

Big data in database marketing

ANAR SULEYMANOV

Kapital Bank, Customer Management Dep.

Customer Data Analyst

Table of Contents

- What is data?
- What is big data?
- Big data technologies
- Database marketing
- Big data and machine learning in database marketing

What is data?

- Data is a set of values
 - Structured
 - Semi-structured
 - Unstructured

Structured data

Customer Information table

Customer_ID	Customer_Full_Name	Gender	Date_of_birth	Place_of_Birth	Phone	Marital_Status	Registered Branch
243105	Anar Suleymanov	M	18.06.1988	Baku	9061846	M	Narimanov
892393	Samir Karimov	M	13.03.1987	Lankaran	5063674	M	Ganjlik
127696	Shamil Suleymanov	M	20.12.1993	Gusar	6976606	M	Ganjlik
786594	Ilqar Mammadov	M	05.06.1987	Baku	8473937	M	Khatai
898806	Sehran Mammadov	M	04.04.1995	Goychay	8377357	S	Surakhani
106583	Nurane Mehdiyeva	F	12.07.1992	Baku	6766636	M	Narimanov
277820	Parviz Mammadov	M	01.11.1985	Kurdamir	8942111	S	Sumgait
377074	Farid Islamov	M	28.09.1994	Baku	9950339	M	28 May
920279	Guljan Nazarli	F	09.12.1992	Ganja	9161688	S	Kapaz

Product table

Customer_ID	Account	Product_Name	Product_Open_Date	Product_Close_Date	Price_AZN	Credit_Amount_AZN
243105	88744334222	Credit card	19.04.2018	19.04.2023	35	5000
243105	41520914104	Salary card	01.02.2017	01.02.2019	10	
786594	36503857902	Salary card	06.05.2016	06.05.2018	10	
786594	58394473640	Salary card	06.05.2018	06.05.2018	10	
377074	98537032699	Cash loan	10.12.2016	10.12.2018		10000
920279	21096104834	Credit card	19.10.2016	19.10.2021	15	3000
920279	52087297484	Cash loan	12.08.2017	19.08.2020		5000
898806	15790285723	Auto credit	14.03.2016	14.03.2020		35000

Transactions table

Account	Operation_Date	Device	Operation Type	Amount_AZN	Merchant_Desc	Location
88744334222	21.04.2018	POS	POS transaction	42	Supermarkets	Bazarstore
88744334222	01.05.2018		Non cash inflow	1300		
88744334222	03.05.2018	ATM	ATM Withdrawal	500	ATM Withdrawal	Narimanov
88744334222	10.05.2018	POS	POS Mobile	250	Utilities	
41520914104	18.07.2017		Non cash inflow	500		
41520914104	19.07.2017	ATM	ATM Withdrawal	500	ATM Withdrawal	Ganklik



Unstructured data



- Photos
- Messages and posts
- Voice calls
- Videos

What is big data?

- Variety

Structured

Unstructured

Semi-structured

- Velocity

Real-time

Near-time

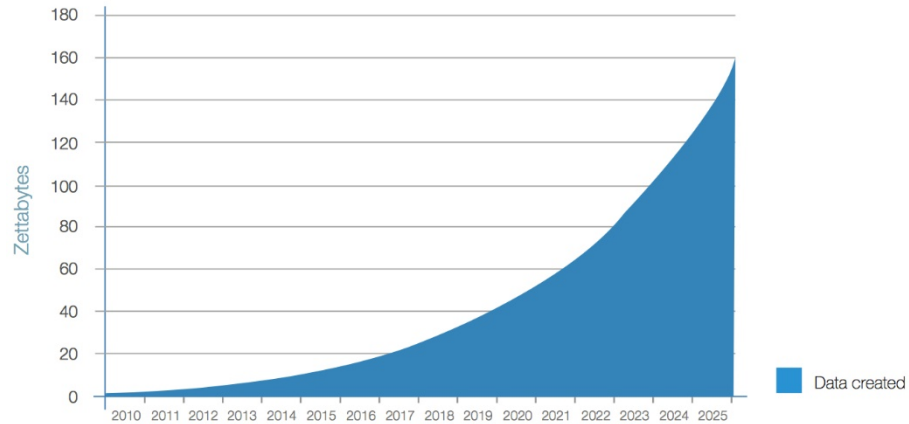
Periodic

- Volume

Amount of data

Transactions table

Size of generated data in globe



37 zettabyte
=
37 billion



Big data technologies

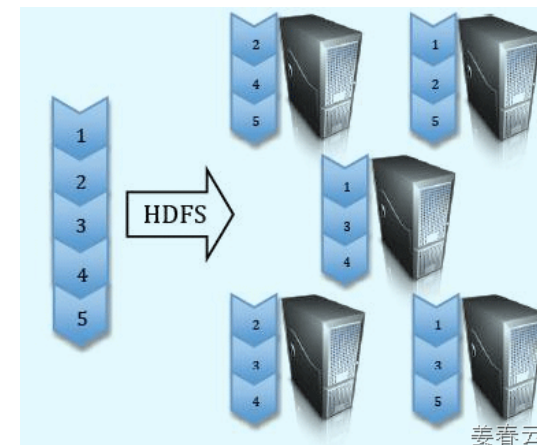


Hadoop



- Open source
- Java based
- Storage of extremely large data

Hadoop Distributed File System filesystem designed for storing very large files with streaming data access patterns, running on clusters of commodity hardwares



