

# 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 následující tabulce.

## 2 Výsledky analýzy

### 2.1 Počty zločinů

Počty jednotlivých zločinů 1 jistě nijak zarážející nejsou. Naprosto převažují krádeže, v tomto případě se z velké části pravděpodobně jedná hlavně o kapsářství a podobné drobné krádeže. Počet vloupání do domu je oproti obyčejným krádežím méně než pětina. Násilné činy tak časté nejsou, přesto se jich několik objevuje.

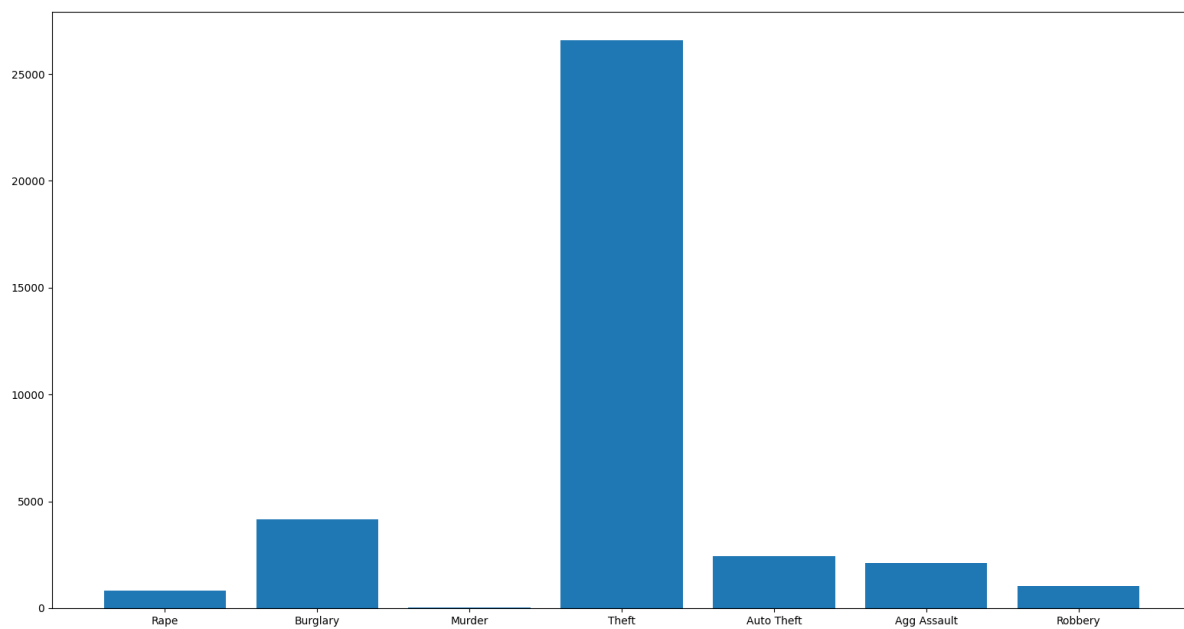
Velmi znepokojující jsou počty nahlášeného násilí na dětech 2. Austin je velké město, a proto se s vysokou pravděpodobností objeví zločiny všech možných typů, přesto řádově stovky případů zneužívání dětí je alarmující číslo.

Počet nahlášených zločinů podle kalendářního měsíce 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 nedošlo k nárůstu zločinnosti v letních měsících.

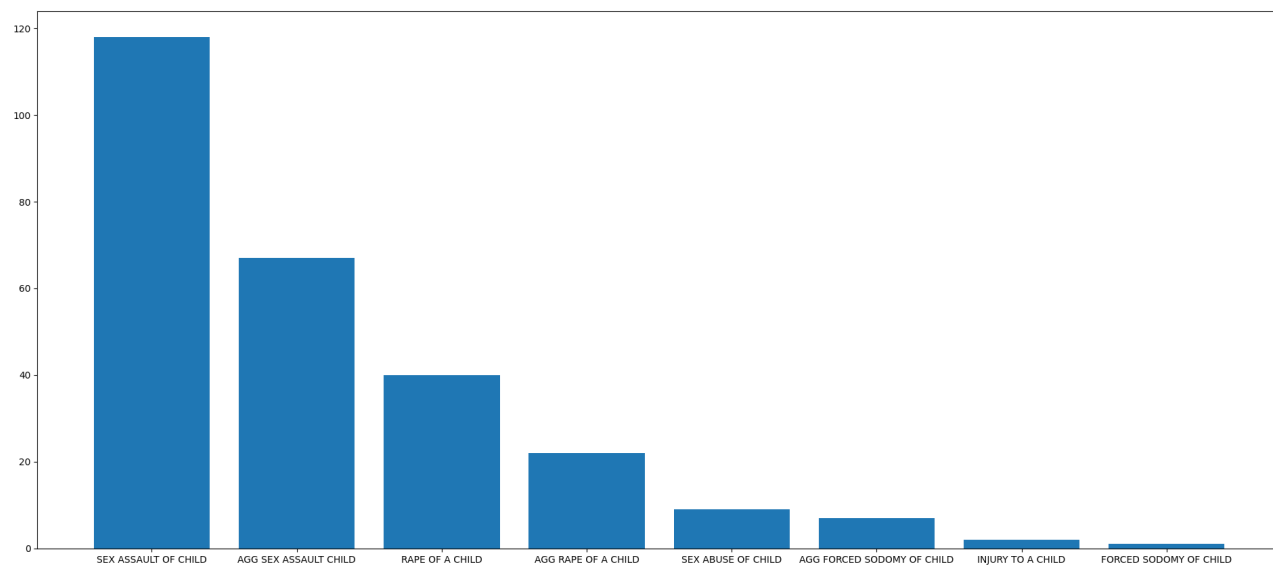
Drobný nárůst zločinnosti v říjnu může vzbudit určitou zvědavost, při podrobnějším prozkoumání ovšem zjistíme, že se neobjevily žádné zajímavé zločiny (např. hromadné nahlášení daňových úniků), jednalo se převážně o drobné krádeže. Přesto tato skutečnost může být zajímavá pro policejní oddělení města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

