

**Howard Steven Friedman  
Akshay Swaminathan**

# **ANALIZA DANYCH**

**dla menedżerów**

Jak zrozumieć i wykorzystać  
dane dostępne w twojej firmie

Howard Steven Friedman  
Akshay Swaminathan

# ANALIZA DANYCH dla menedżerów

Jak zrozumieć i wykorzystać  
dane dostępne w twojej firmie

Przekład: Anita Doroba

# Spis treści

|                                                   |     |
|---------------------------------------------------|-----|
| Podziękowania                                     | 9   |
| Analiza danych dla menedżerów                     | 13  |
| Wstęp                                             | 15  |
| 1    Narzędzia pracy                              | 25  |
| 2    Projekt „data science”                       | 57  |
| 3    Podstawy nauki o danych                      | 87  |
| 4    Podejmowanie decyzji w oparciu o dane        | 133 |
| 5    Klasteryzacja, segmentacja i redukcja szumów | 169 |
| 6    Budowanie pierwszego modelu                  | 191 |
| 7    Narzędzia uczenia maszynowego                | 227 |
| 8    Złożenie wszystkiego w całość                | 251 |
| 9    Etyka                                        | 291 |
| Podsumowanie                                      | 317 |
| Przypisy                                          | 325 |

# Wstęp

Nieważne, czy jesteś świeżo upieczonym absolwentem studiów MBA, menedżerem ds. rozwoju biznesu w start-upie, kierownikiem projektu w firmie z listy Fortune 500 – praca z danymi i analitykami danych będzie nieodłącznym elementem twojej kariery. To, jak dobrze potrafisz współpracować z takim zespołem i jaką wartość potraficie wspólnie wytworzyć, będzie decydowało zarówno o twoim zawodowym sukcesie, jak i o sukcesie twojej firmy.

We współczesnym środowisku korporacyjnym nie można prosperować bez pewnego zrozumienia tzw. data science\* oraz jej zastosowań. Nie oznacza to, że każda praca będzie wymagała od ciebie umiejętności programowania w języku Python, tworzenia własnych rozwiązań opartych na operacjach w chmurze czy wiedzy na poziomie doktoratu w dziedzinie algorytmów uczenia maszynowego.

Oznacza to, że musisz nauczyć się tego przedmiotu na tyle dobrze, aby mówić jego językiem i zadawać dobre pytania. Co rozumiemy przez dobre pytania? To pytania, które zwiększają szanse,

---

\* Dziedzina badań łącząca wiedzę specjalistyczną, umiejętności programistyczne oraz znajomość matematyki i statystyki, zajmująca się poszukiwaniem ukrytych wzorców, trendów oraz informacji, które mogą stać się kluczem do podejmowania przyszłościowych decyzji; w książce będziemy po polsku nazywać ją „nauką o danych” – *przyp. tłum.*

że proponowany tok postępowania będzie stanowił rozwiązanie twoich problemów, pozwalając równocześnie uniknąć zbędnych wydatków i niepotrzebnego wysiłku.

Wyzwania pracy z zespołem specjalistów ds. danych są podobne do wyzwań pracy z innymi zespołami technicznymi. Klient musi rozumieć, czego chce, i potrafić zakomunikować te cele. Zespół specjalistów ds. danych powinien potrafić zakomunikować, jakie są możliwe rozwiązania, oraz wyjaśnić ich wady i zalety. Klient i zespół specjalistów ds. danych muszą współpracować, aby wspólnie stworzyć wartość. Wymaga to od każdej strony na tyle dobrego zrozumienia, by móc zadawać odpowiednie pytania i aktywnie słuchać, jakie są obawy drugiej strony, niezależnie od tego, czy zespół „data science” siedzi w pomieszczeniu za ścianą, czy na drugim końcu świata.

