



Operačný program
**Efektívna
verejná správa**



Európska únia
Európsky sociálny fond

Data – driven organizácia vo verejnej správe



DATALAB

**Výstup č. 2 - Vypracovanie
návrhu pre Web presense -
online príručky pre využívanie
údajov vo verejnej správe**

Zmluva o dielo 263/2018

Projekt:

**Zlepšenie využívania údajov vo verejnej
správe**

ITMS kód projektu:

314011S979

Tento projekt je podporený z Európskeho sociálneho fondu.

OBSAH

1	<i>Manažérske zhrnutie</i>	3
2	<i>Vízia Data-driven organizácie vo verejnej správe</i>	4
3	<i>Kvalita dát</i>	7
4	<i>Zber dát</i>	9
5	<i>Analýza dát</i>	11
6	<i>Návrh výkonnostných ukazovateľov</i>	13
7	<i>A/B testovanie</i>	15
8	<i>Príbehy a vizualizácia</i>	17
9	<i>Lepšie rozhodovanie na základe dát</i>	20
10	<i>Kultúra</i>	23



1 Manažérske zhrnutie

“Topíme sa v dátach. Náš problém nie je: ‚Ako získame viac údajov?‘ Naším problémom je: ‚ako ich spracujeme?‘” - Anthony Scriffignano, Chief Data Scientist, Dun & Bradstreet

Mnohé organizácie si myslia, že jednoducho preto, že generujú množstvo reportov alebo majú mnoho dashboardov, sú automaticky riadené údajmi. Hoci tieto činnosti sú súčasťou toho, čo by mali robiť, nezabezpečí im to automaticky maximálnu penetráciu využívania údajov v praxi. Tento fakt je rovnako platný aj pre inštitúcie verejnej správy.

V tejto príručke sa venujem dvom kľúčovým otázkam:

- Čo pre organizáciu znamená byť ‘založený na údajoch’?
- A ako toho dosiahnuť?

Samotné údaje sú často len deklaráciou minulých alebo súčasných faktov bez veľkého kontextu, bez vysvetlenia, prečo sa niečo stalo a bez odporúčaní, čo ďalej. Na druhej strane máme k dispozícii analýzy, ako napríklad prediktívne modely, ktoré zahŕňajú odpovede na otázky „prečo?“, odporúčania a predpovede a rozprávanie príbehov o zisteniach. Tieto poznatky a odporúčania sú často kľúčovou hnacou silou v organizácii založenej na údajoch.

Kľúčom k úspechu je aj kultúra. Ide o mnohostranný problém, ktorý zahŕňa kvalitu a zdieľanie údajov, prijímanie a školenie analytikov, komunikáciu, analytickú organizačnú štruktúru, dizajn metrík, A / B testovanie, rozhodovacie procesy a ďalšie. Okrem toho v tejto príručke zdôrazňujeme úlohu, ktorú môžu hrať analytici a ich manažéri. Organizácia založená na údajoch a požadovaná kultúra by sa mala budovať nielen z vedenia zhora nadol, ale aj zdola nahor. V skratke, prvé kapitoly tejto príručky sa zaoberajú samotnými údajmi, najmä výberom správnych zdrojov údajov a zabezpečením ich vysokej kvality a dôveryhodnosti. Ďalším krokom je analýza, pre ktorú potrebujete správnych ľudí so správnymi zručnosťami a nástrojmi, ktoré zabezpečí generovanie efektívnych výsledkov. Nasledujúce kapitoly sa venujú samotnej analytickej práci, ako sú analýzy, návrhy metrík, A / B testovanie a rozprávanie príbehov. Ďalej nasleduje problematika aplikácie výsledkov analýz a novo objavených prístupov. Tu sa zaoberáme tým, čo robí rozhodovanie náročné a ako ho možno zlepšiť.

Vo všetkých kapitolách je veľmi silné posolstvo a téma: byť ‘data-driven’ nie je len o údajoch alebo najnovších súboroch Big Data, ale je to predovšetkým kultúra v organizácii. Kultúra je dominantným aspektom, ktorý stanovuje očakávania o tom, do akej miery sú údaje demokratizované, ako sa využívajú v rámci organizácie a ako sa zdroje a odborná príprava investujú do využívania údajov ako strategického aktíva. Kultúru však možno formovať a ovplyvňovať aj zdola nahor. V celej príručke sa teda priamo obraciame na analytikov a ich manažérov, pričom zdôrazňujeme, čo môžu urobiť pre ovplyvnenie tejto kultúry a maximalizáciu ich vplyvu na organizáciu.

“Začnite s údajmi zaobchádzať ako s aktívami a nie ako s vedľajšími produktmi.” - Peter Aiken, majiteľ Data Blueprint

Tento dokument vznikol ako podklad webového obsahu pre stránku Dátovej kancelárie. Slúži ako návod pre inštitúcie verejnej správy ako riešiť kľúčové otázky dátovej transformácie a zároveň ako katalóg dôležitých odkazov.



2 Vízia Data-driven organizácie vo verejnej správe

Aby bola organizácia založená na údajoch účinná a navigovala sa v konkurenčnom prostredí, získa, spracúva a využíva údaje najlepším možným spôsobom. Existuje mnoho spôsobov, ako posúdiť, či je organizácia riadená údajmi. Niektorí merajú, koľko údajov sa generuje. Iní sa zameriavajú na sofistikovanosť údajov, ktoré používajú, alebo na proces internalizácie údajov. Čo potrebujete, aby ste sa stali dátami riadenou organizáciou? Oveľa viac ako prístup k Big Dátam alebo tím dátových vedcov. Vyžaduje si to vytvorenie efektívnej, hlboko zakorenenej kultúry orientovanej na údaje, čo zahŕňa:

- Z začiatku zdola nahor: naučiť sa správne zhromažďovať správne údaje
- Najímať analytikov s potrebnými zručnosťami a organizovať ich do tímov
- Používať štatistické a vizualizačné nástroje a metódy vychádzajúce z faktov
- Zhromažďovať a analyzovať údaje a rešpektovať súkromie a etiku
- Pochopiť, ako môžu analytici a ich manažéri pomôcť podnietiť kultúru založenú na údajoch
- Pochopiť dôležitosť pozícií, ako sú vedúci oddelenia údajov (chief data officer) a vedúci analytik (chief analytics officer)

Predpoklad #1: Organizácia musí zbierať údaje

Kľúčovou zložkou sú nepochybne údaje. Samozrejme, musia to byť správne údaje. Súbor údajov musí byť relevantný, včasný, presný, čistý, objektívny; a čo je najdôležitejšie, musí byť dôveryhodný. Niektorí veľkí predajcovia údajov a poskytovatelia služieb inzerujú veľké údaje ako všeliek: ak zbierate všetko, niekde tam sú aj diamanty, ktoré urobia každú spoločnosť úspešnou. Pravdou je, že samotné údaje nestačia. Malé množstvo čistých, dôveryhodných údajov môžu byť oveľa cennejšie ako nespočet nezmyslov.

Predpoklad #2: Údaje musia byť pripojiteľné, zdieľateľné a vyhľadateľné.

Presné, včasné a relevantné údaje však nie sú dostatočné. Musia byť tiež:

- Pripojiteľné – údaje musia byť v takej forme, aby sa mohli v prípade potreby pripojiť k iným údajom. Existuje mnoho možností na správu neštruktúrovaných a distribuovaných dát, ako sú relačné databázy, NoSQL alebo Hadoop. Lepšie nástroje vám umožňujú vytvárať hlbšie a bohatšie analýzy.
- Zdieľateľné – v rámci organizácie musí existovať kultúra zdieľania údajov. Údaje v silách vždy obmedzujú rozsah toho, čo sa dá dosiahnuť.
- Vyhľadateľné – všetky prehľady a analýzy vyžadujú filtrovanie, zoskupovanie a zhromažďovanie údajov, aby sa veľké množstvá nespracovaných údajov zmenšili na menšiu množinu údajov vyššej úrovne, ktoré nám pomáhajú pochopiť, čo sa deje. Analytici musia mať nástroje, ktoré im umožňujú relatívne ľahko vypočítať tieto metriky.

Predpoklad #3: Potrebujete ľudí so správnymi schopnosťami na používanie údajov

Stručne povedané, údaje samotné nezachránia vašu organizáciu. Aby organizácia bola riadená údajmi, musia existovať ľudia, ktorí kladú správne otázky týkajúce sa údajov, ktorí majú zručnosti na získanie správnych údajov a metrík.



Čo robí dobrého dátového vedca:

- Technická odborná znalosť: najlepší vedci v oblasti údajov majú zvyčajne hlboké odborné znalosti v niektorých vedeckých disciplínach.
- Zvedavosť: túžba ísť pod povrch a objaviť a oddestilovať problém do veľmi jasného súboru hypotéz, ktoré možno testovať.
- Storytelling: schopnosť používať údaje na rozprávanie príbehu a na efektívne komunikovanie.
- Múdrosť: schopnosť pozrieť sa na problém rôznymi, kreatívnymi spôsobmi.

