Detecting fraud and outliers using R

Tomaž Kaštrun

BUDAPEST BIFORUM
October 25-27, 2016



About



- BI Developer and data analyst (SQL Server, SAS, R, Python, C#, SAP, SPSS)
- 15 years experience with MSSQL Server
- 15 years experience data analysis and DM
- Working: Spar ICS Austria, Spar Slovenija
- MVP, MCT, MCPT, MCSE SQL Server
- M in tomaz.kastrun@gmail.com
- y @tomaz_tsql
- https://tomaztsql.wordpress.com
- Frequent community speaker at SQL and Microsoft events
- Blogger, Avid Coffee drinker, Bicycle junkie







My VISA account transaction overview

Moje kartice - arhiv prometa

Znesek prometa v EUR	Znesek v originalni valuti	Prodajno mesto	Datum nakupa
50,00		WHO INTERNET	21.09.2016
100,00		WHO INTERNET	21.09.2016
50,00		WHO INTERNET	21.09.2016
1,88	2,09 USD	WWW.ALIEXPRESS.COM	04.10.2016
Skupaj v breme: 201,88			
Skupaj v dobro: 0,00			
Za plačilo: 201,88			

Noticed on 18.10.2016

Želim pregled prometa za kartico:

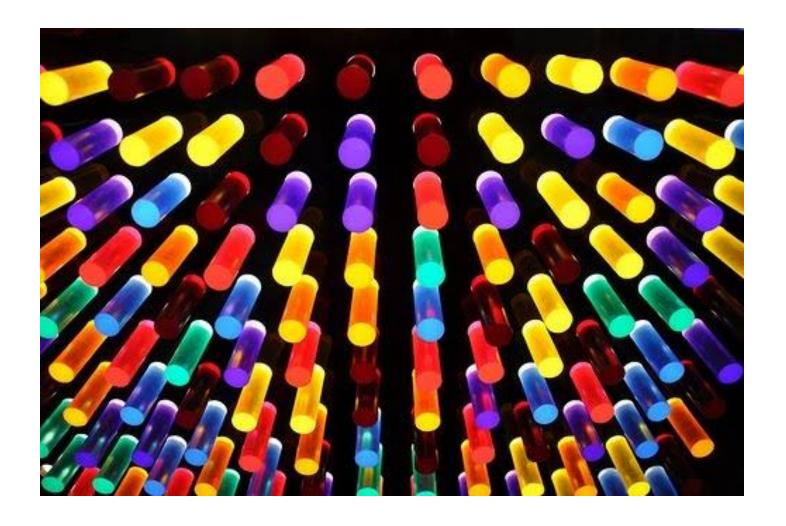
Visa 4106 6900 5046 4190 (Tomaž Kaštrun) ~

Mesec: 18.10.2016 ~





Finding Patterns







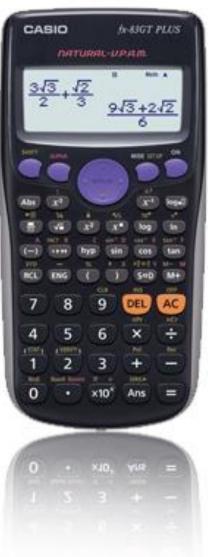
Fitting right values







Applying mathematics









File the "report"





Pritožba imetnika plačilne oz. BA/Maestro kartice/zahtevek za fotokopijo potrdila o nakupu (slip)

Pollijte na nastov: Sektur sprantijana postovanija s flativinist esebar Oddelnja se kantična si betekantiritio premjavo Polec, OJ 400 54, 31 pa kajatna austra Polec, OJ 400 54, 31 pa kajatna austra Polec, OJ 400 54, 32 pa pravne oseba Stativinista SOJ

Obr.: PLK-20. (junij 2012). Nova Ljubljanska banka d.d., Ljubljana. Trg republike 2. 1000 Ljubljana

