

MONA KHALIL

SZTUKA ANALIZY DANYCH

Twarde i miękkie umiejętności
w czasach sztucznej inteligencji



Tytuł oryginału: Effective Data Analysis: Hard and soft skills

Tłumaczenie: Piotr Pilch

Projekt okładki: Studio Gravite / Olsztyn; Obarek, Pokoński, Pazdrijowski, Zaprucki;
fotografia: Adobe Stock

ISBN: 978-83-289-3306-4

© Helion S.A. 2026

Authorized translation of the English edition © 2025 Manning Publications.
This translation is published and sold by permission of Manning Publications,
the owner of all rights to publish and sell the same.

All rights reserved. No part of this book may be reproduced or transmitted
in any form or by any means, electronic or mechanical, including photocopying,
recording or by any information storage retrieval system, without permission
from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości
lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione.
Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie
książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie
praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi
bądź towarowymi ich właścicieli.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce
informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności
ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw
patentowych lub autorskich. Autor oraz wydawca nie ponoszą również żadnej
odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji
zawartych w książce.

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
helion.pl/user/opinie/szanda

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Helion S.A.

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 230 98 63

e-mail: helion@helion.pl

WWW: helion.pl (księgarnia internetowa, katalog książek)

Printed in Poland.

- Kup książkę
- Poleć książkę
- Oceń książkę

- Księgarnia internetowa
- Lubię to! » Nasza społeczność

Spis treści

<i>Słowo wstępne</i>	11
<i>Przedmowa</i>	13
<i>Podziękowania</i>	15
<i>Informacje o książce</i>	17
<i>O autorce</i>	20
CZĘŚĆ I. ZADAWANIE PYTAŃ	21
1. <i>Czym zajmuje się analityk?</i>	23
1.1. Czym jest analityka?	24
1.1.1. Analiza biznesowa	24
1.1.2. Analityka marketingowa	27
1.1.3. Analityka finansowa	29
1.1.4. Analityka produktowa	30
1.1.5. Jak bardzo odrębne są te dziedziny?	32
1.2. Zestaw narzędzi analityka danych	33
1.2.1. Arkusze kalkulacyjne	33
1.2.2. Język zapytań	33
1.2.3. Języki programowania do zastosowań statystycznych	35
1.2.4. Narzędzia do wizualizacji danych	37
1.2.5. Rozbudowa zestawu narzędzi	38
1.3. Przygotowanie się do swojej roli	38
1.3.1. Czego możesz się spodziewać jako analityk?	38
1.3.2. Czego nauczysz się z tej książki?	40
Podsumowanie	40
2. <i>Od pytania do rezultatu</i>	41
2.1. Cykl życia projektu analitycznego	42
2.1.1. Pytania i hipotezy	42
2.1.2. Źródła danych	44
2.1.3. Miary i metody analizy	49

2.1.4. Interpretacja wyników	52
2.1.5. Ćwiczenia	53
2.2. Komunikacja z interesariuszami	54
2.2.1. Wytyczne dotyczące interpretacji wyników	55
2.2.2. Wyniki, które nie potwierdzają hipotez	57
2.2.3. Ćwiczenia	62
2.3. Odtwarzalność	63
2.3.1. Czym jest odtwarzalność?	63
2.3.2. Dokumentowanie efektów pracy	64
2.3.3. Ćwiczenia	69
Podsumowanie	70
3. Testowanie i ocena hipotez	72
3.1. Formułowanie hipotezy	73
3.1.1. Gromadzenie podstawowych informacji	74
3.1.2. Formułowanie hipotezy	78
3.1.3. Ujęcie ilościowe hipotezy	80
3.1.4. Ćwiczenia	81
3.2. Metody gromadzenia dowodów	83
3.2.1. Opisy	83
3.2.2. Korelacje	88
3.2.3. Eksperymenty	92
3.2.4. Typy projektów badawczych	96
3.2.5. Ćwiczenia	99
3.3. Typy programów badawczych	100
3.3.1. Badania podstawowe i stosowane	100
3.3.2. Testy A/B	101
3.3.3. Ocena programów	102
Podsumowanie	102
CZĘŚĆ II. POMIARY	105
4. Metody statystyczne, których (prawdopodobnie)	
 się nauczyłeś: testy t, analiza wariancji i korelacje	107
4.1. Logika statystyk opisowych	108
4.1.1. Podsumowywanie właściwości danych	109
4.1.2. Podsumowanie	119
4.1.3. Ćwiczenia	120
4.2. Wnioskowanie: porównywanie grup	121
4.2.1. Testy parametryczne	122
4.2.2. Ćwiczenia	133
4.3. Wnioskowanie: korelacja i regresja	134
4.3.1. Współczynniki korelacji	135
4.3.2. Modelowanie regresji	137

4.3.3. Raportowanie korelacji i regresji	138
4.3.4. Ćwiczenia	139
Podsumowanie	140
5. Statystyki, które (prawdopodobnie) przeoczyłeś: metody nieparametryczne i ich interpretacja	142
5.1. Obraz statystyki	143
5.1.1. Ewolucja metod statystycznych	144
5.1.2. Odpowiedzialne wybieranie metod	147
5.2. Statystyka nieparametryczna	148
5.2.1. Porównania między grupami dla danych ciągłych lub porządkowych	148
5.2.2. Ćwiczenia	161
5.2.3. Porównywanie danych kategoriycznych	162
5.2.4. Podsumowanie	167
5.3. Odpowiedzialna interpretacja	167
5.3.1. Błędy statystyczne	168
5.3.2. Manipulowanie danymi	175
Podsumowanie	177
6. Czy mierzysz to, co postanowiłeś mierzyć?	179
6.1. Teoria pomiaru	180
6.1.1. Przekształcanie pomysłów w koncepcje	181
6.1.2. Typowe narzędzia pomiarowe	184
6.2. Wybór metody zbierania danych	186
6.2.1. Typy miar	186
6.2.2. Tworzenie miar samooceny	192
6.2.3. Interpretacja dostępnych danych	200
6.2.4. Ćwiczenia	203
6.3. Rzetelność i trafność	205
6.3.1. Rzetelność	205
6.3.2. Trafność	210
6.3.3. Ćwiczenia	212
Podsumowanie	213
7. Sztuka używania wskaźników: monitorowanie wyników pod kątem sukcesu firmy	215
7.1. Rola wskaźników w podejmowaniu decyzji	216
7.1.1. Monitorowanie wydajności	218
7.1.2. Kształtowanie strategii organizacyjnej	219
7.1.3. Promowanie odpowiedzialności	222
7.1.4. Ćwiczenia	223

7.2. Kluczowe zasady projektowania wskaźników	224
7.2.1. Użycie koncepcji SMART	224
7.2.2. Ustalanie punktów odniesienia i celów	230
7.2.3. Ćwiczenia	240
7.3. Unikanie pułapek związanych ze wskaźnikami	241
7.3.1. Reprezentacja	242
7.3.2. Wizualizacja	249
7.3.3. Ćwiczenia	252
Podsumowanie	254

CZĘŚĆ III. ZESTAW NARZĘDZI ANALITYKA 257

8. Zajmowanie się danymi wrażliwymi i chronionymi	259
8.1. Jednomyślność w badaniach	261
8.1.1. Krótka lekcja historii	262
8.1.2. Świadoma zgoda	263
8.1.3. Ćwiczenia	268
8.2. Obecny stan prawny	269
8.2.1. Regulacje dotyczące ochrony danych	269
8.2.2. Stronniczość w systemach zautomatyzowanych	274
8.2.3. Ćwiczenia	280
8.3. Analiza danych wrażliwych	281
8.3.1. Minimalizacja danych	282
8.3.2. Anonimizacja i pseudonimizacja danych	284
8.3.3. Zapobieganie deanonimizacji	290
8.3.4. Ćwiczenia	294
Podsumowanie	294
9. Świat modelowania statystycznego	296
9.1. Wiele obliczy modelowania statystycznego	297
9.1.1. Klasy modeli statystycznych	299
9.1.2. Dane wynikowe modelu i diagnostyka	307
9.1.3. Ćwiczenia	315
9.2. Proces modelowania	316
9.2.1. Analiza eksploracyjna	318
9.2.2. Dopasowanie modelu	324
9.2.3. Ćwiczenia	331
9.3. Model statystyczny i jego wartość	332
9.3.1. Modele objaśniające	333
9.3.2. Modele predykcyjne	336
9.3.3. Ćwiczenia	341
Podsumowanie	342

10.	<i>Uwzględnianie w analizach danych zewnętrznych</i>	345
10.1.	Korzystanie z interfejsów API	346
10.1.1	Pobieranie danych: porównanie interfejsów API z interfejsem przeglądarki	347
10.1.2.	Ustalanie wartości korzystania z interfejsu API	353
10.1.3.	Ćwiczenia	355
10.2.	Uzyskiwanie danych z witryn internetowych	356
10.2.1.	Uzyskiwanie danych z sieci	357
10.2.2.	Wyodrębnianie niezbędnych danych	360
10.2.3.	Określanie wartości metody uzyskiwania danych z witryn	365
10.2.4.	Ćwiczenia	367
10.3.	Korzystanie z publicznych źródeł danych	368
10.3.1.	Kiedy dane publiczne stały się tak popularne?	369
10.3.2.	Typy publicznych źródeł danych	371
10.3.3.	Dostęp do danych publicznych	376
10.3.4.	Ćwiczenia	379
	Podsumowanie	380
11.	<i>Magia danych o właściwej strukturze</i>	383
11.1.	Dylemat analityka	384
11.2.	Przegląd narzędzi do pracy z danymi	386
11.2.1.	Gromadzenie danych	387
11.2.2.	Przechowywanie danych	388
11.2.3.	Przetwarzanie danych	391
11.2.4.	Ćwiczenia	394
11.3.	Praktyki dotyczące zarządzania danymi	395
11.3.1.	Struktura danych	396
11.3.2.	Jakość danych	400
11.3.3.	Użycie metadanych	404
11.3.4.	Ćwiczenia	406
11.4.	Dane do analizy	409
11.4.1.	Inżynieria analityczna	410
11.4.2.	Modelowanie danych w schematach gwiazdy	413
11.4.3.	Ćwiczenia	417
	Podsumowanie	418
12.	<i>Narzędzia i technologie do nowoczesnej analizy danych</i>	421
12.1.	Ewolucja analityki danych	422
12.2.	Narzędzia do analizy i raportowania	424
12.2.1.	Typy narzędzi	424
12.2.2.	Biblioteki analiz	428
12.2.3.	Ćwiczenia	434

12.3. Analityka samoobsługowa	435
12.3.1. Metody zapewniania samoobsługi	435
12.3.2. Tworzenie strategii	437
12.3.3. Ćwiczenia	439
12.4. Sztuczna inteligencja	440
12.4.1. Generatywna sztuczna inteligencja	441
12.4.2. Kierunki na przyszłość	446
12.4.3. Końcowe przemyślenia	448
12.4.4. Ćwiczenia	448
Podsumowanie	449
 <i>Bibliografia</i>	 451



Sztuka używania wskaźników: monitorowanie wyników pod kątem sukcesu firmy

Zagadnienia omówione w rozdziale:

- Wartość wskaźników dla firmy.
- Identyfikowanie miar sukcesu do wykorzystania jako wskaźniki.
- Projektowanie wskaźników SMART do efektywnego monitorowania i podejmowania decyzji.
- Identyfikacja oraz minimalizowanie typowych pułapek i błędów podczas tworzenia wskaźników.
- Skuteczne informowanie interesariuszy o postępach i wnioskach.

Oglądając wiadomości lub czytając prasę, regularnie natrafiasz na wykresy śledzące informacje z upływem czasu. Być może jesteś zaznajomiony z wieloma z nich (notowania giełdowe, indeks jakości powietrza, stopa bezrobocia) i raczej rozumiesz, co oznacza wzrost lub spadek tych wartości. Wskaźniki te pozwalają zrozumieć świat i zachodzące w nim zmiany.

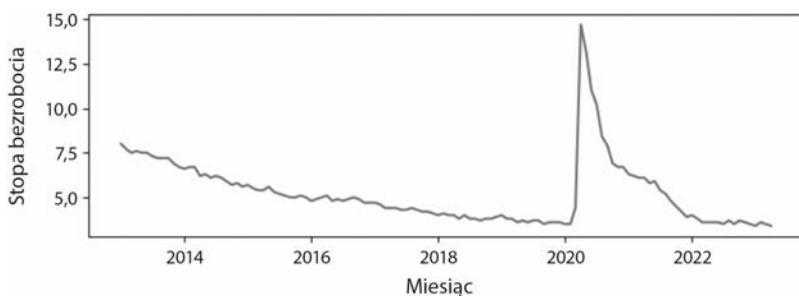
Wskaźniki to znormalizowane miary ilościowe, które monitorują w czasie procesy, wyniki lub działania. Te niezwykle cenne narzędzia pozwalają nam

zrozumieć zarówno natychmiastowe, jak i długoterminowe zmiany, które wpływają na niemal każdy aspekt naszego życia.

W tym rozdziale omówiono umiejętności niezbędne do tworzenia podobnie przejrzystych, praktycznych i skutecznych wskaźników (bez konieczności poświęcania na to całego stulecia!). Dowiesz się, jak identyfikować wartościowe procesy pod kątem monitorowania, przekładać je na wskaźniki, unikać typowych pułapek związanych z obliczeniami wskaźników oraz efektywnie dzielić się wynikami z interesariuszami. Opracowane przez Ciebie wskaźniki pozwolą Twojej firmie zrozumieć wpływ realizowanych działań, podejmować strategiczne decyzje oraz pojąć złożone procesy wpływające na jej funkcjonowanie.

7.1. Rola wskaźników w podejmowaniu decyzji

Wskaźniki są prezentowane niemal *wszystkim*, niezależnie od ich doświadczenia z danymi. Na przykład *stopa bezrobocia* to wskaźnik ekonomiczny powszechnie rozpoznawalny w wielu krajach, który jest definiowany jako *odsetek siły roboczej aktualnie niepracującej i aktywnie poszukującej zatrudnienia*. W Stanach Zjednoczonych Biuro Statystyki Pracy oblicza stopę bezrobocia co miesiąc, podając ją publicznie jako wskaźnik kondycji rynku pracy (rysunek 7.1). Ludzie używają tego wskaźnika do podejmowania kluczowych decyzji życiowych, takich jak poszukiwanie nowej pracy lub zakup domu. Pełne zrozumienie kwestii *bezrobocia* wymaga jednak analizy wielu różnych wskaźników, które dostarczają użytecznych informacji o stanie rynku pracy w danym momencie.



Rysunek 7.1. Oficjalna stopa bezrobocia raportowana co miesiąc w latach 2013 – 2023

Aby przedstawić pełny obraz bezrobocia w Stanach Zjednoczonych, Biuro Statystyk Pracy publikuje co miesiąc *sześć wskaźników*. Oficjalna stopa bezrobocia, która jest znana jako U-3, obejmuje *wszystkie* osoby w sile roboczej, które są bezrobotne, poszukują pracy i są gotowe do jej podjęcia. Pozostałe pięć wskaźników uwzględnia grupy populacji nieuwzględnione w ramach wskaźnika U-3 (np. wskaźnik U-4 obejmuje *zniechęconych* pracowników, którzy nie szukają aktywnie pracy, ponieważ uważają, że nie ma dla nich dostępnych miejsc pracy). Każdy z sześciu wskaźników zapewnia konkretne informacje, a wszystkie one są

niezbędne do zrozumienia uczestnictwa w rynku pracy. Wskaźniki bezrobocia Biura Statystyk Pracy istnieją w pewnej formie od prawie stu lat. Dekady badań pozwoliły na udoskonalenie metod obliczania i zbierania danych, co doprowadziło do opracowania sześciu wskaźników przedstawionych na rysunku 7.2.

	U-1	U-2	U-3	U-4	U-5	U-6
Osoby bez zatrudnienia od dłuższego czasu	✓		✓	✓	✓	✓
Osoby bez zatrudnienia od niedawna		✓	✓	✓	✓	✓
Osoby z zakończonym tymczasowym zatrudnieniem		✓	✓	✓	✓	✓
Osoby zniechęcone lub nieszukające już ofert				✓	✓	✓
Osoby czasami szukające ofert lub marginalnie nimi zainteresowane					✓	✓
Osoby zatrudnione w niepełnym wymiarze godzin						✓

Rysunek 7.2. Każdy wskaźnik bezrobocia uwzględnia różne typy osób bezrobotnych lub zatrudnionych w niepełnym wymiarze godzin

Wskaźniki zapewniają konkretny sposób oceny wydajności i trendów w czasie, stanowiąc niezbędne narzędzia do podejmowania decyzji, planowania strategicznego i wyznaczania celów. Często wskaźniki przedstawia się w formie wykresu liniowego lub słupkowego, gdzie oś y reprezentuje zagregowaną miarę, a oś x przedstawia zmianę tej miary w czasie (rysunek 7.3).



