Тијана Лазић (ИН 3/2021)

Анализа саобраћајних несрећа са повредама у Сан Франциску: Увиди кроз експлораторну анализу података (EDA)  
(СФ\_Саобраћајне\_Несреће)

Извештај за практично истраживање

# Увод

## Предмет истраживања

Предмет истраживања је детаљна анализа скупа података о саобраћајним несрећама са повредама у Сан Франциску, с фокусом на експлоративну анализу података (EDA). Ова анализа ће обухватити различите аспекте несрећа, укључујући временске трендове, анализу локација, учеснике у несрећама, као и услове у којима су се несреће догодиле.

## Циљеви истраживања

Циљ истраживања је разумевање основних карактеристика података о саобраћајним несрећама, идентификација узорака и трендова у овим несрећама, као и откривање потенцијалних фактора који доприносе повредама. Такође, намера је да се кроз анализу пруже препоруке за унапређење безбедности у саобраћају у Сан Франциску.

## Задаци истраживања

Задаци истраживања укључују:

* Сређивање података ради обезбеђивања тачности.
* Визуелизацију и анализа података кроз различите графике ради лакшег разумевања:
* временска анализа
* анализа локација
* анализа учесника
* анализа фактора несрећа
* анализа последица несрећа
* Извођење закључака и давање препорука за побољшање безбедности у саобраћају.

## Очекивани резултати истраживања

Очекивани резултати истраживања су идентификација кључних фактора повезаних са саобраћајним несрећама, као и разумевање њихових међусобних веза кроз EDA приступ. Ови резултати ће пружити основе за препоруке и интервенције које могу допринети побољшању безбедности у саобраћају у Сан Франциску.

# Методологија

## Коришћени подаци

Скуп података о саобраћајним несрећама са повредама у Сан Франциску садржи све несреће које су резултирале повредама.

Податке о несрећама са смртним исходом прикупља Канцеларија главног медицинског испитивања (OME), док се подаци о несрећама са повредама добијају из више извора укључујући привремени систем за сударе (SFPD’s Interim Collision System) за период од 2018. до данас, базу података о саобраћајним сударима (Crossroads Software Traffic Collision Database) за период од 2013. до 2017. године и државни интегрисани систем евиденције саобраћаја (SWITRS) којим управља Калифорнијска патрола за све године пре 2013.

Сви подаци који садрже валидне географске информације су мапирани на поједностављени модел саобраћајне мреже Сан Франциска који одржава Служба за јавне радове (SFDPW). Подаци се ажурирају на кварталном нивоу. Скуп података садржи преко 59,000 записа са 55 различитих атрибута, што га чини веома богатим и детаљним извором информација за анализу.

## Претходна истраживања других особа над коришћеним подацима

Претходна испитиванја на која сам наишла су углавном анализе од стране државних органа у Сан Франциску. Истраживања која су рађена обично су се фокусирала на следеће аспекте:

* Саобраћајне несреће са смртним исходом или повредама
* Ударци и бекство са смртним исходом, повредама или оштећењем имовине
* Несреће са возилима без возача
* Несреће са градским возилима или оштећењем градске имовине
* Несреће са школским аутобусима
* Саобраћајне несреће које укључују хапшења
* Возачи под утицајем алкохола или дрога
* Несреће са бициклистима и пешацима
* Несреће са аутономним возилима

## Методе истраживања

Методе истраживања коришћене у овој анализи укључују:

* **Сређивање и припрема података**: Обрада и чишћење података ради уклањања недостајућих вредности и обезбеђивања тачности. Рачунато је колико има колона са недостајућим вредностима, проценат редова који садрже недостајуће вредности, као и филтрирање редова са негативним вредностима у колонама 'number\_killed' i 'number\_injured'. Такође је и дефинисана функција за парсирање датума. У 49.83% редова је било недостајућих вредности, тако да је након уклањања тих редова за анализу преостало 29403 редова, што и даље богат скуп података.
* **Експлоративна анализа података (EDA)**: Коришћење различитих техника за анализу података како би се идентификовале основне карактеристике и обрасци у подацима. Овај корак обухвата:
  + **Статистичке анализе**: Израчунавање основних статистичких мера као што су средња вредност, медијана и стандардна девијација.
  + **Анализа учесника**: Истраживање типова учесника, као и превозног средства са којим се несрећа догодила.
  + **Временска анализа**: Анализа трендова у несрећама током времена, укључујући годишњу, месечну и седмичну анализу.
  + **Анализа локација**: Ово укључује приказивање локација несрећа на мапи Сан Франциска (није рађена мапа, али су коришћене географска ширина и дужина претворене у број) ради идентификације места несреће са смртним исходом и оног где је само дошло до повреда, као и проценат повређених у односу на близину раскрснице. Такође је анализиран број учесника по полицијским дистриктима.
  + **Анализа фактора несрећа**: Истраживање различитих фактора који доприносе саобраћајним несрећама. Ово укључује анализу различитих временских услова, осветљење, услове на путу, али и стања пута.
  + **Анализа последица несрећа**: Ово укључује анализу типова несрећа, озбиљности повреда и врста повреда.
* **Визуализација података**: Креирање графика као што су хистограми (кутијасти графикони), линијски, стубичасти и тортни. Такође је коришћен и географски приказ.
* **Закључци и препоруке**: На основу добијених резултата, формулисање препорука за побољшање саобраћајне безбедности у Сан Франциску. Ово укључује предлоге за инфраструктурне измене, мере безбедности и едукацију јавности.

