Who are your best clients? Finding unusually strong buyer-seller links using the power of statistics

Miklós Koren
CEU Economics
#ceubusinessanalytics

CEU Webinar

M.Sc. in Business Analytics

CEU highlights

- CEU is a world class American university in the heart of Central Europe.
- One of the most internationally diverse school in the world:
  - 1,500 students and 300 faculty from 130 countries alumni chapters in 72 countries, representatives in 150 countries
  - All programs accredited in U.S.
- Faculty with Ph.D. from elite universities (e.g., Harvard, Yale, MIT, Princeton) and industry experience (e.g., IBM, Epoch, CARD.com, Prezi, Rapidminer).

M.Sc. in Business Analytics

- 1-year program, accredited in U.S.
  - full time or part time
- Courses in four domains
  - Technology: managing data
  - Statistics: analyzing data
  - Economics: interpreting fata
  - Management: acting on data
- Capstone project
  - Complete analytics project (solving a business need) at one of our industry partners.
  - Current examples: machine learning based segmentation of tumors in CT scans, residential real estate demand, creditworthiness from payment patterns, optimizing travel costs at scale

Part 1: The power of statistics

Data science

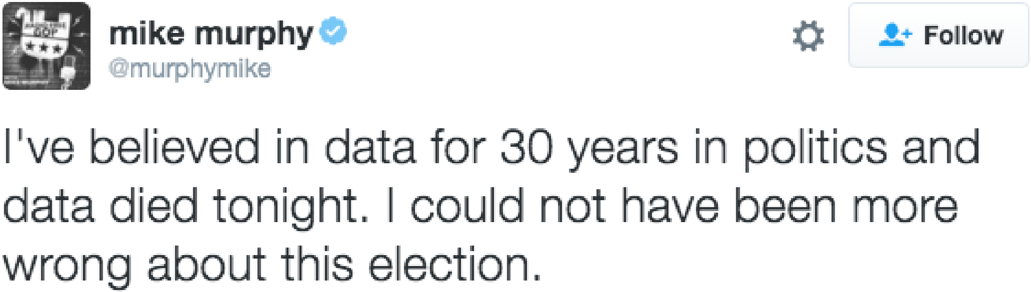
"procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data." (Tukey,)

Data science

"procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data." (Tukey, 1962)

A data scientist works with data to make predictions.

“Data died tonight.”



The power of statistics

Statistics helps us

1. discover patterns in data
2. quantify trends
3. predict the future

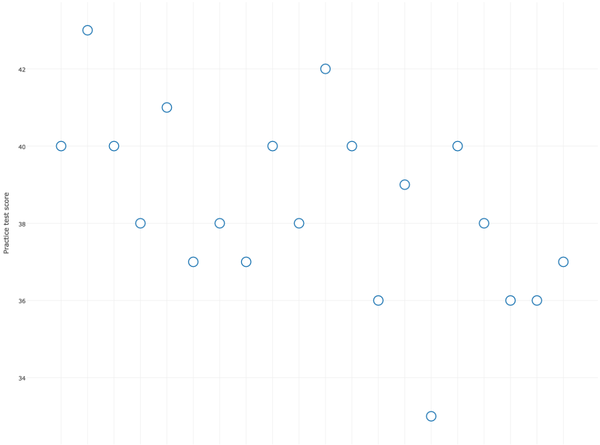
Cross Industry Standard Process for Data Mining (CRISP-DM)

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

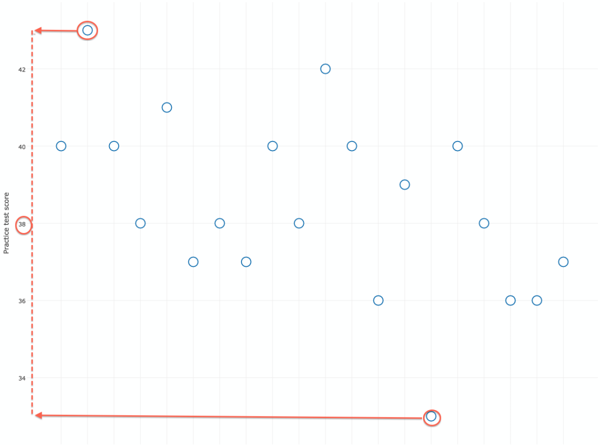
1. Problem

You want to predict your score on a standardized test.

