

Introduction to Data Analysis and Business Intelligence (BI)

Santiago Paiva

ConFoo.CA
WEB TECHNO CONFERENCE

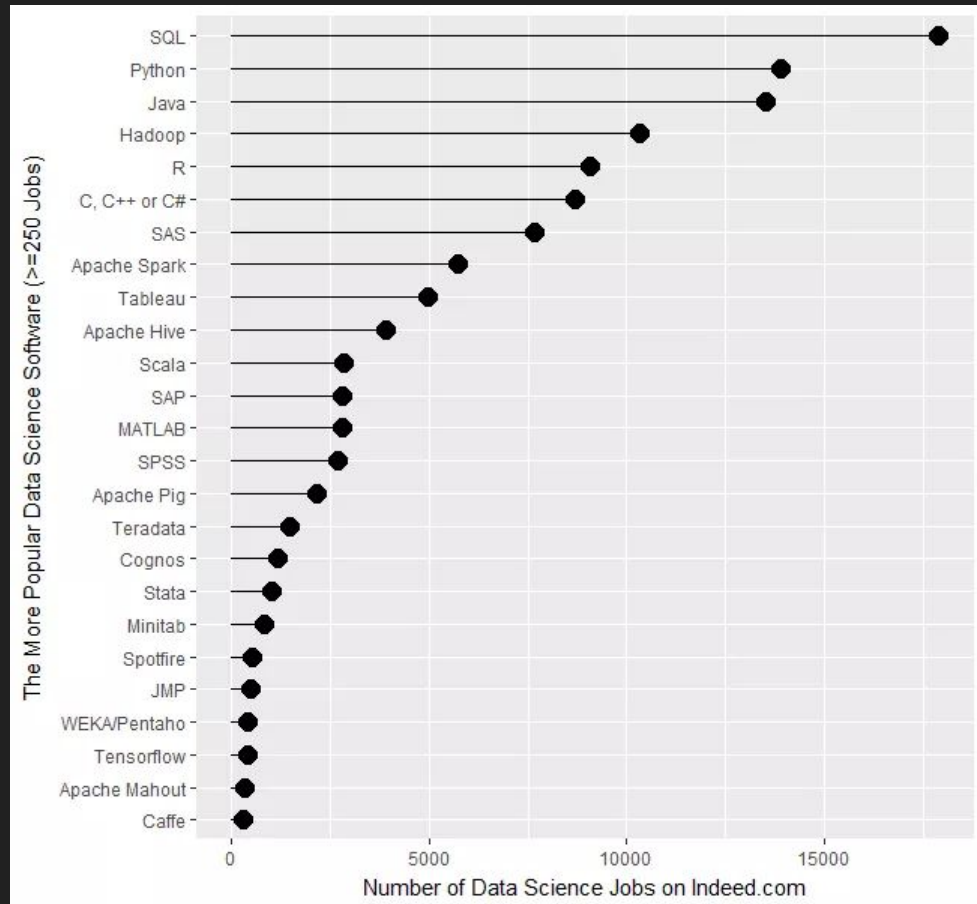
Presentation Overview

- Introduction (4 min)
- Part 1: **Overview of BI Tools** (10 min)
- Part 2: **Google Webmaster Tools** (10 min)
- Part 3: **Google Adwords** (10 min)
- Conclusion (1 min)
- Q & A (10 min)

Slides: <https://www.github.com/spaiva/confoo-2017>

About Me

- Currently at Montreal Neurological Institute (Neuroscience)
- **Startups**
 - Busbud (Travel)
 - Frank + Oak (Retail)
 - Sharethebus (Travel)
- **Research**
 - Polytechnique (Astrophysics)
 - McGill University (AI)
 - MIT (Energy)
- **Confoo**
 - 2014: <https://www.github.com/spaiva/confoo-2014>
 - 2016: <https://www.github.com/spaiva/confoo-2016>





78% are planning to increase the use of cloud for BI and data management in the next twelve months.



46% of organizations prefer public cloud platforms for cloud BI, analytics and data management deployments.



78% are planning to increase the use of cloud for BI and data management in the next twelve months.



46% of organizations prefer public cloud platforms for cloud BI, analytics and data management deployments.

