



Usage of DBSCAN clusterization method to the set of housing sale offers

Paweł Szoltysek

18 May 2009

Agenda



Agenda

Motivation
Housing offers
Clusterization
Results
Summary

Agenda
Motivation
Housing offers
Clusterization
Results
Summary



Preface

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

What's the point in dealing with housing sale offers?

- ✓ Many of us are looking into housing prices.
- ✓ The properties are treated as investments.

Real estate market is instable.

Preface

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

What's the point in dealing with housing sale offers?

- ✓ Many of us are looking into housing prices.
- ✓ The properties are treated as investments.

Real estate market is instable.

Preface

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

What's the point in dealing with housing sale offers?

- ✓ Many of us are looking into housing prices.
- ✓ The properties are treated as investments.

Real estate market is instable.

Preface

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

What's the point in dealing with housing sale offers?

- ✓ Many of us are looking into housing prices.
- ✓ The properties are treated as investments.

Real estate market is instable.

Situation in 1992

Agenda

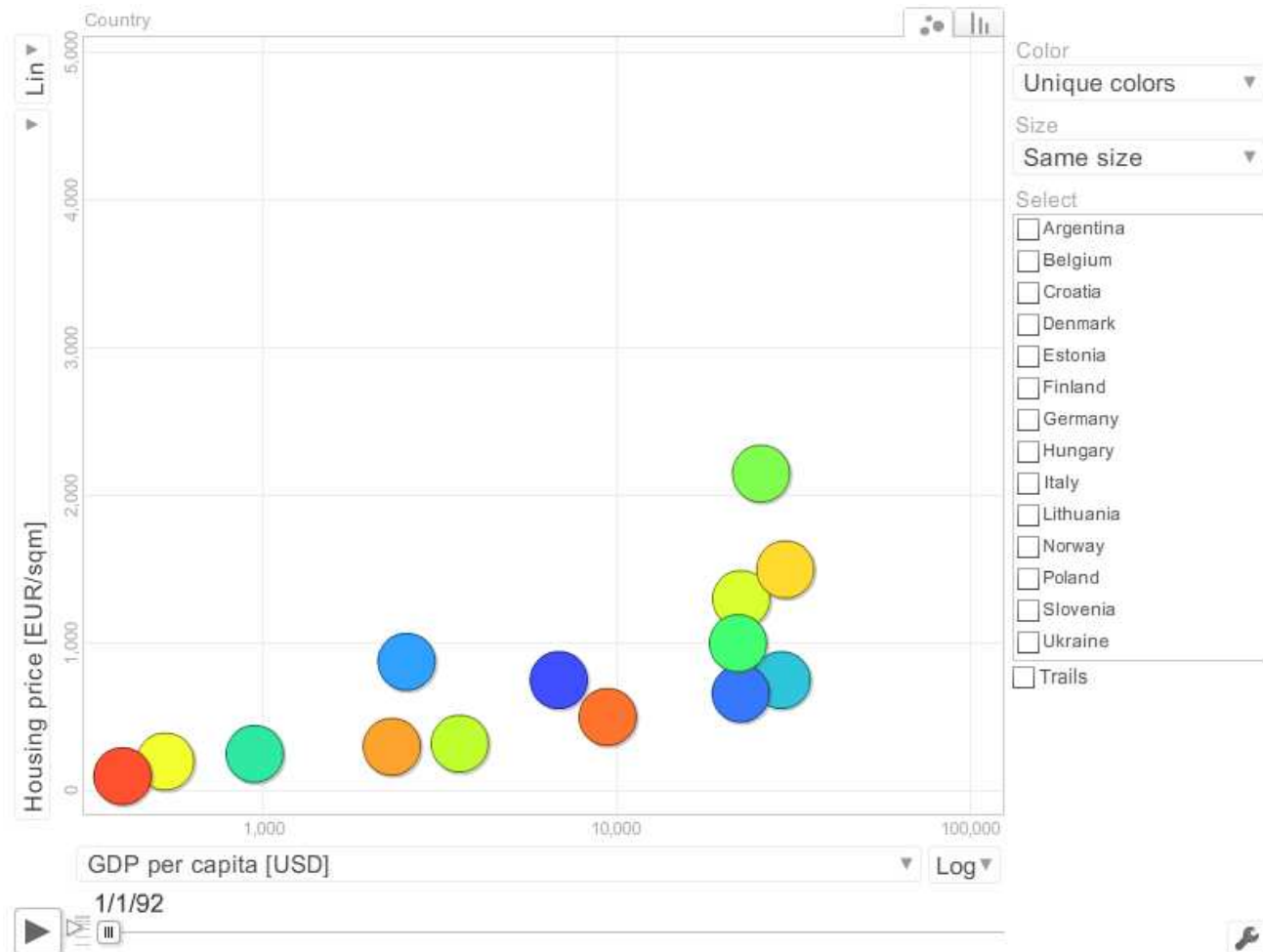
Motivation

Housing offers

Clusterization

Results

Summary



Situation in 1992

Agenda

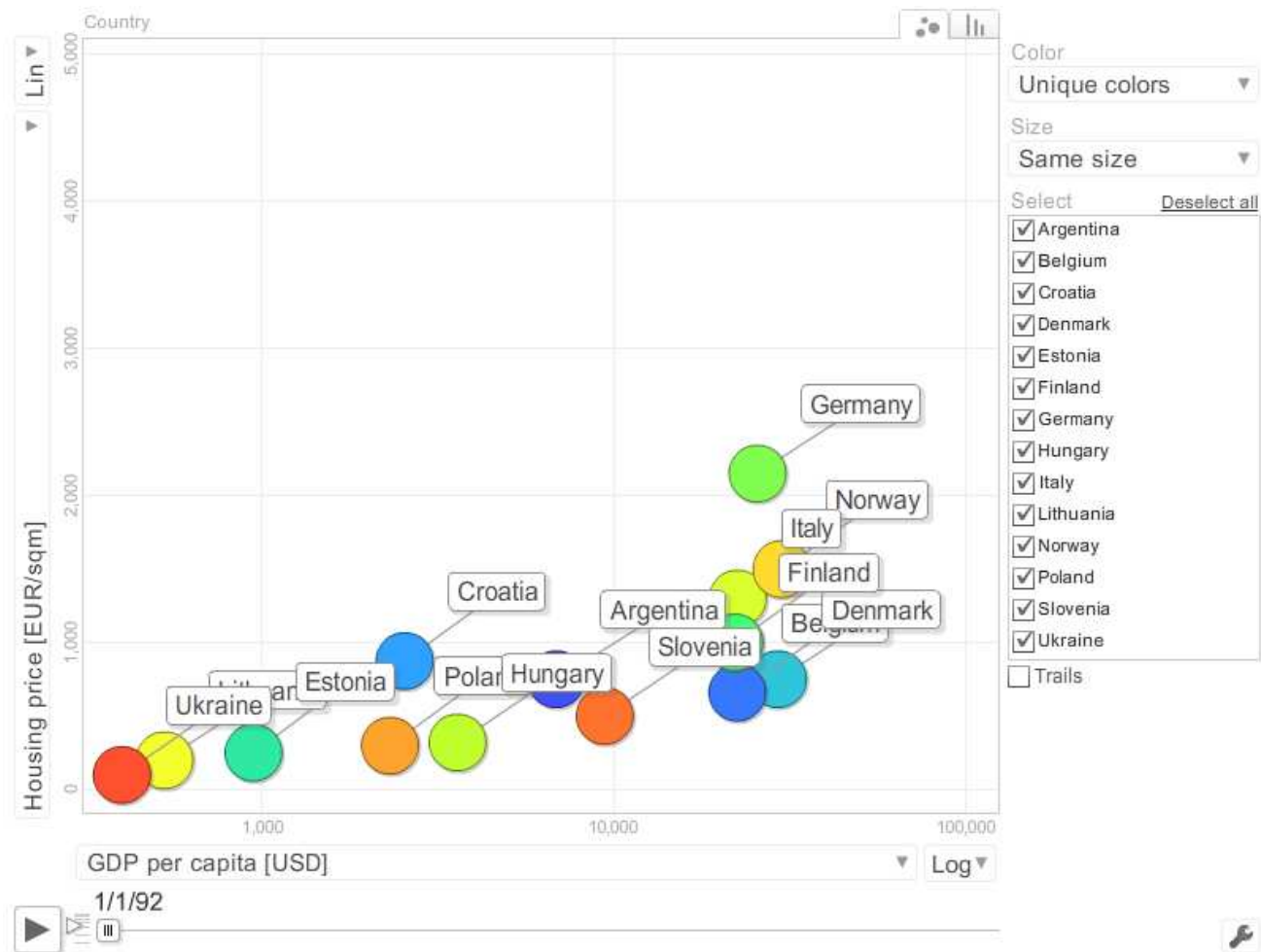
Motivation

Housing offers

Clusterization

Results

Summary



1992 - 1999 : Norway overtakes Germany

