





Using Natural Language Processing to Convert Mud-Log Chip Descriptions to Useful Data Tables

Marcelo Guarido, David Emery, and Kris Innanen

Banff, December 2nd, 2022

www.crewes.org







Today's Agenda

Part 2:

Part 3:



Chip Descriptions to Table

Synthetic DTC

Conclusions

election at the end -ad _ob.select= 1 er_ob.select=1 ntext.scene.objects.acti "Selected" + str(modific irror_ob.select = 0 bpy.context.selected_ob ata.objects[one.name].sel

int("please select exactle

ypes.Operator): X mirror to the select ject.mirror_mirror_x" ror X"

context): active_object is not



Part 1

CHIP DESCRIPTIONS TO TABLE

Chip Descriptions in Mud-Logging

- Part of the mud-logging report
- Descriptive mineral analysis per depth interval
- Used by interpreters
- Supporting document
- Usually in PDF files
- Hard to use as data





ConocoPhillips





ConocoPhillips

ConocoPhillips **Cuttings Descriptions Report**

Well Name :	Poseid	on-2 Print Date 8/07/2010		
Wellsite Geolog	ist(s) :	M Boyd M Ortiz S Phillips J Bardelosa M Ortiz M Warring	Iton	
Interval (m)	%	Lithology / Show Descriptions	Ca (%)	Mg (%)
Main				
2429.0 - 2430.0	100	ARGILLACEOUS CALCILUTITE: (Sample from Bit Junk Slot), dark greenish grey, firm, sticky, plastic, 40-50% clay material, strongly calcareous, grading to Marl?.	52	0
2430.0 - 2440.0	70 30	CALCARENITE: yellowish grey, very pale brown to very pale orange, trace off white to very light grey, moderately hard, blocky to sub-blocky, sucrosic in part, occasional fine calcite grains, 2% pyrite nodules, rare very fine disseminated pyrite in part. CEMENT:	78	0
2440.0 - 2450.0	95 5	CALCARENITE: as above, very pale orange to very pale brown, trace to 2% pyrite nodules & forams, 3% chert in part, very fine quartz grains in part. CALCILUTITE: white, off white, firm to generally moderately hard, blocky to subblocky, cryptocrystalline.	79	1
2450.0 - 2460.0	100	CALCARENITE: very pale brown to very pale orange, rare off white to white, moderately hard, sub-blocky to trace blocky, cryptocrystalline when white, occasionally sucrosic, trace pyrite nodules, trace forams and fossil fragments.	84	1
2460.0 - 2470.0	100	CALCARENITE: as above, very fine quartz grain inclusions in part.		
2470.0 - 2480.0	95 5	CALCARENITE: as above, very pale brown to very pale orange, moderately hard, sub-blocky to blocky, occasionally sucrosic, trace pyrite nodules. CALCILUTITE: medium light grey, medium grey, firm to moderately hard, sub-blocky to blocky, slight to moderately argillaceous,		
2480.0 - 2490.0	95	CALCARENITE:	84	1
	5	CALCILUTITE: as above.		
2490.0 - 2500.0	97 3	CALCARENITE: as above, 3% chert. CHERT: very pale to pale orange, occasionally translucent, hard, angular, conchoidal fracture.	87	1
2500.0 - 2510.0	95 5	CALCARENITE: as above, very pale brown to very pale orange, rare off white to white, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 5% chert. CHERT: as above.	87	1
2510.0 - 2520.0	97 5	CALCARENITE: as above. CHERT: as above.	85	1
2520.0 - 2530.0	92 5 3	CALCARENITE: as above, very pale brown to very pale orange, rare off white and pale orange, moderately hard to brittle, sub-blocky to trace blocky, predominant fine fragments, cryptocrystalline to sucrosic, 3-5% chert fragments. CALCILUTITE: medium light grey, medium grey, medium bluish grey, firm, sub-blocky to occasionally blocky, slightly argillaceous,		
2520.0 - 2540.0	00			
2000.0 2040.0	5	CALCILUTITE: as above		
	5	CHERT: as above.		
2540.0 - 2550.0	95	CALCARENITE: as above.	82	1
	5	CHERT: as above.		
2550.0 - 2560.0	98 2	CALCARENITE: very pale brown, yellowish grey, rare white, moderately hard, predominantly fine fragments, occasionally medium, sub-blocky to blocky, nil to trace chert, Claystone and orange translucent fragments. CHERT: as above.		
Copyright 2002 - Petr	oleum Data	a Systems International Pty Ltd	Page	:1 of 3