# Резултати

## Приказ резултата

**Статистичке анализе**

Груписање података по типу раскрснице и сумирање броја погинулих и повређених:

intersection number\_killed number\_injured

Intersection 339.0 49347

Midblock 221.0 19083

Intersection Rear End 10.0 5671

Undefined 0.0 1

Израчунавање статистичких параметара за колону 'number\_ injured':

mean 1.255775

median 1.000000

std 0.702304

Израчунавање статистичких параметара за колону 'number\_killed':

mean 0.009660

median 0.000000

std 0.100881

Бројање јединствених вредности у колони 'road\_surface':

Dry 51786

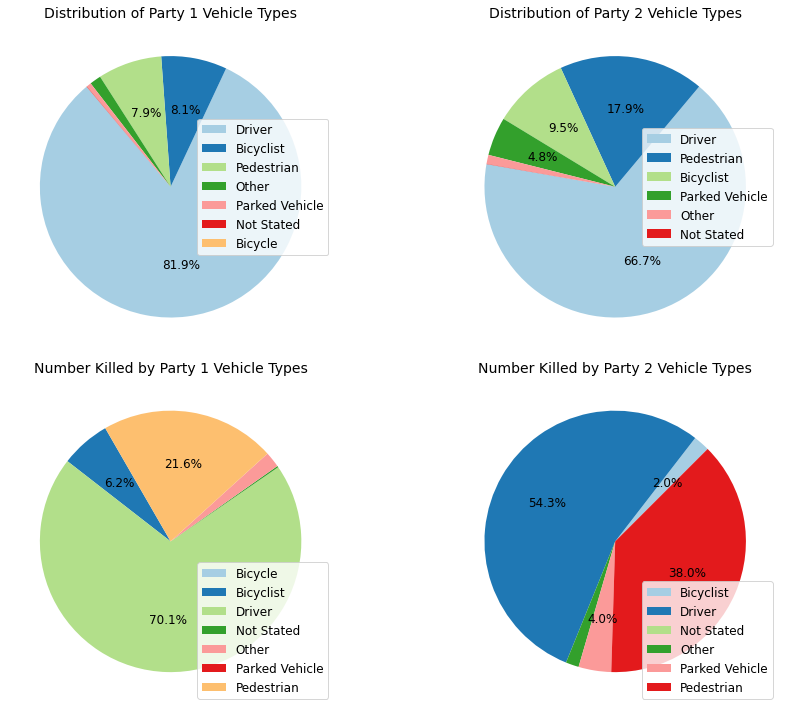
Wet 6063

Not Stated 1045

Slippery 81

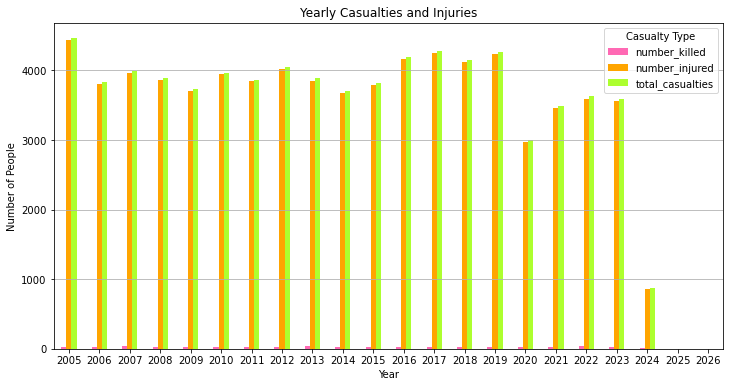
Snowy or Icy 34

**Анализа учесника**

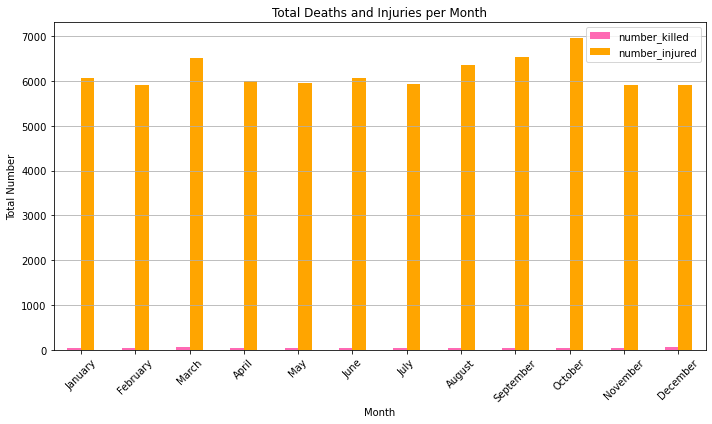
****

Слике 1. Проценат типова возила у саобраћајним несрећама и проценат преминулих од стране њих