Tabulka 1: Popis atributů datové sady

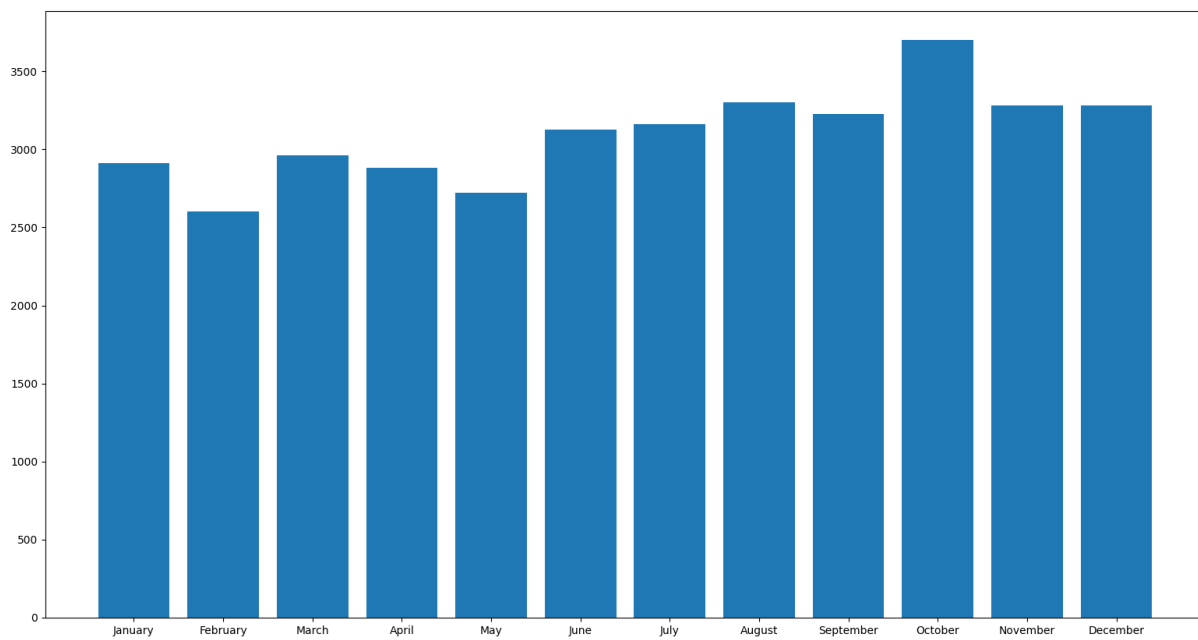
Název	Popis	Typ	Datový typ	Povinný
GO Primary Key	ID incidentu	numerický	celé číslo	ano
Council District	městská čtvrť	kategoriální	celé číslo	ne
GO Highest Offense	zločin	kategoriální	řetězec	ano
Crime Type	typ zločinu	kategoriální	řetězec	ano
GO Report Date	den nahlášení	numerický	datum	ano
GO Location	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