×1	AutoSave	Off 🗄 chip_descriptio	n_Resampled	d_poseidon-2	• Saved \checkmark			h (Alt+Q)					Marcelo Guai	ido 🝘 🖉 -	- 0	×
File	e Hom	e Insert Draw Page Lay	vout Forr	mulas Data	Review	View Hel	р								Commen	ts 🖻 Share	•
C U	Pas	te Clipboard Γs	ri I <u>U</u> ~	√11 √	A^ A` ~ <u>A</u> ~			Center ~	Ge \$	eneral ✓ ✓ % 9 ←0 .00 Number Is	Conditional Formatting ~	Format as Cell Table ~ Styles ~	Insert Delete Format	∑ AutoSum ~ A Z Z / Fill ~ Sort & Fi ♦ Clear ~ Filter ~ Se Editing	Analyze Data Analysis		~
E60	\ \	$r : \times \checkmark f_x 0$															~
	А	В		С	D	E	F	G		Н		I		J	К	L	
1	Depth (I	n) Argillaceous Calcilu	tite Calc	arenite Ce	ement C	Calcilutite C	alcarenite Calciluti	te Che	ert Ca	alcarenite Chert C	alcilutite A	rgillaceous Ca	alcarenite Argillace	ous Calcarenite Argilla	ceous Calcisiltite	Calcilutite	1
2	24	29	100	0	0	0		0	0	0		_	0	0	0		0
3	2429	.5	100	0	0	0		0	0	0			0	0	0		0
4	24	30	0	70	30	0		0	0	0			0	0	0		0
5	2430	0.5	0	70	30	0		0	0	0			0	0	0		0
6	24	31	0	70	30	0		0	0	0			0	0	0		0
7	2431	5	0	70	30	0		0	0	0			0	0	0		0
8	24	32	0	70	30	0		0	0	0			0	0	0		0
9	2432	5	0	70	30	0		0	0	0			0	0	0		0
10	24	33	0	70	30	0		0	0	0			0	0	0		0
11	2433	.5	0	70	30	0		0	0	0			0	0	0		0
12	24	34	0	70	30	0		0	0	0			0	0	0		0
13	2434	.5	0	70	30	0		0	0	0			0	0	0		0
14	24	35	0	70	30	0		0	0	0			0	0	0		0
15	2435	.5	0	70	30	0		0	0	0			0	0	0		0
16	24	36	0	70	30	0		0	0	0			0	0	0		0
17	2436	i.5	0	70	30	0		0	0	0			0	0	0		0
18	24	37	0	70	30	0		0	0	0			0	0	0		0
19	2437	.5	0	70	30	0		0	0	0			0	0	0		0
20	24	38	0	70	30	0		0	0	0			0	0	0		0
21	2438	9.5	0	70	30	0		0	0	0			0	0	0		0
22	24	39	0	70	30	0		0	0	0			0	0	0		0
23	2439	0.5	0	70	30	0		0	0	0			0	0	0		0
24	24	40	0	95	0	5		0	0	0			0	0	0		0
25	2440	0.5	0	95	0	5		0	0	0			0	0	0		0
- Death	chip	_description_Resampled_po	se 🕂		-	-		-	-	-	:			m	· ه س		120%
Keady	K? Acces	soliity: onavaliable	_	_	_	_	_	_	-	_	_	_		ш.		+	130%