Ta książka przygotowuje cię do tego, byś został dobrym klientem zespołów specjalistów ds. danych, które będą ci pomagać w wydobyciu jak największej wartości z ich wiedzy, a także – miejmy nadzieję – w unikaniu największych pułapek, przez które marnujemy pieniądze i czas. Poprowadzi cię przez wiele podstawowych idei, które powinieneś poznać, aby wykorzystać naukę o danych w twojej firmie, poprawić skuteczność i wydajność zespołu specjalistów ds. danych oraz zdobyć podstawy niezbędne do dobrej komunikacji z zespołami technicznymi. Dzięki książce zrozumiesz podstawowy żargon, nauczysz się rozpoznawać w zespole specjalistów ds. danych różne typy talentów i zadawać zespołowi właściwe pytania, aby dzięki nim zyskać przemyślenia, spostrzeżenia i możliwości. Pod koniec tej książki będziesz potrafił odpowiedzieć na podstawowe pytania, między innymi o główne narzędzia sprzętowe i software’owe wykorzystywane w analizie danych, o to, kim są różni gracze w zespole specjalistów ds. danych i jakie modele powinny być brane pod uwagę w konkretnych projektach.

Co najważniejsze, zostaniesz również wyposażony w zasób kluczowych pytań, które będziesz mógł wykorzystać, aby dokładniej ocenić analityków danych, statystyków, specjalistów w dziedzinie data science i innych ekspertów technicznych. Chcąc wydobyć jak najwięcej wartości z tej książki, powinieneś zrozumieć podstawy statystyki opisowej (średnie, mediany i dominanty), potrafić czytać podstawowe grafy oraz posiadać pewne doświadczenie z arkuszami kalkulacyjnymi i programami takimi jak Excel czy Google Sheets. Nie jest konieczne doświadczenie w programowaniu, zaawansowana znajomość statystyki czy inne umiejętności obliczeniowe. Na praktycznych przykładach zgłębimy niektóre fascynujące metody stworzone przez naukowców zajmujących się danymi, ale nie będziemy obciążać cię równaniami i koncepcjami, których nie musisz rozumieć.

Mając to na uwadze, chcemy jasno stwierdzić, czym ta książka na pewno nie jest.

Nie jest podręcznikiem dla tych, którzy liczą na zdobycie umiejętności lub wiedzy niezbędnej, by zostać „mistrzem danych”, jak czasami nazywa się specjalistów ds. danych. Istnieje całe mnóstwo doskonałych podręczników i materiałów dostępnych w sieci, z których można się nauczyć technicznych pojęć z dziedziny nauki o danych (często nazywanej z angielska również w tekstach w języku polskim „data science”). Wprowadzimy tu wiele pojęć podstawowych z punktu widzenia klienta, aby było jasne, w jaki sposób każda idea jest istotna dla rozwiązania jego problemu.

To nie jest książka o programowaniu. Tu też istnieje już mnóstwo świetnych materiałów dostępnych dla tych, którzy chcą się nauczyć języków programowania takich jak Python, R, SQL, Java i innych. Zasoby sieciowe, takie jak na przykład konkursy Kaggle czy akademie kodowania, są łatwo dostępne dla tych, którzy chcą się uczyć.

Nasza książka zaczyna się od omówienia narzędzi stosowanych w tej dziedzinie, podstawowych informacji, których potrzebujesz, aby zostać dobrym klientem specjalistów w dziedzinie nauki o danych. Niezależnie od tego, czy jesteś klientem wewnętrznym, czy kimś, kto chce zaangażować zewnętrzny zespół specjalistów ds. danych, budżet twojego projektu nie powinien zawierać wydatków związanych z inwestycją w nieodpowiednie oprogramowanie i narzędzia. To byłaby strata cennego czasu i pieniędzy. Zastosowanie właściwych narzędzi jest najważniejszym krokiem, ponieważ nie da się uprawiać nauki o danych bez odpowiednich możliwości przechowywania danych i oprogramowania analitycznego. Dzięki zrozumieniu podstaw systemów danych, takich jak chmura i pamięć lokalna czy jeziora i hurtownie danych, zapoznasz się z językiem technicznym, koncepcjami i aplikacjami. Kiedy zespół specjalistów od danych zaproponuje ci sposób analizy zbioru danych lub dostarczenia rozwiązania, będziesz mógł dokładniej to sprawdzić, aby zwiększyć swoje szanse na otrzymanie czegoś, co pasuje do twoich potrzeb. Języki programowania i oprogramowanie to podstawowe narzędzia pracy każdego zespołu specjalistów ds. danych. Chcąc pomóc ci w zwiększeniu skuteczności współpracy z zespołem, wprowadzimy niektóre koncepcje kluczowych języków chronionych i niechronionych prawem autorskim.