Rysunek 7.3. Wskaźniki są zazwyczaj przedstawiane jako zagregowane miary w czasie

Każda firma może skorzystać ze wskaźników, aby zrozumieć wpływ swoich działań i decyzji na wyniki. Gdy firma inwestuje czas w opracowywanie, śledzenie

i prezentowanie wskaźników interesariuszom, zyskuje przewagę konkurencyjną w zakresie podejmowania strategicznych decyzji. W tym podrozdziale zostaną omówione następujące trzy poziomy wskaźników, które firmy mogą monitorować:

- *Wskaźniki wydajności* dostarczają kluczowych informacji zwrotnych na temat stanu firmy. Są to podstawowe wskaźniki monitorowane w wielu typach organizacji, które ostatecznie pozwalają stwierdzić, czy firma odnosi sukces.
- *Wskaźniki strategii organizacyjnej* zapewniają informacje o wpływie konkretnych działań podejmowanych w firmie. Pozwalają one zespołom wejść o poziom głębiej niż w przypadku wskaźników wydajności i zrozumieć działania, które korelują z wynikami lub je przewidują.
- *Wskaźniki odpowiedzialności* czynią pojedyncze osoby i zespoły odpowiedzialnymi za osiągnięcie rezultatów. Często pokrywają się one ze wskaźnikami wydajności, ale są zaprojektowane tak, aby codzienne działania, procesy i sukcesy poszczególnych zespołów były jasne i przejrzyste dla całej firmy.

Firma, która dzięki wskaźnikom jest świadoma swojej wydajności, jest zwykle przygotowana do osiągania lepszych celów długoterminowych. Ustalenie odpowiednich wskaźników wymaga inwestycji w dogłębne zrozumienie danych. Firma, która *nie* inwestuje w pomiary i rozwój wskaźników, z łatwością może napotkać trudności w realizacji swoich długoterminowych celów.

Podejmując odpowiednie kroki, dowolnej wielkości firmy na każdym etapie rozwoju mogą nauczyć się kierować swoją strategią za pomocą właściwych wskaźników. Kluczem jest wybór miar, które są zgodne z ich celami i dostarczają praktyczne wnioski. Gdy wskaźniki te zostaną dobrze wdrożone, umożliwiają firmie pewne poruszanie się w złożonym środowisku decyzyjnym.

7.1.1. Monitorowanie wydajności

Firma powinna zazwyczaj zacząć od stosowania wskaźników do śledzenia miar, które wyraźnie wskazują na jej wyniki. Miary te najczęściej odpowiadają bezpośrednio kluczowym celom firmy, jej zadaniom i zdolności do dalszego odnoszenia sukcesów (np. przychody i zysk). Po ustaleniu tych podstawowych wskaźników firma może zagłębić się w dodatkowe wskaźniki sukcesu operacyjnego.

W mniejszych lub nowszych firmach Twoja rola jako analityka może polegać na opracowywaniu wskaźników organizacyjnych od podstaw. Większość organizacji skorzysta na ustanowieniu wskaźników wydajności w następującej kolejności:

- *Wskaźniki finansowe*. Są niezbędne do zrozumienia kondycji finansowej firmy. Większość firm będzie śledzić i monitorować przychody (tak jak na rysunku 7.4), marże zysku i koszty pozyskania klienta. Firmy nastawione na zysk dysponują dobrze ustalonymi zestawami wskaźników finansowych

w zależności od typu działalności biznesowej (np. model B2C) i metody uzyskiwania przychodów (np. comiesięczne subskrypcje) [1]. Organizacje non-profit i rządowe również śledzą wpływające fundusze i przychody w odpowiedni sposób. Często muszą jednak dostosować związane z tym obliczenia, aby odzwierciedlały sposób otrzymywania środków. Wskaźniki te są *kluczowe* dla inwestorów, członków zarządu i udziałowców bezpośrednio zainteresowanych wynikami finansowymi firmy.

- **Wskaźniki operacyjne.** Mierzą efektywność procesów operacyjnych organizacji. Mogą obejmować takie miary jak wielkość produkcji, czas niedostępności aplikacji lub poziomy akceptacji ofert zatrudnienia. Wskaźniki operacyjne są bardziej szczegółowe niż wiele wskaźników finansowych i często obliczane jako stosunek produktów lub usług przypadających na pracownika albo jako czas potrzebny na wykonanie zadania. Wskaźniki te są niezbędne w przypadku zespołów, które identyfikują obszary nieefektywności i wąskie gardła, co pozwala im zwiększyć produktywność i zmniejszyć koszty związane z codziennymi zadaniami.
- **Wskaźniki dotyczące klientów.** Koncentrują się na doświadczeniach i interakcjach klientów z firmą. Przykłady obejmują wskaźnik satysfakcji klienta (CSAT — *Customer Satisfaction Score*), wskaźniki rezygnacji klientów oraz wartość życiową klienta (LTV — *Lifetime Value*). Wskaźniki te dostarczają informacji o tym, jak dobrze firma spełnia potrzeby i oczekiwania swoich klientów. Często kierują one strategią organizacyjną, co zostanie omówione w dalszej części rozdziału.



Rysunek 7.4. Cykliczny przychód roczny lub cykliczny przychód miesięczny to standardowe wskaźniki finansowe stosowane w firmach opartych na modelu subskrypcyjnym

7.1.2. Kształtowanie strategii organizacyjnej

Po ustaleniu wskaźników wydajności firmy mogą dokładniej analizować konkretne obszary działalności i czynniki wpływające na sukces każdego z tych wskaźników. Mogą to być miary produktywności pracowników, efektywności procesów

i zachowań klientów. Zrozumienie tej kolejnej warstwy wskaźników i ich związków z wynikami firmy pozwala stworzyć strategię organizacyjną opartą na danych.

Wskaźniki wydajności zazwyczaj nie wyjaśniają przyczyny obserwowanej wartości. Jeśli przychody Twojej firmy spadną w danym miesiącu, Ty i Twoi interesariusze możecie jedynie snuć przypuszczenia co do wszelkich potencjalnych powodów zaistniałej zmiany. Wskaźniki *strategii organizacyjnej* umożliwiają analizę czynników wpływających na wyniki i podjęcie działań korzystnych dla firmy.

Istnieje kilka typów wskaźników, które mogą kształtować strategię organizacyjną i potencjalnie korelować ze wskaźnikami efektywności. W większości przypadków można je podzielić na następujące kategorie:

- **Wskaźniki dotyczące klientów.** Śledzą one zachowania, doświadczenia i opinie klientów Twojej firmy. Zrozumienie ich potrzeb i doświadczeń związanych z firmą dostarcza informacji niezbędnych do utrzymania i powiększenia bazy klientów. Przykładami takich wskaźników są: poziom zaangażowania lub aktywności, wskaźnik poleceń lub wysiłek klienta. Każdy z tych wskaźników może dostarczyć cennych informacji na temat tego, jak Twoja firma angażuje się w relacje z klientami.
- **Wskaźniki produktowe.** Są stosowane w firmach oferujących *produkt* lub *usługę*. Pozwalają zrozumieć zachowania użytkowników i sposób korzystania przez nich z produktu lub usługi, a ponadto mogą wpływać na strategię rozwoju produktu. Przykładami tych wskaźników są wskaźniki użytkowania, wskaźniki wdrożenia funkcji oraz czas potrzebny na realizację przepływów pracy.
- **Wskaźniki rynkowe.** Umożliwiają zrozumienie ekosystemu, w którym działa Twoja firma. Wskaźniki te obejmują zwykle dane gromadzone poza firmą, takie jak wiele wskaźników podaży, popytu i cen dostarczanych w Stanach Zjednoczonych przez Biuro Statystyki Pracy lub podobne organizacje rządowe na całym świecie. Dane te mogą być pomocne w podejmowaniu decyzji dotyczących sprzedaży nowym klientom, dostosowywania strategii cenowych lub inwestowania w reklamę.

Identyfikacja odpowiednich wskaźników dotyczących strategii, które przyczynią się do odniesienia sukcesu, może wymagać znacznego nakładu czasu. Jako analityk musisz formułować hipotezy, badać relacje ze wskaźnikami wydajności i rekomendować, które z nich śledzić. Ważne jest również regularne *ocenianie* i *udoskonalanie* wskaźników organizacyjnych, których znaczenie dla ogólnej wydajności może się zmieniać z upływem czasu. Można oczekiwać, że wskaźniki dotyczące strategii będą wymagały częstszych iteracji i korygowania niż wskaźniki wydajności.

Tam, gdzie jest to możliwe, ustalenie związku przyczynowego między używanymi wskaźnikami wydajności a wskaźnikami dotyczącymi strategii może znacząco

usprawnić proces podejmowania decyzji w Twojej firmie. Identyfikacja *przyczynowości* oznacza stwierdzenie, czy zmiany w inicjatywach strategicznych prowadzą bezpośrednio do zmian w wydajności. Wykazanie przyczynowości może być złożone. Napisano całe książki na temat metod wnioskowania przyczynowego, po które możesz sięgnąć, jeśli rozważasz pogłębienie swojej wiedzy w tym zakresie:

- W książce *Causal Inference: The Mixtape* autorstwa Scotta Cunninghama [2] omówiono wiele zaawansowanych podejść do ustalania przyczynowości w trakcie pracy ze złożonymi zbiorami danych. Książka zawiera przykłady kodu napisanego w języku R, a ponadto w wielu z nich wykorzystano dane ekonomiczne.
- Książka *Causality* autorstwa Judei Pearla [3] obejmuje szeroki zakres podejść do wnioskowania przyczynowego stosowanych w dziedzinach akademickich i pozaakademickich.

Jeśli masz ograniczone możliwości użycia zaawansowanych metod statystycznych, to wiedz, że nawet proste podejścia mogą dostarczyć cennych informacji:

- Porównywanie wskaźników wydajności *przed i po* działaniu lub zmianie ujętej w zastosowanych wskaźnikach dotyczących strategii (np. przeprowadzenie analizy regresji z wykorzystaniem tych wskaźników w roli predyktorów i wskaźników wydajności jako danych wynikowych, a następnie wizualizacja ich trendów w sposób pokazany na rysunku 7.5).
- Porównywanie wskaźników wydajności *między grupami*, które zostały i nie zostały dotknięte efektem działania lub zmiany (np. uwzględnienie wskaźników grupowych jako predyktorów w modelu regresji i przeprowadzenie testu *t* lub analizy wariancji).



Rysunek 7.5. Wizualne przedstawienie silnych relacji między wskaźnikami wydajności a wskaźnikami dotyczącymi strategii może pomóc interesariuszom zrozumieć długoterminowe trendy oraz to, czy z upływem czasu dana relacja wzmacnia się czy słabnie

Modele regresji zapewniają dodatkową rzetelność dokonanej analizy, ale pamiętaj, że korelacja *nie* oznacza przyczynowości. Na wyniki organizacji wpływa wiele czynników. Jako analityk będziesz nieustannie starał się zrozumieć relacje, których jeszcze *nie* zmierzyłeś. Z czasem zbudujesz kompleksowy obraz procesów i miar, które składają się na środowisko Twojej firmy.

7.1.3. Promowanie odpowiedzialności

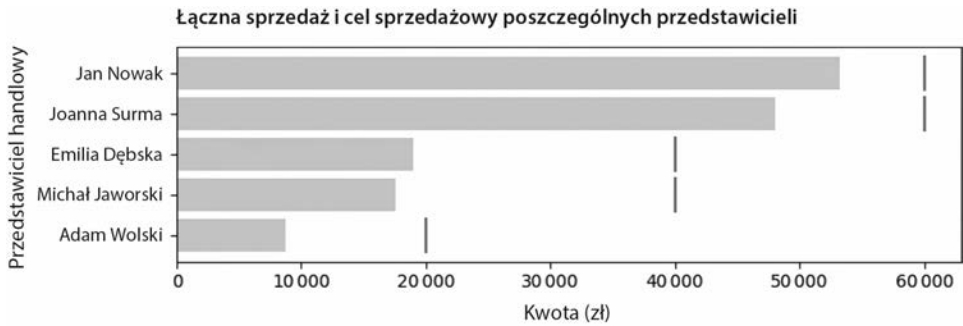
Wskaźniki często służą do mierzenia wydajności poszczególnych osób i zespołów w kontekście zapewnienia firmie sukcesu. Te *wskaźniki odpowiedzialności*, które często są nazywane kluczowymi wskaźnikami efektywności (KPI — *Key Performance Indicator*), oferują namacalny sposób ustalania oczekiwań na wielu poziomach i mierzenia postępów w realizacji celów i obowiązków. Firmy często używają tych wskaźników do ułatwienia oceny wyników, podejmowania decyzji o awansach, przyznawania premii i opracowywania strategii rozwoju zespołu.

W przypadku pojedynczych osób i zespołów wskaźniki te są przeważnie powiązane z terminowym i dokładnym wykonywaniem zadań i projektów. Na przykład zespół obsługi klienta może śledzić procent zgłoszeń klientów zrealizowanych w ciągu 24 godzin. Aby zrozumieć wyniki każdej osoby, menedżer zespołu prawdopodobnie przeanalizuje ogólną wartość wskaźnika dla zespołu, a następnie dokona rozbicia na poszczególnych przedstawicieli handlowych. Informacje te pokazują, którzy członkowie zespołu osiągają cele, a którzy mogą potrzebować dodatkowego wsparcia, szkolenia lub zmiany przydzielonych zadań.

Wskaźniki odpowiedzialności w firmie dla poszczególnych zespołów mogą wyglądać tak:

- *Procent realizacji wyznaczonej wartości sprzedaży* przez przedstawicieli handlowych w zespole sprzedaży mierzony co kwartał (rysunek 7.6).
- *Średni czas obsadzania wolnych stanowisk* przez poszczególne osoby rekrutujące w zespole pozyskiwania talentów (mierzone miesiąc do miesiąca).
- *Czas usuwania błędów* (efekty prac wykazujące błędy lub problemy w witrynie lub aplikacji) w zespole inżynierów oprogramowania (mierzone co kwartał).
- *Liczba kwalifikowanych leadów* (potencjalni klienci zainteresowani ewentualnym zakupem produktu lub usług) generowanych w ramach kampanii marketingowych (mierzone co sześć miesięcy).

W firmach wskaźniki odpowiedzialności są częścią szerszych celów strategicznych, takich jak osiągnięcie określonej marży zysku, redukcja kosztów ogólnych lub poprawa doświadczeń klientów. Wskaźniki te ukierunkowują wysiłki różnych zespołów na ten sam cel i pomagają zapewnić, że firma pozostaje skupiona na swoich strategicznych celach.



Rysunek 7.6. Wykres przedstawiający wyniki przedstawicieli handlowych w porównaniu z ich celami sprzedażowymi zależnie od stażu pracy w firmie

Podobnie jak w przypadku wskaźników strategii organizacyjnej, wskaźniki odpowiedzialności powinny być regularnie weryfikowane w celu upewnienia się, że są odpowiednio dopasowane do ogólnych wskaźników wydajności. W miarę rozwoju i zmian w firmie wskaźniki, według których oceniane są zespoły, powinny ewoluować tak, by odzwierciedlały ich najważniejszą część pracy. Ostatecznie, przemyślane podejście do kwestii pomiaru wykonywanej pracy w związku z odpowiedzialnością może sprzyjać poprawie wydajności i kształtowaniu kultury przejrzystości oraz ciągłemu doskonaleniu.

7.1.4. Ćwiczenia

Jesteś analitykiem zaczynającym nową pracę na stanowisku wspierającym pracę działu marketingu firmy. Twój zespół dysponuje już pewnymi panelami kontrolnymi i dokumentacją dotyczącą wskaźników przedstawionych w tabeli 7.1, które są obecnie śledzone lub wcześniej były uznawane za kluczowe dla firmy.

