

Marketing Mix Modeling

Przetwarzanie danych: czynniki zewnętrzne – cz.1

Wydział Nauk Ekonomicznych | MediaCom Business Science
09 | 11 | 2022



UNIwersYTET WARSZAWSKI
Wydział Nauk Ekonomicznych



UNIwersYTET
WARSZAWSKI

mediacom
business science



- 01 COVID
- 02 Święta i efekty kalendarzowe
- 03 Dane makroekonomiczne/
demograficzne/zewnętrzne
- 04 Dane trade'owe
- 05 Dane pogodowe
- 06 Słownik danych

W skład bazy wchodzi dane z kilkunastu różnych źródeł

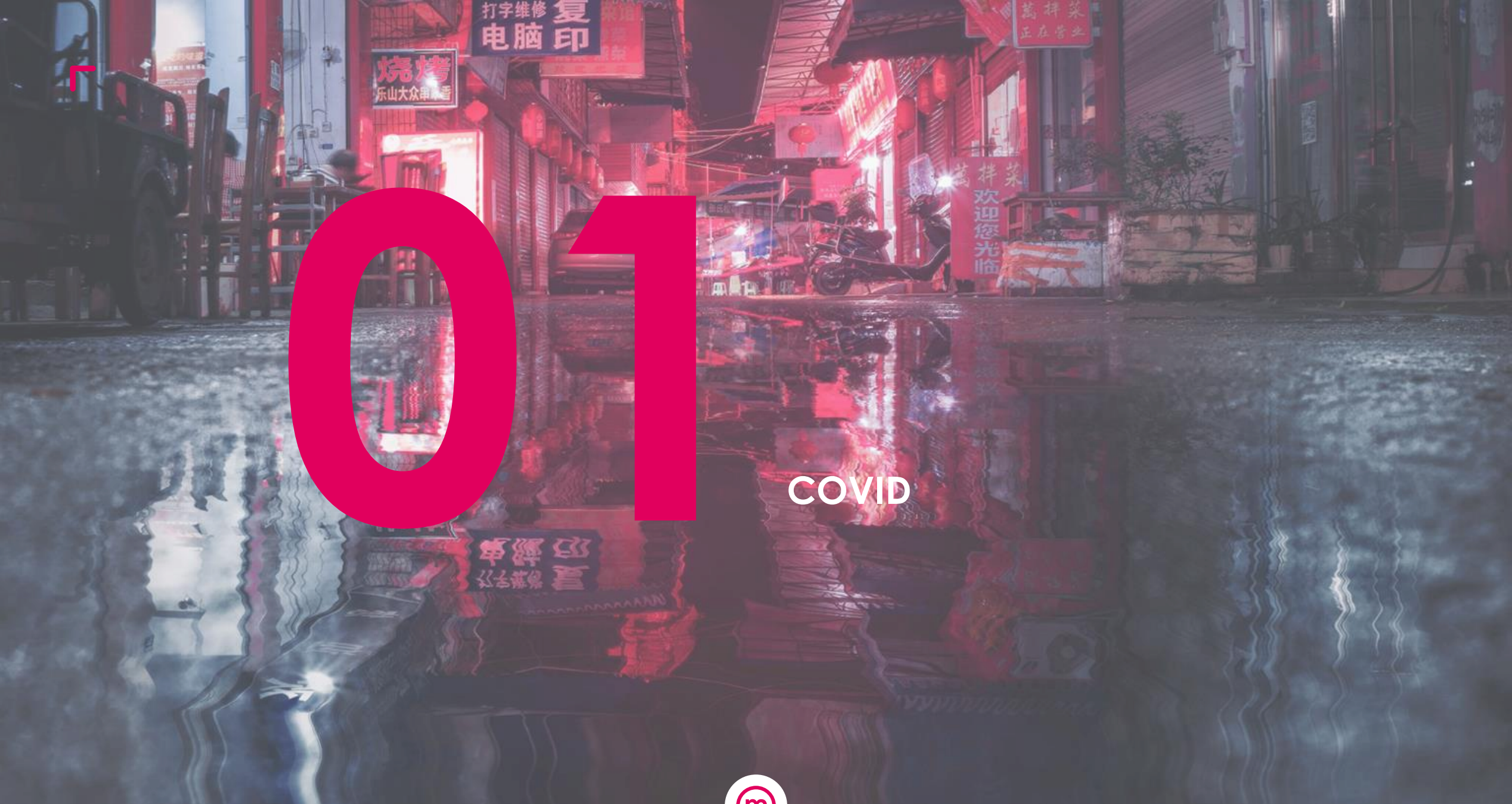
Obszar	Zmienne	Częstotliwość	Źródło
Dane sprzedażowe	Wartość, wolumen, dystrybucja, cena, liczba sklepów	Dz./tyg.	Klient (retail), Nielsen (FMCG), IQVIA (pharma), GfK (FMCG)
Aktywności tradeowe	Standy, faceing, płachty, katalogi, sampling, ulotki	Tyg./mies.	Klient, Nielsen, FOCUS
Aktywności mediowe	Telewizja, radio, social, search, display, VOD	Dz./tyg.	Nielsen, Google, Facebook, TikTok, Radio Track, inni dostawcy
Dane ekonomiczne	CPI, Konsumpcja, urodzenia, CCI	Mies./kw.	GUS, OECD, strony rządowe
Święta i efekty kalendarzowe	Święta, dni handlowe,	Dz./tyg.	Kalendarz, strony rządowe
Pogoda	Opady, temperatura, nasłonecznienie	Dz./tyg.	IMGW, strony rządowe, Dark Sky
Czynniki zewnętrzne	Trendy konsumenckie, COVID	Dz./tyg.	Google Trends, Google Mobility, WHO, agencje badawcze, dane rządowe
Sezonowość	Cykl sezonowości	Dz./tyg.	