Projekty związane z nauką o danych są po prostu projektami. Omówimy podstawy zarządzania projektami dotyczącymi danych, w tym kluczowe kroki niezbędne do osiągnięcia sukcesu. Podstawą sukcesu każdego projektu są ludzie. Jeśli zatrudnimy niewłaściwych, pogrążymy nasz projekt analizy danych. Na podstawie samych tytułów zawodowych, takich jak „data scientist”, czyli specjalista ds. danych, inżynier danych, analityk danych, inżynier uczenia maszynowego czy statystyk, często trudno jest się domyślić, kto dysponuje zestawem umiejętności niezbędnych do osiągnięcia

sukcesu. Podobnie jak drużyna koszykówki składa się z różnych graczy, posiadających różne umiejętności, zespół specjalistów ds. danych tworzą różni specjaliści. Poznamy graczy zaangażowanych w zespole specjalistów ds. danych, zrozumiemy, na czym polegają ich umiejętności, i zobaczymy, jak pracują, aby dostarczać produkty i usługi. Wyjaśnimy role inżynierów danych, specjalistów uczenia maszynowego, ekspertów w dziedzinie wizualizacji danych i innych specjalistów. Powiemy także, jak w projekcie ustalać priorytety, aby odpowiednio alokować ograniczone zasoby czasu i pieniędzy.

Częstym błędem popełnianym przez firmy stawiające pierwsze kroki w dziedzinie nauki o danych jest „strzelanie z karabinu maszynowego do komarów”. Głębokiemu uczeniu i sztucznej inteligencji (AI) poświęca się w nich tak wiele uwagi, że często rozprasza to głównych decydentów. Małym, brudnym sekretem nauki o danych jest to, że najbardziej zaawansowane i najnowocześniejsze metody modelowania często nie są niezbędne do uzyskania największej wartości z danych. Przyjrzymy się, jak klienci nauki o danych pracujący w firmach, organizacjach i instytucjach rządowych mogą najefektywniej przejść od podstawowej analizy przez proste modele predykcyjne do zaawansowanego modelowania. Pokażemy tę ścieżkę, wyjaśniając niektóre podstawowe koncepcje nauki o danych i sposób ich zastosowania.

Jednym z celów tej książki jest nauczenie cię podstawowych podejść do różnych problemów z danymi. Później opiszemy niektóre z najważniejszych technik modelowania. Nie ma to na celu uczynienia z ciebie pełnoprawnego analityka danych ani nawet programisty średniego szczebla. Skupimy się raczej na tym, abyś został dobrze przygotowany do dyskusji ze specjalistami od danych, abyś zrozumiał co nieco z powszechnie używanego żargonu, rozpoznał najbardziej odpowiednie rozwiązania różnych problemów i był gotowy na zadawanie dobrych pytań.



