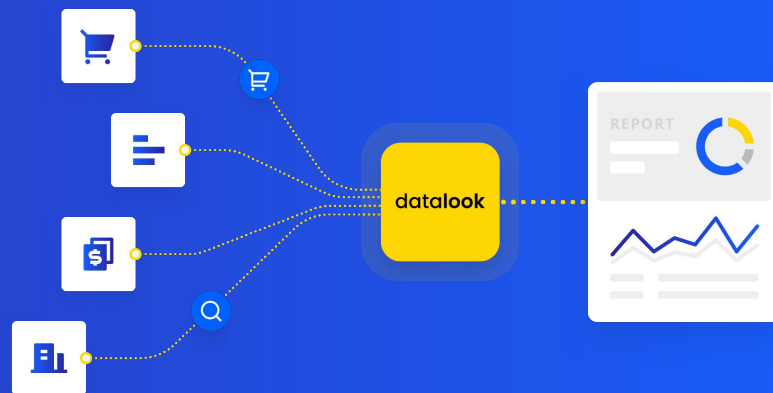


datalook

Datalook

Kombinatorika v segmentaci zákazníků



Prezentuje



Tomáš Zísler

CEO

Tomáš je CEO a co-founder Datalook. Od roku 2004 se věnuje vývoji a poradenství online projektů, e-commerce a systémům řízení. Chce pomoci podnikatelům neřídít firmy naslepo.

datalook

**Rozšířený pohled na analýzu
zákaznických dat a proč jich chtít
co nejvíce.**

Co je segmentace

- Definice segmentace zákazníků
- Důležitost segmentace v marketingu
- Hlavní cíle segmentace

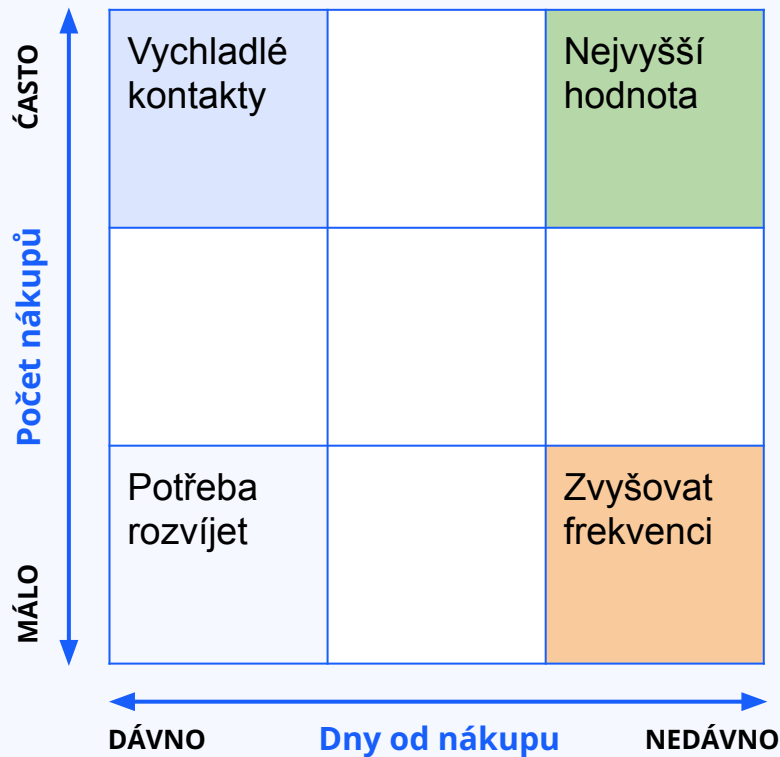


RFM

Recency, Frequency, Monetary Value

- R: Počet dní od posledního nákupu
- F: Počet nákupů
- M: Celková hodnota konverzí

Vzniká nám 3D matice, která rozprostře zákazníky do jednotlivých skupin.

