

#### 1. ŁAńCUCH DANYCH

Kluczem do zrozumienia sposobu pracy z danymi jest przepływ danych lub **łańcuch danych**. Ogólnie są one postrzegane jako sekwencyjne. Niektóre osoby wolą korzystać z wizualizacji cyklu życia, ma to pewne zalety. W obu przypadkach powinieneś przyjąć holistyczny pogląd: nawet gdy poruszasz się krok po kroku, zawsze potrzebujesz jasnego pojęcia, gdzie chcesz zakończyć, zanim zaczniesz.

#### 2. KOLEKCJONOWANIE DANYCH: ŹRÓDŁA POCHODZENIA

Dane mogą pochodzić zarówno ze źródeł wewnętrznych, jak i zewnętrznych.

Jakie korzyści wynikają z każdego z tych źródeł?

- Wewnętrzne źródła danych, gdzie dane tworzone są przez firmę, są zwykle łatwiejsze w kolekcjonowaniu i mogą być bardziej przydatne firmie w zakresie jej własnych celów<sup>1</sup>. Są one darmowe. Firma może uruchamiać na ich podstawie duże inicjatywy nie wyglądając nawet poza własne ściany.
- Dane zewnętrzne pozyskiwane są spoza firmy.
  - Dane osób trzecich - mogą być adekwatne dla twojego rynku lub klientów, ale trudne do zdobycia i/lub drogie.
  - Otwarte dane - np. statystyki rządowe, dane dotyczące wpływu społecznego. Mogą mieć dużą wartość, ale często wymagają większych inwestycji w analizę, aby zapewnić przydatny wgląd.
  - Dane generowane zbiorowo. O wiele więcej firm bierze udział w spotkaniach i tzw. hackathonach.

#### Studium przypadku – czy zewnętrzne dane są warte swojej ceny?

- 7 małych firm w Irlandii Północnej sprzedających przetworzone produkty spożywcze uzyskało dostęp do danych dużej sieci supermarketów w takich obszarach, jak etap życia konsumenta i styl życia, analiza koszyka rynkowego i najlepiej funkcjonujące sklepy sprzedające produkty małych firm.
- Właściciele i menedżerowie brali udział w warsztatach, żeby dowiedzieć się, jak wyszukiwać i analizować najistotniejsze dane z bazy danych, aby odpowiedzieć na pytania takie jak „Jak radzi sobie moja kategoria?” „Jaki jest najbardziej popularny smak chleba?” „Jaki typ konsumentów kupuje produkt podobny do mojego?”
- Producent jogurtów nauczył się, że dorośli są kluczowym rynkiem, więc kiedy odwiedzali supermarkety z degustacją w sklepie, nie próbowali już zachęcać młodszych klientów, a zamiast tego skupiali się na osobach starszych. Taktyka poprawiła wydajność wydarzeń.
- Inne firmy były w stanie przewidzieć innowacje dalekiego zasięgu, zamiast po prostu patrzeć na inne produkty na rynku i próbować je naśladować lub postępować zgodnie z wytycznymi określonymi przez dużych nabywców detalicznych.

---

<sup>1</sup> <http://www.aanalytics.com/understanding-analytics-part-1-top-internal-sources-of-big-data/>

- Dane wzmocniły naturalny charakter przedsiębiorczości firm. Miejsca pracy stały się bardziej kolegalne: większość właścicieli-menedżerów udostępniła informacje o swojej firmie i zachęciła pracowników do zaangażowania się i oferowania nowych pomysłów.<sup>2</sup>

Właściciele małych firm napotykały przeszkody w inwestowaniu w dane zewnętrzne. Bariery obejmują koszty, brak specjalistycznej wiedzy lub wiary we własną wiedzę rynkową. Po uzyskaniu dostępu do danych większość firm szybko sformalizowała swoje podejście do planowania marketingowego.

W tym przykładzie małe firmy miały bezpłatny dostęp do danych zewnętrznych. Ile mogły by zapłacić za taki dostęp, aby nadal był opłacalny?

