



Forest Grove  
Data-Driven Decision Making

Optimising technology to  
unlock insights that drive  
better business decisions

# Forest Grove Customer Conference 2023

[www.forestgt.com.au](http://www.forestgt.com.au) | [info@forestgt.com.au](mailto:info@forestgt.com.au) | 1300 559 827



Forest Grove

Data-Driven Decision Making

Optimising technology to  
unlock insights that drive  
better business decisions

# [D2S3] Unconventional data sources: unlocking insight gold from forgotten sources (images & PDF extraction)

**Presented by:**

Forest Grove Analytics Manager, Angus Veitch & Marketing Manager,  
Akanksha Sinha

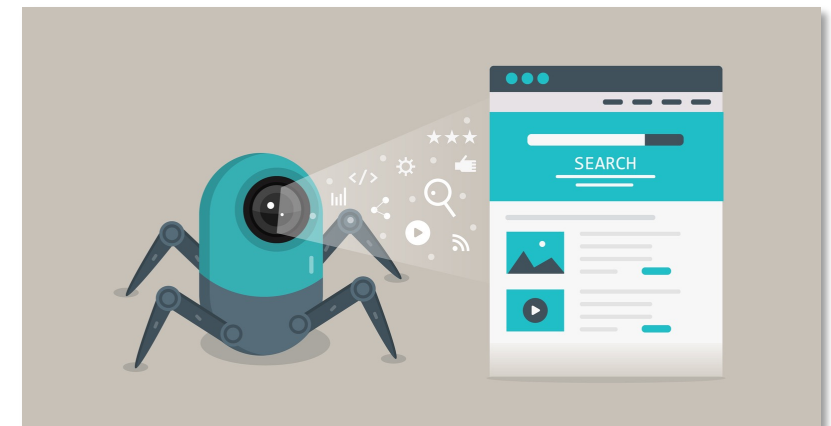
**About the session:**

Modern businesses are held together by high-quality data about their customers, personnel, financials, inventory, and so on, typically maintained by ERPs, CRMs and other dedicated systems. However, value can be found in other types and sources of data as well, including text, images, PDFs, and web pages. This session will explore the opportunities presented by unconventional data sources, featuring a Forest Grove use case in which internal PDF documents were mined for keywords to inform search engine optimisation.

- Use case: Improve Forest Grove's Google ranking (SEO)
- I.E - When finance or data professionals across Australia, like yourself, search on Google for our solutions, Forest Grove should be at the top of the search results for them

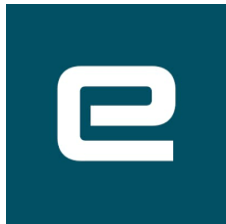


- Google ranking  $\neq$  Google Ads
- Multiple factors contribute to Google ranking, but your website CONTENT is the most important one
- Google's bot crawl the website to determine the relevance of your content to the keywords entered into Google search by someone



- 
- The more relevant the content, the higher the website appears on Google search results
  - So, the BIG Question was - how do we make our content relevant so that we are top of Google search results
  - Easy to get lost in industry lingos and abbreviations, but this is not necessarily how customers talk
  - Right approach: “Hang out where customers are hanging out”
  - RFP or Request for Proposal – business document that announces the start of a project at an organisation
  - In simple terms, a document that lays down business challenges and the solution they are seeking
  - What RFPs could help me with – keywords/ strings of words/ phrases + frequency + context
  - But there were challenges – dozens of RFPs, length, complexity, hard to always add context to findings,  
Control + F doesn't work
  - How we tackled this challenge head-on and delivered data that informed our SEO strategy

