

# Moduł 8

## Analiza danych

### iBigWorld: Innovations for Big Data in a Real World

TSNUK team

**Disclaimer:** Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the National Agency (NA). Neither the European Union nor NA can be held responsible for them.



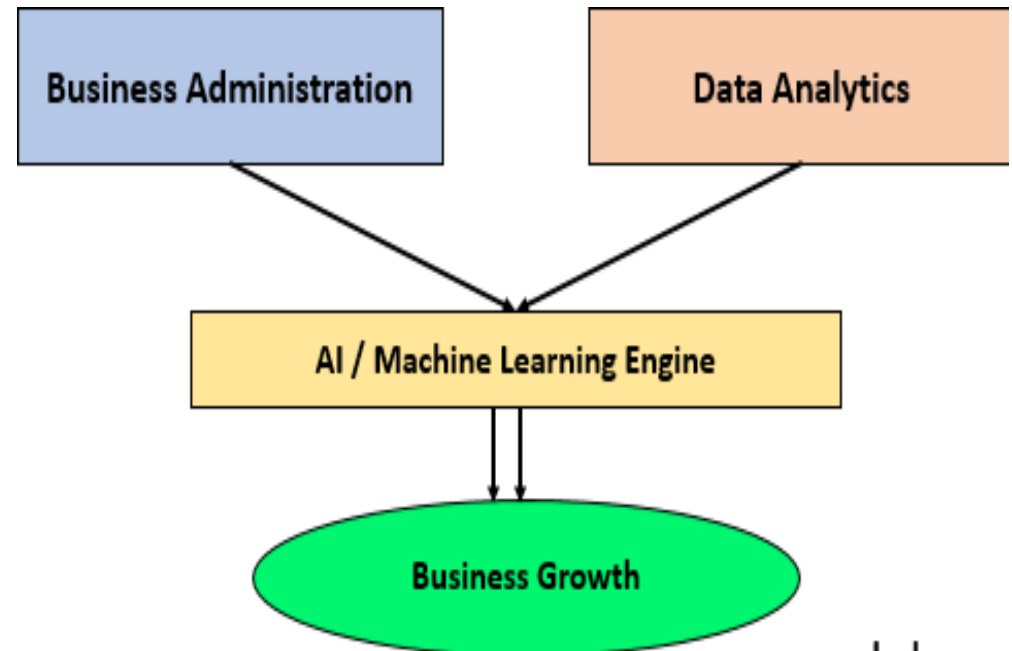
