

# PV251: Vizualizácia

## Projekt report

## Motiváciu

Ako som hľadal dáta pre tento projekt povedal som si, že spravím vizualizáciu niečoho čo mi pomáha počas karantény a celkovo počas pandémie ostať zdravý (mentálne) a to je pivo. Začalo to už minulý rok počas prvej “karantény”, kde sa zatvorili školy a výučba sa stala distančnou. Keďže vďaka tomuto som ostával doma a pri PC ešte viac ako doteraz tak, som sa s kamarátom rozhodol, že začneme chodiť na prechádzku. No keďže sa chodí s prázdnyimi rukami zle, tak sme zavítali do pivotéky, kde predávajú obrovské množstvo druhov pív, pričom tieto piva sú hlavne z menších pivovarov. Keďže pivo je veľmi veľa tak som začal hľadať na Kaggle nejaký dataset, kde by sa dali nájsť nejaké druhy spolu s ich základnými detailmi ako miesto výroby, meno piva a množstvo alkoholu.

## Opis dát

Prvý dataset „open-beer-dataset“ obsahoval veľa stĺpcov pričom väčšia bola pre mňa zbytočná preto som sa rozhodol to osekať na tieto stĺpce Názov piva (Name), Alkohol podľa objemu (Alcohol by volumn), IBU (International Bitterness Units), Druh (Style), Výrobca (Brewer), Mesto(City), Štát (State) a Krajina kde sa pivo vyrobilo (Country). Tento dataset obsahoval približne 5973 dát.

Stĺpec Name, predstavuje meno piva. Tento stĺpec obsahuje hodnoty textového typu, pričom z celkového počtu nenulových ich bolo 5963.

Stĺpec Alcohol by volumn, tento stĺpec predstavuje množstvo alkoholu podľa objemu piva. Obsahuje hodnoty typu float, pričom dosť ich bolo nulových čo som prvotne nevidel až neskôr. Hodnôt, ktoré vôbec neboli definované, bolo okolo 5948 a celkovo hodnôt, ktoré boli nenulové, bolo okolo 2500.

O stĺpci International Bitterness Units môžeme povedať, že je to medzinárodná stupnica, ktorá hovorí o kvantite horkosti piva. Tento stĺpec obsahoval až príliš veľa nulových dát.

Stĺpec Style, hovorí o druhu piva, a teda aké pivo to vlastne je, teda či je to ALE, Porter, Stout alebo veľa iných, celkovo tam bolo cez 160 unikátnych hodnôt.

Stĺpec Brewer je názov pivovaru, ktorý pivo vyrába, tento obsahoval pomerne veľa nenulových hodnôt.

Ostatné stĺpce obsahovali už len mesto, štát a krajinu výroby piva, tieto hodnoty som použil hlavne na vizualizáciu na mape, aby sme si vedeli predstaviť, kde sa aké piva vyrábajú.

Čo sa týka druhého datasetu, ten bol tvorený dvoma súbormi pričom prvý obsahoval veľmi podobné dáta ako prvý, ale bez názvu výrobcu a krajiny, keďže tento dataset je len z USA túto hodnotu som tam musel dodávať. Prvý súbor obsahoval id výrobcov, pričom druhý dataset obsahoval mená týchto výrobcov.

## Predspracovanie dát

Celková veľkosť prvého datasetu bola okolo 6000, pričom vo viacerých stĺpcoch boli nulové hodnoty, tie som musel vymazať keďže ich vizualizácia by bola veľmi náročná, hlavne to boli dáta kde nebol druh piva,

alkohol podľa objemu definovaný. Po odstránení všetkých nedefinovaných hodnôt (NAN) a odstránení zbytočných stĺpcov som získal okolo 2500 dát. Táto veľkosť sa mi zdala celkom malá, tak som hľadal dataset, ktorým by som mohol rozšíriť tento pôvodný.