Przejrzymy techniki nienadzorowanego uczenia maszynowego. Te metody nie starają się przewidywać wyniku, lecz raczej grupować ludzi na podstawie danych. Na przykład zamiast traktować miliony klientów jako jedną grupę, specjaliści od danych mogą znajdować klastry klientów mających podobne upodobania, wzorce zachowań lub inne kluczowe cechy. Klastry klientów w branży restauracyjnej mogą obejmować osoby, które wybierają się na cotygodniowe wieczorne wyjścia, gości na kolacjach urodzinowych (rzadko pozwalających sobie na kilka drogich posiłków), rodziny (grupy składające się, powiedzmy, z co najmniej czterech osób, które często jadają w tych samych kilku restauracjach) i gości jednorazowych (którzy przychodzą raz do każdej restauracji i nigdy potem nie wracają). Klastry pacjentów w branży ubezpieczeń zdrowotnych mogą obejmować starszych, chorych pacjentów z dużymi potrzebami w zakresie zabiegów chirurgicznych; młodszych, zdrowych pacjentów z ograniczonymi potrzebami medycznymi i prawie w ogóle niekorzystających ze świadczeń zdrowotnych; aktywnych, zdrowych seniorów potrzebujących jedynie rutynowych wizyt u lekarza; oraz młode osoby mające skłonności do ryzyka i często odwiedzające oddziały ratunkowe. Te klastry, stanowiące podgrupy całej bazy klientów, mają ważne wspólne cechy. Myśląc o takich klastrach klientów, a nie o całej bazie, łatwiej jest precyzyjnie kierować reklamę, sprzedaż krzyżową i inne rekomendacje do konkretnych odbiorców.

Następnie przejdziemy do nadzorowanego uczenia maszynowego. Te metody są stosowane do prognozowania wyniku w obszarze zainteresowań. Bez względu na to, czy staramy się przewidzieć, kto odpowie na reklamę, jak długo pacjent będzie przebywał w szpitalu albo czy klient terminowo spłaci zadłużenie karty kredytowej, kluczowe koncepcje modelowania i stosowane metody są często podobne w różnych branżach. Metody

obejmujące między innymi regresję liniową, regresję logistyczną, drzewa klasyfikacyjne i regresyjne, lasy losowe i uczenie maszynowe z podbijaniem gradientu zostaną zaprezentowane wraz z pewnymi trikami stosowanymi przez specjalistów od danych w celu poprawy przewidywalności modeli.

W jednym z rozdziałów książki poruszymy kilka zagadnień specjalistycznych związanych z nauką o danych. Rozdział ten skupia się na takich interesujących tematach, jak analiza sieci, analiza przestrzenna, głębokie uczenie i sztuczna inteligencja. Wprowadzona zostanie analiza zachowań grupy wykorzystująca takie narzędzia jak analiza sieci, co stanowi istotną część analizy Facebooka, LinkedIna i Twittera\*. Pomoże to określić charakter powiązań między różnymi użytkownikami, zidentyfikować głównych łączników w grupie i stwierdzić, czy sieć jest połączona ściśle, czy bardzo luźno. Omówimy główne miary zrozumienia sieci, między innymi takie pojęcia jak gęstość i centralność sieci. Wprowadzimy analizę przestrzenną, służącą określaniu, w jaki sposób są połączone ze sobą różne elementy w przestrzeni fizycznej. Kierowcy Ubera nie są jedynymi osobami, które muszą się zastanawiać, jak dotrzeć z jednego miejsca w inne. Kierowcy karetek muszą znać najlepsze trasy przewozu pacjentów do szpitala. Ludzie myślący o zakupie mieszkania mogą preferować takie, które znajdują się najbliżej metra lub autobusu, a większość chce wiedzieć, czy w pobliżu znajdują się sklepy.