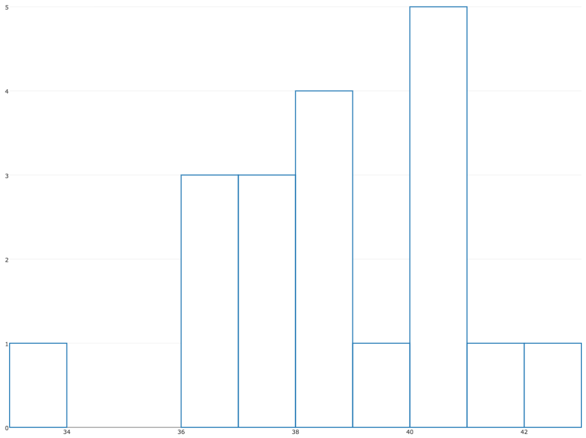
2. Data: Practice Test Scores



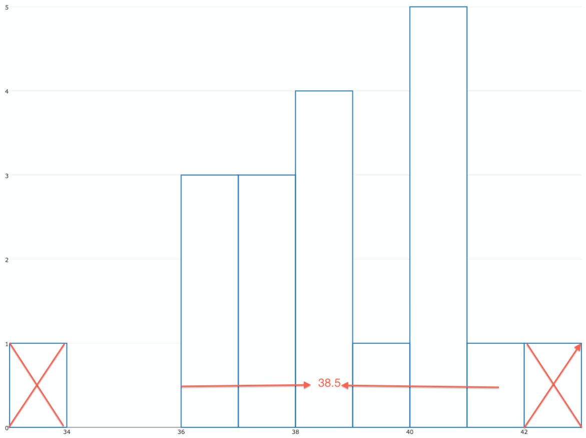
3. One Predictive Model



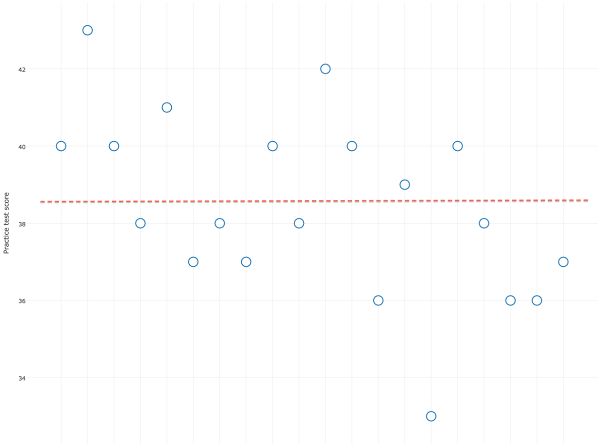
3. The Histogram of Test Scores



3. Another Predictive Model



3. Regression to the Mean



Predict the future

Most statistical predictions are based on the idea that things will look like they did in the past.

Predict the future

Most statistical predictions are based on the idea that things will look like they did in the past.
This is captured by *regression to the mean.*

Predict the future

Most statistical predictions are based on the idea that things will look like they did in the past.
This is captured by *regression to the mean.*
But not all predictions are like that.

North Sea flood of 1953

Predicting extreme values

Notation	$\mathrm{GEV}(\mu, \sigma, \xi)$
Parameters	$\mu \in \mathbf{R}$ — location, $\sigma > 0$ — scale, $\xi \in \mathbf{R}$ — shape.
Support	$x \in [\mu - \sigma/\xi, +\infty)$ when $\xi > 0$, $x \in (-\infty, +\infty)$ when $\xi = 0$, $x \in (-\infty, \mu - \sigma/\xi]$ when $\xi < 0$.
pdf	$\frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}$, where $t(x) = \begin{cases} \left(1 + \left(\frac{x-\mu}{\sigma}\right)\xi\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & \text{if } \xi = 0 \end{cases}$
CDF	$e^{-t(x)}$, for $x \in$ support
Mean	$\begin{cases} \mu + \sigma \frac{\Gamma(1-\xi)-1}{\xi} & \text{if } \xi \neq 0, \xi < 1, \\ \mu + \sigma\gamma & \text{if } \xi = 0, \\ \infty & \text{if } \xi \geq 1, \end{cases}$ where γ is Euler's constant.

Predicting extreme values

Potential use cases:

- risk analysis
- fraud detection

Lessons learned

1. Statistics can help discover patterns and make predictions.
2. Predictions are as good as past data.
3. You can also use statistics to predict the *unprecedented*.

