Droga do monetyzacji danych

Drogowskazy i przykłady

Kamil Cerazy
@CDO Forum Club
21 kwietnia 2022 r.
kamil.cerazy@gmail.com

Zacznijmy od trzech oczywistości

O danych i o ludziach

Dane są cenne

- o ile są dobrej jakości
- o ile jest ich dużo
- o ile traktujemy je intencjonalnie (zbieramy, przechowujemy, wykorzystujemy)

Dane są cenne, **kiedy je wykorzystujemy**

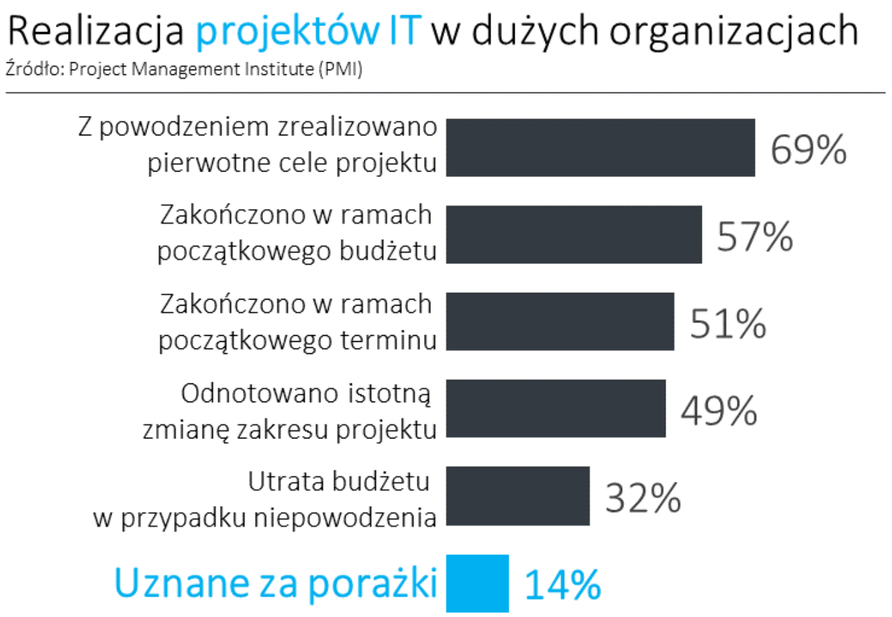
- o ile wiemy jak
- pozwalają ograniczać koszty i optymalizować zyski
- pozwalają budować przewagę nad konkurencją

Ludzie są najcenniejsi

- są kreatywni - generują pomysły
- przekładają dane na wnioski

Dzień dobry, chcemy robić machine learning

Nie zawsze jest kolorowo - wyzwania w realizacji projektów



vs.

87%

projektów obszaru ML

przyniesie błędne wyniki ze względu na błąd w danych, algorytmach lub zespołach odpowiedzialnych za zarządzanie nimi.

Dzień dobry, chcemy robić machine learning

Nie zawsze jest kolorowo - wyzwania w realizacji projektów

vs.

87%

projektów obszaru ML

przyniesie błędne wyniki ze względu na błąd w danych, algorytmach lub zespołach odpowiedzialnych za zarządzanie nimi.

Ok. **14%** realizacji projektów informatycznych zakończyła się całkowitą porażką.

≠

Ok. **87%** realizacji projektów zaawansowanej analityki danych nie zostanie wdrożona na produkcję.

Projekty analityki danych (metod ML/AI), różnią się strukturalnie od projektów informatycznych!

Dzień dobry, chcemy robić machine learning

Powody problemów w realizacji projektów zaawansowanej analityki danych

87%

projektów obszaru ML

przyniesie błędne wyniki ze względu na błąd w danych, algorytmach lub zespołach odpowiedzialnych za zarządzanie nimi.

