TO DO:

1) Proces PRODEJ\_BYTY rozšířit i na zbylé tři kategorie (PRONÁJEM\_DOMY)

2) Ceny za metr nějak agregovat na oblast – město? Aby se daly porovnávat změny i jinak než mediánem inzerátů v oblasti

3) Heat mapy? Jinde než v PowerBI?¨

4) Porovnávání starých a nových inzerátů pomocí URL\_ID

5) Zaokrouhlování i SHORT\_COORDS kvůli lepšímu mapování města, hledání duplikátů, sledování starých inzerátů?

6) Zpřesnit hledání duplikátů i na drobné odchylky u COORDS a u CEN?

7) Popasovat se s více byty na stejné adres – sice mají unikátní URL\_ID, ale třeba v mapě se ten bod potom třeba čtvrtí v bublině.

8) Histogramy, Shluky, Dendrogramy..?

9) Vizualizace v Pythonu potřebují interaktivní Wudgety

10) Interaktivní Obsah

11) Variable Inepsctor

12) Highlight pivot extremy <https://pandas.pydata.org/pandas-docs/stable/user_guide/style.html>

13) Geodata a lepší mapy než powerBI?

14) obecně práce s extrémy, z textu odhalit, kdy jde jen o pronájem nebo půlku nemovitosti

15) Motivace pro náš projekt – Prezentace - Prezi + PowerBI + Příběh + Excel s výpočty?

16) spočítat průměry cen za kraje, aby to v PBI rovnou šlo líp do mapy na obarvení bodů. Nebo alternativně to udělat jako metriku v PBI

17) Predikce, Oceňování podle parametrů, Neuronová síť …

18) Problém !!!! U stahování PRONÁJEM BYTY mi to správně detekuje počet inzerátů a stran, ale ukládání URLs skončí po první stránce !! jen20 záznamů !!!

19) Problém: Barevné sloupce nejde groupovat do side by side plotu :((

20) Přejít z Excelů na CSV !!

21) SCRAPING funkce: Volba mezi tím nechat to defaultní, nebo si to předem vždycky nadefinovat.. ? Nadefinovat, abych mohl správně pojmenovat Excely, i když zatím jedu jen jednu kategorii pořád...

22) Nápad - udělat něco jako indikátor toho, kde jsem skončil, resp kde se SCRAP pokazil. Resp při spuštění se fce koukne (TRY) který soubor jde načíst, tedy jestli už mám uložené něco z prvního pokusu o scrapping, pokud ano, načte ho a uloží nějaké číslo do proměnné, což bude pak indikovat, na který krok dál skočit. Bude to brát odzadu. takže pokud se nevytvořil žádný file, sjede nakoenc to, co už zná, celý scrap od kraje.

23) Průběžně si data během scrapingu ukládat do bloků, časově to moc výsledek neovlivní?

Kód na skrývající HTML tlačítko:

from IPython.display import HTML

HTML('''<script>

code\_show=true;

function code\_toggle() {

if (code\_show){

$('div.input').hide();

} else {

$('div.input').show();

}

code\_show = !code\_show

}

$( document ).ready(code\_toggle);

</script>

<form action="javascript:code\_toggle()"><input type="submit" value="Ukaž/Schovej všechen kód"></form>''')

# df.isna().any() vyhodí seznam sloupců, ve kterém je NAN hodnota

#data.isin([-1]).any() mi najde -1 kdekoliv v df

# df.dropna()

# df.fillna(0)

# data[data.duplicated(subset='url\_id', keep=False)].sort\_values(by=['url\_id']) = Hledání duplikátů

#data.nunique() = Rychlý přehled u všech sloupcích o počtu unikátních hodnot

#data\_all.plocha.value\_counts()

#data.groupby("plocha\_2")["popis"].count()

# Tohle mi to řadí i podle počtu, což ale nuztně nechci, groupby mi to hodí spíš podle abecedy, tím najdu divné hodnoty

# Ještě by byl ořešení přes nunique(), ale to it neřekne které to jsou, jen ukáže celkový počet druhů