

Marketing Mix Modeling

Przetwarzanie danych: czynniki
zewnętrzne – cz.2

05 | 12 | 2023



UNIwersYTET WARSZAWSKI
Wydział Nauk Ekonomicznych



UNIwersYTET
WARSZAWSKI

essencemedia**com**
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 trade'owe	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.	



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 trade'owe	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.	





04

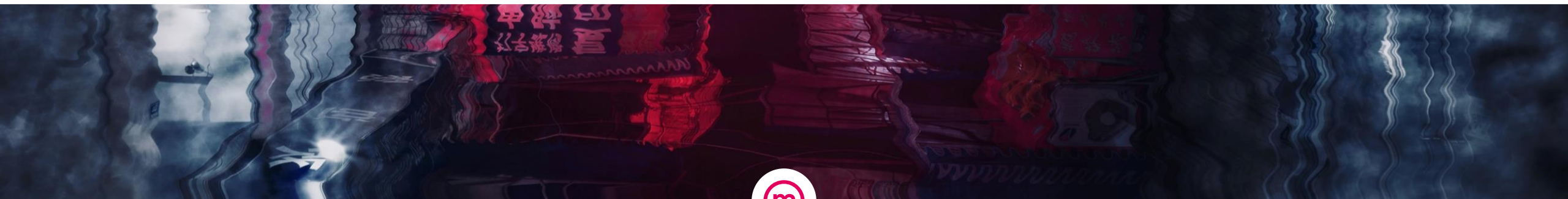
Dane trade'owe



Przykłady danych trade'owych

Przykłady danych:

- Ekspozycje w sklepach, standy
- Samplingi
- Gazetki sklepowe, często również dostępne online, katalogi, ulotki
- Obrandowane lodówki





05

Dane pogodowe



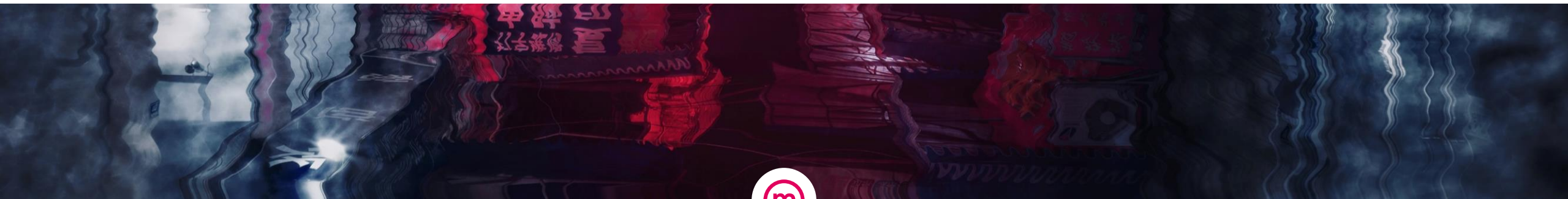
Przygotowanie danych pogodowych

Zmienna temperatura:

- Dane o temperaturze (maksymalnej, średniej, minimalnej), wyciągamy w podziale na miasta o częstotliwości dziennej
- Dane wyciągamy zarówno dla temperatury rzeczywistej jak i odczuwalnej
- Dane tygodniowe temperatur tworzymy w poprzez uśrednienie dziennych wartości
- Liczymy średnią ważoną temperaturę dla danego kraju, stosując do tego temperatury z poszczególnych miast/województw, ważąc ich populacjami

Zmienna opady:

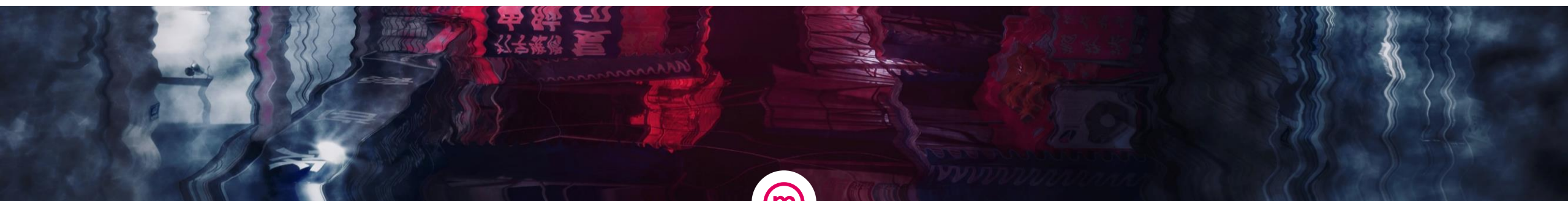
- Dane o opadach (deszcz, śnieg – opad w mm), wyciągamy w podziale na miasta o częstotliwości dziennej
- Dane tygodniowe opadów tworzymy w poprzez zsumowanie dziennych wartości
- Liczymy średni ważony opad dla danego kraju, stosując do tego opady z poszczególnych miast/województw ważąc ich populacjami