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

(sampl1.mpg)



2000-2003 : Argentine economic crisis

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

(sampl2.mpg)

2004 : EU expansion; Orange Eevolution in Ukraine

Agenda

Motivation

Housing offers

Clusterization

Results

Summary

(sampl3.mpg)

2008 : Global financial crisis

Agenda

Motivation

Housing offers

Clusterization

Results

Summary



(sampl4.mpg)

House price changes, 5 years (%)

Agenda

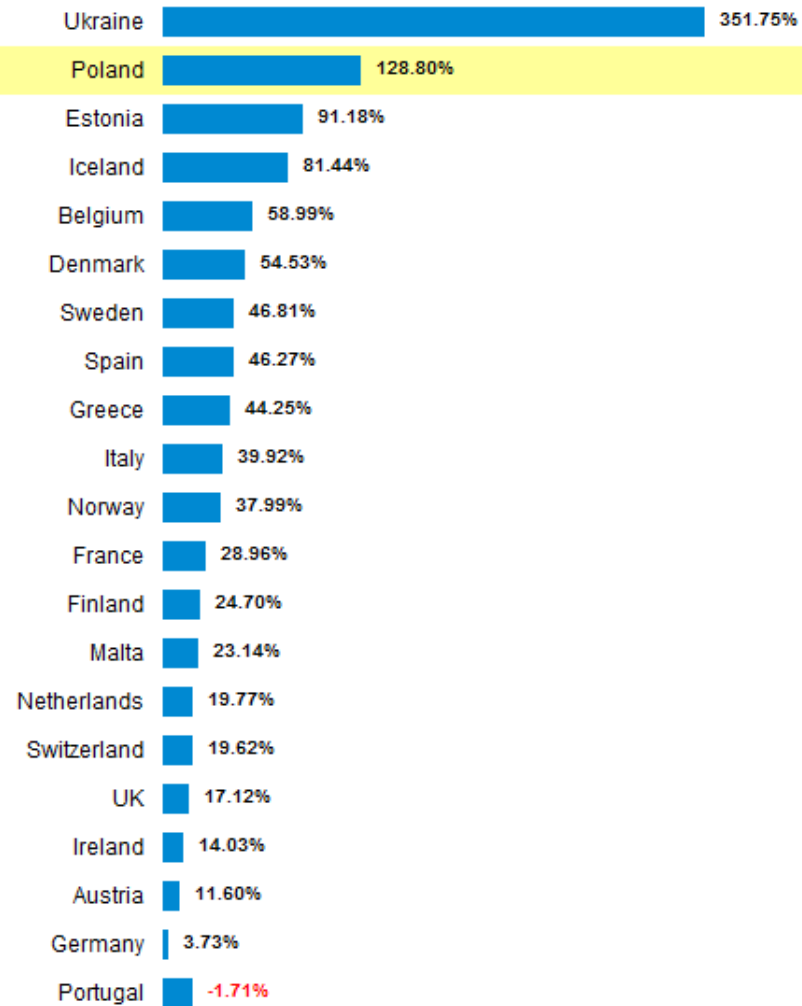
Motivation

Housing offers

Clusterization

Results



Summary



Housing offers data

- Agenda
- Motivation
- Housing offers**
- Clusterization
- Results
- Summary

Data were taken from local web advertisement services, such as otodom, emieszkania and domiporta.