All Rights Reserved







Vell N	lam	ie :	Poseide	on-2
Nellsi	ite	Geolog	ist(s) :	M Boyd M Ortiz S F
In	terv (m)	/al	%	Lithology / Show
ain				
429.0	-	2430.0	100	ARGILLACEOUS CALCILUTITE: (S firm, sticky, plastic, 40-50% clay mate
430.0	-	2440.0	70	CALCARENITE: yellowish grey, very
			20	fine calcite grains, 2% pyrite nodules,
			30	
440.0	-	2450.0	95	CALCARENITE: as above, very pale nodules & forams 3% chert in part ve
			5	CALCILUTITE: white, off white, firm
				subblocky, cryptocrystalline.
450.0	-	2460.0	100	CALCARENITE: very pale brown to moderately hard, sub-blocky to trace I occasionally sucrosic, trace pyrite noc
460.0	-	2470.0	100	CALCARENITE: as above, very fine
470.0	-	2480.0	95	CALCARENITE: as above, very pale
			5	CALCILUTITE: medium light grey, m sub-blocky to blocky, slight to modera
480.0	-	2490.0	95	CALCARENITE:
			5	CALCILUTITE: as above.
490.0	-	2500.0	97	CALCARENITE: as above 3% chert
			3	CHERT: very pale to pale orange, or conchoidal fracture.
500.0	-	2510.0	95	CALCARENITE: as above, very pale
				white, moderately hard to brittle, sub-
			5	CHERT: as above.
510.0	_	2520.0	07	
510.0	-	2020.0	5	CHERT as above
520.0	-	2530.0	92	CALCARENITE: as above, very pale pale orange, moderately hard to brittle fragments, cryptocrystalline to sucrosi
			5	CALCILUTITE: medium light grey, m
			3	CHERT: as above.
530.0	-	2540.0	90	CALCARENITE: as above.
			5	CALCILUTITE: as above.
			5	CHERT: as above.
540.0	-	2550.0	95	CALCARENITE: as above.
			5	CHERT: as above.
550.0	-	2560.0	98	CALCARENITE: very pale brown, yellowish grey, rare white, r predominantly fine fragments, occasionally medium, sub-blocky
			2	chert, Claystone and orange translucent fragments.

Cancel

Click and drag to select an area. Click 'Done' to accept.