Lessons learned

1. Statistics can help discover patterns and make predictions.
2. Predictions are as good as past data.
3. You can also use statistics to predict the *unprecedented*.

Questions?

Part 2: Finding unusually strong buyer-supplier links

Problem

Who are your best clients? Do some sales reps have particularly strong links to some of your buyers?

CRISP-DM

1. Business Understanding: Does each account belong to only one sales rep? Is your sales segmented by region/products?

Problem

Who are your best clients? Do some sales reps have particularly strong links to some of your buyers?

CRISP-DM

1. Business Understanding: Does each account belong to only one sales rep? Is your sales segmented by region/products?
2. Data Understanding: Does your CRM uniquely identify clients and sales reps? Can you augment it with outside information on buyers?

Problem

Who are your best clients? Do some sales reps have particularly strong links to some of your buyers?

CRISP-DM

1. Business Understanding: Does each account belong to only one sales rep? Is your sales segmented by region/products?
2. Data Understanding: Does your CRM uniquely identify clients and sales reps? Can you augment it with outside information on buyers?
3. Data Preparation: How to best represent the data for statistical modeling?

Problem

Who are your best clients? Do some sales reps have particularly strong links to some of your buyers?

CRISP-DM

1. Business Understanding: Does each account belong to only one sales rep? Is your sales segmented by region/products?
2. Data Understanding: Does your CRM uniquely identify clients and sales reps? Can you augment it with outside information on buyers?
3. Data Preparation: How to best represent the data for statistical modeling?
4. Modeling: What is the statistical model of normal/unusual behavior?

Problem

Who are your best clients? Do some sales reps have particularly strong links to some of your buyers?

CRISP-DM

1. Business Understanding: Does each account belong to only one sales rep? Is your sales segmented by region/products?
2. Data Understanding: Does your CRM uniquely identify clients and sales reps? Can you augment it with outside information on buyers?
3. Data Preparation: How to best represent the data for statistical modeling?
4. Modeling: What is the statistical model of normal/unusual behavior?
5. Evaluation: Evaluate model fit and performance.
6. Deployment: What actions will you take? Flag strong links? Follow up with clients, sales reps? Give additional incentives?

Data Preparation

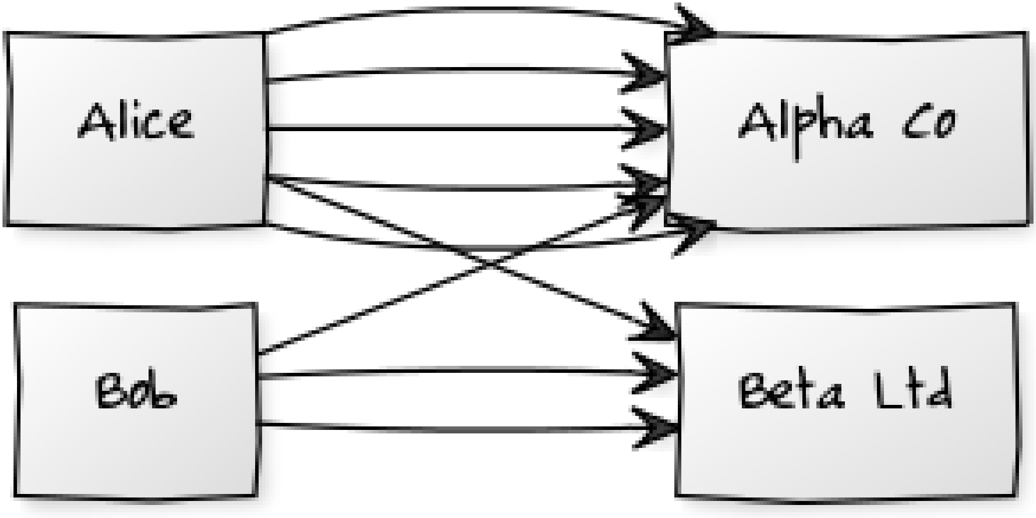
SQL

```
SELECT sales_rep, buyer, COUNT(contract) AS number_of_contracts FROM
crm_table GROUP BY sales_rep, buyer;
```

Table representation

sales_rep	buyer	number_of_contracts
Alice	Alpha Co	5
Alice	Beta Ltd	1
Bob	Alpha Co	1
Bob	Beta Ltd	2

Network representation



Matrix representation

	Alpha Co	Beta Ltd
Alice	5	1
Bob	1	2

Modeling: What is usual?

