

Myślenie jak zespół zajmujący się analizą danych

Skoro już wiesz, jak ustrukturyzować swój zespół i podzielić obszary odpowiedzialności, w jaki sposób możesz upewnić się, że zespół ds. analityki danych myśli jak zespół? W tym rozdziale pomogę ci, analizując kilka typowych sposobów na utrzymanie zespołu na dobrej drodze. Najpierw omówię, jak powstrzymać się od zgłaszania bez uzasadnienia, zadając ciekawe pytania. Następnie zbadam, jak znaleźć właściwy sposób myślenia dla całego zespołu. I na koniec dowiesz się, jak zrozumieć dane i uzyskasz kilka wskazówek, jak znaleźć wyjście z zawieszonych zespołu.

Unikanie raportowania bez uzasadnienia

Jeśli nie znasz się na statystykach, świetnym miejscem na rozpoczęcie jest *Naked Statistics: Stripping the Dread from the Data* autorstwa Charlesa Wheelana, byłego profesora polityki publicznej z Uniwersytetu w Chicago. To przyjemna lektura i świetne wprowadzenie do analizy statystycznej. W książce omawia niebezpieczeństwa związane z używaniem raportów do wyciągania niechlujnych wniosków. Jeśli wiesz, czego szukać, wszędzie możesz zobaczyć niechlujne wnioski. Jednym z miejsc, w których to zobaczysz, są witryny internetowe z wiadomościami. Profesor Wheelan wyobraża sobie serwis informacyjny z przyciągającym uwagę nagłówkiem: „Ludzie, którzy robią krótkie przerwy w pracy, znacznie częściej umierają na raka”. Brzmi dość przerażająco. Według tego badania przeprowadzonego na 36 000 pracowników, ci, którzy zgłaszali, że robili kilka dziesięciominutowych przerw każdego dnia, byli o 41% bardziej narażeni na zachorowanie na raka w ciągu następujących pięciu lat. Niedbały zespół analityków danych może zgłosić to reszcie organizacji, a firma może podjąć dziwaczne kroki, takie jak zamknięcie drzwi, aby zapewnić wszystkim bezpieczeństwo. To może wydawać się skrajnym przykładem, ale tak jest znacznie częściej niż myślisz. Odpowiedzialny zespół data science nigdy nie wyciągałby wniosków z tego typu raportu. Zamiast tego prowadzący badania wykorzystałby je do zadawania interesujących pytań. Dlaczego wstawanie z biurka miało być tak zabójcze? Kim są ci ludzie, którzy robią kilka 10-minutowych przerw? Dlaczego ludzie, którzy nie robią tych przerw, są o wiele bezpieczniejsi? Być może kierownik badań będzie musiał zbadać te pytania, rozmawiając z osobami, które znają tych pracowników. Może ktoś rozpozna to zachowanie. Jeśli pracowałeś w biurze, prawdopodobnie wiesz, że każdy, kto wstaje kilka razy dziennie przez 10 minut, prawdopodobnie wychodzi na zewnątrz, aby zapalić papierosa. Pamiętaj, że korelacja nie oznacza związku przyczynowego. W tym przypadku problem polegał na tym, że pracownicy ci byli częstymi palaczami. Połączenie z nimi wstawanie co 10 minut było po prostu przypadkowe. To palenie było prawdziwym niebezpieczeństwem. Pracując w zespole ds. analityki danych, pamiętaj, że najlepszą ochroną przed niechlujnym raportowaniem jest praca z prowadzącym badaniem w celu zadawania interesujących pytań. Pamiętaj, że wnioski są łatwe. To eksploracja i rozumowanie są trudne. Będą one źródłem twoich największych spostrzeżeń. Między kierownikiem badań a analitykiem danych powinno istnieć zdrowe napięcie. Analityk danych będzie szukał danych do wyciągnięcia wniosków. Kierownik badania zawsze będzie miał dalsze pytania. Analityk danych przedstawi raporty, a kierownik badania przetestuje słabe strony tych raportów. W końcu powinno to pomóc w stworzeniu silniejszej analizy. Niedbałe rozumowanie jest znacznie większym problemem, niż mogłoby się wydawać. Poprzedni przykład był tylko fałszywym nagłówkiem internetowym, ale czasami rzeczywistość może być niebezpieczna. W latach 90. lekarze zaobserwowali, że kobiety, które przyjmowały suplementy estrogenowe, były o jedną trzecią mniej narażone na ataki serca. Masowe badanie 122 000 kobiet wykazało, że istnieje ujemna korelacja między suplementami estrogenowymi a zawałami serca. Wielu lekarzy stwierdziło, że estrogen ma działanie ochronne, które może być korzystne dla zdrowia kobiet. Do 2001 roku 15 milionów kobiet przyjmowało suplementy estrogenowe, powszechnie znane jako hormonalna terapia zastępcza. Potem wydarzyła się dziwna rzecz. Lekarze zaczęli zauważać, że kobiety, które stosowały hormonalną terapię zastępczą, były znacznie bardziej narażone na udar, chorobę serca