Rozširujúci dataset pôvodne obsahoval 2410 dát, pričom tu už bolo oveľa menej nulových hodnôt ako v predošlom, no stále tam boli.

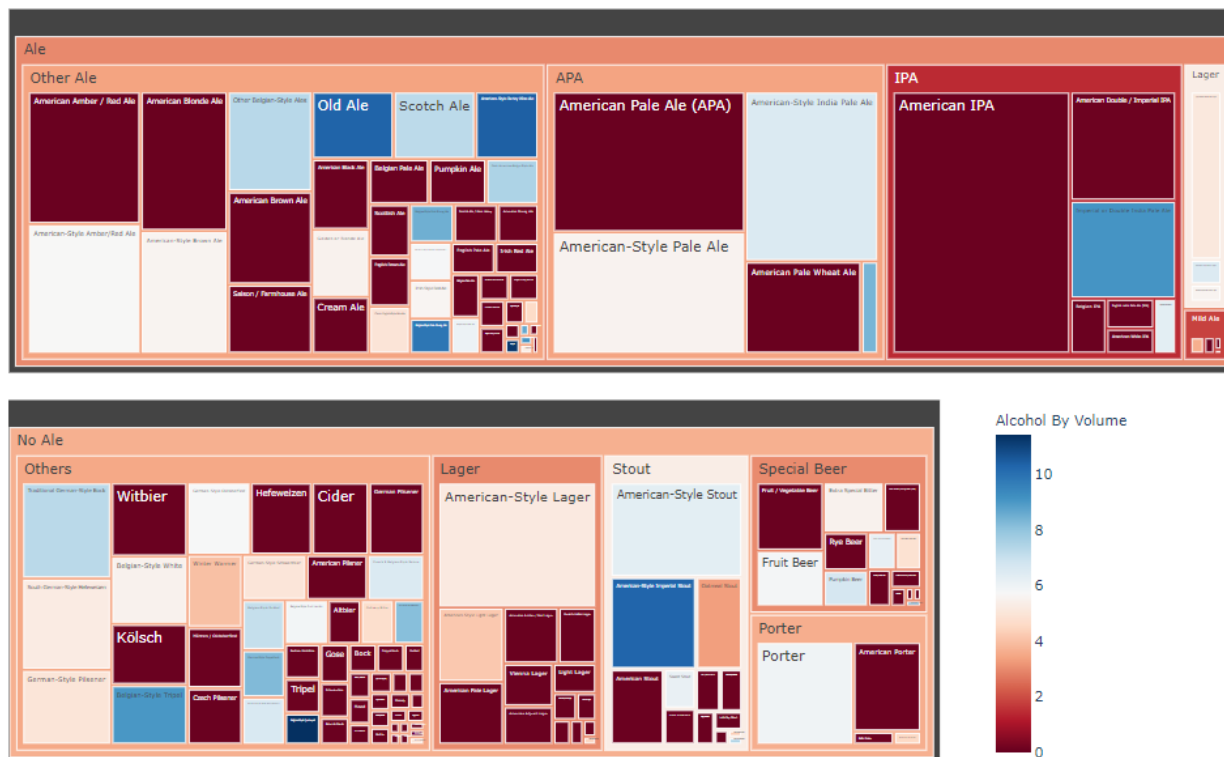
Celkové predspracovanie dát pre prvý dataset spočívalo v tom, že som najprv odstránil všetky stĺpce ktoré som nepotreboval ('Category', 'Description', 'id', 'brewery\_id', 'cat\_id', 'Universal Product Code', 'Address', 'Standard Reference Method', 'style\_id', 'filepath', 'last\_mod', 'Coordinates', 'Website', 'add\_user'), následne som dataset prefiltroval, kde nie je alkohol nulový (všetky piva čo mali nulový alkohol ho nemali mať, boli to piva, ktoré som overoval, že majú aspoň nejaké percentá alkoholu, to mi aj potvrdilo keď niektoré piva mali alkohol okolo 5-6 desatinných miest %), následne som pozaokrúhľoval hodnoty tohto alkoholu nech sú krajšie viditeľné v tabuľke. Ďalej som odstránil riadky kde nulové hodnoty boli v stĺpci Style.

