

Metody analýzy dat I - analýza datasetu

Filip Peterek, PET0342

28. prosinec 2021

1 Popis datové sady

Datová sada obsahuje seznam nahlášených zločinů za rok 2018 ve městě Austin v americkém státě Texas. Data byla sbírána policií města Austin a následně veřejně publikována [1]. Datová sada obsahuje 37156 záznamů odpovídajících 37156 nahlášeným zločinům.

1.1 Atributy datové sady

V datové sadě se objevuje 13 atributů, které jsou popsány v tabulce 1.

2 Výsledky analýzy

2.1 Počty zločinů

Počty jednotlivých zločinů, viz graf 1, jistě nijak zarážející nejsou. Naprosto převažují krádeže, v tomto případě se z velké části s vysokou pravděpodobností jedná převážně o kapsářství a podobné drobné krádeže. Počet vloupání do domu je v porovnání s drobnějšími, méně závažnými krádežemi méně než pětina. Násilné činy tak časté nejsou, přesto se jich několik objevuje.

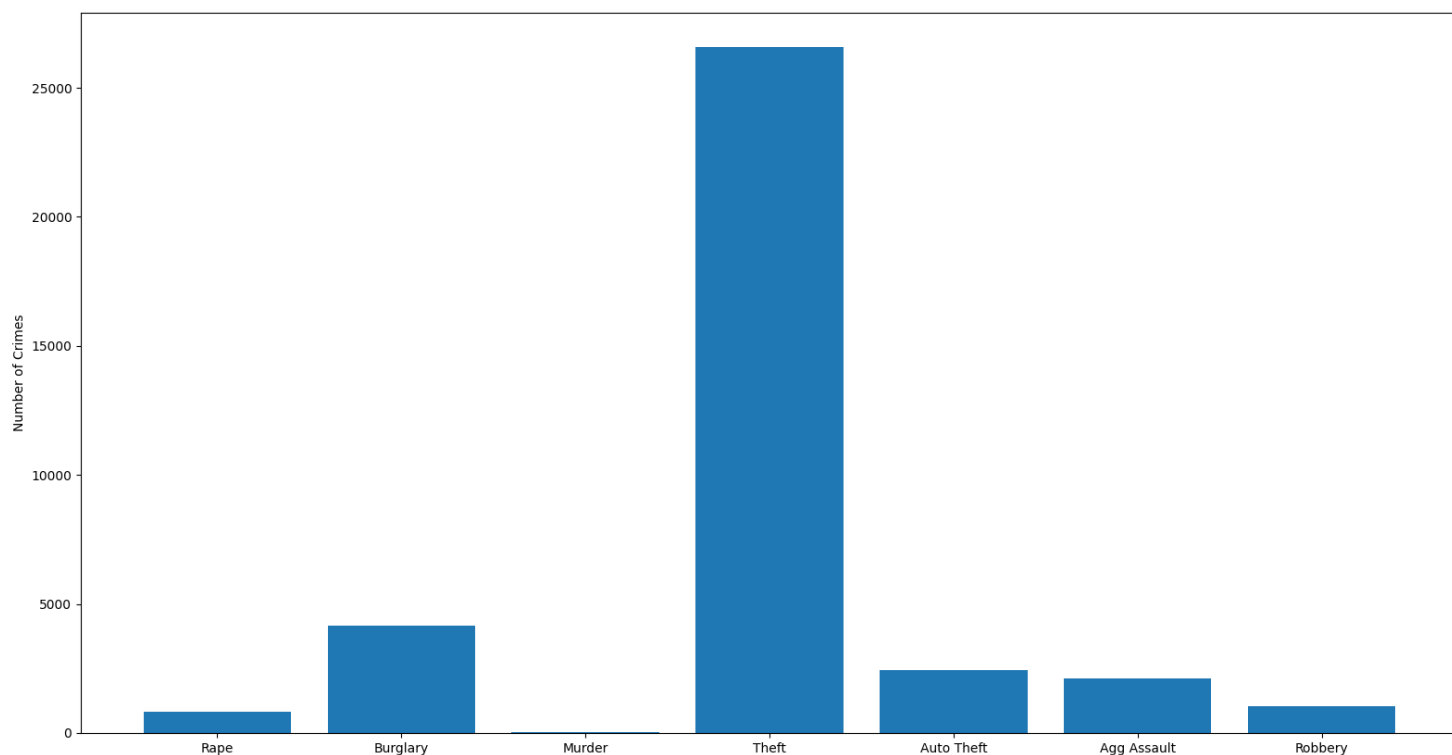
Velmi znepokojující jsou počty nahlášeného násilí na dětech (graf 2). Austin je velké město, a proto se s vysokou pravděpodobností objeví zločiny všech možných typů, přesto řádově stovky případů zneužívání dětí je alarmující číslo.

Počet nahlášených zločinů podle kalendářního měsíce (graf 3) není nijak zajímavý. Nejméně zločinů bylo nahlášeno v únoru – nepřekvapivě, neboť se jedná o nejkratší měsíc v roce. Nejvíce zločinů bylo nahlášeno v říjnu. Překvapivě může působit fakt, že nedošlo k nárůstu zločinnosti v letních měsících. Austin sice nepatří mezi nejvyhledávanější turistické destinace ve Spojených státech, přesto se však jedná o město s bohatým kulturním vyžitím. I tak ale v létě zločinnost nestoupá.

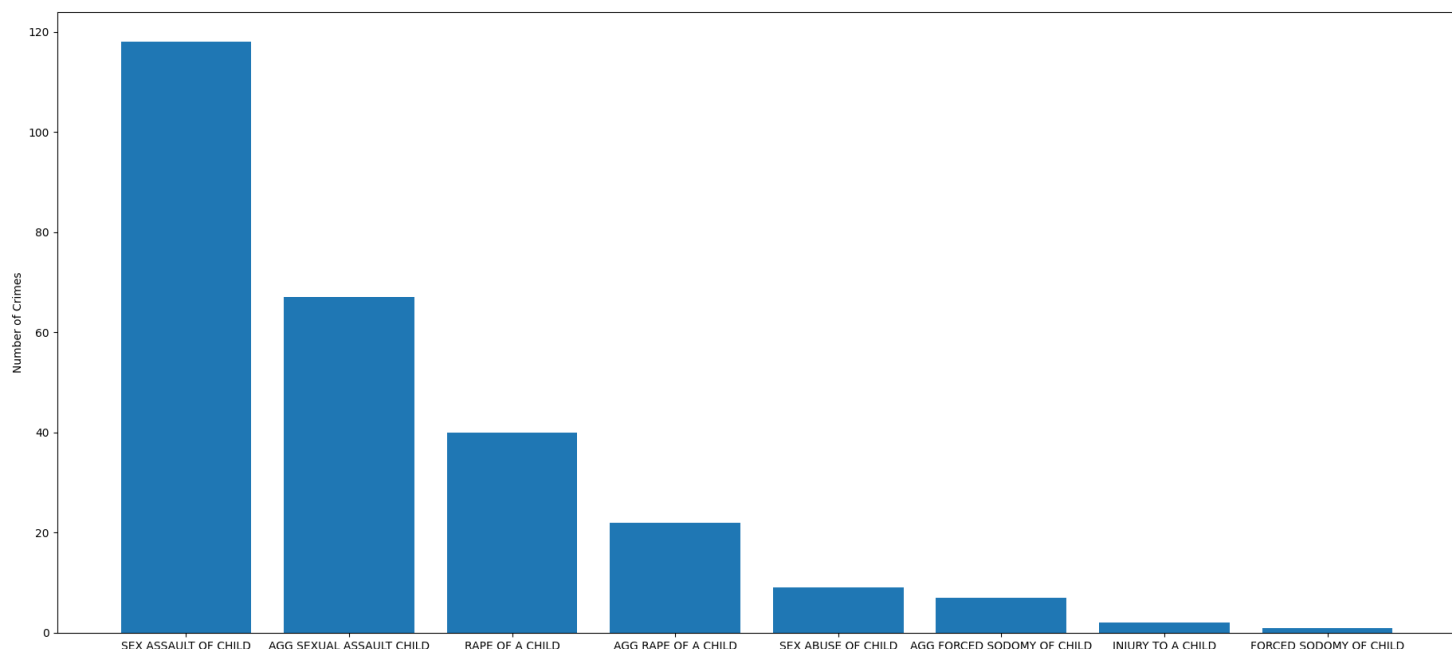
Drobný nárůst zločinnosti v říjnu může vzbudit určitou zvědavost, při podrobnějším prozkoumání ovšem zjistíme, že se neobjevily žádné zajímavé zločiny (např. hromadné nahlášení daňových úniků), jednalo se převážně o drobné krádeže. Přesto tato skutečnost může být zajímavá pro policejní oddělení města Austin.