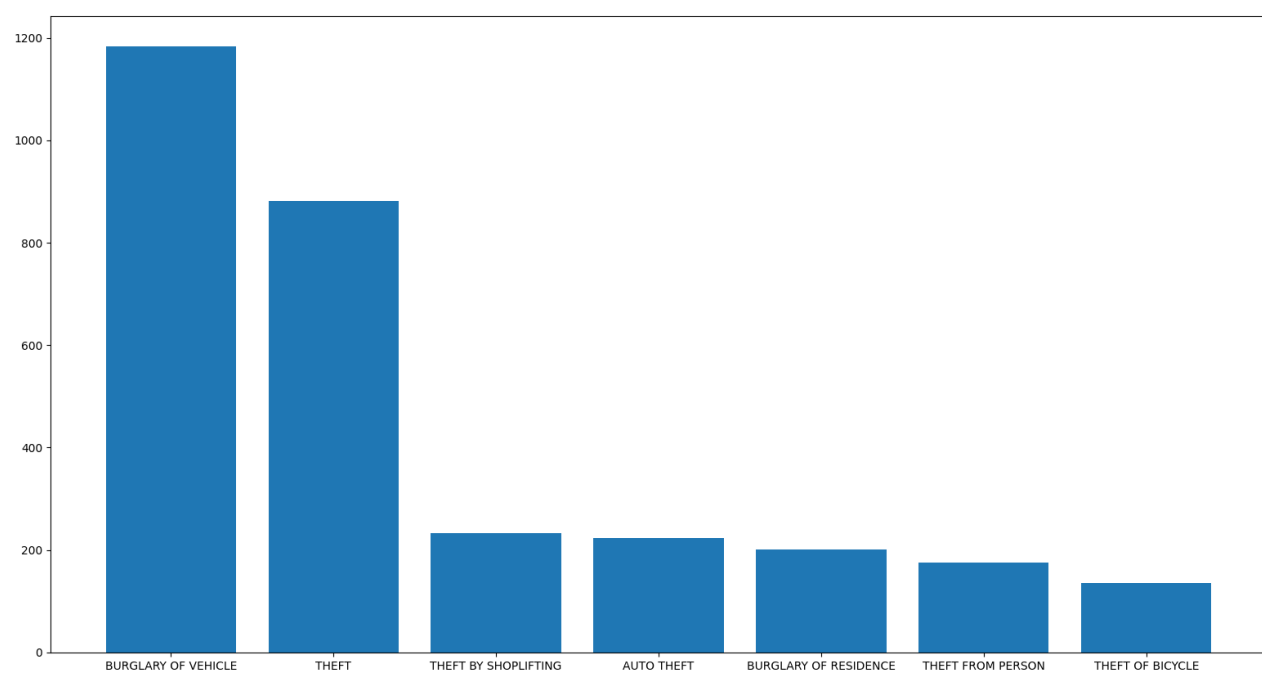
Obrázek 1: Počet nahlášených zločinů podle kategorie



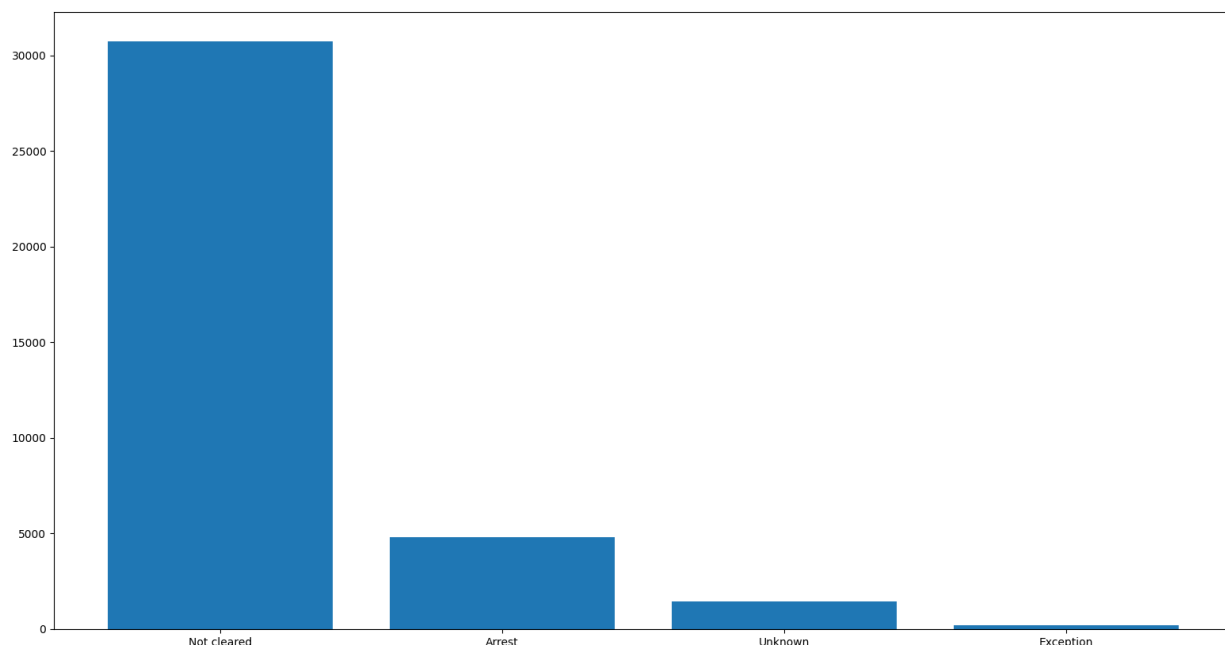
Obrázek 2: Počet nahlášených zločinů proti dětem