Website Geologizatiji M Boydi M Oriz S Philips J Barderica M Oriz M Warrington Vestaria Name Cancol M Name Nam Name Nam	Well I	Kaim	e ;	Poseid	on-2 Print Date B/07/	2010		
No. Linkshopp f Show Descriptions Car (N) Mag (N) Main - Addit_LACCOURD CALLUTTE: disapse from the Jack Show Descriptions 42 0 24008 - 24008 100 Addit_LACCOURD CALLUTTE: disapse from the Jack Show Descriptions 42 0 24008 - 2440.0 100 AddAtement Fig. starshow Descriptions 74 0 24008 - 2440.0 100 CALCAPENTE: starshow Descriptions 74 0 2440.0 - CALCAPENTE: starshow Descriptions in part for description descriptions in part for description descriptions in part for descriptions in part for descriptions in part for description description description description description description	Wells	Re (Geologi	ist(s) :	M Boyd M Ortiz S Phillips J Bardelosa M Ortiz M	Warringto	n	
North North North North North North 24000 2,2000 100 Amilia ACEDOR CACULATER: dispatch to the vary late longing, but of the dispatch to the vary late longing. But of the dispatch to the vary late longing, but of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. But of the dispatch to the vary late longing. The dispatch to the vary late longing longing. The dispatch longing. The dispatch longi	,	itter s	<i></i>	8	Lithology / Show Descriptions		Car(%)	Mg (%)
Market Baseline Justice International and State St		(int)						-
24880 - Desk Product Set Set Set Set Set Set Set Set Set Se	34280		2430-0	NKR	ARGALLACEOUS CALCELETTRE: champs for Br. Amir Bist, dam genater gene ten, node, pleads, 40:505, deg maleral, energy calcements, geneting to Merit.		44	ġ.
Jakebol Josephol The state of the state	243614		29994030	. jtů 50	CALCARENTE: yonowoh grey, vary pasi tenser to vary pasi noinge, taon off wir to vary type grey, readmanany haid, becky to sub-mocky, sucrease in part, occasion the optice greys. 2% pyrite notices, tao very fire deservices of pyrite or part. Createrst:		**	4
Answer Answer Note States & Borney, This states to gate the space to gate to state the stage the space to state to stage the space to state to state the state to state the space to to state sp			1000			sisisisin		a fa fa fa fa fa
38568 - 108 CALCAMENTE: -					incidence & focuers, 25 check in part, very fire query previous per could be & focuers, 25 check in part, very fire query previous report CALCELUTTE: while, off while, firm to generally moderably hard, blocky to could actually constructed firm.			*
34800 - 34710 100 CALCAMENTE: as alone, very pair local quere in classer or my in. Image: classe of the cla	246610	11111	linini di	ideniden Kimi	CALCARENTE: very gain trover to very pain orange, rate of white to white, nederately hard, autorolocky to trace brone, organorystalline when white, consistential autoroloc, trace bride notates, those traces and there they warners.			*
34700 • 24800 0 CALCAMENTE: as above, very task brown to very pak toopag, moderakity hard, add books to tooch, socialisation and sociality to tooch, add to tooch add too Image: social too tooch, add to tooch, add too 34800 • 30000 50 CALCAMENTE: as above, 3% of too 34 7 34800 • 20050 97 CALCAMENTE: as above, 3% of tool 37 7 34800 • 20050 97 CALCAMENTE: as above, 3% of tool 37 7 34800 • 20050 97 CALCAMENTE: as above, 3% of tool 37 7 34800 • 20050 97 CALCAMENTE: as above, 5% of tool 37 7 35800 • 20050 97 CALCAMENTE: as above, 5% of tool 37 7 35800 • 20050 97 CALCAMENTE: as above, 5% of tool 37 7 7 35800 • 20050 97 CALCAMENTE: as above, 5% of tool 37 7 7 35800 <	3460.0	M	21475.05	1945	CALCARENTE: as elenno, very fine spanie grain indusions in puri.	ananananan 		
Jakeso - See CallCommetter: See CallCommetter: See T 2656.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2556.07 - 2557.07 - - 2556.07 - 2557.07 - - 2557.07 -	34700		2489.9	96 3	CALCARENTTE: as stored, very pairs brown to early pairs orange, residenately hant, sub-blocky to blocky, boccisionally audicisis, these pythe hockates. CALCELUTIE: residentially togrey, methon gray, timit to moderately hand, sub-blocky to tocky, store to moderately anglikastrose.			
24562 23552 247 CALCANTERT as above 47 1 24562 - 28525 97 CALCANTENTS: as above 2% creat 47 1 25505 - 28525 95 CALCANTENTS: as above 2% creat 487 1 25505 - 28525 95 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25505 - 28525 97 CALCANTENTS: as above 45 1 25506 - 28585 97 CALCANTENTS: as above 45 1 25506 - 28585 97 CALCANTENTS: as above 45 1 25506 - 28	2480(0	*	2490,0	96	CALCARENTE	*****	84	*
24562 - 28029 97 CALCARENTE: as above 97 1 25509 - 2019 15 CALCARENTE: or above 87 1 25509 - 2019 15 CALCARENTE: or above 85 5 25509 - 2019 97 CALCARENTE: or above 85 5 25509 - 2029 97 CALCARENTE: or above 65 5 25509 - 2029 97 CALCARENTE: or above 65 5 25509 - 2029 97 CALCARENTE: or above 65 5 25509 - 2039 97 CALCARENTE: or above 65 5 25509 - 2039 97 CALCARENTE: or above 65 5 25509 - 2039 67 CALCARENTE: or above 65 5 25509 - 2039 67 CALCARENTE: or above 65 5 25509 -					CALCILI/PITE: all desires			
25500 - 2010 10 CALCARENTE: an aform, oney gate brance, south and angular, that, angular, in the south and angular, the south and angular, the south and angular, the south angular, the southang the southang the south angular, the southa	2466)2	8	29900-0	97	GALCAREN/TE: as above 2's crest		87	*
25500 - 25100 - 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2					CHERT: very pain to gate transpir, docasiantaly translation, next, angular, constructed inscrum.			
25%8.2 - 25	2560-þ		28+4-4	985 5	CALCARENTE: or access, very pare tensors very pare orange, one off write to when inconcessly lined to table, each desire to table, predominant like treprenets, cognizorganities to success, 5% cheet. CHERT: so above.		82	
2550.9 - 2550.9 - <td>25100</td> <td></td> <td>28880-10</td> <td></td> <td>CALCARENTE an exce</td> <td></td> <td></td> <td></td>	25100		28880-10		CALCARENTE an exce			
2550.0. - 2550.0.					CHERTS IN MILLION			
25505 - 2740.0 - 2740.0 - 2740.0 - 1 1 1 25505 - 2750.0 - 2750.0 - 2750.0 - 1 </td <td>2529-0</td> <td></td> <td>2530-6</td> <td>**</td> <td>CALCARENTE: as above, very pade brown to very pade transpe, rate of white an pate stranspe, modestery hard to britle, sub-blocks to trace model, predostrane fire hypreens, cryptocystalizes to success. 3-5% creat dispreses. CALCENTIFE: resultantiple gray, modelan gray, resident toward gray, for which blocks to incoming all blocky, alignly angliteness. CHERE: to show.</td> <td></td> <td></td> <td></td>	2529-0		2530-6	**	CALCARENTE: as above, very pade brown to very pade transpe, rate of white an pate stranspe, modestery hard to britle, sub-blocks to trace model, predostrane fire hypreens, cryptocystalizes to success. 3-5% creat dispreses. CALCENTIFE: resultantiple gray, modelan gray, resident toward gray, for which blocks to incoming all blocky, alignly angliteness. CHERE: to show.			
2560.0 2560.0 2560.0 1 2560.0 2560.0 2560.0 <td>253614</td> <td>1111 #</td> <td>2/542-05</td> <td></td> <td>CALCARENTTE as above</td> <td>ninininini)</td> <td></td> <td></td>	253614	1111 #	2/542-05		CALCARENTTE as above	ninininini)		
2550.5 - 2550.6 46 CALCARENTE: as ablease 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.5 - 2550.6 10 - 2550.6 - 2550.6 10 - 2550.7 - 2550.6 10 - 2550.7 - 2550.6 10 - 2550.7 - 2550.6 10 <td< td=""><td></td><td></td><td></td><td></td><td>CALCHUTTE: sectore.</td><td></td><td></td><td></td></td<>					CALCHUTTE: sectore.			
2560.5 - 250.5 45 CALCARENTE: as above. 47 1 2550.5 - 250.5 48 CALCARENTE: set above. 48 1 2550.5 - 250.5 48 CALCARENTE: set above. 48 1 2550.5 - 250.5 48 CALCARENTE: set above. 48 1 2550.5 - 250.5 48 CALCARENTE: set above. 48 1 2550.5 - 250.5 10 1 7 Very coarse fragme 10 10 1 1 7 CALCARENTE: calculation of the set o				*	CHERT: an above			
2552 * 2002 8 CHERT: we shown 2552 * 2002 1 CALCARENTE: very sele brown, selebrate, grauswerkty mediane, sub-blacky, to technical totechnical to technical to technical to technica	2540.0	de:	1259595-38	465	CALCARENTE: as allows		882	*
2250.5 · 2010.0 vit graduation of the filter integration of the second rest of the moderated phase. Image: Second rest of the filter integration of the fi					CHERT: do aldon			
Bigging 2002 - Petersberre Datas Systems International Phy Ltd Page 1 of 1 82 1 7 CHERT: as above CHERT: as above 2700.0 - 2710.0 95 CALCARENTE: a coarse fragments,	29640-0	M	2000KG (AS	WE X	CALCARENITE: very pairs brown, yellowish goty, rook offile, maximisity famil, predominantly fine fragmente, occlassionally medium, with blocky to Modely, vil to have offert. Chapteries and energy transmission fragments. CHERT: an above.			
Name Name Very coarse fragme 82 1 7 CHERT: as above 2700.0 - 2710.0 95 CALCARENTE: a coarse fragments, i coar	Copyri	givi 21	KiQ - Park	dinorn Data	a Shudatros inversational Pby 1.1d		Pag	0.10/15
82 1 7 CHERT: as above 2700.0 - 2710.0 95 CALCARENTE: a coarse fragments, coarse fragments,	NI RIGH	r Pan	Markied					very coarse fragments
2700.0 - 2710.0 95 CALCARENITE: a coarse fragments,				82	1		7	CHERT: as above.
					2700.0 - 2	2710.0	95	CALCARENITE: as a coarse fragments, 3%
trace 5 CHERT: as above	trace						5	CHERT: as above.
					2710.0 - 2	2720.0	93	CALCARENITE: as



