**Slide 1**

Zdravo, moje ime je Ožbej Golob. Pri izdelavi projektne naloge mi pomagajo še kolegi Bian Klančnik, Gal Lindič, Tadej Ločniškar in Jan Weissenbach.

Naslov naše projektne naloge je »Analiza prometnih nesreč leta 2019 (v Sloveniji). Vir naših podatkov smo našli na uradni strani slovenske policije in sicer v oddelku s statistiko. Policija zbira te podatke, da jih lahko na koncu leta analizira, da spremlja stanje na cestah glede na prejšnja leta in tako dalje.

Podatki v naši zbirki so v obliki »Comma Separated Values« in vsebujejo 33896 primerov in 35 atributov. Podatkovna zbirka je dokaj celovita, tu pa tam je pri kakšni nesreči kakšen atribut brez vpisane vrednosti. Pri statistični analizi smo takšen primer izpustili. V zbirki je tudi podatek o dolžini vozniškega staža, vendar pri vrednosti 0 let in 0 mesecev nismo vedeli, ali so notri upoštevani vozniki brez izpita in podobno. Tudi takšne primere smo pri analizi izpustili.

Celotno zbirko smo prebrali s pomočjo pythonove knjižnice DictReader v en slovar. Pri posamezni nalogi si je nato vsak ustvaril poseben slovar oziroma tabelo z vrednostmi potrebnih atributov in podatkov.

Ker smo analizirali samo en vir podatkov, le teh nismo rabili združevati s čim drugim.

Pri izdelavi projektne naloge smo si zastavili naslednja vprašanja:

* Zanima nas, kateri vsi dejavniki vplivajo na število povzročenih nesreč
* Zanima nas tudi, kaj vse vpliva na število nesreč s smrtnim izidom
* Želimo odkriti najbolj izpostavljeno skupino udeležencev v prometu

**Slide 2**

Pri cilju, kjer želimo odkriti najbolj izpostavljeno skupino v prometu želimo ustvariti napovedni model za prometne nesreče različnih tipov. Na primer, s pomočjo naivnega Bayesa želimo preveriti, kakšna je verjetnost, da se je nesreča končala s smrtnim izidom, če je bil udeležen motorist, ali če je bil udeležen pijan voznik, če je udeleženec uporabljal varnostni pas in podobno.

Želimo si tudi poiskati povezanost med atributi, kjer bi preverili odvisnost atributov med sabo s pomočjo regresije, pearsonovega koeficienta in drugih metod.

Do sedaj smo analizirali vpliv alkohola, ure, spola in vozniškega staža na število prometnih nesreč. Pregledali smo tudi, kako uporaba varnostnega pasu vpliva na izid prometne nesreče ter kako je tip poškodb udeležencev odvisen od tipa prometne nesreče (na primer, ali je trčenje čelno, bočno, naletno in podobno). Vse ugotovitve smo prikazali s pomočjo vizualizacij, ki si jih lahko ogledamo na naslednjem slajdu.

**Slide 3**

Za začetek se osredotočimo na zgornji levi graf, ki prikazuje vrste poškodb glede na uporabo varnostnega pasu. Modri stolpci prikazujejo udeležence, ki so uporabljali varnostni pas, oranžni pa tiste, ki ga niso. Kot vidimo, je večji delež udeležencev brez poškodb tistih, ki so bili pripasani. Po tem pa vidimo, da je večji delež lažje, huje poškodovanih in smrtnih žrtev udeležencev tistih, ki niso bili pripasani.

Pod tem grafom vidimo dva grafa, ki prikazujeta najprej število vseh nesreč glede na uro, nato pa število nesreč, kjer je povzročitelj pijan, glede na uro.

Graf desno zgoraj pa prikazuje število povzročenih nesreč glede na dolžino vozniškega staža. Kot lahko vidimo, se število povzročenih nesreč z dolžino vozniškega staža drastično zmanjša. Nad temi podatki smo izračunali pearsonov koeficient, ki je enak **-0.95**. To pomeni, da sta si dolžina vozniškega staža in število povzročenih nesreč obratno sorazmerna.

Pod tem grafom levo je graf, ki prikazuje število povzročenih nesreč glede na spol povzročitelja. Kot lahko vidimo, so moški povzročili približno dvakrat več nesreč kot ženske.

Desno od tega grafa pa vidimo še graf, ki prikazuje število smrti glede na tip udeleženca. Kot vidimo v vizualizaciji, je v resnici večje število (približno 60%) mrtvih povzročiteljev.

**Slide 4**

Za konec smo pripravili še GIF, ki prikazuje razpršenost in število nesreč po mesecih. Na grafu ena točka predstavlja eno prometno nesrečo. Iz grafa lahko vidimo, kje in kateri mesec se zgodi največ nesreč.

**To je iz naše strani vse. Hvala za pozornost.**