Predpoklad #4: Špecifické činnosti pre dátami riadené organizácie:

- Dátami riadené organizácie neustále testujú. Môže to byť testovanie A / B, tok na webovej stránke alebo testovanie vplyvu v kampani. LinkedIn napríklad vykonáva 200 experimentov za deň. Testy môžu tiež zahŕňať testovanie priamo s užívateľmi s cieľom získať priamu spätnú väzbu o možných nových funkciách alebo službách.
- Dátami riadené organizácie sú zapojené do prediktívneho modelovania, prognózovania, ale čo je najdôležitejšie, kŕmia predikčné chyby a iné učenia späť do modelov, aby sa zlepšili.
- Dátami riadené organizácie takmer určite budú vyberať medzi budúcimi možnosťami alebo akciami na základe údajov. Analytici musí informovať a ovplyvňovať tých, ktorí rozhodujú. Technológia a školenia môžu urobiť prvú časť: umožniť analytikom vykonávať analýzy a zverejňovať svoje zistenia. Je to však kultúra, ktorá vytvára spôsob myslenia, v ktorom môžu byť údaje dôverované a použité na určenie ďalších krokov

Predpoklad #5: Potrebná IT podpora

Staršie IT štruktúry môžu brániť novým typom získavania, ukladania a analýze údajov. Existujúce architektúry IT môžu zabrániť integrácii informácií a správa neštruktúrovaných údajov často zostáva nad rámec tradičných možností IT. Plné riešenie týchto problémov často trvá roky. Vedúci organizácie však môžu riešiť krátkodobé potreby pomocou rýchlej identifikácie a prepojenia najdôležitejších údajov na použitie v analytike a následnej montáži čistiacej operácie na synchronizáciu a zlúčenie prekrývajúcich sa údajov a na prácu s chýbajúcimi informáciami.

Predpoklad #6: Zabudovaná analýza v jednoduchých nástrojoch pre predné línie

Manažéri potrebujú transparentné metódy na každodenné používanie nových modelov a algoritmov. Potrebné sú terabajty dát a sofistikované modelovanie, aby sa zlepšilo riadenie rizík a operácie. Kľúčom k úspechu je oddeliť štatistických expertov od manažérov, ktorí využívajú informácie založené na údajoch. Cieľom je poskytnúť manažérom frontline intuitívne nástroje a rozhrania, ktoré im pomôžu s ich prácou.

Predpoklad #7: Rozvíjanie schopností vedúcich k plnému využívaniu veľkých dát

Aj pri jednoduchých a použiteľných modeloch bude väčšina organizácií potrebovať aktualizovať svoje analytické schopnosti a gramotnosť. Aby analytika bola súčasťou štruktúry každodenných operácií, manažéri ju musia vnímať ako ústrednú pri riešení problémov. Úprava kultúr a myslenia zvyčajne vyžaduje mnohostranný prístup, ktorý zahŕňa vzdelávanie, modelovanie úloh a metriky na posilnenie správneho správania. Zamestnanci napríklad profitujú z prístupu, v ktorom sa zúčastňujú na reálnych rozhodnutiach na pracovisku založených na analytických údajoch, ktoré im umožňujú učiť sa činnosťou.

Dodatočné zdroje:

„[For Big-Data Scientists, ‘Janitor Work’ Is Key Hurdle to Insights](#)“ publikované v The New York Times.

„[Data Scientists Driving Business Advantage](#)“ publikované v The Information Management



[“Google Exec Marissa Mayer Explains Why There Aren’t More Girl Geeks”](#) publikované v The Huffington Post

„[Reporting vs. Analysis: What’s the Difference?](#)“ publikované v The Adobe Blog

„[Big Data, Analytics and the Path From Insights to Value](#)“ publikované v MIT Sloan Management Review

„[Data Quality and the Bottom Line](#)“, autor: Wayne W. Eckerson

„[Data Analysis with Open Source Tools](#)“, autor: Philipp Janert

„[Building Data Science Teams](#)“, autor: DJ Patil

„[Predictive Modeling With Big Data: Is Bigger Really Better?](#)“, autor: Enric Junqué de Fortuny

„[3D Data Management: Controlling Data Volume, Velocity, and Variety](#)“ publikované v Meta Group



3 Kvalita dát

Aspekty kvality dát

Existujú stupne kvality, niektoré problémy sú vážnejšie ako iné. Závažnosť týchto problémov však môže závisieť od kontextu analýzy, ktorá sa má vykonať s údajmi. Konkrétnejšie, kvalita údajov má niekoľko aspektov. Údaje by mali byť:

Prístupné – Analytik musí mať prístup k údajom. Týka sa to nielen oprávnení, ale aj príslušných nástrojov, ktoré umožňujú použiteľnosť údajov. Všetky údaje musia byť dostupné v databáze pre analytikov, aby mohli údaje používať.

Presné – Hodnoty predstavujú skutočnú hodnotu alebo stav subjektu. Nepresné údaje predstavujú napríklad nesprávne kalibrovaný teplomer, nesprávny dátum narodenia alebo zastaraná adresa.

- Hodnotenie presnosti) Metódy hodnotenia sa značne líšia v nákladoch a účinnosti. Ideálne porovnanie údajov so skutočnými údajmi býva drahé, zatiaľ čo najmenej nákladné možnosti môžu prinášať zavádzajúce výsledky. Vytvorenie šablóny pre záverečnú správu o presnosti môže pomôcť udržať postupy zamerané na konečný, užitočný výsledok. Počas procesu hodnotenia databázu neaktualizujte.
- Prevencia nepresnosti) Akýkoľvek model nepresnosti predstavuje príležitosť na zlepšenie procesov, ktoré nepresnosti vytvorili. Analýzy koreňovej príčiny z aktuálneho alebo akéhokoľvek predchádzajúceho dátového projektu môžu priniesť ďalšie prehľady.

Koherentné – Údaje sa dajú presným spôsobom kombinovať s inými relevantnými údajmi. Koherencia je riadená súborom ID alebo kľúčov, ktoré spolu viažu údaje v rôznych častiach databázy. Súlad medzi súbormi údajov v rámci jedného oddelenia je výsledkom vnútornej štandardizácie. Konzistentnosť medzi súbormi údajov kontrolovanými rôznymi organizáciami môže byť zložitejšia. Čím ľahšie je kombinovanie vašich údajov s inými údajmi, tým väčšia hodnota sa dá extrahovať, tým pádom väčšia konzistentnosť prináša verejné výhody.

- Vnútorňá štandardizácia) Mali by ste si vybrať jednotný štandard, ktorý budú zdieľať všetci, špecifikovať referenčné údaje a zoznam zainteresovaných strán, ktorých sa zmena bude týkať. Mali by ste sledovať všetky príčiny prekážok štandardizácie.
- Externá štandardizácia) Existujú vzťahy medzi vašimi údajmi a údajmi, ktoré majú iné organizácie vrátane iných vlád. Môžu existovať normy podporované profesijnými združeniami, akademickými inštitúciami, vládou alebo konzorciami menších vlád. Pochopenie štandardov a postupov umožňuje informovanejšie a cieľenejšie rozhodovanie o vlastných postupoch a štandardoch v oblasti údajov.

Úplné – Chýbajúce údaje neexistujú. To môže znamenať jeden údaj v rámci jedného záznamu, napríklad chýbajúce meno v zázname zákazníka, alebo úplné záznamy, ktoré chýbajú, napríklad celý záznam používateľov, ktorý sa nepodaril uložiť do databázy.

- Posúdenie úplnosti) Pomocou nástroja na profilovanie údajov skontrolujte, či sa počet riadkov zhoduje s realitou, a vykonajte iné kontroly údajov na základe zdravého rozumu. Akékoľvek segmentovanie vašich údajov (podľa roku, adresy, typu atď.) možno použiť na vyhľadanie chýbajúcich údajov.

Konzistentné – Údaje sú v zhode. Napríklad adresa pre konkrétneho používateľa v jednom zdroji údajov sa zhoduje s adresou pre toho istého používateľa v inom zdroji údajov. Ak sa vyskytnú konflikty, jeden zdroj by sa mal považovať za hlavný zdroj alebo oba zdroje sa nepoužijú, kým sa neporozumie zdroju nezhôd.

Definované – Jednotlivé dátové polia majú vždy presne definovaný, jednoznačný význam. Dobře pomenované polia sprevádzané dátovým slovníkom zaručujú kvalitu.



Spôľahlivé – Údaje sú kompletne (máte všetky údaje, ktoré by ste mali očakávať) a sú presné (údaje poskytujú správne informácie).

Včasné – Medzi zberom údajov a dostupnosťou alebo uvoľnením analytikom je krátka alebo primeraná doba trvania. Šesť fáz životného cyklu, o ktorých by sme si mali byť vedomí, zahŕňa: plánovanie, získavanie, skladovanie a zdieľanie, udržiavanie, používanie, likvidáciu.

Proaktívna hygiena údajov

Opravte príčiny problémov – Riešenie niektorých alebo všetkých koreňových príčin problémov šetrí zdroje a čas. Vylepšenia sa môžu pohybovať od jednoduchých zlepšení procesov v rámci vašej pôsobnosti až po hlavné iniciatívy, ktoré si zaslúžia byť vnímané ako nové projekty na zlepšenie kvality údajov, ktoré si vyžadujú nové plánovanie.