datki imetnika plačilne kartice e in priimėk. DMAZ (ASTRUM) Davona številka ralov. MAROUTOVA Y 1000 WUBWANM Tel: 0 40 474 514 priika kartice YUBK 6 900 50 46 4000 ANDERNET 21.09.304 50.00 6 1 ATERNET 21.09.304 WHO INTERNET 21.09.304 50.00 6 1 ATERNET 21.09.304 WHO INTERNET 21.09.304 50.00 6 1 ATERNET 21.09.304 WHO INTERNET 21.09.304 50.00 6 1 ATERNET 21.09.304 Tourn merces Tourn merces Tourn merce	1620 Ljubljana				
prolita kartice ### AROUTOVA 4 106 5000 50 46 4100 #### AROUTOVA 4 106 5000 50 46 4100 ##################################	datki imetnika plačilne karti	co			
povika kartice	e in priimek	142 HASTRU	ν	Davčna številka	
In the provided provided potential of a production of the provided	slov MA	ROUTOVA 4,	1000 WUBWA	NA Tel: 040 6	74 519
WHO INTERNET 21.93.200 270 200 6 NTERNET 21.03.200 Color relega 21.0	evilka kartice	4186	6900 51	046 4190	
Zgoraj navedenet transakcije nisem opravli jaz, niti nisem nikogar pooblastit, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne	klamacija nakupa/storitve				
Zgoraj navedenet transakcije nisem opravli jaz, niti nisem nikogar pooblastit, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne	WHO INTERNE	T 21.09.201	€ 50.00€	INTERNET	21.09.201
Zgoraj navedenet transakcije nisem opravli jaz, niti nisem nikogar pooblastit, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne	WHO INDERNE	21.09.20	16 50.006	NTERNET	21.09.201
Zgoraj navedenet transakcije nisem opravli jaz, niti nisem nikogar pooblastit, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne	WHO INTERNE	T 21.09.20	(100,00E	IN TERNET	U. O. Vo
Zgoraj navedonet transakcije nisem opravi jaz, niti nisem nikogar pooblastil, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne	outorns receits	Optiver notional	Znesek notkupa	Krai nakupa	Datum bremenilve
Zgoraj navedonsi transakcije nisem opravil jaz, niti nisem nikogar pooblastit, da opravi nakup namesto mene. Kartice nisem izgubil, niti mi ni bila ukradena, ves čas sem jo imel pri sebi. Kartico sem biokiral dne 18.40.2046 Na zgoraj navedonem prodajnem mestu sem opravil nakup v višini dne nisem opravil jaz. Kartice nisem izgubil, niti mi ni bila ukrađena, ves čas sem jo imel pri sebi. Zgoraj navedeno transakcijo sem na prodajnem mestu opravil samo enkrat. Prodajno mesto je sprovedlo potrdilo o odobritvi (kreditni slip), vendar do danes za zgoraj omenjeni znesek meja kartica še ni bila odobrena. Prilagam dokazilo. Prodajno mesto je sprovedlo potrdilo o odobritvi (kreditni slip), vendar je bila moja kartica bremenjena zanesek meja kartica še ni bila odobrena. Prilagam dokazilo. Moja kartica je bila obremenjena za znesek višini. V prilogi vem pošilijam fotokopijo računa oziroma potrdila o nakupu. Sporna transakcija je bila plačana z drugim pišcilnim sredetvom, in slor. "Okaja kartica je bila plačana z drugim pišcilnim sredetvom, in slor. "Okaja kartica je bila plačana z drugim pišcilnim sredetvom, in slor. "Denaj navedene transakcija je bila plačana z drugim pišcilnim sredetvom, in slor. "Prilagam dokazilo, da sem zadove poskušal na prodajnem mestu uredili sam, ter dokazilo o plačilu. Poraja navedene transakcija en priznam, saj na tem prodajnem mestu risem opravil nakupa. Dne sem kartico preklical zaradi (ustrezno obkroži): Notelsko rezervacijo sem pisno / telefonsko proklical dne Preklicna številka je Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu uredili sam, ter dokazilo o plačilu. Pri rezervacijo sem pisno / telefonsko proklical dne vendili sam, ter dokazilo o plačilu. Pri rezervacijo sem pisno / telefonsko proklical dne dokazilo o plačilu. Pri rezervacije nene pisno / telefonsko proklical dne odpovem rezervacije, nisem bil obvešćen. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu uredili sam, ter dokazilo o plačilu. Nikoli disem uporabili storitev v ravedenem hot	,		acres on - so reque		
ukradena, vee čas sem jo imol pri sebi. Kariico sem biokiral dne	,	•			
No zgoraj navedenom prodajnem mestu sem opravil nakup v višini	Zgoraj navedene transakcije r	nisem opravil jaz, niti nisem ni			nlsem izgubil, niti mi ni bila
nisem opravil jaz. Kartice nisem izgubil, niti mi ni bila ukrađena, ves čas sem jo imel pri sebi. Zgoraj navedeno transakcijo sem na prodajnem mesku opravil samo enkrat. Prodajno mesto je sprovedlo potrdilo o odobritvi (kreditni slip), vendar do danes za zgoraj omenjeni znesek moja kartica še ni bila odobrena. Prilagam dokazilo. Prodajno mesto je sprovedlo potrdilo o odobritvi (kreditni slip), vendar je bila moja kartica bremenjena namesto odobrena. Prilagam dokazilo. Moja kartica je bila obremenjena za znesek v višini. V prilogi vem pošiljam fotokopijo računa oziroma potrdila o nakupu. Sporna transakcija je bila plačana z drugim plačilnim sredetvom, in sicer	ukradena, ves čas sem jo ime	l pri sebi. Kartico sem biokiral	dne_ 18.10.20	16	
Do mel pri sebi. Zgoraj navedeno transakcijo sem na prodajnem mestu opravil samo enkrat. Prodajno masto je sprovedlo potrdio o odobritvi (kreditni slip), vendar do danes za zgoraj omenjeni znesek moja kartica še ni bila odobrena. Prodajno masto je sprovedlo potrdio o odobritvi (kreditni slip), vendar je bila moja kartica bremenjena za nasek vrišini. Prodajno masto je sprovedlo potrdio o odobritvi (kreditni slip), vendar je bila moja kartica bremenjena za nasek vrišini. Naja kartica je bila obremenjena za znesek vrišini. Anmesto. V prilogi vem pošlijami fotokopijo računa oziroma potrdišio nakupu. Sporna transakcija je bila plačana z drugim plačihim sredetvom, in sicer. Prilagam dokazilo, da sem zadove poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Zgoraj navedene transakcija en priznam, saj na tem prodajnem mestu nisem opravil nakupa. Dne (ustrezno obkroži): a) kraje. b) ugube. c) ziorabe. d) poneverbe. Netiske storitve Hotelsko rezervacijo sem pisno / telefonsko proklical dne Preklicna števlika je. Preklicna števlika je. Preklicna števlike nisem prejel. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Pri rezervacijo sem pisno / telefonsko proklical dne pre sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Pri rezervacije, nisem bil obvešćen. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Nikoli disem uporabil storitev v navedenam hotelu, prav tako nisem opravil nobene hotelske rezervacije.	Na zgoraj navedenem prodajr	nem mestu sem opravil nakup	v višini	dne	