W skład bazy wchodzi dane z kilkunastu różnych źródeł

Obszar	Zmienne	Częstotliwość	Źródło
Dane sprzedażowe	Wartość, wolumen, dystrybucja, cena, liczba sklepów	Dz./tyg.	Klient (retail), Nielsen (FMCG), IQVIA (pharma), GfK (FMCG)
Aktywności tradeowe	Standy, faceing, płachty, katalogi, sampling, ulotki	Tyg./mies.	Klient, Nielsen, FOCUS
Aktywności mediowe	Telewizja, radio, social, search, display, VOD	Dz./tyg.	Nielsen, Google, Facebook, TikTok, Radio Track, inni dostawcy
Dane ekonomiczne	CPI, Konsumpcja, urodzenia, CCI	Mies./kw.	GUS, OECD, Euromonitor, strony rządowe
Święta i efekty kalendarzowe	Święta, dni handlowe,	Dz./tyg.	Kalendarz, strony rządowe
Pogoda	Opady, temperatura, nasłonecznienie	Dz./tyg.	IMGW, strony rządowe, Dark Sky
Czynniki zewnętrzne	Trendy konsumenckie, COVID	Dz./tyg.	Google Trends, Google Mobility, WHO, agencje badawcze, dane rządowe
Sezonowość	Cykl sezonowości	Dz./tyg.	



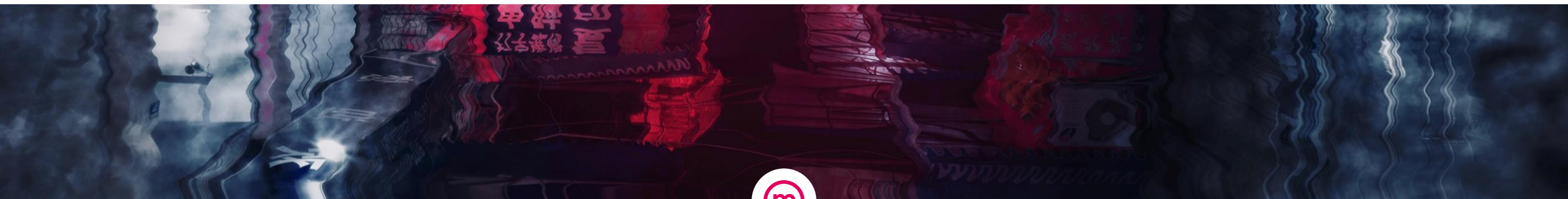


Wybór zmiennej odzwierciedlającej pandemię:

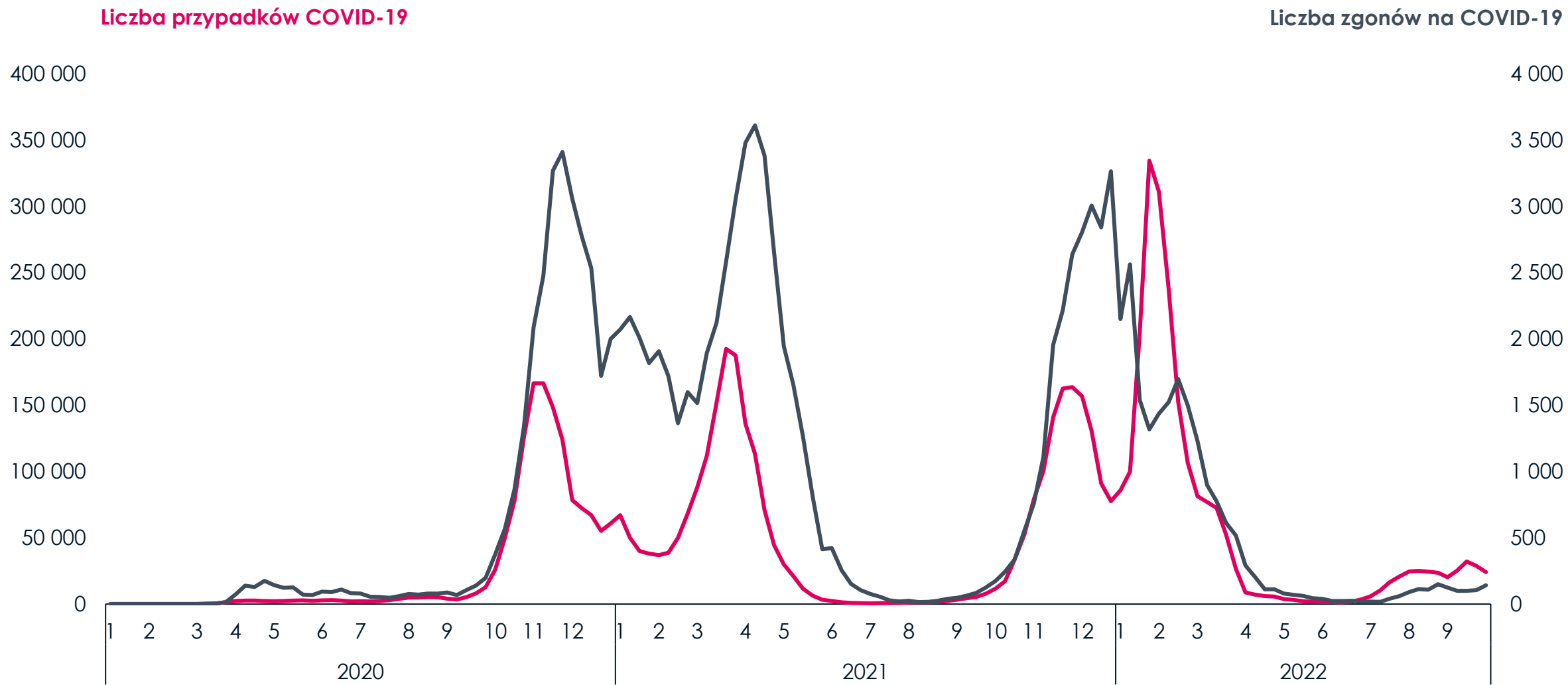
- Metryka niezawierająca w sobie efektów niezwiązanych z COVID-em
- Metryka mająca bezpośredni wpływ na zmienną objaśnianą
- Metryka mająca interpretację biznesową oraz zdrowotno-rozsządkową

Przykład danych

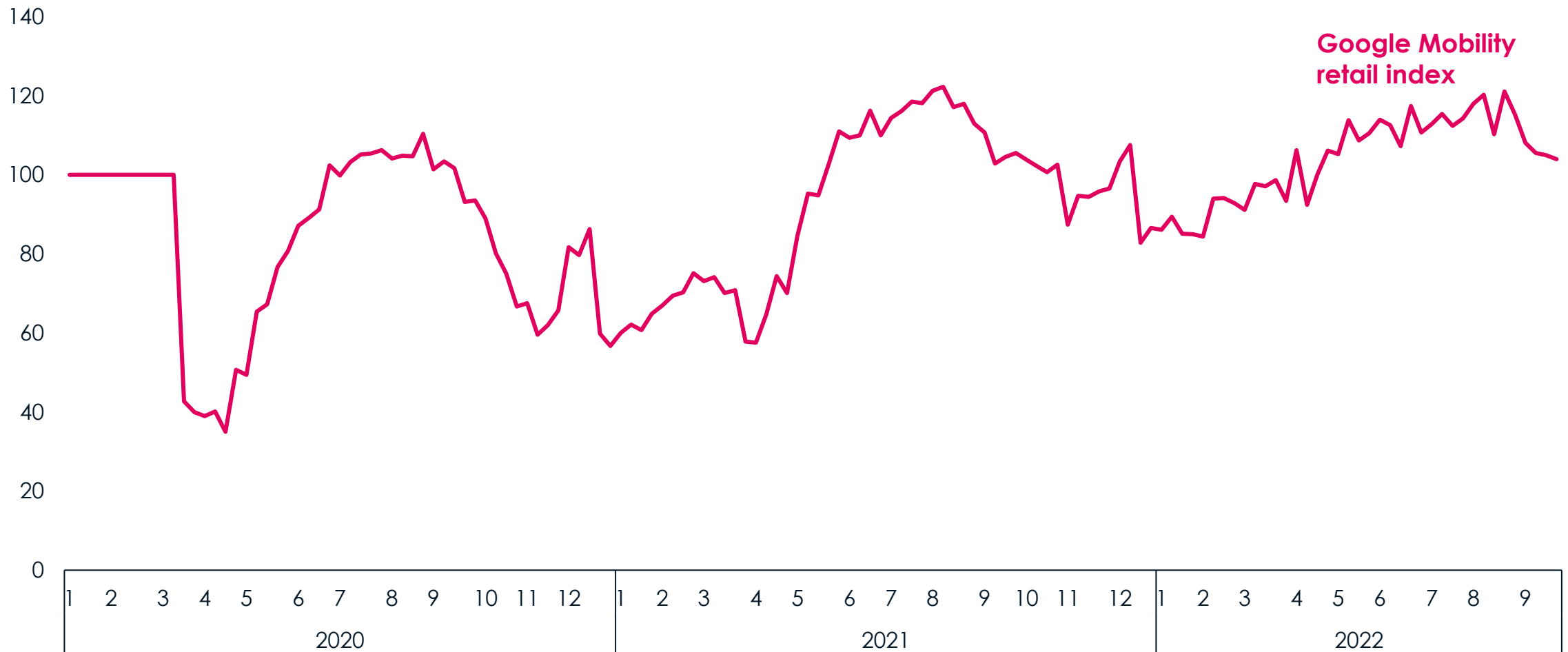
Liczba chorych
Liczba nowych przypadków
Liczba szczepień
Liczba zajętych respiratorów
Mobilność ludności
Wyszukania słów kluczowych w Google
Lockdowny (0-1)
Zamknięcia sklepów (0-1)
Stringency Index