Tabulka 1: Popis atributů datové sady

Název	Popis	Typ	Datový typ	Povinný
GO Primary Key	ID incidentu	numerický	celé číslo	ano
Council District	městská čtvrť	kategoriální	celé číslo	ne
GO Highest Offense	zločin	kategoriální	řetězec	ano
Crime Type	typ zločinu	kategoriální	řetězec	ano
GO Report Date	den nahlášení	numerický	datum	ano
GO Location	místo odehrání zločinu	kategoriální	řetězec	ne
GO X Coordinate	místo odehrání zločinu	numerický	desetinné číslo	ne
GO Y Coordinate	místo odehrání zločinu	numerický	desetinné číslo	ne
Clearance Status	stav policejního pátrání	kategoriální	řetězec	ne
Clearance Date	datum vyřešení případu	numerický	datum	ne
GO District	policejní sektor	kategoriální	řetězec	ano
GO Location Zip	poštovní směrovací číslo	kategoriální	celé číslo	ne
GO Census Tract	čtvrť podle počtu obyvatel	kategoriální	desetinné číslo	ne



Obrázek 1: Počet nahlášených zločinů podle kategorie



Obrázek 2: Počet nahlášených zločinů proti dětem

2.2 Řešení zločinů

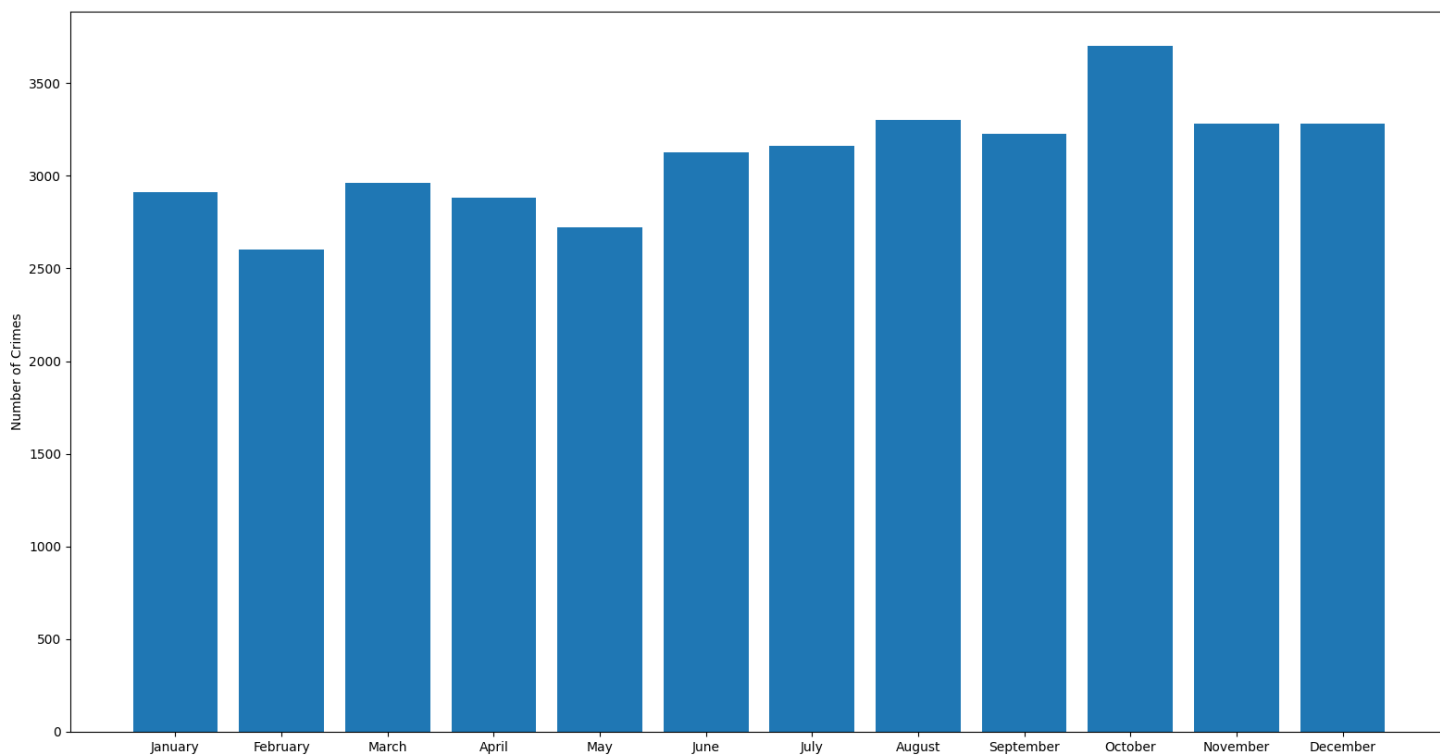
Vzorného občana určitě znepokojí množství nevyřešených případů. Z grafu 5 lze vyčíst, že naprostá většina případů zůstává nevyřešena.

Z grafu průměrné délky řešení případu (graf 6) zjistíme, že nejkratší dobu trvá řešení krádeží. Graf znázorňuje pouze úspěšně vyřešené případy. Nejrychleji jsou v průměru řešeny krádeže. Nejdéle jsou řešeny případy znásilnění, jejichž řešení trvá v průměru přes 80 dní. Možná překvapivě jsou případy vražd a zabití vyřešeny v průměru do jednoho měsíce.

Pokud se podíváme na pět nejdéle řešených případů (graf 7), zjistíme, že byly všechny řešeny déle než rok. Jinak ale nezjistíme žádný zajímavější trend, mezi nejdéle řešenými případy najdeme různé typy zločinů. Překvapivě působí pouze fakt, že jsou délky řešení jednotlivých případů daleko za průměrem úspěšného řešení konkrétního typu zločinu.

