1. ZROZUMIENIE I INTERPRETACJA DANYCH

Autor: Dario Šebalj

W tym rozdziale wyjaśniono podstawowe koncepcje rozumienia i interpretacji danych. Dane stanowią podstawę skutecznego podejmowania decyzji i odgrywają kluczową rolę w napędzaniu sukcesu organizacji. Dzięki biegłości w analizie i interpretacji danych, osoby (analitycy danych, menedżerowie, przedsiębiorcy, entuzjaści danych itp.) mogą nabyć umiejętności niezbędnych do odnajdywania cennych spostrzeżeń z ogromnych ilości dostępnych informacji. Dane zapewniają analizę z faktycznym fundamentem. Umożliwiają organizacjom podejmowanie decyzji, które są bardziej obiektywne i oparte na faktach, umożliwiając im wyjście poza założenia i intuicję. Organizacje mogą wykorzystywać dane do znajdowania korelacji, trendów i wzorców, których w przeciwnym razie mogłyby nie zauważyć.

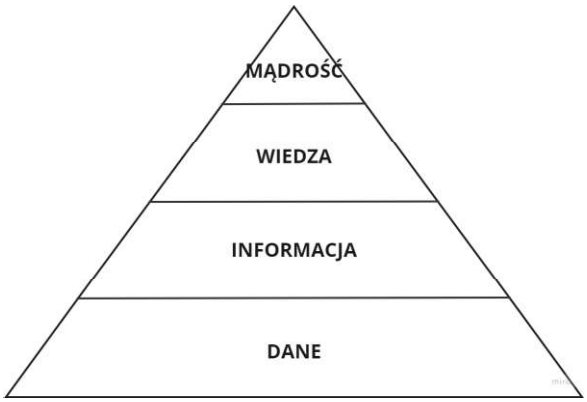
Znajdowanie nieefektywności organizacyjnych i obszarów do poprawy to kolejna korzyść analizy danych. Organizacje mogą znajdować wąskie gardła, usprawniać przepływy pracy i zwiększać ogólną wydajność procesów biznesowych poprzez ocenę danych operacyjnych. Ponadto analiza danych ułatwia ocenę sukcesu projektów lub strategii, które zostały wdrożone, co znacząco przyspiesza podejmowanie decyzji i usprawnia planowanie.

Procedura analizy danych musi być poprzedzona kluczowym opisaniem różnych typów i źródeł danych, omówieniem modelowania danych i podkreśleniem znaczenia jakości danych. Ten rozdział stworzy solidne podstawy Business Intelligence, zapewniając użytkownikom narzędzia do wykorzystania mocy płynącej z analizy danych i podejmowania świadomych decyzji.

1.1. Dane, informacje, wiedza, mądrość

Definiując termin dane, często jako punkt wyjścia stosuje się znaną piramidę danych-informacji-wiedzy-mądrości (DIWM) lub hierarchię DIWM. Rowley (2007) stwierdza, że hierarchia służy do identyfikowania i opisywania procesów zaangażowanych w transformację jednostki na niższym poziomie hierarchii (takiego jak dane) w jednostkę na wyższym poziomie hierarchii (takiego jak informacje), a także kontekstualizowania danych, informacji, wiedzy

i okazjonalnie mądrości w odniesieniu do siebie nawzajem. Każdy poziom piramidy opiera się na poziomie poniżej, a aby podejmowanie decyzji oparte na danych było skuteczne, wymagane są wszystkie cztery poziomy (Cotton, 2023). Rysunek 1.1 przedstawia piramidę DIKW.



Rysunek 1.1 Piramida DIWM

Źródło: Rowley (2007)

Dane stanowią surowy materiał, który nie ma znaczenia. To treść, którą można bezpośrednio zaobserwować lub zweryfikować, niezorganizowany fakt, który jest wyrwany z kontekstu i trudny do zrozumienia (Brackett, 2015; Dalkir, 2023). Dane mogą mieć formę liczb, tekstu, obrazów itp. Bez interpretacji dane pozostaną bezsensowne. Przykład danych: zbiór danych zawierający odczyty temperatury zebrane ze stacji pogodowych.

Zbiór danych w kontekście, który jest istotny dla jednej lub więcej osób w danym momencie lub przez określony czas, nazywa się **informacją**. Są to przetworzone, zorganizowane, ustrukturyzowane i kontekstualizowane dane. Odpowiedzi na typowe pytania, takie jak „kto", „co", „gdzie" i „kiedy" można znaleźć w informacjach (Brackett, 2015; Cotton, 2023). Przykład informacji: Analizując dane dotyczące temperatury, można zauważyć, że średnia temperatura w poprzednim miesiącu jest wyższa niż w tym samym okresie w zeszłym roku.