Komunikacja między biznesem a technologią

Biznes przyjmuje **nierealistyczne założenia** dot. możliwości wykorzystania technologii

Departamenty technologiczne / analityczne zdecydowanie **zorientowane na rozwiązaniach a nie problemach biznesu**

Problemy z jakością i ilością danych (biznes / technologia)

Sam fakt **posiadania „jakichś" danych nie wystarcza** – ważna jest ich jakość i ilość!

Niedokładne, nieaktualne, niekompletne bądź nieustrukturyzowane dane stanowią **hamulec lub barierę** wdrażania rozwiązań analitycznych

Niedopasowanie umiejętności zespołu wewnętrznego

Wysoce **interdyscyplinarny obszar** (informatyka, statystyka, ekonometria, matematyka)

Stanowisko specjalisty od AI **zajęło pierwsze miejsce** na liście najbardziej poszukiwanych zawodów według LinkedIn w 2020 r. w USA

Dzień dobry, chcemy robić machine learning

Trzy metody realizacji projektu zaawansowanej analityki danych

Dostawca zewnętrzny

- rozwiązanie mało rozwojowe ale powinno być szybkie i skuteczne
- potrzebny jest dobrze zdefiniowany problem do rozwiązania
- możemy dostać black-box’a i uzależnić się od dostawcy

Rozwój zespołu

- rozwiązanie przyszłościowe
- rezultaty w średnim-długim okresie
- wymagane jest zapewnienie przestrzeni do eksperymentowania
- mogą pojawić się problemy z przekuciem teorii w praktykę

Współpraca z kimś doświadczonym

- realizacja PoC
- łączy benefity korzystania z doświadczonego doradcy i pozwala na rozwój zespołu
- w efekcie powinniśmy dostać rozwiązanie problemu i zespół, który je rozumie oraz potencjalnie umie zreplikować projekt

Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail

Klient: Firma odzieżowa

Kanały: lokalnie i online

Przychody: ponad $500 mln rocznie

Problem: Prognozowanie wielkości sprzedaży, przy uwzględnieniu planowanych poziomów promocji