Obrázek 3: Počet nahlášených zločinů podle kalendářního měsíce



Obrázek 4: Sedm nejčastějších zločinů v říjnu 2018



Obrázek 5: Četnosti statusů policejního pátrání

## 2.2 Řešení zločinů

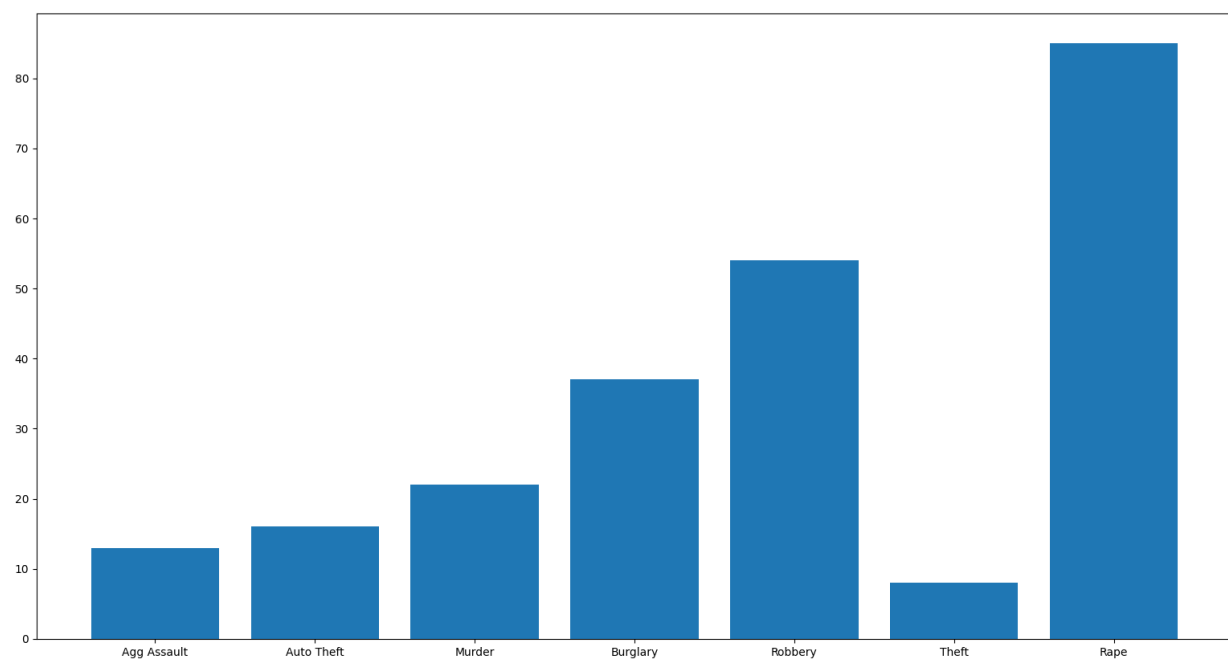
Vzorného občana určitě znepokojí množství nevyřešených případů. Z grafu 5 lze vyčíst, že naprostá většina případů zůstává nevyřešena.

Z grafu průměrné délky řešení případu 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 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 případů daleko za průměrem úspěšného řešení konkrétního typu zločinu.

