**Сара Сапунџија (ИН14/2021)**

**Ранка Радуловић (ИН16/2021)**

Анализа злочина мотивисаних предрасудама у Калифорнији  
(Анализа злочина)

Извештај за практично истраживање

# Увод

## Предмет истраживања

Истраживачки пројекат на тему злочина мотивисаних предрасудама фокусира се пре свега на детаљну анализу криминалних активности у граду Сан Гранциско, Калифорнија. Да бисмо ову анализу спровели, користимо се скупом података који обухвата злочине пријављене од стране полицијског департмана Сан Франциска.

У истраживању такође користимо и скуп података који обухвата процене популације и демографске податке који су изведени из САД пописа како бисмо закључили колики проценат становништва је потенцијално угрожен.

## Циљеви истраживања

Циљ је детаљно разумевање оба скупа података, идентификација фактора који доприносе оваквим злочинима, идентификација сличности и потенцијалног шаблона међу злочинима и процена броја жртава у будућности.

Закључивање о шаблонима подразумева идентификацију најчешће коришћених орзжија, најучесталијих локација, најчешће мотивације. Овим подацима бићемо у могућности да боље разумемо профил жртве, али и починитеља злочина.

## Задаци истраживања

Најизазовнији задатак је само форматирање података и спремање истих за анализирање, агрегирање. Имплементација машинског модела за предикцију броја жртава и израда адекватног приказа за интерактивну анализу резултата у програмском језику Пајтон.

## Очекивани резултати истраживања

Очекиван резултат је прегледан приказ наших скупова података који другима омогућавају лако и брзо сналажење и закључивање. Могућност дубље анализе проблема и налажење потенцијалних решења за исти.

# Методологија

## Коришћени подаци

* <https://data.sfgov.org/Public-Safety/Police-Department-Investigated-Hate-Crimes/huqr-9p9x/about_data>

Скуп података који обухвата злочине пријављене од стране полицијског департмана Сан Франциска. Укључују пријавњени злочин, тип пристрасности, време и место злочина, број жртава и број осумњичених. Подаци не садрже географску локацију злочина већ тип локације (продавница, улица…), као ни било какве податке којима би се жртве или осумњичени могли идентификовати. Обухватају период од 2001. до 2023. године. Подаци су погодни за експлоративну анализу као и за предвиђање одређених параметара.

* <https://data.sfgov.org/Economy-and-Community/San-Francisco-Population-and-Demographic-Census-da/4qbq-hvtt/about_data>

Скуп података обухвата процене популације и демографске податке који су изведени из САД пописа. Подаци су већ агрегирани по раси и старосном добу, а садрже и податке о томе колико година су били прикупљани, као и објашњење како је процена изведена.

## Претходна истраживања других особа над коришћеним подацима

Нисмо налазили слична истраживања над скуповима података које смо ми користиле.

Иако су истраживања на ову тему многобројна, углавном се користе још обимнији скупови података, који обухватају целу површину САД-а.

## Методе истраживања

Коришћењем библиотека numpy и pandas учитавамо податке које затим обрађујемо различитим стратегијама.

Било је важно добро се упознати са скуповима података, свим њиховим колонама и вредностима како би се ефективно одлучило над којим подацима је могуће и смислено спровести даље анализе. Исписивањем колона, типова, јединствених вредности података закључујемо са каквим подацима радимо и колико су адекватни за наше анализе.

У овом моменту било је очигледно да демографски подаци нису били идеални за наше потребе, с обзиром да су у једној колони имали више потпуно различитих вредности, те је било потребно много више времена и труда очистити и агрегирати овакве податке.

Проблем су представљале и колоне о старости починитеља и жртава у скупу о злочинима где је постојала огромна количина недостајућих вредности. Ипак, ова колона је била занимљива иако није садржала велики број података, те је била задржана у скупу података за даљу анализу.

Скуп о злочинима се показао детаљнији него што смо очекивале, те је било потребно и генерализовати садржај неких његових колона, док смо неке друге класификовале на више како бисмо дошле до детањнијег приказа.

Прешле смо на предикције. Користећи алгоритме за линеарну регресију и алгоритам насумичних стабала покушале смо да превидимо трендове у нашим подацима.

Линеарна регресија је једноставан начин моделовања односа између једне зависне и више независниг променљивих.