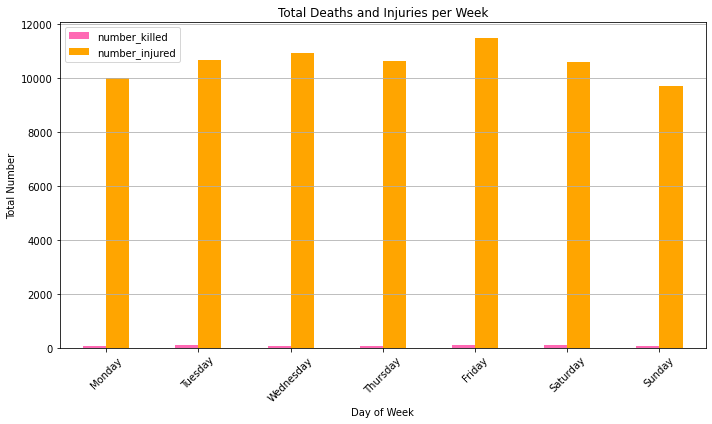
**Временска анализа**



Слике 2. Годишнји преглед учесника у саобраћајним несрећама



Слике 3. Месечни преглед учесника у саобраћајним несрећама



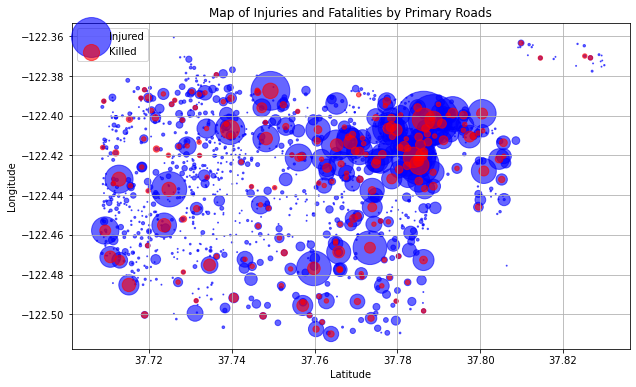
Слике 4. Седмични преглед учесника у саобраћајним несрећама

**Анализа локација**

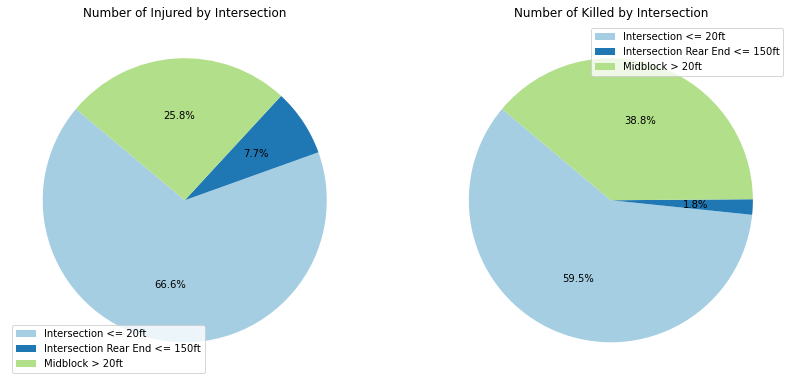
* Географска ширина: приближно 37.7749° С
* Географска дужина: приближно 122.4194° З



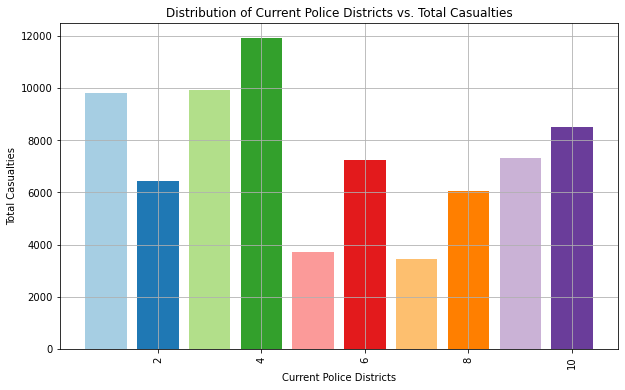
Слика 5. Мапа града Сан Франциска



Слика 6. Графикон распршености броја повређених и броја погинулих у односу на географску дужину и ширину