Hive

facebook

MapReduce Scripts Examples

```
add file page_url_to_id.py;
add file my_python_session_cutter.py;
FROM
    (SELECT TRANSFORM(uhash, page_url, unix_time)
     USING 'page_url_to_id.py'
     AS (uhash, page_id, unix_time)
    FROM mylog
    DISTRIBUTE BY uhash
    SORT BY uhash, unix_time) mylog2
SELECT TRANSFORM(uhash, page_id, unix_time)
    USING 'my_python_session_cutter.py'
    AS (uhash, session_info);
```

Mahout

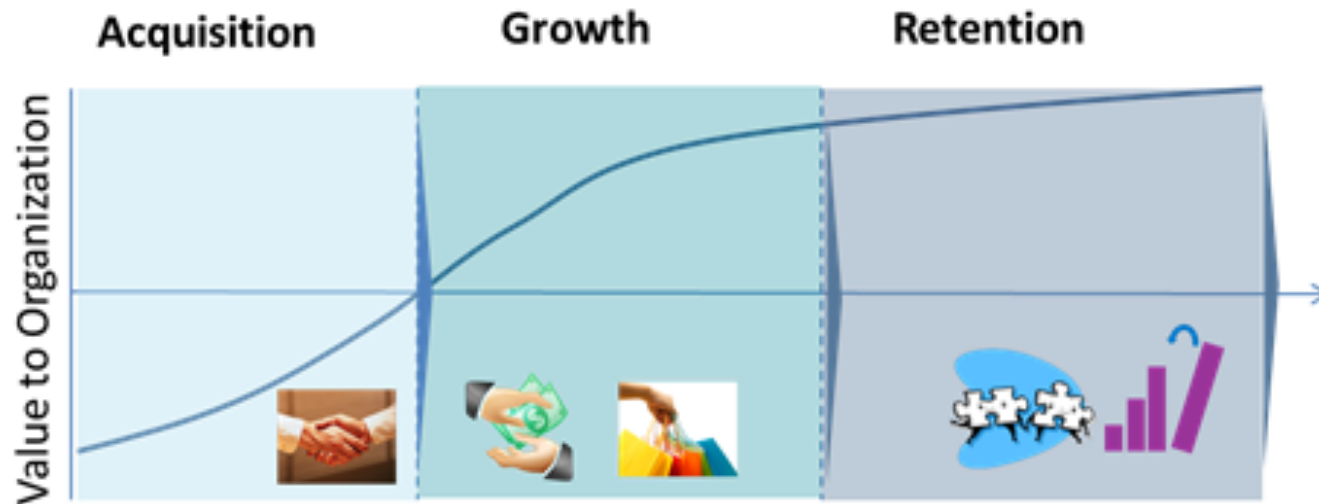


Apache Mahout is a machine-learning engine that fits into the data mining category of the big data landscape

- Recommendation systems
- Predictive modelling
- Clustering

What is database marketing?

Database marketing is a form of **direct** marketing using databases of customers or potential customers to generate personalized communications in order to promote a product or service for marketing purposes.

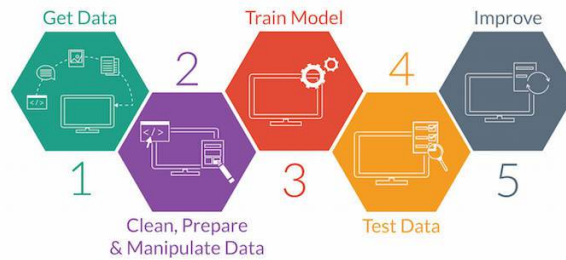


What is database marketing?