Cotton (2023) twierdzi, że **wiedza** jest wynikiem analizy i interpretacji informacji, która ujawnia wzorce, trendy i powiązania. Oferuje wgląd w to, „jak" i „dlaczego" mają miejsce określone zdarzenia. Pociąga to za sobą głębsze zrozumienie podstawowych idei. Chaffey i Wood (2005, cytowani w Rowley, 2007) definiują wiedzę jako „połączenie danych i informacji, do których dodano opinię ekspertów, umiejętności i doświadczenie, co skutkuje cennym zasobem, który może być wykorzystany do wspomagania podejmowania decyzji". Przykład wiedzy: Dzięki wiedzy uzyskanej z analizy historycznych danych dotyczących temperatury meteorolog może przewidzieć warunki pogodowe na nadchodzący tydzień.

Zdolność do zrozumienia podstawowych faktów i podejmowania świadomych decyzji oraz skutecznych działań jest znana jako **mądrość** (Cotton, 2023). Przykład mądrości: Korzystając z prognoz pogody i rozumiejąc lokalne warunki klimatyczne, rolnik może podjąć decyzję o zasianiu określonego rodzaju uprawy.

Zrozumienie wzajemnego oddziaływania danych, informacji, wiedzy i mądrości stanowi fundamentalną podstawę do wykorzystania potencjału Business Intelligence. W dalszej części tej książki zrozumienie to posłuży jako podstawa, w ramach której można wykorzystać moc danych, aby przekształcić je w praktyczne spostrzeżenia, kierując organizacje w stronę sukcesu w ciągle ewoluującym krajobrazie ery informacji.

1.2. Źródła danych i typy danych

Współcześnie dane są często określane jako „nowa ropa naftowa" – cenny zasób, który napędza innowacje i podejmowanie decyzji. W sercu każdego przedsięwzięcia opartego na danych leży spektrum typów danych, z których każde ma swoje unikalne cechy i znaczenie. Każdego dnia produkowana jest ogromna ilość danych. Obecne prognozy mówią, że na świecie jest 97 zettabajtów danych, a każdego dnia tworzonych jest ponad 2,5 kwintyliona bajtów danych. 90% danych na świecie wygenerowano w ciągu ostatnich dwóch lat (Marr, b.d.).

Według Kenetta & Shmueli (2016) „dane mogą pochodzić z różnych instrumentów gromadzenia: ankiet, testów laboratoryjnych, eksperymentów terenowych, eksperymentów komputerowych, symulacji, wyszukiwań w sieci, nagrań mobilnych itp. Dane mogą być pierwotne, zbierane na potrzeby badania, lub wtórne, zbierane z innego powodu. Dane mogą być jednowymiarowe lub wielowymiarowe, dyskretne, ciągłe lub mieszane. Dane mogą zawierać semantyczne niestrukturyzowane informacje w postaci tekstu, obrazów, dźwięku i wideo. Dane mogą mieć różne struktury, w tym dane przekrojowe, szeregi czasowe, dane panelowe, dane sieciowe, dane geograficzne itp. Dane mogą obejmować informacje z jednego źródła lub z wielu źródeł. Dane mogą mieć dowolny rozmiar i dowolny wymiar".

Dane są generowane szybciej i nieprzerwanie. Nowe dane są generowane i muszą być gdzieś przechowywane. Sieci społecznościowe, smartfony i technologia obrazowania wykorzystywana w diagnostyce medycznej to tylko kilka przykładów. Urządzenia i czujniki automatycznie tworzą dane diagnostyczne, które muszą być analizowane i przechowywane natychmiast. Utrzymanie tej ogromnej ilości danych jest wystarczająco trudne, ale ich analiza w celu zidentyfikowania trendów i wydobycia cennych informacji jest o wiele trudniejsza,

szczególnie gdy dane nie są zgodne z tradycyjnymi pojęciami struktury danych (EMC Education Services, 2015).

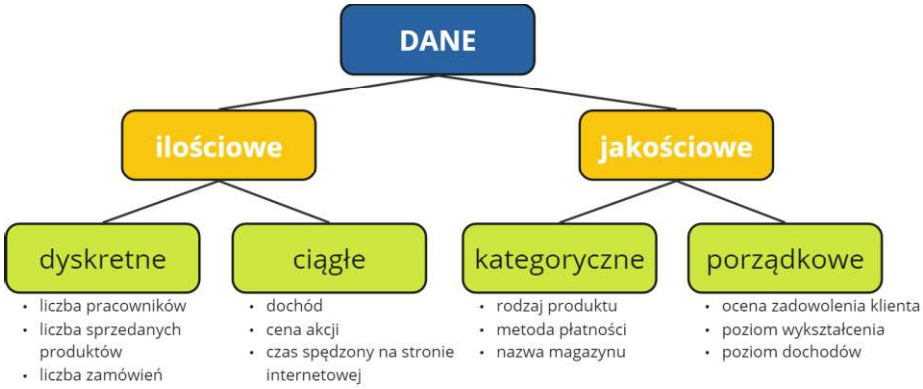
Według Blazqueza i Domenecha (2018) coraz częściej technologie związane z Internetem, smartfonami i inteligentnymi czujnikami są włączane do większości codziennych operacji biznesowych i osobistych. Na przykład wiele firm korzysta z mediów społecznościowych, aby promować swoje marki, oferować produkty online, używać smartfonów do śledzenia tras pokonywanych przez sprzedawców lub używać specjalistycznych czujników do rejestrowania pracy maszyn. Z drugiej strony ludzie używają komputerów, smartfonów i tabletów do przeglądania Internetu w poszukiwaniu produktów, komunikowania się ze znajomymi, dzielenia się opiniami i znajdowania drogi. Ponadto czujniki rozmieszczone w miastach, na autostradach i w miejscach publicznych, takich jak supermarkety, rejestrują codzienne ruchy i działania obywateli. Z tego powodu wszystkie te technologie produkują ogromną ilość nowych zdigitalizowanych i świeżych danych o działaniach ludzi i przedsiębiorstw. Po odpowiedniej analizie dane te mogą pomóc zidentyfikować trendy i śledzić skalę zachowań ekonomicznych, przemysłowych i społecznych. Dane te są ogromne i stale aktualizowane.