Ovšem podíváme-li se na průměrnou délku řešení případu, medián délky, a varianci (graf 2) (včetně uzavření nevyřešených případů), zjistíme, že průměr (a ani medián), nejsou příliš vypovídající.

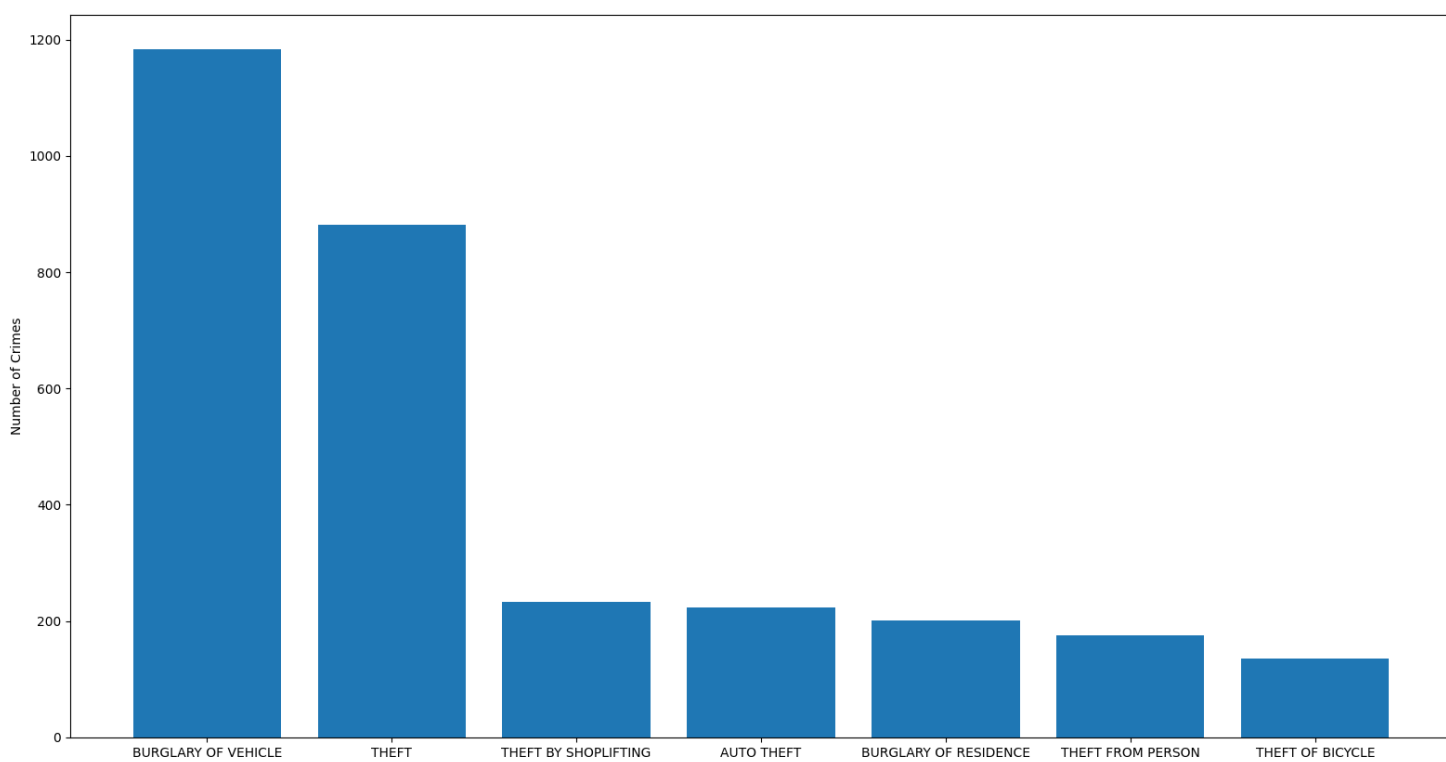
Počty uzavřených případů seskupených podle kalendářního měsíce (graf 8) převážně kopírují počty nahlášených případů. Nejvíce případů bylo uzavřeno v říjnu a listopadu, což pouze reflektuje fakt, že bylo v říjnu nahlášeno nejvíce případů – z velké části se jednalo převážně o drobné krádeže, které jsou zpravidla



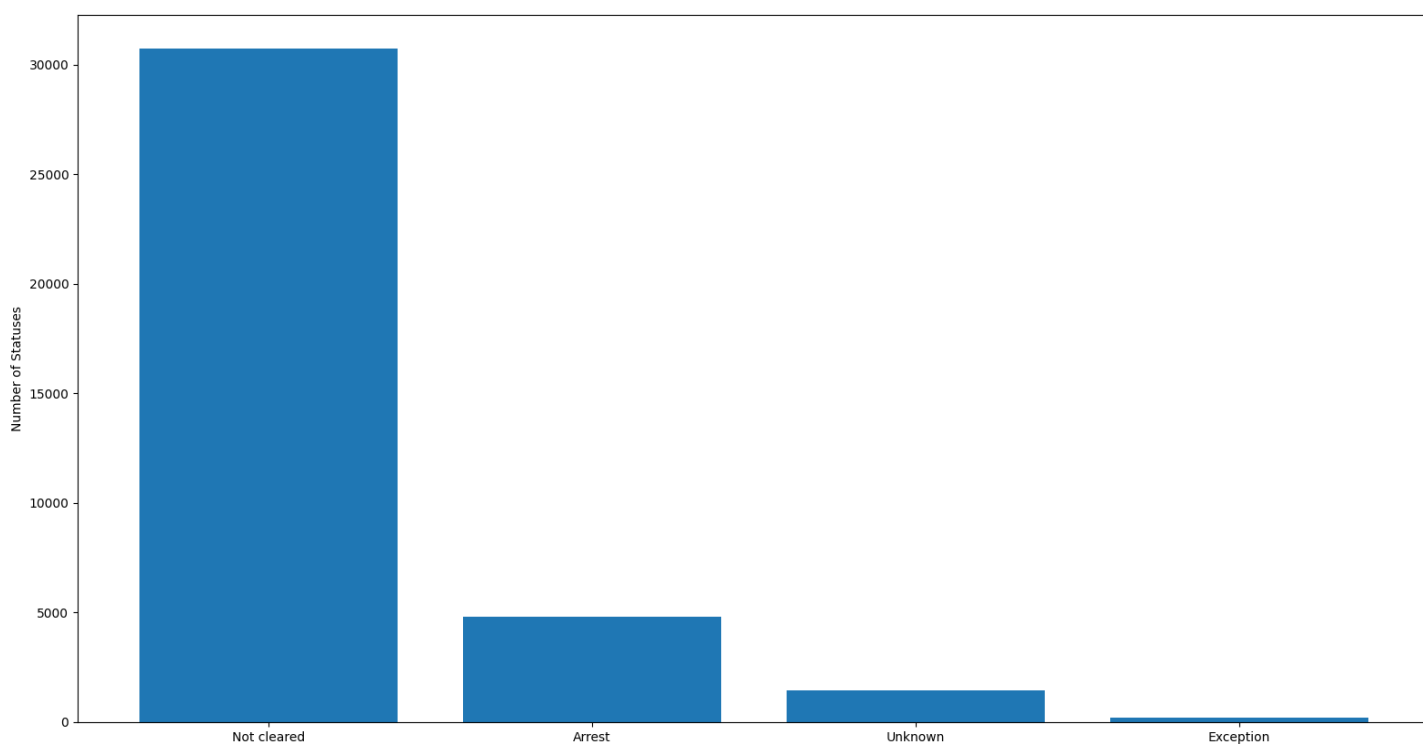
Obrázek 3: Počet nahlášených zločinů podle kalendářního měsíce