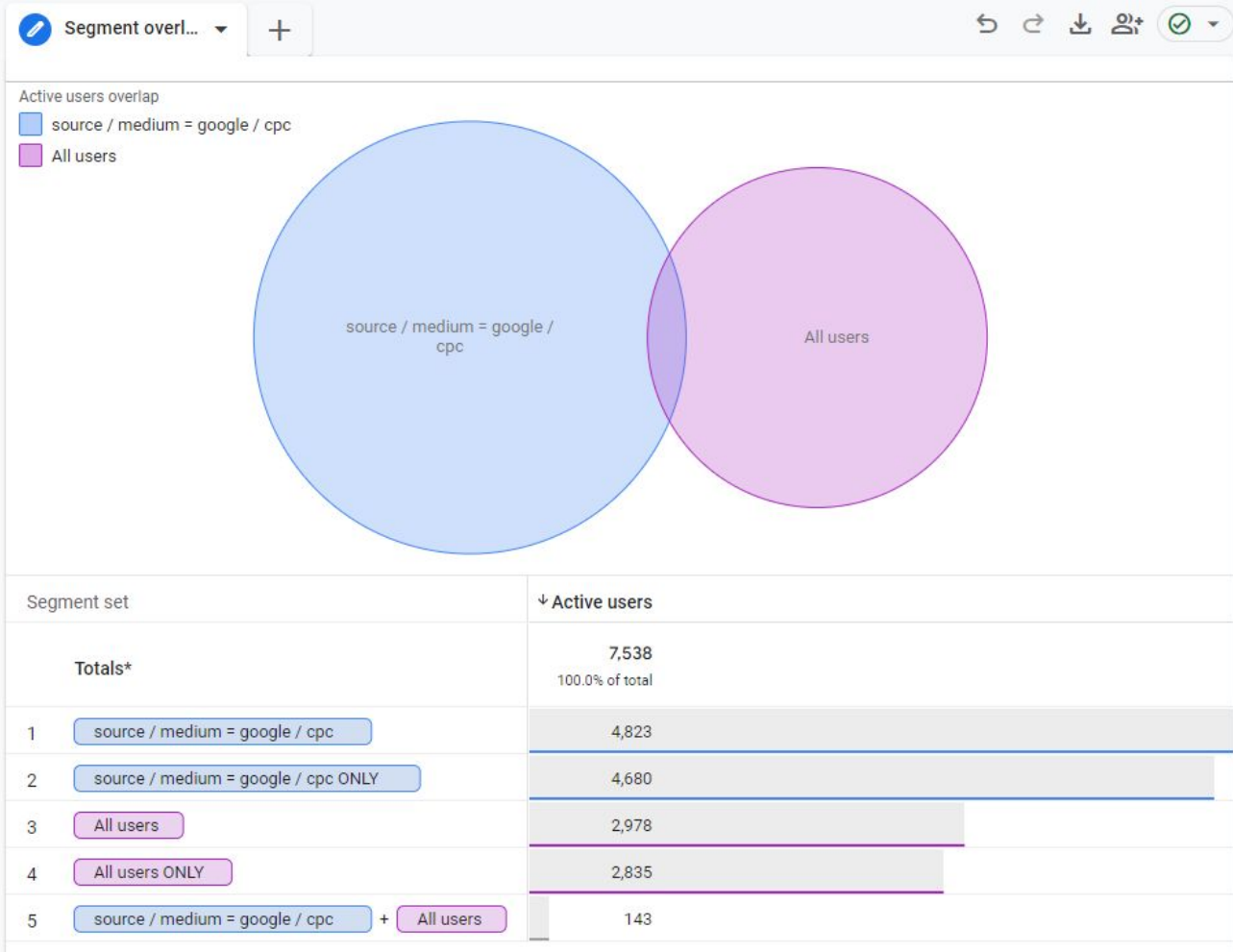


Příklad jednoduché RFM

type	segment	max	min	customers
frequency	1	28	1	3274
frequency	2	1	1	3274
frequency	3	1	1	3274
frequency	4	1	1	3273
monetary	1	5500	3000	3274
monetary	2	82.4	34.896	3274
monetary	3	34.896	18.144	3274
monetary	4	18.128	0.56	3273
recency	1	2021-07-28 14:23:58	2021-03-19 19:2	3274
recency	2	2021-03-19 19:14:03	2021-02-10 05:2	3274
recency	3	2021-02-10 03:48:13	2020-08-09 04:3	3274
recency	4	2020-08-08 21:59:40	2018-10-14 19:5	3273

R	F	M	Segment Name
1	1	1	Champions
2	1	1	Can't lose them
3	1	1	Can't lose them
4	1	1	Lost
5	1	1	Lost
1	2	1	Loyal
2	2	1	Can't lose them
3	2	1	Can't lose them
4	2	1	Lost
5	2	1	Lost
1	3	1	New customers
2	3	1	About to sleep
3	3	1	About to sleep
4	3	1	Lost
5	3	1	Lost
1	4	1	New customers
2	4	1	About to lose
3	4	1	About to lose
4	4	1	Lost
5	4	1	Lost
1	5	1	New customers
2	5	1	About to lose
3	5	1	About to lose
4	5	1	Lost
5	5	1	Lost
1	1	2	Loyal
2	1	2	Need attention
3	1	2	Need attention
4	1	2	Lost
5	1	2	Lost
1	2	2	Loyal
2	2	2	About to sleep
3	2	2	About to sleep