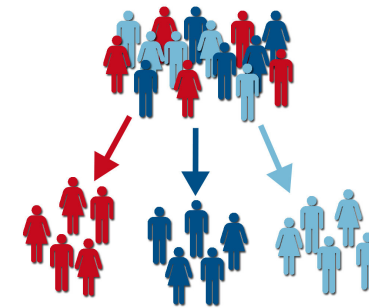
Machine learning applications



Recommendation systems
& market basket analysis



Predictive modelling



Clustering

Recommendation systems

amazon

NETFLIX

YouTube

Quora

ebay

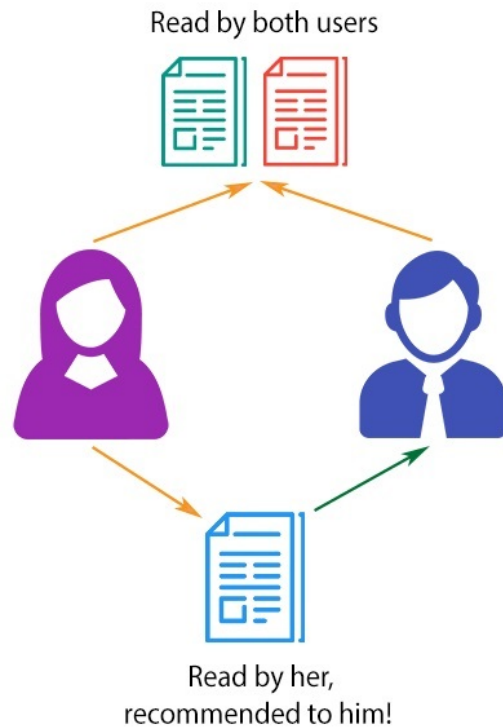


coursera

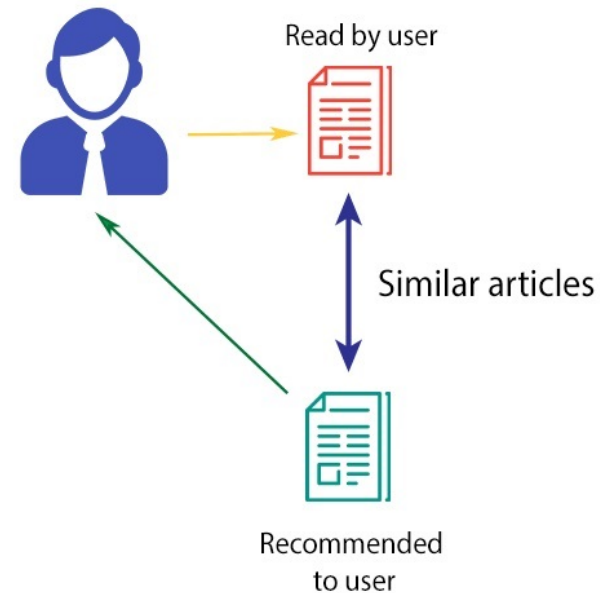
facebook®

Recommendation systems

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING



Recommendation systems

Grant, Welcome to Your Amazon.com (If you're not Grant Ingersoll, click here.)

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#).


 Principles of Data Mining (A... by David J... ★★★★☆ (17) \$52.00	 Python in a Nutshell, Secon... by Alex Mart... ★★★★☆ (40) \$26.39	 Introductory Statistics wit... by Peter Dal... ★★★★☆ (20) \$48.56
--	---	---



Recommendations for you.

We combed our catalog and found courses and Specializations that we think match your interests. Browse our recommendations below, and start learning something new today!

Degrees + Professional Certificates



Earn your Master's degree on Coursera
Build high-demand skills in accounting, business leadership, or data science and receive a graduate degree from a top university

[Learn more](#)

Specializations + Courses

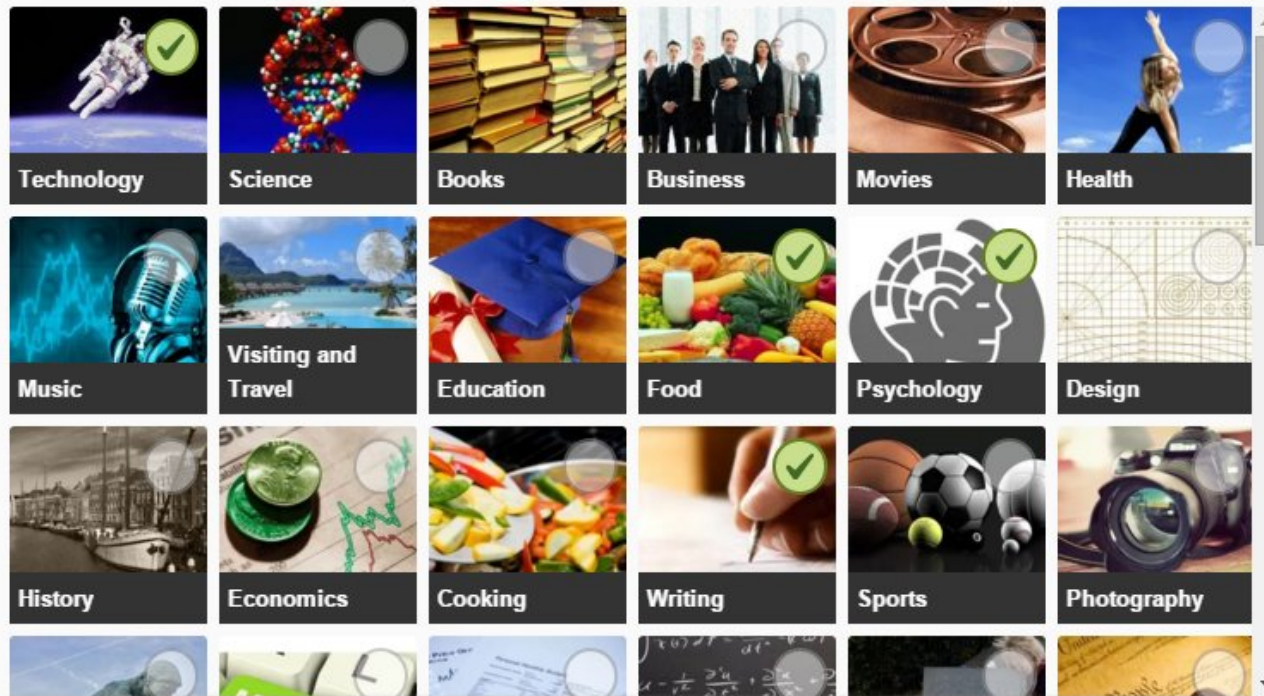
	Qualitative Research University of California, Davis Starts May 14, 2018
	Search Engine Optimization (SEO) University of California, Davis 6 Course Specialization Starts May 14, 2018