### KOLEKCJONOWANIE DANYCH: RODZAJE DANYCH

Dane są prezentowane jako ZESTAW DANYCH. **Zestaw danych** to grupa informacji, które są ze sobą powiązane.

- Zestawy danych mogą mieć charakter ilościowy lub jakościowy. Różnica wydaje się wyraźna, ale może istnieć wiele sposobów ich interakcji. Dane jakościowe mogą być trudne do dokładnego zmierzenia i analizy. Dane mogą mieć postać opisowych słów, które można zbadać pod kątem wzorców lub znaczenia, czasami za pomocą kodowania. Kodowanie pozwala badaczowi kategoryzować dane jakościowe w celu zidentyfikowania tematów odpowiadających pytaniom badawczym i przeprowadzenia analizy ilościowej.
- Dane mogą mieć charakter oddzielny lub ciągły.
- Możemy wyróżnić dane:
  - **Nominalne:** dane nominalne są danymi jakościowymi; do danych nie przypisano żadnej skali ani wartości.
  - **Porządkowe:** dane posiadające porządek, takie jak skala Likerta, gdzie 1 oznacza „nienawidzę!”, a 5 oznacza „uwielbiam to!”. Pamiętaj tylko, że różnica między rzędnymi (np. między 1 a 2 lub między 3 a 4) niekoniecznie jest równa. Często odległość między skrajnymi lubieniem i nienubieniem jest większa niż między odczuwaniem neutralności lub opinią nieznacznie pozytywną albo negatywną.
  - **Interwałowe:** dane, które mają naturalny porządek, a odległość między każdą wartością jest równa (np. temperatura).
  - **Ratio:** dane o naturalnym porządku, równej odległości między wartościami i naturalnym punkcie zerowym.<sup>3</sup>

### 3. MAGAZYNOWANIE DANYCH

Planując gromadzenie danych, musisz myśleć z wyprzedzeniem. Zastanów się, jak wykorzystasz gromadzone dane i jakie decyzje chcesz na ich podstawie podjąć. Chcesz uzyskać informacje o sytuacji w czasie rzeczywistym? Czy chcesz zrozumieć przyczynowość? A może chcesz porównać dwie zmienne i wybrać najlepszą?

Przechowywanie na miejscu vs przechowywanie na odległość

- Tradycyjnie dane przechowywane były na miejscu, ale rosnąca ich ilość i malejące koszty usług świadczonych przez firmy zajmujące się magazynowaniem danych czynią magazynowanie danych w chmurze dużo bardziej atrakcyjnym.

<sup>2</sup> <https://www.conference-board.org/blog/post.cfm?post=2526>

<sup>3</sup> <https://communitymedicine4asses.com/2013/01/13/scales-of-measurement-nominal-ordinal-interval-ratio/>

- Rynek pamięci masowej w chmurze jest zdominowany przez Amazon Web Services, Google i Microsoft Azure.
- Nie wszystkie chmury pamięci są sobie równe. Niektóre chmury są zoptymalizowane do obsługi archiwizacji, inne mają wydajność i stabilność niezbędną do tego by działać jako zaplecze dla podstawowego systemu przechowywania danych, a jeszcze inne nie są warte żadnego ryzyka. Chmura o najniższej cenie może w dłuższej perspektywie okazać się znacznie droższa, jeśli dane zostaną utracone lub staną się nieosiągalne.
- Wiele rozwiązań zapewnia elastyczność wyboru miejsca przechowywania danych: na miejscu i/lub w chmurze.<sup>4</sup>

### **Jak wybrać metodę przechowywania danych?**

Jeśli chodzi o przechowywanie danych, nie ma jednego rozwiązania idealnego dla każdego odbiorcy. Firmy muszą rozumieć ilość i rodzaj danych, które posiadają, oraz przyczyny przechowywania informacji.

Kto potrzebuje dostępu do jakiego rodzaju informacji? Jak często i jak szybko potrzebny jest nam dostęp do danych?