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

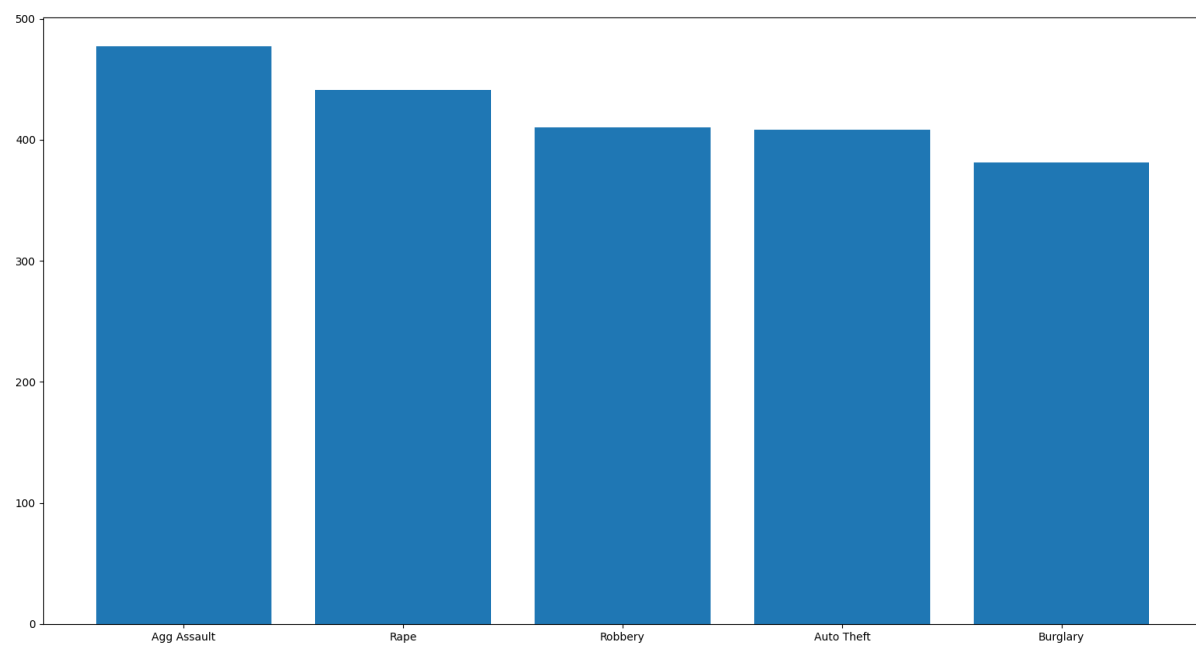
Počty uzavřených případů podle kalendářního měsíce 8 převážně kopírují počty nahlášených případů. Nejvíce případů bylo uzavřeno v říjnu a listopadu, což pouze reflektuje fakt, že bylo v říjnu nahlášeno nejvíce případů – z velké části se jednalo převážně o drobné krádeže, které jsou zpravidla uzavírány



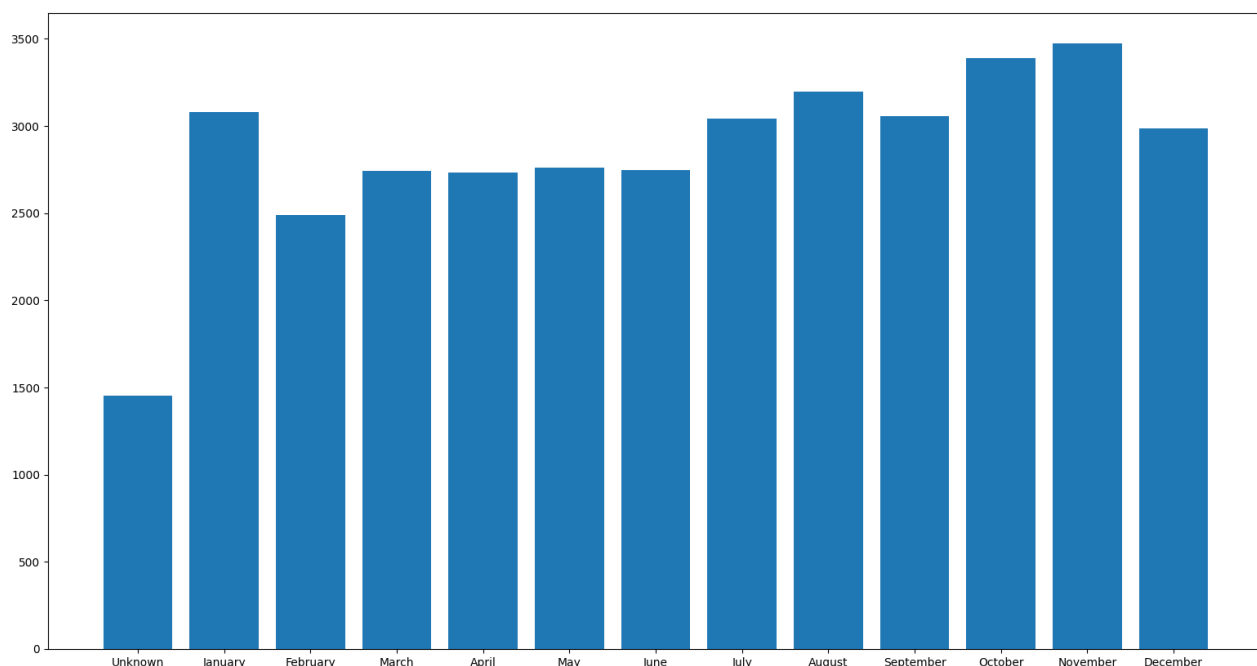
Obrázek 6: Průměrná doba úspěšného řešení