# 'Conventional' data sources



PRONTO

myob



2CLOUDNINE

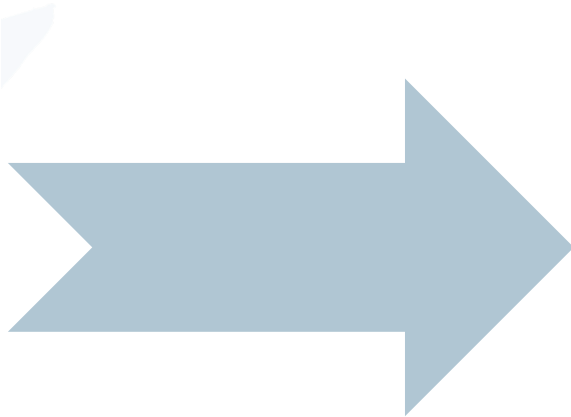
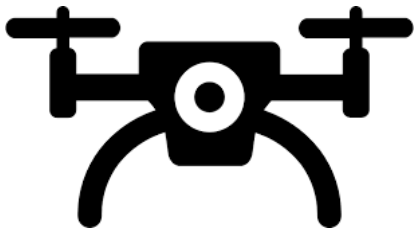
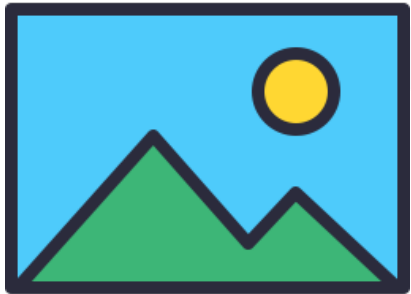


	POK	POB	POD	POA	POV	POE	POF	POG	POH	POI	POJ	POK	POL	POM	PON	POO	POP	POQ	POR	POS	POV	POW	POX	POY
131181AKONOR,AL	131181-131112	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103
131182AKONOR,AL	131182-131112	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
131183AKONOR,AL	131183-131112	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
131184AKONOR,AL	131184-131112	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
131185AKONOR,AL	131185-131112	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
131186AKONOR,AL	131186-131112	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
131187AKONOR,AL	131187-131112	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
131188AKONOR,AL	131188-131112	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
131189AKONOR,AL	131189-131112	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
131190AKONOR,AL	131190-131112	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
131191AKONOR,AL	131191-131112	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103
131192AKONOR,AL	131192-131112	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
131193AKONOR,AL	131193-131112	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
131194AKONOR,AL	131194-131112	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
131195AKONOR,AL	131195-131112	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
131196AKONOR,AL	131196-131112	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
131197AKONOR,AL	131197-131112	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
131198AKONOR,AL	131198-131112	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
131199AKONOR,AL	131199-131112	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
131200AKONOR,AL	131200-131112	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113
131201AKONOR,AL	131201-131112	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103
131202AKONOR,AL	131202-131112	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93

# 'Unconventional' data sources



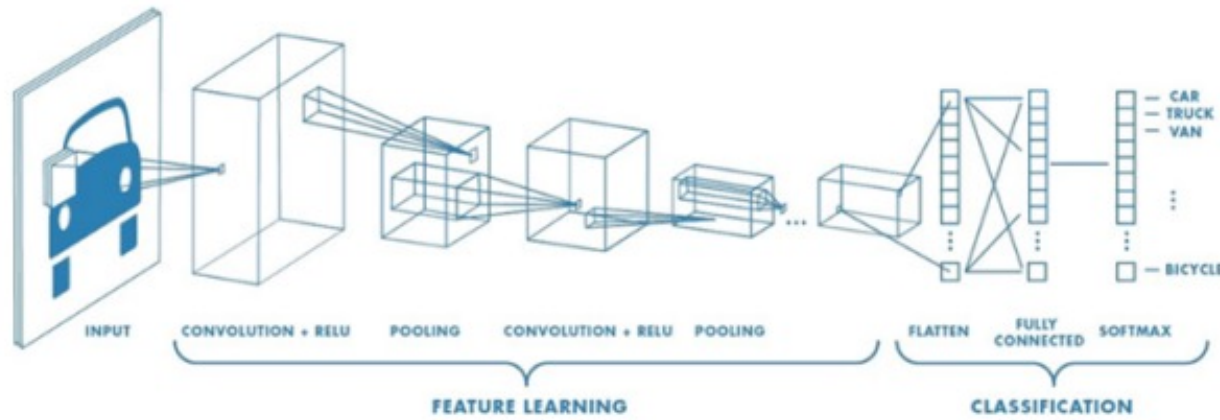
>Lorem ipsum dolor sit amet, consectetur adipiscing elit. posuere tortor vitae elit. Sed vitae metus a elit bibendum malesuada cras pulvinar. Quisque pellentesque nibh in sem. Curabitur ligula. Suspendisse potenti. Duis sit amet augue eu arcu ultrices auctor. Suspendisse elementum, nunc ut molestie elementum, neque augue vulputate elit, eu blandit enim velit vitae nulla. Duis sed.