# Spis treści

## M8\_2. Analiza danych

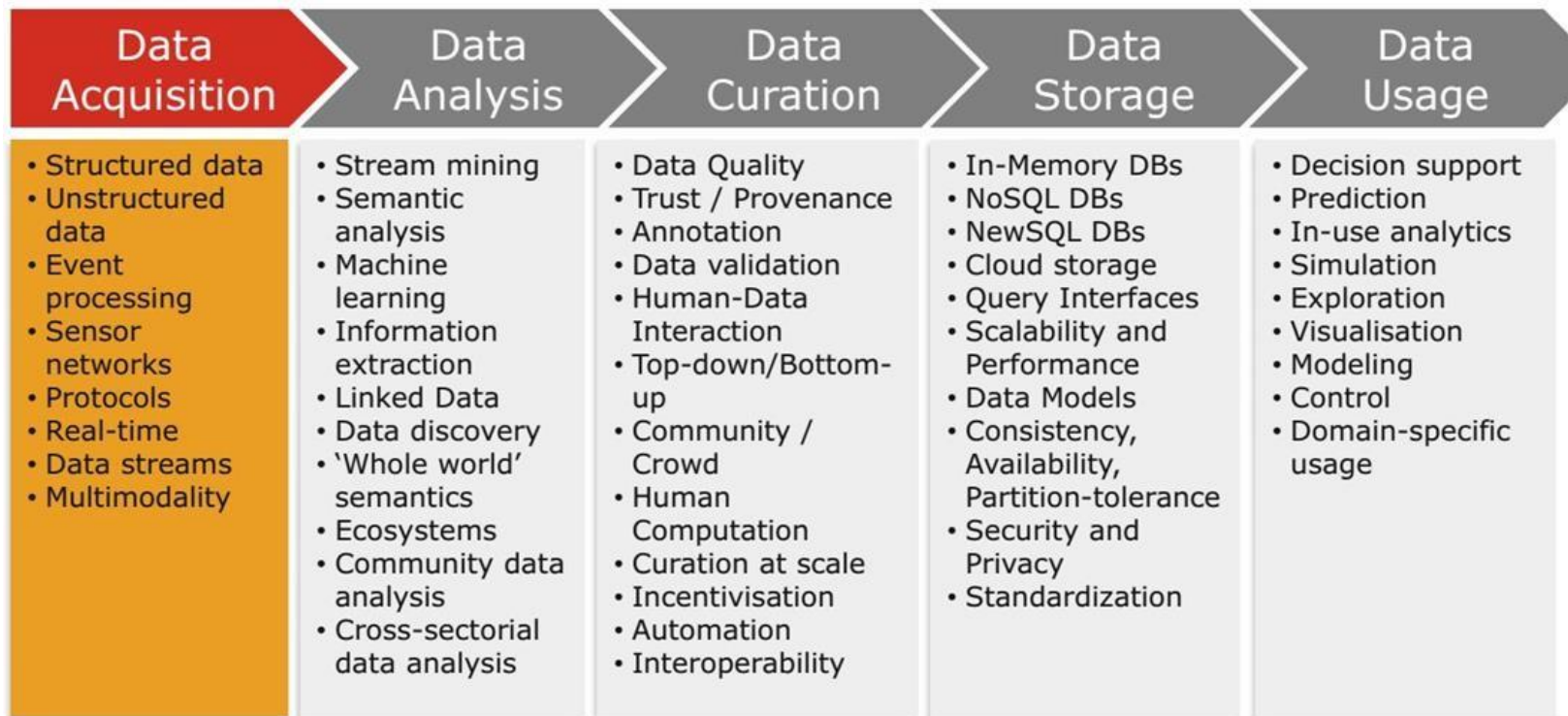
1. Cele tej części
2. Duże dane
3. Podstawowe etapy analizy danych
4. Rodzaje analizy danych
5. Wykorzystane popularne narzędzia
6. Proces analizy danych
7. Metody analizy danych wtórnych
8. Zalety stosowania analizy danych wtórnych
9. Przyszłe wymagania i nowe trendy w analizie dużych danych