Recommendation systems

Follow topics to see the best answers about them

Select all the topics you're interested in.



6 More Topics to Continue

Quora

Market basket analysis



Market basket analysis



Market basket analysis

Market Basket Analysis in Kapital Bank



Supermarkets



Perfume Stores



Utilities



Beauty Salon

Predictive modelling

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



Kashmir Hill, FORBES STAFF ✓

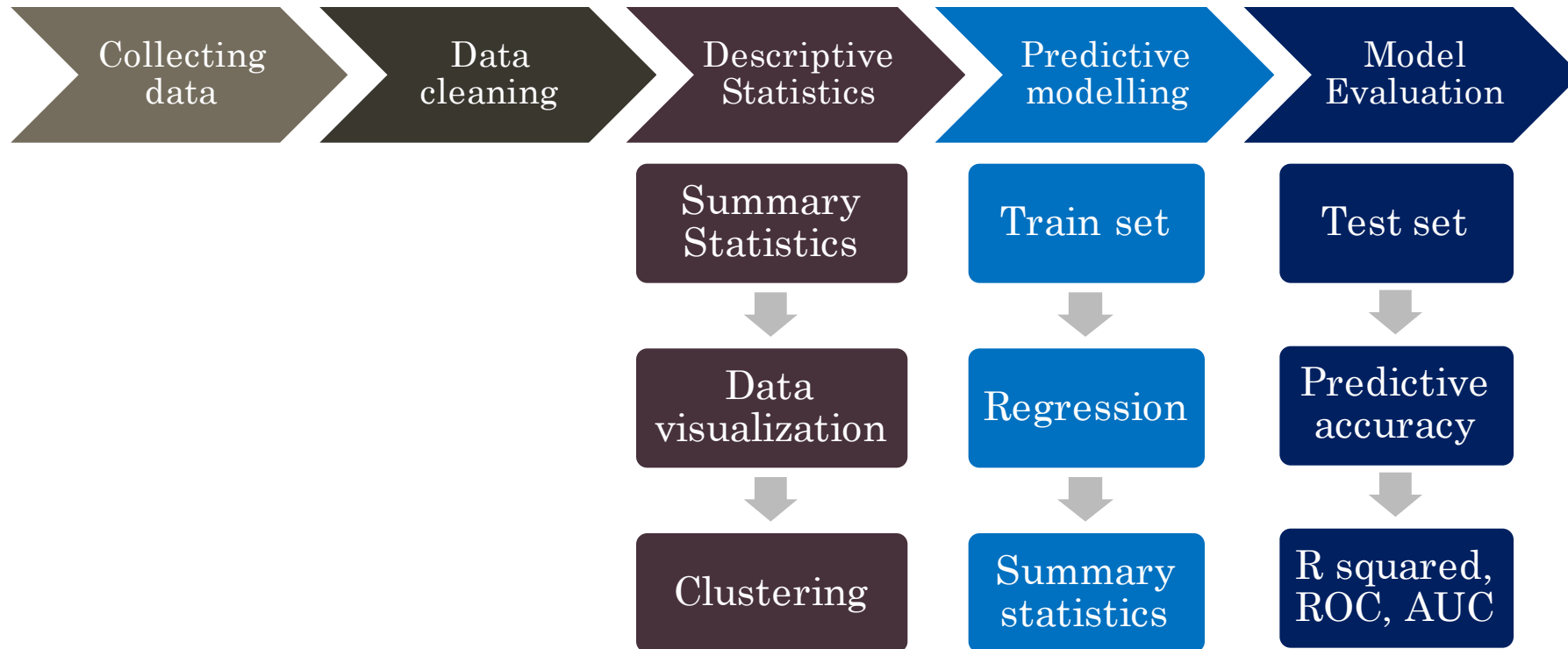
Welcome to The Not-So Private Parts where technology & privacy collide [FULL BIO](#) ✓

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. [Target](#) TGT +0.8%, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a