i raka piersi. Wiele lat później lekarze zaczęli analizować wyniki. Wielu z nich stwierdziło, że we wstępnym badaniu kobiety, które wydawały się zdrowsze, miały czynniki zewnętrzne - były bogatsze, bardziej świadome zdrowia i miały większą szansę na doskonałą opiekę medyczną. Estrogen nie był prawdopodobną przyczyną ich dobrego zdrowia. Te dane są nadal przedmiotem dyskusji. Akceptuje się jednak fakt, że ta niechlujna analiza przyczyniła się do śmierci tysięcy kobiet. Najlepszą obroną przed niechlujnym rozumowaniem jest twórcze napięcie między kierownikiem badania a analitykiem danych. Jeśli nie zadajesz właściwych pytań, istnieje większe prawdopodobieństwo, że wyciągniesz błędne wnioski.

Posiadanie właściwego nastawienia

Widziałeś niebezpieczeństwa niechlujnego raportowania. Zastanówmy się teraz przez chwilę nad właściwym nastawieniem. Wiele organizacji uważa, że nauka o danych to tylko umiejętność analityczna, że w pokoju jest grupa analityków, którzy po prostu składają sprawozdania ze swoich danych liczbowych. Te liczby są postrzegane jako prawda, ponieważ liczby nie kłamią, ale liczby kłamią. Widziałeś przykłady statystyk opowiadających niewłaściwą historię w Części 4, gdzie dwóch polityków opowiadało różne historie na podstawie tych samych liczb. Badania mogą prowadzić do fałszywych wniosków. Ludzie mogą błędnie interpretować dane. Ich historia może być niepełna. Widziałeś już, że zadawanie pytań jest najlepszą obroną przed niechlujnymi wnioskami. Jak więc wymyślić lepsze pytania? Aby się tam dostać, musisz mieć właściwy sposób myślenia. Dobrą wiadomością jest to, że w ciągu ostatnich kilku lat wykonano w tym obszarze dużo pracy. Wiele różnych dziedzin połączyło się, aby lepiej widzieć, jak myślą ludzie. Inżynierowie komputerowi stosują je do uczenia maszynowego i sztucznej inteligencji. Badania wykazały, że myślenie analityczne niekoniecznie będzie dobrze służyć nauce o danych. Raporty i dane to dopiero pierwszy krok. Następnym krokiem jest myślenie koncepcyjne – umiejętność patrzenia na dane i łączenia ich z własną intuicją. Myślenie koncepcyjne pomoże zespołowi zidentyfikować obszary, na których należy skoncentrować swoje pytania. Istnieje interesująca książka na ten temat zatytułowana *A Whole New Mind: Why Right-Brainers Will Rule the Future* autorstwa Daniela Pinka. 3 W książce twierdzi, że zbliżamy się do końca ery informacji, a skupianie się tylko na liczbach i raportach nie jest tak cenne. Prawdziwa wartość będzie pochodzić z wiedzy, którą tworzymy. Nazywa to wiekiem konceptualnym. Nauka o danych znajduje się na samym skraju tej ery konceptualnej. W pewnym momencie narzędzia analityczne będą na tyle proste, że dostęp do danych będzie miało znacznie więcej osób. Wkrótce użytkownicy uzyskają dostęp do narzędzi do analizy danych, które są tak samo dostępne jak WordPress czy LinkedIn firmy Microsoft. Dzięki temu więcej osób uzyska dostęp do interesujących informacji. Na razie zespół data science odpowiada zarówno za dane, jak i leżące u ich podstaw koncepcje, co pod wieloma względami jest znacznie trudniejsze. Zespół będzie musiał wykorzystać swoje umiejętności analityczne jako punkt wyjścia, a następnie użyć zupełnie nowego zestawu umiejętności koncepcyjnych. W książce Daniela Pinka przechodzi przez kilka z tych umiejętności. Nazywa je „zmysłami” wieku konceptualnego. Zaadaptowałem te zmysły do trzech wartości zespołowych. Te wartości powinny pomóc Twojemu zespołowi myśleć o danych na poziomie konceptyjnym.