Tabela 7.1. Wskaźniki firmowe w prezentacjach i na pulpitach nawigacyjnych

Wskaźniki w prezentacjach	Wskaźniki na panelach kontrolnych
Liczba odwiedzających witrynę	Średni czas spędzony na stronie (podany w minutach)
Liczba pozyskanych leadów	Wartość życiowa klienta (LTV)
Liczba uczestników webinarów	Koszt pozyskania klienta (CAC)
Koszt pozyskania leada	Wzrost przychodów (odsetek wzrostu)
Współczynnik otwarcia wiadomości e-mail (odsetek wiadomości wyświetlonych przez klientów)	Współczynnik konwersji (odsetek leadów, które stają się klientami)
Zwrot z wydatków na reklamę	Wskaźnik zaangażowania w mediach społecznościowych (liczba wyświetleń, udostępnień i kliknięć poszczególnych postów)
Współczynnik spamu (odsetek wiadomości e-mail oznaczonych jako spam przez klienta)	Wskaźnik terminowego uruchamiania kampanii

1. Używając listy wskaźników z tabeli 7.1, określ, które z nich mogą należeć do trzech omówionych kategorii wskaźników: *wydajności*, *operacyjnych* i *odpowiedzialności*. Wskaźniki mogą należeć do więcej niż jednej kategorii.
2. W przypadku wskaźników najbardziej odpowiednich w roli wskaźników operacyjnych lub odpowiedzialności, które z nich mogą wymagać osobnego pomiaru dla poszczególnych zespołów zamiast dla całej firmy marketingowej?
3. Zaproponuj hipotezę dotyczącą tego, które wskaźniki operacyjne mogą wpływać na ogólne wskaźniki wydajności. W jaki sposób byś to przetestował?

7.2. Kluczowe zasady projektowania wskaźników

Projektowanie wskaźników to kluczowa umiejętność analityka. Wymaga ona dogłębnego zrozumienia celów i zadań firmy oraz kompleksowej wiedzy o procesach i działaniach, które są mierzone. W niemal każdej organizacji można zastosować podstawowe zasady projektowania wskaźników, aby wybrać odpowiednie i praktyczne cele do monitorowania i oceny sukcesu. Przyjrzymy się bliżej koncepcji SMART stosowanej do projektowania wskaźników. Zagwarantuje ona możliwość powiązania bardziej szczegółowych wskaźników z celami organizacyjnymi i ustalenia właściwych celów z myślą o zapewnieniu kontekstu do oceny poziomu sukcesu.

7.2.1. Użycie koncepcji SMART

Koncepcja SMART (*Specific, Measurable, Achievable, Relevant, Time-bound*, czyli konkretne, umożliwiające pomiar, osiągalne, istotne i określone w czasie) to skuteczne i szeroko stosowane podejście do projektowania wskaźników pozwalających określić cele w firmie możliwe do zmierzenia. SMART to akronim obejmujący kluczowe cechy niezbędne do efektywnego śledzenia wyników. Stosując tę koncepcję, analityk może opracować właściwie zdefiniowane i bardzo skuteczne wskaźniki, które pomogą firmie w zwiększaniu wydajności i podejmowaniu decyzji.

KONKRETNE

Wskaźnik powinien jasno i zwięźle reprezentować *konkretny* aspekt wydajności lub celu Twojej firmy. Należy unikać niejasnych i wieloznacznych sformułowań w definicjach, aby wszyscy interesariusze rozumieli, co jest mierzone. Wskaźnik powinien określać, co trzeba osiągnąć, czym objawia się sukces i jakie są oczekiwane rezultaty. W tabeli 7.2 pokazano różnicę między ogólnym pojęciem a konkretnym językiem używanym w definicji wskaźnika.

Tabela 7.2. Wiele wskaźników jest potocznie omawianych jako koncepcje, dlatego wymagają one współpracy interesariuszy w celu stworzenia konkretnych definicji operacyjnych

Ogólna koncepcja	Konkretny wskaźnik
Zespół inżynierów jest produktywny.	Odsetek zadań ukończonych na czas.
Klienci otrzymują wsparcie odpowiadające ich potrzebom.	Odsetek problemów i pytań klientów rozwiązanych przy pierwszym kontakcie (FCR — <i>First Call Resolution</i>).
Tworzenie łatwej w obsłudze aplikacji.	Mediana czasu niezbędnego nowym użytkownikom na ukończenie procesu konfiguracji.

Jeśli na przykład interesariusze poproszą Cię o zmierzenie poziomu satysfakcji klienta, będziesz musiał współpracować z nimi, aby określić *definicję operacyjną* tego pojęcia. Wybrany wskaźnik powinien być *wiarygodną i trafną* miarą ilościową (zajrzyj do rozdziałów 2. i 6.) uzgodnioną przez zespoły, które zamierzają użyć go do podejmowania strategicznych decyzji. Przy wykorzystaniu danych przekazanych przez samych klientów Twoim konkretnym wskaźnikiem mógłby być odsetek klientów zgłaszających, że w trakcie użytkowania byli „Zadowoleni” lub „Bardzo zadowoleni”. Taka precyzyjna definicja operacyjna przyspieszy proces uzgadniania z interesariuszami.

UMOŻLIWIAJĄCE POMIAR

Wskaźnik musi umożliwiać ujęcie ilościowe i obiektywny *pomiar*. Gdy Ty i interesariusze uzgodnicie ilościowy i obiektywny pomiar, Twoim zadaniem jako analityka jest zapewnienie, że Twoja firma będzie rzetelnie i dokładnie gromadzić dane z upływem czasu. W praktyce wymaga to współpracy z wieloma zespołami w celu zadbania o to, by dane spełniały następujące kryteria:

- *Dane są zbierane w regularnych odstępach czasu.* Jeśli na przykład korzystasz z pytań z ankiety poziomu satysfakcji klienta przeprowadzanej raz na kwartał, powinieneś oczekiwać podobnej liczby odpowiedzi rejestrowanych w każdym takim okresie.
- *Firma jest zaangażowana w ciągłe inwestowanie w zbieranie danych.* Wskaźniki są gromadzone w średnich lub długich okresach (np. co miesiąc przez rok lub dłużej). W związku z tym Twoja firma będzie musiała zbierać dane przez minimalny uzgodniony okres (np. dwa lata) bez dokonywania zmian w podstawowych danych lub częstotliwości ich gromadzenia.
- *Dane mają spójne, zdefiniowane jednostki miary, na których można polegać przy agregacji danych.* Na przykład formaty skal odpowiedzi w ankiecie poziomu satysfakcji klienta powinny pozostać niezmiennie. Ponadto tabele w hurtowni danych zawierające odpowiedzi powinny być modyfikowane tylko wtedy, gdy jest to konieczne i z odpowiednim wyprzedzeniem, które umożliwi dostosowanie zapytań i paneli kontrolnych.

- *Dane są wysokiej jakości i dokładne.* Ty i interesariusze powinniście móc ufać temu, że wartości są właściwą reprezentacją zebranych informacji. W przypadku danych kiepskiej jakości (np. nieodpowiednich typów danych albo nieprawidłowych lub zduplikowanych rekordów; rysunek 7.7) należy podjąć odpowiednie kroki w celu usunięcia problemów, które mogłyby wpłynąć na wartości używanych wskaźników.

	ticket_id	customer_id	category	priority	status	created_date
0	1527	cid_901	General	Low	Open	2023-06-10
1	1643	cid_986	General	Low	Closed	2023-05-17
2	3390	cid_569	General	High	Open	1023-01-01
3	4389	cid_754	Billing	Low	Resolved	2023-05-28
4	4389	cid_886	Product Inquiry	High	Open	2023-06-07

Rysunek 7.7. Ten zbiór danych zawiera zduplikowane klucze główne (kolumna `ticket_id`) oraz nieprawidłową wartość daty, co może obniżyć jakość wszelkich tworzonych wskaźników

OSIĄGALNE

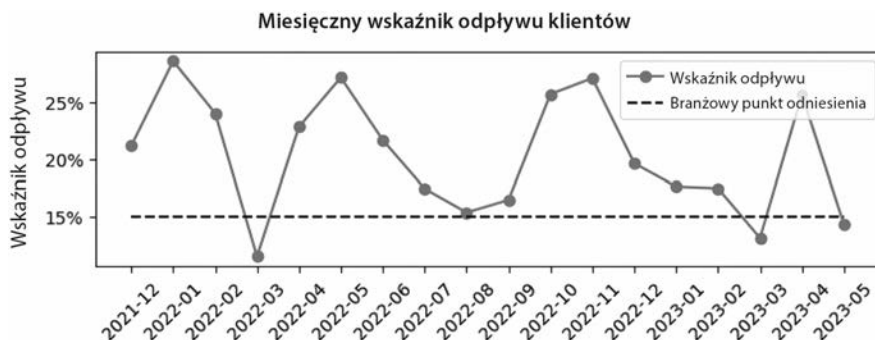
Cele firmy wyrażone za pomocą wskaźników powinny być *osiągalne* przy dostępnych zasobach. Wskaźniki wydajności i odpowiedzialności organizacyjnej mogą być bardzo motywujące, *jeśli* członkowie zespołów wierzą, że są w stanie je osiągnąć. Zbyt ambitny lub nieosiągalny cel może być frustrujący i demotywujący, prowadząc do wypalenia i obniżenia wydajności pracowników i zespołów. Z kolei wyznaczanie zbyt łatwych celów raczej nie przyniesie znaczącej poprawy.

Ustalanie osiągalnych i realistycznych celów wymaga równowagi i ciągłej iteracji. Ty i Twój interesariusze z łatwością możecie zauważyć, że wskaźnik „100% utrzymania klientów” jest nierealistyczny. Określenie odpowiedniego celu wymaga jednak głębszego zrozumienia Twoich klientów i ich potrzeb. Jeżeli nie masz konkretnego pojęcia o działaniach, które wpływają na wskaźnik rezygnacji klientów, mało prawdopodobne jest, że będziesz w stanie na niego wpłynąć.

Aby wyznaczyć odpowiednie i osiągalne cele, przeważnie niezbędne są następujące informacje:

- Długoterminowe trendy w przypadku proponowanego wskaźnika w dłuższych okresach (najlepiej co najmniej dwóch lat, aby móc porównać zmiany sezonowe).
- Porównania z punktem odniesienia proponowanego wskaźnika dotyczące firm o podobnych cechach (np. z tej samej branży lub o podobnej wielkości, lokalizacji geograficznej i typie usługi). Takie porównania są cenne z punktu widzenia zrozumienia zakresu możliwych do osiągnięcia celów dla interesującego nas wskaźnika. Rysunek 7.8 prezentuje przykład tego, jak w firmie można dokonać wizualizacji wskaźnika i wartości referencyjnych.

- Czynniki, które korelują lub pozwalają przewidzieć wartość proponowanego wskaźnika. Jeśli chcesz wyznaczyć cele służące poprawie wartości wskaźnika, Twoja firma musi być świadoma tego, jakie działania strategiczne umożliwią osiągnięcie tych celów.
- Informacje o kluczowych czynnikach zewnętrznych będących poza Twoją kontrolą, które wpływają na Twoją firmę, jej główne wskaźniki i wyznaczane cele. Na przykład cele produkcyjne w firmie odzieżowej będą zwykle obniżane w przypadku niedoboru niezbędnych materiałów.



Rysunek 7.8. Wykres przedstawia wskaźnik odpływu klientów firmy w porównaniu z branżowymi punktami odniesienia. W tym przypadku wskaźnik odpływu będzie raczej wyższy niż punkt odniesienia, co sugeruje, że może istnieć przestrzeń do poprawy wskaźnika utrzymania klientów

W punkcie 7.2.2 zostaną omówione strategie ustalania punktów odniesienia i identyfikowania benchmarków niezbędnych do wyznaczania osiągalnych celów.

ISTOTNE

Inwestycje wielofunkcyjne są niezbędne do opracowania wskaźników, informowania o nich i wykorzystywania ich do ciągłego monitorowania postępów. Aby firma mogła uzyskać wartość w wyniku tych działań, wskaźnik powinien być zgodny z jej strategicznymi celami i priorytetami. Powinien bezpośrednio odnosić się do obszaru objętego zakresem pomiaru i dostarczać znaczących informacji, które przyczyniają się do podejmowania decyzji i poprawy wyników firmy.

Organizacje mogą z łatwością określić istotność swoich *wskaźników wydajności*. Wskaźniki te są powiązane z ostatecznymi wynikami (np. w postaci stanu finansów, efektywności i pozyskanych klientów). Identyfikacja najbardziej odpowiednich wskaźników *odpowiedzialności* i wskaźników dotyczących *strategii* wymaga dodatkowej analizy w celu zrozumienia związku między tymi wskaźnikami a ogólnymi wynikami w ramach prowadzonej działalności.

Na przykład zespół inżynierów chce zrozumieć swoją skuteczność we wspieraniu celów firmy. Do zmierzenia produktywności jej kierownictwo proponuje analizę *całkowitej liczby wierszy kodu* utworzonych przez każdego inżyniera w ciągu kwartału. Choć można to łatwo zmierzyć, a zespół może ustalić cele na

podstawie tego wskaźnika, jest mało prawdopodobne, aby przyniosło to firmie znaczące rezultaty. Liczba napisanych wierszy kodu *nie* pozwala zmierzyć bezpośrednio wydajności, użyteczności ani wartości oprogramowania dla użytkownika, a ponadto nie dostarcza praktycznych wniosków na potrzeby rozwoju produktu. Zamiast tego zespół inżynierów prawdopodobnie skorzysta bardziej na skupieniu się na wskaźnikach, które odzwierciedlają dobrze zdefiniowane miary sukcesu dla korzystających z nich użytkowników.

Przykłady w tabeli 7.3 pokazują, że miara może być konkretna, możliwa do zmierzenia i osiągalna, a mimo to *nieistotna* z punktu widzenia celów firmy. Często nieodpowiednie wskaźniki są wybierane ze względu na ich wygodę, dostępność i pozorne powiązanie z działaniami podejmowanymi przez pojedyncze osoby i zespoły. Może to skutkować miesiącami zmarnowanego wysiłku w dążeniu do uzyskania celów, które nie przynoszą pożądanego efektu dla firmy. Należy zachować ostrożność i upewnić się, że przed przejściem do ostatniego etapu dany wskaźnik spełnia *wszystkie* omówione do tej pory kryteria koncepcji SMART!

Tabela 7.3. Bądź gotów do analizy trafności stosowanych wskaźników i upewnij się, że są one powiązane z sukcesem Twojego zespołu i firmy

Dziedzina	Nieistotne	Istotne
Inżynieria	Liczba wierszy kodu	Liczba błędów przypadających na wiersz kodu
Produkt	Liczba wdrożonych opcji	Wskaźnik utrzymania klientów
Marketing	Liczba osób obserwujących w mediach społecznościowych	Koszt pozyskania klienta
Finanse	Wskaźnik absencji pracowników	Marża zysku brutto

OKREŚLONE W CZASIE

Wskaźnik powinien mieć *określony okres* pomiaru i ustalania celów. Powinien być powiązany z konkretnym przedziałem czasu, takim jak dzień, tydzień, miesiąc lub kwartał. Pozwala to zespołom i firmom monitorować i porównywać wskaźniki w celu określenia postępów. Ustalenie wyraźnych okresów dla celów wskaźników (np. wyniki raportowane na koniec kwartału) tworzy poczucie pilności. Pomaga to zespołom ustalać priorytety zadań i przydzielać zasoby na podstawie postępów w realizacji celu.

Ustalanie wskaźników z określonymi ramami czasowymi wymaga zrozumienia okresu, w jakim zachodzą zmiany w badanym zachowaniu. Jako analityk powinieneś uwzględnić szereg czynników związanych z danymi szeregów czasowych użytego wskaźnika:

- *Czas niezbędny do pełnego sprawdzenia danego zachowania lub procesu wpływa na okres, jaki powinieneś wybrać dla swojego wskaźnika.* Jeśli stworzysz wskaźnik dla czasu ukończenia przepływu pracy, który trwa średnio dwa miesiące, w celu zauważenia znaczących zmian

prawdopodobnie będziesz musiał agregować dane w ujęciu kwartalnym, a nie tygodniowym lub miesięcznym.

- *Przy wyborze okresu należy wziąć pod uwagę oczekiwany czas trwania inicjatyw mających wpłynąć na cele wskaźników.* Jeśli przeprowadzasz test A/B o przewidywanym czasie trwania dwóch tygodni, wskaźnik kwartalny może nie mieć odpowiedniej szczegółowości, aby odzwierciedlić rzeczywiste efekty testu.
- *Wiele procesów podlega sezonowym wahaniom, które powodują cykliczne zmiany w skali dziennej, tygodniowej lub miesięcznej.* Aby ustalić cele dające podstawy do działania, trzeba rozważyć okres agregacji, który wyeliminuje kwestię sezonowości.
- *Oprócz wzorców sezonowych wskaźniki różnią się stabilnością w różnych okresach.* Aby wyznaczyć cele, które można wiarygodnie osiągnąć, trzeba znaleźć przedział czasowy, który zminimalizuje niestabilność bez eliminowania przy tym zmienności. Widać to wyraźnie na rysunku 7.8, w przypadku którego wskaźnik rezygnacji klientów wykazuje dużą, losową zmienność utrudniającą wyznaczenie celów dla pojedynczego miesiąca.