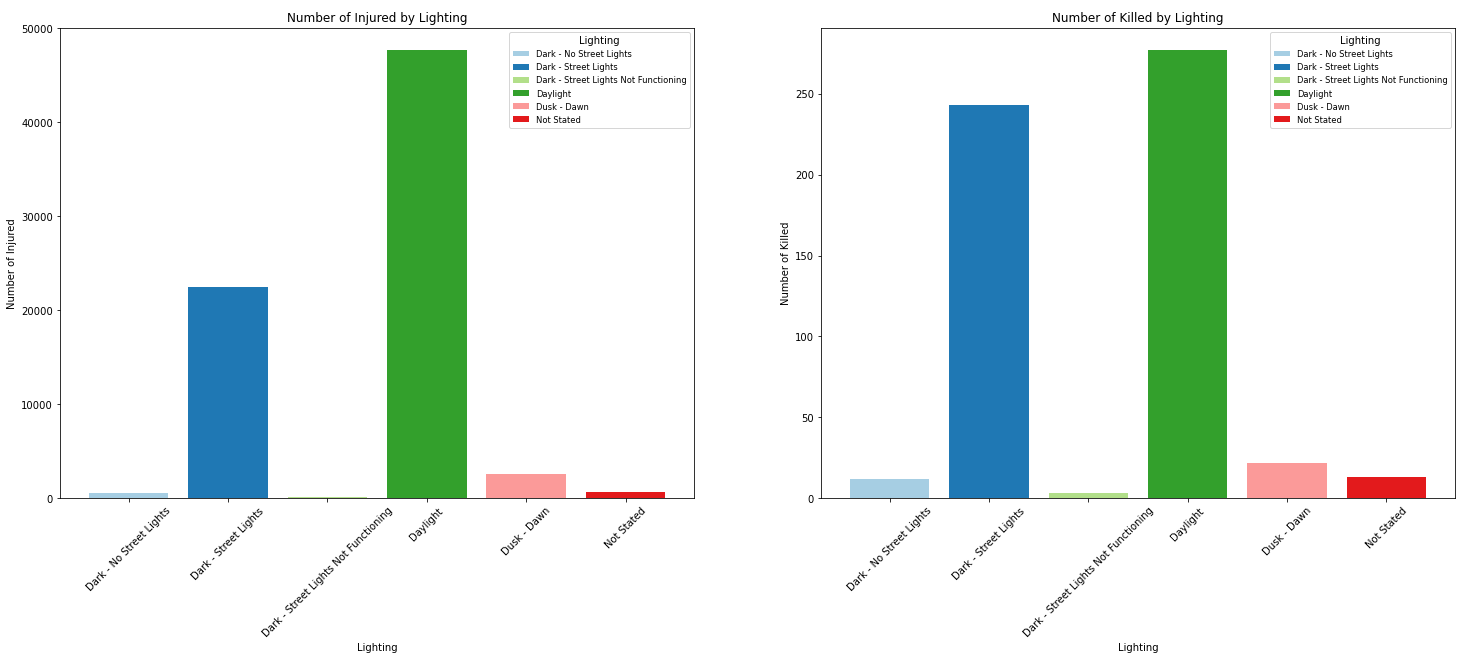


Слика 7. Проценат повређених и погинулих у близини раскрсница

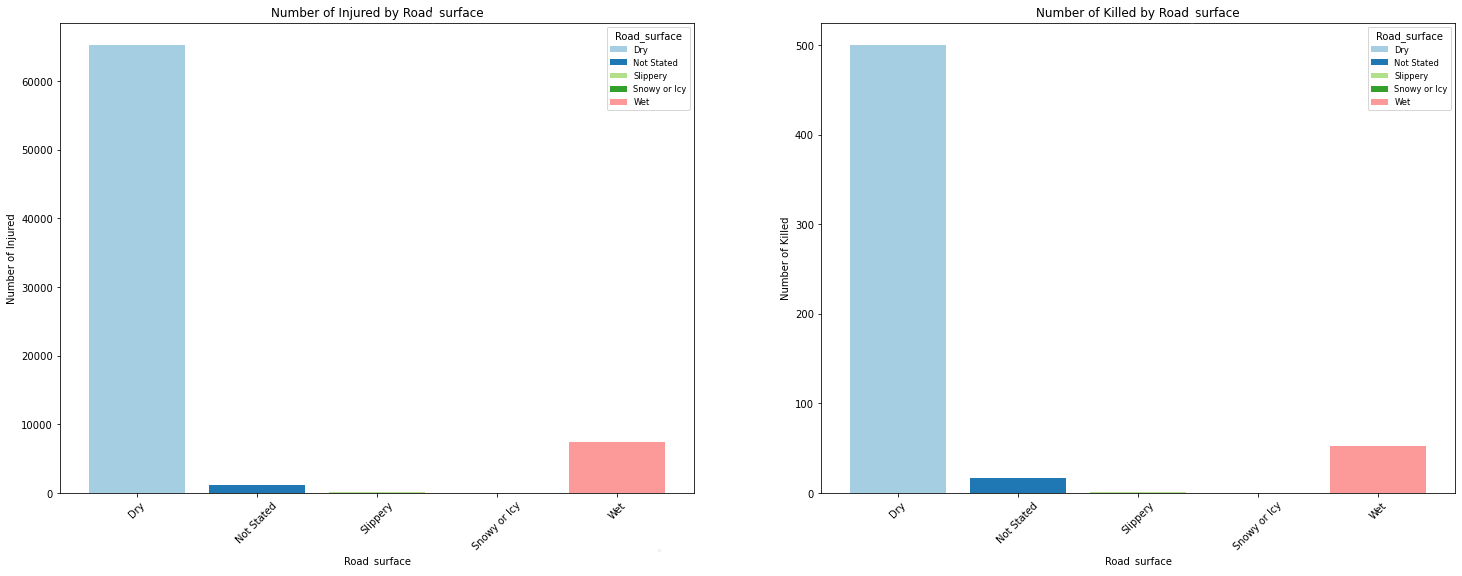


Слике 8. Укупан број учесника по полицијским дистриктима

