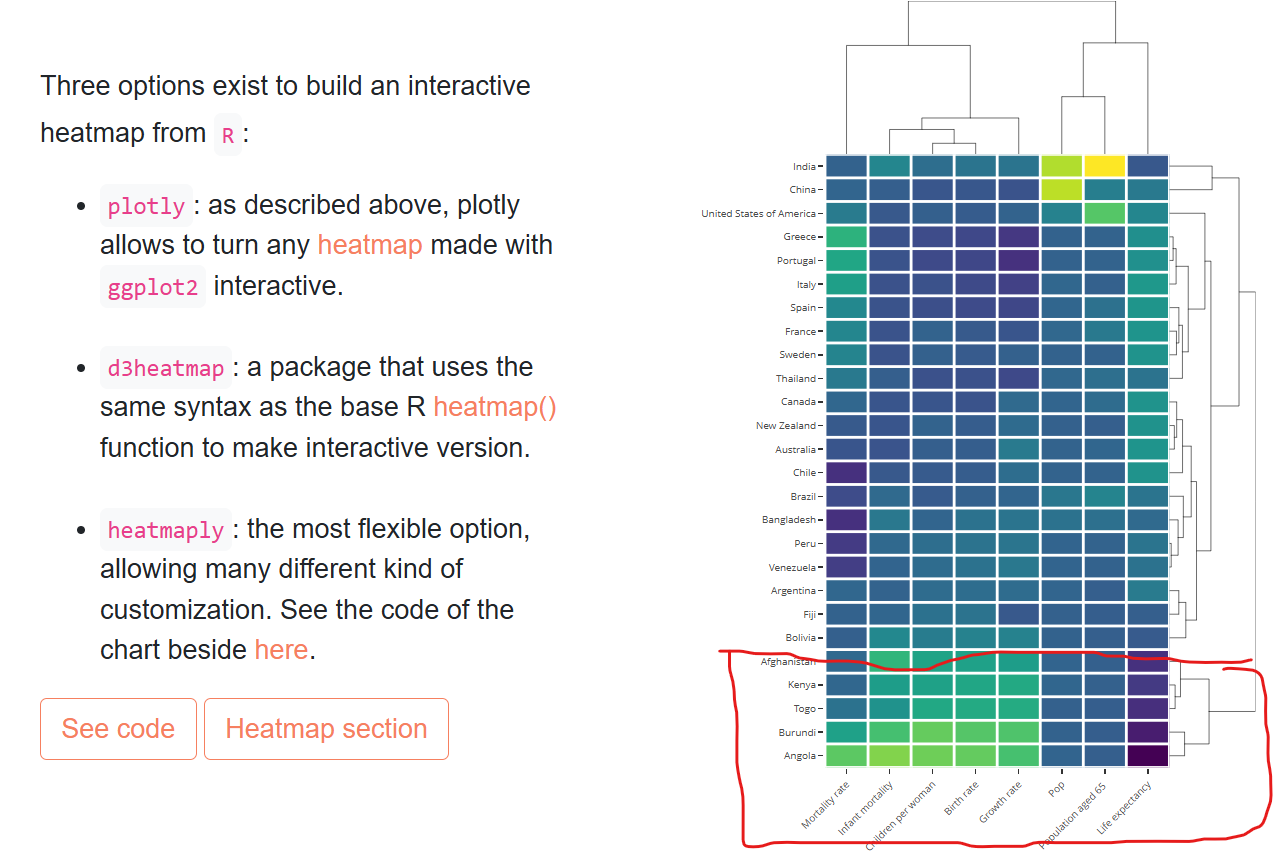
Pomysły do realizacji w ramach projektu

**1. Analiza wpływu lokalizacji i odległości od centrum miasta na ceny mieszkań**

* Zbadanie, jak odległość od centrum miasta (pole centreDistance) wpływa na cenę sprzedaży mieszkań (cena). – rozpoczęte, do doszlifowania 😊
* Wykorzystanie wizualizacji takich jak wykresy punktowe oraz ~~mapy cieplne.~~
* *~~Wybrać 5 do tego co niżej i zbudować heatmap?~~*