Tabulka 2: Délka řešení případu

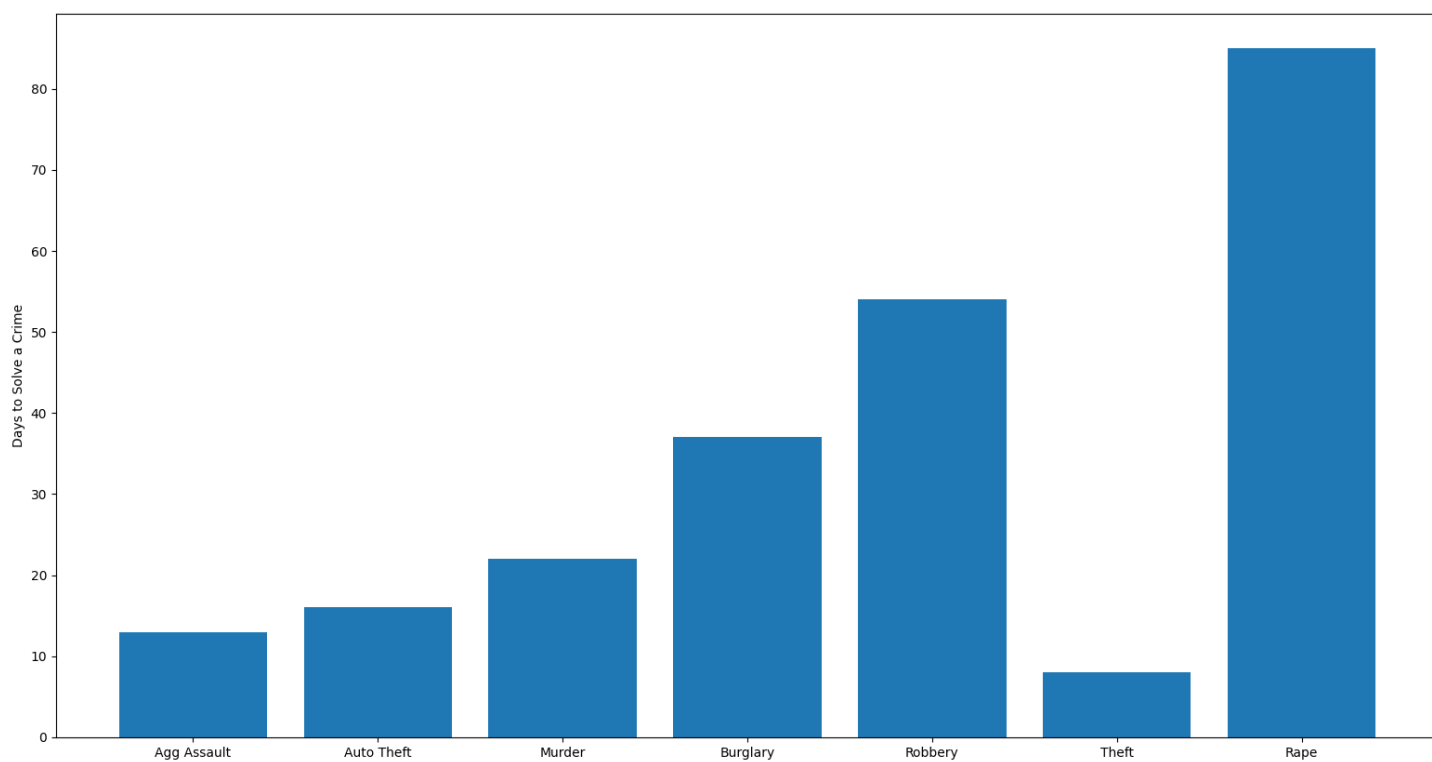
Průměr	16.0
Medián	8.0
Variance	1172



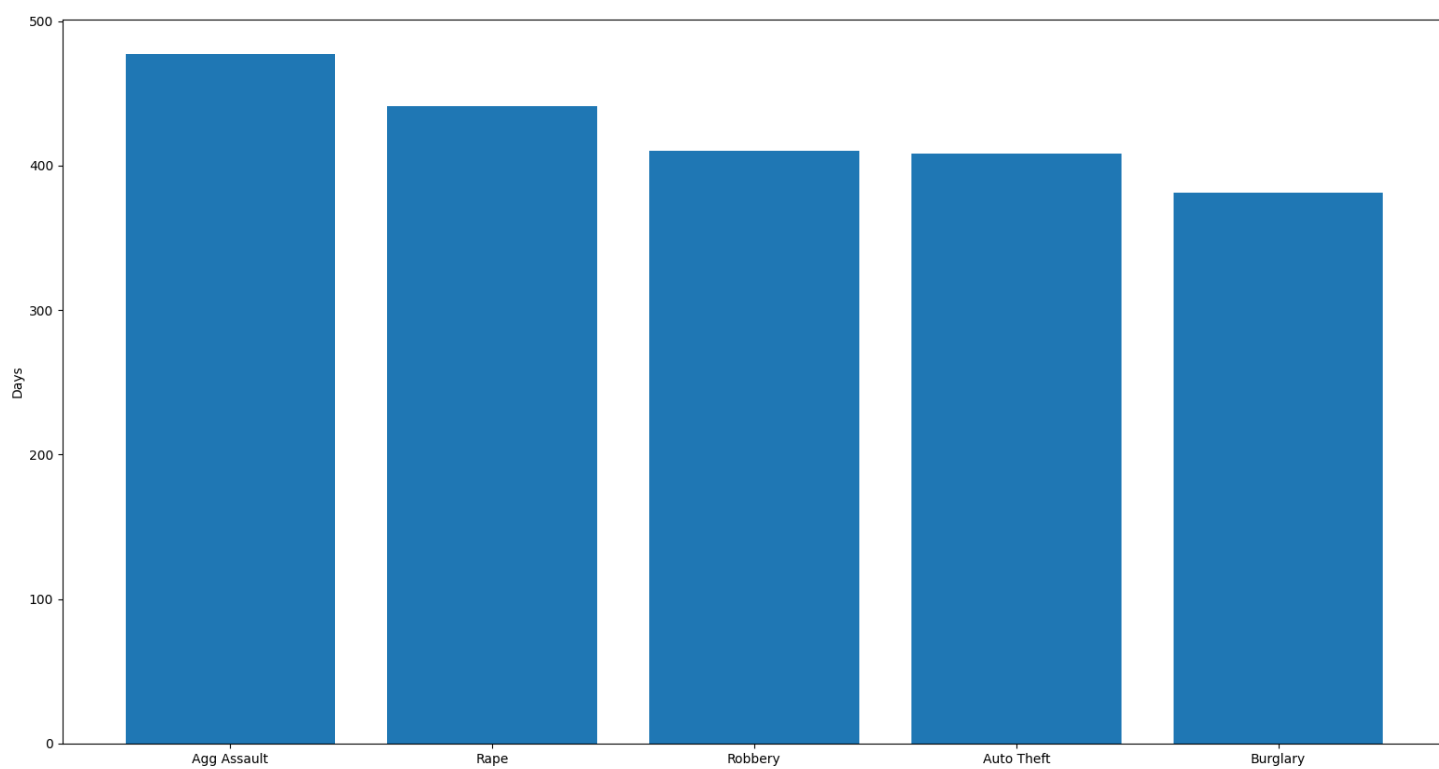
Obrázek 4: Sedm nejčastějších zločinů v říjnu 2018