Predspracovanie dát pre druhý dataset bolo také, že som najprv spojil spomínané dva súbory kde jeden obsahoval id výrobcu piva a druhý názov tohto výrobcu. Následne som odstránil zbytočné stĺpce ('id\_x', 'id\_y', 'ounces', 'brewery\_id'), pridal som stĺpec s názvom Country a hodnotou „United States“. Znova odstránil riadky, ktoré mali nulovú hodnotu v stĺpci Style a Alcohol by Volume.

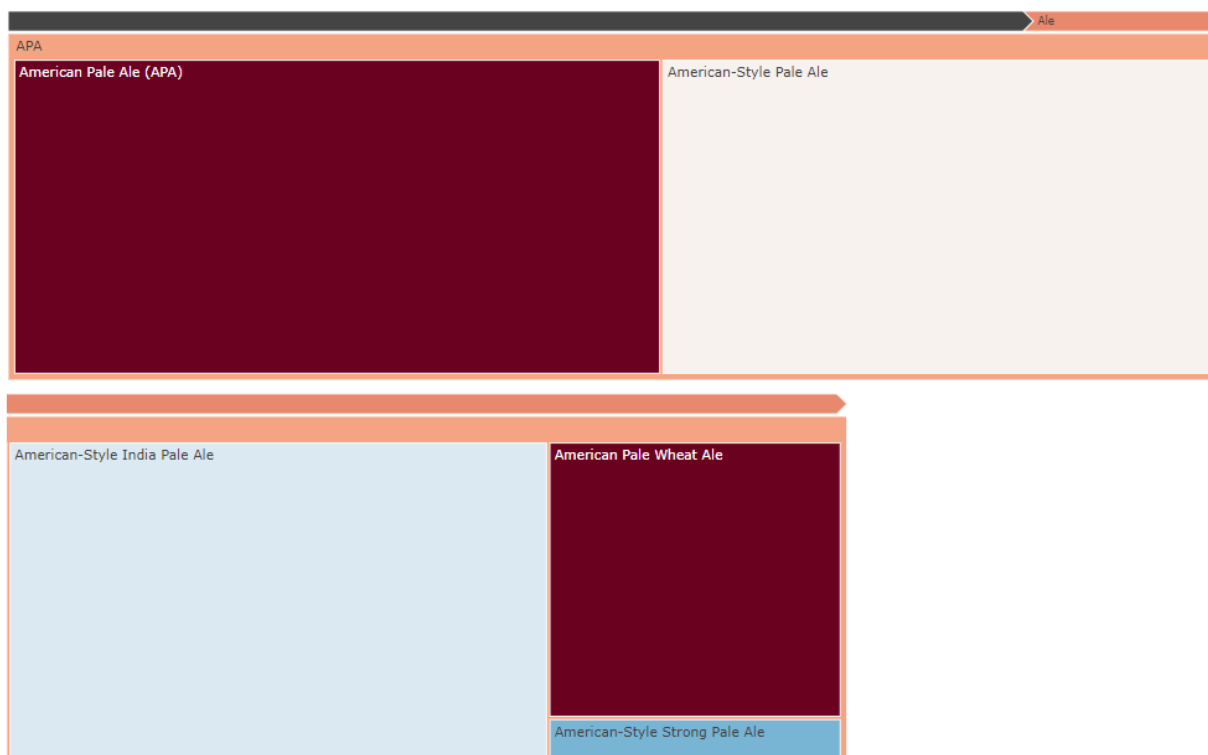
Po tomto som datasety spojil do jedného a začalo predspracovanie pre treemap. Ten potreboval mať piva rozdelené do viacerých úrovní a tie som vytvoril pridaním umelých stĺpcov. Najprv som pridal stĺpec kde boli hodnoty len 1 a 0, pričom tieto hodnoty som priraďoval na základe toho či pivo je ALE alebo nie. Následne som ich rozdelil do skupín podľa typu, pričom tento stĺpec obsahoval hodnoty „Porter, Stout, Lager, APA(alebo jej varianty), Mild ALE, IPA (jej varianty) a Špeciálne piva“. Predspracovanie dát som dosť zhrnul, no celkovo to predstavovalo viacero námahy ako vytvorenie vizualizácie.