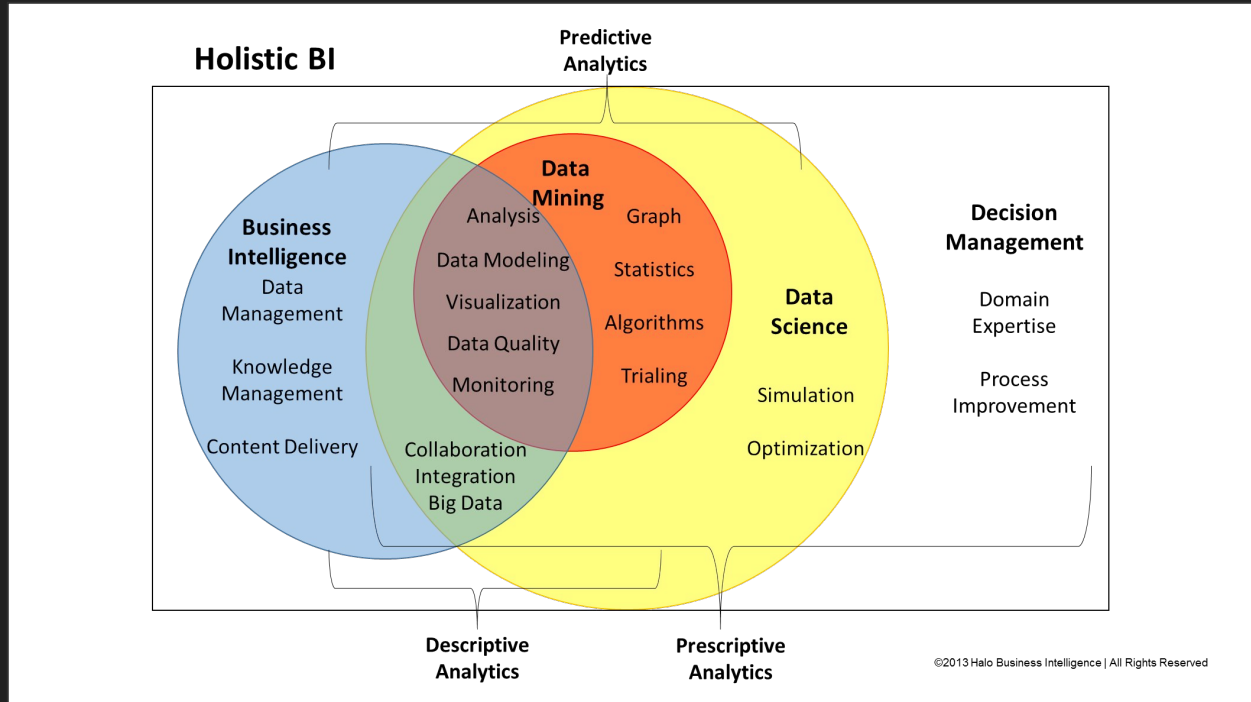


Cloud BI adoption increased in respondent companies from **29%** to **43%** from 2013 to 2016.



Almost half of organizations using cloud BI (**46%**) use a public cloud for BI and data management compared to less than a third (**30%**) for hybrid cloud and **24%** for private cloud.

Business Intelligence vs Data Scientist



The rise of the Full-Stack Data Analyst

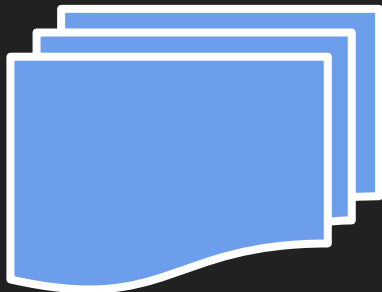
Data Management



MySQL
PostgreSQL
MongoDB
Redis
Cassandra
DW



Data Science / BI



Excel
SQL
Python
R
Hadoop

...



Data Visualization



Excel
D3
HighCharts
Google Visualization
Grafana
...