Tabulka 2: Délka řešení případu

Průměr	16.0
Medián	8.0
Variance	1172



Obrázek 7: Nejdéle řešené jednotlivé případy



Obrázek 8: Počet uzavřených případů podle kalendářního měsíce

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

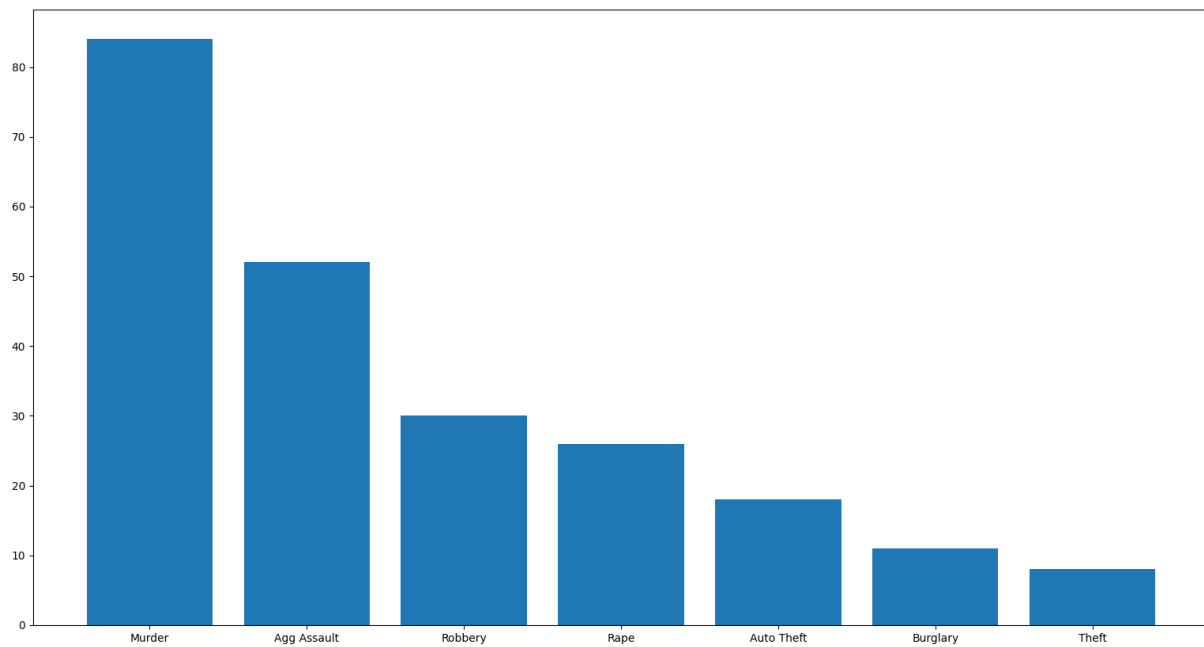
Děsivě může působit poměr vyřešených případů v poměru k celkovému počtu případů 9. Méně než pětina vloupání, přibližně pětina krádeží aut, a přibližně čtvrtina znásilnění skončila úspěšným vyřešením případu (zatčením pachatele nebo výjimečným řešením). Statistika násilných napadení je poněkud lepší,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

## 2.3 Bezpečnost městských částí

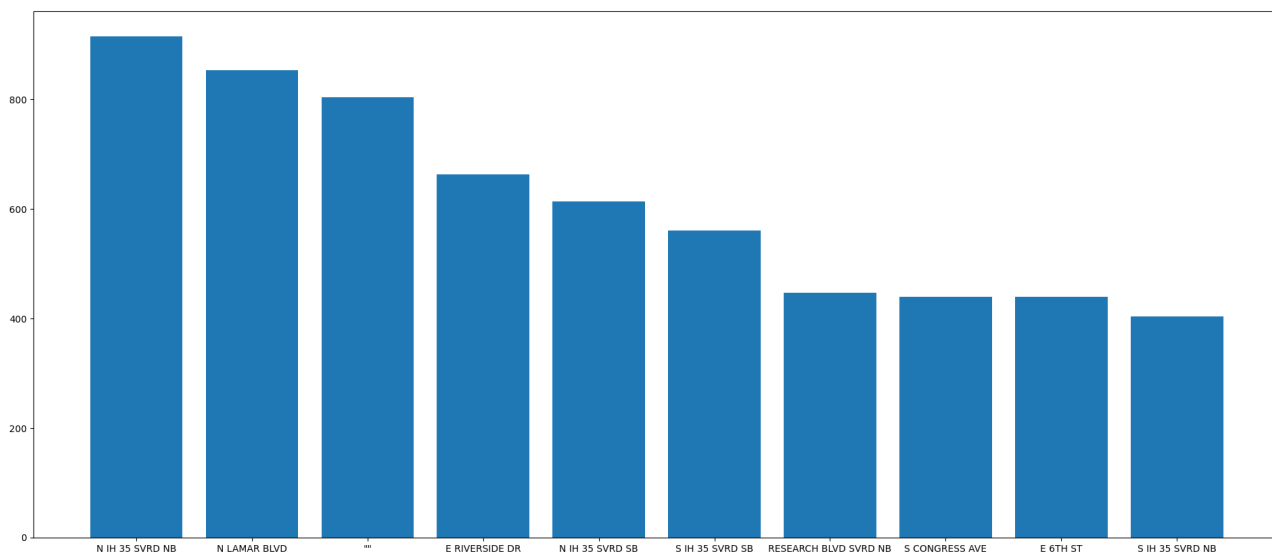
Případného návštěvníka města Austin může zajímat statistika nejméně bezpečných ulic 10. Při bližším prozkoumání však zjistíme, že ty nejméně bezpečné ulice jsou zároveň jedny z nejdelších ulic v Austinu, a jedná se o ulice, které se většinou táhnou přes polovinu města.

