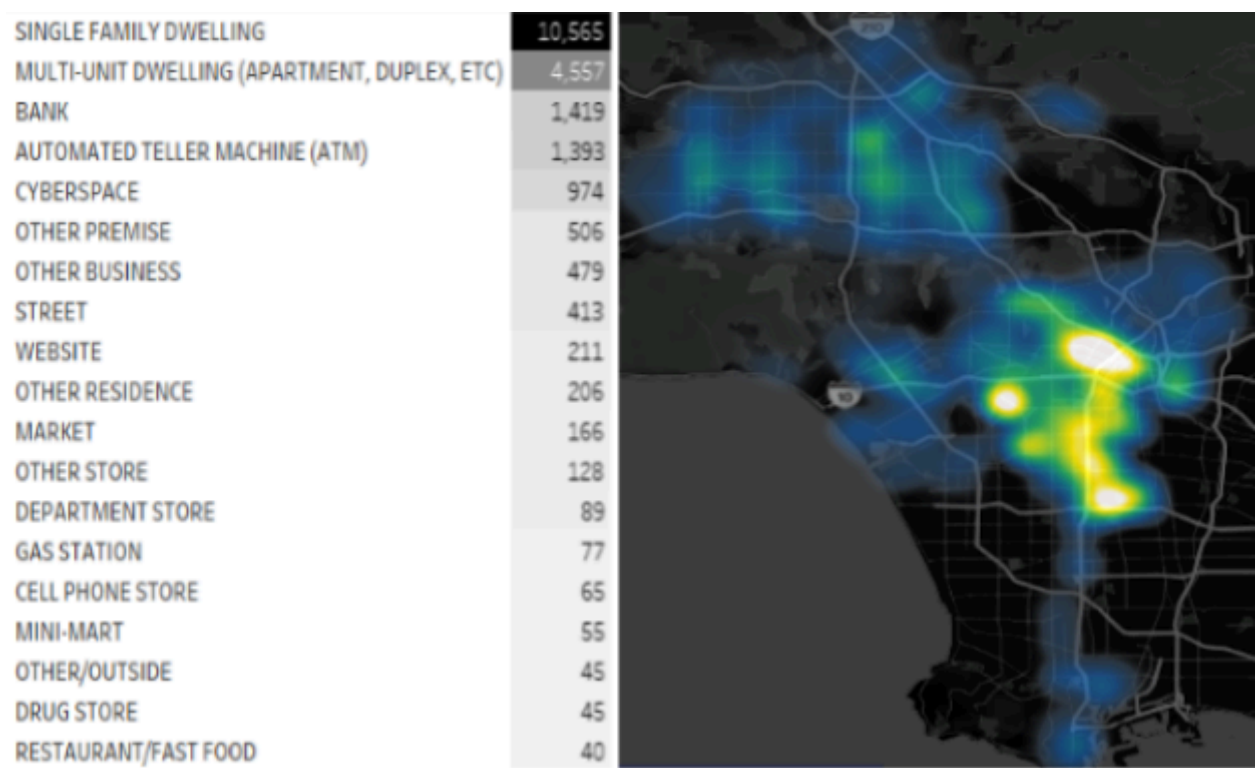


Figure 16.

Single-family homes, apartments, banks and ATMs are the majority of places where identity theft was reported to happen. Looking at the heat map most thefts that are reported are concentrated around the city centre and spread south, but keep in mind that the city centre by itself contains three very next to each other districts (Figure 16.). While they might not have huge numbers of identity theft reports, districts being close to each other creates an overlap.

As with districts' numbers, we have to see if there is a correlation between the *Theft of Identity* numbers and premises or specific locations. For premises, we are going to use

logistic regression to classify if a premise or multiple of them could be classified as a place where identity theft occurs. As with linear regression results, we are going to isolate the most significant values⁴. *Banks, ATMs, and Cyber Space* are the places which have significant links to identity theft. (Table 4.)

| | <i>Coefficient</i> | <i>Std. Error</i> | <i>z</i> | <i>p-value</i> |
|-------------|--------------------|-------------------|----------|----------------|
| const | -2.4405 | 0,008 | -318.376 | <0,0001 |
| BANK | 4.0528 | 0,066 | 61.825 | <0,0001 |
| ATM | 5,4170 | 0,122 | 44.436 | <0,0001 |
| CYBER SPACE | 5,3329 | 0,140 | 38.089 | <0,0001 |

Table 4.

Identity theft based on the Department of Justice (DOJ) Criminal Divisions Fraud Section's definition "Identity theft and identity fraud are terms used to refer to all types of crime in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud or deception, typically for economic gain."⁵ Definition makes sense for our data and clarifies the motivation behind identity theft.

Having premises in context we can see what crimes they might commit once they have stolen someone's identity (Figure 17.) Note that *Bunco* could mean a squad dealing with fraud or in street slang it means to cheat someone by selling false goods.

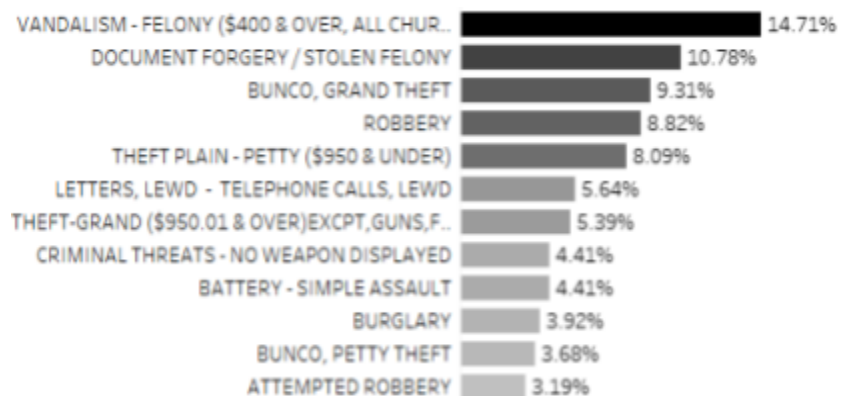


Figure 17. (crimes > 3%)

We can see that a lot of crimes

include other thefts and forgery. Document forgery could be reinforced by DOJ which links identity theft to."