Oczywiście żadna dyskusja na temat nauki o danych nie jest kompletna bez omówienia AI i głębokiego uczenia. Niezależnie od tego, czy mówimy o samochodach autonomicznych, tłumaczeniu języków, wirtualnych asystentach czy wykrywaniu

---

\* 23 lipca 2023 roku zmieniono nazwę Twittera na X, jednak autorzy książki konsekwentnie stosują poprzednią nazwę – *przyp. red.*

nowotworów, w tych nowatorskich obszarach nieustannie dokonuje się niesamowity postęp. Wprowadzimy podstawowe koncepcje stojące za najpotężniejszymi z tych metod, między innymi takie jak splotowe sieci neuronowe, rekurencyjne sieci neuronowe i duże modele językowe. Potem opiszemy jedno z najpopularniejszych zastosowań sztucznej inteligencji, obejmujące widzenie komputerowe. Pomyśl o niesamowitym postępie technologicznym umożliwiającym komputerom dokładne oznaczanie obrazów jako zawierających drastyczne treści. Podobnie można przeskanować miliony odręcznie zapisanych dokumentów i cyfrowo wydobyć z nich informacje przy użyciu widzenia komputerowego. Zastąpi to pracę ogromnych zespołów ludzi starających się rozszyfrować czyjeś niewyraźne pismo.

Wielu menedżerów wyższego szczebla wykazuje mniejsze zainteresowanie zrozumieniem szczegółów struktury danych i modelowania decyzji, skupiając się raczej na tym, aby zobaczyć wyniki. Powinieneś być przygotowany do udzielenia odpowiedzi na takie proste pytanie: „Czy miliony dolarów, które wydajemy na inwestycje związane z naszym badaniem danych, zapewniają nam dobry wskaźnik zwrotu z inwestycji?”. Będziesz musiał mierzyć wpływ zarówno takimi prostymi metodami, jak testowanie A/B, jak i bardziej zaawansowanymi metodami wnioskowania przyczynowego, takimi jak różnica w różnicach, przerywane szeregi czasowe i nieciągłość regresji, aby pokazać wpływ polityk i programów. W tej książce zostaną wprowadzane podstawowe koncepcje pomiaru wpływu, a następnie ważne pytania do rozważenia w związku z bardziej zaawansowanymi metodami.

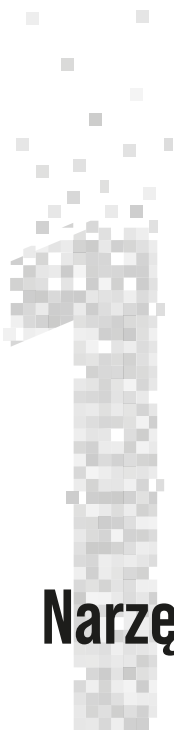
Ważną sprawą, często pomijaną przez specjalistów ds. danych, jest etyka. Ostatnią rzeczą, jakiej chciałaby twoja firma, jest znalezienie się na pierwszej stronie „New York Timesa” z powodu jej algorytmu wzmacniającego dyskryminację z powodu rasy, płci lub

innych uprzedzeń. Twój zespół specjalistów ds. danych musi mieć świadomość konsekwencji swojego modelowania i zadbać o to, aby w modelach nie pojawiły się wprost lub w zawoalowany sposób żadne uprzedzenia. Na przykład modele stosowane w ocenie wiarygodności kredytowej mogą przypadkowo i niebezpośrednio uwzględniać takie czynniki jak rasa i płeć, co jest nielegalne. Modele prognozowania talentów mogą tworzyć zamknięte pętle, w których uczniowie ubiegający się o przyjęcie do szkoły i ludzie starający się o pracę na określonych stanowiskach są niesprawiedliwie karani za niedopasowanie do wzorców historycznych. Poznasz niektóre podstawowe problemy etyczne, sposoby ich unikania oraz pewne najlepsze praktyki etyczne.