Deduplikácia – Duplikácie prichádzajú v mnohých formách. Niekedy sú dva riadky v tabuľke presne identické alebo sa líšia len v ich primárnom kľúči. Niekedy môže byť jedna položka opísaná dvakrát v databáze s rôznymi pravopismi. Deduplikácia jednej dátovej tabuľky ovplyvní všetky súvisiace tabuľky. Deduplikácia môže byť náročná na prácu a ak sa vykoná príliš rýchlo alebo nedbalo, môže spôsobiť veľké škody. Pred vykonaním akýchkoľvek nezvratných zmien údajov by ste mali vždy starostlivo naplánovať, pripraviť a otestovať metodiku. Pred aplikáciou metodiky na kmeňové dáta si aspoň jeden ďalší znalec preštuduje metodiku a výsledky testovania metodiky.

Prevenca duplikátov – Profilovanie a deduplikovanie údajov môže odhaliť vzory duplikácie. Či duplikáty boli vytvorené ľuďmi, ktorí zadali jednotlivé záznamy alebo manipuláciou s väčšími údajmi. Z dlhodobého hľadiska je zvyčajne efektívnejšia, aby sa zabránilo vytváraniu duplikácií, než ich opravovať.

Monitorovanie kvality údajov – Pravidelné monitorovanie kvality údajov vám pomôže zachytiť a opraviť problémy skôr, ako spôsobia veľké problémy.

Kultúra dátovej kvality – Dlhodobé, udržateľné zlepšovanie kvality údajov je uskutočniteľnejšie, keď ľudia vo vašej agentúre alebo vláde chápu hodnotu kvalitných údajov a negatívny vplyv údajov s nízkou kvalitou. Tento druh kultúrnej zmeny nemôže byť nútený, ale môže byť podporovaný rôznymi spôsobmi.



4 Zber dát

Čím viac údajov zbierate, tým väčšia je šanca, že pochopíte správanie používateľov a, čo je dôležité, aj kontext ich správania. Čím viac organizácia chápe o jednotlivých užívateľoch, ich potrebách, zámeroch a túžbach, tým viac môže zlepšiť užívateľskú skúsenosť.

Dátami riadená organizácia, bude však pravdepodobne myslieť oveľa širšie: data-driven marketing, data-driven služby, data-driven zákaznícky servis, data-driven rozhodovanie, data-driven HR. Ak má každá z nich sadu interných a externých zdrojov údajov s rôznymi formátmi, latenciou, problémami s kvalitou údajov, požiadavkami na zabezpečenie a súlad atď., potom táto prax zhromažďovania maximálneho množstva údajov zlyhá. Navyše nie je zadarmo. Na udržanie týchto systémov, zálohovanie údajov a na integráciu týchto zdrojov údajov s cieľom poskytnúť holistický pohľad sú vysoké náklady. Zostavujete zdroje údajov, spojenia medzi nimi a schopnosti spracovania v priebehu času. Musíte mať racionálnu, dobre premyslenú stratégiu získavania údajov. Dátové tímy vo väčšine organizácií sú navyše obmedzené na zdroje; nemôžu robiť všetko naraz, a preto musia uprednostniť, ktoré zdroje údajov sa majú veľmi starostlivo riešiť.

Dátami riadená organizácia by sa mala zamerať na hodnotový aspekt. Primárnou motiváciou dátového tímu by malo byť uspokojenie potrieb analytikov. Tímy musia mať svoje základné údaje, navyše môžu mať iný súbor údajov, ktoré rozširujú ich základné údaje. Tieto iné zdroje údajov sú pravdepodobne cenné a účinné, ale nie rozhodujúce.

Keďže sa do analýzy zapája viac ľudí a organizácie sa zaoberajú veľkými údajmi, cloudom a ďalšími novými technológiami (napr. Stream Mining), je potrebné vyvíjať postupy. Napríklad, ako viac ľudí analyzuje údaje, existuje väčšia potreba konzistentných slovníkov a flexibilného prístupu k údajom. Keď do organizácie prichádzajú nové druhy údajov, podnik musí určiť, ako s nimi zaobchádzať. Pre organizácie je rovnako dôležité, aby mali politiky na implementáciu prediktívnych modelov.

Z technologického hľadiska sú nové nástroje vylepšené funkciami, ktoré pomáhajú s riadením údajov a analýz. Niektoré dôležité funkcie zahŕňajú:

- Dátové katalógy, glosáre a slovníky. Dátový katalóg organizácie je súpisom údajov (vrátane metaúdajov), ktoré umožňujú organizáciám stať sa riadenými údajmi tým, že pomáhajú používateľom porozumieť a budovať dôveru v údaje. Podobne aj dátové slovníky pomáhajú používateľom pochopiť bežné slovníky, ktoré sú dôležité pri analýze údajov. Novšie nástroje automatizujú niektoré kroky a postupy na aktualizáciu katalógov, ako aj objavujú metaúdaje z existujúcich súborov údajov, aby sa o nich dozvedeli podrobnosti. Nástroje môžu tiež označovať údaje podľa definícií a pravidiel organizácie na vyššej úrovni a vyhľadávať a používať existujúcu dokumentáciu.
- Dátová línia. Dátová línia umožňuje používateľom dôverovať ich BI a analytike. Pomocou dátovej línie s metaúdajmi môžu organizácie pochopiť, odkiaľ údaje pochádzajú a sledovať, ako sa zmenili a transformovali. To by mohlo zahŕňať, ako by vplyv zmien na jeden dátový prvok mohol ovplyvniť iné systémy.
- Modelové riadenie. Akonáhle je prediktívny model postavený, môže sa časom stratiť a degradovať, takže kontinuálne sledovanie je dôležitou súčasťou procesu riadenia analýzy. Niektoré riešenia nástrojov automatizujú monitorovanie; môžu naplánovať aktualizácie pre model budovy, aby sa ubezpečil, že model je stále aktuálny. Iné nástroje poskytujú automatické upozornenia, keď je model degradujúci.

Zapojenie IT a zainteresovaných strán je kľúčom k spoľahlivému riadeniu. Tento tím, ktorý riadi údaje, sa musí zamerať na svoju flexibilitu, ktorá je dôležitá na pomoc pri budovaní vzťahov medzi vlastníckmi údajov, správcami údajov a inými subjektmi zapojenými do riadenia.

Mali by ste používať nespracované alebo agregované údaje?



Niektorí dodávatelia poskytujú dashboardy, kde boli údaje agregované až po kľúčové metriky, ktoré analytici potrebujú. Tie môžu poskytnúť významnú hodnotu. Analytici chcú ďalej rozrezať a nakrájať údaje spôsobom, ktorý dodávateľský panel predajcu nepodporuje. Aj keď tieto dashboardy môžu byť ľahko použiteľné, zvyčajne je lepšie použiť nespracované údaje tam, kde je to možné, pretože môžete vždy agregovať, ale nie naopak. S nespracovanými údajmi môžete v zásade pracovať podľa vašich predstáv. Hoci existujú prípady, keď nie je možné spotrebovať nespracované údaje, ako je napríklad to, že prvotné údaje sú obrovské a nákladné. Surové údaje sú zvyčajne vždy lepšou voľbou.

Aj keď existuje množstvo informácií z interných dátových systémov organizácie, môžu byť doplnené verejne dostupnými údajmi od tretích strán. Existuje mnoho dôvodov, prečo by ste mali nakupovať externé súbory údajov. Ak vám to zdroje umožňujú, môžete si zakúpiť celú sadu údajov. Nie je to tak vždy. Môžete však byť v situácii, keď platíte mesačný poplatok za informačný kanál. Môžete spustiť experimenty na určenie hodnoty údajov a rozhodnúť sa, či to vedie k pozitívnym výsledkom alebo nie. Ak nie, zrušte službu.

Dodatočné zdroje:

„[Creating a Data-Driven Organization](#)“, autor: Carl Anderson

„[Bigger Is Better, But At What Cost?](#)“, autor: Brian Dalessandro

„[What Is A Data Scientist?](#)“, autor: Carl Anderson



5 Analýza dát

Jeffrey Leek, asistent profesora biostatistiky na Univerzite Johns Hopkins a spoluautor blogu o štatistikách sa domnieva, že existuje šesť typov analýz, usporiadané od najjednoduchších až po najzložitejšie: i.) Deskriptívna analýza; ii.) Prieskumná analýza; iii.) Inferenciálna analýza; iv.) Prediktívna analýza; v.) Kauzálna analýza; vi.) Mechanická analýza;

Deskriptívna analýza je najjednoduchší typ analýzy. Cieľom deskriptívnej analýzy je numericky opísať kľúčové vlastnosti vzorky. Často môže tvoriť údaje, ktoré sa zobrazujú na dashboardoch. Môžu tiež poskytnúť prehľad alebo pohodlie, pokiaľ ide o kvalitu údajov. Deskriptívna analýza je zvyčajne prvým krokom, šancou získať zmysel pre dáta k hlbšej analýze.