Модел насумичних стабала је алгоритам машинског учења који користи случајан одабир стабла одлуке како би предвидео параметре.

Тренирањем ова два модела и њиховим прилагођавањем покушале смо да дођемо до адекватних резултата.

Након детањне обраде података прешле смо на приказ користећи првенствено библиотеку matplotlib. Многобројни графикони су нам били од велике помоћи за боље разумевање и уочавање трендова, али кључни за следећи корак - визуализацију.

Визуализација и израда дашборда је нешто са чиме до сада нисмо имале додира.

На самом почетку покушале смо да користимо Dash радни оквир за израду апликација са Python-ом. Нисмо биле успешне у нашој потрази за начином како да конвертујемо matplotlib графиконе у оне који су прикладни за Dash - plotly. Те проналазимо друго решење.

Тkinter је стандардни пакет за израду корисничких интерфејса у језику Python. Омогућава брзо и једноставно креирање апликација са графичким елементима и веома је прикладан за почетнике. Било је веома једноставно креирати дашборд који изгледа веома једноставно али прегледан.

# Резултати

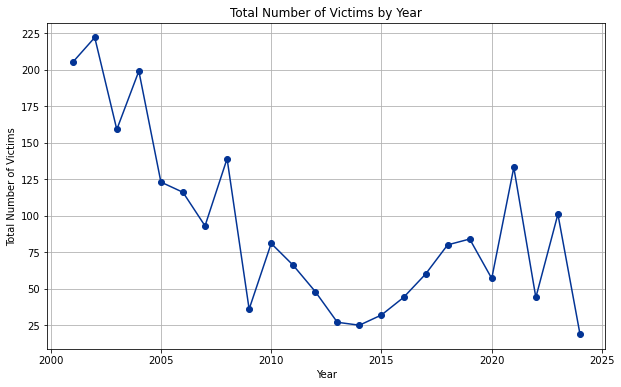
## Приказ резултата

Приказ резултата обухвата сам кориснички интерфејс односно дашборд где се кроз графиконе и табеле може стећи боље разумевање образаца и трендова, као и најугроженијих друштвених група и мета злочина мотивисаним предрасудама у Сан Франциску.

## Тумачење резултата

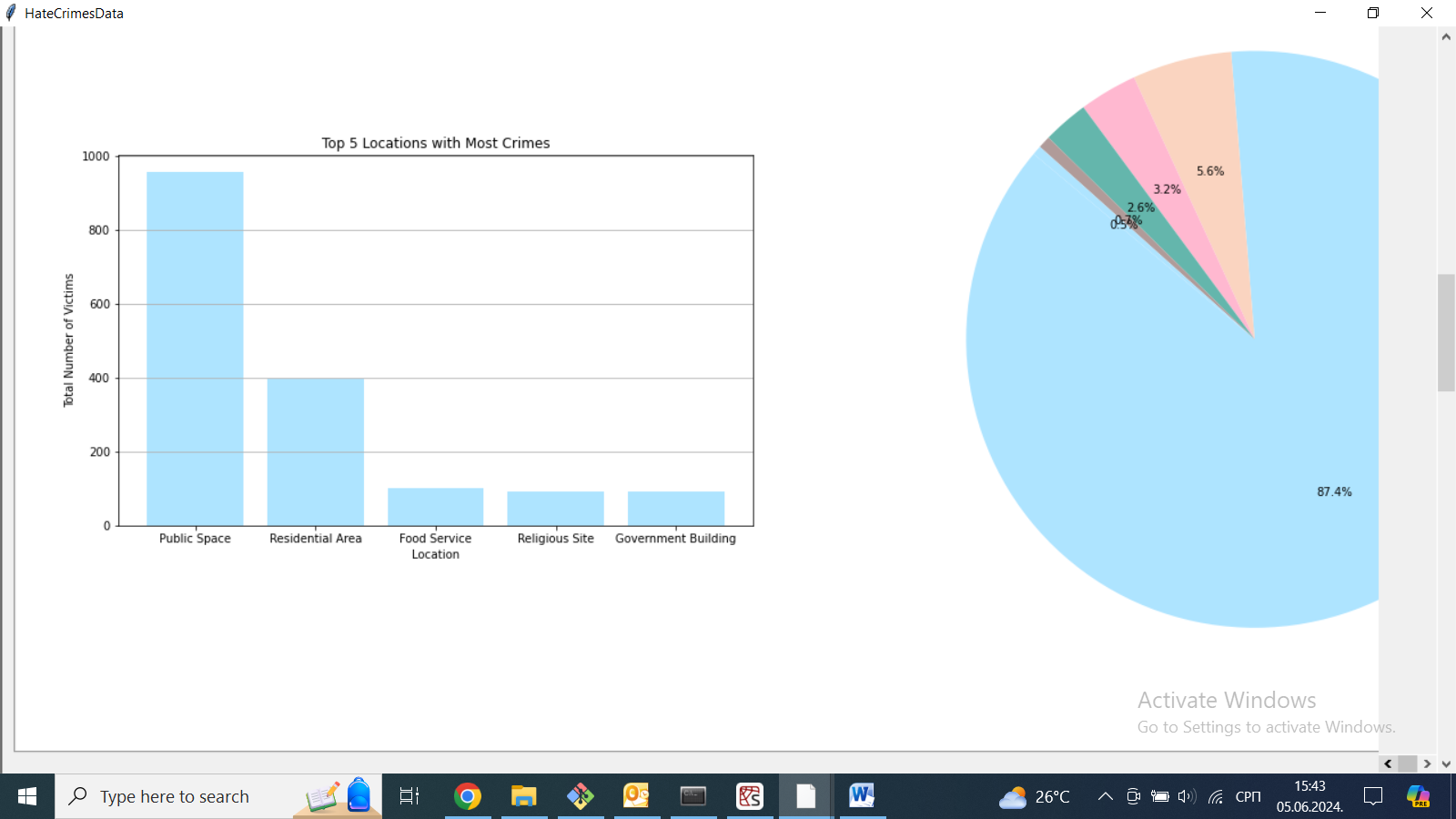
Кроз графиконе јасно можемо видети све информације које су нам у скупу података биле не тако очигледне.