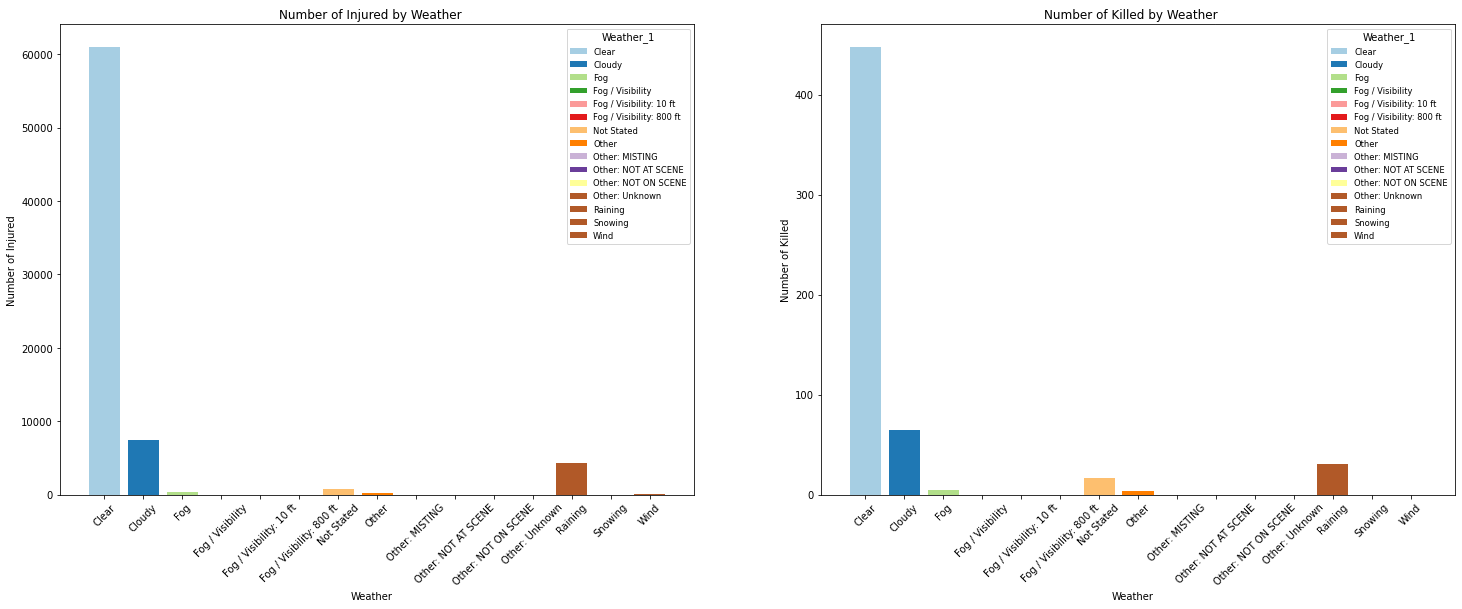
**Анализа фактора несрећа**

****

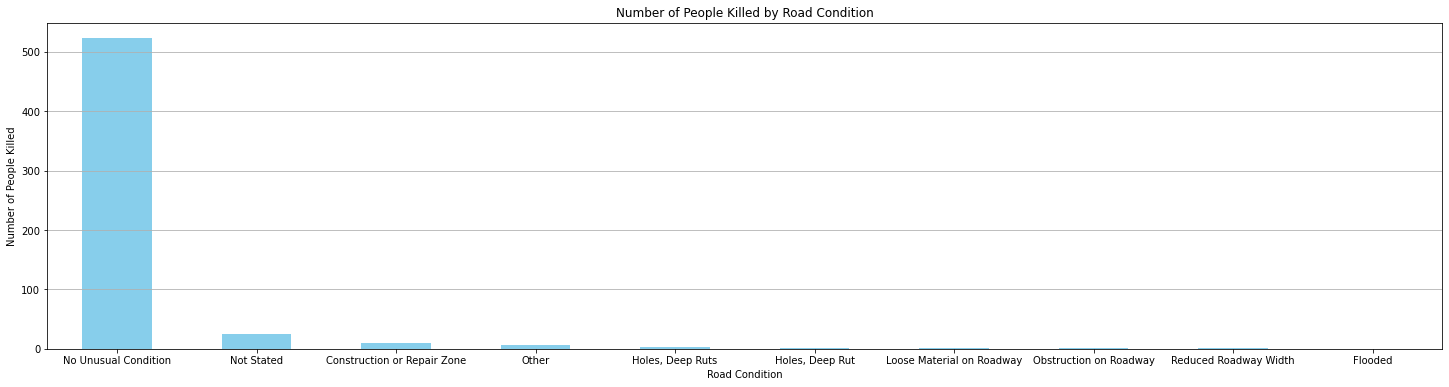
Слике 9. Број повређених и погинулих у односу на осветљење