Dimel pri sebi. Zgoraj navedeno transakcijo sem na prodajnem mostu opravil samo enkrat. Prodajno mosto je sprovedlo potrdilo o odobritvi (kredilni slip), vendar do danes za zgoraj omenjeni znesek moja kartica še ni bila odobrena. Prilagam dokazilo. Prodajno mosto je sprovedlo potrdilo e odobritvi (kredilni slip), vendar je bila moja kartica bremenjena namesto odobrena. Prilagam dokazilo. Nicja kartica je bila obremenjena za znesek vvišini	spornega nakupa v višini	/	nisem opravil iaz.	Kartice nisem izgubil, niti mi ni i	bila ukradena, ves čas sem
Prodajno mesto je sprovedlo potrdilo o odobritvi (krediini slip), vendar do danes za zgoraj omenjeni znesek moja kartica še ni bila odobrena. Prilagam dokazilo. Prodajno mesto je sprovedlo potrdilo o odobritvi (krediini slip), vendar je bila moja kartica bremerijena namesto odobrena. Prilagam dokazilo. Moja kartica je bila obremerijena za znesek v višini. V prilogi vem posilijami lotokopijo računa ozirona potrdila o nakupu. Sgorna transakcija je bila plačana z drugim plačinim sredetvom, in sicer			, ,		
Pritagam dokazilo. Prodajno mesto je sprovedio potrdio o odobrkvi (kroditni slip), vendar je bila moja kartica bremerjena namesto odobrena. Prilagam dokazilo. Moja kartica je bila obremerjena za znesek v višini. V prilogi vem pošiljam fotokopijo ražuna oziroma potrdila o nakupu. Sporna transakcija je bila plačana z drugim plačitnim sredetvom, in sicer	Zgoraj navedeno transakcijo	sem na prodajnem mestu opr	avil samo enkrat.		
Moja kartica je bila obremenjena za znesek v višini		potrdilo o odobritvi (kreditni sli	p), vendar do danes za zg	oraj omenjeni znesek moja ka	rtica še ni bila odobrena.
V prłogi vem poślijam fotokopijo računa ceżroma potrdila o nakupu. Sporna trensakcija je bila plaćana z drugim plačilnim sredetvom, in sicor	Prodajno mesto je sprovedlo j	cotrdiia e adobritvi (kreditni sli	p), vendar je bila moja kar	tica bremenjena namesto odob	orena. Prilagam dokazilo.
dokazilo, da sem zadovo poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Zporaj navudene transakcije ne priznam, saj na tem prodajnem mestu risem opravil nakupa. Dne			upu.	namesto	
Zgoraj navedene transakcije ne priznam, saj na tem prodajnem mestu nisem opravil nakupa. Dnesem karlioo preklical zaradi (ustrezno obkroži): a) kraje. b) izgube, c) zlorabe, d) poneverbe. Nelske storitve Hotelsko rezervacijo sem pisno / telefonsko preklical dnePreklicna števlika jePreklicna števlika jePreklicna števlika jePreklicna števlika je					Prilagam
(ustrezno obkroži): a) kraje, b) izgube, c) zlorabe, d) poneverbe. Nelske storitve Hotelsko rezervacijo sem pisno / telafonsko proklical dne Preklicna številka je Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačiliu. Hotelsko rezervacijo sem pisno / telafonsko preklical dna, vendar preklicne številke nisem prejel. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Pri rezervacije i hotela sem kot garacijo za rezervacijo navedel številko kartice. O tem, da mi hotel zaračuna stroške nočitve za en dan, če ne odpovem rezervacije, nisem bil obvešćen. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Nikoli nisem uporabil storitev v navedenem hoteku, prav tako nisem opravil nobene hotelske rezervacije.					
Hotelsko rezervacijo sem pisno / felsfonsko prektical dne Prekticna številka je Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Hotelsko rezervacijo sem pisno / telefonsko prektical dna venden prekticne številke nisem prejel. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Pri rezervacija i hotela sem kot grancijo za rezervacijo navedel številko kartico. O tem, da mil hotel zaračuna stroške nočitve za en dan, če ne odpovem rezervacije, nisem bil obvešćen. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Nikoli disem uporabil storilev v navedenem hoteku, prav tako nisem opravil nobene hotelske rezervacije.					sem kartico preklical zarsol
Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Hotelsko rezervacijo sem pisno / telefonsko preklical dna prodajnem mestu urediti sam, ter dokazilo o plačilu. Pri rezervacijo i hotela sem kot garacijo za rezervacijo navedel številko karlice. O tem, da mi hotel zaračuna stročke nočitve za en dan, će ne odpovem rezervacije, nisem bil obvešćen. Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilo o plačilu. Nikoli nisem uporabil storikev v navedenem hoteku, prav tako nisem opravil nobene hotelske rezervacije.	otelske storitve				
sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilio o plačiliu. Pri rezervaciji hotela sem kot garancijo za rezervacijo navedel številko kartice. O tem, da mi hotel zaračuna stroške nočitve za en dan, če ne odpovem rezervacije, nisem bil obveščen. Prilagam dokazilio, da sem zadeve poskušal na prodajnem mestu urediti sam, ter dokazilio o plačiliu, Nikoli nisem uporabil storilev v navedenem hotelu, prav tako nisom opravil nobene hotelske rezervacije.					
odpovem rezervacije, nisem bil obvešćon, Prilagam dokazilo, da sem zadeve poskušal na prodajnem mestu urediti sam, tar dokazilo o plačilu. Nikoli disem uporabil storilev v navedenem holeku, prav tako nisem opravil nobene hotelske rezervacije.				ar prekticne številke nisem pre	jel. Prilagam dokazilo, da
	Pri rezervaciji hotela sem kot odpovem rezervacije, nisem l	garancijo za rezervacijo nave pil obveščen, Prilagam dokazi	del številko kartice. O tem lo, da sem zadeve poskuš	, da mi hotel zaračuna stroške ial na prodajnem mestu urediti	nočítve za en dan, če ne sam, ter dokazilo o plačilu,
Obmite!	Nikoli nisem uporabil storitev	v navedenem hotelu, prav tak	o nisem opravil nobene ho	otelske rezervacije.	
					Obmite!