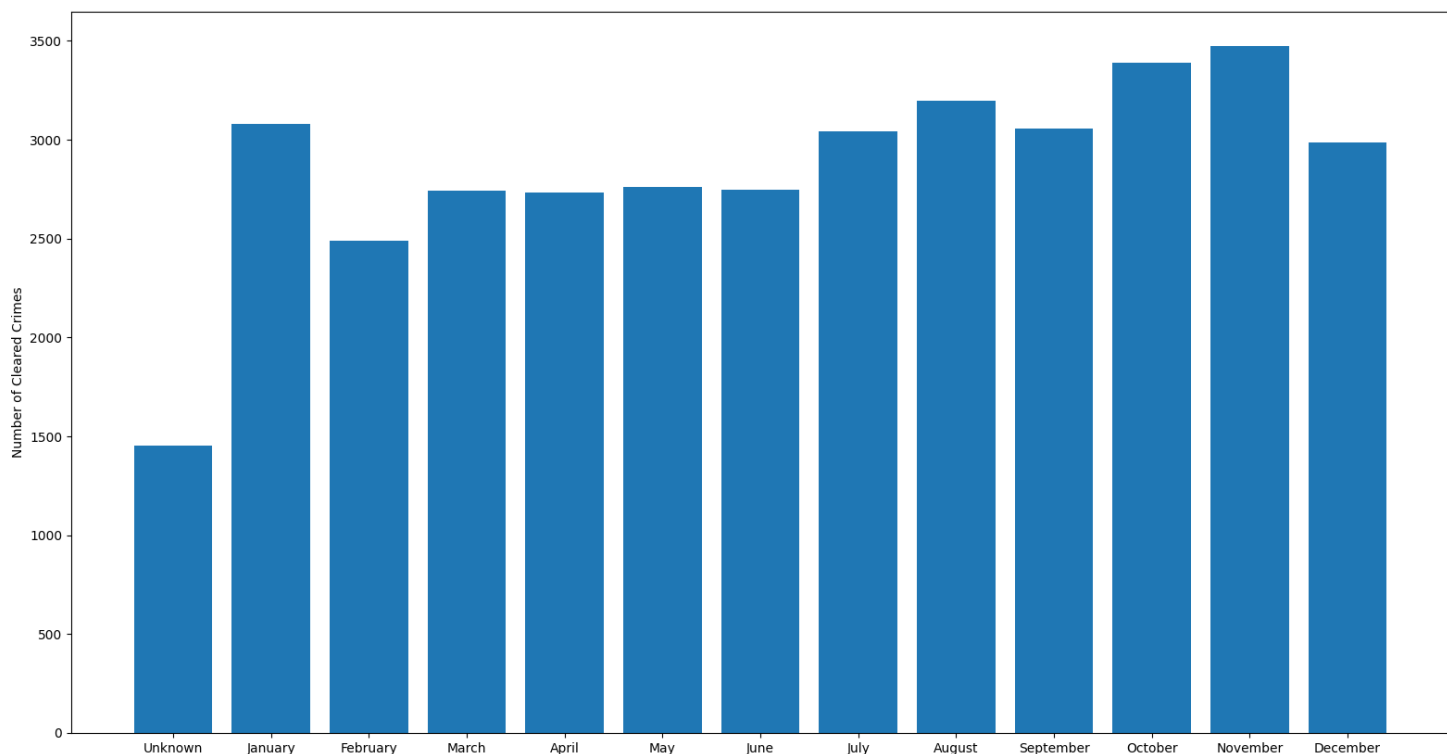
Obrázek 5: Četnosti statusů policejního pátrání



Obrázek 6: Průměrná doba úspěšného řešení



Obrázek 7: Nejdéle řešené jednotlivé případy



Obrázek 8: Počet uzavřených případů podle kalendářního měsíce

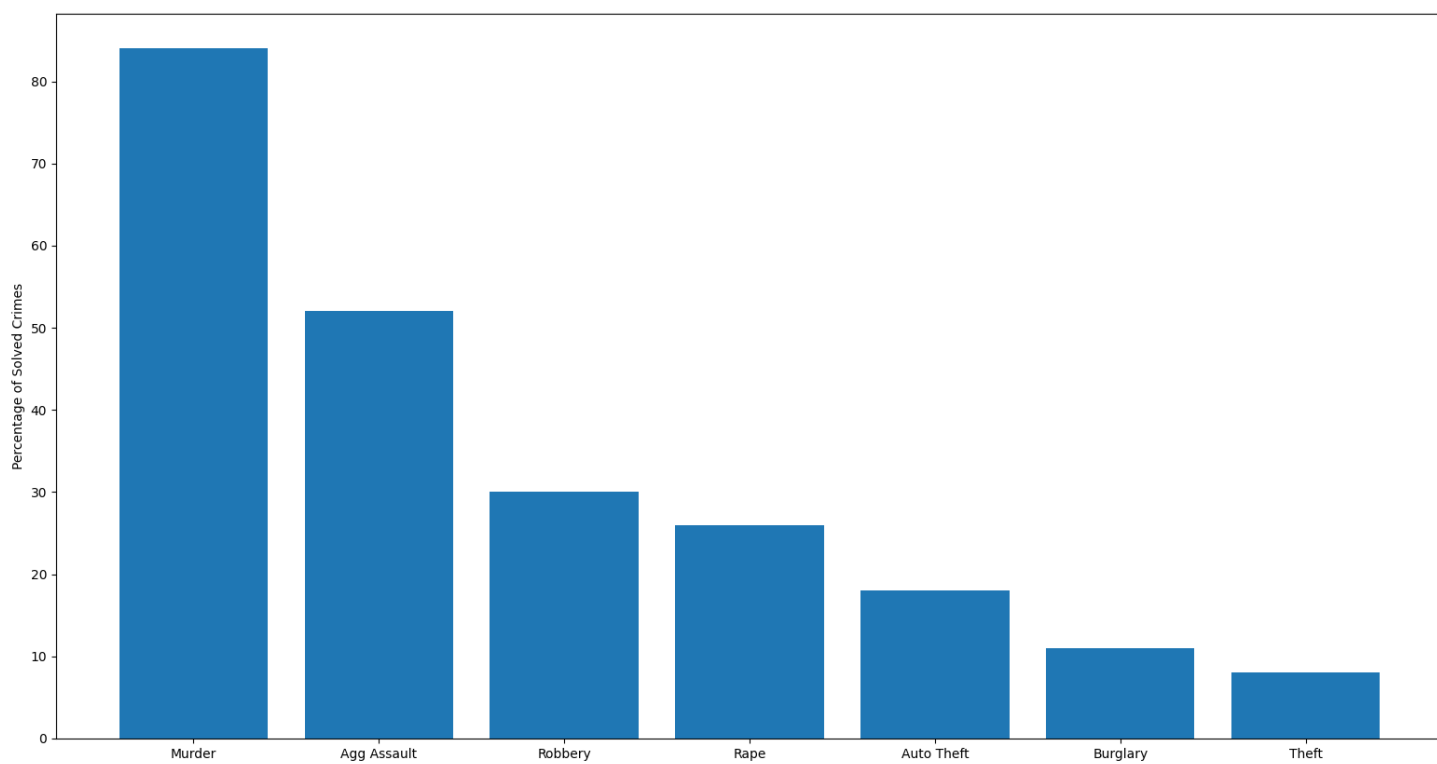
uzavírány velmi rychle. Nejméně případů bylo vyřešeno v únoru. Tato informace je také nepřekvapivá, jelikož se jedná o nejkratší měsíc v roce.