## Použité vizualizačné metódy / metódy vizualizácie

Ako prvé môžeme vidieť treemap, kde sú všetky piva rozdelené do 3 úrovní, prvá úroveň je ALE a No Ale, druhá úroveň je pre každú kategóriu iná, pre ALE to je OtherAle, APA, IPA, Lager(niektoré piva boli definované ako Lager or Ale), Mild Ale. Pre No ALE to je Lager, Stout, Special Beer a Porter. Tento treemap môžeme vidieť na obrázku č. 1. Tento je rozdelený podľa počtu pív, čím viac je pív, tým je okno s pivom väčšie. Taktiež je rozdelené na farby a teda čím modrejšia je farba okna, tým viac má pivo alkoholu, legendu môžeme vidieť aj vedľa celého treemapu. Tento treemap je plne klikateľný a naň sú naviazané ďalšie elementy, Mapa a aj tabuľka. Pokiaľ viackrát klikneme na jeden element, nič sa nestane, teda treemap sa vráti späť, lenže dáta sa nezmenia, to je spôsobené tým, že tento treemap by sa takto nemal správať, lenže celá implementácia je robená v dash python, a do jednotlivých elementov som zasahovať nechcel. No pokiaľ klikneme inam ako na ten element na ktorý sme klikli, dáta a elementy sa upravujú. Ako príklad sa správa treemap môžeme vidieť na obrázku číslo 2. Taktiež môžeme prejsť nad jednotlivými položkami a budeme vidieť presný počet pív a presnú hodnotu alkoholu. Príklad tohto detailu môžeme vidieť na obrázku číslo 3.



Obrázok č. 1 Treemap rozdelená na dve časti



Obrázok č. 2 Kliknutie na položku APA v Treemape

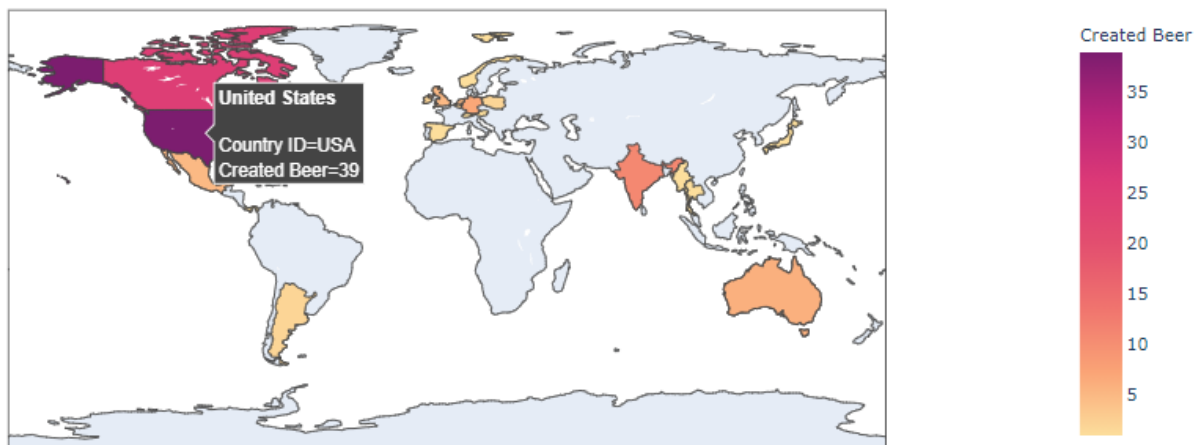
```
labels=Others  
count=948  
parent=No Ale  
id=No Ale/Others  
Style=(?)  
Alcohol By Volume=3.512623417721519
```