Bank	romati
- Bar	nkomat mi želenega zneska, za katerega je bila bremenjena moja karlica, ni izplačal.
	lačan je bil le delni znesek dviga na bankomatu v vlšininamesto kolikor je bila moja kartica tudi bremenjena.
Zaht	evek za kopijo potrdila o nakupu (slipa) za zgoraj omenjeno transakcijo (obkrožil):
b) c) d) e)	z bremenjenim zneskom se ne strinjam ne prepoznam transakcije za lastno evidenco zaradi suma zlorabe drugo ji razlogi za zavrnitev s kratkim opisom.
K	REDITNE GRETICE LEDNO NE YOUTBLONM, RAJEN
	U PRIMERU POTOVANS ALI NOKUPOU PREGO AVAZONA,
	MIEXPRESSA. DNE 21.09.2016, NO SO QUE 2500/16
21	OLABNE (FRAUD) TRANSAKCINE, SEM BIL V SLOVENS)/LYDINA
No	TO POUSEY NOVADEN DAN DPORTEWAY USAK DAN MAESTRO
2	A PLACILA. NA TA DAN, 21.05.2016, SEM OPRAVIL NAKUPE

ZA PUTCILY.	NH THE WHO, WISH BOLD, SEN VIETVIC NATIONAL
2 MESTROM	(SPAR, BENCINSKY SERVIS), A NOBENIH SPLETNIH
NAFUPOU, Z	VISA KARTICO.
NAKUPEV (02.	BREMENITUE) NA VISA KARTICI NISEM DERIVIL
DAZ IN SEM	OHAZIL STELE DANES - 18.10.2016 - KO SOMI TREATILITY TREJE ZNESEK V VISTIVI 200€.
Priloge	17 TREJE ZNESEK V USTW1 200€.
	ANA, 18,10,2016 Podpis imetrilis
	Izpolni banka
Enote banke 0,19 Nove	tale tubulance capacific in la shushaman acido
Telefon 01 587 - 48	-88 Faks 01/5217-214 elektronski naslov
ŧ	,