Děsivě může působit poměr vyřešených případů v poměru k celkovému počtu případů (graf 9). Méně než pětina vloupání, přibližně pětina krádeží aut, a přibližně čtvrtina znásilnění skončila úspěšným vyřešením případu (zatčením pachatele nebo výjimečným řešením). Statistika násilných napadení je poněkud lepší, více než polovina případů skončila úspěšným vyřešením. Nejlepší statistiku nalezneme u zabití (ať už úmyslných nebo neúmyslných). 27 z celkového počtu 32 případů zabití, tedy přes 80 %, skončilo nalezením pachatele. Z grafu tedy vyplývá, že trváte-li nutně na dopadení pachatele, musíte se nechat zavraždit.

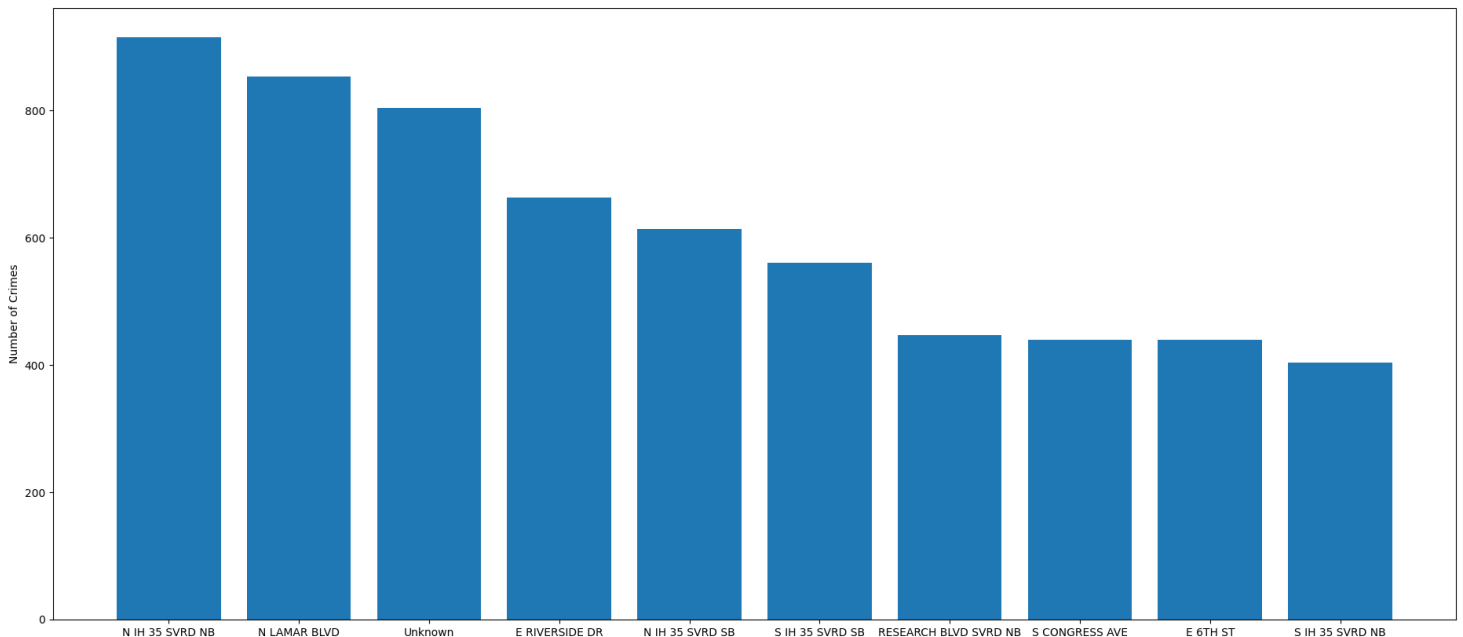
2.3 Bezpečnost městských částí

Případného návštěvníka či obyvatele texaského města Austin může zajímat statistika nejméně bezpečných ulic (graf 10). Při bližším prozkoumání však zjistíme, že ty nejméně bezpečné ulice jsou zároveň jedny z nejdelších ulic v Austinu, a jedná se o ulice, které se většinou táhnou přes polovinu města.

Z charakteristiky datové sady vyplývá, že datová sada nebude obsahovat



Obrázek 9: Poměr úspěšně vyřešených případů (v procentech)



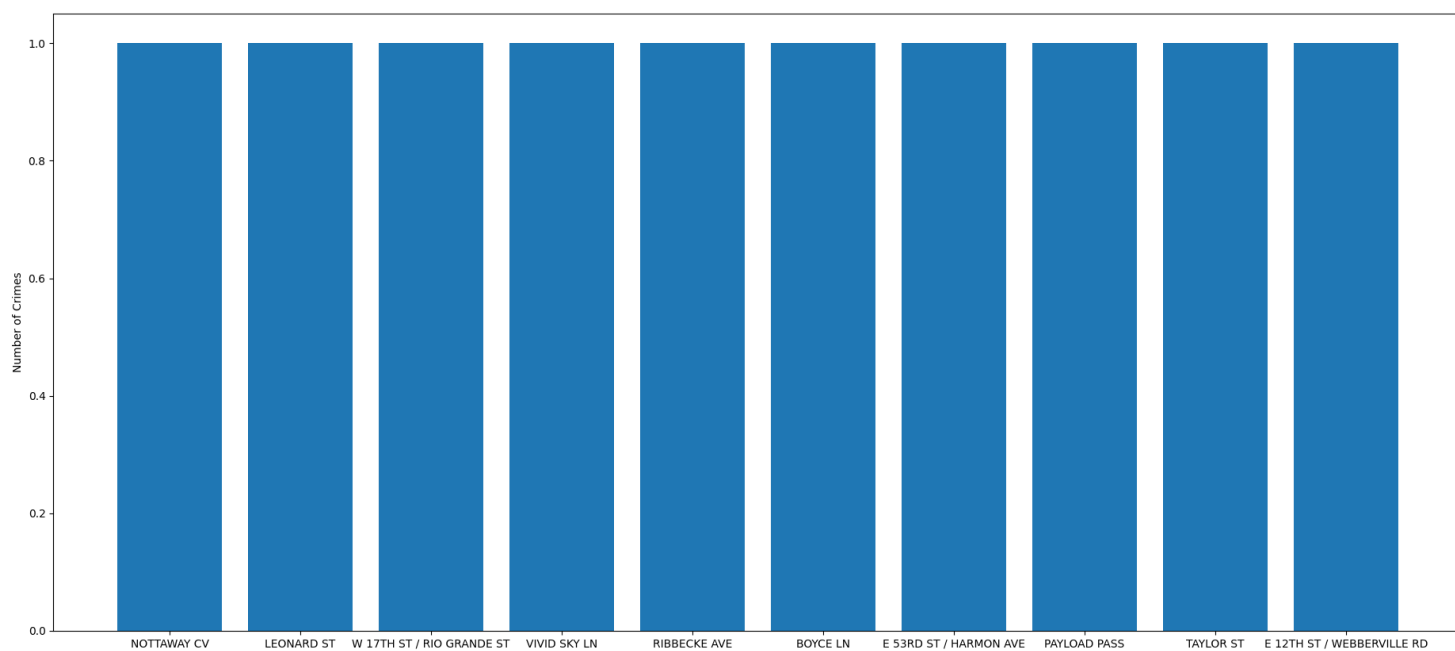
Obrázek 10: Ulice s nejvyšším počtem nahlášených zločinů