Použitie grafiky na preskúmanie a vizualizáciu údajov sa nazýva **prieskumná analýza** (z angl. Exploratory Data Analysis, ďalej len EDA). Grafy nám pomáhajú vidieť väčší obrázok dátového súboru a pomáhajú nám odhaliť zjavné alebo nezvyčajné vzory. EDA nám zvyčajne poskytuje množstvo možností, s ktorými sa dá študovať systém. To zase pomáha analytikovi, aby prišiel s novými hypotézami o tom, čo by sa mohlo diať a s ktorými pákami musíte pracovať. EDA môže tiež vyzdvihnúť medzery v našich vedomostiach a pomôcť určiť, ktoré experimenty by mohli zmysluplne vyplniť tieto medzery.

Cieľom **Inferenciálnej analýzy** je odvodiť niektoré informácie, ktorými môžu byť parametre, distribúcie alebo vzťahy o širšej populácii, z ktorej vzorka prišla. Stanovuje tiež štádium pre testovanie hypotéz, v ktorom navrhujeme a spúšťame experimenty na testovanie a analýzu nášho chápania základných mechanizmov. Typy otázok, na ktoré je možné odpovedať vďaka inferenciálnej analýze zahŕňajú: štandardnú chybu, intervaly spoľahlivosti a mieru odchýlky; Očakávaný priemer jednej vzorky; Výpočty veľkosti vzoriek; regresie a asociácie.

Prediktívna analýza vychádza z inferenciálnej analýzy. Cieľom je poukázať na vzťahy medzi premennými z existujúceho súboru údajov a vytvoriť štatistický model, ktorý dokáže predpovedať hodnoty atribútov pre nové, neúplné alebo budúce dátové body. Prediktívna analýza sa potom môže použiť na generovanie prognóz a budúcich predpovedí v časových radoch, ktoré sa môžu použiť na generovanie plánov. Aplikácie prediktívnej analýzy sú obrovské s obrovskými vplyvmi, napríklad: softvér na predikciu akcií; Filtre proti spamu; Sociálne siete; prediktívny policajný dozor; Predpovedanie dopytu. Kým prediktívna analýza nie je nevyhnutne náročná. Jedným z najťažších krokov je získanie dobrého, čistého súboru údajov. S dobrou knižnicou údajov môže implementácia základného modelu zahŕňať doslovne niekoľko riadkov kódu. Získanie dobrého modelu však často vyžaduje viac úsilia a iterácie a proces nazývaný Feature engineering. Features sú vstupy modelu. Môžu zahŕňať základné surové údaje, ktoré ste zhromaždili, jednoduché odvodené premenné, ako aj zložitejšie abstraktné funkcie.

Osvedčené postupy

Ako analytik: Mali by ste sa snažiť rozšíriť svoju sadu nástrojov, aby ste sa stali efektívnejším a hodnotnejším analytikom. Vyhodnoňte štatistické a vizualizačné schopnosti, ktoré v súčasnosti používate. Poskytol by ďalší analytický prístup na vyššej úrovni väčší prehľad a vplyv na váš projekt? Čo by to znamenalo naučiť sa tejto zručnosti?

Ako manažment: Dávajte pozor na situácie, kde by ďalšie analytické typy mohli poskytnúť lepšiu predstavu a vplyv na organizáciu. Môžete urobiť viac experimentov, ktoré prehĺbia inštitucionálne znalosti kauzálnych faktorov? Pomáhajte analytikom rozvíjať svoje zručnosti a otvorene ich podporujte školením a mentorstvom. Nech experimentujú s inými softvérovými nástrojmi, ktoré by mohli poskytnúť nové typy vhládov. Dodávatelia poskytujú ľahko použiteľné samoobslužné nástroje, ktoré pomáhajú tímom s takými úlohami, ako je príprava a analýza údajov. Mnohé z týchto nástrojov majú v sebe zabudované pokročilé analytické funkcie (napríklad strojové učenie) a sú navrhnuté tak,



aby fungovali v celom životnom cykle analýzy, od zberu údajov a profilovania až po monitorovanie pokročilých analytických modelov v implementácii. Medzi dôležité princípy tejto novej triedy nástrojov patrí automatizácia, opakovaná použiteľnosť a čitateľnosť:

- **Automatizácia.** Automatizácia, ktorá zahŕňa zníženie ľudského zásahu, prichádza v mnohých smeroch. Napríklad, pretože informatici a štatistickí (ľudia, ktorí vytvárajú prediktívne modely) sú v nedostatočnej ponuke, existujú nástroje, ktoré automatizujú proces vytvárania modelu. Analytik napríklad jednoducho špecifikuje cieľovú (alebo výslednú) premennú, ktorá je predmetom záujmu, a softvér vytvorí najlepší model s použitím poskytnutých atribútov. Niektoré nástroje na prípravu údajov uľahčujú vyhľadávanie, zhromažďovanie, integráciu, profilovanie a transformáciu údajov. Niektoré nástroje využívajú strojové učenie a spracovanie prirodzeného jazyka na pochopenie sémantiky a urýchlenie porovnávania údajov. Niektoré nástroje tiež umožňujú tvorcom modelu, aby určil pravidlá pre systém na spustenie výstrah, keď je model degradujúci a je potrebné ho preškoliť. Iné nástroje idú ďalej a vykonávajú automatickú detekciu degradácie modelu na základe určitého parametra.
- **Opakovaná použiteľnosť.** Ďalším aspektom automatizácie je opätovné použitie toho, čo už bolo vytvorené pre správu dát a analýzu. Napríklad pracovné postupy sa často vytvárajú pomocou rozhraní drag-and-drop na zostavovanie dátových potrubí zo zdroja na cieľ. Ten istý pracovný tok možno uložiť a vložiť do analytického pracovného postupu, aby sa vytvoril prediktívny model. Mnohé nástroje poskytujú možnosť ukladania a opätovného použitia pracovných postupov. Niektoré poskytujú aj možnosti plánovania.
- **Čitateľnosť.** Keďže nástroje sa ľahšie používajú, tí, ktorí používajú nástroje, musia pochopiť, čo robia. Napríklad používateľ, ktorý vytvára prediktívny model s automatizovaným nástrojom, musí pochopiť, čo výstup predstavuje. Na tento účel nástroje často obsahujú vysvetlenia toho, čo urobili. Napríklad analytický nástroj využívajúci technológiu neurónovej siete (Neural Network Technology) môže s výsledkami obsahovať vysvetlenie, ako boli výsledky získané.

Dodatočné zdroje:

[“Analyzing the Analyzers”](#), autor: Harlan Harris

[“SQL Tutorial”](#) publikované v w3schools

[“Using SQLite”](#), autor: Jay Kreibich

[“Big Impact from Big Data”](#), autor: Ken Rudin

[“Data Analysis with Open Source Tools”](#), autor: Philipp Janert

[“To Centralize Analytics or Not, That is the Question”](#) publikované v The Forbes

[“Big Data, Analytics and the Path From Insights to Value”](#) publikované v MIT Sloan Management Review

[“Predictive Modeling With Big Data: Is Bigger Really Better?”](#), autor: Enric Junqué de Fortuny



6 Návrh výkonnostných ukazovateľov

Organizácia založená na údajoch musí stanoviť jasnú stratégiu a potom stanoviť súbor metrík na sledovanie toho, či podnik skutočne smeruje správnym smerom a na monitorovanie pokroku a úspechu. Je nevyhnutné, aby boli metriky navrhnuté správne. Mali by pôsobiť ako pravdivé a presné kompas. Metriky sú tiež výstupmi z experimentov a testov A / B, ktoré, ak sú dobre inštruované, poskytnú najväčší základ pre postrehy a stratégie založené na údajoch. V ideálnom svete by metriky mali vykazovať niekoľko nasledujúcich znakov:

Jednoduché metriky sú z definície jednoducho definované, čo znamená, že sú:

- jednoduchšie sprostredkovať ostatným: je menšia šanca na zmätok,
- jednoduchšie na implementáciu: je pravdepodobnejšie, že budú vypočítané správne,
- pravdepodobnejšie, že budú porovnateľné s inými tímami alebo organizáciami.

Štandardizované metriky spôsobia menší zmätok, najmä pre kolegov, ktorí sa pripoja k vašim tímom z iných organizácií. To tiež uľahčí porovnanie vašich metrík s ostatnými v tom istom sektore. Najlepším postupom je mať centralizovaný, automatizovaný, zdokumentovaný jediný zdroj štandardizácie. Vďaka tomu je jednoduchšie vybudovať úložisko, ktorému môže organizácia dôverovať, využívať ho a stavať na ňom.

Metriky by mali byť **presné**. To znamená, že ich priemerná číselná hodnota by mala byť blízka skutočnej strednej hodnote. Pri navrhovaní metrík sa pokúste zamyslieť nad možnými zdrojmi zaujatosti v samotných údajoch aj v metrike.

Metriky by tiež mali byť **precízne**. To znamená, že by mali vrátiť podobné hodnoty, ak sa opakujú za rovnakých podmienok. Jeden nástroj alebo páka, ktorú musíte ovládať, je veľkosť vzorky. Čím väčšia je veľkosť vzorky, tým menšia je štandardná chyba. Pretože štandardná chyba priemeru sa rovná štandardnej odchýlke delenej druhou odmocninou veľkosti vzorky, aby sa štandardná chyba znížila na polovicu, potrebujeme štvornásobnú veľkosť vzorky.