Z charakteristiky datové sady vyplývá, že datová sada nebude obsahovat 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 11. Proto statistika nejbezpečnějších ulic není příliš vy-

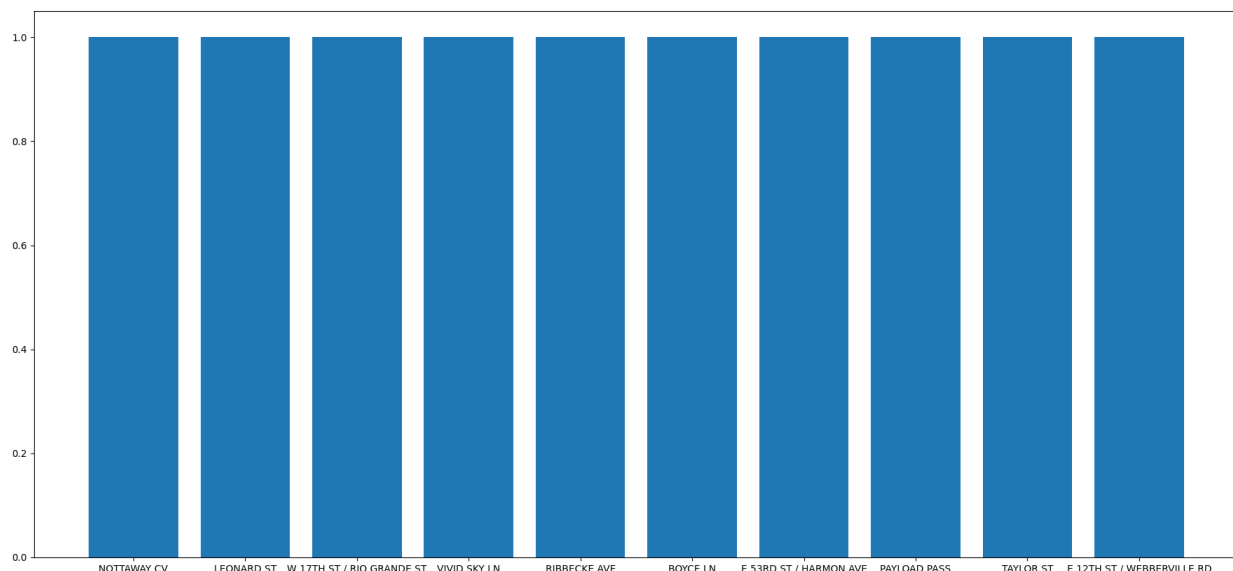




Obrázek 9: Poměr úspěšně vyřešených případů (v procentech)



Obrázek 10: Ulice s nejvyšším počtem nahlášených zločinů



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

povídající.