Przekroje danych:
transakcje x lokalizacje x produkty

Okres danych: 2010-2020

Horyzont prognozy

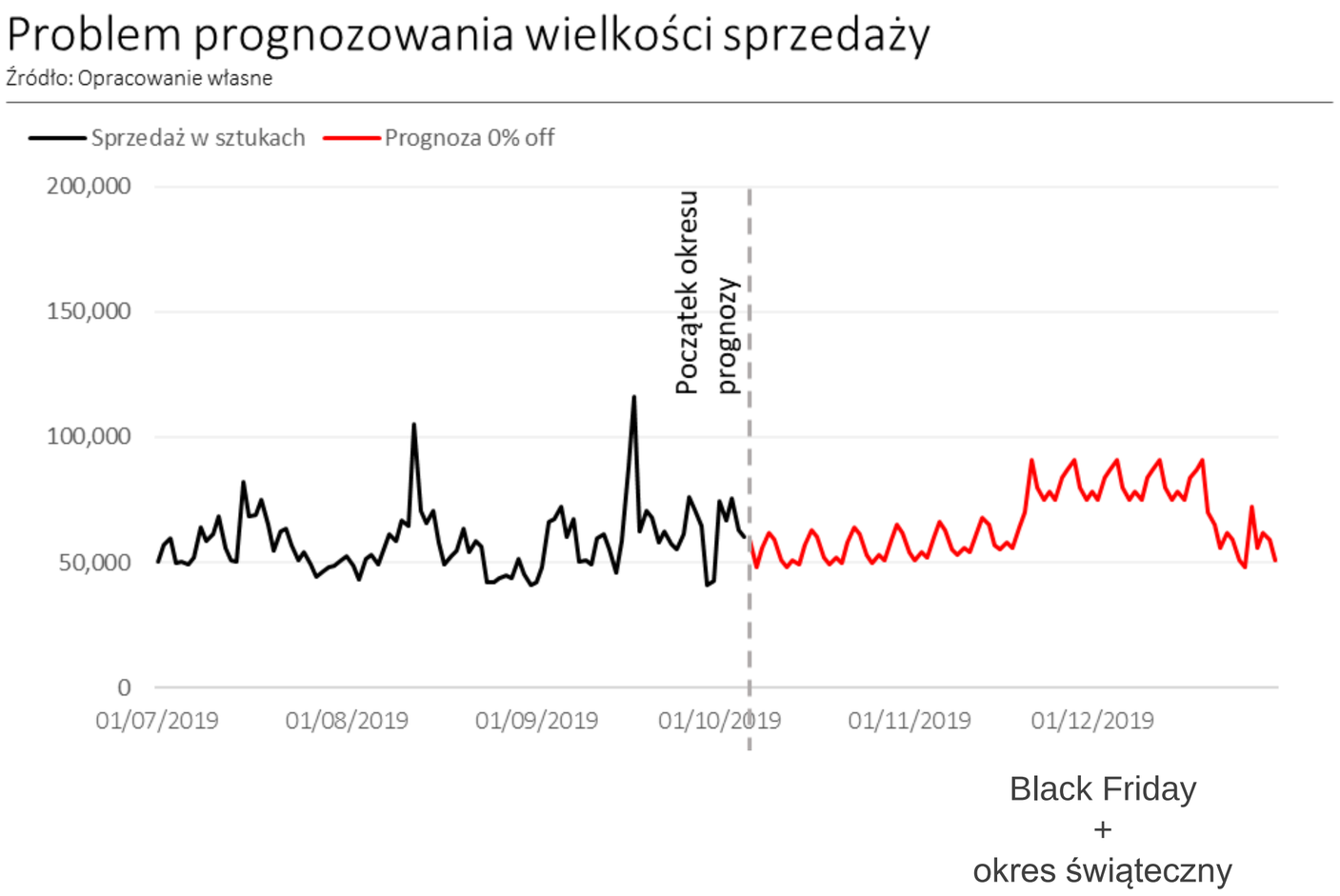
3 miesiące

Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail

Pierwszy model nie przewiduje żadnej akcji promocyjnej w danym okresie – wzrost poziomu sprzedaży związany jest z podwyższoną sprzedażą w okresie świątecznym.