W podróży opisanej w tej książce będą ci towarzyszyć dwie osoby: Steve i Kamala. Oboje starają rozwijać swoją karierę i oboje muszą wydobywać wartość z wiedzy o badaniu danych w ich firmach. Steve pracuje w instytucji udzielającej kredytów konsumenckich, kolejno w Departamencie ds. Oszustw, Departamencie Windykacji i Dziale Nieruchomości. Kamala, zarządzająca niższego szczebla w firmie ubezpieczeń zdrowotnych, pełni rolę zarówno w strategii klinicznej, jak i w marketingu, gdzie musi zadbać o równowagę między zapewnieniem pacjentom dobrej opieki i utrzymaniem zyskowności firmy. Choć w tej książce będziemy zgłębiać tematykę związaną z kredytami konsumenckimi i ubezpieczeniami zdrowotnymi, wnioski dotyczące bycia dobrym klientem mają charakter ogólny i są przydatne dla wszystkich branż, dlatego niezależnie od tego, na czym skupiasz się w swoim życiu zawodowym, znajdziesz w niej użyteczne wskazówki.

Na co czekać? Zaczynamy.





# Narzędzia pracy

## Steve i Shu Money Financial

W domu Steve był zawsze trochę złotą rączką. Wykonywał drobne prace hydrauliczne i elektryczne, a jego dumą i radością był zestaw szafek, które samodzielnie zaprojektował, zbudował i zamontował. Na jego stole warsztatowym można było znaleźć niemal wszystkie narzędzia, jakie można sobie wyobrazić: elektryczne szlifierki, pełny zestaw wiertel i frezów, strugi, ponad trzydzieści typów śrubokrętów i pilę stołową. Umiejętnie posługiwał się każdym narzędziem i wiedział, kiedy i gdzie każde z nich powinno być używane, a kiedy nie powinno.

Ale Steve nie siedział teraz w domu. Wpatrywał się w ekran komputera, będąc na półmetku swojej dwuletniej rotacyjnej pracy w Shu Money Financial. Była to jego pierwsza prawdziwa praca zaraz po ukończeniu studiów MBA w Columbia Business School. Dlaczego podjął tę pracę w branży kredytów konsumenckich?

Firma Shu Money Financial niezmiennie była zaliczana do grona najlepszych pracodawców w Stanach Zjednoczonych. Dużo inwestuje w swoich pracowników i traktuje ponad dwudziestotysięczną załogę jak jedną wielką rodzinę. „Shusterzy” – jak mówią o sobie pracownicy Shu – są dumni z tego, że korzystają z najnowocześniejszej technologii, rzucając akronimami i żargonem, który Steve z trudem mógł zrozumieć.

Głównym zadaniem Steve’a było kierowanie opracowywaniem nowej strategii priorytetyzacji spraw wysyłanych do Departamentu Windykacji. Przepływ klientów do tego departamentu jest prosty. Jeśli klient nie spłaci żadnej części zadłużenia z ostatnich sześciu miesięcy, dług jest uznawany za „zły”. Tak dzieje się w przypadku każdego produktu finansowego oferowanego przez Shu Money Financial – zadłużenia karty kredytowej, linii kredytowej, kredytu samochodowego lub hipotecznego. Klient ze złym długiem jest wtedy przekazywany do Departamentu Windykacji, którego zadaniem jest odzyskanie możliwie największej kwoty<sup>1</sup>.

Tego ranka Steve spotkał się z firmowym zespołem specjalistów ds. danych. Na początek koledzy pokazali mu oszałamiająco długą listę języków programowania, narzędzi programistycznych oraz metod wdrażania stosowanych w ciągu ostatnich trzech lat. Oczy mu zabłyśły, kiedy członkowie zespołu opowiadali o swoich produktach danych, które dostarczyli innym jednostkom w firmie. To był pierwszy projekt, który zespół specjalistów ds. danych realizował dla Departamentu Windykacji.

Steve był poruszony i przytłoczony. Chciał zrozumieć wszystko, o czym wspominali członkowie zespołu specjalistów ds. danych, i nie wprawić się w zakłopotanie. Chciał, żeby postrzegali go jako kompetentnego przyszłego lidera firmy. Zaczął myśleć o dyskusji na temat badania danych i znów poczuł ucisk w piersiach.

Jednak zaraz przypomniał sobie, że to nie on tworzy rozwiązanie. On jest klientem.