## Cele tej części

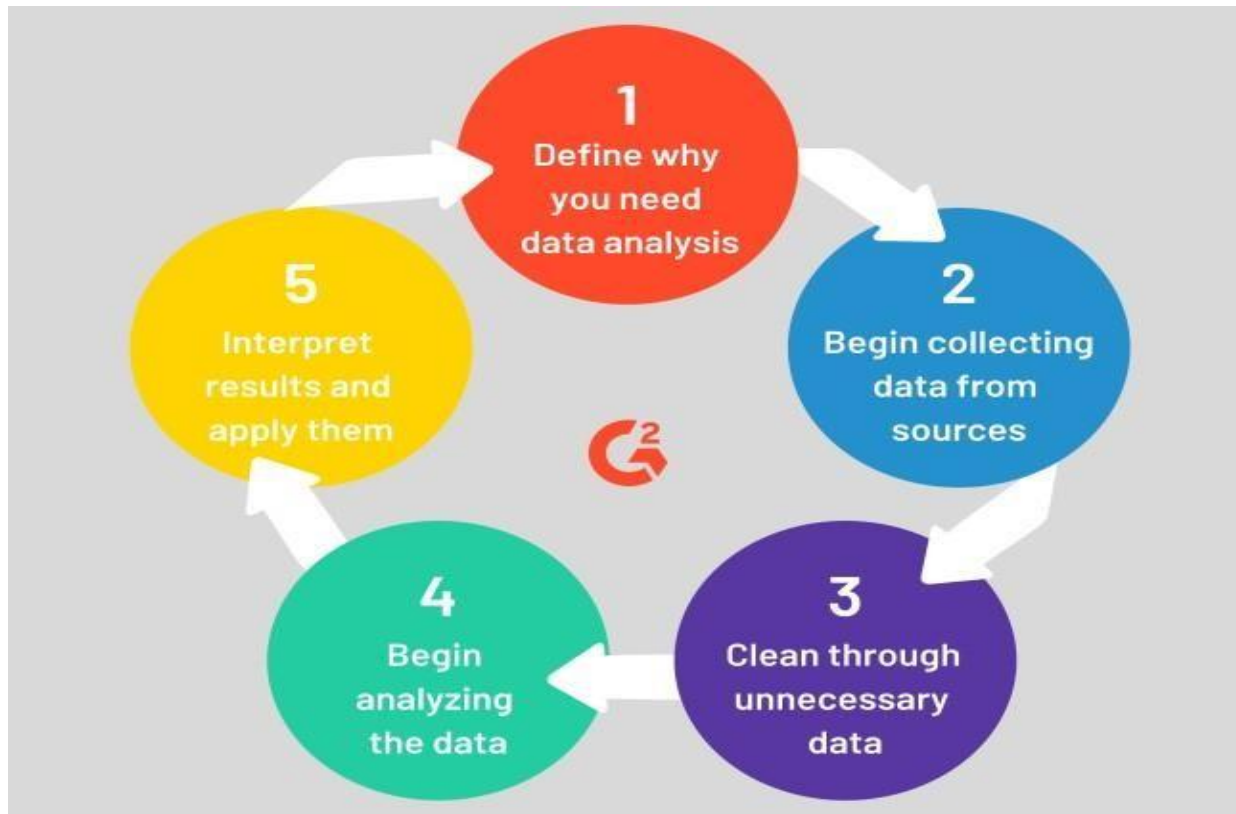
- Zbieraj ukryte informacje.
- Generuj raporty na podstawie dostępnych danych.
- Przeprowadź analizę rynku.
- Poprawić strategię biznesowe.



# Łańcuch wartości Big Data



# Główne etapy analizy danych



# Główne etapy analizy danych

- Krok 1: Głównym zadaniem będzie tutaj profilowanie danych.
- Krok 2: Następnie musimy przechowywać te dane przy użyciu dowolnej nieustrukturyzowanej metody przechowywania danych.
- Krok 3. Kolejnym krokiem jest zbudowanie modelu po kategoryzacji i pogrupowaniu danych. Po przygotowaniu modelu danych system zacznie otrzymywać informacje.
- Krok 4: Gdy dane zaczną napływać, różne dane, takie jak dane dotyczące interakcji i dane transakcyjne, mogą być skorelowane i przetworzone w celu ustalenia wzorca

# Rodzaje analizy danych

**Analiza opisowa**

**Analiza  
diagnostyczna**

**Analiza  
danych**

**Analiza  
predykcyjna**

**Analiza nakazowa**

# Używane popularne narzędzia

- **Tableau:** Możliwość tworzenia wizualizacji danych, pulpitu nawigacyjnego i raportu analitycznego po połączeniu z różnymi źródłami danych. To narzędzie działa z danymi nieustrukturyzowanymi, więc jest kompatybilne z Big Data.
- **Power BI:** kiedyś był rozszerzeniem MS Excel, później stał się samodzielny narzędziem. Jest lekki i często aktualizowany.
- **R i Python:** Jeśli lubisz niestandardowe kodowanie i dostosowywanie, R i Python to opcja dla Ciebie. R lepiej nadaje się do analizy statystycznej, podczas gdy Python ma wbudowane biblioteki analizy danych.
- **Apache Spark:** Apache Spark to szybki, lekki i wielkoskalowy procesor danych, który wykonuje dane z dużych klastrów danych i może szybko przetwarzać ogromne porcje danych.