Traktujemy tę krzywą, jako model sprzedaży bazowej



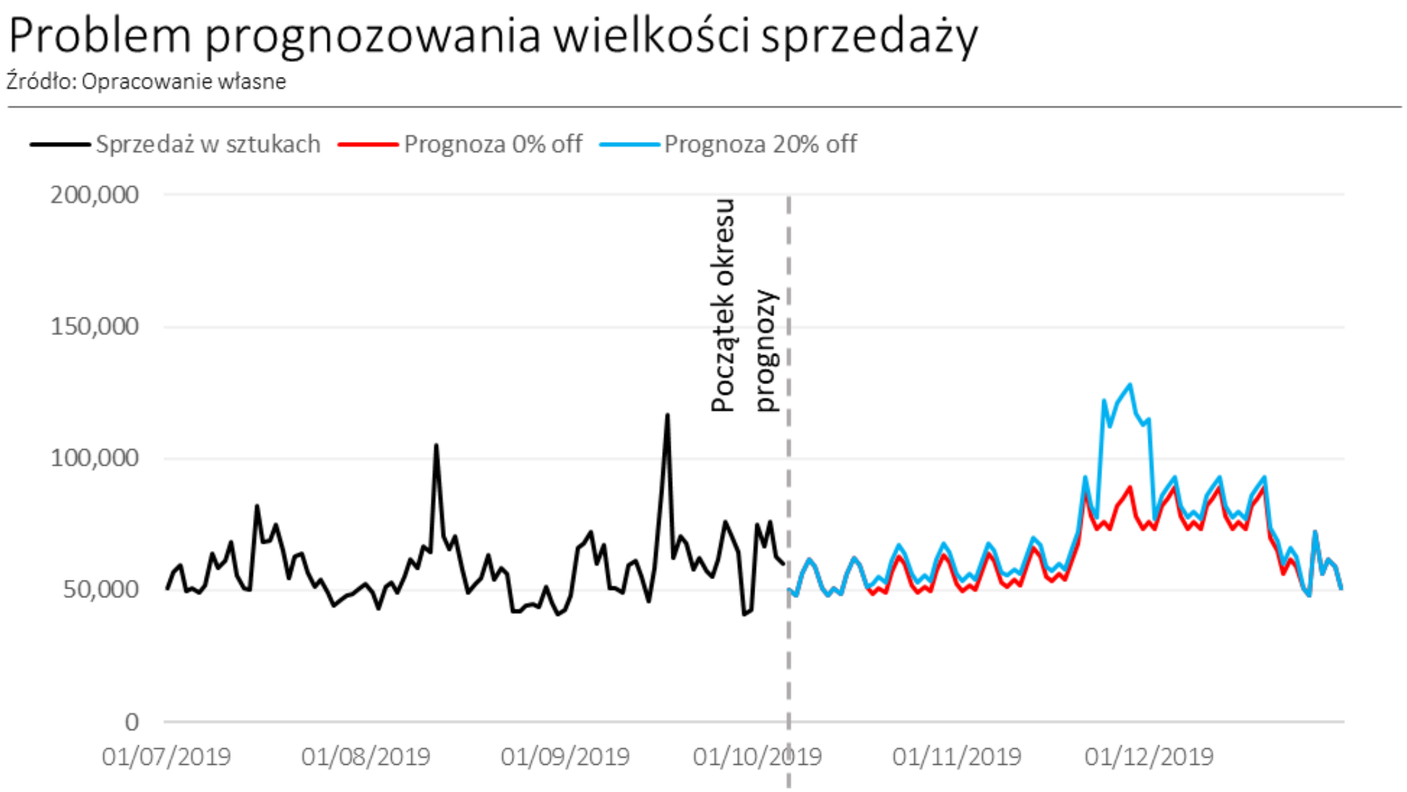
Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail

Drugi model zakłada, że klient wprowadzi promocję -20%.

W okresie Black Friday widać istotny wpływ wprowadzenia promocji względem **modelu bazowego**.