Opowiadanie historii nad raportowaniem

Pierwsza wartość to opowiadanie historii nad raportowaniem. Twój zespół ds. analizy danych powinien starać się dostarczać ciekawe historie na temat danych. Powinieneś opowiedzieć przekonującą narrację. Najczęściej Twoje dane będą interpretowane, co oznacza, że może być więcej niż jedna historia. Łatwiej wymyślić pytania, jeśli myślisz o ciekawej historii. Pomyśl o danych jak o postaciach w sztuce. Zapytaj, dlaczego robią jedną rzecz zamiast drugiej, a następnie zadaj pytania dotyczące ich zachowania.

Symfonia nad szczegółami

Druga wartość to symfonia nad detalem. Jednym z najlepszych sposobów uczestnictwa w erze informacji była specjalizacja. Widać to w sposobie zatrudniania firm – na przykład inżynierowie baz danych często specjalizują się w zaledwie kilku platformach. Będziesz chciał odejść od tej specjalizacji w swoim zespole analityków danych. Chcesz, aby zespół stworzył symfonię, łącząc kilka różnych historii i tworząc większy obraz. Widziałeś już to na kilku przykładach: widząc ogólny obraz, w którym palacze odchodzą od biurka i rozumiejąc, dlaczego klienci kupują buty do biegania na początku roku. Te historie wymagają wprowadzenia wielu różnych rodzajów danych, aby uzyskać lepsze zrozumienie.

Empatia nad pewnością

Ostateczną wartością jest empatia nad pewnością. Wiedza o tym, co motywuje ludzi, jest jednym z najlepszych sposobów na zadawanie pytań dotyczących Twoich danych. Twój zespół ds. analityki danych będzie chciał zrozumieć, o czym myśli Twój klient i co jest dla niego ważne. Pamiętaj, że data science może analizować działania milionów ludzi. Jeśli Twój zespół zrozumie, co go motywuje, może zadawać ciekawsze pytania. Należy pamiętać, że aby odnieść sukces, zespół zajmujący się analizą danych będzie musiał użyć zupełnie nowego zestawu umiejętności. Aby zadawać dobre pytania, musisz myśleć koncepcyjnie. Spróbuj wykorzystać te wartości zespołowe jako sposób na przypomnienie sobie, że nauka o danych nie polega na prostym raportowaniu. Twój zespół musi wykorzystać swoje umiejętności koncepcyjne, aby zadawać dobre pytania i tworzyć wiedzę organizacyjną.

Zagłębianie się w „tworzenie sensu”