- We have to model usual link formation.

Modeling: What is usual?

- We have to model usual link formation.
- This relates to the more general problem of "unsupervised fraud detection."

Modeling: What is usual?

- We have to model usual link formation.
- This relates to the more general problem of "unsupervised fraud detection."
- Challenges:
  - "usual" depends on context
  - changing behavior

Modeling: What is usual?

- We have to model usual link formation.
- This relates to the more general problem of “unsupervised fraud detection.”
- Challenges:
  - “usual” depends on context
  - changing behavior
- With normative model of “usual,” we can quantify value lost to unusual links.

The buyer-supplier sudoku

	Alpha Co	Beta Ltd	total
Alice			6
Bob			3
total	6	3	9

The buyer-supplier sudoku (one solution)

	Alpha Co	Beta Ltd	total
Alice	6		6
Bob		3	3
total	6	3	9

The buyer-supplier sudoku (another solution)

	Alpha Co	Beta Ltd	total
Alice	4	2	6
Bob	2	1	3
total	6	3	9

Statistical model: balls falling into bins (Armenter and Koren, 2013)

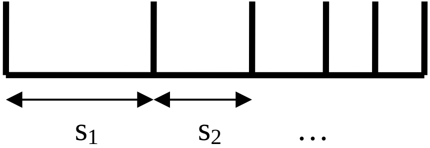


Fig. 1A

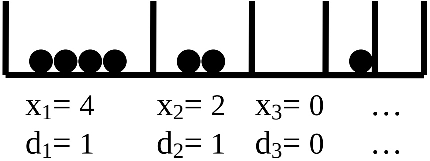
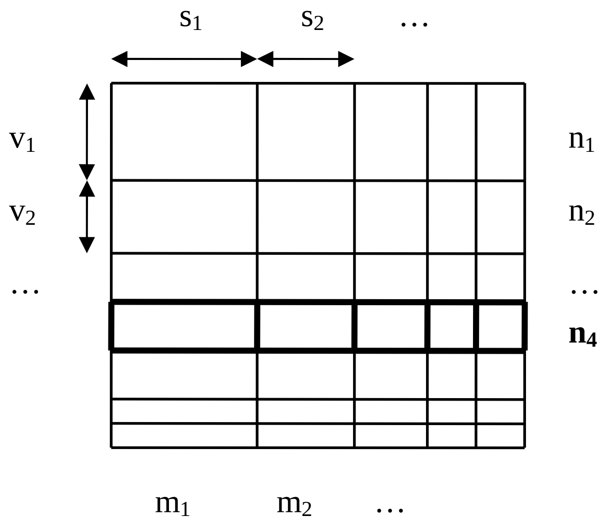
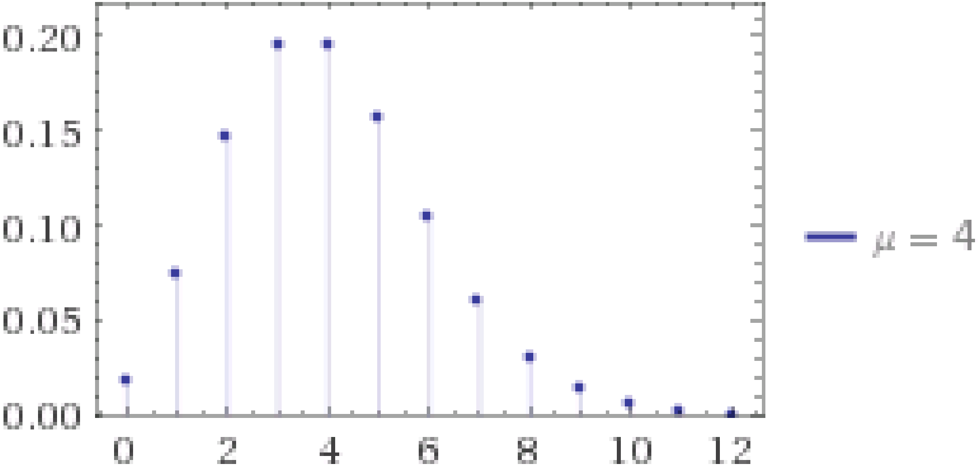


Fig. 1B

“Normal” is governed by bin size (Armenter and Koren, 2013)



Under this model, Alice could expect 3-4 contracts from Alpha Co.