* Porównanie wyników pomiędzy miastami. – *wybrano 6 miast*

**2. Model predykcji cen mieszkań (po wybraniu testów stat.)**

* Opracowanie modelu regresji liniowej lub wielokrotnej do przewidywania cen mieszkań na podstawie zmiennych takich jak:
  + Powierzchnia (squareMeters),
  + Liczba pokoi (rooms),
  + Odległość od centrum (centreDistance),
  + Rok budowy (buildYear).
* *~~Weryfikacja jakości modelu poprzez podział danych na zestaw treningowy i testowy.~~*
  + *~~Iza do weryfikacji i potwierdzenia h~~*

**~~3. Porównanie cen mieszkań w różnych miastach~~**

* ~~Stworzenie interaktywnych wykresów porównujących średnie ceny mieszkań w 15 miastach.~~ *~~(mysle że analizy dla 15 miast to może być za dużo, projekt będzie się ciągnąć w nieskończoność a na zajęciach prowadzący mówił że nie o to w tym chodzi)~~*
  + *~~Można dać bubble/scatterplot/scatter – bez potrzeby większej analizy. Po prostu pokazać na wykresie i zinterpretować~~*
* ~~Uwzględnienie czynników, takich jak powierzchnia mieszkań, liczba pokoi i typ budynku (type).~~

**4. Wpływ udogodnień na wartość nieruchomości**

* Analiza, jak obecność udogodnień (has[features], np. balkon, parking) wpływa na cenę mieszkań.
* Porównanie mieszkań z podobnymi atrybutami, ale różnymi udogodnieniami.
* *Klasteryzacja (np. wpływ odległości od centrum czy liczby pokoi na podział na klastry.) (jako smaczek)*

**5. Profil mieszkań kupowanych w różnych miastach**

* Opracowanie statystyk opisowych dla poszczególnych miast:
  + Średnia powierzchnia mieszkań,
  + ~~Dominujący rok budowy (buildYear),~~
  + Najczęstsza liczba pokoi (rooms).
* Wizualizacja wyników w formie raportów tabelarycznych i wykresów.

**~~6. Analiza trendów czasowych~~**

* ~~Jeśli w danych uwzględniono okres od sierpnia 2023 do czerwca 2024, można przeanalizować zmiany średnich cen mieszkań w czasie.~~
* ~~Wizualizacja trendów za pomocą wykresów liniowych.~~

**7. Interaktywna mapa cen mieszkań**

* Wykorzystanie współrzędnych geograficznych (latitude, longitude) do stworzenia mapy przedstawiającej ceny mieszkań w różnych lokalizacjach.

**8. Zróżnicowanie cen mieszkań w zależności od liczby pokoi i stanu mieszkania**

* Analiza wpływu liczby pokoi (rooms) i stanu mieszkania (condition) na cenę.
* Porównanie średnich cen w grupach mieszkań o różnych stanach (np. nowe, do remontu).

~~9. Wykonać analizy dla top 5 miast, które mają najwyższe kwoty za metr kwadratowy.~~ 6 miast /

~~10. Dekompozycja na podstawie modelu regresji (chodzi o szeregi czasowe, możemy ocenić które zmienne/co najbardziej wpływa na ceny nieruchomości ale również można to zrobić zwykłą regresją)~~

~~12.~~ *~~Może dodatkowa analiza miesięczna dla tych top 5 miast? (szereg czasowy ok)~~* ~~To jest, w jakich dniach miesiąca drożeją, a w jakich tanieją mieszkania: czy pod koniec, czy na początku miesiąca.~~

~~13. Zwrócenie uwagi też na rodzaj przy analizie: czy to są mieszkania nowe; do remontu, a może kamienice?~~ Powtórzenie punkt 8

14. Można te miasta rozłożyć jeszcze pod kątem 3-5 najdroższych dzielnic, albo też przedstawić te najtańsze dzielnice dla kontrastu. Mogłoby być całkiem ciekawie, gdyby w drodze tego, okazało się, że najtańsza dzielnica w Krakowie czy Gdańsku, odpowiada cenowo najdroższej w takiej Bydgoszczy. Nie sądzę, żeby aż taki był rozstrzał, ale mogłoby być to ciekawe.