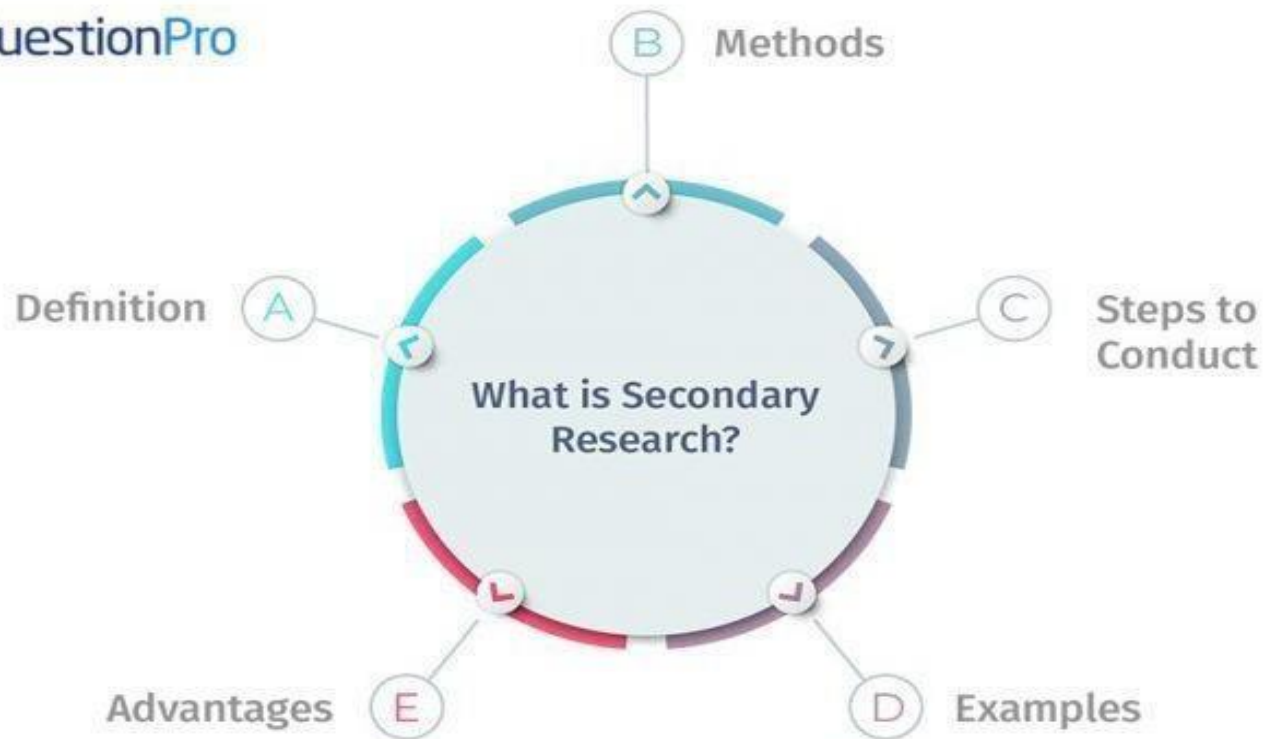


# Proces analizy danych

---

1. Zrozumienie biznesu
2. Uzyskaj surowe dane
3. Wyodrębnienie danych
4. Transformacja danych
5. Wizualizacja danych
6. Analiza statystyczna
7. Opracowanie modelu danych
8. Rekomendacje/Raport/Historia

# Analiza danych wtórna



# Metodyki wtórnej analizy danych. Kroki stosowane przez badaczy

- **Cel badań:** ...
- **Integralność danych:** należy zweryfikować uprawnienia badacza, który zebrał dane do innych celów. Należy również wziąć pod uwagę kwestie takie jak spójność danych i powiązane kwestie. Badacze muszą wiedzieć, dlaczego prowadzą badania na podstawie gromadzonych danych.
- **Śledzenie** danych: badacze mogą korzystać Internet w celu uzyskania dostępu do najnowocześniejszych technik gromadzenia danych w innym celu.
- **Znaczenie** danych: Badacze muszą znać podstawowe procesy stosowane przy gromadzeniu danych, takie jak wypełnianie strategii próbkowania, kiedy i jak zostały zebrane.
- **Analiza** danych: do wyciągania wniosków i odpowiednich wzorców można wykorzystać różne procesy statystyczne.