Model na tym etapie umożliwia ekstrakcję wpływu promocji na poziomie 20% na wielkość sprzedaży!



Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail

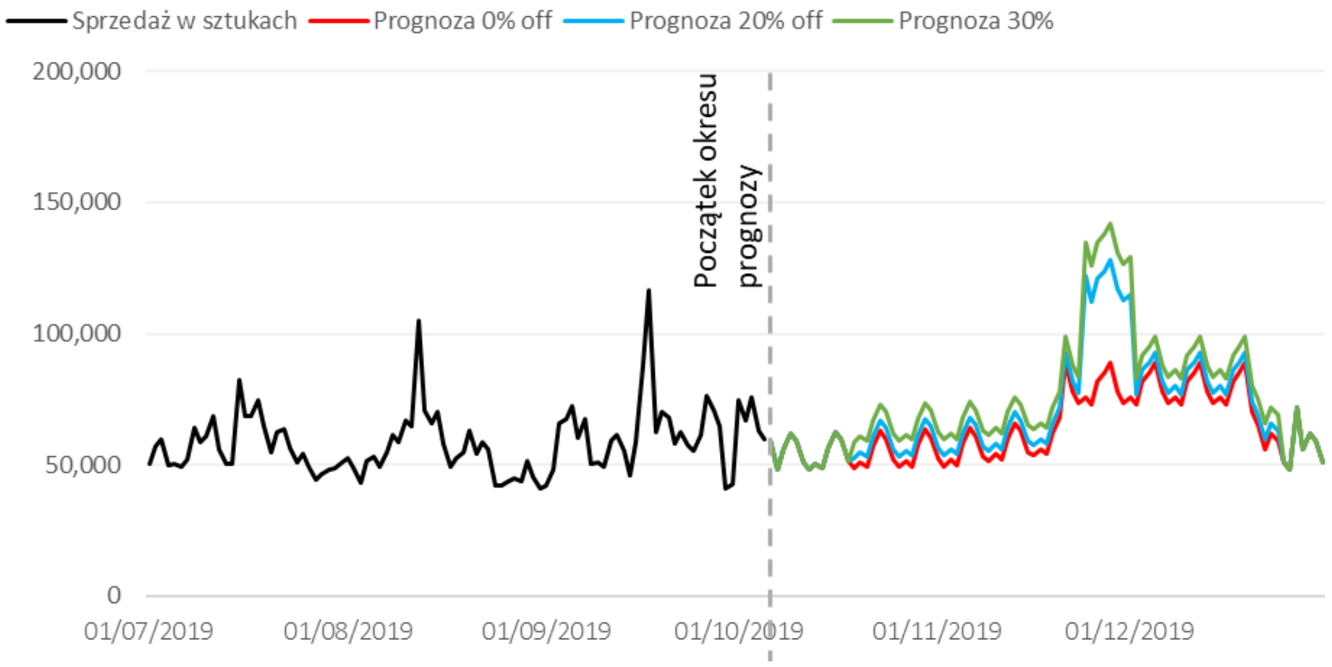
Ostateczny model przedstawia wpływ promocji na poziomie 30% na wielkość sprzedaży.