	PKW	DAK	TEB	WBA	MMK	MAF	WBB	WPL	WBO	WBP	WBT	WBU	WBS
1.0118181818181818	10000-10000	81	81	81	81	81	100	100	100	100	100	100	100
1.0118181818181818	10000-10000	71	81	80	82	86	100	100	100	100	100	100	100
1.0118181818181818	10000-10000	84	84	88	84	100	100	100	100	100	100	100	100
1.0118181818181818	10000-10000	83	83	88	84	88	100	100	100	100	100	100	100
1.0118181818181818	10000-10000	80	80	80	80	80	80	80	80	80	80	80	80
1.0118181818181818	10000-10000	86	86	86	86	86	86	86	86	86	86	86	86
1.0118181818181818	10000-10000	88	88	88	88	88	88	88	88	88	88	88	88
1.0118181818181818	10000-10000	89	89	89	89	89	89	89	89	89	89	89	89
1.0118181818181818	10000-10000	87	87	87	87	87	87	87	87	87	87	87	87
1.0118181818181818	10000-10000	85	85	85	85	85	85	85	85	85	85	85	85
1.0118181818181818	10000-10000	82	82	82	82	82	82	82	82	82	82	82	82
1.0118181818181818	10000-10000	80	80	80	80	80	80	80	80	80	80	80	80
1.0118181818181818	10000-10000	83	83	83	83	83	83	83	83	83	83	83	83
1.0118181818181818	10000-10000	84	84	84	84	84	84	84	84	84	84	84	84
1.0118181818181818	10000-10000	85	85	85	85	85	85	85	85	85	85	85	85
1.0118181818181818	10000-10000	86	86	86	86	86	86	86	86	86	86	86	86
1.0118181818181818	10000-10000	87	87	87	87	87	87	87	87	87	87	87	87
1.0118181818181818	10000-10000	88	88	88	88	88	88	88	88	88	88	88	88
1.0118181818181818	10000-10000	89	89	89	89	89	89	89	89	89	89	89	89
1.0118181818181818	10000-10000	90	90	90	90	90	90	90	90	90	90	90	90
1.0118181818181818	10000-10000	91	91	91	91	91	91	91	91	91	91	91	91
1.0118181818181818	10000-10000	92	92	92	92	92	92	92	92	92	92	92	92
1.0118181818181818	10000-10000	93	93	93	93	93	93	93	93	93	93	93	93
1.0118181818181818	10000-10000	94	94	94	94	94	94	94	94	94	94	94	94
1.0118181818181818	10000-10000	95	95	95	95	95	95	95	95	95	95	95	95
1.0118181818181818	10000-10000	96	96	96	96	96	96	96	96	96	96	96	96
1.0118181818181818	10000-10000	97	97	97	97	97	97	97	97	97	97	97	97
1.0118181818181818	10000-10000	98	98	98	98	98	98	98	98	98	98	98	98
1.0118181818181818	10000-10000	99	99	99	99	99	99	99	99	99	99	99	99
1.0118181818181818	10000-10000	100	100	100	100	100	100	100	100	100	100	100	100



# 'Unstructured' data - Images



Architecture of a convolutional neural network. Image by: [Sumit Saha](#), 2018.

The screenshot shows a web application titled "Cats or dogs? Let's check the tiles!". It displays a grid of 10 image classification results. Each result includes an image, a filename, a probability score, and a prediction.

Filename	Image	Probability	Prediction
10207.jpg		0.09	Dog
10440.jpg		0.58	Cat
10553.jpg		0.81	Cat
10714.jpg		0.59	Cat
1082.jpg		0.14	Dog
10981.jpg		0.91	Cat
11080.jpg		0.21	Dog
113.jpg		0.38	Dog
11372.jpg		0.78	Cat
11509.jpg		0.45	Dog

Showing 1 to 10 of 100 entries

Navigation: Previous | 1 | 2 | 3 | 4 | 5 | ... | 10 | Next

Buttons: Reset | Apply | Close

<https://www.knime.com/blog/how-to-classify-images-codeless-cnn>



# Web scraping

WINEENTHUSIAST

[355,982 total results] READ ARTICLES [8,426]