Done			
	Cutt	ings Descrip	tion Report
	criptions	Ca (%)	Mg (%)
	fragments, 30% medium fragments.		
	, translucent, hard, conchoidal fracture, dissemination in part.		
	iragments.	82	1
	n, occasionally very light grey to light		
	grey, translucent.		
	fragments, 20% medium fragments, trace	84	1
	plive grey, translucent.		
	o coarse fragments, rare very coarse		
	grey, rare light brown.	87	1
	re very coarse fragments, sub-blocky, black specks in part.		
	grey, rare light brown, translucent, hard, occasional trace black specks, trace		
	in to white, grading to Calcilutite, 10%		
	nents, 30% medium fragments, 10%		
	nents, 10% medium fragments, 30%		
	grey, translucent.		
	n, 40% fine fragments, 20% medium		
	in to yellowish grey, rare light grey, trace coarse fragments, 40% very coarse a forams, trace pyrite nodules, trace black	87	1
	grey, rare light brown, translucent to , hackly fracture in part, occasional trace part.		
fragments	nents, 40% 30% coarse fragments, 10%	73	0
s above.			
IITE: as above, 62% fine frag ments, 3% very coarse fragmen s above.	ments, 30% medium fragments, 5% nts.		
IITE: as above, very light brow	wn, minor white.		
rnational Pty Ltd		Page	:2 of 32