Jak widzieliście, bardzo ważne jest, aby zespół miał właściwe nastawienie. Twój zespół powinien myśleć koncepcyjnie o danych. To nie jest łatwe zadanie. Pamiętaj, że większość analityków danych wywodzi się z matematyki lub statystyki, czyli dziedzin, które zwykle opierają się na ustrukturyzowanych danych. Myślenie koncepcyjne często wymaga stylu opowiadania historii, który jest bardziej kreatywny i artystyczny. Wielu analityków danych musi polegać na swoich umiejętnościach, a nie na szkoleniu. Gdy wszyscy są już we właściwym nastawieniu, mogą zacząć koncentrować się na tworzeniu sensu. Tworzenie sensu jest kluczową częścią nauki o danych. Wszyscy w taki czy inny sposób tworzymy sens. Kiedy wrócisz z wakacji, poczta e-mail w Twojej skrzynce odbiorczej może Cię przytłoczyć. Możesz zdecydować się na utworzenie podfolderów oznaczonych jako „ważne”, „stare” lub „ogłoszenia” lub możesz zdecydować się na sortowanie wszystkich wiadomości e-mail według nadawcy. Każda osoba może inaczej podejść do danych. Jako zespołowi jeszcze trudniej jest mieć wspólny sens danych. Każda osoba ma swoją własną technikę tworzenia sensu. Te poglądy nie mogą się pokrywać. To, co ma sens dla jednej osoby, może wyglądać jak strata czasu dla drugiej. Tworzenie sensu może być ogromnym wyzwaniem dla zespołów zajmujących się analizą danych. Zespoły te będą próbowały zrozumieć bardzo duże zbiory danych, a informacje mogą być przytłaczające. Może to prowadzić do zamrożenia zespołu, co oznacza, że zespół ma tak dużo danych, że nie jest pewien, od czego zacząć. Jeśli jesteś poza zespołem, może być trudno zobaczyć, kiedy zespół się zawiesza. Kiedyś pracowałem dla firmy, która próbowała nadać sens ogromnemu nowemu zestawowi danych. Firma zebrała niewielkie ilości danych, a następnie zakupiła duże zestawy danych od firmy zewnętrznej. Mieli nadzieję, że będą mogli połączyć swoje istniejące dane z tymi większymi zestawami danych, aby lepiej zrozumieć swojego klienta. Z pomocą drugiej firmy byli w stanie szybko wprowadzić te nowe zestawy danych do swojego klastra. Problem polegał na tym, że zespół analityków danych nie mógł dowiedzieć się, od czego zacząć. Zmagali się z nadawaniem sensu. Zamarli i po prostu stworzyli raporty, które pokazywały, ile mają danych. Były zamrożone na miesiące. Za każdym razem, gdy mieli spotkanie biznesowe, zespół po prostu pokazywał ładniejsze raporty danych o danych. Pobrali drogie narzędzia do wizualizacji i stworzyli wspaniałe wykresy. Ale w końcu nie zadawali żadnych interesujących pytań. Jeśli jesteś w

zespołe zajmującym się analizą danych, spróbuj poszukać oznak, że możesz mieć problemy z nadawaniem sensu. Jeśli wszyscy koncentrują się na narzędziach, może to oznaczać, że Twój zespół jest przytłoczony danymi. Uważaj na puste demonstracje, które zastępują naukę o danych. Jednym ze sposobów na ominięcie zamrożenia danych jest czerpanie większej przyjemności z danych. Możesz wymyślić śmieszne pytania, które mogą wydawać się bezwartościowe. Spróbuj wymyślić pytania, aby sprawdzić, czy ktoś jest psem, czy kotem. Może spróbuj odgadnąć czyjś wzrost, po prostu patrząc na jego historię zakupów. Pamiętaj, że pytania często prowadzą do większej liczby pytań. Aby zacząć, często wystarczy zdobyć dane i bawić się nimi. Duże odkrycia twojego zespołu będą pochodzić z przypadku, co oznacza, że im więcej bawisz się danymi, tym większe prawdopodobieństwo, że znajdziesz coś interesującego. Nie bój się bawić (ale nie nazywaj tego ludźmi spoza zespołu). Wewnątrz zespołu po prostu odbijaj pytania, aż coś się przyklei. Często istnieje cienka granica między eksperymentami a graniem w gry. Im więcej jesteś w danych, tym lepiej wyczuwasz interesujące pytania. Postaraj się zrozumieć, że każdy członek Twojego zespołu może podejść do danych w inny sposób. To kluczowa część tworzenia sensu. Powinieneś wywołać to wyzwanie i poszukać sposobu na wspólne zrozumienie danych. Ważne jest również, aby rozpoznać, kiedy Twój zespół się zawiesza. Wiele zespołów zajmujących się analizą danych utknie, patrząc na narzędzia podczas tworzenia raportów. Na koniec pamiętaj o dobrej zabawie. Możesz uzyskać najlepsze perspektywy, bawiąc się danymi.

PODSUMOWANIE

Poznałeś kilka typowych sposobów na utrzymanie zespołu na dobrej drodze. Obejmują one, jak powstrzymać ich przed zgłaszaniem bez uzasadnienia, upewniając się, że zadają interesujące pytania. Dowiedziałeś się również, jak mieć właściwe nastawienie do nauki o danych. Zespół musi zrozumieć dane, myśląc koncepcyjnie i ucząc się opowiadać historie