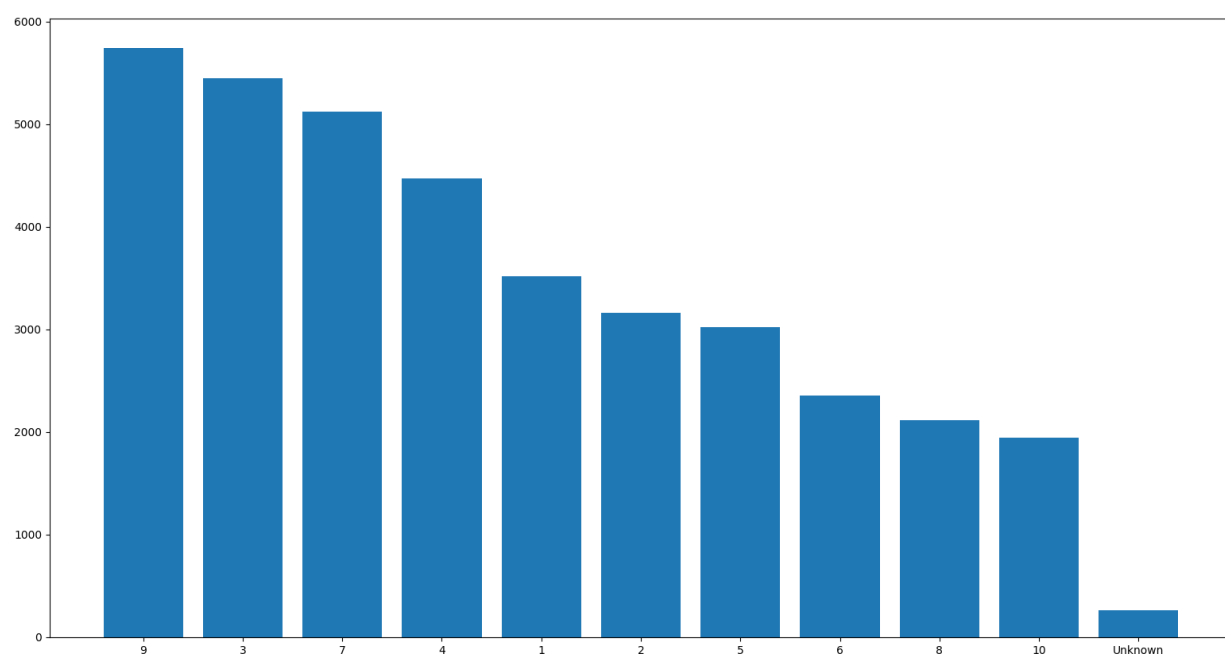
Zákony dodržujícího občana bude zajímat také seřazení čtvrtí podle bezpečnosti. Nejmeně bezpečnou čtvrtí je čtvrt' číslo devět. Čtvrť číslo devět je ovšem také centrum města Austin. Nejbezpečnějšími čtvrtěmi jsou poté čtvrti šest, osm a deset. V případě těchto čtvrtí se zase jedná o odlehlejší a klidnější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

## 2.4 Slaughter lane

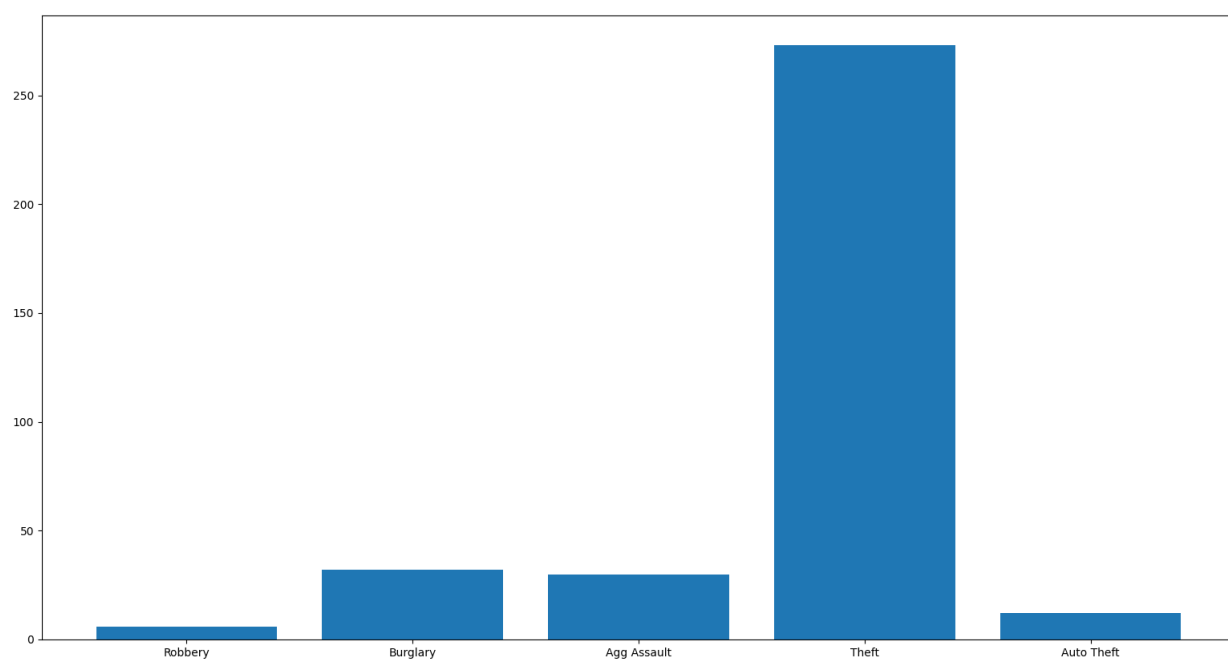
Určitou kuriozitou je existence ulice s názvem *Slaughter Lane*. Opět se jedná o relativně dlouhou ulici, proto nikoho nemůže překvapit vyšší počet nahlášených zločinů. Kuriózní je ovšem fakt, že na východní ani západní *Slaughter Lane* nebyla nahlášena ani jedna vražda nebo zabití. Lze tedy říct, že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

## 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



Obrázek 12: Čtvrťi seřazené podle počtu nahlášených zločinů



Obrázek 13: Četnosti zločinů na ulici Slaughter lane