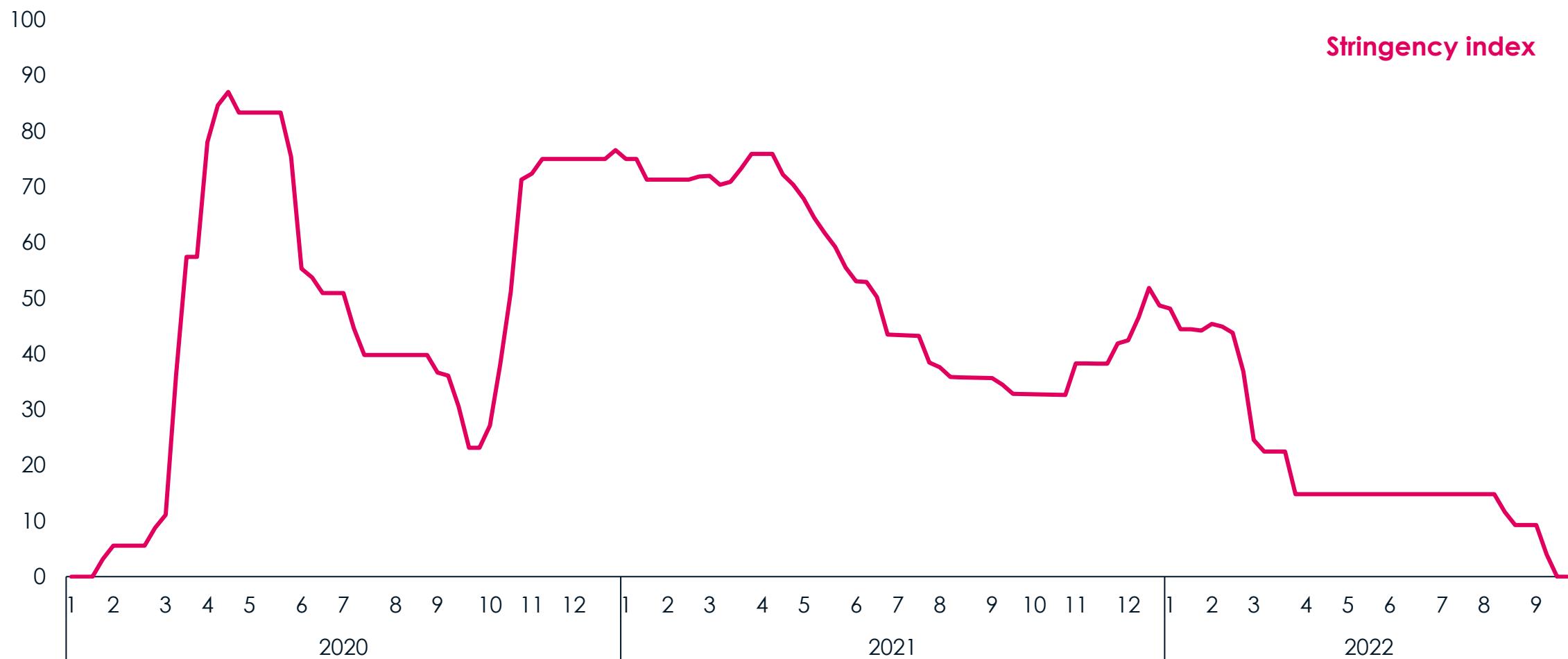
Przykłady danych COVID-owych – liczba przypadków i zgonów



Przykłady danych COVID-owych – Google Mobility Index

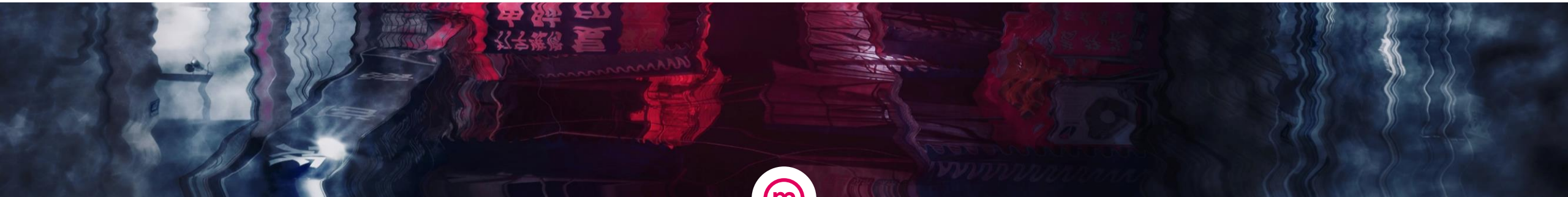


Przykłady danych COVID-owych – Stringency Index



Zadania

- 1) Wczytaj dane covidowe (covid_google_mobility_data.csv)
W osobnej ramce danych wybierz zmienną date i zmienną Mobility index dla retailu (kolumna retail_and_recreation_percent_change_from_baseline) i nazwij ją MOBILITY_RETAIL (1 pkt)
Zmień częstotliwość zmiennej z dziennej na tygodniową, licząc średnią z 7 dni dla każdego tygodnia (2 pkt)
- 2) Zestaw zmienne mobility index retail i residential (w częstotliwości dziennej) na wykresie liniowym
Zadbaj o to, by wykres był czytelny (1 pkt)