Zamiast zestawiać arkusze papieru ściernego o różnej ziarnistości, drewno, wiertła i piły, musiał zachowywać się jak klient. Gdyby kupował szafki i wynajmował kogoś do ich zamontowania, wiedziałby, jakie powinny mieć wymiary, gdzie je zamontować i jak dokładnie je użytkować. Zleciłby specjalistom od szafek przygotowanie rozwiązania i zadbałby, żeby produkty zaspokajały jego potrzeby.

To go trochę uspokoiło. Nie musiał być ekspertem od każdego narzędzia w zestawie do badania danych. Jego zadaniem było skupienie się na problemie, który miał do rozwiązania. Choć jest prawdą, że różne narzędzia lepiej pasują do różnych rozwiązań, nie zamierzał podejmować decyzji co do tego, z jakich urządzeń do przechowywania danych, programów, języków programowania i innych narzędzi skorzystać. Musiał jedynie zrozumieć opcje i konsekwencje rekomendacji zespołu specjalistów ds. danych. Gdyby miał jakieś szczególne potrzeby lub ograniczenia, musiałby zadbać o to, by członkowie zespołu specjalistów ds. danych dobrze je zrozumieli i żeby wymagania co do projektu zostały uwzględnione w ich rekomendacjach i rozwiązaniach. Na przykład to, że być może będzie trzeba wdrożyć to rozwiązanie w ciągu najbliższych dwóch miesięcy, że będzie potrzebował jasnego określenia kamieni milowych i oczekiwanych efektów i że koszt całego projektu nie może przekroczyć 50 tysięcy dolarów.

Głównym celem Steve'a było rozwiązanie problemu biznesowego. W tym wypadku jego zadaniem była lepsza priorytetyzacja pracy Departamentu Windykacji.

Myśląc o Departamencie Windykacji jak o firmie z własnym rachunkiem zysków i strat, Steve chciał zmaksymalizować sumę odzyskanych środków, równocześnie minimalizując niezbędne



koszty operacyjne. Departament miał środki wystarczające jedynie na podjęcie poważniejszych działań wobec ograniczonej liczby przypadków. W innych klienci po prostu otrzymywali automatyczny e-mail albo powiadomienie telefoniczne, co zazwyczaj skutkowało brakiem kontaktu z klientem i – co nie dziwi – niedzyskaniem środków. Większość długów była sprzedawana zewnętrznym agencjom windykacyjnym<sup>2</sup>. Dysponując ograniczonymi środkami, Steve nie mógł po prostu zwiększyć dwu- lub trzykrotnie zatrudnienia w Departamencie Windykacji, ponieważ to najprawdopodobniej jeszcze bardziej obniżyłoby zyski jednostki.

Chciał przede wszystkim zrozumieć aktualnie stosowane metody priorytetyzacji klientów i przypisywania ich opiekunom z Departamentu Windykacji. Odkrył, że obecna metoda jest dość prosta. Każdy klient, który kiedykolwiek dokonał jakiejś płatności, był zaliczany do kategorii 1. Każdy klient, który nigdy nie zapłacił, ale utrzymywał jakiś kontakt e-mailowy lub telefoniczny z Departamentem Obsługi Klienta, Departamentem Windykacji lub inną jednostką banku, trafiał do kategorii 2. Wszyscy klienci, którzy nigdy niczego nie zapłacili i w ogóle nie nawiązali kontaktu z żadnym z tych departamentów, stanowili kategorię 3. Te zasady klasyfikacji zostały stworzone przez poprzedniego dyrektora Departamentu Windykacji w oparciu o intuicję i kilkudziesięcioletnie doświadczenie zawodowe.