1) Model poprawnie identyfikuje **monotoniczność funkcji sprzedaży**, względem promocji.
2) Model poprawnie identyfikuje **elastyczność cenową popytu** tzw. zasada malejącej krańcowej użyteczności.

Opracowane modele umożliwiają prowadzenie **optymalizacji prowadzenia akcji promocyjnej** (np. funkcja celu to poziom zysku).

Problem prognozowania wielkości sprzedaży

Źródło: Opracowanie własne

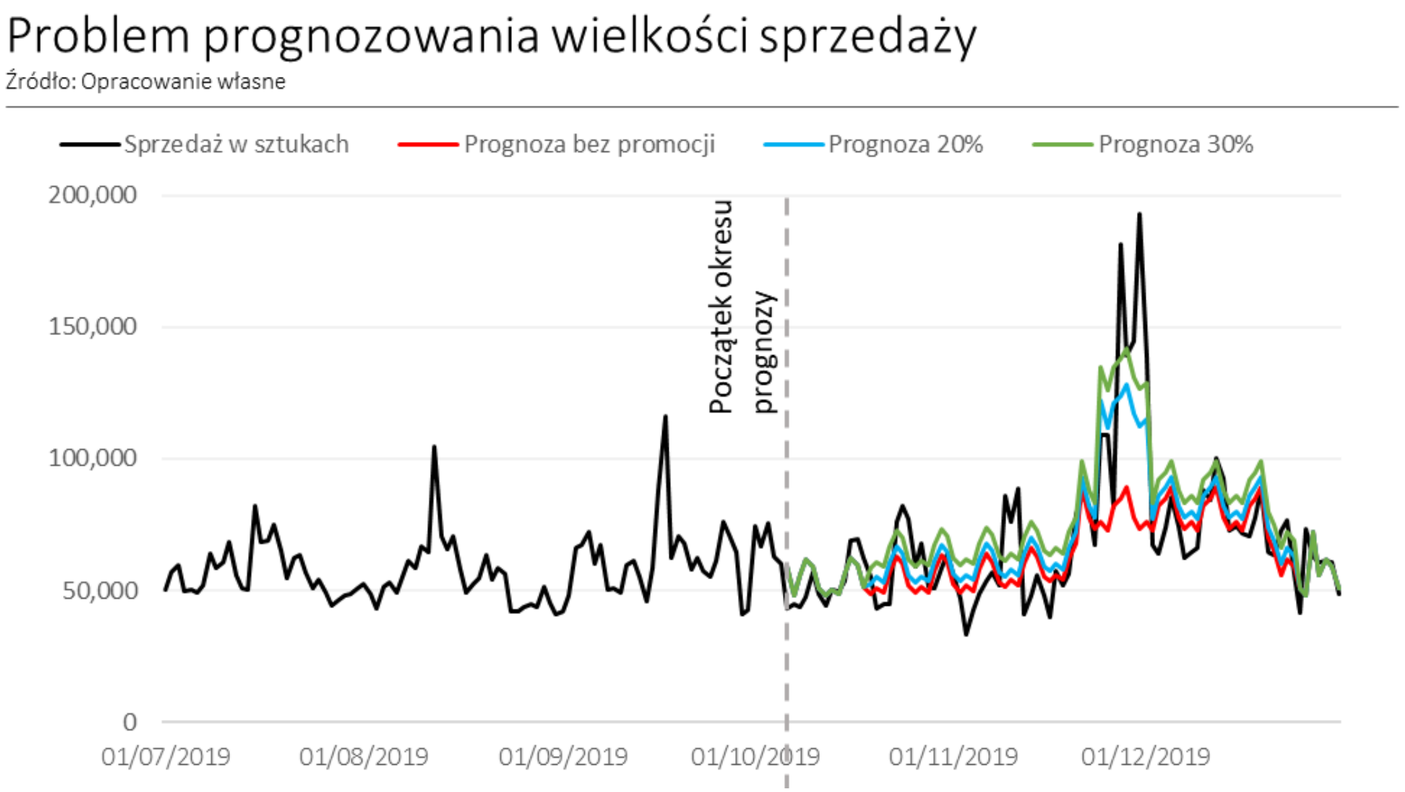


Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail

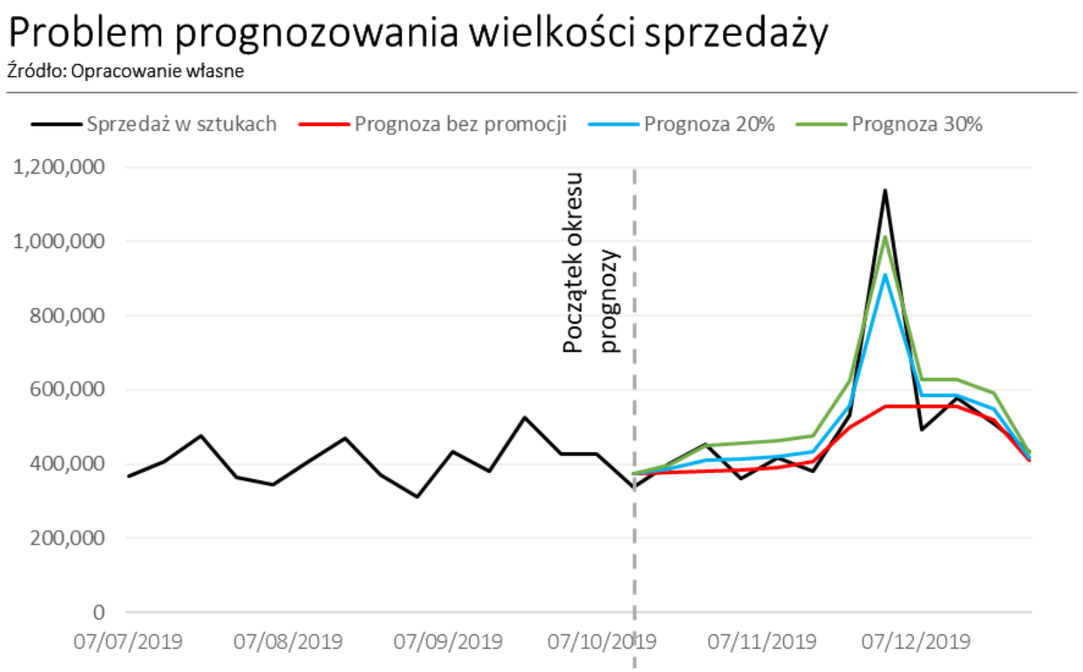
Klient w rzeczywistości przyjął **30% poziom promocji**. Modele prognostyczne potrafiły w poprawny sposób zidentyfikować poziomy sprzedaży.