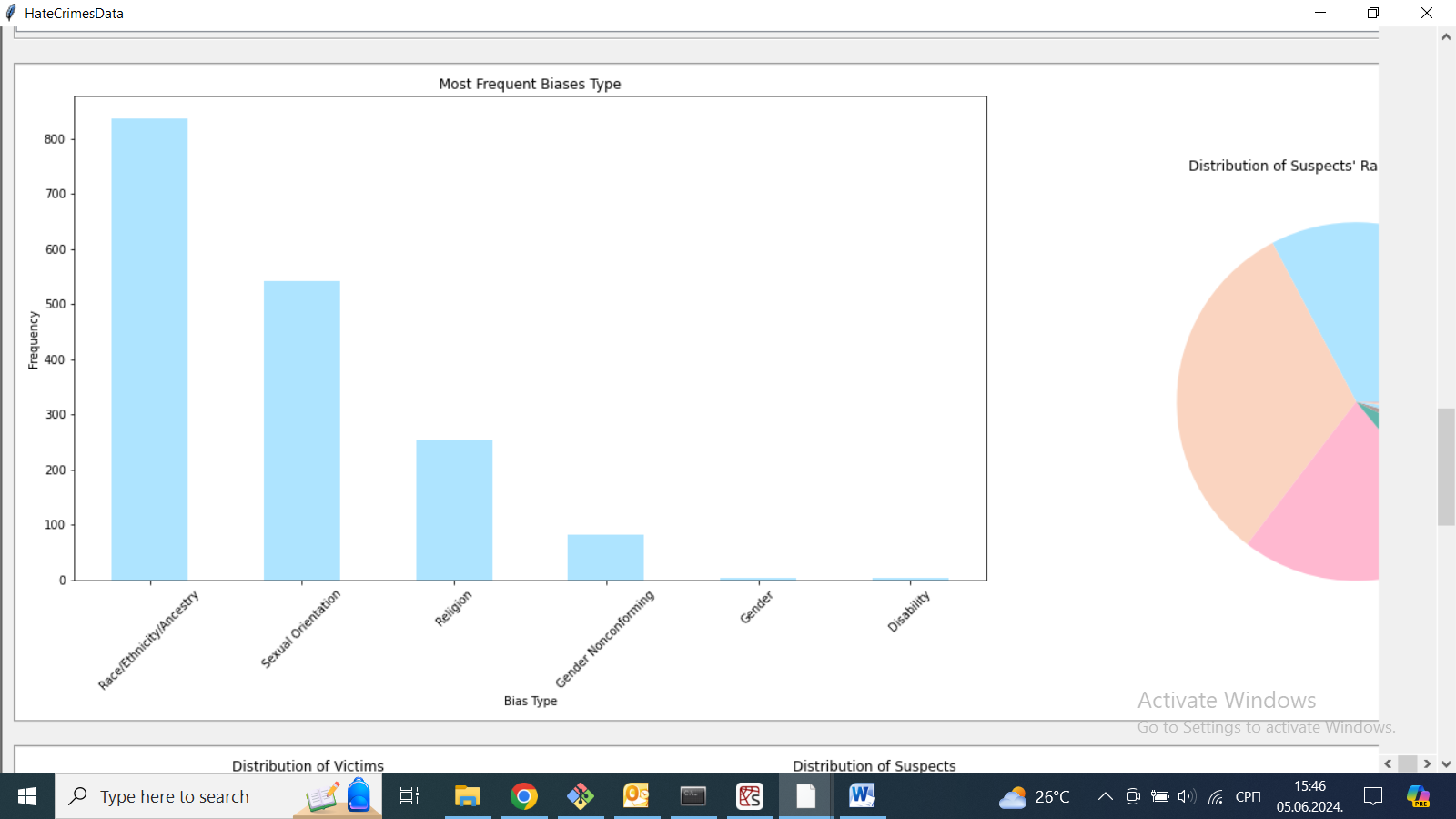
Издвојићемо неколико графикона.