RELATED REVIEWS 1-6 of 347,556 Filter by Sort by

- Clif Family 2019 Cima Cabernet Sauvignon (Howell Mountain)** NAPA  
Handsome, intricate oak spices top off this complex and layered wine beautifully, ...  
96 Points  
\$150
- Clif Family 2019 Cold Springs Vineyard Estate Cabernet Sauvignon (Howell Mountain)** NAPA  
An alluring, sexy, spicy, cigar-box aroma and impossibly rich black fruit and ...  
95 Points  
\$150
- Robert Renzoni 2019 Riserva Meritage (Temecula Valley)** SOUTH COAST  
This Bordeaux blend shows off Temecula Meritage in a very promising light, ...  
94 Points  
\$76
- Bodega Garzón 2019 Petit Clos Block #212 Tannat (Maldonado)** URUGUAY  
This well-structured wine offers aromas of green bell pepper and plum accented ...  
94 Points  
\$60
- Dunites 2021 Syrah-Grenache** CALIFORNIA OTHER  
Crisp aromas of rose, carnation and black-raspberry sorbet are both juicy and ...  
94 Points

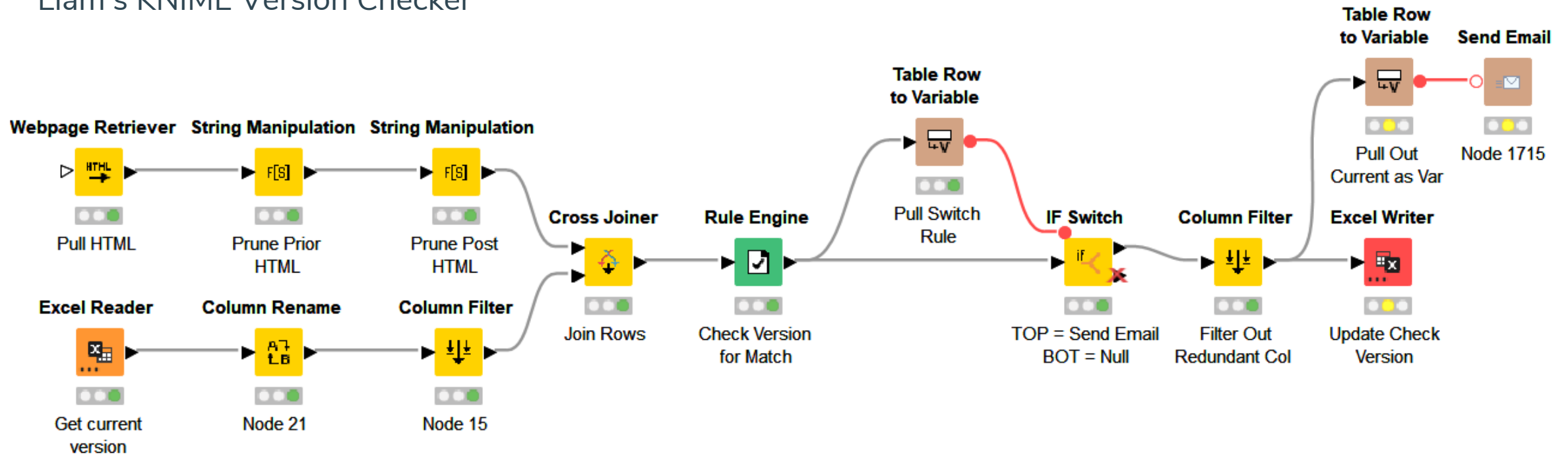


S	Country	S	Province	S	Title	S	Description	I	Price	I	Points
	Italy		Sicily & Sardinia		Nicosia 2013 Vulkà Bianco (Etna)		Aromas include tropical fruit, broom, brimstone and ...	?			8
	Portugal		Douro		Quinta dos Avidagos 2011 Avid...		This is ripe and fruity, a wine that is smooth while sti...	15			8
	US		Oregon		Rainstorm 2013 Pinot Gris (Will...		Tart and snappy, the flavors of lime flesh and rind d...	14			8
	US		Michigan		St. Julian 2013 Reserve Late H...		Pineapple rind, lemon pith and orange blossom start ...	13			8
	US		Oregon		Sweet Cheeks 2012 Vintner's R...		Much like the regular bottling from 2012, this comes ...	65			8
	Spain		Northern Spain		Tandem 2011 Ars In Vitro Tem...		Blackberry and raspberry aromas show a typical Na...	15			8
	Italy		Sicily & Sardinia		Terre di Giurfo 2013 Belsito Fra...		Here's a bright, informal red that opens with aromas...	16			8
	France		Alsace		Trimbach 2012 Gewurztraminer...		This dry and restrained wine offers spice in profusio...	24			8
	Germany		Rheinhessen		Heinz Eifel 2013 Shine Gewürzt...		Savory dried thyme notes accent sunnier flavors of ...	12			8
	France		Alsace		Jean-Baptiste Adam 2012 Les ...		This has great depth of flavor with its fresh apple a...	27			8
	US		California		Kirkland Signature 2011 Mount...		Soft, supple plum envelopes an oaky structure in thi...	19			8
	France		Alsace		Leon Beyer 2012 Gewurztramin...		This is a dry wine, very spicy, with a tight, taut text...	30			8
	US		California		Louis M. Martini 2012 Cabernet...		Slightly reduced, this wine offers a chalky, tannic ba...	34			8
	Italy		Sicily & Sardinia		Masseria Setteporte 2012 Ross...		This is dominated by oak and oak-driven aromas tha...	?			8