Obrázok č. 3 Príklad pre kategóriu Others

Ako ďalšou metódou pre vizualizáciu som zvolil mapu, táto mapa je plne prepojená s tabuľkou, a teda aj s treemapou. Teda pokiaľ sa upraví dáta v tabuľke upraví sa aj táto mapa, a to tak, že vyznačí jednotlivé pív na mape, a tiež aj ich počet. Pokiaľ prejdeme nad jednotlivú krajinu uvidíme názov krajiny a presný počet pív, ktoré sa vytvorili v tejto krajine. Príklad tejto mapy pre všetky dáta môžeme vidieť na obrázku číslo 4, a na obrázku číslo 5 môžeme vidieť mapu pre kategóriu Lager a s detailom na USA.



Obrázok č. 4 Mapa pre všetky dáta



Obrázok č. 5 Mapa pre kategóriu Lager s detailom na USA

Ako posledná metóda vizualizácie je tabuľka, pričom táto tabuľka je plne prepojená na treemap, a teda pokiaľ vyberieme nejakú kategóriu v treemape zobrazia sa nám dáta len pre túto tabuľku. Tabuľka obsahuje stĺpce Meno piva, Alkohol, IBU, Kategória, Výrobca, Mesto a Krajinu. V tejto tabuľke vieme zoradiť každý stĺpec, taktiež tabuľka obsahuje plne funkčné stránkovanie a teda listovanie medzi jednotlivými záznamami. Nad tabuľkou sa nachádza dropdown menu, kde môžeme filtrovať podľa jednotlivých krajín a podľa jednotlivých výrobcov. Príklad tejto tabuľky, kde zobrazujeme všetky dáta môžeme vidieť na obrázku číslo 6 a príklad tabuľky pre kategóriu Stout krajiny USA a výrobcu môžeme vidieť na obrázku číslo 7.

### Beers detail

Country to search						
Brewer to search						
Name	Alcohol By Volume	IBU	Style	Brewer	City	Country
Possession Porter	5.6	0	Porter	Diamond Knot Brewery & Alehouse	Mukilteo	United States
Free Bike Amber	4.5	0	American-Style Amber/Red Ale	Skagit River Brewing	Mount Vernon	United States
Potrero ESB	5.2	0	Special Bitter or Best Bitter	21st Amendment Brewery Cafe	San Francisco	United States
Twisted Kilt Scotch Ale	5	0	Scotch Ale	Thirsty Dog Brewing	Akron	United States
Grolsch Amber Ale	5.3	0	German-Style Brown Ale/Altbier	Grolsche Bierbrouwerij	Enschede	Netherlands
Harvest Ale Fresh Hop Ale	7.2	0	American-Style Pale Ale	Lift Bridge Brewery	Stillwater	United States
Big Hoppy Monster	8.75	0	American-Style Amber/Red Ale	Terrapin Beer Company	Athens	United States
Brigantine Smoked Porter	8.54	0	Porter	South County Brewing Co.	Fawn Grove	United States

Obrázok č. 6 Tabuľka dát

### Beers detail

United States						
21st Amendment Brewery Cafe						
Name	Alcohol By Volume	IBU	Style	Brewer	City	Country
563 Stout	5	0	American-Style Stout	21st Amendment Brewery Cafe	San Francisco	United States
Oyster Point Oyster Stout	5.9	0	American-Style Stout	21st Amendment Brewery Cafe	San Francisco	United States

Obrázok č. 7 Tabuľka pre kategóriu Stout, krajinu USA a výrobcu 21 Amendment Brewery Cafe

Celkový vzhľad stránky môžeme vidieť na obrázku číslo 8, táto stránka je na heroku na linku <https://beer-visualization.herokuapp.com/>.

Select your favorite beer