Segmentace v Google Analytics



Kde můžete segmentovat



Excel



Google Analytics



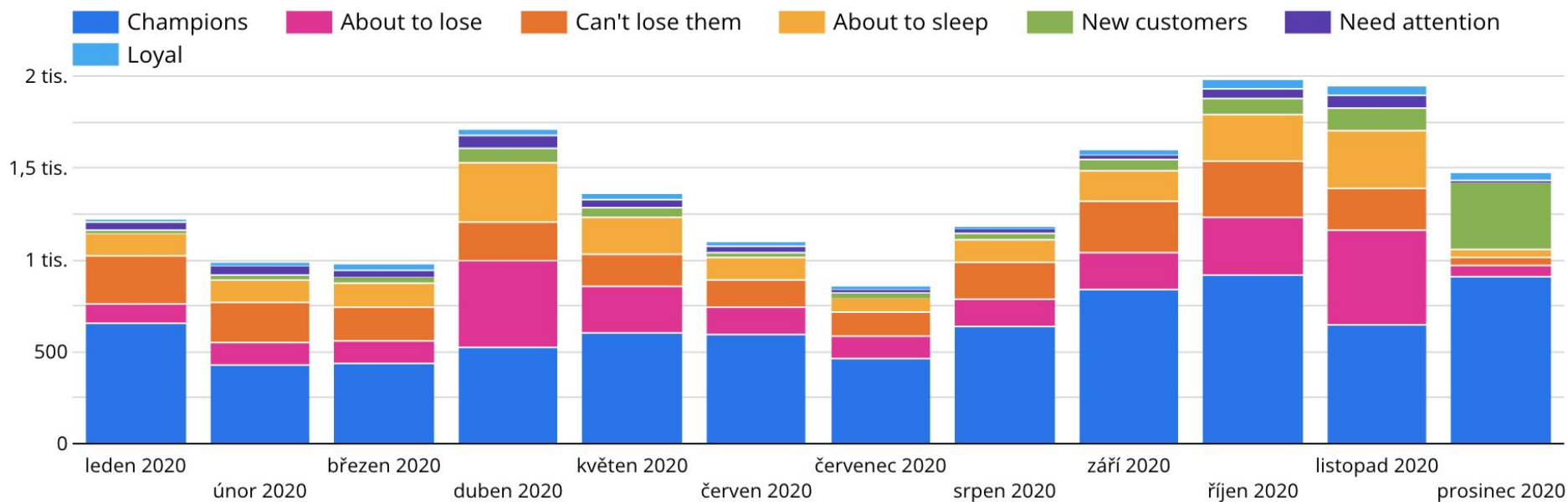
Targito

datalook!

Datalook

Segment	# zákazníků	% zákazníků	Obrat	% obratu	# objednávek	% objednávek	ø hodnota obj.
Champions	1 664	20,86 %	9 681 207 Kč	52,64 %	7 647	46,55 %	1 266 Kč
About to lose	2 591	32,48 %	2 320 238 Kč	12,62 %	2 591	15,77 %	895 Kč
Can't lose them	1 126	14,11 %	3 096 252 Kč	16,84 %	2 385	14,52 %	1 298 Kč
About to sleep	1 355	16,98 %	1 943 554 Kč	10,57 %	2 006	12,21 %	969 Kč
New customers	776	9,73 %	923 949 Kč	5,02 %	913	5,56 %	1 012 Kč

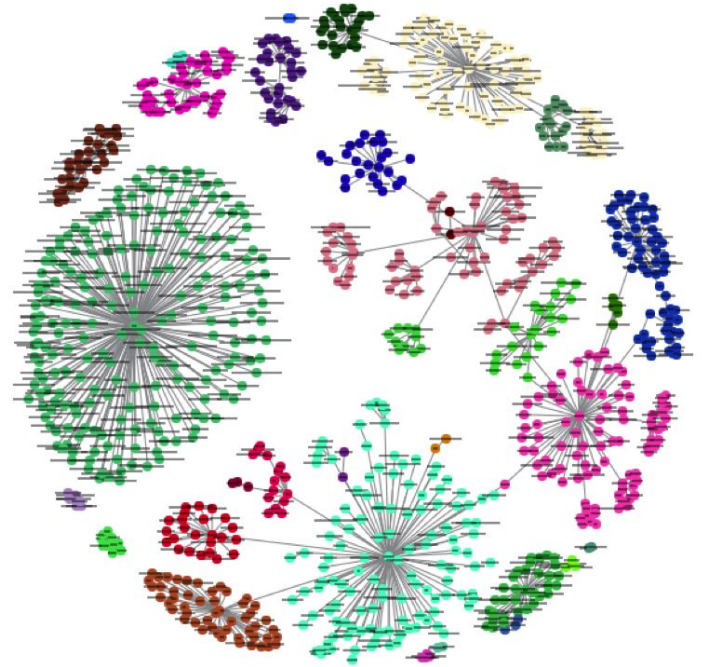
Objednávky dle segmentů



Kdy RFM nestačí

- Omezený pohled jen na 3 metriky (R,F,M)
- Nepředjímá jemnější preference zákazníků
- Nezohledňuje kontext v případě sezónnosti
- Nemá schopnost predikce

CLUSTERIZACE Dat



datalook

Jak funguje CLUSTERIZACE?