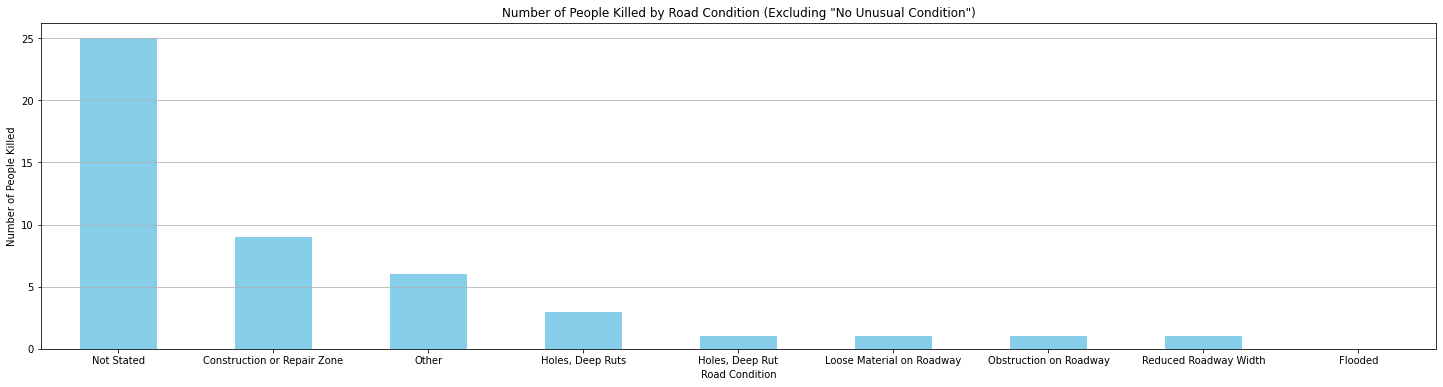


Слике 10. Број повређених и погинулих у односу на површину пута



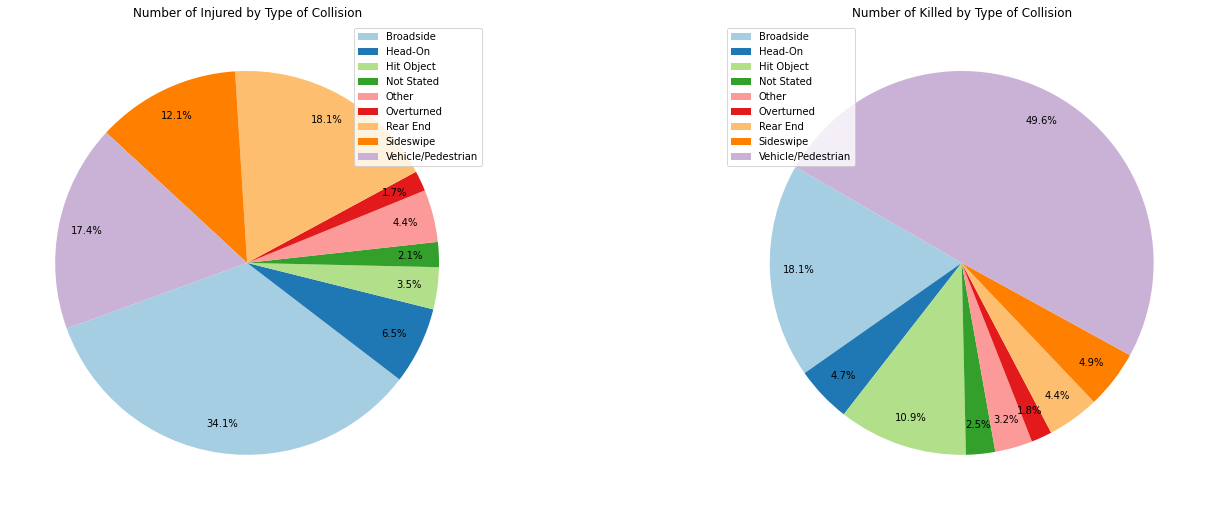
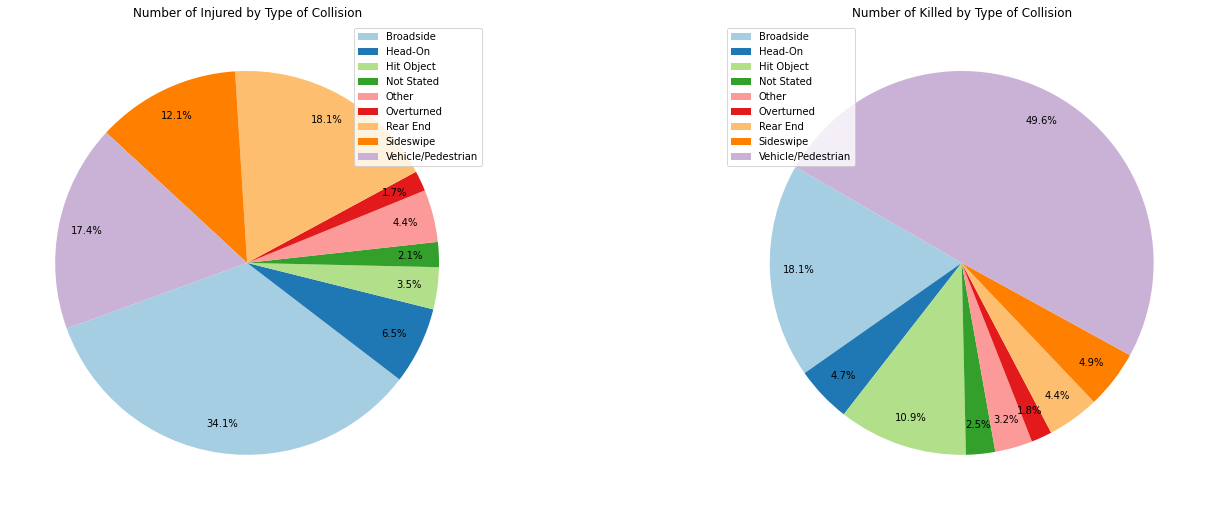
Слике 11. Број повређених и погинулих у односу на временске услове у тренутку несреће