Part I: BI Tools

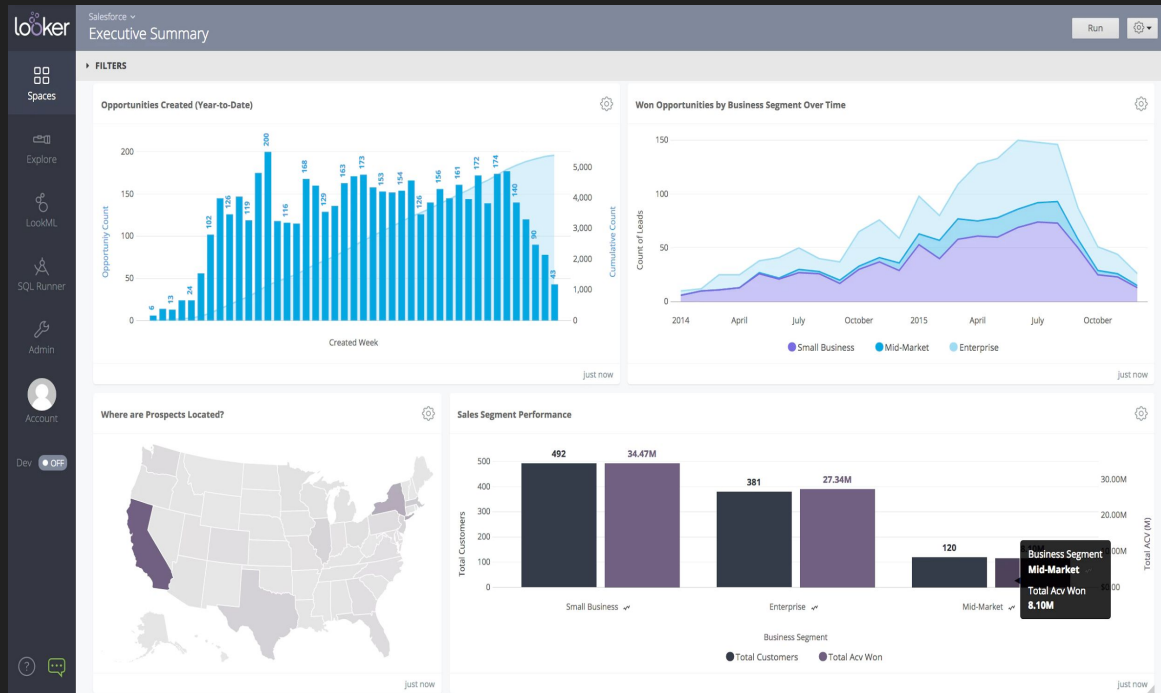
Tool #1: Microsoft Excel

- Still very important!
- Powerful Macros
- VLOOKUP and Pivot Tables
- Rapid prototyping



<https://apandre.wordpress.com/2011/06/05/excel-as-bi-platform/>

Tool #2: Looker



- LookML Language
- Custom Dashboards
- Very common in e-commerce
- Examples: Frank + Oak and Breather

<https://www.getapp.com/business-intelligence-analytics-software/a/looker/>

Tool #3: Tableau



- Popularity ↑↑↑
- Pretty User Interface
- Get moving quickly
- Rapid prototyping



Tool #4: TIBCO Spotfire



- Vivek Ranadivé
- Heavy lifting
- Requires knowledge of Statistics!

TIBCOTM
Spotfire[®]

<http://spotfire.tibco.com/>

Tool #5: Stata



- Simulation
- Modelling
- Heavy on Statistics
- Highly Recommended

The screenshot shows the Stata 14.0 interface with the following components:

- Review Panel:** Shows the command `. describe` and its output.
- Command Window:** Displays the output of the `. describe` command.
- Variables Panel:** Lists all variables in the dataset with their names and labels.
- Properties Panel:** Shows the properties of the selected variable 'make'.

Command Window Output:

```
. describe
Contains data from http://www.stata-press.com/data/r14/auto.dta
obs:      74      1978 Automobile Data
vars:     12      13 Apr 2014 17:45
size:     3,182   (_dta has notes)
```

variable name	storage type	display format	value label	variable label
make	str18	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%6.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

Sorted by: foreign

Variables Panel:

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. in.)
gear_ratio	Gear Ratio
foreign	Car type

Properties Panel (make):

Property	Value
Name	make
Label	Make and Model
Type	str18
Format	%-18s
Value label	
Notes	

Data Panel:

Property	Value
Filename	auto.dta
Label	1978 Automobile Data
Notes	

Tool #5: Stata



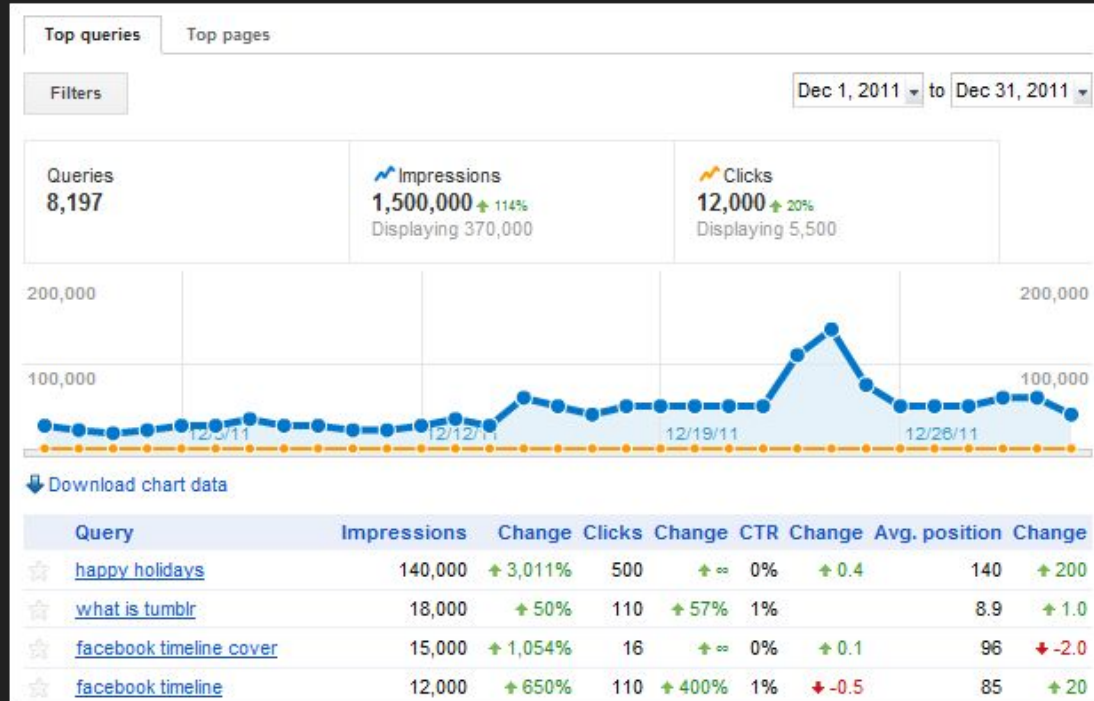
- Previously at NY Times
- Massive Monte-Carlo Simulations on Stata during American Elections
- Now doing Sport Analytics
- <https://FiveThirtyEight.com>

Attention

We can build our own BI tools!

Part II: Google Webmaster Tools

Google Webmaster Tools (GWT)



<http://amintor.com/1/post/2013/11/technology-extend-python-library-for-google-web-master-tools.html>

Google Webmaster Tools (GWT)

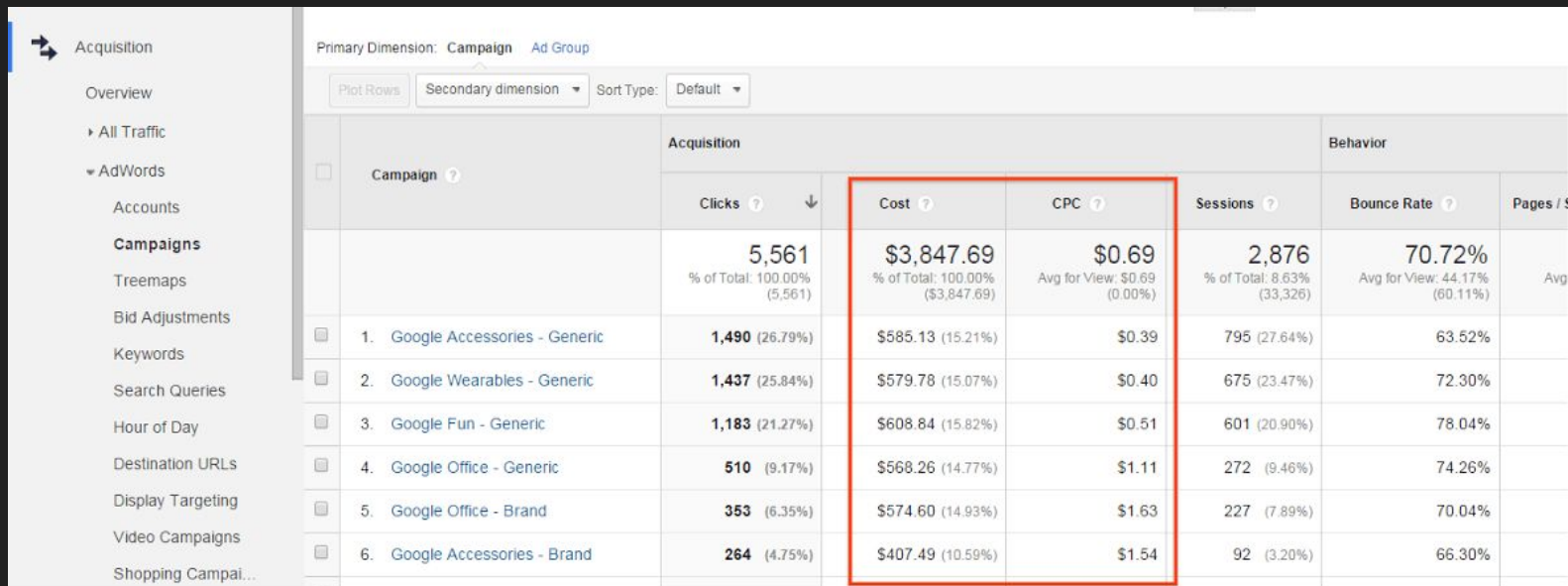
URL: <https://www.google.com/webmasters/>