Choć wiele problemów można złagodzić za pomocą dłuższego przedziału czasowego (np. miesięcznego zamiast tygodniowego), może to spowodować ukrycie efektów działań podjętych w celu wpłynięcia na dany wskaźnik. Inicjatywa rozpoczęta w trzecim miesiącu kwartału będzie mieć minimalny wpływ na wskaźnik tego kwartału, a efekty mogą być widoczne dopiero w *kolejnym* kwartale. Najlepszy okres do wybrania będzie zależeć od procesu, potrzeb Twojej firmy i charakterystyki danych. W tabeli 7.4 przedstawiono przykłady odpowiednich okresów dla różnych wskaźników.

Tabela 7.4. Zalecane kwestie do rozważenia przy wyborze okresu pod kątem agregowania wskaźników

Przedział czasowy	Widoczna sezonowość	Kiedy stosować?	Kiedy unikać?
Dzienny	Tygodniowa, miesięczna	Śledzenie krótkoterminowe, procesy wymagające szybkiej reakcji	Śledzenie długoterminowe
Tygodniowy	Kwartałna, roczna	Śledzenie krótko- i średnioterminowe, gdy dane dzienne zawierają zbyt dużo zbędnych danych	Śledzenie długoterminowe, śledzenie krótkoterminowe, gdy dzienne wahania są znaczące
Miesięczny	Roczna, wieloletnia	Śledzenie i prognozowanie średnio- i długoterminowe	Śledzenie krótkoterminowe i wieloletnie (powyżej 3 lat)
Kwartalny	Wieloletnia	Śledzenie i prognozowanie długoterminowe	Śledzenie krótkoterminowe

Tworzenie wskaźników SMART w firmie tworzącej oprogramowanie

W studium przypadku z tego rozdziału Artur jest analitykiem sprzedaży w szybko rozwijającej się firmie z branży SaaS (oprogramowanie jako usługa). Zespół ds. sprzedaży chce poszerzyć bazę potencjalnych klientów, ale ma trudności z określeniem, które działania marketingowe są najbardziej skuteczne. Artur otrzymuje zadanie zbadania i zaproponowania wskaźnika, który pomoże rozwiązać ten problem. Postanawia zastosować koncepcję SMART do oceny potencjalnych wskaźników.

- *Konkretne.* Artur stwierdza, że zespół ds. sprzedaży potrzebuje wskaźnika, który odzwierciedla wpływ działań marketingowych konkretnie na procesy sprzedażowe. W związku z tym postanawia się skupić na tym, czy leady pozyskane przez ten zespół są *kwalfikowane* (oznacza to prawdopodobieństwo zakupu oprogramowania). Proponuje wskaźnik *Odsetek leadów kwalifikowanych jako sprzedażowe*.
- *Umożliwiające pomiar.* Artur ustala, że wskaźnik ten można zmierzyć w sposób ilościowy, śledząc liczbę leadów pochodzących z każdego źródła marketingowego, które zespół ds. sprzedaży uznał za bardzo wartościowe (kwalfikowane leady sprzedażowe). Dane te są dostępne w systemie firmy rejestrującym źródło leadów.
- *Osiągalne.* Artur proponuje wprowadzenie nowego wskaźnika z wykorzystaniem istniejących procesów śledzenia leadów. Jedyną niezbędną zmianą byłoby zapewnienie, że zespół ds. sprzedaży rejestruje w systemie czynniki decydujące o tym, czy lead jest *kwalfikowany*. Zauważa, że zespół można łatwo przeszkolić, ale dane są wprowadzane ręcznie i mogą wymagać monitorowania pod kątem jakości.
- *Istotne.* Artur jest przekonany, że proponowany wskaźnik stanowi bezpośrednią odpowiedź na wyzwanie zespołu ds. sprzedaży dotyczące określenia skuteczności działań marketingowych. Może on dostarczyć informacji o tym, skąd pochodzą najlepsze leady, umożliwiając bardziej ukierunkowany i efektywny marketing.
- *Określone w czasie.* Artur proponuje śledzenie wskaźnika co miesiąc przez trzy miesiące w celu zebrania wystarczającej ilości danych do wstępnej analizy. Wie, że wszystkie wskaźniki sprzedaży w firmie są oceniane w cyklach miesięcznych lub kwartalnych, a liczba transakcji w skali tygodnia jest zbyt mała, by krótszy okres analizy okazał się wartościowy. Po tym wstępnym okresie Artur proponuje ocenę skuteczności wskaźnika wraz z zespołem marketingowym i wprowadzenie wszelkich niezbędnych korekt.

PODSUMOWANIE

Projektowanie skutecznych wskaźników wymaga strategicznego podejścia do celów firmy, jej wyników i definicji sukcesu. Zastosowanie przejrzystych i związanych ram, takich jak koncepcja SMART, zapewnia, że wybrane wskaźniki są zgodne z tymi celami, a ponadto mogą być objęte pomiarem oraz są osiągalne i istotne w dłuższej perspektywie. Ogólnie rzecz biorąc, koncepcja SMART oferuje przejrzystą strukturę na potrzeby badań i opracowywania wskaźników, co w przeciwnym razie może okazać się niejednoznaczne i utrudnione.

7.2.2. Ustalanie punktów odniesienia i celów

Tworzenie w firmie wskaźników SMART wymaga wielu podstawowych informacji, aby móc wyznaczyć cele i podjąć działania zmierzające do ich osiągnięcia. Po zidentyfikowaniu wskaźnika konieczne jest ustalenie *obecnego stanu punktów*

odniesienia do zrozumienia poziomów wydajności przed dokonaniem jakichkolwiek zmian lub interwencji. Służy to za punkt wyjścia do podejmowania strategicznych decyzji. Określenie odpowiednich wartości bazowych dla wskaźników wymaga dokładnej analizy danych historycznych, recenzowanych badań i standardów branżowych (jeśli są dostępne).

Analiza przeprowadzona podczas tworzenia punktów odniesienia dla wskaźników nadaje strukturę i przejrzystość działaniom mającym na celu poprawę wydajności. Pozwala to interesariuszom w spójny sposób zrozumieć wcześniejsze trendy, zachowania użytkowników i procesy zewnętrzne związane z celami Twojej firmy. Z tej perspektywy bazującej na danych zespoły mogą łatwiej uzgodnić, co jest możliwe do osiągnięcia i jakim nakładem pracy.

ANALIZUJ DANE HISTORYCZNE

Po zaproponowaniu *konkretnego* wskaźnika *umożliwiającego pomiar* rolę analityka jest kompleksowe zbadanie danych historycznych dotyczących podstawowej miary. Pozwala to zapewnić wiedzę ekspercką na temat reprezentowanego zachowania i zaproponować strategiczne metody poprawy wartości wskaźnika. Uzyskane w ten sposób wnioski mogą pomóc w określeniu, czy wyznaczone cele są *osiągalne*, a także tego, jaka *agregacja czasowa* będzie najbardziej odpowiednia i pozostanie użyteczna w dłuższej perspektywie.

Analiza danych historycznych dla proponowanego wskaźnika wymaga zbadania ich z *wielu* różnych perspektyw. Obejmuje to badanie kształtu, trendów, anomalii, segmentów, danych szeregów czasowych oraz korelacji z innymi znanymi miarami i wskaźnikami. Żadne dwie analizy nie są identyczne, ponieważ wyniki prowadzą do coraz bardziej szczegółowych badań. Zalecam, aby podczas analizy danych wskaźników historycznych rozważyć co najmniej następujące kwestie:

- *Jaki jest kształt danych?* Aby dokonać wizualizacji rozkładu, przyjrzyj się danym za pomocą histogramu lub wykresu pudełkowego. Czy dane są skośne, czy mają rozkład normalny, czy nie dotyczy ich żadna z tych rzeczy? Jak bardzo dane są skupione wokół mediany (określa się to też mianem *kurtozy*)? Cechy te ujawnią, co jest typowe i możliwe w kontekście wartości wskaźników.
- *Czy w danych występują wartości odstające?* Jeśli tak, w jakim odbywa się to kierunku? Czy są one na tyle skrajne, że zniekształcają średnią zbioru danych zagregowanych dla danego okresu? Nietypowe rekordy w rozkładzie wskazują na podgrupy użytkowników lub klientów, którzy mogą „wypadać przez szczeliny” albo są superużytkownikami produktu lub usługi.
- *Czy istnieją różnice między segmentami?* Zastanów się, jak Twoja firma dzieli swoich klientów i użytkowników (np. według wielkości firmy, regionu geograficznego lub grupy wiekowej). Istniejący w firmie podział bazy

użytkowników to znakomity punkt wyjścia do badania różnic w nowych wskaźnikach. Spróbuj porównać miary tendencji centralnej i nakładające się rozkłady między segmentami, aby zrozumieć, gdzie występują znaczące różnice i czy potwierdzają one istniejące hipotezy dotyczące tych segmentów.

- *Jak dane zmieniają się w czasie?* Wartości wskaźników rzadko pozostają niezmiennie w długich przedziałach czasowych, w których zwiększają się i ewoluują. Spróbuj podzielić dane na lata lub inne znaczące okresy. Czy dostrzegasz jakieś nowe trendy?
- *Jak wartości wskaźników zmieniają się w trakcie współpracy klienta lub użytkownika z Twoją firmą?* Jakie trendy zauważasz, jeśli zaczniesz analizować dane od momentu, gdy dana osoba dołączyła do Twojej aplikacji, zakupiła produkt lub usługę albo rozpoczęła współpracę z Twoją firmą? Czy na przykład dla nowych klientów występują niższe wartości wskaźnika niż dla lojalnych i długoletnich klientów? Czy widać wyraźne trendy w ciągu pierwszych 7, 30, 60 lub 90 dni współpracy z klientem?
- *Czy wartości wskaźników są powiązane z którymkolwiek z poniższych elementów?*
 - pora roku (numer miesiąca, numer tygodnia) wskazująca na sezonowość,
 - długość stażu klienta wskazująca na wzrost lub spadek zależnie od czasu współpracy z firmą,
 - wskaźniki aktywności użytkownika (np. liczba dni aktywności w tygodniu) oznaczające powiązanie z co najmniej jedną inną formą aktywności oraz wariantami użytkowania.

Zaimportujmy zbiór danych dotyczący aktywności klientów, aby przekonać się, jak za pomocą tych kroków można sprawdzić wskaźnik. Każdy wiersz w zbiorze zawiera liczbę dni w miesiącu, w których klient się zalogował. Dostępne są też informacje o regionie każdego klienta jako potencjalnym segmencie:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

← Importowanie bibliotek pandas i seaborn.

```
logins = pd.read_csv("customer_logins.csv")
logins["month"] = pd.to_datetime(logins["month"], format="%Y-%m-%d")
print(logins.head())
```

← Wczytanie zbioru danych dotyczących wskaźnika aktywności klientów.

← Wyświetlenie pięciu pierwszych wierszy zbioru danych.

← Przekształcenie w celu późniejszego użycia kolumny miesiący w format daty i godziny.

Oto wynikowe dane wyjściowe:

	customer_id	region	month	login_days
0	93	Europe	2020-01-01	21
1	346	Europe	2020-01-01	12

2	404	Asia	2020-01-01	22
3	347	North America	2020-01-01	15
4	403	Asia	2020-01-01	30

Można zacząć analizować wzorce w danych wskaźnika od wygenerowania nowej kolumny reprezentującej *procent* dni w miesiącu, w których klient się zalogował. Pozwoli to znormalizować niewielkie różnice w liczbie dni poszczególnych miesięcy. Można następnie utworzyć histogram wartości procentowych zarejestrowanych *dla każdego klienta w każdym miesiącu*:

```
logins["n_days"] = logins["month"].dt.days_in_month
logins["login_days_pct"] = logins["login_days"]/logins["n_days"]

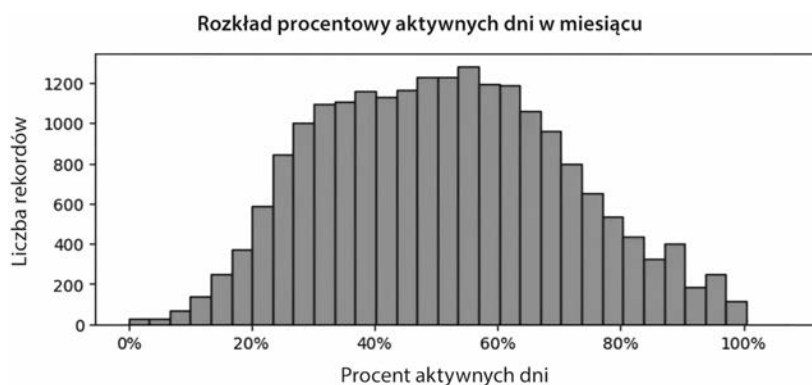
sns.histplot(x=logins["login_days_pct"], bins=32)
plt.title("Rozkład procentowy aktywnych dni w miesiącu")
plt.xlabel("Procent aktywnych dni")
plt.ylabel("Liczba rekordów")
```

Ustalenie liczby dni w każdym miesiącu.

Obliczenie procentu aktywnych dni w miesiącu.

Utworzenie histogramu wartości procentowych.

Histogram przedstawiony na rysunku 7.9 pokazuje, że większość klientów jest aktywna przez 25% – 75% dni w miesiącu. W tym szerokim rozkładzie można zauważyć potencjalnie dwie lub więcej wartości modalnych. W przypadku bardzo niewielu klientów występują miesiące, w których ich aktywność przekracza 80% lub spada poniżej 20% dni.



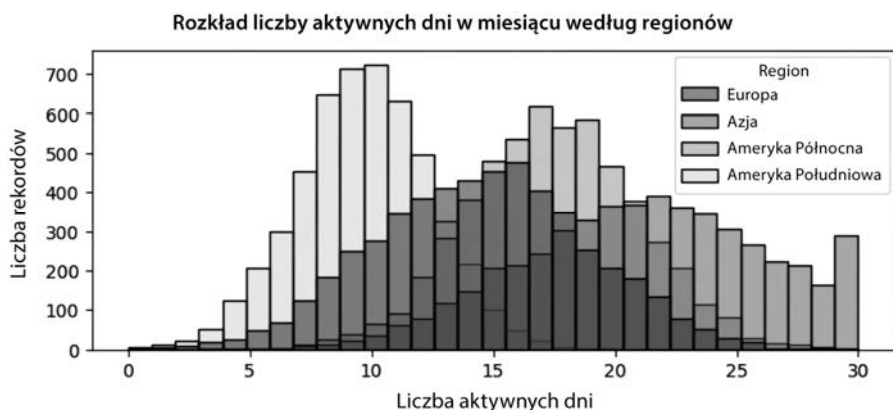
Rysunek 7.9. Histogram przedstawiający procent aktywnych dni w miesiącu wskazuje na możliwy rozkład bimodalny

Aby uzyskać bardziej subtelny obraz w porównaniu z oryginalnym histogramem, utwórzmy wykres *liczby* aktywnych dni w podziale na segmenty regionów dostępne w zbiorze danych:

```
sns.histplot(x=logins["login_days"], hue=logins["region"], bins=31)
```

Na rysunku 7.10 można zaobserwować, że rozbieżność wartości wskaźnika na segmenty powoduje dodanie *wielu* dodatkowych niuansów. Dla każdego regionu występują odrębne rozkłady bazowe. Klienci w Azji wydają się mieć najwyższą

medianę procentu aktywnych dni, a tuż za nimi plasują się klienci z Ameryki Północnej. Klienci z Ameryki Południowej wykazują zwykle najmniejszą aktywność. W ramach badań mających na celu ustalenie punktów odniesienia to odkrycie sugeruje, że w przypadku tego wskaźnika dla każdego regionu należy wyznaczyć *osobne cele i śledzenie*.



Rysunek 7.10. Histogram przedstawiający liczbę aktywnych dni w podziale na segmenty regionów umożliwia lepsze zrozumienie danych

W dalszej kolejności można przyrzeć się danym szeregu czasowego w podziale na segmenty regionów. Zaczniemy od analizy wartości miesięcznych dla wszystkich trzech lat ujętych w zbiorze danych:

```
avg_logins = (
    logins.groupby(["month", "region"])["login_days_pct"].mean().reset_index()
)

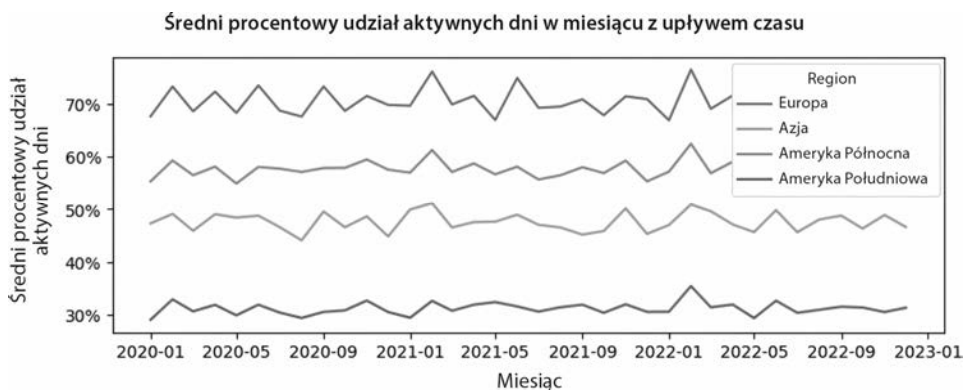
sns.lineplot(x="month",
              y="login_days_pct",
              hue="region",
              data=avg_logins)
plt.xlabel("Miesiąc")
plt.ylabel("Średni procentowy udział aktywnych dni")
plt.title("Średni procentowy udział aktywnych dni w miesiącu z upływem czasu")
```

← **Utworzenie wykresu średniej liczby logowań w miesiącu z upływem czasu.**

Obliczenie średniej liczby logowań w miesiącu według regionu.