- Podczas konfigurowania procesów w firmie należy zidentyfikować najważniejsze dane organizacji i odpowiednio uszeregować zasoby zarządzania pamięcią masową. E-mail może być najwyższym priorytetem firmy, ale przechowywanie i archiwizowanie danych e-mail dla jednej konkretnej grupy, np. kierownictwa, może być ważniejsze niż w przypadku innych grup.
- „Oszczędzaj pieniądze, wykorzystując tylko najszybszą pamięć masową, taką jak SSD, do danych, których aktywnie używasz, i korzystaj z tańszych platform, takich jak chmura, do przechowywania danych archiwalnych lub zapasowych”.

Jaki rodzaj danych musimy przechowywać?

- Czy będziemy mogli połączyć ze sobą różne zestawy danych w przyszłości?

Jak wiele danych mamy teraz o ile zwiększy się ich ilość w przyszłości?

- Nie przechowuj zbędnych danych. Wiele firm przechwytuje i przechowuje zbędne dane. To był mniejszy problem, gdy dysk był drogi, ale obecnie, w dobie niskich cen, mamy tendencję do przechowywania wszystkiego.

Jak długo powinniśmy przechowywać dane?

- Polityka przechowywania danych jest niezbędna zarówno dla wewnętrznego zarządzania danymi, jak i zgodności z prawem. Niektóre dane muszą być przechowywane przez wiele lat, podczas gdy inne mogą być potrzebne tylko przez kilka dni.

JAK BEZPIECZNE muszą być dane? Jakie wymagania prawne należy spełnić?

- Przygotuj plan odzyskiwania w razie ewentualnej awarii. Niezależnie od wybranej metody lub metod tworzenia kopii zapasowych, należy je przetestować i upewnić się, że można przy ich pomocy odzyskać dane.

## **4. ANALIZA DANYCH**

Aby odpowiedzieć na strategiczne pytania przewodnie i lepiej zrozumieć swoją działalność, musisz wziąć surowe dane i przekształcić je za pomocą kombinacji podstawowych statystyk arytmetycznych

---

<sup>4</sup> <https://techcloudlink.com/14-things-you-need-to-know-about-data-storage-management/>

i opisowych w celu uzyskania wzorców - zarówno tych, które spodziewasz się zobaczyć, jak i tych, które mogą być cię zaskoczyć.<sup>5</sup>

Na co zwracać uwagę: fakty, statystyki, trendy są ważne, choć każde z nich z innego powodu.

Uwaga: mówiąc o wizualizacji danych możemy ją postrzegać jako CZĘŚĆ procesu analitycznego i oddzielnie, jako środek do prezentacji wyników. Te dwie formy są ze sobą nierozzerwalnie powiązane i zazwyczaj jedno może wspierać drugie.

**Przykłady wykresu sprzedaży i rentowności.** Jest to często typowy wynik analizy danych. (Źródło: Forbes)

Wykres sprzedaży i zyskowności autorstwa Ann Jackson jest wizualnie atrakcyjnym, przejrzystym podsumowaniem danych, które pokazuje zmiany w czasie, różnice geograficzne, straty dla niektórych kategorii produktów i podsumowuje kluczowe wskaźniki wydajności w postaci liczb. Czy Ann przekaże ten wykres swoim interesariuszom i uzna sprawę za zamkniętą? Nie, ponieważ dopiero teraz zaczyna się prawdziwa dyskusja. Ann może usiąść z publicznością i zagłębić się w dalsze szczegóły, aby dowiedzieć się, dlaczego osiągnięto takie wyniki i zidentyfikować możliwości poprawy wyników biznesowych. Dalsze badanie danych, eksploracja ich z odbiorcami – to jest wartość, którą dodajesz jako analityk poza samą produkcją wykresów i zestawień. Te działania dają ci doskonałą podstawę do dyskusji, ale nie powinny być punktem końcowym.<sup>6</sup>

## 5. INTERPRETACJA I WYKORZYSTANIE DANYCH

Posiadanie danych to nie wszystko. Analizowanie danych to jeszcze nie koniec. Konieczne jest przełożenie spostrzeżeń uzyskanych z danych na działania wspierające rozwój firmy.