Sherman (2015) podkreśla, że problem może stanowić sytuacja, w której przedsiębiorstwo generuje więcej danych, niż jest w stanie zarządzać. Poprzez codzienne interakcje z klientami, partnerami i dostawcami firmy gromadzą ogromne ilości danych zarówno wewnętrznie, jak i zewnętrznie. Prowadzą badania rynku i śledzą informacje o swoich konkurentach. Witryny z kodami śledzącymi umożliwiają im śledzenie dokładnej liczby odwiedzających i ich pochodzenia. Przechowują i zarządzają danymi potrzebnymi do inicjatyw branżowych i wymogów rządowych. Obecnie istnieje Internet Rzeczy (IoT), który gromadzi dane z czujników wbudowanych w przedmioty ze świata rzeczywistego, takie jak obroże dla psów, rozruszniki serca i termostaty. To potop danych. Według EMC Education Services (2015) wśród źródeł Big Data o najszybszym tempie wzrostu znajdują się media społecznościowe i sekwencjonowanie genetyczne, które są przykładami wykorzystania niestandardowych źródeł danych do analizy.

Howson (2014) uważa, że sukces inicjatywy z zakresu Business Intelligence (BI) zależy od dostępności istotnych danych wysokiej jakości, obejmujących szeroki zakres źródeł danych niezbędnych do informowania procesów podejmowania decyzji.

Dane występują w różnych formach. Główne typy danych obejmują (Rysunek 1.2):

1. **Dane ilościowe (numeryczne)** – dane wyrażone w liczbach. Można je dalej podzielić na dane dyskretne i ciągłe. Dane dyskretne to dane, które mogą mieć tylko określoną wartość (np. liczba pracowników). Dane ciągłe mogą mieć nieskończoną liczbę możliwych wartości (np. cena produktu).
2. **Dane jakościowe** – ten typ danych można podzielić na dane kategoryczne i porządkowe. Dane kategoryczne reprezentują dane tekstowe, które można grupować w odrębne kategorie (np. miejsce urodzenia, region, kategoria produktu), podczas gdy dane porządkowe mogą być uszeregowane lub uporządkowane (np. satysfakcja klienta – bardzo niezadowolony, niezadowolony, neutralny, zadowolony, bardzo zadowolony; poziom wykształcenia – podstawowy, średni, licencjat, magister, doktorat).



Rysunek 1.2 Główne typy danych

Źródło: Opracowanie własne

Inną klasyfikacją danych stanowią dane ustrukturyzowane, półustrukturyzowane i nieustrukturyzowane.

Ustrukturyzowane dane mogą być przechowywane, przetwarzane i manipulowane w tradycyjnym systemie zarządzania relacyjną bazą danych. Dane te pochodzą z różnych źródeł, w tym strumieni kliknięć, formularzy internetowych, transakcji w punktach sprzedaży, czujników i maszyn. Mogą być wytwarzane przez ludzi lub maszyny. Mają one z góry określony format, rodzaj i organizację (EMC Education Services, 2015; Person & Porway, 2015).

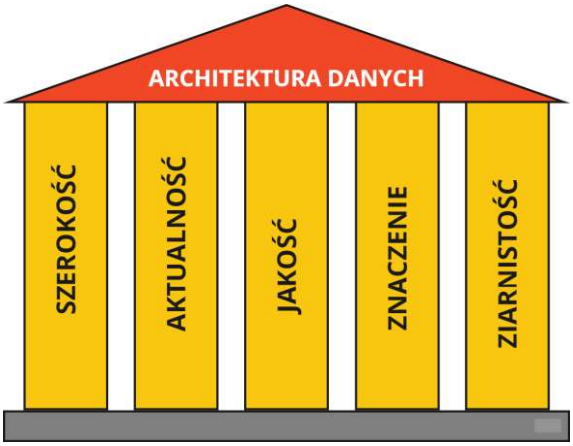
Dane półustrukturyzowane są organizowane za pomocą tagów, które pomagają nadać danym hierarchię i porządek, nawet jeśli nie pasują do ustrukturyzowanego systemu baz danych. Bazy danych i systemy plików często zawierają dane półustrukturyzowane. Pliki dziennika, tekst oznaczony tagami HTML, pliki XML i pliki danych JSON to niektóre z możliwych formatów przechowywania (McKinsey Global Institute, 2011; Person & Porway, 2015).