Definujte metriky, ktoré sú štatisticky „**robustné**“, to znamená, že sú relatívne necitlivé voči jednotlivým extrémnym hodnotám. Treba poznamenať, že existujú časy, keď chcete navrhnuť metriku, ktorá je špecificky citlivá na hraničné hodnoty. Robustnosť môžete odhadnúť alebo vizualizovať pomocou prevzorkovania. Vezmite podmnožinu údajov a vypočítajte metriku. Opakujte mnohokrát, aby ste získali množinu metrických hodnôt a vykreslili ich distribúciu. Najlepšou praxou je použiť prieskumnú analýzu dát (napríklad histogramy a bivariantné rozptylové grafy) na získanie pocitu pre údaje a použiť ich na usmernenie výberu vhodných robustných metrík a opatrení.

Kľúčové ukazovatele výkonnosti (KPIs) sú opatrenia najvyššej úrovne spojené so strategickými cieľmi spoločnosti. Sú to opatrenia, ktoré vám pomôžu porozumieť tomu, ako robíte proti svojim cieľom. Je dôležité, aby KPI boli SMART:

- Špecifické (Specific)
- Merateľné (Measurable)
- Dosiahnuteľné (Achievable)
- zamerané na výsledky (Result-oriented)
- časovo viazané (Time-bound)

Ďalšie charakteristiky výkonnostných opatrení:



- ľahko pochopiteľné: KPI by mali byť intuitívne a ľahko stráviteľné zainteresovanými stranami a zamestnancami bez veľkého vysvetlenia.
- transparentné pre zúčastnené strany: Spôsob, akým meriate výkon, by mal byť jasne viditeľný pre všetkých zúčastnených. Takáto transparentnosť pomáha každému pochopiť ich prínos k celkovému výkonu organizácie.
- odrážajú to, čo sa organizácia snaží dosiahnuť - Je príliš ľahké dostať sa do pasce sledovania toho, čo sa dá ľahko merať, keď skutočným cieľom môže byť niekde inde.

Bežné chyby pri navrhovaní KPI zahŕňajú:

- Príliš veľa ukazovateľov: Ak máte príliš veľa KPI, zamestnanci sa rozdelia. Namiesto toho by sa malo zamerať na menší, ucelenejší súbor cieľov a KPI, ktorým môže každý porozumieť.
- Nepripojené k strategickým prioritám: Organizácie, ktoré nemôžu prepojiť ukazovatele výkonnosti s konečným cieľom, sú oveľa menej pravdepodobné, že dorazia na miesto určenia včas a bez úhony.
- Nereálne: Niekedy sú KPI jednoducho mimo realitu. To je veľmi bežné pri KPI výstupov, ktoré znejú veľmi dobre teoreticky, ale v praxi sú oveľa ťažšie dosáhnúť.

Dodatočné zdroje:

„[The 75 KPIs Every Manager Needs To Know](#)“, autor: Bernard Marr



7 A/B testovanie

Jedna z pekných vecí o testovaní A / B je, že nemusíte mať a priori kauzálne vysvetlenie, prečo by niečo malo fungovať; stačí len vyskúšať, preskúmať a nájsť tie zlepšenia, ktoré majú pozitívny vplyv. V spoločnosti Microsoft dve tretiny experimentov nepreukázali vylepšenia. Jediné zlepšenie môže znamenať obrovský rozdiel v konečnom dôsledku.

Osvedčený postup #1: Pred začiatkom testu by ste mali jasne definovať metriky úspechu.

Je dôležité mať jasnú predstavu o cieľoch. Môžete zbierať množstvo údajov, ale pokiaľ ide o analýzu, existuje veľká šanca, že budete vykazovať metriky a výsledky, ktoré len vyzerajú dobre. Tieto prístupy nebudú organizácii prinášať dlhodobú hodnotu.

Osvedčený postup #2: zahrňte A/A testovanie.

Test A/A kontroluje kontrolnú skupinu proti inej kontrolnej skupine, pričom obe sú testované na rovnakom produkte. Existuje niekoľko výhod:

- môžete ho použiť na otestovanie testovacej infraštruktúry a procesov priradenia. Ak narazíte na veľmi rozdielne veľkosti vzoriek v dvoch skupinách, znamená to, že v procese priradenia je niečo nesprávne,
- ak vidíte porovnateľné veľkosti vzoriek, ale veľmi rozdielne výkonnostné metriky, môže to znamenať problém v sledovaní udalostí, analýze alebo v prehľadoch,
- výsledky testu môžete použiť na odhad variability vašej metriky vo vašej kontrole.

Osvedčený postup #3: Používajte kalkulačku veľkosti vzorky.

Pre rôzne situácie existujú rôzne kalkulačky veľkosti vzorky. Určenie veľkosti vzorky je dôležité. Ak testujete s menšou ako potrebnou veľkosťou vzorky, pravdepodobne budete vystavení falošným výsledkom. Ak podceňujete experiment, zvýšite chybovosť. Stručne povedané, budete podliehať viac falošným negatívam a falošným pozitívam.

Osvedčený postup #4: mali by ste usilovať o 50%.

Keď začnete test, cieľom je odkloniť 50% návštevnosti na test. Problém je v tom, že ak sa vyskytnú nejaké veľké softvérové chyby, vystavujete polovicu návštevnosti stránok tejto negatívnej skúsenosti. Namiesto toho môžete prijať bezpečnejší prístup:

- 1% počas 4 hodín
- 5% počas 4 hodín
- 20% počas 4 hodín
- 50% pre zvyšok experimentu

Osvedčený postup #5: Zastavenie testov hneď po tom, čo uvidíte výsledky, vytvorí falošné falošné pozitívne výsledky.

Niektoré programy na testovanie A / B sú navrhnuté takým spôsobom, že je prirodzené neustále monitorovať výsledky testu a zastaviť test hneď, ako sa dosiahne významný výsledok. Falošné pozitívne výsledky môžu ohroziť až 80% výsledkov testov, keď sa testy vykonávajú týmto nesprávnym spôsobom. Niektoré A / B testovacie programy sa v poslednej dobe snažia s týmto problémom vyrovnávať.

Osvedčený postup #6: Viacnásobné testovanie

Nebezpečná nedávna móda bola vykonávať mnoho A / B testov súčasne v nádeji, že niečo bude úspešné. Ďalším variantom je vykonanie „post-test segmentácie“ - rozdelenie vzorky po vykonaní testu, kým nedosiahnete pozitívny výsledok. Jednoduchá aritmetika ukazuje, prečo je to zlý nápad. Každý test



má 5% pravdepodobnosť, že bude mať pozitívny výsledok, aj keď neexistuje žiadny skutočný dôvod. Ak spustíte 20 testov, alebo vyskúšate 20 segmentácií, budete mať v priemere jeden víťazný test a ak sa pokúsite 40, budete mať dva, aj keď varianty nevytvárajú žiadny záver. Namiesto rozptýleného prístupu je najlepšie vykonať malý počet sústredených a dobre uzemnených testov, z ktorých všetky majú primeranú štatistickú silu. Ak naozaj máte zdroje na plytvanie, pokračujte a urobte veľa testov A / A alebo jeden veľký test A / A / A /... Vo všetkých ostatných prípadoch stačí investovať prostriedky do správnych testov A / B alebo A / B / N. Ak si len chcete byť istí v záveroch testov, stačí zvýšiť požadovanú štatistickú úroveň významnosti.

Dodatočné zdroje:

[“The importance of A/A Testing”](#), autor: Jordi Cabot

[“Challenging Data Driven Design”](#), uator: PJ McCormick

[“24 of the Most Surprising A/B Tests of All Time”](#), autor: Elisa Gabbert

[“Planning, Running, and Analyzing Controlled Experiments on the Web”](#), autor: Kohavi, R.

[“How Not To Run an A/B Test”](#), autor: Evan Miller

[“Most Winning A/B Test Results Are Illusory”](#), autor: Martin Goodson

[“Should you do A/A, A/A/B or A/A/B/B tests in CRO?”](#), autor: Georgi Georgiev



8 Príbehy a vizualizácia

Každý súbor údajov, každá databáza, každá tabuľka má príbeh. Úlohou analytika je nájsť tento príbeh, interpretovať ho a komunikovať. Hľadanie príbehu a jeho interpretácia zahŕňa celý rad analytických techník, ktoré zvyčajne zahŕňajú vizualizáciu údajov pomocou grafov a tabuliek. Vizualizácia dát má mnoho komponentov, z ktorých každá vyžaduje starostlivé zváženie. Jediná zlá voľba, ako sú nízko kontrastné farby, malá veľkosť písma alebo nevhodná voľba grafu, môže urobiť celú vizualizáciu škaredou a zbytočnou..