Steve rozumiał, że jego celem jest zwiększenie zyskowności Departamentu Windykacji przez poprawę nadawania priorytetów w pracy. Natychmiast zauważył, że priorytetyzacja dokonywana według aktualnych zasad nie opiera się na podejściu, w którym na podstawie danych można by prognozować, którzy klienci najprawdopodobniej spłacą przynajmniej część zadłużenia oraz jakiej wielkości spłaty można się spodziewać. Aktualna polityka w tym

zakresie polegała na uproszczonym podejściu do priorytetyzacji obłożenia pracą z wykorzystaniem wiedzy o dotychczasowych zachowaniach klienta do podjęcia decyzji, w jaki sposób odzyskiwać dług obecnie i w przyszłości. Zmienne stosowane w tych zasadach biznesowych mogą mieć pewne zalety, na przykład słuszne wydaje się założenie, że klienci, którzy wcześniej płacili Shu Money Financial, prawdopodobnie będą bardziej skłonni zapłacić w przyszłości. Może jednak istnieć wiele innych czynników pozwalających prognozować prawdopodobieństwo otrzymania przyszłych wpłat, ich przewidywaną wysokość oraz to, jak będzie reagował klient i jakie będą najkorzystniejsze sposoby współpracy z nim. Na takie pytania można poszukiwać odpowiedzi, stosując podejście oparte na danych. W tym celu trzeba zebrać odpowiednie dane o klientach, a potem systematycznie je analizować, nie polegając wyłącznie na intuicji i doświadczeniu. Steve uświadomił sobie, że gdyby potrafił przewidzieć, którzy klienci z największym prawdopodobieństwem spłacą najwyższe kwoty, mógłby zoptymalizować priorytetyzację klientów pod kątem zasobów ludzkich, czasu i pieniędzy, jakimi dysponował jego departament. W tym schemacie długi o najniższej prawdopodobnej kwocie do odzyskania otrzymywałyby najniższy priorytet albo byłyby sprzedawane zewnętrznej agencji windykacyjnej.

Steve umówił kolejne spotkanie z zespołem specjalistów ds. danych, aby przedyskutować swój problem i zrozumieć, na czym polegają proponowane rozwiązania. Przed spotkaniem przejrzał podstawowe informacje na temat danych, systemów danych oraz architektury danych. Jego skromnym celem była możliwość zadawania dobrych pytań.

**Twoi analitycy przedstawiają skomplikowane modele, ale ich rekomendacje nie przekładają się na realne działania? Raporty pełne wskaźników i wykresów nie dają ci jasnych odpowiedzi? Czujesz, że dane w twojej firmie mają ogromny potencjał, ale nie wiesz, jak go wykorzystać?**

*Analiza danych dla menedżerów* to przewodnik, który pozwoli ci przekształcić dane w konkretne decyzje biznesowe.

Dzięki tej książce:

- zrozumiesz kluczowe pojęcia analityki danych i zaczniesz skutecznie komunikować się z zespołem specjalistów
- nauczysz się zadawać właściwe pytania, które pomogą ci oddzielić wartościowe analizy od zbędnego szumu informacyjnego
- poznasz każdy etap zarządzania projektem związanym z analizą danych, od zrozumienia podziału ról w zespole po dobór odpowiedniego oprogramowania
- dowiesz się, jak ocenić jakość prezentowanych wyników, tak by podejmować trafne decyzje i unikać kosztownych błędów
- odkryjesz praktyczne metody wykorzystywania danych w strategii biznesowej i zwiększysz efektywność swojej organizacji.

Howard Steven Friedman i Akshay Swaminathan przekazują wiedzę w sposób prosty i przystępny. Zamiast skomplikowanego języka technicznego oferują jasne, praktyczne wskazówki, które można zastosować od razu.

**Nie pozwól, by analityka danych była w twojej firmie czarną skrzynką.  
Zdobądź kompetencje, które pozwolą ci lepiej wykorzystywać dane,  
optymalizować procesy i osiągać lepsze wyniki.**

Patroni:

THINKTANK®

sprawnymarketing

Książka dostępna także jako e-book.

**mtbiznes.pl**

ISBN 978-83-8231-469-4



9 788382 314694

MT24021

Cena 74,90 zł