Dane przedstawione na rysunku 7.11 pokazują wyraźne różnice między regionami, które pozostają stabilne przez trzy lata objęte tym zbiorem danych. Widoczne są pewne losowe wahania, ale brak jest zauważalnego trendu z miesiąca na miesiąc. Nie można stwierdzić, że istnieją dowody na jakikolwiek wzrost lub spadek, który nastąpił w okresie objętym pomiarem.

Zbadajmy następnie potencjalne trendy sezonowe, agregując dane w ujęciu miesięcznym i rocznym. Aby ułatwić interpretację wykresu, numery miesięcy zostaną też przekształcone w ich nazwy:



Rysunek 7.11. Średni procentowy udział aktywnych dni w miesiącu zagregowany według miesięcy w okresie trzech lat

```
logins["month_num"] = logins["month"].dt.month
logins["year"] = logins["month"].dt.year
```

← Zapisanie miesiąca i roku
jako nowe kolumny.

```
avg_logins_m = (
    logins.groupby(["year",
                    "month_num"])["login_days_pct"].mean().reset_index()
)
```

← Obliczenie średniej liczby logowań
dla poszczególnych numerów
miesiący i roku.

```
sns.lineplot(
    x="month_num",
    y="login_days_pct",
    hue="year",
    data=avg_logins_m
)
plt.xticks(rotation=45)
plt.title("Średni procentowy udział aktywnych dni w miesiącu w ujęciu miesięcznym  
i rocznym")
```

← Utworzenie wykresu średniej liczby
logowań dla poszczególnych nazw
miesiący i roku.

Rysunek 7.12 prezentuje potencjalną sezonowość, w przypadku której odsetek aktywnych dni był nieco wyższy w lutym niż w innych miesiącach. Różnica jest jednak niewielka i zauważalna tylko ze względu na mały zakres osi *y* (50% – 58%). Warto zweryfikować ten trend z zespołami odpowiedzialnymi za produkty, marketing i obsługę klienta, aby ustalić, czy zwiększona aktywność ma związek z jakimkolwiek znanym powodem (np. coroczną promocją).

W firmie, z której pochodzi ten zbiór danych, kolejnym krokiem jest zbadanie potencjalnych powiązań z istniejącymi wskaźnikami. Ze wskaźnikami aktywności klientów należy porównać wskaźniki wydajności, takie jak wielkość kontraktów, dodatkowe zakupy i wskaźniki odnowienia umów. Umiarkowane lub silne korelacje mogą stanowić podstawę do formułowania hipotez na temat długoterminowych efektów działań mających na celu poprawę wartości nowego wskaźnika.

Zależności korelacyjne można wykorzystać do tworzenia hipotez dotyczących sposobów osiągania celów związanych ze wskaźnikiem aktywności klientów oraz ustalania realistycznych celów. Korelacje z innymi procesami nie powinny być



Rysunek 7.12. Liczba aktywnych dni w miesiącu przedstawiona dla poszczególnych nazw miesięcy i roku w celu zbadania potencjalnej sezonowości

jednak używane do określania *istotności* danego wskaźnika. W rzeczywistości wskaźnik o bardzo wysokiej korelacji (np. powyżej 0,8) jest prawdopodobnie nadmiarowy i nie wniesie nowych informacji do strategii firmy.

Podstawowe zrozumienie historycznych wzorców wskaźników pomaga interesariuszom efektywniej korzystać z nich i wyznaczać cele. Stosując standardową metodę analizy, taką jak właśnie zaprezentowana, umożliwisz lepsze i bardziej strategiczne podejście do kwestii zrozumienia wskaźnika i zachowań użytkowników, a także zapewnisz, że spełnia on kryteria koncepcji SMART.

IDENTYFIKACJA ODPOWIEDNIH PUNKTÓW ODNIESIENIA

Punkty odniesienia to wartości referencyjne służące do porównywania z wartościami Twoich wskaźników. *Analiza porównawcza* odnosi się do procesu zestawiania Twoich wskaźników z określonymi punktami odniesienia. Ponieważ taki punkt reprezentuje cel wydajnościowy, który został wcześniej osiągnięty, zastosowanie tych punktów danych stanowi bardzo skuteczny punkt zaczepienia przy ustalaniu celów dla Twojej firmy.

Wiele źródeł danych może służyć jako punkty odniesienia w przypadku nadawania kontekstu wartościom Twoich wskaźników:

- *Branżowe dane porównawcze* zapewniają firmie średnie wartości dotyczące wydajności. Takie punkty odniesienia są dostępne dla wskaźników powszechnie mierzonych w danej branży (np. wskaźników wygranych w przypadku transakcji sprzedaży oraz przedziałów wynagrodzeń dla stanowiska, na które rekrutuje Twoja firma). Dane te są zazwyczaj gromadzone przez stowarzyszenia branżowe lub firmy zewnętrzne, które udostępniają lub sprzedają te informacje po wyższej cenie.

- *Recenzowane dane porównawcze* są cenne dla firm działających w branżach, w których prowadzone są intensywne badania naukowe. Ten typ danych jest dostępny w czasopismach naukowych. Na przykład firma z branży technologii dla służby zdrowia może porównać dokładność nowego urządzenia, które projektuje, z czasem diagnostyki i dokładnością w bazie danych takiej jak PubMed.
- *Rządy oraz organizacje pozarządowe i non-profit* gromadzą ogromne ilości danych udostępnianych do użytku publicznego. Jeśli uznajesz za wartościowe porównywanie ze wskaźnikami społecznymi i ekonomicznymi, istnieje duże prawdopodobieństwo, że znajdziesz informacje udostępnione dla swojej lokalizacji geograficznej. Większość tych źródeł danych pochodzi z badań przygotowanych tak, by dostarczać szacunków zjawisk na poziomie populacji. Na przykład wskaźnik bezrobocia, który omówiono na początku tego rozdziału, jest aktualizowany co miesiąc wraz z dziesiątkami innych wskaźników dostępnych w witrynie internetowej amerykańskiego Biura Statystyki Pracy (np. wydatki na obuwie przedstawione na rysunku 7.13). Takie miary jak oferty pracy, informacje o zdrowiu publicznym, postawy społeczne, zyski kwartalne spółek publicznych i inne są bezpłatnie dostępne dla Ciebie i Twojej firmy jako punkty odniesienia. Te cenne źródła danych mogą pomóc Ci zrozumieć szerszy system, w którym działa firma.



Rysunek 7.13. Roczna sprzedaż firmy obuwniczej w porównaniu z referencyjnymi szacunkowymi wydatkami na obuwie według danych z badania wydatków konsumentów przeprowadzonego przez Biuro Statystyki Pracy

Twoja firma skorzysta na użyciu wielu źródeł danych porównawczych do wyznaczania celów strategicznych. Załóżmy, że masz możliwość segmentacji i filtrowania tych danych, aby lepiej odzwierciedlały charakterystykę Twojej firmy (np. jej branżę i wielkość) albo bazę jej użytkowników lub klientów (np. wiek i lokalizację geograficzną). W tym przypadku Twoje cele będą znacznie bardziej realistyczne i osiągalne.

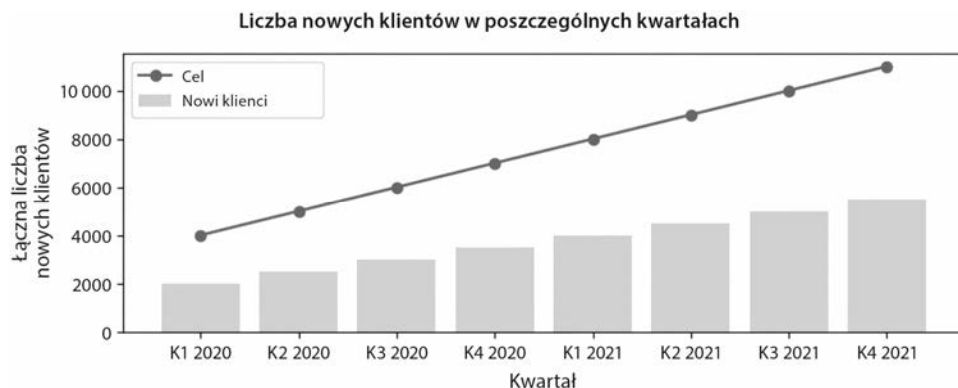
WYZNACZ REALISTYCZNE CELE

Po ustaleniu punktu odniesienia kolejnym krokiem jest wyznaczenie celów. *Cel* to pożądaný przyszły stan lub poziom wydajności ustalanego wskaźnika. Cel reprezentuje poprawę dotyczącą procesu lub doświadczenia. Cele powinny być ambitne i realistyczne, a ponadto muszą być zgodne z celami strategicznymi firmy.

Firmy przeprowadzają dogłębną analizę danych historycznych i porównują wskaźniki, aby wyznaczyć świadome cele i zrozumieć środowisko, w którym działają. Pomaga to zapewnić, że cele są *osiągalne* i *istotne* w procesie napędzania wzrostu i dokonywania ulepszeń. Proponując cele, rozważ następujące pytania, aby ułatwić określenie ich adekwatności dla zespołów, które będą odpowiedzialne za ich realizację:

- Jak przekonane są zespoły i interesariusze do działań, które należy podjąć w celu wprowadzenia zmiany? Czy istnieją znane czynniki, które korelują lub przewidują zmianę we wskaźniku?
- Czy istnieje dowód na to, że wskaźnik da się zmienić? Czy jest on stabilny, czy też waha się losowo mimo wysiłków mających na celu stymulowanie wzrostu?
- Czy zespoły i interesariusze mają możliwości napędzania wzrostu i dokonywania ulepszeń na poziomach, do których dążą? Czy będą w stanie poprawić *jakość* swoich działań, czy też oczekuje się od nich po prostu zwiększenia *ilości* wykonywanej pracy?
- Czy wyznaczone cele będą motywować zespoły do ich osiągnięcia, czy obniżą morale i pogorszą ich wydajność, jeśli nie będą w stanie ich zrealizować?

Na przykład na rysunku 7.14 pokazano cel firmy polegający na pozyskiwaniu nowych klientów netto, który z czasem staje się coraz trudniejszy do zrealizowania.



Rysunek 7.14. Wyznaczenie celu w postaci podwojenia kwartalnej liczby rejestracji nowych klientów jest raczej mało realistyczne lub nie zmotywuje zespołów odpowiedzialnych za jego realizację

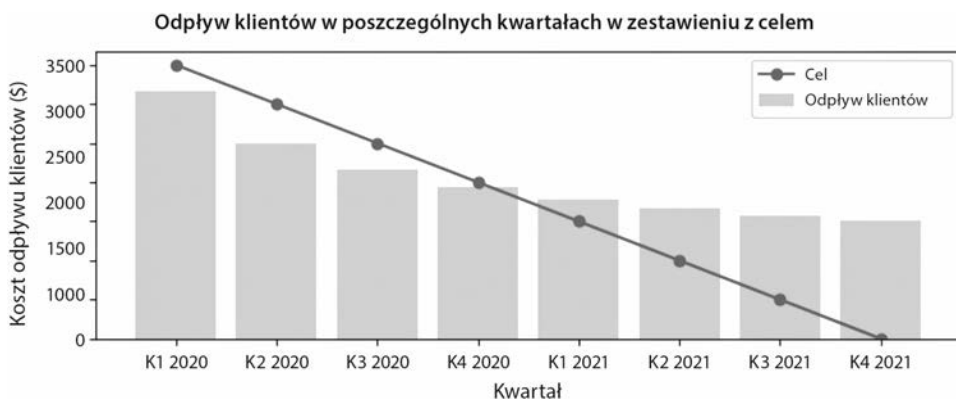
Twój zespół analityczny prawdopodobnie nie będzie ustalał celów dla *całej* firmy, ale Twoja wiedza na temat danych pozwoli Ci sformułować przemyślane rekomendacje dotyczące tych celów. Jeśli istnieje taka możliwość, *warto zabiegać o to, by członek zespołu analitycznego uczestniczył w rozmowach dotyczących planowania strategicznego*. Pozwoli to na świadome ustalanie celów dla firmy, zespołów i poszczególnych pracowników.

ITERACJA

Wyznaczanie celów nie jest jednorazowym zadaniem. Firmy i ich otoczenie stale się zmieniają, co wymaga ponownego przeglądu i odpowiedniego dostosowywania wskaźników oraz celów. Takie iteracyjne podejście zapewnia, że wskaźniki pozostają aktualne i gwarantują wartościowe informacje oraz strategiczne wytyczne.

Cele dotyczące usprawnień rzadko powinny pozostawać niezmiennie w kolejnych okresach. Na przykład dążenie do zmniejszenia odpływu klientów o 5% w każdym kwartale nie może być kontynuowane aż do całkowitego wyeliminowania tego zjawiska. Klienci będą odchodzić, nowi będą dołączać, a zachowania użytkowników będą się zmieniać. Nierealistyczne jest założenie, że poprawa dokonana w jednym procesie będzie trwać bez końca. Ty i Twoi interesariusze będziecie rozczarowani, a firma będzie marnować cenne zasoby na zadanie o malejących korzyściach. W ramach iteracji korzystne jest przesuwanie z upływem czasu uwagi między istotnymi procesami. Takie zrównoważone podejście zapewnia, że firma działa w oparciu o najnowsze i najbardziej znaczące informacje.

Podczas wyznaczania celów warto przypomnieć interesariuszom, że trajektorie wzrostu nie mogą trwać w nieskończoność (rysunek 7.15). Określenie, *kiedy* tempo poprawy zacznie zwalniać, jest trudne, ale świadomość tego, że w końcu to nastąpi, jest niezbędną.



Rysunek 7.15. Trajektorie wzrostu nie mogą być liniowe w nieskończoność. W ciągu czterech kwartałów statyczny cel zmniejszania odpływu klientów z kwartału na kwartał przynosi coraz mniejsze efekty i w końcu staje się nieosiągalny

Na przykład duże firmy technologiczne, takie jak Facebook i Netflix, mają miliony lub miliardy użytkowników. Ich wskaźniki będą prawdopodobnie obejmować cele polegające na zwiększeniu bazy użytkowników i ich zaangażowania. Biorąc pod uwagę skalę tych firm, można stwierdzić, że ich całkowita baza użytkowników jest całkiem dosłownie ograniczona liczbą populacji ludzi mających dostęp do internetu. Po osiągnięciu tych poziomów firmy te muszą wkraczać w nowe branże lub pogodzić się z ograniczeniami swojego potencjału biznesowego.

Przygotowywanie obliczeń pod kątem wskaźnika

Kolejnym krokiem Artura jest przeprowadzenie analizy eksploracyjnej danych historycznych na potrzeby nowego, proponowanego wskaźnika *Odsetek leadów kwalifikowanych jako sprzedażowe*. Obejmuje to przegląd historii danych sprzedażowych firmy w celu zrozumienia wzorców i trendów dotyczących typów leadów uznawanych za kwalifikowane, a także tego, jak zmieniała się ich definicja oraz ile leadów spełnia poszczególne kryteria.

Artur zaczyna od zbadania rozkładu wskaźnika za pomocą histogramu. Odkrywa, że rozkład cechuje się silną skośnością prawostronną, co wskazuje, że bardzo niewiele leadów ma wysokie wyniki w porównaniu z większością leadów, których wyniki są dość niskie. Mierzony w czasie i grupowany kwartalnie odsetek kwalifikowanych leadów waha się, ale nie widać wyraźnego wzrostu ani spadku na przestrzeni trzech lat.

Następnie Artur dzieli rozkład na trzy segmenty istotne dla firmy, czyli branżę, wielkość firmy i region geograficzny. Stwierdza, że procent kwalifikowanych leadów znacznie różni się zależnie od branży, przy czym najwyższe wyniki obserwuje się wśród firm technologicznych. Sugeruje to, że firma jest dobrze przygotowana do sprzedaży swojego produktu w branży technologicznej.