Andrew Pole

Predictive modelling



Predictive modelling

Train set

ID	Age	Gender	Purchased
609120	23	Male	Yes
931157	34	Female	Yes
323263	56	Male	No
364971	21	Male	Yes
939706	18	Male	Yes
171702	71	Male	No
922208	43	Female	No
270314	27	Male	Yes
938345	46	Female	Yes
318302	54	Male	Yes
942375	31	Male	No

Test set

ID	Age	Gender	Purchased	Predicted Score	Predicted Output
603010	46	Female	Yes	0.6	Yes
543606	39	Female	Yes	0.7	Yes
402867	65	Male	No	0.8	Yes
444470	21	Male	Yes	0.3	No
564456	59	Male	Yes	0.6	Yes
897401	43	Female	No	0.2	No

Predicted Output

	No	Yes
Purchased No	1	1
Purchased Yes	1	3

Logistic regression output

Linear formula = $B_0 + B_1 * \text{Age} + B_2 * \text{Gender}(\text{Male})$

Probability of Purchase = $1 / (1 + \exp(-1 * (B_0 + B_1 * \text{Age} + B_2 * \text{Gender}(\text{Male})))$

Predictive accuracy = 0.67 or 67%

Predictive modelling

Application set

ID	Age	Gender	Predicted Scores	Predicted Output
2576290	31	Female	0,4	No
9202372	43	Male	0,7	Yes
5887922	37	Male	0,7	Yes
8007730	45	Male	0,6	Yes
4909796	39	Female	0,4	No
8836760	39	Male	0,1	No
8491362	30	Female	0,2	No
6407487	67	Female	0,9	Yes
1206382	36	Male	0,6	Yes
3695979	58	Male	0,7	Yes
9831207	45	Female	0,8	Yes

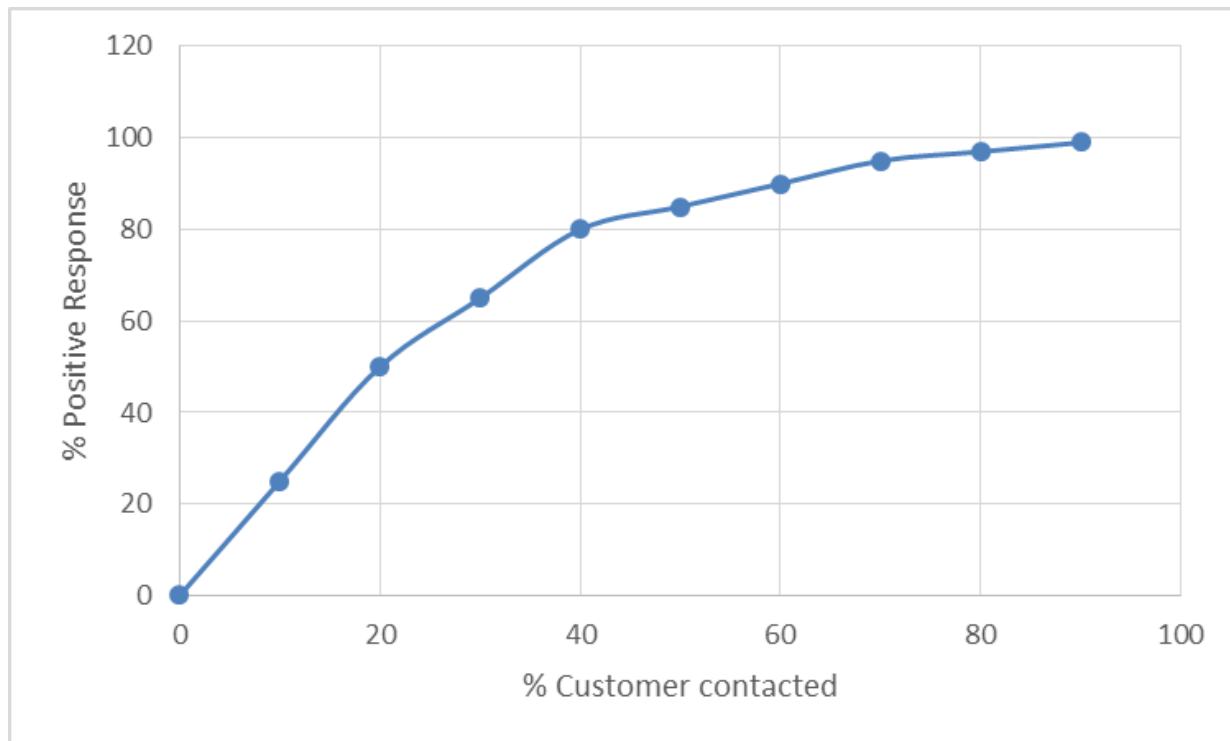
Logistic regression output

Linear formula = $B_0 + B_1 * \text{Age} + B_2 * \text{Gender}(\text{Male})$

Probability of Purchase = $1 / (1 + \exp(-1 * (B_0 + B_1 * \text{Age} + B_2 * \text{Gender}(\text{Male})))$