Kontrolný zoznam vizualizácie údajov

Text

- 6-12 slovný názov v ľavom hornom rohu
- Podtitul alebo anotácie poskytujú ďalšie informácie
- Veľkosť textu je hierarchická a čitateľná
- Text je vodorovný
- Údaje sú označené priamo

Usporiadanie

- Pomery sú presné
- Údaje sú zámerne usporiadané
- Intervaly osí sú rovnako vzdialené
- Graf je dvojrozmerný

Farba

- Farebná schéma je úmyselná
- Farba sa používa na zvýraznenie vzorov
- Farba textu je čitateľná, ak je vytlačená čiernobielo
- Text dostatočne kontrastuje pozadie

Čiary

- Mriežky, ak sú prítomné, sú stlmené
- Graf nemá hraničnú čiaru
- Osi nemajú zbytočné zaškrťavacie značky
- Graf má jednu horizontálnu a jednu vertikálnu os

Ostatné

- Typ grafu je vhodný pre dáta
- Graf má primeranú úroveň presnosti
- Sú prítomné kontextualizované alebo porovnávacie údaje



Infografiky

Všeobecne platí, že organizácia riadená údajmi bude mať malé využitie pre infografiku interne a určite nie pre rozhodovanie. Zaujímavé je, že nedávny výskum zistil, že "chartjunk", piktogramy, farba a kontrast (z ktorých všetky sú prominentné v infografike) uľahčujú ich zapamätanie. Aby sme však zopakovali hlavné posolstvo, cieľom je komunikácia s cieľom podnietiť činnosť. Vedenie potrebuje vysokokvalitné informácie, ktoré im umožnia nielen vidieť a zapamätať si údaje, ale aj hodnotiť na ich základe a uistiť sa, že rozhodnutie je správne.

Dashboardy

Mnohé organizácie mylne merajú svoju údajovú orientáciu podľa počtu správ a dashboardov, ktoré produkujú. Dashboardy sú veľmi užitočné a môžu podporovať celý rad aktivít, ako je napríklad poskytovanie rozhrania na vyhľadávanie údajov a upozornenia, ako aj prístup k predpovediam a prediktívnym modelom. Dashboardy patria do troch kategórií: i.) výkonné alebo strategické; ii.) analytické; iii.) operatívne;

Výkonné dashboardy poskytujú vysokoúrovňový pohľad na organizáciu a zvyčajne sa zameriavajú na scorecard (KPI a ich ciele). Informačný panel by mal poskytnúť rýchly a jednoduchý prehľad o tom, či organizácia plní svoje ciele.

Analytické zaujímajú pohľad na nižšiu úroveň, zachytávajú kľúčové trendy a ukazovatele v rámci jednej jednotky. Typicky interaktívne, čo umožňuje zúžiť pohľad na neobvyklé trendy a objavovať údaje.

Operatívne dashboardy poskytujú nižšiu latenciu, podrobnejšie pohľady na veľmi špecifické časti. Mal by existovať jasný prípad použitia a presvedčivý dôvod pre každú tabuľku alebo číslo, ktoré sa objaví na dashboarde. Menej je viac. Možno má zmysel mať viac dashboardov pokrývajúcich rovnaké údaje, ale pre rôzne časové rámce.

Storytelling

Obsah založený na údajoch nie je len o prezentovaní čísel a pochopení. Ide aj o vytváranie vzťahov s cieľovými skupinami. Každé slovo alebo vyhlásenie nemožno považovať za údaje. Môžu to byť len faktické informácie, ktoré sú výsledkom testovania alebo experimentov, ktoré tvoria záver alebo ponúkajú riešenie. Rôzni výskumníci a experimenty na tej istej téme môžu ponúknuť rôzne údaje, ale pokiaľ údaje pochádzajú zo spoľahlivých zdrojov, váš obsah založený na údajoch bude hodnotnejší v očiach cieľovej skupiny.

Efektívne príbehy založené na údajoch začínajú so silnými otázkami, ktoré sú relevantné pre vaše publikum a ak môžete kvantifikovať svoje odpovede na tieto otázky, môžete vytvoriť základ pre svoj obsah založený na údajoch. Prečo je rozprávanie založené na dátach tak dôležité? Nájdenie odpovedí musí predchádzať pochopenie bariér, ktoré bránia oslovenie publika:

- Ľudia sú zahlteniu nadbytkom obsahu a médiami.
- Marketingové správy, mobilné zariadenia a ďalšie obsahy denne útočia na pozornosť spotrebiteľa a multitasking sa stal novou kultúrnou normou.
- Ľudia nemajú dostatok pozornosti. Mozog spotrebiteľov nie je nafukovací a môže v jednom okamihu vstrebať len určité množstvo informácií tak, aby ešte dávali zmysel. Konzumácia obsahu vyžaduje tunelové videnie. Spotrebiteľia sa zameriavajú len na obsah, ktorý je pre nich ešte relevantné v danom okamihu a všetko ostatné odfiltrujú.
- Cesta spotrebiteľa je nepredvídateľná. Spotrebiteľia obsah spracúvajú bez špecifickej rutiny. Ich rozhodovací proces pri nákupe je nepredvídateľný, čo značne znižuje schopnosť značiek je osloviť.



Ak zistíte, aké sú najdôležitejšie otázky cieľovej skupiny, budete musieť posúdiť, či sú k dispozícii údaje, ktoré vám pomôžu odpovedať na tieto otázky. Na zhromažďovanie, odhadovanie a filtrovanie informácií budete potrebovať postup. Hľadanie vhodných údajov na podporu vášho príbehu je kľúčové, pretože nechcete vytvárať legitímny obsah, ktorý nemôže presvedčiť vaše publikum. Hlavnými komponentmi dobrého zdroja údajov sú:

- Originalita - Vaše údaje by mali vždy pochádzať z primárneho zdroja. Ak narazíte na zaujímavé pozorovanie v článku alebo prípadovej štúdií, nájdite jeho pôvodný zdroj. Ak údaje neskúmate, nikdy neviete, či je to pravda alebo nie.
- Dôkladnosť – Váš zdroj údajov by mal byť dostatočne informatívny, aby zodpovedal otázky vašich zákazníkov.
- Včasnosť – Údaje sa rýchlo menia. Informácie sú zastarané v mesiacoch. To je dôvod, prečo by ste mali pre svoju stratégiu obsahu používať iba najnovšie zdroje.
- Spoľahlivosť – Váš zdroj údajov by mal byť čo najrelevantnejší. Skôr ako si vyberiete zdroj, opýtajte sa sami seba, kto to napísal, kto to povedal alebo kedy bola naposledy aktualizovaná, ako aj účel zverejnených informácií. (Akademické časopisy, univerzitné stránky, štúdie a výskumné správy sú vo všeobecnosti výbornými zdrojmi údajov, zatiaľ čo väčšina blogov na druhej strane nie sú.)

Napriek skutočnosti, že mnohé zo špecifik, ktoré sa týkajú rozprávania príbehov údajov, závisia od konkrétneho prípadu použitia, zámeru a problému, ktorý sa rieši, existujú niektoré všeobecné zásady, ktoré predstavujú najlepšie postupy a ktoré sú podrobne opísané v nasledujúcom poradí:

- Nastavenie: Počiatočné okamihy prezentácie údajov by mali začať zaujímavým otváračom, ktorý podporí zvedavosť. Najlepším príkladom je ilustrovanie problému alebo jeho účinkov na organizáciu.
- Kontext: Je potrebné ilustrovať súčasné skutočnosti, ktoré sa týkajú údajov alebo problému, ktorý vyžaduje, aby vyšetrovatelia vybrali vizualizácie, určili zámer akcie a identifikovali nástroje a iné logistické informácie.
- Možnosti: Tento krok predstavuje bod v príbehu, v ktorom moderátor presvedča poslucháčov o užitočnosti svojho návrhu založeného na analytike. Priamo zahŕňa dôsledky údajov a je vizuálne reprezentovaný. Podmienené obavy zahŕňajú údaje, ktoré sú k dispozícii na použitie, a všetky procesy riadenia, ktorých sa môžu týkať.
- Výzva na akciu: výzva na akciu zdôrazňuje, čo presne organizácia môže urobiť, aby poukázala na príčinu problému alebo poskytla jej riešenie.

Dodatočné zdroje:

[“Interactive Data Visualization for the Web”](#), autor: Scott Murray

[“Data Visualization: A New Language for Storytelling”](#), autor: Mike Barlow

[“The Graphic Continuum”](#) autor: Jon Schwabish

[“Gallery Of D3 Visualizations”](#), publikované v GitHub

[“Introducing the Data Visualization Checklist”](#), autor: Stephanie Evergreen

[“The Rise of the Y-Axis-Zero Fundamentalists”](#), autor: Justin Fox

[“A Better Definition of Chart Junk”](#), autor: Robert Kosara

[“Data Driven”](#), autor: Hilary Mason, DJ Patil



9 Lepšie rozhodovanie na základe dát

Údaje nemajú mozog, a preto nemôžu nič riadiť. Rozhodovanie založené na údajoch je len čiastočne o údajoch. Jasná a spoločná vízia a vedenie zohrávajú hlavnú úlohu pri rozhodovaní. Ako už bolo uvedené, údaje musia byť včasné, relevantné a dôveryhodné. Ak tomu tak nie je, rozhodovacie orgány majú obmedzené možnosti. Ďalšou významnou prekážkou efektívneho rozhodovania založeného na údajoch je fakt, že sme ďaleko od dokonalých tvorcov rozhodnutí. Nie vždy pristupujeme k problémom najobjektívnejším spôsobom, často prinášame vlastné subjektívne skúsenosti a zameriavame sa na nepodstatné detaily. To vedie k chybným, nelogickým mysleniam a rozhodnutiam. Tieto vplyvy a mechanizmy sa nazývajú kognitívne predsudky. prečo si nemôžeme vždy dôverovať?