	US		California		Mirassou 2012 Chardonnay (C...		Building on 150 years and six generations of winema...	12			8
	Germany		Mosel		Richard Böcking 2013 Devon Ri...		Zesty orange peels and apple notes abound in this s...	24			8
	Argentina		Other		Felix Lavaque 2010 Felix Malbe...		Baked plum, molasses, balsamic vinegar and cheesy ...	30			8
	Argentina		Mendoza Pro...		Gaucha Andino 2011 Winemak...		Raw black-cherry aromas are direct and simple but g...	13			8
	Spain		Northern Spain		Pradorey 2010 Vendimia Selecc...		Desiccated blackberry, leather, charred wood and m...	28			8
	US		Virginia		Quiévreumont 2012 Meritage (Vi...		Red fruit aromas pervade on the nose, with cigar bo...	32			8
	US		Virginia		Quiévreumont 2012 Vin de Mais...		Ripe aromas of dark berries mingle with ample notes...	23			8
	US		Oregon		Acrobat 2013 Pinot Noir (Oregon)		A sleek mix of tart berry, stem and herb, along with ...	20			8
	Italy		Sicily & Sardinia		Baglio di Pianetto 2007 Ficigno...		Delicate aromas recall white flower and citrus. The p...	19			8
	US		California		Bianchi 2011 Signature Selectio...		This wine from the Geneseo district offers aromas of...	22			8
	Italy		Sicily & Sardinia		Canicattì 2009 Aynat Nero d'A...		Aromas of prune, blackcurrant, toast and oak carry ...	35			8

<https://www.kaggle.com/datasets/zynicide/wine-reviews>

# Web scraping

## Liam's KNIME Version Checker



# PDFs

NATIONAL SALES MANAGER		TOTAL SALARY (\$p.a)			
	No. Cases	LQ	Median	UQ	Avg
<b>ANNUAL TURNOVER</b>					
\$10 million to \$20 million	5	115,844	124,808	146,280	127,322
\$20 million to \$100 million	38	146,884	159,367	176,313	163,268
\$100 million to \$200 million	11	173,951	187,348	211,814	188,255
Over \$200 million	27	93,880	93,880	93,467	93,268
<b>NUMBER OF EMPLOYEES</b>					
Less than 26	6	128,846	146,932	162,154	145,246
51 to 100	50	93,880	116,842	136,761	126,465
101 to 200	17	146,880	176,792	187,469	176,542
201 to 500	6	176,880	176,828	213,824	188,225
<b>LOCATION</b>					
NSW/ACT	19	93,323	126,261	148,777	148,747
Metro	15	93,880	123,880	148,761	138,465
VIC/TAS	36	93,814	146,824	181,412	148,221
Metro	35	93,807	146,800	181,418	148,415
QLD	20	94,880	150,222	162,428	137,880
Metro	17	94,880	150,480	164,128	136,911
WA	7	93,278	141,880	195,869	148,887
Metro	7	93,278	141,880	195,869	148,887
<b>INDUSTRY GROUP</b>					
Wholesale - Other	10	176,880	182,846	188,828	177,462
Hospitality and Recreation	24	93,880	93,880	93,280	93,465
Government/Institutions/Social Services	12	188,222	196,747	148,840	189,248
Manufacturing - Metal/Automotive	12	131,228	162,154	176,182	164,107
Manufacturing - Other	6	146,880	168,284	172,888	159,222