ulice, na kterých k žádnému zločinu nedošlo. Mezi ulicemi, na kterých byl nahlášen alespoň jeden zločin, najdeme spoustu ulic, kde došlo k nahlášení právě jednoho zločinu (graf 11). Proto statistika nejbezpečnějších ulic není příliš vypovídající.

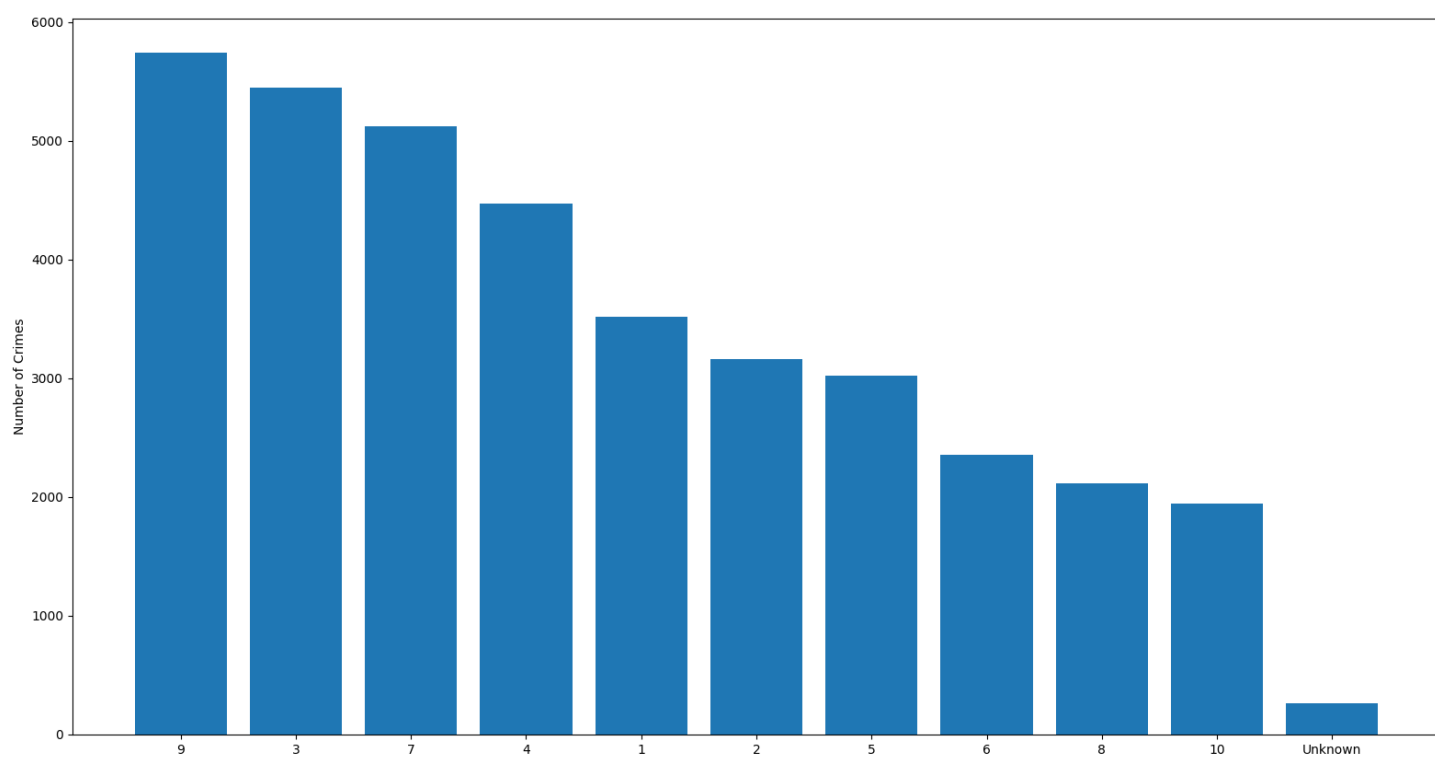
Zákony dodržujícího občana bude zajímat také seřazení čtvrtí podle jejich bezpečnosti. Nejméně bezpečnou čtvrtí je čtvrt' číslo devět. Čtvrt' číslo devět je ovšem také centrum města Austin. Nejbezpečnějšími čtvrtěmi jsou poté čtvrti šest, osm a deset. V případě těchto čtvrtí se zase jedná o odlehlejší a klidnější čtvrti. [2]