~~15. Analiza pozostałych czynników, wpływających na ceny: czy np. miasto jest studenckie, możliwości pracy, ile ludzi przyjezdnych zostaje finalnie w mieście, rozrost miasta i życia w nim, inwestycje w infrastrukturę, dostępność komunikacyjna, czy są parki blisko.~~

~~16. Pora roku a wahania cen,~~ ~~turystyczność miasta,~~ *uwarunkowania geograficzne*. ~~Czy wyższe ceny mieszkań mogą być w jakiś sposób zależne od popularności krótkoterminowego wynajmu?~~

~~17. Korelacja między liczbą ofert, a średnią ceną w danym miesiącu~~

~~18. Przebadanie szybkości sprzedaży między miastami: w których miastach mieszkania szybciej sprzedają i dlaczego?~~

19. Średni dochód w mieście, a średnia cena mieszkania.

* *Można też przeanalizować zmiany tego wskaźnika*

~~20. Inflacja, stopy procentowe i ich wpływ na średnie ceny mieszkań~~

~~21. Struktura wiekowa i liczba mieszkańców a wpływ na ceny mieszkań.~~

~~22. Koszty zakupu a koszty wynajmu~~

~~23. Które miasta mają największy potencjał wzrostu cen na podstawie przeanalizowanych danych?~~

24. Wnioskowanie (testy statystyczne)

**1. Wnioskowanie statystyczne:**

**a. Testy hipotez**

* **Porównanie średnich cen mieszkań:**
  + **Test t-Studenta**: Porównanie średnich cen mieszkań z balkonem i bez balkonu, z windą i bez windy, itp.
  + **ANOVA**: Porównanie średnich cen w różnych miastach (np. Warszawa vs Kraków vs Gdańsk)
* **Wpływ udogodnień na cenę:**
  + Test dla proporcji (dla zmiennych binarnych, np. czy mieszkania z parkingiem są droższe od mieszkań bez parkingu?).

**b. Analiza korelacji**

* Korelacja między zmiennymi ilościowymi, np.:
  + Powierzchnia mieszkania (squareMeters) vs cena (price).
  + Odległość od centrum (centreDistance) vs cena (price).
  + Liczba pokoi (rooms) vs cena (price).
  + Rok budowy (buildYear) vs cena (price).
* Test korelacji Pearsona (sprawdzamy czy jest statystycznie związek między zmiennymi)

**c. Testy zależności**

* **Chi-kwadrat**: Test zależności między zmiennymi jakościowymi, np. czy obecność parkingu zależy od lokalizacji (miasta).
* **Logit/Probit**: Modelowanie wpływu cech binarnych (np. obecności windy) na prawdopodobieństwo, że mieszkanie osiągnie określoną cenę.

**2. Modelowanie statystyczne:**

**a. Regresja liniowa**

* Analiza wpływu cech mieszkania (powierzchnia, liczba pokoi, obecność windy, parkingu, itd.) na cenę (price).
* Model: B0 + B1 \* square\_meters + B2 \* rooms + … + C

**b. Regresja wielomianowa**

* Rozszerzenie regresji liniowej o nieliniowe zależności (np. wpływ odległości od centrum na cenę, gdzie zależność może nie być liniowa).

**c. Regresja logistyczna**

* Modelowanie prawdopodobieństwa wystąpienia określonego zdarzenia, np. czy mieszkanie z określonymi cechami kosztuje powyżej mediany cen.

**3. Klasyfikacja i segmentacja:**

**a. Analiza skupień (k-means, hierarchiczna)**

* Grupowanie mieszkań na podstawie cech takich jak cena, lokalizacja, powierzchnia, rok budowy.

**b. Analiza dyskryminacyjna**

* Klasyfikacja mieszkań w kategorie na podstawie ich cech (np. tanie, średnie, drogie).

**3. Testy jakości modelu:**

* **AIC/BIC/AICC:** Ocena dopasowania modeli regresji.
* **R²:** Ocena jakości wyjaśnienia zmienności przez modele regresyjne.

**4. Test Shapiro – Wilka**

Sprawdzić czy dane mają rozkład normalny (do naszych histogramów)