02

Święta i efekty kalendarzowe

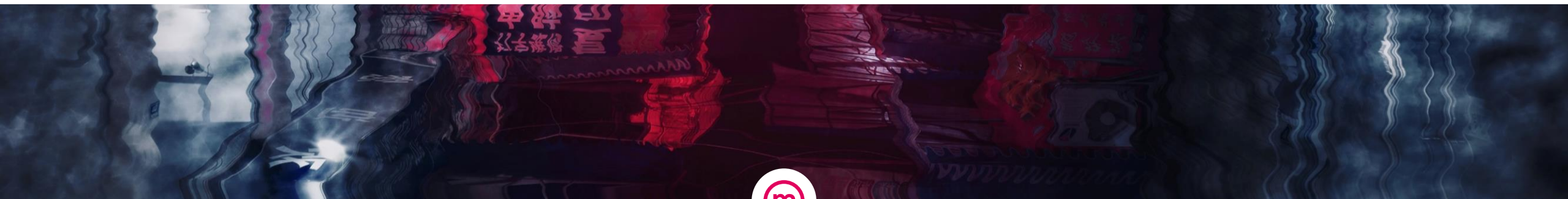
Święta i efekty kalendarzowe

Przykłady

- Zmienna przyjmująca wartość 1 w każdym tygodniu wielkanocnym lub majówkowym
- Zmienna przyjmująca wartość 1 w tygodniach obowiązywania narodowej kwarantanny
- Model dzienny: zmienna przyjmująca wartość 1 w każdy piątek

Best practise

- Zmienne binarne (i zero-jedynkowe) bardzo silnie oddziałują na statystyki modelu, trzeba więc być ostrożnym i nie używać ich liberalnie
- Nigdy nie wykorzystujemy w modelu zmiennych z wartością nierówną jeden w pojedynczym tygodniu (oczywiście są wyjątki)
- Zmienna binarna ZAWSZE musi mieć interpretację w modelu

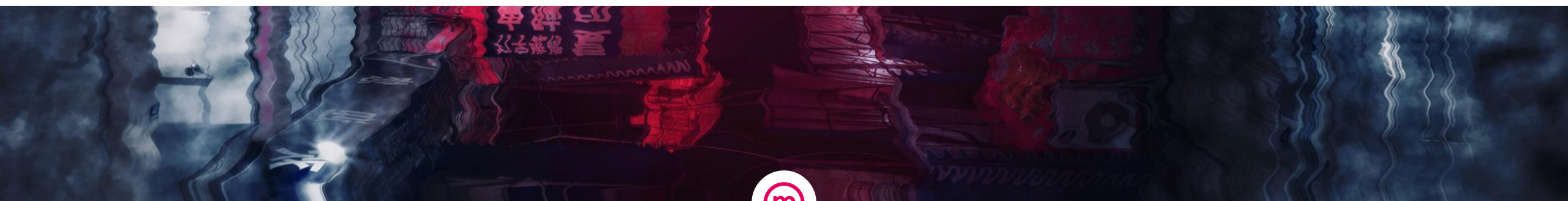


Zadania

3) Stwórz data frame ze zmienną Date, która przyjmuje wartości tygodniowe dat od 2021-01-04 do 2022-05-23. Możesz zastosować funkcję **seq()**. (1 pkt)

4) Stwórz 3 zmienne w przedziale dat podanym w zadaniu 2: (2 pkt)

- **XMAS** – która przyjmuje wartość 1, dla 25-tego grudnia każdego roku
- **EASTER** – która przyjmuje wartość 1 dla Niedzieli Wielkanocnej w każdym roku
- **Y2022** – która przyjmuje wartości 1 dla wszystkich dat w 2022 roku.