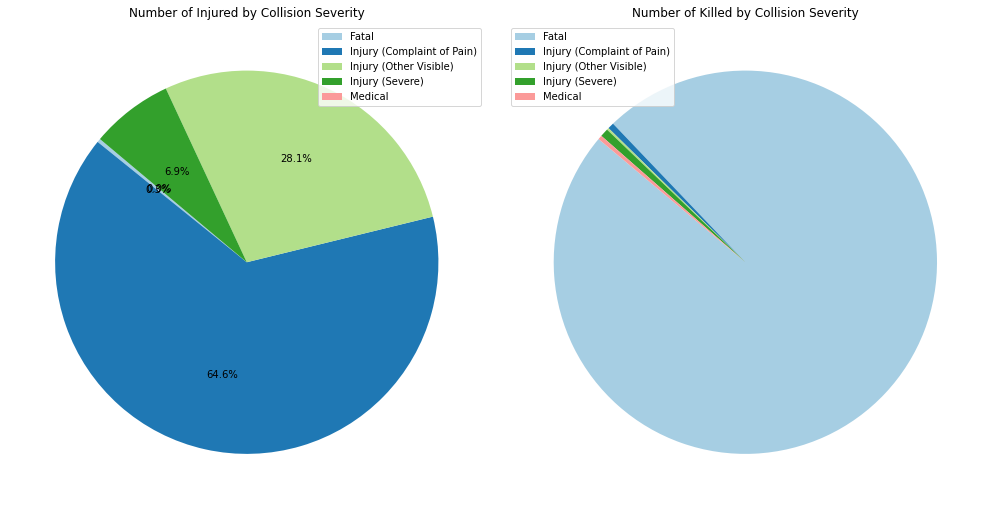


Слике 12 и 13. Број повређених у односу на стање пута

**Анализа последица несрећа**



Слика 14. Проценат повређених и погинулих према типу несреће



Слика 15. Проценат повређених и погинулих по новоу повреде у несрећи

## Тумачење резултата

**Анализа учесника**

Са прва два тортна графикона на слици 1 можемо да закључимо да су најчешћи учесници у саобраћајним несрећама возачи, пешаци и бициклисти, а потом паркирана возила. Такође, са друга два торна графикона можемо да видимо да су и лица која су погинула у саобраћајној несрећи у највећој мери возачи, потом пешаци, а за нјим бициклисти.

**Временска анализа**

Код годишне прегледа (слика 2) можемо да видимо да је до 2020. године углавном био сличан број учесника у саобраћајним несрећама. Код 2020. године видимо мањи пад, па можемо да претпоставити да је то због вируса COVID-19, који је задесио цео свет. Углвном је у свим државама било ограничено кретање, а са сигурношћу то знамо за Сан Франциско. Слабије кретање становништва је довело до мањих саобраћајних несрећа.

Код месечног прегледа (слика 3) видимо да је месец са највише несрећа септембар и октобар, али и март. Септембар и октобар су месеци када се ученици и студенти враћају назад у школе и на универзитете, али се и запослени враћају са годишњих одмора, па је промет у граду већи. Такође велики утицај има и присуство магле у овим месецима у току године, што може смањити видљивост и самим тим изазвати саобраћајне несреће.

Код седмичног прегледа (слика 4) благо се истиче петак као дан са највише саобраћајних несрећа. Петком се углавном људи са стране враћају и одлазе у град, али и туристи долазе туристички у Сан Франциско, па можемо претпоставити да се због тога петак истиче у седмичном прегледу.

**Анализа локација**

Код анализе локација са слике 7 можемо да видимо да је највећи број саобраћајних несрећа у раскрсницама. Многи возачи не поштују правила, између осталог и стоп када је црвено светло. Потребна је едукација и стално подстицање грађана широм света да је јако важно поштовање саобраћајних правила, јер непоштовање угрожава све. Делови Сан Франциска са највећом стопом саобраћајних несрећа су Union Square и Civic Center.

**Анализа фактора несрећа**

Што се тиче анализе броја повређених и погинулих у односу на осветљење, а графике (слиак 9) можемо да видимо да до несреће долази током дана, на дневној светлости, али и ноћу или у условима слабог осветљења, али са укљученим уличним светлом. Постоје и случајеви када је мрак и када нема уличног осветљења. Тада је број саобраћајних несрећа доста мањи, па из тог можемо претпоставити да је опрезност возача тада много већа него када је дан.

Код анализе саобраћајних несрећа где сагледавамо аспекте површине пута на којој је дошло до несреће (слика 10) можемо видети да је коловоз сув у великом случају несрећа. Ту је присутан и мокар коловоз, али са доста мање несрећа. Претпоставка би била или велика опрезност учесника у саобраћају или мање Кишних дана у Сан Франциску.

Анализа саобраћајних несрећа према временским условима у тренутку несреће (слика 11) је слична и за ситуације у којима је дошло до повред и у ситуацијама са смртним исходом. Са графикона можемо врло јасно закључити да је током ведрог времена највећи број саобраћајних несрећа. Такође су присутни случајеви где је време облачно, магловито и кишовито.

Што се тиче анализе стања пута (слика 12 и 13) можемо видети да нема неких необичних услова пута због којих долази до саобраћајних несрећа. Претпоставка је да се Сан Франциско због своје и прометности и популарности одржава, бар у овом случају коловоз.

**Анализа последица несрећа**

Код анализе повређених према типу несреће (слика 14, први тортни графикон) видимо да је највише повређених у сударима где предњи део једног возила удара у бочну страну другог возила. Ово се подудара са нашом тврдњом да возачи не поштују црвено светло на семафору. Након тога судар у којем возило удара у заднји део другог возила испред њега, несрећа у којој возило удара пешака или неко друго возило, судар у којем се бочни делови два возила додирују или ударају док се крећу у истом или супротном смеру и судар у којем се преднји делови два возила сударају. У мањој мери су присутни ударанје неког предмета и превртање возила.