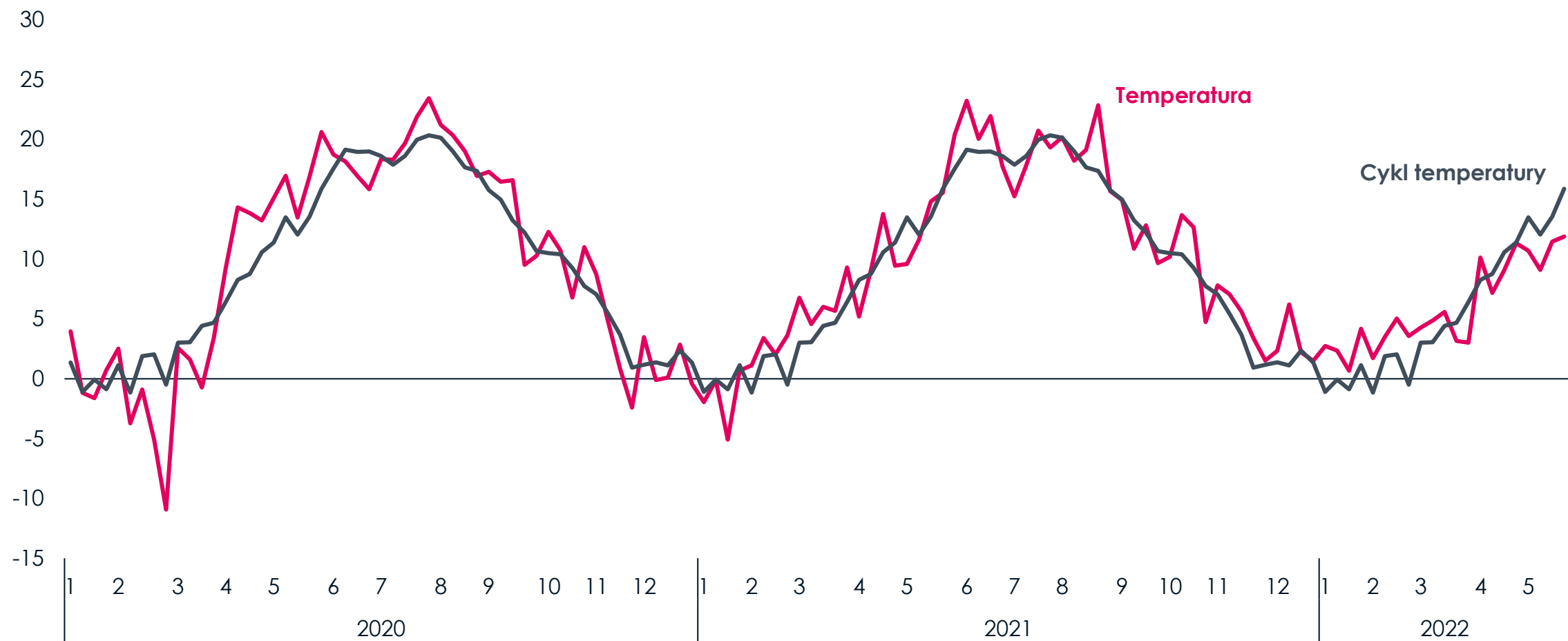
Przygotowanie danych pogodowych

Cykl temperatury i opadów:

- Tworzony na dłuższym okresie danych, aby faktycznie reprezentował cykl pogodowy
- Cykl temperatury powstaje poprzez policzenia średniej arytmetycznej dla każdego numeru tygodnia w roku, analogicznie dla opadów
- Wartości dla danego tygodnia powinny być takie same w każdym roku



