

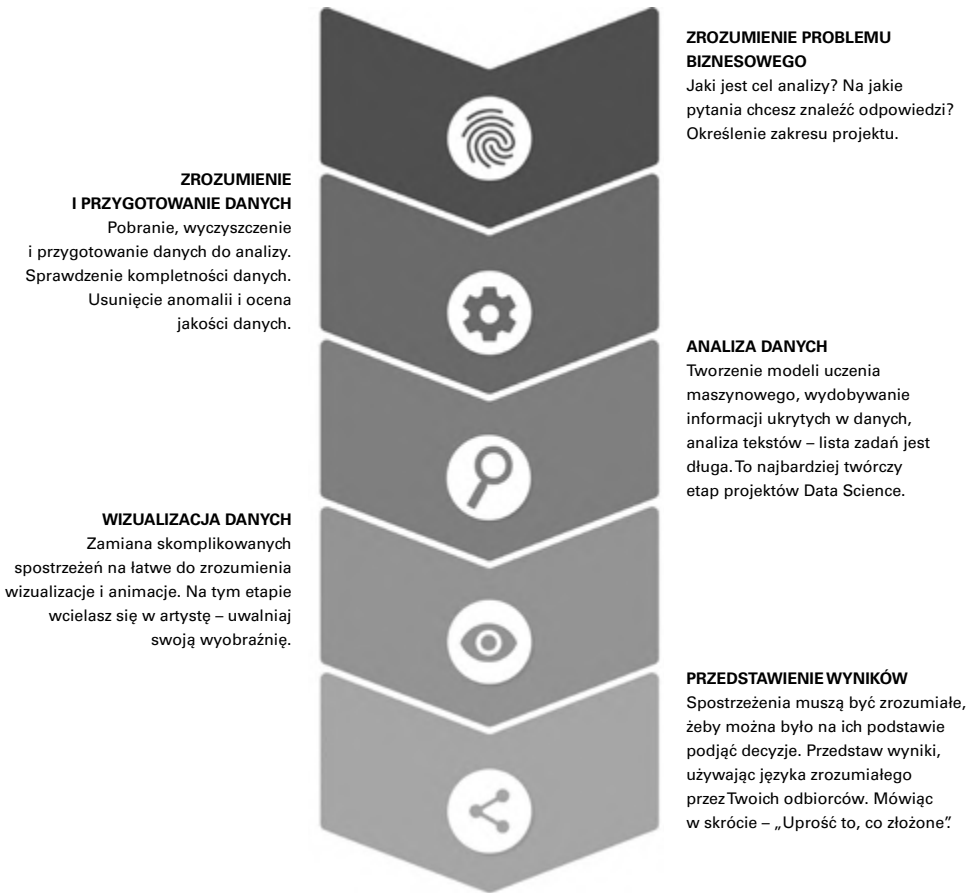
Część II

„Gdzie i jak je znajdę?” Gromadzenie i analiza danych

Jak to zwykle bywa, to co najlepsze, jest trudne do osiągnięcia. Projekty, w których używa się danych, doskonale ilustrują tę tezę. Celem tego typu projektów jest wydobycie z danych czegoś *nowego*, jako że zadaniem specjalistów danych jest rozwiązanie istniejących *problemów*. Zaczynając nowy projekt, lubię wyobrazić sobie, że rozmawiam z danymi. Poznaję je, żeby mieć pewność, że będę mógł je uczciwie zaprezentować i że moi klienci oraz sponsorzy projektu odniosą z nich maksimum korzyści. Z własnego doświadczenia, oraz z doświadczeń moich współpracowników, wiem, że samo poznanie danych potrafi otworzyć oczy i dostrzec możliwości zmian na lepsze. Dotyczy to wszystkich firm, z którymi pracowałem, a zakres zmian zawsze był szeroki – od sposobu wykonywania konkretnych zadań po strukturę organizacyjną. Czasami uzyskane wyniki były bezpośrednio związane z problemami biznesowymi, o których rozwiązanie zostaliśmy poproszeni. Równie często udawało nam się rzucić światło na aspekty działalności biznesowej niewidoczne wcześniej dla klienta.

A zatem w danych tkwi ogromny potencjał. To właśnie sprawia, że praca z nimi jest tak ekscytująca. Dane zawsze coś nam powiedzą, czasami będzie to nowa informacja, a czasami nie. Dane pozwalają badać różne możliwości i testować różne hipotezy. Robimy to, zadając różne pytania, przekształcając dane na różne sposoby i analizując je przy użyciu różnych algorytmów.

Proces Data Science



Rysunek II.1. Proces Data Science

Dane mają ogromny potencjał, ale wydobyć ukrytych w nich informacji może być skomplikowane. Szczególnie, jeżeli danych jest dużo, są one różnorodne i porozrzucane między różnymi systemami, a firma nie wie dokładnie, gdzie i jak przechowuje swoje dane. Wtedy właśnie z pomocą przychodzi proces Data Science. To sprawdzona i niezawodna metoda prowadzenia projektów, pozwalająca uporządkować dane niezależnie od ich ilości i rodzaju. Proces Data Science przeprowadzi Cię przez cały projekt, od jego zaplanowania po wdrożenie. Po raz pierwszy opracowany przez Joe Blitzsteina i Hanspetera Pfistera opisuje on każdy etap projektu, zaczynając od zgromadzenia danych po zaprezentowanie wyników w prosty i przekładający się na decyzje biznesowe sposób.