- Sme rozporuplní – Keď máme k dispozícii rovnaké dôkazy v rôznych časoch, dospievame k iným záverom; a keď sú rôznym ľuďom predložené rovnaké dôkazy, môžeme dospieť k iným záverom
- Pamätáme si na veci, ktoré sa nestali – Naša intuícia je založená na častiach skutočných faktov, obklopených dierami, ktoré máme sklón vyplniť predpokladmi a presvedčeniami.
- Nevzdáme sa zlých údajov – Internalizujeme fakty, budujeme mentálne modely a potom, keď sú nové dôkazy v rozpore s týmito faktami, sme veľmi odolní voči zmene nášho modelu alebo prijatiu nových údajov.
- Sme unavení a hladní – Naše rozhodnutia ovplyvňujú iné vnútorné, vonkajšie faktory, ako je hlad, nálada a energia. V analýze boli sledované rozhodnutia ôsmich sudcov. Najjednoduchším rozhodnutím bolo odmietnuť prepustenie, ktoré dosiahlo až 65% po obede a ku koncu dňa sa postupne znižovalo.

Riešenie

Prvou podmienkou je **motivácia**. Ľudia sú motivovaní rôznymi vecami: peniazmi, možnosťami zlepšenia kariéry, voľno, povýšenie. Dôležitou súčasťou je prepojenie odmeny s motiváciou zamestnanca. Mesto Flagstaff v Arizone má systém, ktorý odmeňuje zamestnancov nasledovne:

- Unikátna menovka
- Vyryté ocenenie
- Finančné ocenenie
- Platená dovolenka
- Povýšenie
- Jednorazový bonus až do výšky 500 USD

Spolupráca a konsenzus. Ak je rozhodnutie zložité alebo pravdepodobne nepopulárne, jedným z prístupov je dosiahnuť konsenzus. Dôležité je, aby sa všetci cítili súčasťou procesu. Rozhodnutie nie je správne, ak ho nikto nepodporuje. Vo svete založenom na údajoch to znamená zabezpečiť, aby každý rozumel cieľu, zozbieraným údajom, metrikám a spôsobu, akým primárny rozhodovací orgán interpretuje dôkazy. Dajte ostatným možnosť predložiť svoje interpretácie a názory:

- Definujte problém.
- Stanovte kritériá.



- zväžte všetky alternatívy.
- Identifikujte najlepšiu alternatívu.
- Vypracujte a realizujte akčný plán.
- Vyhodnoťte a monitorujte riešenie a spätnú väzbu v prípade potreby.

Analytické technológie podporujúce rozhodnutia založené na údajoch

Mnohé technológie môžu podporovať rozhodnutia založené na údajoch v celom rade prípadov použitia. Keď organizácie uvažujú o analytických technológiách, ktoré pomáhajú pri rozhodovaní založenom na údajoch, zvyčajne uvažujú o nástrojoch, ako napríklad:

Riadenie výkonnosti. Metódy a nástroje riadenia výkonnosti pomáhajú používateľom zamerať sa na prístup k údajom, analýzu a v konečnom dôsledku na výkonnosť dôležitých cieľov a cieľov projektu. Ciele sa zvyčajne pohybujú od širších cieľov, ako je zlepšenie spokojnosti zákazníkov až po úzko definované ciele zamerané konkrétne na zodpovednosti zamestnanca, ako napríklad dokončenie procesu v stanovenej lehote. Riadenie výkonnosti je preto dlhodobou hlavným zameraním BI a dátových skladov, aby umožnilo podnikovým používateľom pracovať s metrikami založenými na údajoch efektívne na usmerňovaní rozhodnutí a činností. Namiesto toho, aby používatelia poskytovali objemné prehľady, ktoré vyžadujú, aby vyhľadávali relevantné údaje, môžu metriky výkonnosti uľahčiť používateľom rýchlejšie využívanie údajov v rámci kontextu a rýchlejší prístup k najdôležitejším údajom.

Na presné a konzistentné definovanie KPI a iných metrík musia manažéri preskúmať BI reporty a analýzy. Dátové sklady často hrajú dôležitú úlohu pri poskytovaní prístupu k starostlivo vyčisteným a štruktúrovaným údajom na podporu metrík riadenia výkonnosti. Vďaka dobre definovaným metrikám, ktoré sú aktuálne, môže riadenie výkonnosti pomôcť organizáciám zosúladiť rozhodnutia a akcie s cieľmi. Ciele poskytujú hodnotný kontext pre analýzu údajov. Rozhodovatelia tak potrebujú technológiu, ktorá im umožní interakciu s údajmi relevantnými pre metriky výkonnosti. Ak užívatelia nie sú schopní preskúmať údaje, nemôžu efektívne využívať údaje na zlepšenie výkonu. To môže spôsobiť, že metriky budú menej cenné pre riadenie rozhodnutí a akcií.

Dashboards a scorecards. Dashboards a scorecards sú obvyčajne spôsob, akým organizácie komunikujú ciele riadenia výkonnosti. Väčšina BI riešení podporuje ovládacie panely na mobilných zariadeniach, natívne alebo prostredníctvom webu, ako aj na desktopoch, laptopoch a pracovných staniciach. Dashboards sa vyvinuli tak, aby slúžili širšiemu účelu ako používateľský portál (alebo „informačný panel“) nielen pre metriky výkonnosti, ale aj pre širší rozsah vizualizácií, vrátane grafov, tepelných máp a meradiel a textových informačných kanálov.

Samoobslužná vizuálna analýza a zisťovanie údajov. Riešenia v tejto oblasti sa zameriavajú na riešenie potrieb používateľov, ktoré presahujú spotrebu údajov. Používatelia chcú analyzovať údaje, ale nechcú opustiť jednoduchšie používateľské rozhranie a grafické skúsenosti, ktoré sú charakteristické pre riadiace panely. Riešenia vizuálnej analýzy a vyhľadávania údajov poskytnú väčšiu samoobslužnú funkcionálnosť ako staršie podnikové BI riešenia. S týmito riešeniami môžu používatelia sledovať údaje prostredníctvom moderných grafických rozhraní a vizualizácií, vrátane dashboardov. Schopnosti samoobsluhy im umožňujú vybrať si súbory údajov, dotazovať sa na dáta a vytvárať vlastné vizualizácie. Väčšina riešení prichádza s knižnicami typov vizuálnej reprezentácie, ako sú grafy, tepelné mapy a rozptýlené projekty; Mnohé riešenia umožňujú používateľom rozšíriť svoje možnosti importom vizualizácií z vonkajších zdrojov, vrátane open source.

Príprava a integrácia údajov. Pokroky v príprave údajov sú rozhodujúce pre to, aby používatelia mohli využívať vizuálne analytické riešenia, pretože sa snažia o prístup k väčšiemu množstvu zdrojov údajov. Príprava dát zahŕňa celý rad procesov, ktoré začínajú s prijímaním a zhromažďovaním údajov a prebiehajú prostredníctvom zlepšovania kvality a transformácie. Tieto procesy sú často pomalé a komplikované a vyžadujú značné manuálne úsilie, ktoré môže zabrániť tomu, aby BI a analytika



zohrávali integrálnejšiu úlohu v každodennom rozhodovaní. Užívatelia zvyčajne musia počkať, kým IT pripraví údaje, alebo prevziať úlohu s neštandardnými nástrojmi a menej konzistentnými metódami, ktoré môžu spôsobiť chyby a nezrovnalosti. Samoobslužná funkčnosť pomocou termínov, ako je napríklad miešanie údajov (blednign), wrangling a munging, umožňuje používateľom preskúmať údaje a vybrať súbory údajov, ktoré vyhovujú ich procesom BI a vizuálnej analýzy. Samoobslužná funkcionálnosť je tiež zrelá na použitie pri vývoji dátových katalógov, glosárov a repozitárov metadát. Toto je rozhodujúce pre to, aby používatelia mohli získať kompletný prehľad o údajoch a zdieľať údaje a poznatky založené na údajoch s ostatnými.

Pokročilé analýzy. Pokročilé analytické nástroje zahŕňajú nástroje ako prediktívne modelovanie a strojové učenie. Tieto nástroje môžu pomôcť používateľom nájsť vzory v údajoch, ktoré môžu riadiť rozhodnutia. Tieto nástroje majú miesto vo vznikajúcej oblasti automatizácie, ktorá môže tiež pomôcť organizácii stať sa viac orientovanou na údaje.

Automatizácia. Existuje celý rad nástrojov, ktoré automatizujú životný cyklus analýzy, od prípravy údajov až po budovanie modelov, ktoré využívajú pokročilé analytické metódy, ako napríklad strojové učenie alebo spracovanie prirodzeného jazyka. Tieto produkty môžu pomôcť organizáciám pri rýchlejšom rozhodovaní. Organizácie môžu používať prediktívne analýzy na pomoc pri upozorňovaní, napríklad podvody – systémy môžu smerovať potenciálne podvodné transakcie na špeciálnu vyšetrovaciu jednotku na ďalšie spracovanie.