Ponieważ **niestrukturyzowane dane** są zazwyczaj wytwarzane przez działalność człowieka i nie mieszczą się w ustrukturyzowanym formacie bazy danych, są całkowicie niestrukturyzowane. Dokumenty tekstowe, pliki PDF, wpisy na blogu, wiadomości e-mail, zdjęcia i filmy są przykładami tego typu danych. (EMC Education Services, 2015; Person & Porway, 2015). Według Shermana (2015) niestrukturyzowane i półstrukturyzowane dane muszą być obsługiwane inaczej niż tradycyjne ustrukturyzowane dane.

Ta ogromna ilość ustrukturyzowanych i przede wszystkim nieustrukturyzowanych danych, które są generowane, gromadzone i przetwarzane z dużą prędkością i złożonością, nazywa się Big Data. McKinsey Global Institute (2011) definiuje Big Data jako „dane, których skala, dystrybucja, różnorodność i/lub terminowość wymagają użycia nowych architektur technicznych i analiz, aby umożliwić wgląd, który odblokowuje nowe źródła wartości biznesowej". Według Kitchena i McArdle'a (2016) w 2001 r. Doug Laney szczegółowo opisał, że Big Data charakteryzują trzy cechy (Trzy „V"):

- **objętość** (składająca się z ogromnych ilości danych),
- **prędkość** (tworzona w czasie rzeczywistym),
- **różnorodność** (ustrukturyzowana, półustrukturyzowana i nieustrukturyzowana).

Według Howsona (2014) podstawą udanej analizy biznesowej jest architektura danych (patrz Rysunek 1.3), która składa się z sześciu ważnych aspektów dotyczących danych: szerokości, terminowości, jakości, trafności i szczegółowości. Jakość danych jest filarem centralnym, ponieważ tak wiele wysiłku wkłada się w zapewnienie i poprawę jakości danych.



Rysunek 1.3 Architektura danych jako podstawa udanego BI

Źródło: opracowanie na podstawie Howson (2014)

Jak już wspomniano, dane mogą być zbierane z różnych źródeł. Jest to związane z szerokością danych jako jednym z filarów architektury danych. Odnosi się to do możliwości uzyskania wielu

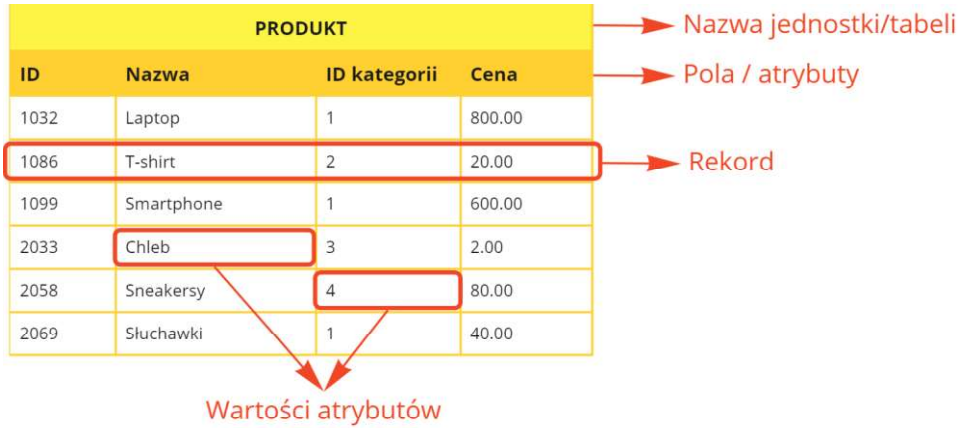
źródeł danych, co jest obecnie, w erze Big Data, powszechnym sposobem zbierania danych. Z drugiej strony łączenie danych z wielu różnych systemów źródłowych również przyczynia się do problemów z jakością danych (Howson, 2014).

W następnym podrozdziale uwaga czytelnika zostanie przesunięta z pojęć związanych z rozumieniem danych na pojęcia związanie z wykorzystaniem ich potencjału. Zostaną zbadane relacyjne systemy zarządzania bazami danych (RDBMS) do przechowywania i pobierania danych, wraz z wizualną reprezentacją struktur baz danych za pomocą diagramów encji-związków (ER). Łącząc nasze zrozumienie źródeł danych i typów danych z modelowaniem i projektowaniem danych, zostanie wykonany znaczący krok w kierunku praktycznego zastosowania Business Intelligence.

1.3. Modelowanie i projektowanie danych

Model danych to formalna reprezentacja danych, z których system biznesowy korzysta i które generuje. (Dennis i in., 2018). Jak już wspomniano, ustrukturyzowane dane są zwykle przechowywane w relacyjnym systemie zarządzania bazą danych (RDBMS). Według Tilleya (2020) „system zarządzania bazą danych to zbiór narzędzi, funkcji i interfejsów, który umożliwia użytkownikom dodawanie, aktualizowanie, zarządzanie, uzyskiwanie dostępu i analizowanie danych”. Niektóre popularne RDBMS to Oracle (Oracle), DB2 (IBM) i SQL Server (Microsoft).