	mieszkanie dwupokojowe Powierzchnia 34 m2	Wrocław Wejherowska	220 000,00 PLN do negocjacji
	mieszkanie dwupokojowe w centrum, Pułaskiego Powierzchnia 54 m2	Wrocław Pułaskiego	360 000,00 PLN do negocjacji
	Mieszkanie Partynice Powierzchnia 115 m2	Wrocław Gen. Stanisława Maczka	8 650,00 PLN do negocjacji
	Mieszkanie Krzyki Borek Powierzchnia 87 m2	Wrocław Burzowa	620 000,00 PLN do negocjacji

Housing offers data

- Agenda
- Motivation
- Housing offers**
- Clusterization
- Results
- Summary

It is assumed that:

- ✓ set is bounded to flats;
- ✓ real estates are located within Wrocław agglomeration;
- ✓ are from secondary market;
- ✓ flats rightfully belongs to seller;
- ✓ are not on debt.

Such assumptions makes clusterization process easier, but still complies to any other situation.

Housing offers data

- Agenda
- Motivation
- Housing offers**
- Clusterization
- Results
- Summary

It is assumed that:

- ✓ set is bounded to flats;
- ✓ real estates are located within Wrocław agglomeration;
- ✓ are from secondary market;
- ✓ flats rightfully belongs to seller;
- ✓ are not on debt.

Such assumptions makes clusterization process easier, but still complies to any other situation.

Set of attributes

- Agenda
- Motivation
- Housing offers**
- Clusterization
- Results
- Summary

Final set of attributes:

- ✓ Localization (eg distance to city centre, access to public means of transport etc);
- ✓ Building (eg number of floors, year of build);
- ✓ House (size, number of rooms and bathrooms/toilets, floor, parking place etc);
- ✓ Price.

In total: 15 attributes, where price is most important one. All values of attributes were normalized, analyzed, tuned and then directed to the clusterization algorithm.

Clusterization

Agenda
Motivation
Housing offers

Clusterization

Results
Summary

In clusterization process, DBSCAN algorithm has been used.

Two approaches to the clusterization:

- ✓ Clustering *with* the prices;
- ✓ Clustering *without* the prices.

Improvements in DBSCAN method are basing on used measure of ϵ - which meant total difference between offer attributes.

Tests has been performed basing on 113 housing offers.

Clusterization

Agenda
Motivation
Housing offers

Clusterization

Results
Summary

In clusterization process, DBSCAN algorithm has been used.

Two approaches to the clusterization:

- ✓ Clustering *with* the prices;
- ✓ Clustering *without* the prices.

Improvements in DBSCAN method are basing on used measure of ϵ - which meant total difference between offer attributes.

Tests has been performed basing on 113 housing offers.

Clusterization

- Agenda
- Motivation
- Housing offers
- Clusterization**
- Results
- Summary

In clusterization process, DBSCAN algorithm has been used.
Two approaches to the clusterization:

- ✓ Clustering *with* the prices;
- ✓ Clustering *without* the prices.

Improvements in DBSCAN method are basing on used measure of ϵ - which meant total difference between offer attributes.
Tests has been performed basing on 113 housing offers.

Clusterization

- Agenda
- Motivation
- Housing offers
- Clusterization**
- Results
- Summary

In clusterization process, DBSCAN algorithm has been used.
Two approaches to the clusterization:

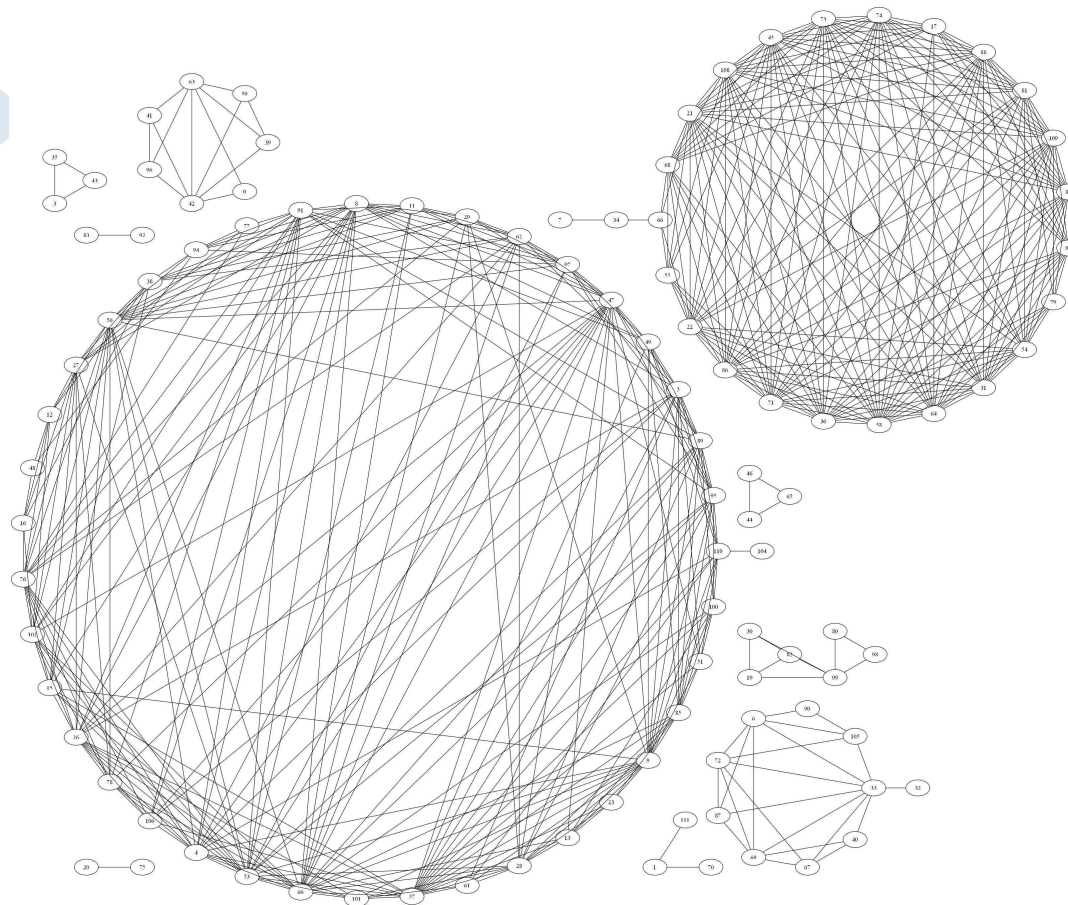
- ✓ Clustering *with* the prices;
- ✓ Clustering *without* the prices.

Improvements in DBSCAN method are basing on used measure of ϵ - which meant total difference between offer attributes.
Tests has been performed basing on 113 housing offers.