Dodatočné zdroje:

„[Gaining trust as well as respect in communicating to motivated audiences about science topics](#)“, autor: Susan T. Fiske

„[What to Do When People Draw Different Conclusions From the Same Data](#)“, autor: Walter Frick

„[The Future of Decision Making: Less Intuition, More Evidence](#)“, autor: Andrew McAfee

„[The Dangers of Faith In Data](#)“, autor: Scott Berkun

„[Analytics: A Blueprint For Value](#)“, publikované v IBM



10 Kultúra

Organizácie založené na údajoch poskytujú široký prístup k údajom. Musí existovať jasný signál, že údaje nie sú „vlastnené“ jednotlivými tímami, ale že patria k organizácii ako celku. Podpora postupov zdieľania údajov by mala mať vysokú prioritu. Každý v organizácii by mal mať prístup k čo najväčšiemu množstvu údajov, ktoré sú právne možné. Prístup môže byť cez reporty a dashboardy, ale aj „aktívny prístup“ alebo dokonca nespracované dáta. To tiež zahŕňa obrovský prvok dôvery. Organizácia musí dôverovať, že údaje nebudú zneužitú, ale budú vhodne použité na podporu celej organizácie. Ak má prístup k údajom, ktoré potrebujú, viac zamestnancov, je možné demokratizovať viac rozhodovania.

Ako vyzerá kultúra organizácie informovaná údajmi?

- Lídri, manažéri a zamestnanci podporujú bezpečnú demokratizáciu údajov a zvyšujú dostupnosť údajov pre tých, ktorí ich potrebujú
- Organizácia má jasné strategické priority, ktorých úspech je kvantifikovateľný a každý chápe, ako ich práca prispieva k týmto prioritám
- Vedúci a vedúci tím rutinne kladie otázky, ktorých odpovede vyžadujú hlbšiu analýzu základnej dynamiky každého problému
- Zamestnanci trávajú viac času analýzou ako vytváraním správ alebo zberom údajov.

Široká dátová gramotnosť

Je zrejmé, že analytici potrebujú školenie v oblasti experimentálneho dizajnu, kritického myslenia, prezentácie údajov, používania nástrojov, štatistiky atď. Potrebujete však, aby manažéri a iní tvorcovia rozhodnutí boli tiež gramotní. Prečo?

- Manažéri nakupujú, inštalujú a udržiavajú nové nástroje alebo službu prediktívneho modelovania. Musia pochopiť hodnotu, ktorú to prinesie organizácii.
- Manažéri robia konečné strategické a taktické rozhodnutia na základe analýz. Musia rozpoznať chyby. Mali by sa neustále pýtať hlbšie otázky týkajúce sa údajov a očakávať od analytikov viac.
- Manažéri musia tiež prekládať svoje zistenia a závery. To znamená, že musia pochopiť nuansy analýzy, mať v ňu dôveru a byť pripravení ju brániť.

Stručne povedané, manažéri nemusia nevyhnutne chápať mechaniku zhromažďovania, čistenia, spracovania a zhromažďovania údajov, ale mali by mať pochopenie dobrého experimentálneho dizajnu, základného štatistického odhadu a nebezpečenstva extrapolácie.

Atmosféra zdravej diskusie

Organizácia založená na údajoch by mala vytvoriť atmosféru zdravej diskusie, kde je možné požiadať o dodatočné informácie, spochybniť predpoklady a prediskutovať odporúčania alebo dodatočné testy. Ak je pre vás vysokou prioritou vytvorenie prostredia riadeného údajmi, mala by byť prijateľná určitá úroveň nesúhlasu, ktorá by v niektorých prípadoch mala byť dokonca odmenená. Nové nápady, ktoré sú overené na základe údajov, sú obrovským východiskovým bodom pre pozitívne inovácie.



Vedenie

Organizácie založené na údajoch vyžadujú silné vedenie. Potrebujú vedenie, ktoré inšpiruje, podporuje kultúru založenú na údajoch a aktívne riadi a podporuje všetky aspekty hodnotového reťazca analytiky. Vedúci musí podporovať analytickú organizáciu ako takú. Tento vedúci musí poskytnúť údaje, nástroje a školenia. Vedúci by mal definovať organizačnú štruktúru a mal by tiež poskytnúť jasnú kariérnu cestu a stimuly pre analytikov, aby boli šťastní, produktívni a odvedli skvelú prácu. Vedúci musí mať podporu od zvyšku vyššieho manažmentu.

Ako prilákať a najat' nový talent?

- Napíšte lepšie popisy úloh. Pochopte, čo skutočne potrebujete pre pozíciu, ktorú sa snažíte prenajať, a podľa toho napíšte popis.
- Zlepšite transparentnosť a citlivosť vašich postupov prijímania zamestnancov. Buďte otvorení od začiatku, ako vyzerá váš proces, aj keď to nie je ideálne.
- Ako motivátori hľadajte alternatívne prínosy alebo výhody. Čo ešte môžete ponúknuť? Premýšľajte o výhodách, ako je vynikajúce pracovisko, flexibilné plány, účasť na konferenciách a možnosti vzdelávania.
- Nečakajte, že budete zamestnávať ľudí navždy. Zatiaľ čo niektorí ľudia môžu zostať niekoľko rokov alebo desať rokov, čo je čoraz zriedkavejšie, mali by ste akceptovať, že ľudia odídu po kratšom čase.
- Vytvorte analytickú komunitu v rámci organizácie a ponúknite ju komukoľvek, kto má vášeň pre používanie údajov na pomoc pri riešení problémov.
- Vytvorte si profesijné dráhy a programy mentorstva pre svojich profesionálnych analytikov, najmä vysoko technicky skúsených.

Dodatočné zdroje:

„[Empowering Analysis Ninjas? 12 Signs To Identify A Data Driven Culture](#)“, autor: Avinash Kaushik

„[Understanding the Chief Data Officer](#)“, autor: Julie Steele

„[Status of the Chief Data Officer: An Update on the CDO Role in Organizations Today](#)“, publikované v DATAVERSITY

„[A Data Culture For Everyone](#)“, autor: Satya Nadella

Záverečné odporúčania

Aplikujte analytiku na riadenie výkonnosti. Hoci mnohé organizácie využívajú BI a dátové sklady na podporu riadenia výkonu. Organizácie by mali na zlepšenie spôsobu, akým odvodzujú, definujú a uplatňujú metriky, využívať analytické nástroje. Mohli by tiež použiť analýzy na preskúmanie údajov súvisiacich s metrikami na predpovedanie budúcich výsledkov a byť proaktívni pri vykonávaní úprav s cieľom dosiahnuť lepšie výsledky.

Dajte veľké dátové platformy a dátové sklady do jednej stratégie. Významné percento organizácií v súčasnosti je, alebo plánuje vyvinúť BI a analytiku na technológiách Apache Hadoop a Spark. Aby sa predišlo oneskoreniu a odpojeniu, organizácie založené na údajoch musia vytvoriť komplexnú multiplatformovú architektúru údajov, ktorá poskytuje stratégiu na integráciu všetkých platforiem.

Pochopte, kedy automatizovať. Niektoré analýzy pre rozhodovanie nemusia nevyhnutne vyžadovať, aby boli ľudia zapojení, len čo sa vytvorí počiatočná budova modelu, napríklad aplikácia na



odhaľovanie podvodov. Organizácie môžu používať aplikácie na rozhodovanie v prípadoch, keď potrebujú urobiť malé, opakovateľné rozhodnutia.

Monitorujte modely. Využívajte pokročilejšie analýzy na vykonávanie rozhodnutí založených na údajoch. Je však mimoriadne dôležité zabezpečiť, aby vás staré modely nespomaľovali.

Premýšľajte o ľuďoch. Vaša organizácia a ľudia v nej si budú musieť osvojiť data-driven mentalitu. To znamená premýšľať o ich motivácii a správne ich motivovať. Organizácie tiež potrebujú pomáhať pracovníkom rozvíjať zručnosti, ako napríklad používať nástroje, zapájať sa do kritického myslenia a využívať výstupy z analýz. To si vyžaduje školenie. Nezabúdajte, že aj keď sa nástroj ľahko používa, stále to znamená, že budete musieť nástroj používať správne.

Modernizujete prípravu dát pre rýchlosť a agilitu. Procesy prípravy dát sú často pomalé, únavné a spotrebúvajú príliš veľa času. Aby organizácie mohli efektívne využívať údaje na riadenie rozhodnutí, procesy prípravy údajov musia využívať trendy smerom k inteligentnejším a automatizovanejším technológiám, ktoré využívajú strojové učenie. Organizácie by mali vyhodnotiť nové technológie na prípravu údajov vrátane prípravy samoobslužných údajov na podporu samoobslužných BI, vizuálnej analýzy a pokročilých analýz s veľkými údajmi.

Modernizujete dátovú platformu. Multiplatformová dátová architektúra (MDA) má zmysel pre pokročilejšie analýzy na podporu rozhodnutí založených na údajoch. Hoci dátové sklady existujú, má zmysel začať uvažovať o tom, čo príde z hľadiska rozšírenia prostredia na správu údajov na podporu pokročilých analytických riešení. MDA by mohli zahŕňať dátový sklad, Hadoop a iné platformy aj vo verejnom cloudu.