S	Role_Description	S	Category_Name	D	SAL_LQ	D	SAL_M...	D	SAL_UQ	D	SAL_Avg	D	PKG_LQ
	NATIONAL SALES MANAGER		Summary		93,880		146,323		176,288		143,822		138,880
	NATIONAL SALES MANAGER		\$10 million to \$20 million		115,844		124,808		146,280		127,322		126,876
	NATIONAL SALES MANAGER		\$20 million to \$100 million		146,884		159,367		176,313		163,268		169,362
	NATIONAL SALES MANAGER		\$100 million to \$200 million		173,951		187,348		211,814		188,262		215,952
	NATIONAL SALES MANAGER		Over \$200 million		93,880		93,880		93,467		93,268		117,385
	NATIONAL SALES MANAGER		Less than 26		128,846		146,932		162,154		145,246		162,907
	NATIONAL SALES MANAGER		51 to 100		93,880		116,842		136,761		126,465		117,385
	NATIONAL SALES MANAGER		101 to 200		146,880		176,792		187,469		176,542		166,478
	NATIONAL SALES MANAGER		201 to 500		176,880		176,828		213,824		188,228		188,128
	NATIONAL SALES MANAGER		NSW/ACT		93,323		126,261		148,777		148,747		122,328
	NATIONAL SALES MANAGER		NSW/ACT Metro		93,880		123,880		148,761		138,467		117,384
	NATIONAL SALES MANAGER		VIC/TAS		93,814		146,824		181,412		148,221		138,485
	NATIONAL SALES MANAGER		VIC/TAS Metro		93,807		146,800		181,418		148,415		138,922
	NATIONAL SALES MANAGER		QLD		94,880		150,222		162,428		137,880		124,764
	NATIONAL SALES MANAGER		QLD Metro		94,880		150,480		164,128		136,911		122,462
	NATIONAL SALES MANAGER		WA		93,278		141,880		195,869		148,887		138,728
	NATIONAL SALES MANAGER		WA Metro		93,278		141,880		195,869		148,887		138,728
	NATIONAL SALES MANAGER		Wholesale - Other		176,880		182,846		188,828		177,461		204,285
	NATIONAL SALES MANAGER		Hospitality and Recreation		93,880		93,880		93,280		93,465		117,385
	NATIONAL SALES MANAGER		Government/Institution...		188,222		196,747		148,840		189,268		180,124
	NATIONAL SALES MANAGER		Manufacturing - Metal/...		131,228		162,154		176,182		164,107		164,128
	NATIONAL SALES MANAGER		Manufacturing - Other		146,880		168,284		172,888		159,228		169,876
	SALES MANAGER - REGION...		Summary		93,448		117,762		141,868		121,762		138,888
	SALES MANAGER - REGION...		Less than \$5 million		94,468		107,076		121,813		117,362		148,374
	SALES MANAGER - REGION...		\$10 million to \$20 million		94,278		125,476		148,348		123,823		124,438
	SALES MANAGER - REGION...		\$20 million to \$100 million		93,880		112,763		127,876		114,951		138,888
	SALES MANAGER - REGION...		\$100 million to \$200 million		136,927		155,424		165,744		148,728		138,821
	SALES MANAGER - REGION...		Over \$200 million		124,765		125,880		148,047		130,163		115,230
	SALES MANAGER - REGION...		Less than 26		101,880		112,876		141,467		124,227		136,737
	SALES MANAGER - REGION...		26 to 50		111,218		121,880		152,416		126,888		151,835
	SALES MANAGER - REGION...		51 to 100		93,880		105,268		127,876		114,281		138,888
	SALES MANAGER - REGION...		101 to 200		124,285		142,796		165,761		144,032		148,465
	SALES MANAGER - REGION...		201 to 500		111,177		126,927		151,114		125,872		151,884

---

## Use Case 2: Mining PDFs for SEO Keywords