На горњем графикону јасно можемо видети да је број злочина генерално опадао, са одређеним скоковима. Занимљиво је приметити да након завршеног карантина 2021. године број жртава расте и бива највећи у последњих дванаест година.

На наредним графиконима ћемо посматрати графиконе који се тичу самих предрасуда, размотрићемо мотивације и пристрасност.

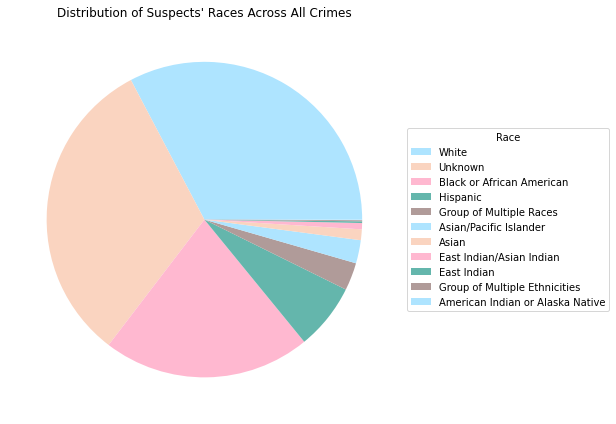




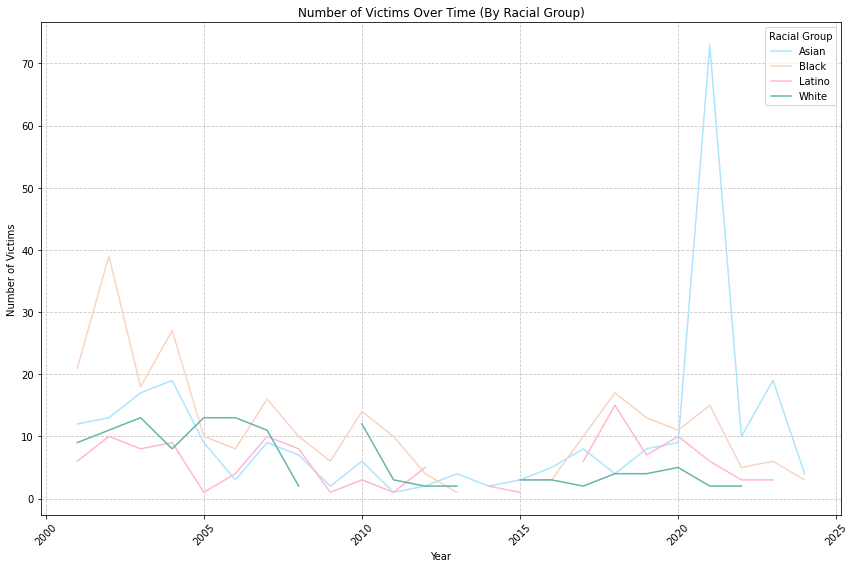
На последња два графикона увиђамо да се злочини најчешће догађају на јавним местима као што су улица, паркин, паркови и слично. Узимајући у обзир да су ово злочини пријављени полицији, овај податак има смисла с обзиром да би на јавним местима био велики број сведока који би могао да пријави злочин.