Predictive modelling

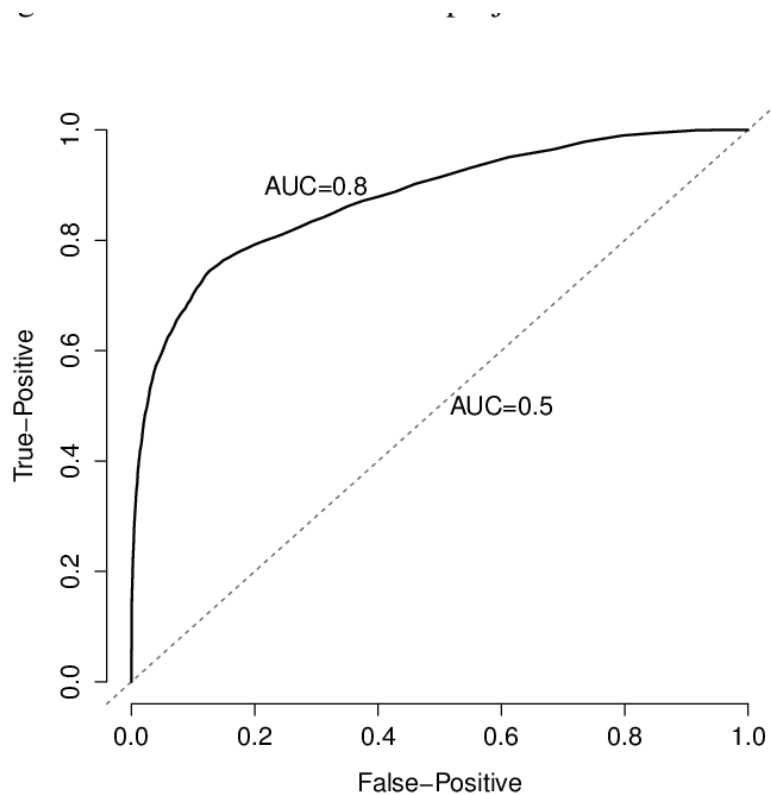
Predictive modelling in Kapital Bank



Cumulative gain graph

Predictive modelling

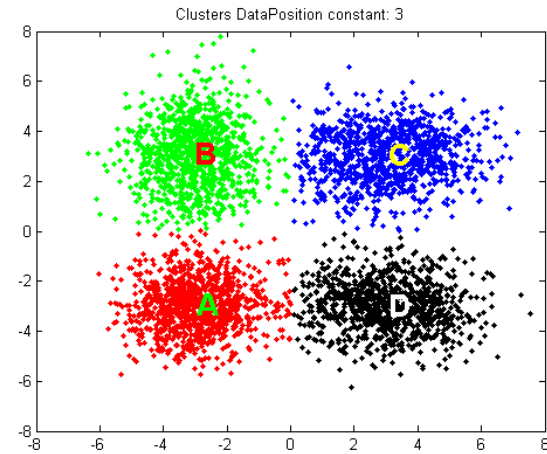
Predictive modelling in Kapital Bank



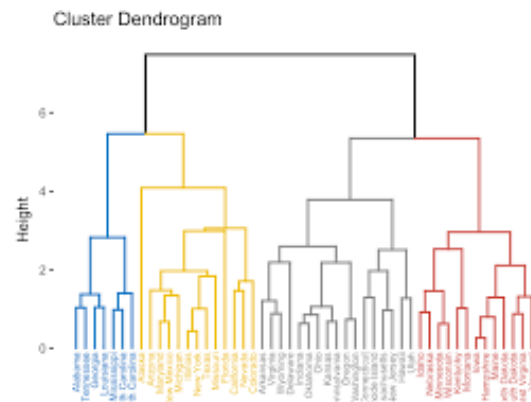
Bir Kart	Cash loan	Final offer
0,7	0,6	Bir Kart
0,4	0,7	Cash loan
0,4	0,1	Bir Kart
0,9	0,8	Bir Kart
0,1	0,6	Cash loan
0,2	0,8	Cash loan
0,9	0,8	Bir Kart
1	0,3	Bir Kart