Lessons learned

- Data representation varies by need.
- Even under no preferential association, we expect wide variation in number of contracts.
- Statistical model can predict usual association, deviation can capture special preference.

Questions?

Part 3: Application to Hungarian procurement contracts

Hungarian procurement contracts

- We study about 200,000 public procurement contracts of Hungarian governments 1997-2014.
- Buyer: procuring agency (e.g., municipal government, school)
- Supplier: firm winning conract
- Additional controls: product being procured (Common Procurement Vocabulary), industry of winner, distance, size measures

But first we have to go from this...

Győr Megyei Jogú Város Polgármesteri Hivatal tájékoztatója az eljárás eredményéről (1123)

1. Az ajánlatkérő neve és címe:
Győr Megyei Jogú Város Polgármesteri Hivatal
9021 Győr, Városház tér 1.
Telefon: 96/500-115
Fax: 96/442-650
nevében és megbízásából:
Győr-Moson-Sopron Megyei Beruházási Kft.
9022 Győr, Czuczor G. u. 18/24.
Telefon: 96/317-366
Fax: 96/312-864

2. a) A választott eljárás: hirdetmény közzététele nélküli tárgyalásos eljárás.
b) Hirdetmény közzététele nélküli tárgyalásos eljárás esetén az eljárás alkalmazásának indokolása: a VII. Kajak-kenu maraton vb időpontja előre kijelölt, és ezen változtatni nem lehet [a Kbt. 70. § d) pontja alapján].

3. Az elbírálás időpontja: 1999. február 15.

4. Az elbírálás szempontjai: az ajánlatkérő az összességében legelőnyösebb ajánlatot választja. Az elbírálás szempontjainak fontossági sorrendje:
vállalási ár,
azonos jellegű vagy hasonló körülmények között végzett referenciák bemutatása,
kivitelezés határideje.

5. A nyertes ajánlattevő neve és címe:
Alapozás és szerkezetépítés:
Horváth és Társa Építő Kft., 9026 Győr, Dózsa rkp. 7.
Acélszerkezetek és lakatosmunka:
Integrál-H Acélszerkezeti Kft., 9028 Győr, Fehérvári út 75.
Kőműves-, szakipari munka, szerelt épületszerkezetek:
Mélyépítő Budapest Építőipari, Kivitelező Kft., 1144 Budapest, Füredi út 74/76.
Tetőfedő és bádogos munka:
ÉPSZITEC Kft., 9026 Győr, Zemplén u. 38.
Villanyszerelés:
Vill-Korr Bt., 9029 Győr, Boglárka u. 3739.
Víz-, gáz-, csatorna- és fűtésszerelés:
Integrál-Hexa Rt., 9022 Győr, Bajcsy-Zs. u. 68.
Nyílászárók:
Ablakcentrum Kft., 9025 Győr, Kossuth L. u. 121.