Na koniec Artur bada siłę korelacji między wskaźnikiem *Odsetek leadów kwalifikowanych jako sprzedażowe* a istniejącymi wskaźnikami sprzedaży, takimi jak współczynnik wygranych (procent pozyskanych leadów) i średnia wartość transakcji (kwota podana w kontrakcie). Identyfikuje umiarkowaną korelację dodatnią między nowym wskaźnikiem, współczynnikiem zamknięcia i średnią wartością transakcji. Sugeruje to, że leady z wyższym wynikiem mają większe szanse na przekształcenie się w sprzedaż, a ponadto że te kontrakty będą miały wyższą wartość. Skłania to Artura do dalszego rekomendowania nowego wskaźnika jako potencjalnego predyktora sukcesu sprzedażowego.

Artur sporządza raport zawierający wszystkie wizualizacje, które stworzył podczas procesu eksploracji, wraz z interpretacją każdego jego kroku. Przedstawia wyniki reszcie zespołu. Na podstawie analizy danych historycznych i punktów odniesienia dla wyników branży technologicznej zespół ds. sprzedaży postanawia wyznaczyć cel w postaci zwiększenia ogólnego odsetka leadów kwalifikowanych jako sprzedażowe do poziomów obserwowanych wśród firm technologicznych. Osiągnięcie tego celu będzie prawdopodobnie wiązało się ze zwiększeniem nacisku na sektor technologiczny przy poszukiwaniu nowych leadów sprzedażowych oraz z przeprowadzeniem bardziej dogłębnej analizy czynników w innych branżach, które klasyfikują je jako leady kwalifikowane.

7.2.3. Ćwiczenia

Po zidentyfikowaniu w przypadku zespołu marketingowego wskaźników wydajności, wskaźników organizacyjnych i wskaźników odpowiedzialności wykorzystajmy je do opracowania wskaźnika SMART. Stworzymy i zweryfikujemy jeden

lub więcej wskaźników reprezentujących ogólny cel, jakim jest pomiar zaangażowania w witrynie internetowej. Załóżmy, że masz dostęp do danych analitycznych witryny, w tym informacji o wyświetleniach stron, czasie trwania sesji, danych o kliknięciach (np. przycisków na stronie) oraz zbiorczych danych o zaangażowaniu związanym z postami w mediach społecznościowych:

1. Doprecyzuj cel, czyli pomiar zaangażowania w witrynie internetowej, aby był bardziej konkretny. Jakie są możliwe *szczegółowe* definicje, które wskażą na zaangażowanie w witrynie? Biorąc pod uwagę dostępne dane, jakie są zalety, wady lub sprzeczne definicje?
2. Zaproponuj jeden lub więcej sposobów pomiaru tego celu. Czy jest on już dostępny na liście wskaźników i zbiorów danych? Czy konieczne jest zebranie jakichkolwiek dodatkowych danych, aby Twoja konkretna definicja *umożliwiała pomiar*?
3. Zastanów się, jak wyglądałaby poprawa w przypadku zaproponowanego przez Ciebie wskaźnika. Co byłoby zarówno znaczące, jak i *osiągalne*?
4. Jak można udokumentować, gdzie rozpatrywany wskaźnik jest *istotny* dla Twojej firmy? Dlaczego jest on ważny z punktu widzenia obecnych celów Twojego zespołu? Jakie według Ciebie będą skutki wyznaczenia celów dla tego wskaźnika?
5. Zaproponuj konkretny harmonogram osiągnięcia celu. Jakie mogą być zalety i wady, jeśli Twój wskaźnik będzie *ograniczony czasowo* do tygodnia, miesiąca lub kwartału?
6. Po ocenie każdego z tych kryteriów udokumentuj swój wskaźnik SMART. Jaka jest nazwa wskaźnika i jego cel, o którym możesz poinformować interesariuszy i szerszą grupę osób w firmie?

7.3. Unikanie pułapek związanych ze wskaźnikami

Nie wszystkie wskaźniki ani ich wizualne reprezentacje są tworzone w jednakowy sposób. Choć wskaźniki w organizacji mogą kierować podejmowaniem decyzji i pokazywać postępy w realizacji celów, mogą też wprowadzać w błąd, dezorientować i zniekształcać rzeczywistość, jeśli nie zostaną odpowiednio skonstruowane lub zaprezentowane. W tym podrozdziale zostaną omówione typowe błędy pojawiające się przy obliczaniu i przedstawianiu wskaźników, które mogą maskować podstawowe trendy, zniekształcać wyniki i dezorientować Ciebie oraz interesariuszy. Aby zapewnić interesariuszom dokładny kierunek działań, oprócz spełnienia kryteriów koncepcji SMART konieczne jest upewnienie się, że wskaźnik unika każdej z tych pułapek.

7.3.1. Reprezentacja

Interpretacja wskaźnika SMART może się różnić w zależności od sposobu jego obliczania i prezentowania użytkownikom. Choć pewne reprezentacje (np. średnia zestawiona z medianą, wykres słupkowy porównany z liniowym) są odpowiednie dla danego kształtu danych, niektórych powszechnych sposobów prezentacji wskaźników należy zawsze unikać. Analitycy zdecydowanie zbyt łatwo wpadają w pułapkę używania nieodpowiednich reprezentacji, które mogą zniekształcać wyniki i odciągać uwagę od strategicznego celu wskaźnika.

WARTOŚCI SKUMULOWANE

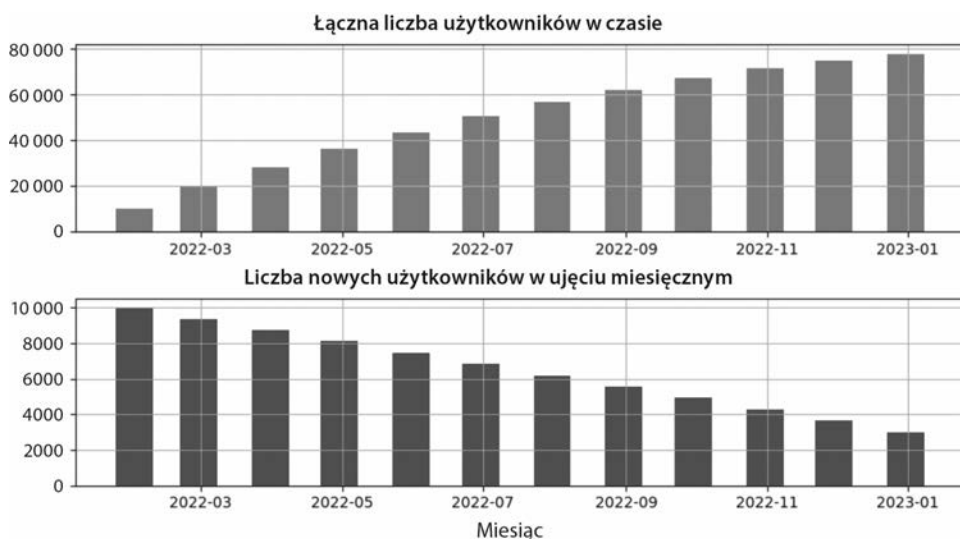
Wskaźniki śledzące w czasie liczbę lub sumę wartości można przedstawić jako obliczenia *przyrostowe* lub *skumulowane*. *Wskaźnik przyrostowy* uwzględnia tylko nowe wartości dla danego okresu (np. miesięczną sprzedaż netto). Ten typ wskaźnika można bezpośrednio porównywać między kolejnymi okresami, aby zrozumieć, jak procesy zmieniają się z upływem czasu. Wskaźniki przyrostowe są korzystne, ponieważ obrazują wahania między poszczególnymi okresami, ostrzegając firmy o zmianach w wydajności, które powinny wpływać na podejmowane działania i motywować do nich.

Wskaźnik skumulowany oblicza wszystkie bieżące i poprzednie wartości dla danego okresu. Skumulowany widok danych ukazuje łączny wynik do danego momentu, a nie rejestruje wartości dla konkretnego przedziału czasu. Choć może to być przydatne do prezentowania całkowitego postępu w czasie, wskaźniki skumulowane mogą ukrywać wahania lub trendy, które w innych okolicznościach powinny motywować w firmie do podejmowania strategicznych decyzji (tabela 7.5).

Tabela 7.5. Przykłady wskaźników skumulowanych i ich bardziej odpowiednie alternatywy przyrostowe

Wskaźnik skumulowany	Wskaźnik przyrostowy
Całkowita liczba klientów w ujęciu miesięcznym	Procentowa zmiana całkowitej liczby klientów w ujęciu miesięcznym
Przychody od początku działalności	Porównanie przychodów z kosztami w ujęciu miesięcznym

Wskaźniki skumulowane dokumentują stan firmy w różnych momentach. Na przykład firmy muszą zaadaptować skumulowany widok danych, aby z upływem czasu śledzić wzrost całkowitych przychodów cyklicznych i bazy klientów. Ilustruje to rysunek 7.16, na którym skumulowana łączna liczba użytkowników w każdym miesiącu odzwierciedla aktualny stan bazy użytkowników i ogólną trajektorię wzrostu. Aby w pełni zrozumieć te trendy, wskaźnikowi skumulowanemu musi towarzyszyć przyrostowa wersja tego samego wskaźnika. Można zauważyć, że choć baza użytkowników zwiększa się, *tempo* wzrostu maleje. Subtelność ta jest trudniejsza do dostrzeżenia wyłącznie na podstawie wykresu skumulowanego.



Rysunek 7.16. Skumulowana liczba użytkowników prezentuje ogólny wzrost, ale może maskować spadek jego tempa

Gdy wskaźnikowi skumulowanemu nie towarzyszy obliczenie przyrostowe, nawet wskaźnik SMART wyznaczony jako wartość skumulowana może wpaść w pułapkę stania się dla firmy *wskaźnikiem pozornego sukcesu*. Tego rodzaju wskaźniki to punkty danych, które dobrze wyglądają na papierze i łatwo z ich okazji świętować, ale zwykle *nie* przyczyniają się do uzyskania praktycznych wniosków. Wykres z ciągle rosnącą wartością (np. skumulowaną bazą użytkowników) może być z łatwością celebrowany, ale *nie* dostarcza informacji przydatnych do określania celów, działań i strategii.

Ostatecznie wskaźniki skumulowane rzadko są tak wartościowe, jak sądzą analitycy i interesariusze. Mogą one błędnie przedstawiać w pozytywnym świetle obszary problemów istniejących w firmie i często maskują subtelne trendy lub zmiany w danych. Upewnij się, że wskaźnikom skumulowanym towarzyszy obliczenie *przyrostowe*, aby interesariusze poznali rzeczywisty status celów.

SZCZEGÓŁOWOŚĆ

Opracowywanie wskaźników często obejmuje wybór odpowiedniej *szczegółowości* danych *przed* wykonaniem obliczeń zagregowanych. *Szczegółowość* odnosi się do poziomu informacji dostępnych w każdym wierszu. Podobnie jak w przypadku kompromisów towarzyszących wyborowi miar tendencji centralnej (zajrzyj do rozdziałów 4. i 5.), ustalenie stopnia dokładności może uwypuklić lub zminimalizować różne cechy danych prezentowanych interesariuszom, co wpływa na podejmowane przez Ciebie decyzje.

Firmy różnią się *znacznie* pod względem postrzegania kwestii szczegółowości, co jest zależne od ich klientów, modelu biznesowego i celów. Oto kilka typowych przykładów, z którymi możesz się spotkać w trakcie projektu:

- W modelu B2B dane są zazwyczaj zbierane na temat *klientów* (innych firm) oraz *użytkowników* każdego klienta. Analityk musi zdecydować, czy odpowiednie jest obliczanie wskaźnika na poziomie użytkownika, czy też w formie zagregowanej na poziomie klienta.
- Większość firm gromadzi *informacje geograficzne* o swoich klientach i użytkownikach. Dane geograficzne można łatwo agregować na różnych poziomach szczegółowości (np. kod pocztowy, miasto, województwo, kraj) oraz według typów obszarów (miejskie lub wiejskie).
- *Kategorie produktów* w firmach zawierają zwykle wiele poziomów szczegółowości i nakładających się klasyfikacji służących do śledzenia i analizy. Na przykład baza danych magazynu w firmie handlowej będzie klasyfikować działy według grup wiekowych (np. odzież dziecięca lub dla dorosłych), płci (męska lub damska) oraz rodzaju odzieży (koszule, spodnie itp.).

Wybór poziomu szczegółowości to ostatecznie kwestia równowagi. Wyższy poziom szczegółowości (np. kod pocztowy zamiast województwa) może dostarczyć bardziej szczegółowych informacji, ale ogranicza możliwość identyfikacji szerszych trendów. Z kolei niższy poziom szczegółowości może zaciemnić trendy i ukryć istotne podzbiory danych. W ramach analizy wskaźników warto segmentować swoje dane na różnych poziomach szczegółowości, aby określić odpowiedni *poziom dokładności*, który zostanie zarekomendowany interesariuszom.

Przeanalizujmy przykład procesu eksploracji i podejmowania decyzji dotyczącej wyboru między agregacją danych na poziomie użytkownika lub na poziomie klienta w firmie z segmentu B2B. Firma ma kilkuset klientów z wielu branż i różnej wielkości.

Na rysunku 7.17 przedstawiono przykładowy fragment tabeli customers z bazy danych firmy. Jeśli chce ona stworzyć i monitorować prosty wskaźnik *aktywności użytkowników*, może zacząć od utworzenia zapytania agregującego, którego wynik pokazano w tabeli na rysunku 7.18.

	id	created_date	status	contract_amount	subscription_type	industry	company_size	referral
0	012	2023-02-17	Churned	13614.88	Pro	Tech	Large	No
1	013	2023-04-18	Active	14798.01	Pro	Tech	Large	Yes
2	014	2023-01-24	Inactive	14781.55	Basic	Finance	Small	Yes
3	015	2023-03-18	Churned	11471.28	Basic	Education	Medium	No
4	016	2023-04-07	Churned	18448.20	Pro	Healthcare	Large	Yes

Rysunek 7.17. Przykładowe dane zawierające informacje o klientach firmy z segmentu B2B

	month	total_users	active_users	pct_active_users
0	2022-08	47077	24800	52.679652
1	2022-09	47054	21504	45.700684
2	2022-10	47060	25568	54.330642
3	2022-11	47035	24700	52.514085
4	2022-12	47040	22185	47.161990
5	2023-01	47042	27391	58.226691
6	2023-02	47060	21809	46.342966

Rysunek 7.18. Przykład wskaźnika aktywnych użytkowników w każdym miesiącu zagregowanego dla wszystkich użytkowników

W tabeli customers obecnych klientów widać, że każda firma zatrudnia od 10 do 10 000 pracowników, co przedstawiono w kolumnie kattegorycznej `company_size` (Small: 10 – 99, Medium: 100 – 999, Large: ponad 1000). Jeśli dla wskaźnika obliczy się ogólną sumę wartości, wynik będzie faworyzował większych klientów. Ponieważ mają oni więcej użytkowników, ich wskaźniki aktywności będą miały znacznie większy wpływ na wskaźnik.

Aby uwzględnić różnice w wielkości firm, można *częściowo zagregować* dane na poziomie firmy, generując wynik aktywności użytkowników *dla każdej firmy w ujęciu miesięcznym*. Taki pośredni zbiór danych można *ponownie* ogólnie zagregować dla każdego miesiąca, skutecznie eliminując wpływ faworyzujący duże firmy.

Na rysunku 7.19 każdy klient jest traktowany jako równorzędna encja w przypadku wskaźnika częściowo zagregowanego zbioru danych. Przy założeniu, że strategia firmy polega na zaangażowaniu całej bazy klientów, wskaźnik średniego procentu aktywnych użytkowników w ujęciu miesięcznym zapewni bardziej dokładne i przydatne wnioski na temat postępów w realizacji celów.

	customer_id	month	n_users	n_active_users	pct_active
0	012	2022-08	697	599	0.859397
1	012	2022-09	692	70	0.101156
2	012	2022-10	695	472	0.679137
3	012	2022-11	695	600	0.863309
4	012	2022-12	695	396	0.569784

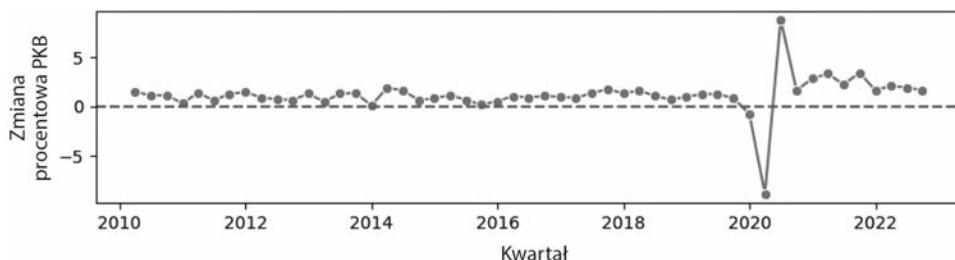
Rysunek 7.19. Częściowo zagregowane dane aktywności użytkowników dla poszczególnych klientów i miesięcy

WSKAŹNIKI ZŁOŻONE

Wskaźnik złożony lub *indeks* łączy co najmniej dwa wskaźniki w pojedynczy, standaryzowany wynik, który można przedstawić użytkownikom lub interesariuszom. Wskaźniki te są często stosowane w pomiarze złożonych procesów, gdy

raportowanie pojedynczych wskaźników mogłoby utrudnić wyciągnięcie praktycznych wniosków. W teorii połączenie wielu wskaźników pozwala na łatwiejsze zrozumienie na poziomie heurystycznym mierzonych podstawowych procesów.