Clustering

K-means

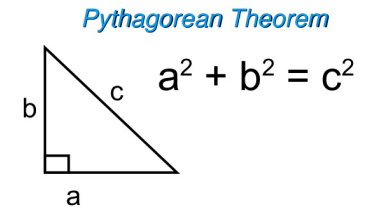
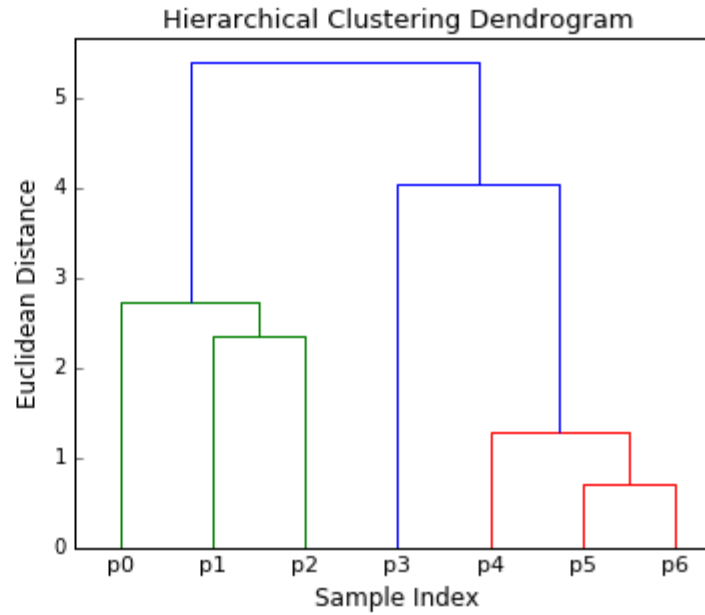
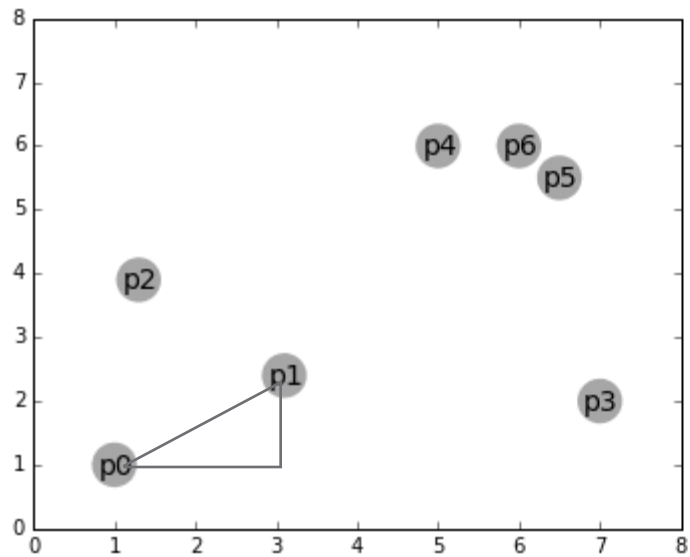


Hierarchical clustering

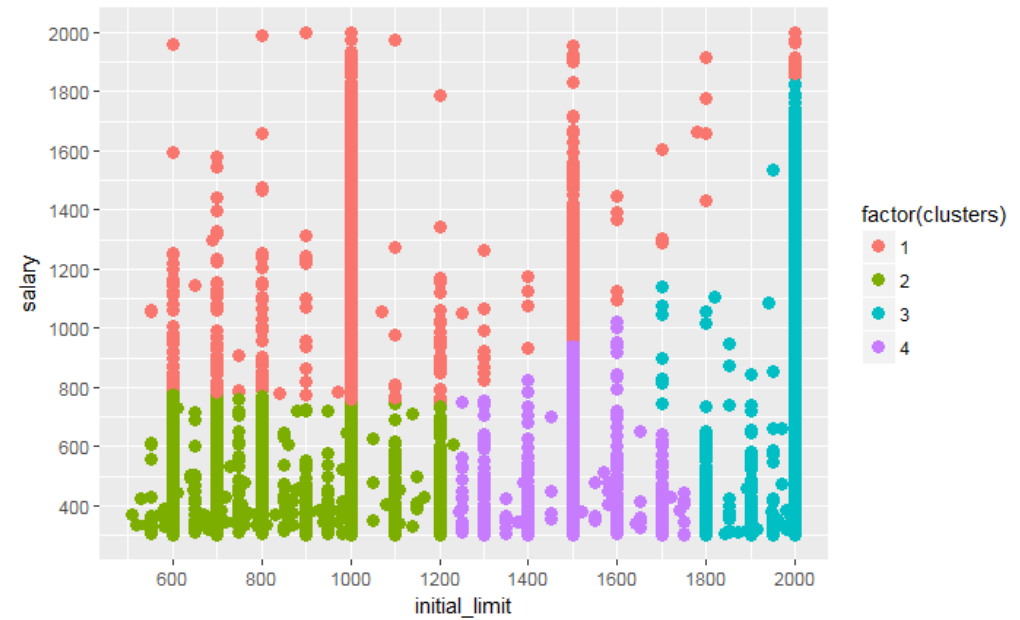
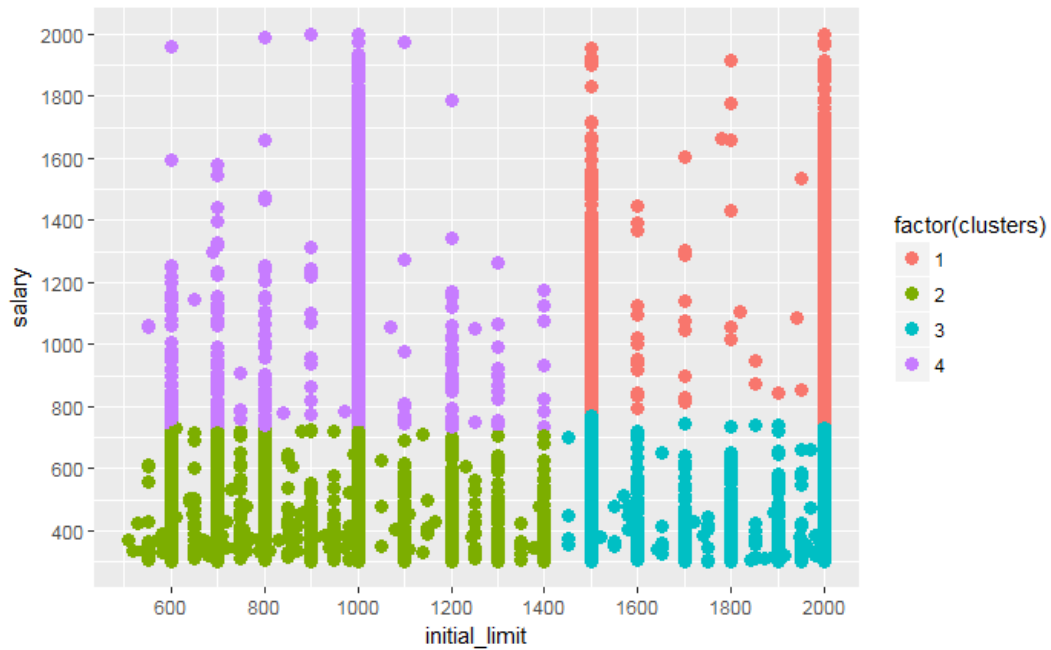