Matematika

Výpočetní modely:

- K-means cluster analysis
- Euklidovská vzdálenost
- Ale existuje dalších 15 variant

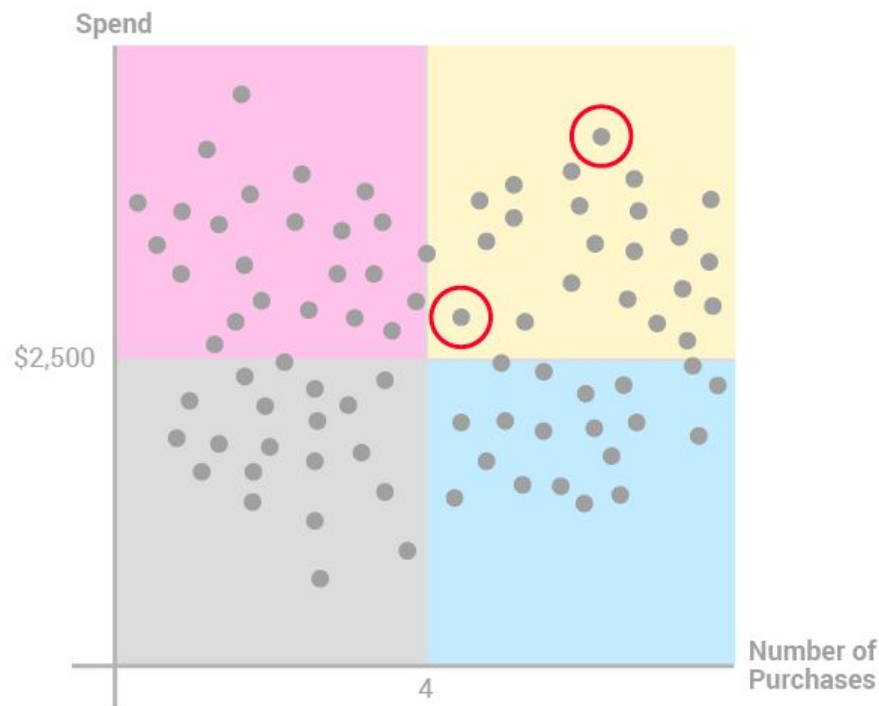


datalook

Cluster analýza je technika pro rozdělení datových bodů do skupin na základě podobnosti těchto bodů.

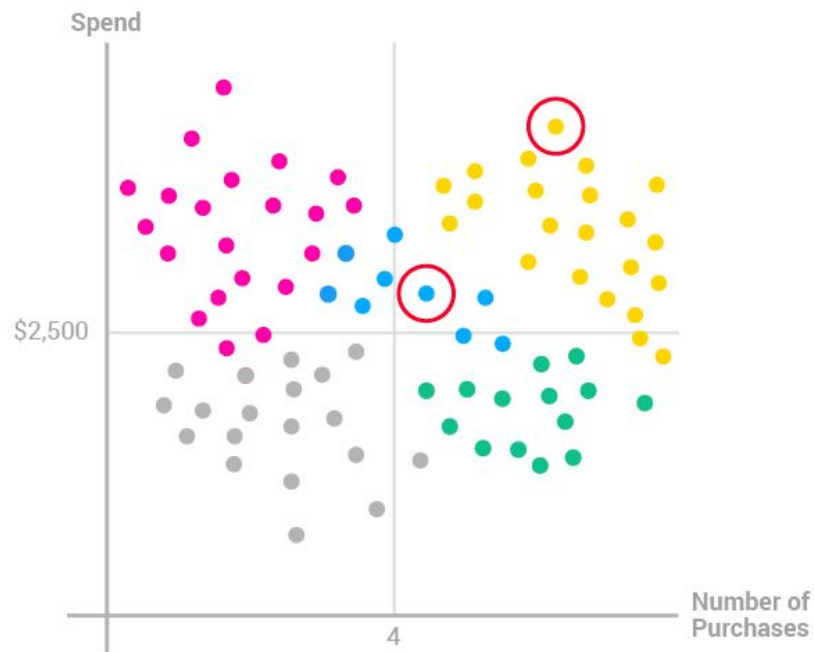
Klasická segmentace (RFM)

- Přednastavené data násilně shlukují
- Značné rozdíly chování v každém segmentu
- Jste omezení často pouze na 2-3 dimenze



Clusterizace

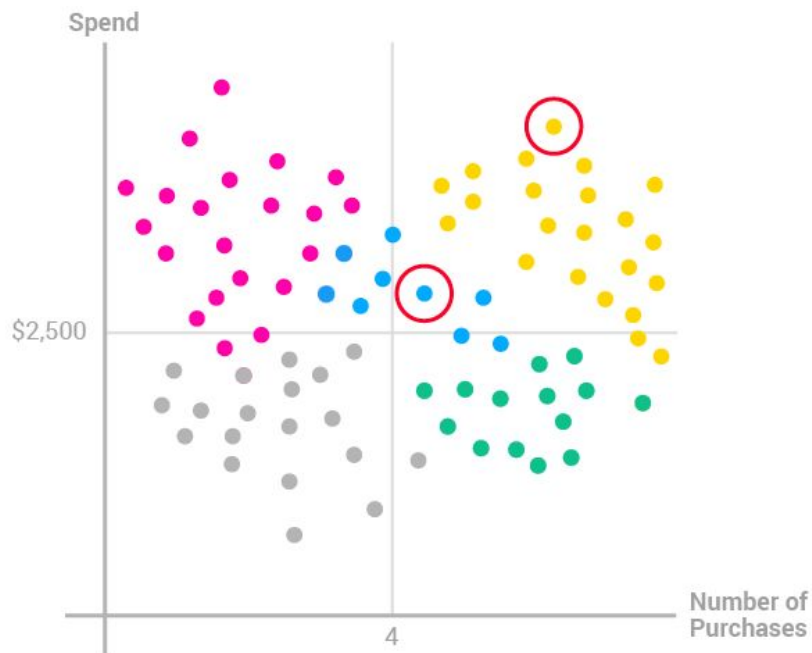
- Přirozené skupiny
- Jsem “neomezen” počtem dimenzí
- Dynamičnost skupin dle aktuálních dat