Indeksy są powszechnie używane do raportowania informacji ekonomicznych i społecznych, takich jak produkt krajowy brutto (PKB) publikowany kwartalnie w Stanach Zjednoczonych przez Biuro Analiz Ekonomicznych. Ze względu na swoją prostotę i powszechność procentowa zmiana wskaźnika PKB może być w przystępny sposób przedstawiana szerokiemu gronu odbiorców. Zakłada się powszechne rozumienie tego wskaźnika. Jego wzrost procentowy jest postrzegany jako pozytywny, a spadek jako negatywny (rysunek 7.20).



Rysunek 7.20. Wskaźnik PKB jest przedstawiany jako wartość w dolarach oraz w postaci wartości procentowej ukazującej względną zmianę w stosunku do poprzedniego kwartału

Indeksy, takie jak PKB, wymagają szeroko zakrojonych badań w celu właściwego połączenia wskaźników. Oczekuje się, że poszczególne elementy będą zmieniać się razem, odzwierciedlając ten sam proces bazowy. Do identyfikacji podobnych miar, które można wykorzystać do stworzenia skali złożonej, można zastosować wskaźniki rzetelności omówione w rozdziale 5. (np. rzetelność wewnętrzną). Jednakże indeksy mają zwykle poważne ograniczenia dotyczące dokładności i możliwości praktycznego użycia dostarczanych przez nie wartości:

- Z każdym elementem dodanym do indeksu zmniejsza się *zmienność wyników* w czasie i między encjami (np. klienci, kraje). Podjęcie próby uproszczenia obliczania wielu wskaźników prawdopodobnie spowoduje zaciemnienie wartościowych trendów, które w przeciwnym razie byłyby widoczne w poszczególnych wskaźnikach tworzących indeks.
- Uwzględnienie w indeksie lub wykluczenie z niego poszczególnych wskaźników jest ostatecznie *arbitralną* decyzją, którą musisz podjąć wraz z interesariuszami. Projektując i stosując indeks, zwłaszcza w celu wewnętrznego śledzenia, zobowiązujesz się do ciągłego jego udoskonalania na podstawie nowych ustaleń, procesów lub usług oferowanych w firmie.
- Przedstawienie *równowagi* między rywalizującymi procesami może być trudne. Możesz tworzyć i raportować wskaźniki przeciwdziałające

(które szerzej zostaną omówione w dalszej części tego rozdziału), ale prawdopodobnie nie będą one postrzegane jako tak samo ważne jak sam indeks.

Każde z tych ograniczeń dotyczy obliczania wskaźnika PKB, który określa wartość wszystkich produktów i usług wytworzonych w określonym czasie. Jest on uważany za kluczowy wskaźnik kondycji gospodarczej. Niemniej jednak wskaźnik ten *pomija* takie informacje jak dystrybucja majątku, deprecjacja wartości wytworzonych produktów lub towary i usługi oferowane na „czarnym” rynku. Każdy z tych czynników potencjalnie wpływa na stan gospodarki oraz jakość życia obywateli i społeczności reprezentowanych w obliczeniach wskaźnika PKB. Teoretycznie każdy z tych czynników *mógłby* zostać uwzględniony w obliczeniu tego wskaźnika. Jednakże żaden pojedynczy indeks nie zapewni pełnego obrazu gospodarki, nadal pozostając wrażliwym na zmiany.

O ile Twoja firma nie dysponuje ogromnymi zasobami, które może przeznaczyć na badania niezbędne do opracowania odpowiednich indeksów opartych na danych, zdecydowanie odradzam ich tworzenie. Indeksy i wskaźniki złożone wydają się atrakcyjne dla interesariuszy ze względu na pozorną prostotę i zdolność do heurystycznego reprezentowania mierzonego pojęcia lub procesu. Często jednak utrudniają podejmowanie decyzji na podstawie danych.

WSKAŹNIKI PRZECIWDZIAŁAJĄCE

Podczas śledzenia pełnego zakresu istotnych procesów w firmie stwierdzisz, że wiele z nich wydaje się być ze sobą w konflikcie. Na przykład firma tworząca oprogramowanie może mieć dwa następujące cele: przyspieszenie wprowadzania nowych opcji w aplikacji oraz zmniejszenie liczby błędów zgłaszanych w niej przez klientów. Są to przykłady *wskaźników przeciwdziałających* lub *ochronnych*, które służą jako mechanizm kontroli i równowagi, a także pomagają ujawnić potencjalne negatywne konsekwencje skupienia się wyłącznie na celu.

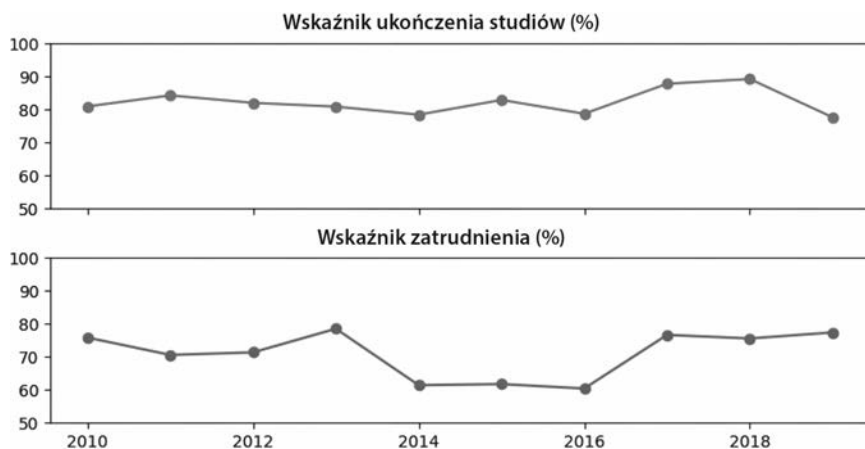
Wskaźniki przeciwdziałające identyfikuje się w *kontekście* projektowania nowego wskaźnika. Często są to już istniejące w firmie wskaźniki monitorowane pod kątem innych celów. Identyfikacja odpowiednich wskaźników przeciwdziałających to proces jakościowy wymagający krytycznego rozważenia nowego wskaźnika oraz związanego z nim zachowania lub procesu. Zadaj interesariuszom następujące pytania: Co może pójść nie tak, jeśli skupimy się tylko na tym wskaźniku? Gdzie możemy poświęcać jakość na rzecz ilości? Jakie procesy potencjalnie stoją w opozycji do tego wskaźnika, a nadal są ważne?

Rozważ następujące przykłady:

- Zespół wsparcia klientów dysponuje wskaźnikiem śledzącym liczbę zgłoszeń rozpatrzonych w ciągu 24 godzin. W ramach wskaźnika przeciwdziałającego zespół raportuje również poziom zadowolenia klienta, aby upewnić się, że nie pogarsza jakości wsparcia na rzecz ilości.

- Zespół inżynierów oprogramowania śledzi liczbę nowo dostarczonych opcji, co jest uważane za jedną z miar produktywności zespołu. Jako wskaźnik przeciwdziałający zespół monitoruje też liczbę nowo wprowadzonych błędów, aby upewnić się, że zwiększanie produktywności nie odbywa się kosztem jakości kodu.
- Organizacja non-profit zajmująca się edukacją śledzi wskaźniki przyjęć na studia wśród uczniów szkół średnich uczestniczących w ich programach. W ramach wskaźnika przeciwdziałającego organizacja monitoruje także wskaźniki ukończenia studiów i zatrudnienia absolwentów w dziedzinie oferowanych studiów. Pomimo tego, że nie dysponuje ona programami bezpośrednio wpływającymi na te procesy, uważa za istotne monitorowanie studentów w dłuższej perspektywie, aby zapewnić ich sukces.

W trakcie przygotowywania opracowania swoich wskaźników (paneli kontrolnych, raportów i prezentacji) zdecydowanie zaleca się uwzględnienie w nim wizualizacji wybranego wskaźnika przeciwdziałającego wraz z wyjaśnieniem jego znaczenia i uzasadnieniem (rysunek 7.21). Ponadto wskaźniki przeciwdziałające powinny być zawsze oceniane równoległe ze wszystkimi innymi wskaźnikami lub miarami stosowanymi w eksperymentach, testach A/B lub ocenach. Dodanie do panelu kontrolnego sekcji wizualizującej wszystkie wskaźniki przeciwdziałające wraz z wyjaśnieniem, *dlaczego* zostały wybrane, pozwoli Tobie i Twoim interesariuszom na ich iteracyjne doskonalenie z upływem czasu oraz śledzenie pełnego obrazu realizowanych celów.



Rysunek 7.21. Prosta wizualizacja wskaźnika ukończenia studiów oraz wskaźnika przeciwdziałającego zatrudnienia używanych przez organizację non-profit

7.3.2. Wizualizacja

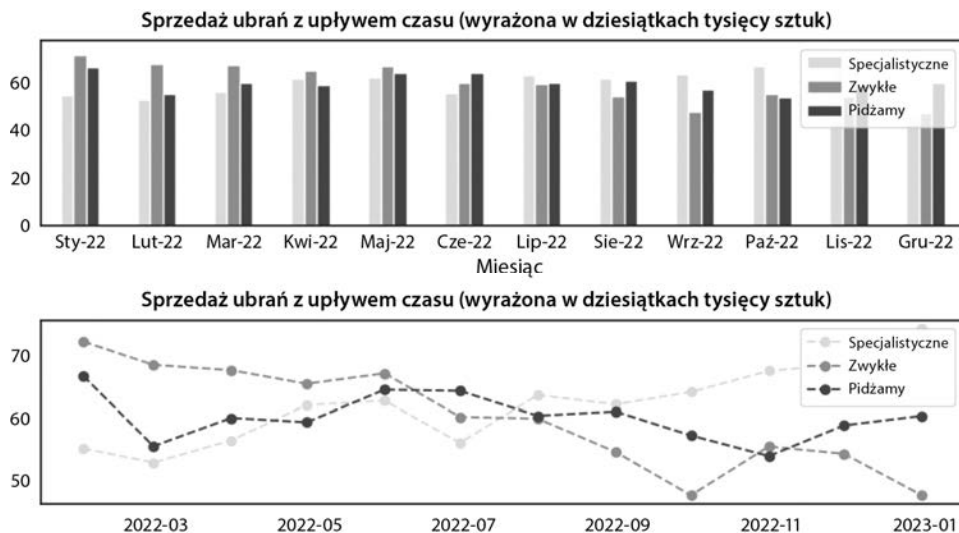
Zasadniczo wskaźnik jest narzędziem służącym do opowiedzenia interesariuszom wartościowej historii, a sposób jego wizualnego przedstawienia stanowi *znaczącą* część tej narracji. Dobrze dobrana wizualizacja może mówić sama za siebie, uwiadamiając trendy, które w przeciwnym razie wymagałyby długich opisów. Kiepska prezentacja wskaźnika może jednak łatwo zniweczyć całą pracę włożoną w zapewnienie, że Twoje wskaźniki są możliwe do zinterpretowania i zastosowania.

Ostatecznie zagadnienia omówione w tym punkcie dotyczą wizualizacji danych i opowiadania historii za ich pomocą. Skupimy się głównie na przedstawianiu wskaźników w postaci *danych szeregów czasowych* (danych prezentowanych na osi x zwykle za pomocą zmiennej czasu). Niemniej jednak uwzględnienie rozważań przedstawionych w tym punkcie może być korzystne w przypadku każdego typu wizualizacji, który nie jest związany z czasem.

WYBÓR NIEWŁAŚCIWEGO WYKRESU

Niewłaściwy dobór wykresów to *bardzo* powszechny problem. Jeśli zwrócisz uwagę na segmenty informacyjne prezentujące wyniki ankiety, badań czy innych danych, zauważysz, że wybrane wykresy nie zawsze są najlepszą opcją w przypadku danych. Nie wszystkie wykresy nadają się do prezentacji wskaźników, a wiele decyzji dotyczących wyboru wykresu może negatywnie wpłynąć na zdolność interpretacji danych:

- Wykres może łatwo *nadmiernie uprościć* trendy, które chcesz zobrazować za pomocą swojego wskaźnika. Pojedynczy wykres słupkowy lub liniowy może być pomocny w uzyskaniu *ogólnego* obrazu wskaźnika w skali całej firmy (widoczne to jest np. na rysunkach 7.8 i 7.16). Bądź jednak przygotowany na to, że różne działy i zespoły będą wymagały bardziej szczegółowego widoku wskaźnika (np. według regionu, branży).
- I odwrotnie, wykres może z łatwością *nadmiernie skomplikować* trend, jeśli zawiera zbyt wiele informacji lub jest to wizualizacja, z którą Twoi interesariusze nie są zaznajomieni. Jeżeli planujesz podzielić swój wskaźnik na segmenty, zdecydowanie zalecam użycie wykresu liniowego zamiast słupkowego (rysunek 7.22). Grupowany lub skumulowany wykres słupkowy przedstawiający dane w czasie może być trudny do interpretacji w porównaniu ze śledzeniem trajektorii wzrostu zestawu linii, pod warunkiem, że są one unikalne pod względem koloru lub reprezentacji.
- Podczas tworzenia panelu kontrolnego lub raportu w celu śledzenia wskaźników w czasie łatwo można wybrać *niewłaściwy wykres*. Wiele typów wykresów dostępnych w narzędziach do analizy biznesowej jest kuszących, zwłaszcza gdy mają być prezentowane wyniki prac. Jeśli jednak spodziewasz się, że będziesz musiał wyjaśniać zawartość wykresu swoim



Rysunek 7.22. Wykresy liniowe są zazwyczaj łatwiejsze do interpretacji w przypadku wizualizacji trendów między grupami z upływem czasu

przełożonym, warto raczej użyć prostszej wizualizacji w opracowaniu. To typ przeprowadzanej analizy powinien determinować wybór wykresu, a nie odwrotnie.

Grupowany wykres słupkowy lub liniowy jest zazwyczaj najlepszą opcją wyboru pod kątem odpowiedniej wizualizacji wskaźnika. W zależności od obliczeń można łatwo grupować lub sumować dane, a większość interesariuszy może interpretować takie wykresy bez dodatkowej pomocy. Unikanie bardziej złożonych wykresów zminimalizuje konieczność dodatkowych działań wykonywanych przez Ciebie i Twój zespół.

ZNIEKSZTAŁCANIE OSI

Kolejną częstą pułapką związaną z wizualizacją jest niewłaściwe użycie osi y . Reprezentuje ona *skalę* użytego wskaźnika, ukazując interesariuszom wielkość zmian w czasie. Bardzo łatwo można zniekształcić znaczenie zmian między przedziałami czasu w wyniku manipulowania osią y (rysunek 7.23).

Wyróżnić można kilka kluczowych metod, w przypadku których zmiany dotyczące osi y mogą zniekształcić wskaźniki:

- **Przycinanie osi y** , czyli zawężanie jej górnych i dolnych granic, może wyolbrzymiać zmiany w danych. Może to sprawić, że trendy wydadzą się bardziej znaczące, niż są w rzeczywistości, potencjalnie prowadząc do decyzji opartych na niewielkich różnicach. Ważne jest, aby to monitorować *nawet wtedy, gdy nie jest to robione celowo*, ponieważ do ustawiania limitów osi y narzędzia do wizualizacji danych używają



Rysunek 7.23. Rysunek przedstawia te same dane co rysunek 7.21, ale z domyślnymi granicami osi y ustawionymi za pomocą biblioteki matplotlib na podstawie minimalnych i maksymalnych wartości w zbiorze danych. Te domyślne ustawienia powodują, że wahania w danych wydają się przesadnie duże

zakresu wartości w zbiorze danych. Jeśli pełny zakres wartości wykracza poza minimum i maksimum, konieczne jest ręczne ustawienie granic.