SANDSTONE: light olive grey to pale yellowish brown, olive grey, translucent, predominantly moderately hard aggregates, predominantly fine to commonly medium, moderately well sorted, sub-rounded to occasionally sub-angular, moderately strong siliceous cement, minor silty matrix in part grading to trace very fine arenaceous SILTSTONE, predominantly fair to good visible porosity, poor porosity in very fine to silty aggregates, no fluorescence. **SILTSTONE:** olive black, commonly grading to olive grey, firm, blocky, weakly calcareous, moderately argillaceous, commonly arenaceous.

 $\textbf{REGEX} \rightarrow \textbf{Regular Expression}$

[A-Z]+:?([A-Z]+[A-Z \d]+:)



Tabulizer

- Select the table on the first page
- Read other pages
- Extract long table

Processing

- Regularize table and merge rows
- Locate minerals with NLP and separate into columns
- Feed the new columns with percentages

Export

- Resample to the desired sample rate
- Export to CSV







Datasets:	Ma	nage	View Visualize	Pivot	E	plore	Transform	Cor	nbine			
ConocoPhillips	Da	ita pr	eview									
Add/edit data description			в	C	D	F	F		G	Н		
Rename data	1 [Depth (m) A	rgillaceous Calcilutite	Calcarenite (Cement	Calcilutite	Calcarenite Calc	ilutite (Chert C	alcarenite Cher	Calcilutite Argillaceous Calcar	renite A
	2	2429	100	0	0	0		0	0	()	0
Display:	3	2429.5	100	0	0	0		0	0	()	0
● preview ○ str ○ summary	4	2430	0	70	30	0		0	0	()	0
	5	2430.5	0	70	30	0		0	0	()	0
	6	2431	0	70	30	0		0	0	()	0
Load data of type:	7	2431.5	0	70	30	0		0	0	()	0
rda Lirda Lirdata —	8	2432	0	70	30	0		0	0	()	0
	9	2432.5	0	70	30	0		0	0	()	0
	10	2433	0	70	30	0		0	0	()	0
Browse No file selected	11	2433.5	0	70	30	0		0	0	()	0
blowse No file selected	12	2434	0	70	30	0		0	0	()	0
	13	2434.5	0	70	30	0		0	0	()	0
	14	2435	0	70	30	0		0	0	()	0
	15	2435.5	0	70	30	0		0	0	()	0
	16	2436	0	70	30	0		0	0	()	0
Save data to type:	17	2436.5	0	70	30	0		0	0	()	0
C01	18	2437	0	70	30	0		0	0	()	0
CSV 👻	19	2437.5	0	70	30	0		0	0	()	0
	20	2438	0	70	30	0		0	0	()	0
Ł Save	21	2438.5	0	70	30	0		0	0	()	0
	22	2439	0	70	30	0		0	0	()	0
	23	2439.5	0	70	30	0		0	0	()	0
C Obarry D and a	24	2440	0	95	0	5		0	0	()	0
Show R-code	25	2440.5	0	95	0	5		0	0	()	0

A dataset containing the prices and other attributes of a sample of 3000 diamonds. The variables are as follows:

Variables

- price = price in US dollars (\$338–\$18,791)
- carat = weight of the diamond (0.2–3.00)
- clarity = a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
- cut = quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- color = diamond color, from J (worst) to D (best)
- depth = total depth percentage = z / mean(x, y) = 2 * z / (x + y) (54.2-70.80)
- table = width of top of diamond relative to widest point (50–69)
- x = length in mm (3.73–9.42)
- y = width in mm (3.71–9.29)
- z = depth in mm (2.33–5.58)
- date = shipment date

Additional information

Diamond search engine





Project: (None)