W RDBMS dane są organizowane w tabelach, które zawierają zbiór rekordów przechowujących informacje o konkretnej jednostce. Tabele są reprezentowane jako dwuwymiarowe struktury z pionowymi kolumnami i poziomymi wierszami. Każda kolumna reprezentuje pole lub atrybut jednostki, podczas gdy każdy wiersz reprezentuje rekord, który jest wystąpieniem jednostki (Tilley, 2020). Rysunek 1.4 pokazuje przykład tabeli Product.



Rysunek 1.4 Przykład tabeli w RDBMS

Źródło: Opracowanie własne

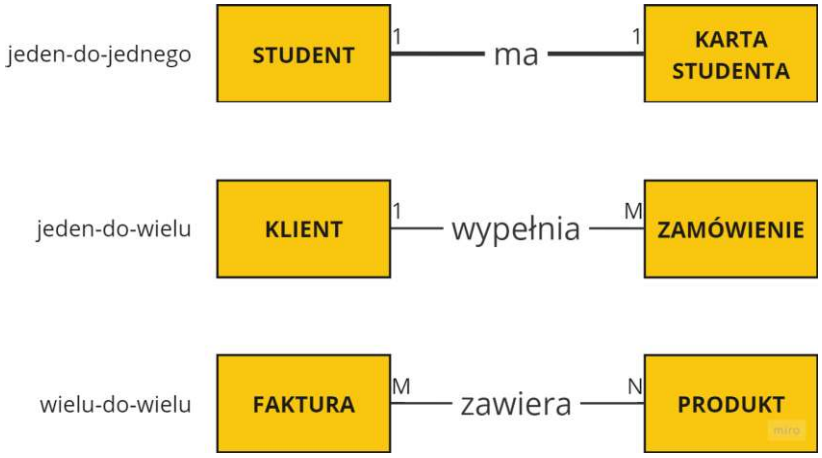
Tabela ta przedstawia prosty katalog produktowy zawierający 6 produktów, z których każdy ma unikalny identyfikator produktu, nazwę, kategorię i cenę.

Atrybut to szczególny rodzaj danych o jednostce. Na przykład identyfikator klienta, imię, nazwisko, adres, kod pocztowy, miasto, kraj, adres e-mail to atrybuty jednostki klienta. Wiersz w tabeli lub zbiór powiązanych pól opisujących pojedyncze wystąpienie jednostki (takiej jak klient) nazywa się **rekordem**.

Każda tabela w bazie danych musi mieć atrybut, który służy jako **klucz podstawowy**. Jest to pole (lub zestaw pól), które nadaje każdemu produktowi w tabeli odrębną tożsamość. Oznacza to, że w tabeli nie może być dwóch produktów o tym samym ID.

Tabele w bazie danych są często połączone z innymi tabelami w bazie danych, tzn. istnieje między nimi relacja. Relacje określają, w jaki sposób dane w jednej tabeli są powiązane z danymi w innej tabeli.

Według Tilleya (2020) między jednostkami mogą istnieć trzy typy relacji: jeden do jednego, jeden do wielu i wiele do wielu. Przykłady tych relacji pokazano na Rysunku 1.5.



Rysunek 1.5 Przykłady relacji pomiędzy jednostkami

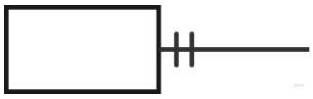
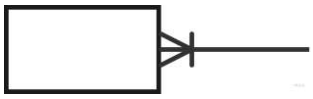
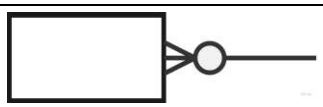
Źródło: Opracowanie własne

Logiczną strukturę bazy danych i relacje między tabelami można przedstawić wizualnie za pomocą **Diagramu Relacji Jednostki (ERD – ang. Entity Relationship Diagram)**.

Istnieją różne notacje do tworzenia ERD, z których najpowszechniejszymi są notacja Chen i notacja Martina („kruczej stopki"). W tej książce zostanie przedstawiona notacja Martina.

Zgodnie z notacją Martina („kruczej stopki"), jednostka (lub inaczej encja) jest reprezentowana przez prostokąt. Może to być osoba, miejsce, wydarzenie lub rzecz, na temat której zbierane są dane. Atrybuty są wymienione jako rzeczowniki w obrębie encji. Relacje między encjami są reprezentowane za pomocą linii, które łączą ze sobą encje. Relacje cechują się kardynalnością, która pokazuje, ile wystąpień jednej encji jest powiązanych z wystąpieniem innej (Dennis i in., 2018). W notacji „kruczej stopki" kardynalności są demonstrowane za pomocą różnych symboli. Na przykład pojedyncza kreska oznacza jeden, podwójna kreska oznacza jeden i tylko jeden, kółko oznacza zero, a „krucza stopka" oznacza wiele. Tabela 1.1. pokazuje różne symbole kardynalności i ich znaczenie.