Clustering

Hierarchical clustering

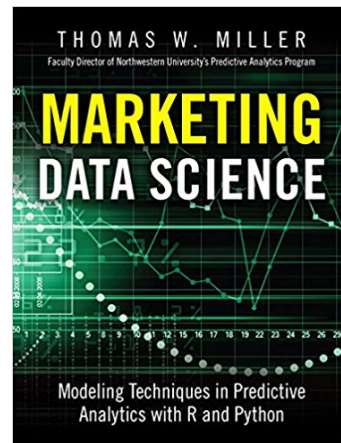
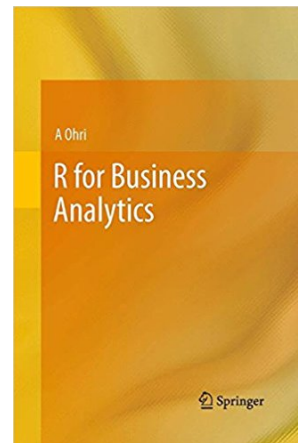
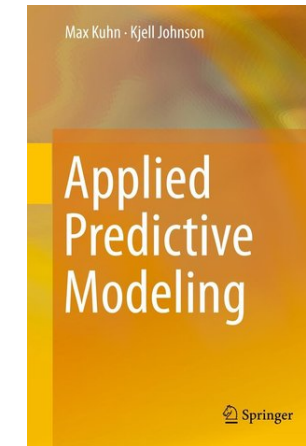
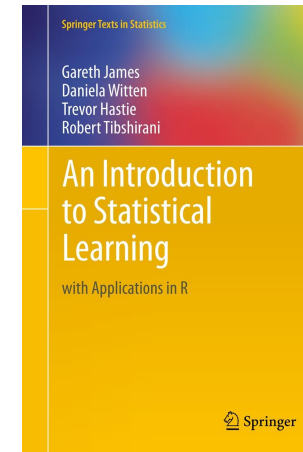
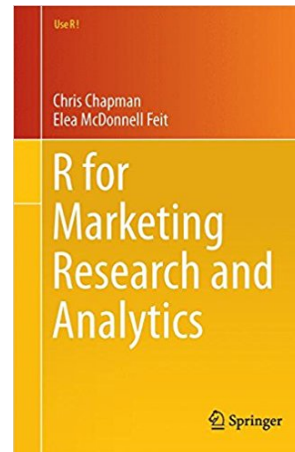
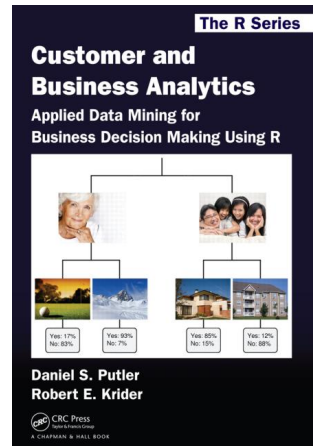
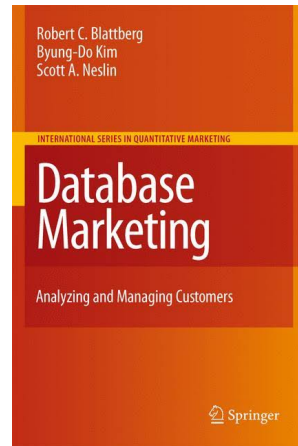


Clustering



Set random seed

Additional sources



Thank you for attention!

E-mail: anar.suleymanov@kapitalbank.az