Clusterization Results - with prices

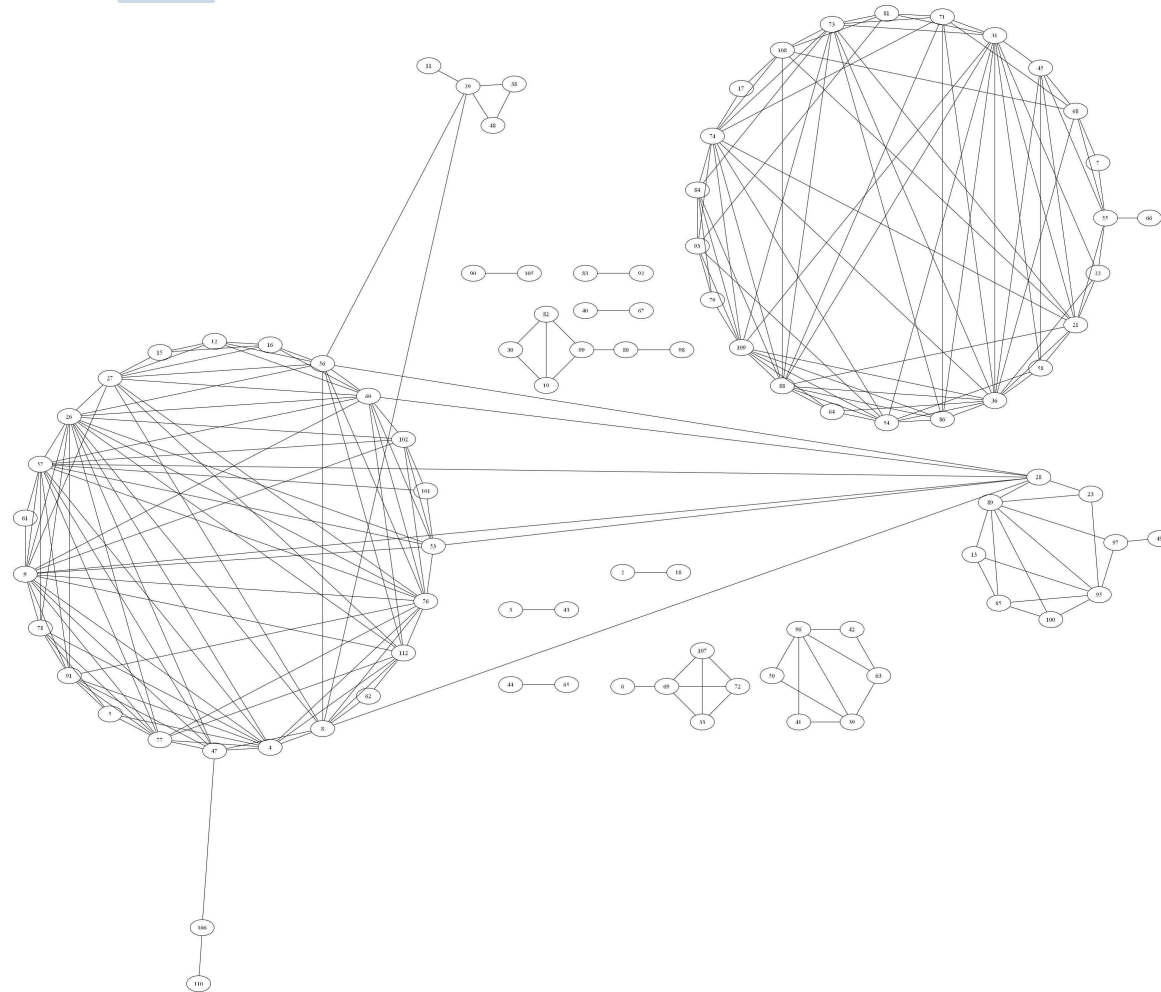
- Agenda
- Motivation
- Housing offers
- Clusterization
- Results**
- Summary

Best results: $\epsilon = 0.2$, standard DBSCAN procedure.
28 offers has been chosen.



Clusterization Results - without prices

Best results: $\epsilon = 0.125$, standard DBSCAN procedure.



Summary

- Agenda
- Motivation
- Housing offers
- Clusterization
- Results
- Summary

There is a need to have tools that will make decision-making process easier, even in real estate market.

Even easy, pure-mathematic methods, like DBSCAN clustering method, can significantly narrow the set of offers that are worthy of interest.

Further research fields.

For provided, random set of offers, the extraction of the best of them was finished with full success. What's more, different approaches to the problem has resulted in similar set of offers.

Summary

- Agenda
- Motivation
- Housing offers
- Clusterization
- Results
- Summary

There is a need to have tools that will make decision-making process easier, even in real estate market.

Even easy, pure-mathematic methods, like DBSCAN clustering method, can significantly narrow the set of offers that are worthy of interest.

Further research fields.

For provided, random set of offers, the extraction of the best of them was finished with full success. What's more, different approaches to the problem has resulted in similar set of offers.

Summary

- Agenda
- Motivation
- Housing offers
- Clusterization
- Results
- Summary

There is a need to have tools that will make decision-making process easier, even in real estate market. Even easy, pure-mathematic methods, like DBSCAN clustering method, can significantly narrow the set of offers that are worthy of interest.

Further research fields.

For provided, random set of offers, the extraction of the best of them was finished with full success. What's more, different approaches to the problem has resulted in similar set of offers.

Summary

- Agenda
- Motivation
- Housing offers
- Clusterization
- Results
- Summary

There is a need to have tools that will make decision-making process easier, even in real estate market. Even easy, pure-mathematic methods, like DBSCAN clustering method, can significantly narrow the set of offers that are worthy of interest.

Further research fields.

For provided, random set of offers, the extraction of the best of them was finished with full success. What's more, different approaches to the problem has resulted in similar set of offers.