03

Dane makroekonomiczne /
demograficzne / zewnętrzne

Brainstorm – jakie czynniki makroekonomiczne, demograficzne mogą wpływać na sprzedaż piwa



Dane makroekonomiczne / demograficzne / zewnętrzne

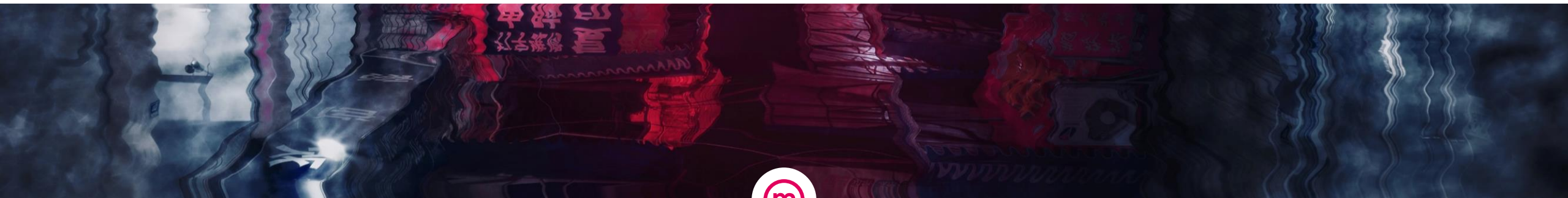
Przykłady

- Inflacja
- Konsumpcja prywatna
- Stopa bezrobocia
- CCI
- Urodzenia
- Wojna w Ukrainie – ruch na granicy

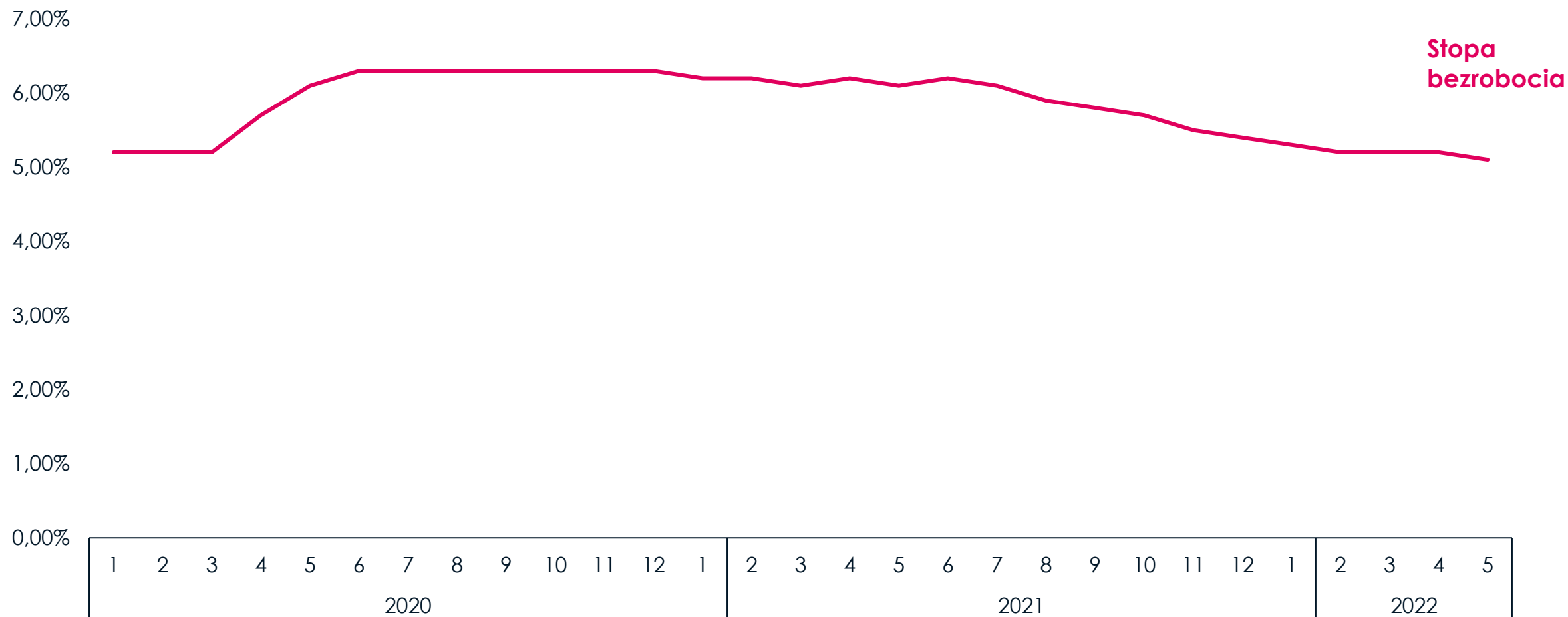
Transformacja danych:

- Zmienne zazwyczaj są w częstotliwości miesięcznej lub kwartalnej,
- W celu użycia ich naszych modelach, musimy w pierwszej kolejności je przekształcić do postaci tygodniowej
- Robimy to poprzez przypisanie wszystkim dat z konkretnego miesiąca jednej wartości, natomiast na przetomie miesięcy liczymy (najczęściej) średnie
- Następnie dane wygładzamy