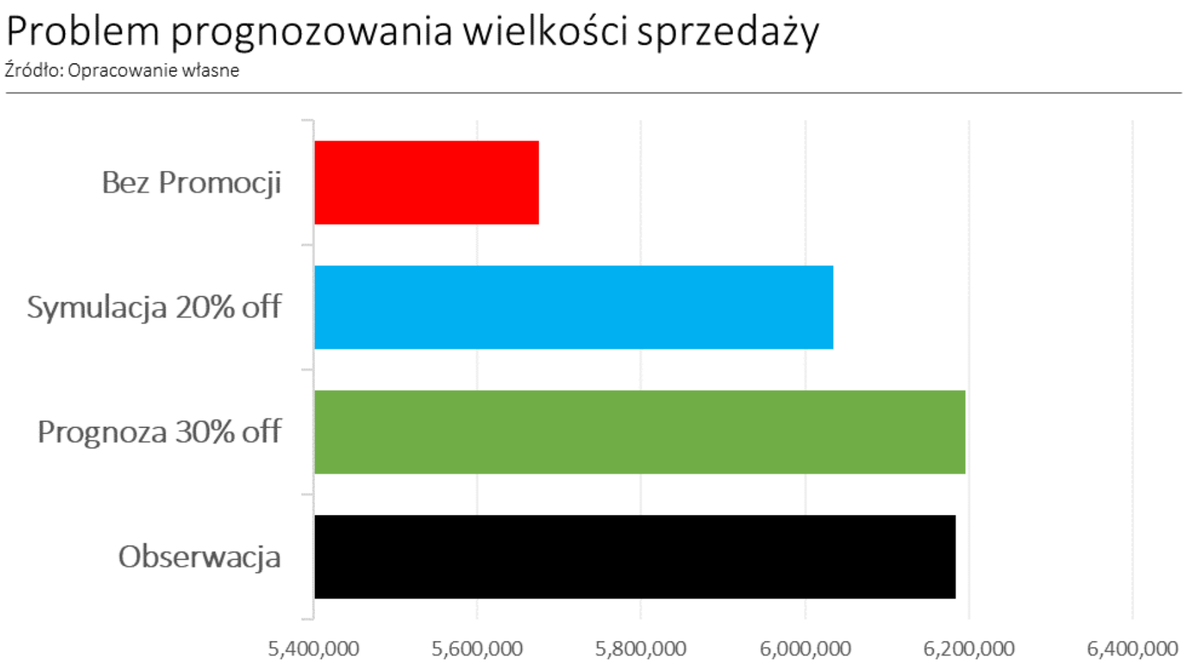
Należy zwrócić uwagę, że dla poszczególnych dat, **poziom prognozowanej sprzedaży** może istotnie różnić się od poziomu obserwowanego…



Success story 1 – Optymalizacja zysku

PoC - Prognoza wpływu promocji na wolumen sprzedaży w sektorze retail





Na wyższych poziomach agregacji, błąd prognozy staje się coraz bardziej pomijalny.

Względny poziom błędu prognozy za okres 3 miesięcy wynosi ok. 2.5%!

Success story 2 – Optymalizacja kosztów

Operacjonalizacja ładu danych

Klient: Scale-up technologiczny

Produkt: aplikacja do CRM

Przychody: ponad $250 mln rocznie

Problem: projekty analityczne są ograniczane przez problemy z danymi

30% czasu pracy analityków ilościowych jest spędzane na problemy niskiej jakości i dostępności danych

5% w organizacjach, które wdrożyły Data Governance, **zagadnienia te zajmują istotnie mniej czasu**!

Koszty złej jakości lub braków danych stanowią aż:

20-35%

przychodów operacyjnych organizacji.

Procesy Data Governance umożliwiają **odzyskanie około połowy tych kosztów**.

Pozostałej części nie można odzyskać, ze względu na ciągłą ewolucję danych i wykorzystywanych technologii

Źródło: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value

Źródło: L.P. English, Information Quality Applied, Wiley, 2009