Tabela 1.1 Przykłady kardynalności

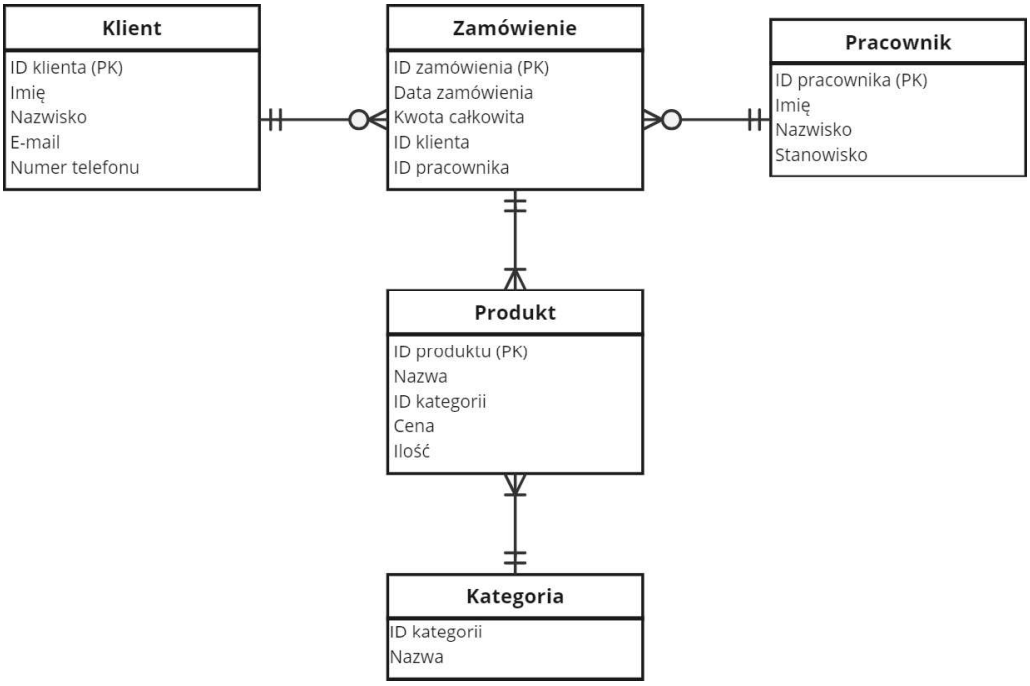
Symbol	Znaczenie
	Jeden i tylko jeden
	Jeden lub wiele
	Zero lub wiele

Symbol	Znaczenie
	Zero lub jeden

Źródło: Tilley (2020)

Według Dennisa i in. (2018) tworzenie ERD składa się z 3 etapów:

1. Zidentyfikuj jednostki.
2. Dodaj atrybuty i przypisz klucze podstawowe.
3. Zidentyfikuj relacje.



Rysunek 1.6 Przykład systemu sprzedaży ERD

Źródło: Opracowanie własne

Rysunek 1.6 przedstawia część systemu sprzedaży ERD. Każda jednostka w tym ERD jest przedstawiona jako prostokąt z listą jej atrybutów znajdującą się wewnątrz. Linie łączące jednostki służą do ukazania relacji jakie zachodzą pomiędzy nimi:

- Klient może złożyć wiele Zamówień. Każde zamówienie ma unikalnego klienta.
- Pojedyncze zamówienie może obejmować więcej niż jeden produkt. Każdy produkt jest powiązany z konkretnym zakupem.
- Zamówienie jest powiązane z Klientem i Pracownikiem, który je obsłużył.
- Produkt jest powiązany z Kategorią, ponieważ każdy produkt należy do pojedynczej kategorii.

Podstawowe zrozumienie modelowania i projektowania danych, jak pokazują ERD i RDBMS, jest niezbędne do efektywnego wykorzystania danych. To zrozumienie stanowi podstawę przejścia do pragmatycznego wdrażania danych poprzez podejmowanie decyzji opartych na danych, w którym systematycznie ustrukturyzowane modele danych dostarczają cennych spostrzeżeń motywujących świadome strategie biznesowe i procesy podejmowania decyzji.

1.4. Podejmowanie decyzji w oparciu o dane

Podejmowanie decyzji w oparciu o dane (DDDM – ang. Data-driven decision-making) to strategiczne podejście, które polega na analizowaniu i interpretowaniu danych w celu kierowania wyborami i działaniami. Wykorzystując spostrzeżenia pochodzące z ogromnych zestawów danych, firmy są w stanie precyzyjnie poruszać się w niepewnych obszarach, minimalizując w ten sposób ryzyko i maksymalizując możliwości. Nelson (2022) definiuje podejmowanie decyzji w oparciu o dane jako proces podejmowania strategicznych decyzji biznesowych, które współgrają z celami, zadaniami i inicjatywami organizacji, wykorzystując fakty, metryki i dane. Podejmowanie decyzji w oparciu o dane, według Provost & Fawcett (2013), to proces dokonywania wyborów, które opierają się bardziej na analizie danych niż na intuicji. Na przykład, specjalista ds. marketingu może wybierać reklamy, wykorzystując wyłącznie swoją rozległą wiedzę branżową i wyczucie tego, co spodoba się konsumentom. Alternatywnie, może oprzeć swoją decyzję na analizie danych pokazującej, jak klienci odbierają różne reklamy.