2.4 Slaughter lane

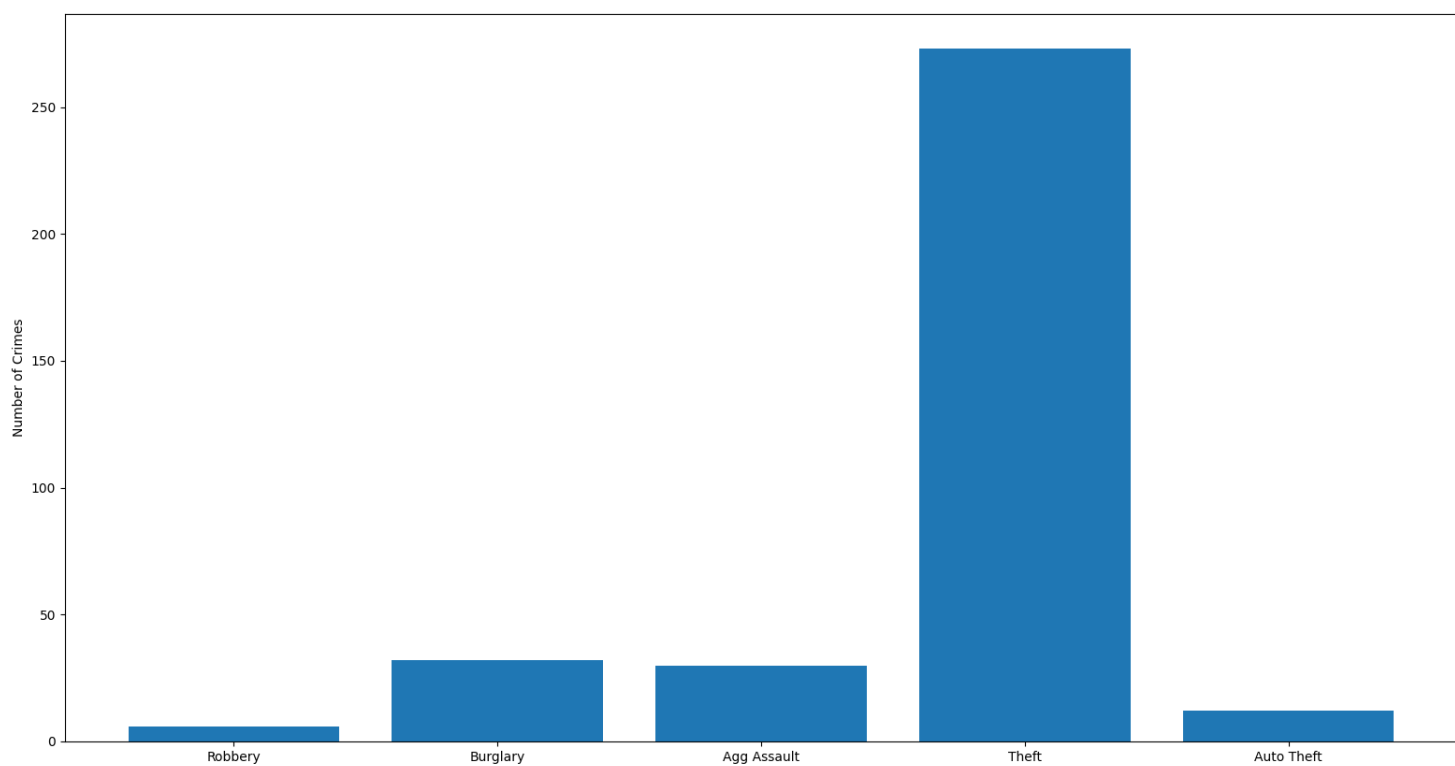
Určitou kuriozitou je existence ulice s názvem *Slaughter Lane*. Opět se jedná o relativně dlouhou ulici, proto nikoho nemůže překvapit vyšší počet nahlášených zločinů. Kuriózní je ovšem fakt, že na východní ani západní *Slaughter Lane* nebyla nahlášena ani jedna vražda nebo zabití. Lze tedy říct, že se jedná o tzv. false advertising a ulice *Slaughter Lane* si své jméno nezaslouží.



Obrázek 11: Ulice s nejnižším počtem zločinů (ulice bez zločinů jsou ignorovány)



Obrázek 12: Čtvrti seřazené podle počtu nahlášených zločinů



Obrázek 13: Četnosti zločinů na ulici Slaughter lane

Tabulka 3: Celoamerická statistika

Zločin	Celkový počet	Počet na 100 tisíc obyv.
Znásilnění	127945	44.4
Vraždy a zabití	14905	5.0
Přepadení	265464	89.9
Napadení	738590	250.0
Vloupání	1097816	371.6
Krádeže vozidla	692172	234.3
Krádeže majetku	4758121	1610.5

3 Srovnání se zbytkem Spojených států

Podle oficiálních dat Spojených států [3] žilo v Austinu v roce 2010 asi 790 tisíc obyvatel. V roce 2020 počet obyvatel vzrostl na přibližně 960 tisíc. Průměrný roční přírůstek tedy byl asi 17 tisíc nových obyvatel. Pro zjednodušení budeme počítat, že Austin rostl lineárně a v roce 2018 měl Austin 926 tisíc obyvatel. Přesná data neznáme, neboť se počet obyvatel v průběhu roku měnil, ale také proto, že v daném roce neproběhlo sčítání obyvatelstva.

Počet obyvatel je důležitý pro relativní srovnání se zbytkem Spojených států. Tabulka 3 vykazuje statistiku zločinu agregovanou pro celé Spojené státy americké. Tabulka 4 poté vykazuje statistiku agregovanou napříč městy ve Spojených státech. Tato statistika vychází z oficiálních dat vydaných FBI [4]. A nakonec, tabulka 5 vykazuje statistiku města Austin, tedy data, jež jsou předmětem této analýzy a jež vychází z dat vydaných městem Austin [1].

V porovnání s celoamerickou statistikou napříč všemi padesáti státy Austin v bezpečnosti očividně pokulhává. Austin má nižší počet napadení a vražd/zabití, v ostatních ohledech ale pokulhává pozadu za průměrem. Tato statistika je určitým způsobem nepřekvapivá – celoamerický průměr obsahuje také odlehle oblasti, menší vesnice, apod., tedy místa, kde je zločinnost relativně nízká. V případě Austinu se naopak jedná o relativně velké město.

Pokud porovnáme město Austin pouze s ostatními americkými městy, statistiky se více srovnají. Austin se vyznačuje nižším počtem napadení, vražd, přepadení a krádeží vozidel. Počet vloupání je naopak o osminu vyšší než průměr amerických měst, počet krádeží je o třetinu vyšší než průměr. Alarmující je poté počet znásilnění, který se rovná skoro dvojnásobku průměru všech měst. Ve prospěch Austinu vypovídá převážně fakt, že počet vražd je oproti průměru o 41 % nižší, počet přepadení je poté nižší o 19 %. Nakonec tak nemůžeme město Austin prohlásit za nadprůměrně bezpečné. Vyložene nebezpečné také není, ale relativně vysoký počet znásilnění je alarmující.

Tabulka 4: Statistika amerických měst

Zločin	Celkový počet	Počet na 100 tisíc obyv.
Znásilnění	96818	48.0
Vraždy a zabití	11890	5.8
Přepadení	236682	114.9
Napadení	585017	283.9
Vloupání	834355	404.9
Krádeže vozidla	564989	274.2
Krádeže majetku	3916991	1900.8

Tabulka 5: Statistika města Austin, TX

Zločin	Celkový počet	Počet na 100 tisíc obyv.
Znásilnění	804	86.8
Vraždy a zabití	32	3.45
Přepadení	1022	110.37
Napadení	2128	229.8
Vloupání	4171	450.4
Krádeže vozidla	2427	262.1
Krádeže majetku	26572	2869.5

Reference

- [1] City of Austin *2018 Annual Crime* [online, cit. 2021-12-28] Dostupné z: <https://data.austintexas.gov/Public-Safety/2018-Annual-Crime/pgvh-cpyq> City of Austin, 6. leden 2020
- [2] City of Austin *Council District Map* [online, cit. 2021-12-29] Dostupné z: <https://www.austintexas.gov/GIS/CouncilDistrictMap/> City of Austin
- [3] United States Census Bureau *U.S. Census Bureau QuickFacts: Austin city, Texas* [online, cit. 2021-12-30] Dostupné z: <https://www.census.gov/quickfacts/fact/table/austincitytexas/LND110210> United States Census Bureau
- [4] Federal Bureau of Investigation *Council District Map* [online, cit. 2021-12-30] Dostupné z: <https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/tables/table-16> Federal Bureau of Investigation