Segmentace



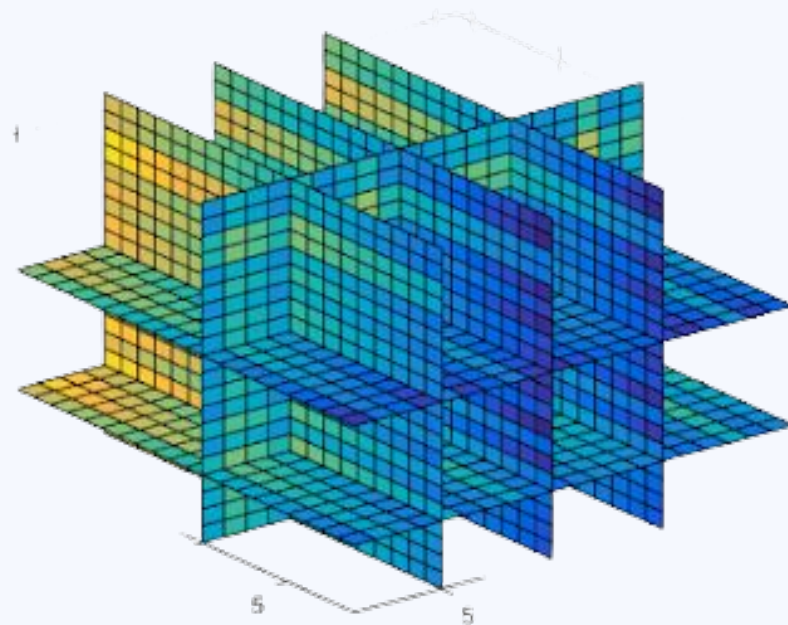
Clusterizace



Problémy s Vizualizací MKS

Jak propojit 12 dimenzí?

- 1 terabyte dat
- 7h výpočetního času
- Bez cloudu se nepohnete



Implementace (SQL, Python a Rko)

Naimplementované K-means

- Bigquery
- Snowflake

Custom a vlastní modely:

- Keboola

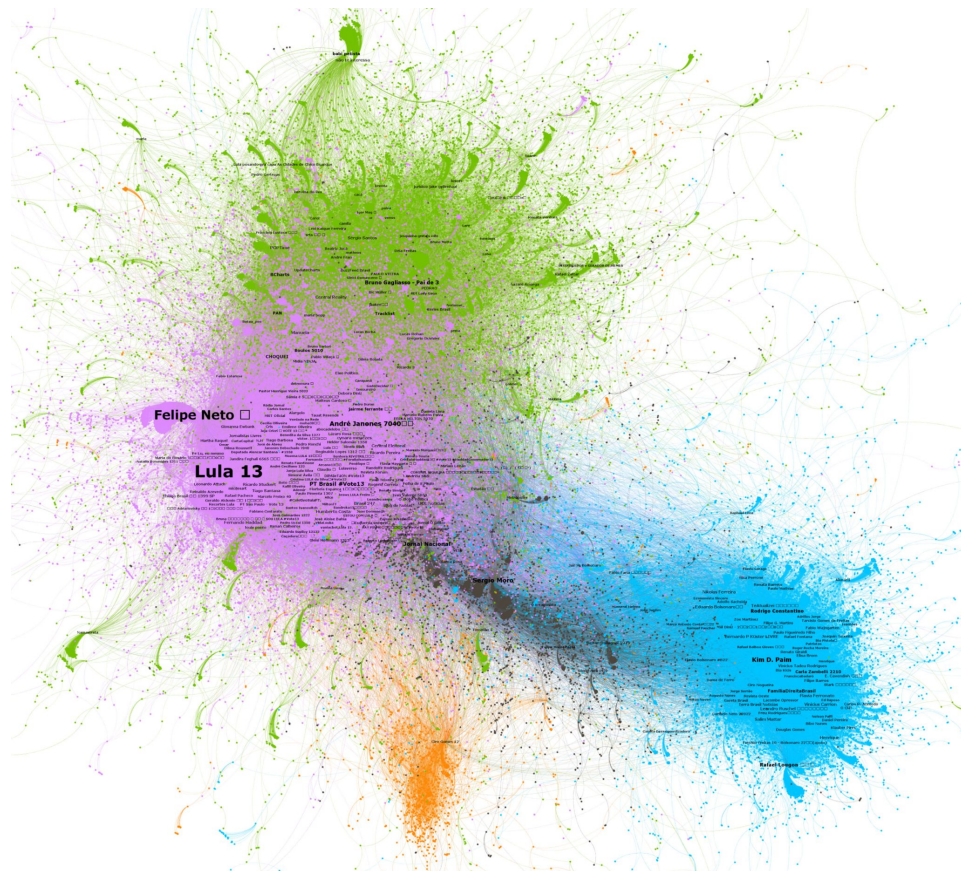
```
CREATE OR REPLACE MODEL demos_eu.london_station_clusters
OPTIONS(model_type='kmeans', num_clusters=4, standardize_features =
true) AS

WITH hs AS (
SELECT
h.start_station_name as station_name,
IF(EXTRACT(DAYOFWEEK FROM h.start_date) = 1 OR
EXTRACT(DAYOFWEEK FROM h.start_date) = 7, "weekend", "weekday") as
isweekday,
h.duration,
s.bikes_count,
ST_DISTANCE(ST_GEOGPOINT(s.longitude, s.latitude),
ST_GEOGPOINT(-0.1, 51.5))/1000 as
distance_from_city_center
FROM `bigquery-public-data.london_bicycles.cycle_hire` as h
JOIN `bigquery-public-data.london_bicycles.cycle_stations` as s
ON h.start_station_id = s.id
WHERE h.start_date BETWEEN CAST('2015-01-01 00:00:00' AS TIMESTAMP)
AND CAST('2016-01-01 00:00:00' AS TIMESTAMP)
),
stationstats AS (
SELECT
```