Temperatura i cykl temperatury



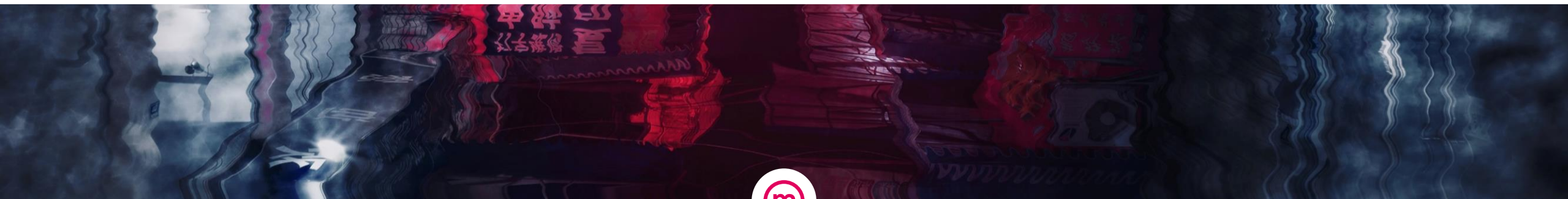
Zadania

Przygotuj dane o temperaturze dla całego kraju:

1. Na podstawie dziennych wartości temperatury z pliku **pogoda.xlsx** stwórz dane o częstotliwości tygodniowej w podziale na miejscowości (`station.name`) dla okresu od 2018-12-31 do 2022-05-28. (1 pkt)
Tip: do stworzenia daty tygodniowej z dziennej możesz zastosować funkcję **`floor_date()`** lub **`cut()`**.
2. Połącz dane pogodowe ze słownikiem województw (zakładka **słownik**) i policz średnią temperaturę per tydzień i województwo (1 pkt)
3. Złącz dane z poprzedniego punktu z wagami określającymi populację województw stosując wagi z zakładki **wagi**. Następnie policz średnią ważoną temperaturę dla całego kraju (1 pkt)

Przygotuj dane o cyklu temperatury, bazując na przygotowanych wcześniej zadaniach:

4. Do dataframe z zadania 3 dołącz numery tygodnia z zakładki **nrtygodnia** (nr tygodnia to cykl od 1 do 52) (1 pkt)
5. Policz cykl temperatury poprzez policzenie średniej temperatury dla każdego numeru tygodnia (1 pkt)
6. Stwórz finalny dataframe z wartościami cyklu temperatury dla okresu od 2018-12-31 do 2022-05-23, łącząc dataframey z punktów 4 i 5 (1 pkt)



06

Słownik danych

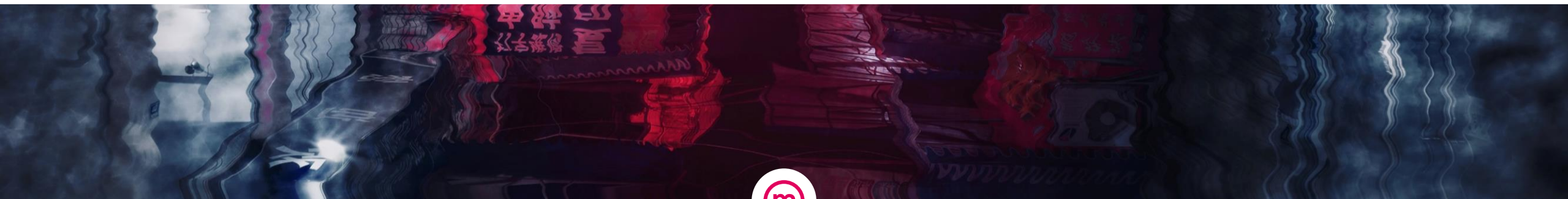
I wreszcie ... słownik danych!

Czym jest słownik?

- Słownik to zbiór informacji o kodowaniu poszczególnych zmiennych użytych w procesie przygotowania i następnie w modelowaniu
- Stosujemy usystematyzowane kodowanie zmiennych – staramy się w ramach kategorii być spójni

Do czego jest potrzebny słownik?

- Brak problemu wejścia w projekt
- Powrót do projektu po jakimś czasie jest znacznie prostszy,





essencemediacom
business science


Jarek Dejneka, Managing Partner
jaroslaw.dejneka@essencemediacom.com

Bartek Kowalski, Business Science Director
bartosz.kowalski@essencemediacom.com

 mbs@mediacom.com

 [@MBSWarsaw](https://www.instagram.com/MBSWarsaw)

 [@MBSWarsaw](https://twitter.com/MBSWarsaw)

 business-science.pl

***„An investment in knowledge
always pays the best interest”***

Benjamin Franklin