Argillaceous Calcisilitie Calcilutite Calcilutite 0 0 0 0 </th <th>J</th> <th>К</th> <th>L</th> <th></th>	J	К	L	
0 0	rgillaceous Calcarenite	Argillaceous Calcisiltite	Calcilutite	1
	0	0		0
	0	0		0
	0	0		0
	0	0		0
	0	0		0
0 0 0 0 0 0 <td>0</td> <td>0</td> <td></td> <td>0</td>	0	0		0
0 0 0 0 0 0	0	0		0
0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0	0	0		0
0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0 0 0 0	0	0		0
0 0 0 0 0 0	0	0		0
0 0 0	0	0		0
	0	0		0







Part 2

SYNTHETIC DTC

Synthetic DTC

- Poseidon 3D (Australia)
- Mud-logging from well Poseidon-2
- Chip descriptions + GR + RDEP (LWD)
- De-trend logs
- Standardize the data (Z-Score)
- Predict DTC
 - Linear regression
 - Step-wise regression
 - XGBoost







Synthetic DTC: LWD 40/60 data split



Significance level: 0.05





Hypothesis Testing

Null hypothesis \rightarrow H₀: w = 0 Alternate hypothesis \rightarrow H_a: w \neq 0

P-value (or *probability value*) is the probability of the Null Hypothesis being true.

The p-value is determined from a statistical value of the population distribution.



But what is the population distribution?

Hypothesis Testing



- t distribution
- Used when:
 - Small amount of data
- t-statistic

t =

Where:

- w is the estimated weight ٠
- •



• Population's standard deviation is unknown

$$=\frac{w-w_0}{se(w)}$$

w₀ is the hypothesis weight (equals zero) se(w) is the standard error of the weight

Synthetic DTC: LWD + Chip Descriptions

40/60 data split

Call: lm(formula = DTC_Detrended ~ Depth	ı - Trend_DT	rc – DTC, d	lata = ti	rain)	
Residuals: Min 1Q Median 3Q -24.2906 -2.3174 -0.0601 2.4379 2	Max 2.1395				
Coefficients:	Fetimata (std Ennon	+	Dn(> + +)	
(Intercent)	-0 07621	0 09041	-0.843	0 399360	
Depth (m)	-5.10753	0.43234	-11.814	< 2e-16	* *
	-0.28764	0.42570	-0.5/5	0 100306	
Calcarenite	-2.53034	0.75655	-3.345	0.000838	* *
Cement	0.34663	0.08092	4.284	1.92e-05	* *
Calcilutite	0.06983	0.44667	0.156	0.875787	
Calcarenite Calcilutite	-0.47073	0.14862	-3.167	0.001560	**
Chert	-0.44993	0.16017	-2.809	0.005010	**
Calcarenite Chert	-0.40398	0.15794	-2.558	0.010602	*
Calcilutite Argillaceous Calcarenite	0.05612	0.14630	0.384	0.701306	
`Argillaceous Calcarenite`	-0.27558	0.12137	-2.271	0.023270	*
Argillaceous Calcisiltite	-0.12594	0.08487	-1.404	0.137975	
Calcilutite 1	0.19880	0.11100	1 770	0.075495	
Calcareous Silisione	-0.13957	0.09888	-1.412	0.158227	
Argillaceous Calcilutite 1	0.34530	0.09703	3.559	0.000381	**
Marl	0.68583	0.26453	2.593	0.009587	* *
Sanusione	0.24984	0.27965	0.893	0.371752	
<u>Marl Calcilutite</u>	0.12855	0.13569	0.947	0.343524	
-Calcarcous Claystone	-0.08277	0.35293	-0.235	0.814598	
Claystone	2.97673	0.60937	4.885	1.11e-06	* *
Claystone 1	0.20035	0.17072	1 174	0.240709	
Contamination	0.29400	0.11964	2.457	0.014075	*
Siltstone	-0.24599	0.51090	0.481	0.630218	
Volcanic	0.38974	0.12842	3.035	0.002434	**
No Lithology`	0.30034	0.13603	2.208	0.027349	*
Ferruginous Volcanic	1.10406	0.20897	5.283	1.39e-07	* *
Argillaccous Volcanic	0.30196	0.3040/	0.993	0.320/85	
<u>Siltstone 1</u>	-0.08152	-0.11029	0.739	0.459887	
`Siltstone Siltstone 1`	0.14090	0.06590	2.138	0.032627	*
`Argillaceous Siltstone`	0.31283	0.13287	2.354	0.018638	*
`Ca (%)`	-1.68043	0.67561	-2.487	0.012945	*
GR	2.27713	0.13907	16.374	< 2e-16	**
RDEP	-4.72598	0.15345	-30.799	< 2e-16	**





Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.281 on 2240 degrees of freedom Multiple R-squared: 0.8235, Adjusted R-squared: 0.821 F-statistic: 326.7 on 32 and 2240 DF, p-value: < 2.2e-16





Synthetic DTC: Step-wise regression

40/60 data split

Step-wise Model Prediction

Call: lm(formula = DTC_Detrended ~ Chert + Calcilutite + Man Depth (m) + Ca (%) +	RDEP + Cl rl + Cemer Calcareni	aystone + GR + Siltstone - it + `Ferruginous Volcanic te + `Argillaceous Calcil	t tite_1	200 -	.metric.est <chr><chr< td="">rmsestarrsqstar</chr<></chr>	imator > hdard hdard
Volcanic + Calcarenite (No Lithology` + `Argilla `Siltstone Siltstone 1` data = train)	aceous Ca + `Calci	Step	Of states	Deviance <dbl></dbl>	Resid. Df <dbl></dbl>	Resid. Dev <dbl></dbl>
Residuals: Min 10 Median -23.7128 -2.3492 -0.0624	3Q 2.4954	+ RDEP + Claystone	NA -1 -1	NA 164201.15914 11905 59089	2272 2271 2270	232626.37 68425.21 56519.62
Coefficients:	Estimat€	+ GR	-1	3638.88705	2269	52880.73
(Intercept) RDEP Claystone	-0.06767 -4.63892	+ Siltstone + Chert	-1 -1	4409.50446 1948.92161	2268 2267	48471.22 46522.30
GR Siltstone	2.26257 -0.61817	+ Calcilutite + Marl	-1 -1	433.89126 411.73326	2266 2265	46088.41 45676.68
Chert Calcilutite Marl	-0.35745 0.29630 0.73528	+ Cement	-1	404.94381	2264	45271.73
Cement `Ferruginous Volcanic` `Denth (m)`	0.35096 1.05934	+ Ferruginous voicanic + `Depth (m)`	-1	370.41554	2263	44864.55
Ca (%) Calcarenite	-2.69692	+ `Ca (%)` + Calcarenite	-1 -1	1190.78662 814.37080	2261 2260	43303.35 42488.98
Volcanic Calcarenite Calcilutite	0.28086 0.319374 -0.36713	- `Argillaceous Calcilutite 1`	-1	222.65322	2259	42266.33
Contamination `Calcarenite Chert` `No Lithology`	0.23988	+ `Calcarenite Calcilutite`	-1	173.82514	2258	42080.39
Argillaceous Calcarenite` `Argillaceous Siltstone`	-0.23653	+ Contamination + `Calcarenite Chert`	-1 -1	107.77090 110.49629	2256 2255	41804.99 41694.50
Calcilutite 1 Argillaceous Calcisiltite	0.21857	+ `No Lithology`	-1	99.43980	2254	41595.06
 Signif. codes: 0 '***' 0.001	1'**'0.	+ `Argillaceous Siltstone`	-1	82.81256	2253	41415.90
Residual standard error: 4.28 Multiple R-squared: 0.8229,	81 on 224 Adjus	+ `Siltstone Siltstone 1` + `Calcilutite 1`	-1 -1	85.12202 82.97880	2251 2250	41330.77 41247.80
F-Statistic: 454.2 off 25 and	2249 DF,	+ `Argillaceous Calcisiltite`	-1	39.81617	2249	41207.98





Synthetic DTC: XGBoost

40/60 data split



XGBoost Model Prediction





Synthetic DTC: XGBoost

5/95 data split



XGBoost Model Prediction

200











Part 3

CONCLUSIONS

Conclusions

PDF to Table

- Mud-logging chip descriptions
- Complex tables in PDF
- Successfully converted to CSV tables
- Redo processes to other templates
- Build the app

Synthetic DTC

- Chip descriptions improve DTC predictions
- Step-wise regression selected statistically significant features
- XGBoost relies on LWD/Wireline logs to predict DTC
- All models were robust



H

Acknowledgments

- CREWES Sponsors
- NSERC
- CFREF and GRI
- All the co-authors
- CREWES students and staff