Podobnie jak w przypadku każdej dobrej strategii, wiedza o tym, kiedy i jak ją zastosować, kiedy ją zmienić i jak zmierzyć jej sukces, ma kluczowe znaczenie dla organizacji, aby mogła ona przejść od organizacji posiadającej dane do organizacji napędzanej danymi.

Cykl korzystania z danych trwa - nie zatrzymuje się, gdy prezentowane są informacje, ale wymaga szczegółowej dyskusji. Analiza danych jest punktem początkowym, a nie końcowym.

### Na którym wykresie różnica jest większa?

Wykorzystanie danych wynika z nauki, ale jest to również sztuka. Dane są obiektywne, ale mogą być prezentowane subiektywnie. Biorąc pod uwagę, że wizualizacja danych jest bardzo ważna dla nadania sensu dużym zestawom danych, ale też dla podejmowania decyzji, ważne jest, aby zrozumieć, jak sobie z tym zagadnieniem poradzić.

Slajd 1: Na pierwszym wykresie wygląda na to, że istnieje duża różnica.

Slajd 2: Na tym wykresie widzimy różnicę między 35,5% a 41,5% - ta sama różnica. To pokazuje że:

- wykresy mogą wprowadzać w błąd
- ważne jest przekazywanie danych w formatach, które są dokładne i w rozsądnej skali.
- Gdy informacje są dodawane w skali, stają się znacznie wyraźniejsze.

<sup>5</sup> <https://www.schusterman.org/playbooks/data/making-meaning/find-patterns-stats/>

<sup>6</sup> <https://www.forbes.com/sites/evamurray/2019/03/29/what-is-the-difference-between-data-analysis-and-data-visualization/>

### **Który sposób prezentowania danych jest najbardziej skuteczny?**

Musisz pomyśleć o tym, w jaki sposób przekazujesz dane ORAZ jak przekazywać je precyzyjnie, aby Twoja wiadomość została zrozumiana.

W tym przykładzie kolorowy wykres jest najłatwiejszy do zrozumienia. Co się zmieniło? Niebieska mapa została „odszyfrowana” - oś i etykiety zostały usunięte, a puste miejsce zmniejszone. Następnie „podkreślono” najistotniejsze szczegóły – dodano kolory, aby odzwierciedlić ciąg odpowiedzi, a liczby dodano, aby pokazać wartości względne.

Wniosek:

Za każdym razem, gdy używasz danych do przekazania czegoś, zasadniczo opowiadasz historię. Może to być kwartalna prezentacja dla Twojej rady, formalne sprawozdanie z oceny programu, roczne sprawozdanie dla wyborców.

Opowiadanie o danych ma swoją moc i wiąże się z kwestiami etycznymi. Danymi można łatwo manipulować, aby opowiedzieć historię, której nie ma, lub zminimalizować historię.<sup>7</sup> Ważne jest, aby zacząć od danych i na ich podstawie opowiedzieć historię, zamiast tworzyć historię z wyprzedzeniem i wybierać tylko te dane, które do niej pasują. Na przykład, jeśli przedstawisz podsumowanie oceny programu, które podkreśla tylko sukcesy i aspekty usługi, które ludzie lubili, i nie obejmuje elementów, które się nie sprawdziły i ujawniły obszary wymagające poprawy, nie opowiadasz całej historii.

Projekt został zrealizowany przy wsparciu finansowym Komisji Europejskiej.

Niniejsza publikacja odzwierciedla jedynie poglądy autora, a Komisja nie ponosi odpowiedzialności za jakiegokolwiek wykorzystanie zawartych w niej informacji

---

<sup>7</sup> <http://www.bdbanalytics.ir/media/1123/storytelling-with-data-cole-nussbaumer-knafllic.pdf>