Explorativní postup

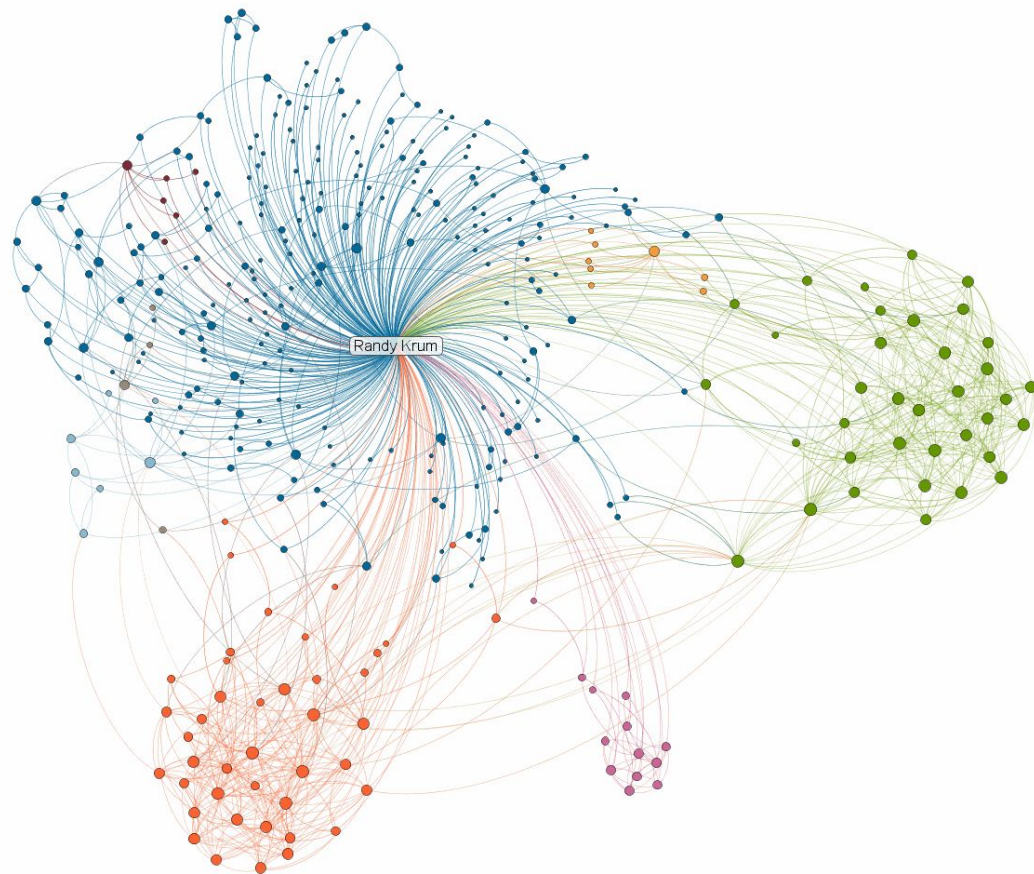
Nevýhoda: nemůžeme predikovat výsledek

Výhoda: dostáváme nové insighty do svých dat



Obecné využití

- Zákazníci
- Produkty
- Finance
- Obchod
- Sociologie
- Zdravotnictví



5 důvodů proč clusterizaci používat

- Identifikace skrytých vzorců
- Segmentace cílového trhu
- Personalizace marketingu
- Odhalování nových příležitostí
- Lepší pochopení zákazníků

datalook

Praktický postup

Co potřebuji definovat?