Kraj in datum L, AP. 10. 2010 Zigini poobi





Moje kartice - arhiv prometa



Patterns

Right values

Mathematics

Želim pregled prometa za kartico:

Visa 4106 6900 5046 4190 (Tomaž Kaštrun) ~





Finding what is "strange"

- "Strange" values can:
 - Simply be out of context
 - Do not comply with business rules
 - Reject mathematical and/or statistical rules
- Methods for "finding" values:
 - Semi-graphical methods
 - Uni-/Multi- variate statistical methods
 - Machine learning methods





Building rules

- Statistical rules:
 - Skewness, Kurtosis, Mean, SD and many more
 - Coefficients, Residuals, Loadings
- Procedural rules:
 - Rejecting all "strange"
 - Con-sequential rejection
- Learning rules:
 - AUC, precision
 - Supervized learning methods
 - Refining the models







Rules example

- Statistical rules
 - Flag if out of +/-3SD
 - Flag if spike is at the end of tail
 - Flag if more than 3 transactions (>50€) occur within 5 minutes
- Procedural rules
 - Flag all that come out of "WHO INTERNET" point of Sales
 - Flag if x MOD 10 = 0
 - Flag if Ratio of Max and 2nd highest > x
- Learning rules
 - Flag if $(x \mid y \& z) > 0.4256$ AND (y & x) = "WHO INTERNET"







Stoping "strange" values

- Outliers are:
 - Business-wise defined; with-in business there are many
 - Have no definite answer
 - Statistically/Mathematically it is not always that straightforward
- Start abritrary
 - Check the data
 - Mark the outlier|strange|fraud
 - Build a model based on rules and knowledge-base
 - Learn from it





Data prep's

- Data preparing takes 50% of the time
- Problem with data is low presence of "strange" data
- Sampling data (over or under sampling)
- SCD and RDBM issues (many to many, one to many relations, SCD I, II, III)
- Common sense; Day of week, holidays, Time of the day, Demographics, etc.





DEMO BI FORUM







Machine learning approaches

- Detecting "strange" values with:
 - GLM
 - Naive bayes
 - Random forest
 - Neural Networks
- Other techniques:
 - Density estimation
 - Time series
 - others





Measuring Efficiency and robustness

- Measure model efficiency:
 - Lift Chart
 - AUC
 - Accuracy, Precision, Recall
- Fraud detection <u>is always work in progress</u> and <u>never 100%</u> <u>automated</u> remember: rules change, business change and people "frauding" you also learn (change)





DEMO BI FORUM







Key Take-aways

- Is a work in progress and constant learning cycle
- Combine statistical and mathematical methods for detecting and preventing "strange" observations
- Always refine the models and look for false positives
- Rules are business-wise different between each others
- No definite definition of what Fraud is and no dedicated program.



A&Q

Demo data and R Code:

https://github.com/tomaztk/Detecting outliers and fraud BBI2016



http://tomaztsql.wordpress.com



tomaz.kastrun@gmail.com



@tomaz_tsql



/in/tomaztsql



https://github.com/tomaztk



BUDAPEST BIFORUM
October 25-27, 2016