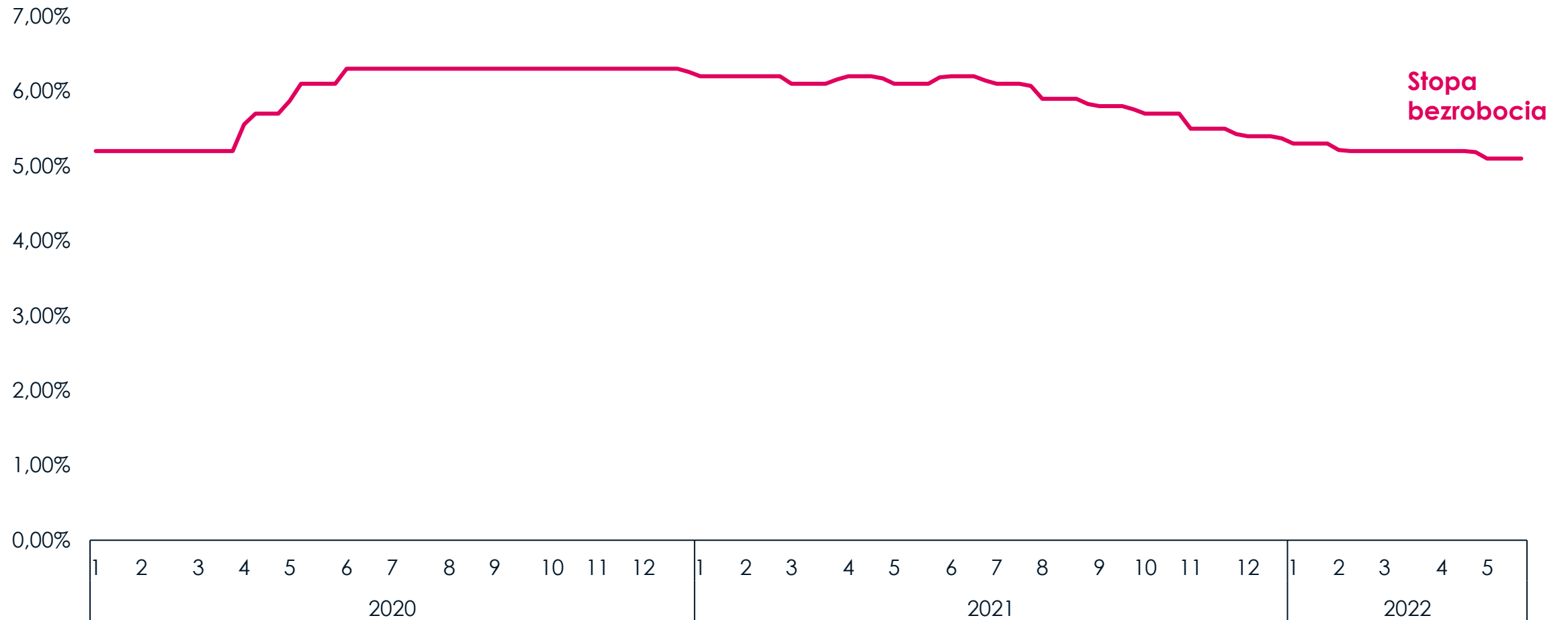
Źródła danych: GUS, Euromonitor, Dane straży granicznej



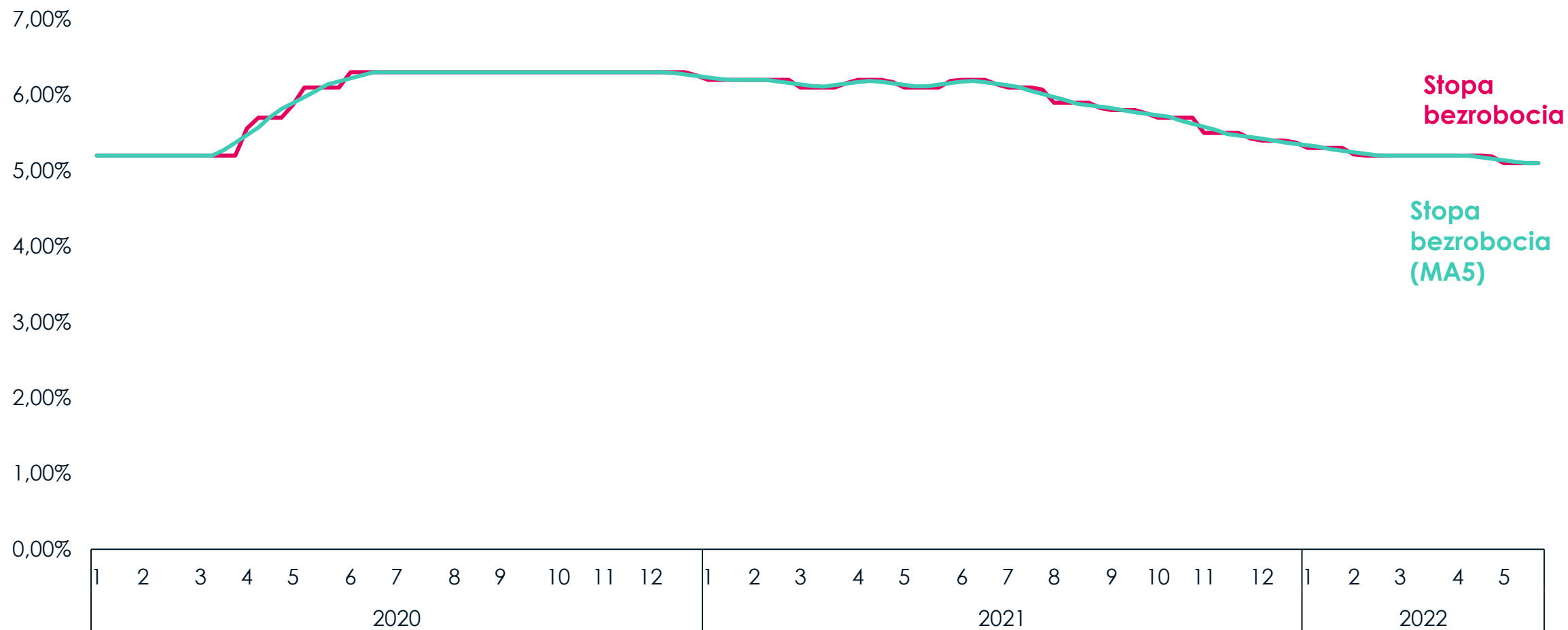
Dane o poziomie bezrobocia – częstotliwość miesięczna