Rozhodující metriky pro dataset

- U všech vstupních dat platí čím více tím lépe
- Očištěná a kvalitní data
- Co nejvíce informací

Fakt: 70% firem má nespolehlivá (špinavá) data.

Co potřebuji definovat?

Customer ID	Name	Email	Gender	Age	Income	Postcode	Marital Status
3453450	James Smith	James.Smith@	M	32	55000	3013	M
3453451	John Davison	John.Davison@	M	29	62000	2122	S
3453452	Sarah Daniels	Sarah.Daniels@	F	22	43000	2522	M
3453453	Sally Jones	Sally.Jones@g	F	29	98000	2000	D
3453454	David Waters	David.Waters@	M	30	72000	3000	M
3453455	Isobel Archer	IsobelArcher@	F	35	53000	3037	S
3453456	Dianne Wilson	Dianne.Wilson@	F	42	67000	3040	S
3453457	Dennis Potter	MrPotterD@d	M	64	20000	3067	D
3453458	Jane Norton	JaneNort1987@	F	28	95000	3055	M
3453459	Janet Wilson	Janet.Wilson@	F	30	110000	3000	M

Co k tomu budu potřebovat?

Databázi zákazníků

- Definujete si metriky, vlastnosti které budou klíčové pro charakteristiku

Co potřebuji definovat?

Customer ID	Name	Email	Gender	Age	Income	Postcode	Marital Status
3453450	James Smith	James.Smith@	M	32	55000	3013	M
3453451	John Davison	John.Davison@	M	29	62000	2122	S
3453452	Sarah Daniels	Sarah.Daniels@	F	22	43000	2522	M
3453453	Sally Jones	Sally.Jones@g	F	29	98000	2000	D
3453454	David Waters	David.Waters@	M	30	72000	3000	M
3453455	Isobel Archer	IsobelArcher@	F	35	53000	3037	S
3453456	Dianne Wilson	Dianne.Wilson@	F	42	67000	3040	S
3453457	Dennis Potter	MrPotterD@d	M	64	20000	3067	D
3453458	Jane Norton	JaneNort1987@	F	28	95000	3055	M
3453459	Janet Wilson	Janet.Wilson@	F	30	110000	3000	M

Finální příprava před analýzou

Kvantifikace: Převedení vybraných informací na čísla

Age	Income	Postcode	Marital Status
32	55000	3013	M
29	62000	2122	S
22	43000	2522	M
29	98000	2000	D
30	72000	3000	M
35	53000	3037	S
42	67000	3040	S
64	20000	3067	D
28	95000	3055	M
30	110000	3000	M



M	1
S	2
D	3

datalook

**Výstupem je důležitost
jednotlivých uživatelů vyjádřená
číslem v daném sloupci.**

datalook

Aplikace analýzy

Co s daty

Jednoduchá cesta

*Získané clustery si vezmu a použiju je pro
“look alike publika”*

Komplexní způsob

*Využiju reverzní engineering a data začnu
analyzovat a hledat souvislosti.*

Jaká uživatelská data mohu použít

- **demografické charakteristiky** (věk, pohlaví, příjem, vzdělání atd.),
- **geografické informace** (kde zákazník žije)
- **psychografické informace** (hodnoty, zájmy, životní styl atd.),
- **informace o chování** (historie nákupů, využití produktu, loajalita ke značce)

Začínáte analyzovat souvislosti

- Transakční data
- Data z registrace
- Soutěžní formuláře
- Věrnostní programy
- Data ze sociálních sítí