- **Rozszerzanie osi y** poza rzeczywisty zakres wartości może zmniejszyć możliwość dostrzegania trendów lub wahań. Wiele wskaźników z faktycznie określonym minimum i maksimum (np. wartościami procentowymi) zwróci wartości tylko dla ograniczonego podzbioru całego zakresu. Wskaźnik ukończenia studiów przedstawiony na rysunkach 7.21 i 7.23 teoretycznie może przyjmować wartości od 0% do 100% (rysunek 7.24), ale dane roczne z okresu dekady nie zawierają wartości poniżej 75%. Jeśli dolna i górna granica osi y zostałyby ustawione, odpowiednio, na 0% i 100%, odbiorcy prawdopodobnie nie zauważyliby wahań, których odkrycie *tak naprawdę* może być znaczące i ważne. Na rysunku 7.21 oś jest ustawiona między 50% a 100%, co pozwala zobrazować zakres wahań bez przesadnego lub niedostatecznego przedstawiania zmian w czasie.

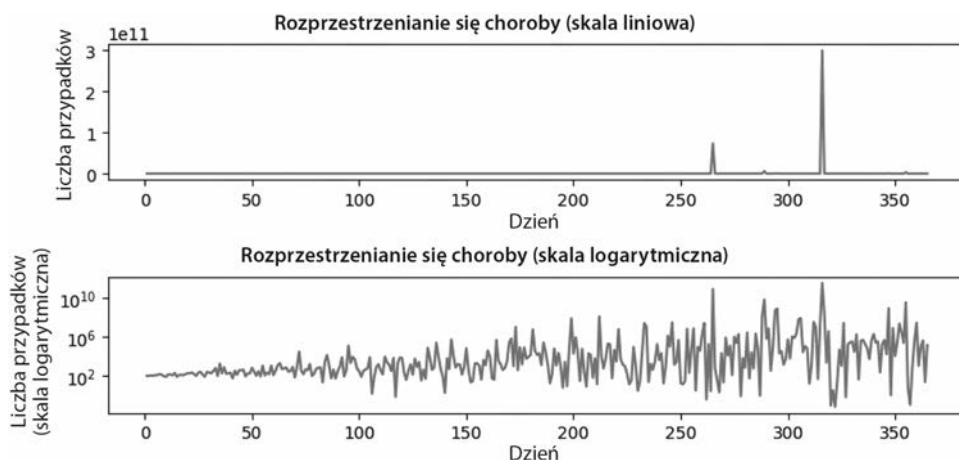


Rysunek 7.24. Ten sam wskaźnik ukończenia studiów wydaje się mieć niewielkie lub nie mieć żadnych wahań, gdy granice osi y są rozszerzone od 0% do 100%

- **Stosowanie podwójnej osi y** może być przydatne w niektórych kontekstach (np. do przedstawienia relacji między dwoma wskaźnikami o takiej samej skali czasowej). Jednakże, częściej niż rzadziej, podwójna oś y okaże się myląca dla interesariuszy i będzie wymagać do interpretacji znacznie więcej czasu i wysiłku, niż gdyby zostały przygotowane dwa osobne wykresy. Interesariusze będą zakładać, że dwa wskaźniki mają taką

samą skalę (np. cale, złotówki lub punkty procentowe), a ponadto mogą błędnie interpretować siłę korelacji na podstawie wizualnej obserwacji wahań.

- **Skale nieliniowe i logarytmiczne** są często stosowane do wizualizacji danych zmieniających się o kilka rzędów wielkości (np. wzrostu populacji). Dane przedstawione w skali logarytmicznej sprawiają wrażenie liniowych. Choć może to ułatwić przedstawienie pełnego zakresu wartości, skale logarytmiczne łatwo zniekształcają postrzeganie występujących zmian. Powinny być one używane *oszczędnie* (tylko wtedy, gdy niemożliwa jest interpretacja skali liniowej) i opatrzone jasnymi objaśnieniami skali dla interesariuszy (rysunek 7.25).



Rysunek 7.25. Skale logarytmiczne mogą pomóc w wizualizacji wskaźników z wykładniczymi wahaniami w danych. Powinny być one jednak stosowane z rozwagą

7.3.3. Ćwiczenia

Kontynuujmy działania związane ze wskaźnikiem SMART, który został opracowany wcześniej w rozdziale:

1. Poszukaj w internecie hasła „wskaźniki zaangażowania w witrynie internetowej”. Jakiego typu wyniki uzyskujesz?
2. Przeprowadź podstawowe badanie, aby zrozumieć inne definicje zaangażowania w witrynie internetowej. Czy istnieją artykuły o tym, jak je mierzyć? Jak się one mają do Twojego wskaźnika?
3. Zidentyfikuj niektóre cechy przykładowych wskaźników, które znalazłeś podczas wyszukiwania. Czy są one mierzone w skali dnia, tygodnia, czy miesiąca? Czy możesz znaleźć jakieś istotne trendy lub punkty odniesienia, które mogłyby być warte wykorzystania?

Identyfikacja najlepszych wskaźników przeciwdziałających

Kończąc prace nad wskaźnikiem *Odsetek leadów kwalifikowanych jako sprzedażowe*, Artur odkrywa kilka kroków, które musi podjąć, aby zapewnić jego użyteczność i skuteczność dla swojego zespołu.

Początkowo Artur rozważa przedstawienie skumulowanego widoku wskaźnika w celu podkreślenia ogólnego wzrostu liczby leadów kwalifikowanych w firmie. Zdaje sobie jednak sprawę z tego, że takie ujęcie maskuje wahania *sezonowe*, w ramach których więcej leadów generowanych w miesiącach letnich jest uznawanych za kwalifikowane. Postanawia zatem zaprezentować skumulowaną liczbę leadów kwalifikowanych jako pojedynczą wartość, aby docenić pracę zespołu, oraz przyrostowy widok wskaźnika dla każdego okresu.

Artur eksploruje następnie najlepsze możliwe sposoby wizualizacji różnic w liczbie kwalifikowanych leadów sprzedażowych w ujęciu miesięcznym. Używa do tego wykresu liniowego, ponieważ pozwala on łatwo zobrazować zmiany rozbite na różne segmenty istotne dla zespołu (np. region, branża).

Oprócz świadomości tych końcowych kwestii Artur zdaje sobie sprawę z tego, że zespół mógłby stracić z oczu inne wartościowe procesy, skupiając się wyłącznie na tym nowym wskaźniku. Aby go zrównoważyć, Artur proponuje dodatkowy miesięczny wskaźnik przeciwdziałający, który będzie raportowany razem ze wskaźnikiem *Odsetek leadów kwalifikowanych jako sprzedażowe*. Wskaźnikiem tym jest *Czas reakcji na lead*, który mierzy, ile czasu zajmuje kontakt z potencjalnym leadem po jego identyfikacji. Zapewnia to, że zespół nie będzie koncentrował się wyłącznie na leadach kwalifikowanych kosztem szybkości reakcji na *wszystkie* leady.

Staranne podejście Artura jest widoczne w jego prezentacji dla zespołu motywującej jego członków do zaadaptowania nowego wskaźnika oraz wskaźnika przeciwdziałającego do śledzenia skuteczności ich pracy.

4. Czy istnieją potencjalne wskaźniki przeciwdziałające, które powinienēs wziąć pod uwagę w przypadku swojego wskaźnika SMART?

Korzystając z wybranego środowiska języka Python (terminal, Jupyter Notebook itp.), zaimportuj plik zbioru danych *website_engagement.csv*. Dostępne są następujące kolumny, które można poddać analizie:

- `website_engagement` — wartość jest mierzona jako procent aktywnych użytkowników w każdym tygodniu,
- `session_duration` — średni czas trwania sesji odwiedzającego witrynę (w minutach),
- `bounce_rate` — procent odwiedzających, którzy opuszczają witrynę po wyświetleniu tylko jednej strony,
- `email_subscribers` — łączna liczba subskrybentów wiadomości e-mail; dostępna jest przyrostowa wersja tej kolumny o nazwie `new_email_subscribers`,
- `social_media_followers` — łączna liczba obserwujących w mediach społecznościowych; dostępna jest przyrostowa wersja tej kolumny o nazwie `new_social_media_followers`,

- `avg_page_views_per_visit` — średnia liczba stron wyświetlonych dla poszczególnych sesji użytkownika,
- `total_items_purchased` — całkowita liczba produktów zakupionych przez użytkowników w danym tygodniu,
- `total_sales` — całkowita wartość sprzedaży w danym tygodniu (w dolarach).

Do wykonania następujących działań muszą zostać zainstalowane biblioteki `numpy`, `matplotlib` i `pandas`:

1. Ustal kompleksowy zestaw bazowych informacji dotyczących zaangażowania użytkowników w witrynie. Jakie trendy (np. sezonowość, kształt rozkładu, zmiany wzdłużne), o których powinni wiedzieć interesariusze, występują w przypadku wskaźnika zaangażowania użytkowników?
2. Czy istnieją powiązania z innymi wskaźnikami, o których warto poinformować interesariuszy?
3. Zidentyfikuj w zbiorze danych jeden lub więcej potencjalnych wskaźników przeciwdziałających. W jaki sposób skupienie się wyłącznie na zaangażowaniu w witrynie internetowej może negatywnie wpłynąć na inne obszary biznesowe?
4. Utwórz jedną lub więcej wizualizacji dla interesariuszy pod kątem monitorowania kolumny `website_engagement`. Uwzględnij informacje o wszelkich punktach odniesienia, o których powinni wiedzieć, a także odpowiednio obserwuj wybrane wskaźniki przeciwdziałające wraz z głównym wskaźnikiem.
5. Korzystając z informacji bazowych, punktów odniesienia i cennych informacji zebranych w tym procesie, zaproponuj interesariuszom pierwszy *osiągalny* cel. W jakim stopniu powinni oni dążyć do zwiększenia zaangażowania w witrynie internetowej?

Podsumowanie

- *Wskaźniki* to znormalizowane miary ilościowe śledzone z upływem czasu. Są one często używane do monitorowania postępów, wyników i działań związanych z efektywnością firmy i jej zespołów.
- Wskaźniki wspierają podejmowanie decyzji na wielu poziomach. *Wskaźniki wydajności* to najszersza kategoria, której firma używa do zrozumienia postępów w realizacji celów. Przykłady obejmują przychody, wartość życiową klienta i efektywność operacyjną.

- *Wskaźniki strategii organizacyjnej* śledzą konkretne elementy efektywności firmy, takie jak zachowania związane z produktami, sentyment klientów i wyniki rynkowe. Wskaźniki z tej kategorii pomagają różnym zespołom w firmie zrozumieć środowisko, w którym działają, i podejmować odpowiednie decyzje na podstawie danych.
- *Wskaźniki odpowiedzialności* określają efektywność i produktywność poszczególnych osób i zespołów. Wskaźniki te umożliwiają *rozliczanie* pracowników z ich obowiązków i są podstawą ocen wydajności, premii lub szkoleń.
- *Ramy koncepcji SMART* to istotny przewodnik podczas definiowania skutecznych wskaźników. Aby zapewnić Twojej firmie możliwość monitorowania i podejmowania strategicznych decyzji opartych na danych, każdy jej wskaźnik powinien być *konkretny* (jasno zdefiniowany), *umożliwiać pomiar* (w sposób ilościowy), być *osiągalny* (realistyczny i w zasięgu), *istotny* (zgodny z celami firmy) i *ograniczony czasowo* (agregowany w odpowiednim przedziale czasu).
- Uświadomienie sobie tego, jak mają zostać wyznaczone cele dla danego wskaźnika, wymaga zebrania informacji z wielu źródeł, aby ustalić *punkt odniesienia* lub w podstawowym zakresie zrozumieć wskaźnik i jego trendy. Źródła te obejmują:
 - *wewnętrzne dane historyczne* firmy analizowane między segmentami, z upływem czasu i w odniesieniu do innych wskaźników biznesowych;
 - *punkty odniesienia lub wskaźniki porównawcze*, które często są pozyskiwane ze źródeł publicznych, takich jak dane rządowe, badania branżowe lub recenzowane publikacje naukowe;
 - *wstępne cele*, które pozwalają przetestować zdolność do wprowadzania zmian w danym wskaźniku i osiągania celów.
- Sposób *reprezentowania* wskaźnika jest równie ważny jak jego definiowanie. Istnieje wiele *pułapek* związanych z obliczaniem i przygotowywaniem wskaźnika, które mogą zmniejszyć wartość, jaką ma z niego firma. Oto kilka przykładów:
 - *Wskaźniki skumulowane* mogą stwarzać iluzję ciągłej poprawy mimo pogarszających się wyników. *Widoki przyrostowe* wskaźników prawie zawsze stanowią bardziej odpowiedni sposób zapewniania wyraźnego obrazu trendów dotyczących wydajności.
 - Poziom *szczegółowości* w używanym zbiorze danych może znacząco wpłynąć na możliwość interpretacji wskaźnika i jego odpowiedzialność. Wskaźniki mierzone i grupowane na niskim poziomie szczegółowości

(np. geograficznie według krajów) mogą ukrywać istotne detale, natomiast bardzo dokładny wskaźnik (np. na poziomie kodu pocztowego) może dezorientować interesariuszy nadmiarem szczegółów. Wybór odpowiedniego poziomu *szczegółowości* to sztuka zachowania równowagi, która wymaga zrozumienia podstawowych informacji.

- *Wskaźnik złożony* łączy co najmniej dwa wskaźniki w celu dokonania pomiaru skomplikowanych koncepcji. Wskaźniki te są zazwyczaj trudne do interpretacji i zastosowania w praktyce, a ponadto mogą maskować sprzeczne trendy wśród pojedynczych wskaźników. Warto w jak największym stopniu ograniczyć starania związane z ich tworzeniem.
- *Wskaźniki przeciwdziałające* chronią przed potencjalnymi negatywnymi skutkami skupiania się tylko na jednym wskaźniku wydajności. Każdemu wskaźnikowi powinny towarzyszyć wskaźniki przeciwdziałające w celu zapewnienia zrównoważonego podejścia Twojej firmy do wyznaczania celów strategicznych.
- *Wizualna reprezentacja* danych jest kluczowa w przekazywaniu informacji o wskaźniku i jego wydajności. *Niewłaściwy typ wykresu* może wprowadzić interesariuszy w błąd i odwrócić uwagę od istotnych trendów i różnic. *Zniekształcenie osi* na wykresach może stworzyć wrażenie pojawienia się znaczącego trendu tam, gdzie w rzeczywistości go nie ma, i odwrotnie. Typ wykresu i osie powinny zostać starannie dobrane i skonfigurowane pod kątem prezentacji wskaźnika.

PROGRAM PARTNERSKI

— GRUPY HELION —



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion

Nigdy wcześniej firmy nie miały dostępu do takiej ilości danych — i nigdy wcześniej tak bardzo nie potrzebowały ludzi, którzy potrafią je właściwie interpretować. W erze sztucznej inteligencji analiza danych stała się kluczowym narzędziem podejmowania decyzji biznesowych.

Ta książka to kompleksowy przewodnik po umiejętnościach twardych i miękkich niezbędnych do odniesienia sukcesu w tej dynamicznie rozwijającej się dziedzinie. Autorka dzieli się wiedzą zdobytą podczas ponad dziesięcioletniej kariery, pokazując, jak radzić sobie z rzeczywistymi wyzwaniami analitycznymi. Zamiast skupiać się wyłącznie na narzędziach i wzorach, opisuje praktyczne aspekty codziennej pracy analityka. Obejmuje przy tym cały proces: od definiowania pytań i metryk, przez dobór metod statystycznych, aż po skuteczną komunikację z interesariuszami. Omawia również zagadnienia związane z modelowaniem statystycznym, wykorzystaniem zewnętrznych źródeł danych, a także nowoczesnymi narzędziami analitycznymi, w tym zastosowaniem sztucznej inteligencji.

W książce:

- Kompleksowy cykl życia projektu analitycznego
- Parametryczne i nieparametryczne testy statystyczne
- Projektowanie rzetelnych wskaźników SMART
- Etyczne i prawne aspekty pracy z danymi wrażliwymi
- Modelowanie statystyczne
- Pozyskiwanie danych z API i web scrapingu
- Zastosowania generatywnej AI w pracy analityka

Miedzy liczbami a ludźmi. Dane, które kształtują biznes

Mona Khalil

jest starszą menedżerką działu inżynierii analitycznej w firmie Justworks, ma ponad dziesięcioletnie doświadczenie w analityce i danologii. Studiowała psychologię na Fordham University i statystykę w Baruch College. Jest entuzjastką udostępniania danych i wniosków wszystkim zainteresowanym, łączy podejście ukierunkowane na człowieka z zaawansowanymi technikami analitycznymi.

Helion 



helion.pl



HELION S.A.
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 96 63
helion@helion.pl

KOD KORZYŚCI
Sięgnij po więcej! ▶



ISBN 978-83-289-3306-4



9 788328 933064

Cena: 119,00 zł