W tym podejściu decyzje nie są podejmowane na podstawie intuicji, ale twardych faktów. Proces decyzyjny polega na gromadzeniu, analizowaniu i interpretowaniu danych w celu identyfikacji wzorców, trendów i korelacji. Niezależnie od tego, czy chodzi o optymalizację wydajności operacyjnej, poprawę doświadczeń klientów czy udoskonalenie strategii produktów, decyzje oparte na danych umożliwiają firmom dostosowanie się do dynamicznie rozwijających się rynków.

Dzięki integracji danych z procesem podejmowania decyzji organizacje stają się bardziej elastyczne, responsywne i odporne na zmiany, wspierając tym samym innowacyjność i zrównoważony wzrost. Duża liczba prac badawczych wykazała, że podejmowanie decyzji w oparciu o dane wiąże się ze zwiększoną produktywnością (np. Brynjolfsson i McElheran, 2019; Sala i in., 2022; Colombari i in., 2023). Jest to powód, dla którego większość organizacji, zwłaszcza dużych, inwestuje w gromadzenie i analizowanie swoich danych. Ponad dwie trzecie

z ponad 300 dyrektorów objętych badaniem Bain & Company (2017) twierdzi, że ich firma inwestuje duże środki w analizę danych, podczas gdy ponad połowa przewiduje transformacyjne zwroty z inwestycji.

Istnieje pięć kroków podejmowania decyzji opartych na danych (Asana, 2022):

1. Zrozumienie wizji firmy.
2. Znajdowanie źródeł danych.
3. Czyszczenie i organizowanie danych.
4. Wykonywanie analizy danych.
5. Wyciąganie wniosków.

McKinsey Global Institute (2014) podaje, że organizacje oparte na danych mają 23 razy większą szansę na pozyskanie klientów, 6 razy większą szansę na utrzymanie klientów i 19 razy większą szansę na osiągnięcie rentowności.

Największe przedsiębiorstwa na świecie, które podejmują decyzje w oparciu o dane to Google, Amazon i Netflix.

Aby organizacja mogła w pełni wykorzystać potencjał Business Intelligence, kluczowe jest uwzględnienie jakości danych, która jest szczegółowo omówiona w poniższej sekcji. Jakość ta zapewnia precyzję i niezawodność wniosków oraz spostrzeżeń uzyskanych z danych.

1.5. Jakość danych

Stopień dokładności, spójności, niezawodności i przydatności do danego celu jest określany jako jakość danych. Jakość danych jest ważna w kontekście business intelligence i analizy danych, ponieważ wnioski i osądy wyciągnięte z danych zależą głównie od ich dokładności i niezawodności. Niska jakość danych może prowadzić do błędnych wniosków, wadliwych strategii i ostatecznie do szkodliwych wyników biznesowych. Według Gartnera (2021) każdego roku słaba jakość danych kosztuje organizacje średnio 12,9 miliona dolarów.

Jakość danych można scharakteryzować za pomocą sześciu najczęściej stosowanych wymiarów (Foote, 2022):

- **Dokładność:** jak dokładne są wartości atrybutów w danych?
- **Kompletność:** czy dane są kompletne, bez brakujących informacji?
- **Spójność:** jak spójne są wartości w bazach danych i między nimi?
- **Aktualność:** jak aktualne są dane?

- **Ważność:** w jaki sposób dane są zgodne z wstępnie zdefiniowanymi regułami biznesowymi?
- **Unikalność:** czy każdy rekord jest jednoznacznie zidentyfikowany, bez zbędnego przechowywania?

Według Shermana (2015) dane można uznać za wysokiej jakości, jeśli spełniają następujące kryteria (pięć „C" danych):

- **Czystość –** odnosi się do brakujących elementów, nieprawidłowych wpisów i innych podobnych problemów,
- **Spójność –** jednolitość i spójność danych w różnych źródłach i w samym zestawie danych,
- **Zgodność –** odnosi się do danych, które są zgodne z wstępnie zdefiniowanymi standardami i zasadami danych,
- **Aktualność –** niezbędne jest korzystanie z najnowszych danych do podejmowania decyzji i analiz,
- **Kompleksowość –** obejmuje wszystkie niezbędne elementy danych wymagane do zamierzonego procesu podejmowania decyzji, nie pomijając krytycznych informacji.

Według Kenetta i Shmueli (2016) niemal wszystkie dane muszą zostać oczyszczone, zanim będzie można ich użyć do dalszej analizy. Jednak cel determinuje stopień czystości i strategię czyszczenia danych. Wysokiej jakości informacje dla jednego celu i niskiej jakości informacje dla innego celu mogą znajdować się w tych samych danych.