Employee ID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
E02002	Kai Lo	Controls Engineer	Engineering	Manufacturing	Male	Asian	47	2/5/2022	\$92,368	0%	United States	Columbus	
E02003	Robert Patel	Analyst	Sales	Corporate	Male	Asian	58	10/23/2013	\$45,703	0%	United States	Chicago	
E02004	Cameron Lo	Network Administrator	IT	Research & Development	Male	Asian	34	3/24/2019	\$83,576	0%	China	Shanghai	
E02005	Harper Castillo	IT Systems Architect	IT	Corporate	Female	Latino	39	4/7/2018	\$98,062	0%	United States	Seattle	
E02006	Harper Dominguez	Director	Engineering	Corporate	Female	Latino	42	6/18/2005	\$175,391	24%	United States	Austin	
E02007	Ezra Vu	Network Administrator	IT	Manufacturing	Male	Asian	62	4/22/2004	\$66,227	0%	United States	Phoenix	2/14/2014
E02008	Jade Hu	Sr. Analyst	Accounting	Specialty Products	Female	Asian	58	6/27/2009	\$85,744	0%	China	Chongqing	
E02009	Miles Chang	Analyst II	Finance	Corporate	Male	Asian	62	2/19/1999	\$69,674	0%	China	Chengdu	
E02010	Gianna Holmes	System Administrator	IT	Manufacturing	Female	Caucasian	38	9/9/2011	\$97,630	0%	United States	Seattle	
E02011	Jameson Thomas	Manager	Finance	Specialty Products	Male	Caucasian	52	2/5/2015	\$105,879	10%	United States	Miami	
E02012	Jameson Pena	Systems Analyst	IT	Manufacturing	Male	Latino	49	10/12/2003	\$40,499	0%	United States	Miami	
E02013	Bella Wu	Sr. Analyst	Finance	Specialty Products	Female	Asian	63	8/3/2014	\$71,418	0%	United States	Phoenix	
E02014	Jose Wong	Director	IT	Manufacturing	Male	Asian	45	11/15/2017	\$150,558	22%	China	Chongqing	
E02015	Lucas Richardson	Manager	Marketing	Corporate	Male	Caucasian	36	7/22/2018	\$118,912	8%	United States	Miami	
E02016	Jacob Moore	Sr. Manager	Marketing	Corporate	Male	Black	42	3/24/2021	\$131,422	15%	United States	Phoenix	
E02017	Luna Lu	IT Systems Architect	IT	Corporate	Female	Asian	62	7/26/1997	\$64,208	0%	United States	Miami	
E02018	Bella Tran	Vice President	Engineering	Specialty Products	Female	Asian	45	8/9/2010	\$254,486	33%	China	Chengdu	
E02019	Ivy Chau	Analyst	Sales	Specialty Products	Female	Asian	61	3/9/2019	\$54,811	0%	China	Chongqing	
E02020	Jordan Kumar	Service Desk Analyst	IT	Specialty Products	Male	Asian	29	11/13/2017	\$95,729	0%	United States	Seattle	
E02021	Sophia Gutierrez	Manager	Accounting	Specialty Products	Female	Latino	63	2/8/2009	\$102,649	6%	United States	Austin	
E02022	Eli Dang	Sr. Manager	Accounting	Specialty Products	Male	Asian	45	11/16/2015	\$122,875	12%	United States	Chicago	
E02023	Lillian Lewis	Technical Architect	IT	Research & Development	Female	Black	43	8/14/2013	\$83,323	0%	United States	Phoenix	3/31/2019
E02024	Serenity Cao	Account Representative	Sales	Manufacturing	Female	Asian	31	10/21/2018	\$66,721	0%	China	Shanghai	
E02025	Parker Lai	Vice President	Accounting	Specialty Products	Male	Asian	48	11/29/2006	\$46,400	36%	China	Chongqing	
E02026	Charles Simmons	Manager	Sales	Specialty Products	Male	Caucasian	55	10/27/1997	\$113,525	6%	United States	Miami	
E02027	Jayden Luu	Director	Accounting	Manufacturing	Male	Asian	64	5/13/2004	\$184,342	22%	China	Beijing	
E02028	Brooks Richardson	Director	Marketing	Specialty Products	Male	Caucasian	58	11/24/2020	\$151,941	22%	United States	Seattle	
E02029	Ivy Thompson	Manager	Marketing	Manufacturing	Female	Caucasian	50	8/11/2004	\$118,900	5%	United States	Seattle	
E02030	Peyton Wright	Sr. Manager	Marketing	Corporate	Female	Black	41	5/13/2017	\$153,370	10%	United States	Chicago	
E02031	Myatt Dinh	System Administrator	IT	Specialty Products	Male	Asian	50	3/15/2002	\$72,860	0%	China	Shanghai	
E02032	Ruby Alexander	Vice President	Finance	Research & Development	Female	Caucasian	59	8/13/2001	\$255,610	36%	United States	Phoenix	

datalook

Příklad aplikace dat

Subscription ve fitness

- Z Freemium na Premium
- Klíčové hodnoty uživatele
- + Zapojení uživatele

Jak to řešíme

- Provedeme cluster analýzu
- Analyzujeme souvislosti
- Vytvoříme si pravidla

Edit report SAVE

Name*
Segmentation - DEV

Categories*
Segmentation ADD CATEGORY

Full width content
Should the content be in full width of the page? Otherwise the width is 1000px.

Bottom iframe bar Right iframe bar

If the size is defined the white bar on the bottom will appear. Size in pixels. If the size is defined the white bar on the right will appear. Size in pixels.

database* schema* table*
DEV SOURCE CUSTOMERS_TEST

Any changes to database, schema or table will result in resetting all standing dimensions. All created segments are placed into 'SEGMENTATION' schema. The source schemas will not be modified in any way.

ID_IRER (string)
 LANGUAGE (string)
 GENDER (string)
 YEAR_OF_BIRTH (number)
 BMI (number)

LOAD TABLE DIMENSIONS

Jak to řešíme my

- Segmentační appka
- Odešleme do MKT nástrojů
- Necháme aktualizovat (volitelné)

Segmentation Overview

Total Customers
10000

Average Order Value

Average Number of Orders

Average Days between Orders

Your segments

ADD SEGMENT

Segments	Users in segment	Exporting in
Male-Age over 30	847	
Targito VIP	229	

Rows per page: 25 ▾ 1-2 of 2 < >

datalook

**Než se do toho pustíte
VYČISTĚTE SVOJE DATA**

Kontakt



Tomáš Zísler
CEO

T. [+420 734 242 847](tel:+420734242847)

E. tomas.zisler@datalook.cz