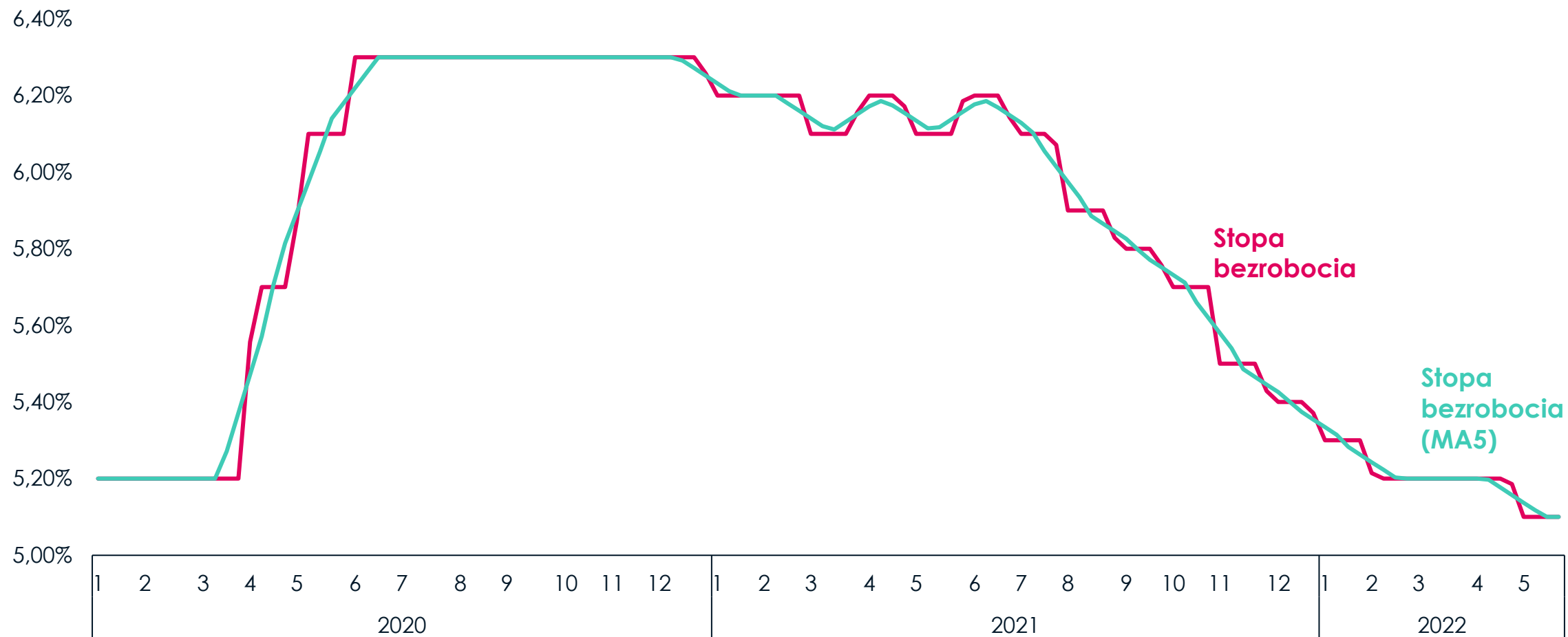
Transformacja danych o bezrobociu z miesięcznych na tygodniowe



Wyglądanie danych tygodniowych



Wyglądanie danych tygodniowych



Zadania

5) Wczytaj dane sales_data.xlsx

Na podstawie zmiennej **INFLACJA_WOW** (współczynnik inflacji tydzień do tygodnia, czyli np. 1.002 w tygodniu T6 znaczy, że poziom cen w T6 wzrósł o 0.2% w porównaniu do T5) przygotuj zmienne:

- o **INFLACJA_BASE**, która będzie zawierała skumulowany współczynnik inflacji w odniesieniu do pierwszego tygodnia przed okresem modelowanym (1 pkt)
- o **INFLACJA_SMOOTH**, która będzie scentrowaną średnią ruchomą z 5 tygodni policzoną na podstawie zmiennej INFLACJA_BASE (1 pkt)

Wynikiem zadania 5 powinna być niezależna ramka danych