Gartner (2023) niedawno zidentyfikował zestaw 12 działań mających na celu poprawę jakości danych. Działania te zostały sklasyfikowane w czterech odrębnych kategoriach, które należy wziąć pod uwagę przy ocenie integralności danych:

- Skup się na właściwych rzeczach, tworząc solidne fundamenty,
- Zastosuj odpowiedzialność za jakość danych,
- Ustal jakość danych „dostosowaną do celu",
- Zintegruj jakość danych z kulturą korporacyjną.

Według Howsona (2014) spójne, kompleksowe i dokładne dane są postrzegane jako mające wysoki stopień jakości. Trudno jest uzyskać dane wysokiej jakości, ponieważ problemy organizacyjne i własnościowe oddziałują na dane w bardzo dużym stopniu.

Umiejętność rozumienia i analizowania danych jest kluczowa dla podejmowania świadomych decyzji w dziedzinie business intelligence. Zrozumienie przejścia od surowych danych do

praktycznych spostrzeżeń, obejmujących różne źródła danych, typy, modelowanie i projektowanie, jest integralną częścią procesu czerpania wartości z informacji.

BIBLIOGRAFIA

1. Asana (2022). Data-driven decision making: A step-by-step guide [dostępne: https://asana.com/resources/data-driven-decision-making, dostęp 5 listopada 2023]
2. Bain & Company (2017). Closing the Results Gap in Advanced Analytics: Lessons from the Front Lines [dostępne: https://www.bain.com/insights/closing-the-results-gap-in-advanced-analytics-lessons-from-the-front-lines/, dostęp 5 listopada 2023]
3. Blazquez, D. & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. Technological Forecasting & Social Change, 130, pp. 99-113.
4. Brackett, M. (2015). The Data-Information-Knowledge Cycle. Dataversity [dostępne: https://www.dataversity.net/the-data-information-knowledge-cycle/, dostęp 5 listopada 2023]
5. Brynjolfsson, E. & McElheran, K. (2019). Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing. Rotman School of Management Working Paper No. 3422397 [dostępne: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3422397, dostęp 5 listopada 2023]
6. Chaffey, D. & Wood, S. (2005). Business Information Management: Improving Performance using Information Systems. FT Prentice Hall.
7. Colombari, R., Geuna, A., Helper, S., Martins, R., Paolucci, E., Ricci, R. & Seamans, R. (2023). International Journal of Production Economics, 255.
8. Cotton, R. (2023). The Data-Information-Knowledge-Wisdom Pyramid. Datacamp [dostępne: https://www.datacamp.com/cheat-sheet/the-data-information-knowledge-wisdom-pyramid, dostęp 5 listopada 2023]
9. Dalkir, K. (2023). Knowledge Management in Theory and Practice, 4th Edition. MIT Press.
10. Dennis, A., Wixom, B. H. & Roth, R. M. (2018). Systems Analysis and Design, 7th Edition. Wiley.
11. EMC Education Services (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley.
12. Foote, K. D. (2022). Data quality dimensions. Dataversity [dostępne: https://www.dataversity.net/data-quality-dimensions/, dostęp 5 listopada 2023]

13. Gartner (2021). How to Improve Your Data Quality [dostępne: https://www.gartner.com/smarterwithgartner/how-to-improve-your-data-quality, dostęp 5 listopada 2023]
14. Gartner (2023). Gartner Identifies 12 Actions to Improve Data Quality [dostępne: https://www.gartner.com/en/newsroom/press-releases/2023-05-22-gartner-identifies-12-actions-to-improve-data-quality, dostęp 5 listopada 2023]
15. Howson, C. (2014). Successful Business Intelligence: Unlock the Value of BI & Big Data, 2nd Edition. McGraw-Hill Education.
16. Kenett, R. S. & Shmueli, G. (2016). Information Quality: The Potential of Data and Analytics to Generate Knowledge. Wiley.
17. Kitchen, R. & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1).
18. Marr, B. (n.d.). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Bernard Marr & Co. [dostępne: https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/, dostęp 5 listopada 2023]
19. McKinsey Global Institute (2011). Big data: The next frontier for innovation, competition, and productivity. [dostępne: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_full_report.pdf, dostęp 5 listopada 2023]
20. McKinsey Global Institute (2014). Five facts: How customer analytics boosts corporate performance [dostępne: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/five-facts-how-customer-analytics-boosts-corporate-performance, dostęp 5 listopada 2023]
21. Nelson, M. (2022). Beyond The Buzzword: What Does Data-Driven Decision-Making Really Mean?. Forbes [dostępne: https://www.forbes.com/sites/tableau/2022/09/23/beyond-the-buzzword-what-does-data-driven-decision-making-really-mean/?sh=2d35c4eb25d6, dostęp 5 listopada 2023]
22. Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big data, 1(1), pp. 51-59.
23. Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science, 33(2), pp. 163–180.

24. Sala, R., Pirola, F., Pezzotta, G. & Cavalieri, S. (2022). Data-Driven Decision Making in Maintenance Service Delivery Process: A Case Study. Applied Sciences, 12(15).
25. Sherman, R. (2015). Business Intelligence Guidebook: From Data Integration to Analytics. Elsevier Inc.
26. Tilley, S. (2020). Systems Analysis and Design, 12th Edition. Cengage.