trough this...


```xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="/templates/tender_detail.xsl"?>

<tender requestor_name="Győr Megyei Jogú Város Önkormányzata Polgármesteri Hivatala" source_url="http://regi.kozbeszerzes.hu/static/KEarchiv/9914/1123.html" url="/tender/1999/14/1123.xml" subject="Céltorony építése: három szerkezeti egységből áll. Egy: alapozás és monolit vasbeton szerkezet Kettő: könnyűszerkezetű torony Három: a céltornyot és a bővítendő Vízitelepet összekötő vasbeton födém, rajta könnyűszerkezetű felépítmény, külön megközelítést biztosító acéllépcsővel. Az épület fszt.-i alapterülete 26 m2, magasság +18,94 és -4,09 Vízitelep bővítése: a meglévő lapos tetős épületre emelet-ráépítés könnyűszerkezettel, lapos tetővel, mintegy 200 m2 hasznosalapterület-növeléssel. Munkarészek: 1. Alapozás és vasbeton szerkezetek 2. Acélszerkezetek 3. Kőműves munkák 4. Szakipari munkák 5. Szerelt épületszerkezetek, belső nyílászárók 6. Tetőfedő-bádogos munkák 7. Villanyszerelés 8. Víz-, gáz-, csatorna- és fűtésszerelés 9. Nyílászáró szerkezetek 10. Lakatos munkák" year="1999" estimated_value="" requestor_city="Győr" id="1999_14_1123">
        <procedureType>hirdetmény közzététele nélküli tárgyalásos eljárás</procedureType>
        <estimatedValue></estimatedValue>
        <press></press>
        <requestor id="p_466004" name="Győr Megyei Jogú Város Önkormányzata Polgármesteri Hivatala" city="Győr" url="/entity/p/466004.xml">
</requestor>
        <parts>
                <part id="1999_14_1123_1">
                        <allBidders>
                                <bidder id="t_11491107" name="SZTANA-PLUSZ KFFT" city="Nyíregyháza" url="/entity/t/11491107.xml">
                                        <bestGuessValue>2080328</bestGuessValue>
                                </bidder>
                                <bidder id="t_10304563" name='ABLAKCENTRUM Kft. "f.a."' city="Győr" url="/entity/t/10304563.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11139681" name='INTEGRÁL-H Acélszerkezeti Kft. "f.a."' city="Győr" url="/entity/t/11139681.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_10481246" name="ReCons Kft." city="Szeged" url="/entity/t/10481246.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11150635" name="HAJDUFERR Fémipari Korlátolt Felelősségű Társaság" city="Debrecen" url="/entity/t/11150635.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_10238842" name='INDUKCIÓ Kft. "f. a."' city="Győr" url="/entity/t/10238842.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="u_a054240657747db1aad24c7ca15b7b87" name="Stukkó Építőipari Kft." city="Győr" url="/entity/u/a054240657747db1aad24c7ca15b7b87.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11611178" name='MITRING Kft. "f. a ."' city="Győr" url="/entity/t/11611178.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="u_1fa497100f42526ff7c1510362cab789" name="Mélyépítő Budapest Építőipari, Kivitelező Kft." city="Budapest" url="/entity/u/1fa497100f42526ff7c1510362cab789.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_22409153" name="VILL-KORR HUNGÁRIA Bt." city="Győr" url="/entity/t/22409153.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11125891" name="INTEGRÁL-HEXA Zrt." city="Győr" url="/entity/t/11125891.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11137799" name="ÉPSZITEC Kft." city="Győr" url="/entity/t/11137799.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                                <bidder id="t_11469830" name="WEST HUNGÁRIA BAU Kft." city="Biatorbágy" url="/entity/t/11469830.xml">
                                        <bestGuessValue></bestGuessValue>
                                </bidder>
                        </allBidders>
                        <currency>HUF</currency>
                        <decisionDate>1999/02/15</decisionDate>
                        <estimatedValue></estimatedValue>
                        <numberOfBidder>14</numberOfBidder>
                        <partId>1</partId>
                        <runnerUpValue></runnerUpValue>
                        <subject>Céltorony építése: három szerkezeti egységből áll. Egy: alapozás és monolit vasbeton szerkezet Kettő: könnyűszerkezetű torony Három: a céltornyot és a bővítendő Vízitelepet összekötő vasbeton födém, rajta könnyűszerkezetű felépítmény, külön megközelítést biztosító acéllépcsővel. Az épület fszt.-i alapterülete 26 m2, magasság +18,94 és -4,09 Vízitelep bővítése: a meglévő lapos tetős épületre emelet-ráépítés könnyűszerkezettel, lapos tetővel, mintegy 200 m2 hasznosalapterület-növeléssel. Munkarészek: 1. Alapozás és vasbeton szerkezetek 2. Acélszerkezetek 3. Kőműves munkák 4. Szakipari munkák 5. Szerelt épületszerkezetek, belső nyílászárók 6. Tetőfedő-bádogos munkák 7. Villanyszerelés 8. Víz-, gáz-, csatorna- és fűtésszerelés 9. Nyílászáró szerkezetek 10. Lakatos munkák</subject>
                        <value>2080328</value>
                        <vat></vat>
                        <winner id="t_11491107" name="SZTANA-PLUSZ KFFT" city="Nyíregyháza" url="/entity/t/11491107.xml">
                                <bestGuessValue>2080328</bestGuessValue>
                        </winner>
                        <best_guess_value>2080328</best_guess_value>
                </part>
```

to this.

KOZBESZERZES.CEU.HU

Mi ez? | Hogyan használd?