Такође, на графикону је приказано само пет најчешћих локација, те је веома забрињавајуће да се велики број злочина одвија у нечијем дому или на јавним местима које имају обезбеђење као што су зграде под управом државе или ресторанима.

Следеће што сазнајемо је да је најчешћи узрок због којег се изврши злочин нетрпрељивост опрема дређеној раси, етничкој припадности или пореклу, а потом сексуална оријентација. Све ово је детаљније истражено у самом пројекту, где су ове колоне разложене на подкатегорије.



На четвртом графикону закључујемо на основу података да су припадници европеидне расе најчешћи осумњичени за извршење злочина, а потом и припадници негроидне расе. Такође наилазимо на проблем јер велики број осумњичених није изјаснио које је расе, али је свеједно занимљиво удубити се у график и упоредити га са наредним...



На петом графику закључујемо да је током времена монголоидна раса постала знатно угроженија у односу на припаднике других раса. Видимо да се велики скок десио након пандемије COVID-19 која је очигледно допринела порасту ксенофобије и расизма према људима азијског порекла.

Са друге стране, видимо да нетрпељивост према припадницима негроидне расе константно опада у протеклих пар година. Ово може бити резултат многобројних едукација о расизму које се дешавају у Америци, потенцијално и протеста Black Lives Matter који су захтевали спровођење правде жртвама расизма.

# Закључак

## Анализа испуњења циљева истраживања

Сматрамо да су наши циљеви већински испуњени. Из података које смо имале на располагању сматрамо да је извучен максимум и да смо дошле до занимљивих запажања и закључака.

Идентификовани су поједини фактори који доприносе оваквим злочинима, а то су велике промене у свету као што су економске кризе, светске пандемије, али и подизање свести о оваквим проблемима који смањују њихову учесталост.

Уочиле смо и образложиле учесталост одређених друштвених групе које су мете оваквих злочина и шаблоне које починитељи свесно или несвесно поштују приликом њихових злочина.

## Анализа остварења очекиваних резултата истраживања

Дошли смо до очекиваних резултата у смислу да је реализован задатак осмишљања детаљног и прегледног графичког приказа наших анализа.

## Могућности за примену истраживања у пракси

Сматрамо да наше истрађивање има широку и може имати веома значајан утицај на друштво.

Самим информисањем о оваквим темама диже се свест становништва о учесталости оваквих злочина. У последњих пет до десет година отвореније се прича о оваквим темама и сматрамо да баш због тога видимо пад броја жртава током година на нашем првом приказаном графикону.

Такође, знање о шаблонима, најучесталијим местима и оружијима може значајно помоћи при превенцији сличних злочина у будућности.

Подаци о учесталости злочина могу донети и промене у области казнене политике, пропорционално кажнавање починилаца пружа правду жртвама и шаље поруку да се овакво понашање не толерише.

## Идеје за побољшање и разраду истраживања

Оно што нам је свакако фалило у нашем истраживању је бољи модел за предикције, нако је метод насумичних стабала био у могућности да предвиди и прати скокове, није био онолико близу колико бисмо желе. То би било нешто што бисмо детаљније истражиле у будућности и нашле прикладније решење.

За разраду нашег истраживања би био идеалан обимнији скуп података. Одређени подаци би знатно допринели нашој анализи. Профилисање починитеља и жртава би било знатно једноставније уколико бисмо имале приступ подацима о њиховом статусу запослености, економском статусу, образовању… Исто важи и за жртве, било би изузетно занимљиво стећи увид о томе ко је особа која почини овакав злочин. Можда бисмо били у стању да боље проценимо да ли су овакви злочини већински последица необразовања, страха или само неоснована мржња према припадницима одређених друштвених заједница.с

# Литература

1. Fredrik Lundh, An Introduction to Tkinter, http://jgaltier.free.fr/Terminale\_S/ISN/TclTk\_Introduction\_To\_Tkinter.pdf
2. https://stefanini.com/en/insights/news/machine-learning-models-for-precise-predictive-analytics
3. Материјал са вежби из предмета Методе и технике науке о подацима и Нумерички алгоритми и нумерички софтвер