Код анализе погинулих према типу несреће (слика 14, други тортни графикон) видимо да је највише усмрћених лица у сударима где возило удара пешака (или неко друго возило), чак скоро половина. Потом судари где предњи део једног возила удара у бочну страну другог возила и ударанје у непокретни објекат. У јако сличном проценту присутни су судари где се предњи делови два возила сударају и судари у којем се бочни делови два возила додирују или ударају док се крећу у истом или супротном смеру.

Код анализе повређених према нивоу повреда (слика 15, први тортни графикон) видимо да је највише присутна повреда где учесници у саобраћајној несрећи пријављују бол, али нема видљивих повреда. Овај тип повреде је најпожељнији ако је већ морало да дође до повреде, али свакако да је најбоље ако није дошло до повреда. Потом следе повреде које су видљиве, али не толико озбиљне да би захтевале продужену хоспитализацију. Тешке повреде које захтевају хоспитализацију су присутне у 6,9%.

# Закључак

## Анализа испуњења циљева истраживања

Циљеви истраживања су били разумевање основних карактеристика података о саобраћајним несрећама, идентификација узорака и трендова, као и откривање фактора који утичу на њих. Истраживање је испунило ове циљеве кроз детаљну експлораторну анализу података (EDA). Анализирани су временски трендови, просторна дистрибуција несрећа, фактори и последице несрећа, типови учесника(нпр. возачи, бициклисти), као и различити типови локација (нпр. раскрснице, средњи део улице). Визуелизације су омогућиле адекватан приказ свих анализа.

## Анализа остварења очекиваних резултата истраживања

Очекивани резултати су укључивали идентификацију кључних фактора повезаних са саобраћајним несрећама и разумевање њихових међусобних веза. Закључно, анализа је успешно остварила своје циљеве, пружајући дубљи увид и нудећи конкретне смернице за побољшање безбедности у саобраћају у Сан Франциску.

На основу датих анализа, можемо извести неколико генералних закључака:

* **Профил учесника у саобраћајним несрећама**: Возачи, пешаци и бициклисти су најчешћи учесници у саобраћајним несрећама, а такође су и најчешће жртве таквих инцидената.
* **Ефекат COVID-19 на саобраћајне несреће**: Увод мера за сузбијање пандемије COVID-19 имао је утицај на смањење броја саобраћајних несрећа, што сугерише да је ограничавање кретања смањило инциденте.
* **Утицај временских услова**: Магла, киша и мрачни услови су повезани са већим ризиком од саобраћајних несрећа због смањене видљивости и услова пута, али су према нашим анализама возачи тада опрезнији или ових услова нема толико често.
* **Локације са највећим ризиком**: Раскрснице представљају места са највећим бројем саобраћајних несрећа, што указује на потребу за бољим поштовањем саобраћајних правила и повећаним контролама.
* **Утицај осветљења и стања пута**: Иако су саобраћајне несреће најчешће условима дана и услојима осветљења, постоје случајеви и у ноћним условима, а када је мокар коловоз учесници су опрезнији у односу на сув коловоз.
* **Последице несрећа**: Повреде у саобраћајним несрећама најчешће су повреде које нису видљиве на први поглед, али свака повреда свакако захтева преглед.
* **Неопходност едукације и поштовања саобраћајних правила**: Како би се смањио број саобраћајних несрећа и њихове последице, неопходно је континуирано образовање и подсећање на важност поштовања саобраћајних правила за све учеснике у саобраћају.

Ови закључци наглашавају потребу за мерама како би се подигла свест о безбедности у саобраћају и смањио број саобраћајних несрећа и њихових последица.

## Могућности за примену истраживања у пракси

* **Доносиоци одлука**: Могу користити резултате за планирање инфраструктурних измена и унапређења безбедности у саобраћају.
* **Јавне агенције**: Могу применити налазе за унапређење политика и програма који се односе на саобраћајну безбедност.
* **Образовне кампање**: Резултати могу бити основа за кампање које имају за циљ едукацију возача, пешака и других учесника у саобраћају.
* **Истраживачи**: Могу користити добијене резултате као полазну тачку за даља истраживања и дубље анализе.

## Идеје за побољшање и разраду истраживања

Постоје бројне могућности за побољшање и разраду овог истраживања:

* **Проширење скупа података**: Укључивање додатних година података и других извора може обезбедити потпунију анализу.
* **Напредни модели машинског учења**: Примена машинског учења и предиктивних модела може помоћи у прецизнијем предвиђању будућих несрећа и њихових узрока.
* **Детаљнија просторна анализа**: Коришћење географских информационих система (GIS) за детаљнију анализу просторно-временских узорака.
* **Анализа утицаја мера безбедности**: Истраживање ефеката конкретних мера које су имплементиране у циљу побољшања безбедности у саобраћају.

# Литература

1. Сан Франциско Градски портал података. (2023). Traffic Crashes Resulting in Injury. [*data.sfgov.org*](https://data.sfgov.org/Public-Safety/Traffic-Crashes-Resulting-in-Injury/ubvf-ztfx/about_data)
2. San Francisco Department of Public Health (SFDPH). (2023). Vision Zero Fatality Protocol. [*sfdph.org*](https://www.sfdph.org)
3. Coursera: Exploratory Data Analysis in Python.

[*Geeks for Geeks: Exploratory Data Analysis in Python.*](https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/)