# Etapy wtórnej analizy danych

- Dane **dokumentacyjne**: są wykorzystywane przez badaczy do realizacji projektów wykorzystujących metody gromadzenia danych pierwotnych. Korzysta z następujących rodzajów dokumentów, tj. dokumentów pisemnych i niepisemnych wymienionych poniżej:
  - Dokumenty **pisemne**: są to protokoły z posiedzeń, sprawozdanie wspólników, transkrypcje przemówień, korespondencja, wiadomości itp.
  - Dokumenty **niepisane**: wykorzystuje cyfrowy format kodowania, taki jak nagrania wideo i kasetowe, programy telewizyjne i filmowe, płyty CD/DVD itp.

# Dane wtórne na podstawie ankiety

- **Spis ludności:** używany przez rząd jako badanie populacji lub frekwencja wyborców, ale uczestnictwo jest opcjonalne. Dostarcza danych porównawczych i kontekstowych do celów badawczych.
- **Przebiegi ciągłe i regularne:** Eliminuje powyższą metodę i powtarza się przez pewien okres czasu.
- **Ankieta specjalna:** nie odbywa się regularnie i jest raczej jednorazowa. Dotyczy to dokładniej określonej dziedziny badań.
- **Dane wtórne z wielu źródeł:** są to dane dokumentacyjne, dane ankietowe lub kombinacja tych dwóch.

# Zalety korzystania z wtórnej analizy danych

- **Rentowność:** Wymagana jest mniejsza inwestycja pieniędzy, czasu lub wysiłku, ponieważ dane są zbierane wcześniej przez innego badacza.
- **Zredukowany zakres:** Wtórna analiza danych pozwala badaczowi na zebranie określonego podzbioru danych zamiast zbierania dużych ilości danych.
- **Poziom profesjonalizmu:** Utrzymuje pewien poziom wiedzy eksperckiej i profesjonalizmu, co może nie być prawdziwe w przypadku osób prowadzących niezależne badania.
- **Nie rzucająca się w oczy:** Nie zakłóca ani nie zakłóca trwających badań prowadzonych przez naukowców.
- **Typ danych:** zapewnia naukowcom dane kontekstowe i porównawcze.
- **Odkrycia naukowe:** może to prowadzić do nieoczekiwanych odkryć i przełomów naukowych, ponieważ wcześniej przeanalizowane dane do innego celu są wykorzystywane jako punkt odniesienia.
- **Spójność danych:** zapewnia spójność danych analizowanych dla określonych celów badawczych.

# Przyszłe wymagania i pojawiające się trendy w analityce big data

---

Celem jest przekształcenie danych w informacje, a informacji w zrozumienie. Oto 7 kluczowych trendów w danych, które dziś skupiają uwagę.

- Analityka predykcyjna i nakazowa
- Sztuczna inteligencja
- Zarządzanie danymi
- Analiza chmury
- Wbudowana analityka
- Analiza wizualna
- Wspólna analityka

# Literatura

---

1. José María Cavanillas, Edward Curry, Wolfgang Wahlster. New Horizons for a Data-Driven Economy. 2016.  
<https://link.springer.com/content/pdf/10.1007/978-3-319-21569-3.pdf>
2. Will Hillier. A Step-by-Step Guide to the Data Analysis Process.  
<https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step/>
3. Karin Kelley. What is Data Analysis: Methods, Process and Types Explained. <https://www.simplilearn.com/data-analysis-methods-process-types-article>
4. Will Hillier A Guide To Secondary Data Analysis.  
<https://careerfoundry.com/en/blog/data-analytics/secondary-data-analysis/>