Keresés

Céltorony építése: három szerkezeti egységből áll. Egy: alapozás és monolit vasbeton szerkezet Kettő: könnyűszerkezetű torony Három: a céltornyot és a bővítendő Vízitelepet összekötő vasbeton födém, rajta könnyűszerkezetű felépítmény, külön megközelítést biztosító acéllépcsővel. Az épület fszt.-i alapterülete 26 m2, magasság +18,94 és -4,09 Vízitelep bővítése: a meglévő lapos tetős épületre emelet-ráépítés könnyűszerkezettel, lapos tetővel, mintegy 200 m2 hasznosalapterület-növeléssel. Munkarészek: 1. Alapozás és vasbeton szerkezetek 2. Acélszerkezetek 3. Kőműves munkák 4. Szakipari munkák 5. Szerelt épületszerkezetek, belső nyílászárók 6. Tetőfedő-bádogos munkák 7. Villanyszerelés 8. Víz-, gáz-, csatorna- és fűtésszerelés 9. Nyílászáró szerkezetek 10. Lakatos munkák

Kiíró: **Győr Megyei Jogú Város Önkormányzata Polgármesteri Hivatala (Győr)**
A kiírás időpontja:

Az eredményhirdetés a kozbeszerzes.hu-n.

Jelezd, ha ez a dokumentum hibás!

1999_14_1123_1

A tenderrész összege: **2 080 328 Ft**
Az eredményhirdetés időpontja: **1999. február 15.**

Ajánlattevők:

- SZTANA-PLUSZ KFFT (Nyíregyháza)
- ABLAKCENTRUM Kft. "f.a." (Győr)
- INTEGRÁL-H Acélszerkezeti Kft. "f.a." (Győr)
- ReCons Kft. (Szeged)
- HAJDUFERR Fémipari Korlátolt Felelősségű Társaság (Debrecen)
- INDUKCIÓ Kft. "f. a." (Győr)
- Stukkó Építőipari Kft. (Győr)
- MITRING Kft. "f. a ." (Győr)
- Mélyépítő Budapest Építőipari, Kivitelező Kft. (Budapest)
- VILL-KORR HUNGÁRIA Bt. (Győr)
- INTEGRÁL-HEXA Zrt. (Győr)
- ÉPSZITEC Kft. (Győr)
- WEST HUNGÁRIA BAU Kft. (Biatorbágy)

Nyertes ajánlattevők:

- SZTANA-PLUSZ KFFT (Nyíregyháza): 2 080 328 Ft

Data cleaning

- Identify buying agencies, contract winners (Entity Resolution).
- Remove duplicates.
- Parse value of contract. (Not used today.)

Sample

- Contracts in construction, manufacturing and business services.
- 500 largest buying agencies.
- 1,000 largest winners.
- 19,500 contracts.

Sample

- Contracts in construction, manufacturing and business services.
- 500 largest buying agencies.
- 1,000 largest winners.
- 19,500 contracts.
- Sparse buyer-supplier network: only 1% of 500,000 potential links exist.

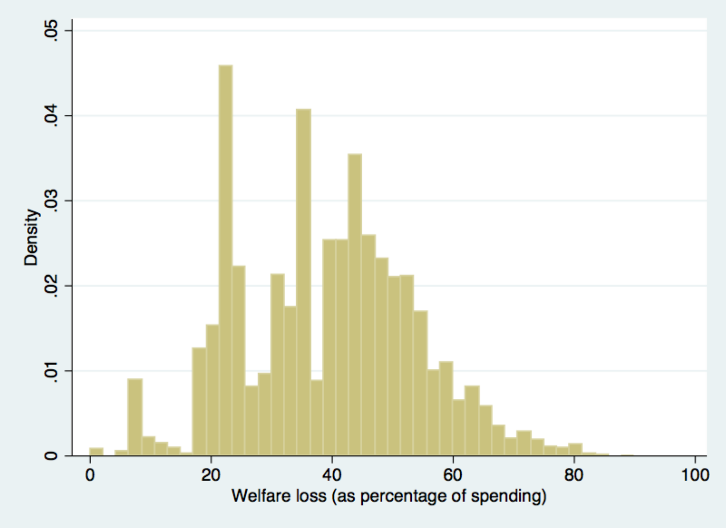
Modeling

- Number of contracts won by i from j follows a Poisson distribution:

$$E(n_{ij}) = \lambda_{ij}$$

$$\lambda_{ij} = F(\text{size}, \text{distance}, \text{purchased product}, \text{political affiliations}, \text{foreign ownership})$$

Many agencies spend very inefficiently



Lessons learned

- Method can be easily applied to transactional data of 200,000 contracts.
- The buyer-supplier network is very sparse.
- Losses from misallocating contracts are substantial.

Questions?