⁴ When p-values are equal for even more precise significance evaluation looking at the coefficient is good way to not only to see if variable is impacting the outcome but also in which direction.

⁵ Identity Theft <https://www.justice.gov/criminal/criminal-fraud/identity-theft/identity-theft-and-identity-fraud>

- False applications for loans and credit cards,
- Fraudulent withdrawals from bank accounts,
- Fraudulent use of telephone calling cards or online accounts, or
- Obtaining other goods or privileges which the criminal might be denied if he were to use his real name.”

All geographical data and checking for correlation gives an overview and ability to pinpoint the streets and premises where identity theft could have been used to commit other crimes. There are close to 43 thousand unique street names in the L.A. metropolitan area. Two of them make up almost 3% of all crimes linked to identity theft. The concern is that both of these streets are part of the *Newton* district and are exclusive to that district. Having that in mind suggestion would be to investigate these street Banks and ATMs if they have good surveillance. Given that

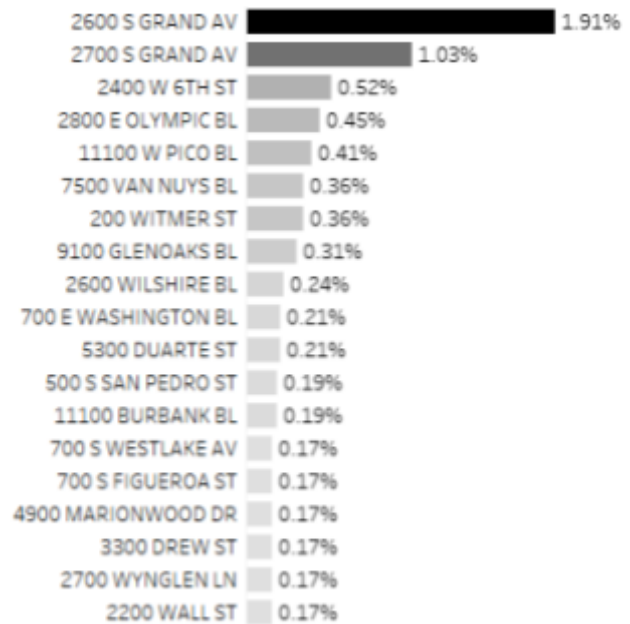


Figure 18.

identity theft could be as simple as using someone's stolen credit card information by simply looking over the shoulder near an ATM it is hard to do something about it but to raise awareness and inform people. For Banks specifically, police can inform them that if their bank is in that ‘problematic’ street perform strict identity checks. The trickiest of places is *CyberSpace* where anonymity is defended by the law, while the hard-to-battle problem of cyberspace, awareness can be spread through social media to build people's literacy about cyber crimes.

This report analysed victims, time and location to target police resources and in summary, based on data we can say that theft of identity is targeted at adults during working hours and half of the time is targeted at people of black descent.

5. Conclusion

From the above analysis we can establish that overall crimes have increased and continue to trend upward. The biggest increase was observed for the *Theft of Identity* crime. This crime was reported in every district throughout every month, day and hour. Time data's cohort analysis shows not only crime's consistency but also that this crime is a major share of reports of all thefts that occur. The victims of said crime tend to be Hispanic/Latin/Mexican ethnic groups in terms of amount but in terms of share of all thefts, identity theft is mostly reported by people of black descent. Females are more likely to report that their identity was stolen and the age of victims is between 21 and 40. *Southeast*, *77th Street*, and *Foothill* district correlate with identity theft and notably *77th Street* district where identity theft was reported the most. *Theft of Identity* is a tool for committing crimes concentrated around ATMs, Banks and Cyberspace. *2600 S Grand AV* and *2700 S Grand AV* are two streets part of *Newton* district where identity theft-related crimes occur the most. Starting to investigate *77 Street* district would help to identify why *Theft of Identity* is so prevalent there and try to figure out if this district has a link with *Newton* district where most of the crimes using stolen identity occur.

6. Recommendations

- Allocate recourses to the most prominent crime areas (districts, streets).
- Inform Banks to tighten up their identity check policies.
- Build residents' awareness through public announcements, city billboards or social media.
- Investigate what is specific about the 2 streets and the district itself.
- Investigate if there is a link between *77th Street* and *Newton* district to commit a crime using a stolen identity.
- Refine data collection procedures to avoid terms Unknown/Other/Undisclosed in your data.
- Perform a deeper analysis using the provided dashboard to investigate other crimes and what is specific to them and perform data-driven decisions.

7. References

1. Demographics of Los Angeles

https://en.wikipedia.org/wiki/Demographics_of_Los_Angeles

2. Data <https://www.kaggle.com/datasets/asaniczka/crimes-in-los-angeles-2020-2023>

3. Official data set

https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data

4. Identity Theft

<https://www.justice.gov/criminal/criminal-fraud/identity-theft/identity-theft-and-identity-fraud>

5. Identity Theft by DOJ

<https://www.justice.gov/criminal/criminal-fraud/identity-theft/identity-theft-and-identity-fraud>