BI Tool: <https://gist.github.com/spaiva/c7519dc3b2c78c94f2b2ee9a05f9785d>

Data Analysis (Pandas): <https://gist.github.com/spaiva/083387046ebfec858316>

Part III: Google Adwords

Google Adwords



The screenshot displays the Google AdWords interface. On the left is a navigation sidebar with categories like Acquisition, Overview, All Traffic, AdWords, Accounts, Campaigns, Treemaps, Bid Adjustments, Keywords, Search Queries, Hour of Day, Destination URLs, Display Targeting, Video Campaigns, and Shopping Campaigns. The main content area shows a table of campaign performance data. At the top, it indicates 'Primary Dimension: Campaign' and 'Ad Group'. Below this are filters for 'Plot Rows', 'Secondary dimension', and 'Sort Type: Default'. The table has columns for Campaign, Clicks, Cost, CPC, Sessions, Bounce Rate, and Pages / Session. The 'Cost' and 'CPC' columns are highlighted with a red box. The data is as follows:

Campaign	Acquisition			Behavior		
	Clicks	Cost	CPC	Sessions	Bounce Rate	Pages / Session
	5,561 % of Total: 100.00% (5,561)	\$3,847.69 % of Total: 100.00% (\$3,847.69)	\$0.69 Avg for View: \$0.69 (0.00%)	2,876 % of Total: 8.63% (33,326)	70.72% Avg for View: 44.17% (60.11%)	
1. Google Accessories - Generic	1,490 (26.79%)	\$585.13 (15.21%)	\$0.39	795 (27.64%)	63.52%	
2. Google Wearables - Generic	1,437 (25.84%)	\$579.78 (15.07%)	\$0.40	675 (23.47%)	72.30%	
3. Google Fun - Generic	1,183 (21.27%)	\$608.84 (15.82%)	\$0.51	601 (20.90%)	78.04%	
4. Google Office - Generic	510 (9.17%)	\$568.26 (14.77%)	\$1.11	272 (9.46%)	74.26%	
5. Google Office - Brand	353 (6.35%)	\$574.60 (14.93%)	\$1.63	227 (7.89%)	70.04%	
6. Google Accessories - Brand	264 (4.75%)	\$407.49 (10.59%)	\$1.54	92 (3.20%)	66.30%	

<https://www.shopify.com/blog/16909640-how-to-spend-your-first-100-on-google-adwords>

Google Adwords

Types of reports:

- ACCOUNT_PERFORMANCE_REPORT
- CRITERIA_PERFORMANCE_REPORT
- KEYWORDS_PERFORMANCE_REPORT
- CAMPAIGN_PERFORMANCE_REPORT
- AD_PERFORMANCE_REPORT
- AD_GROUP_PERFORMANCE_REPORT
- PAID_ORGANIC_QUERY_REPORT
- URLS_PERFORMANCE_REPORT
- There are a total of 43 types of reports!

Source: <https://developers.google.com/adwords/api/docs/appendix/reports>

Google Adwords

URL: <https://adwords.google.com/home/>

BI: <https://gist.github.com/spaiva/745d901675bef85bebac77220df2b08f>

Data Analysis:

<https://gist.github.com/spaiva/135bbb5e6392fa6c567fbe79fe2d297a>

Conclusion

- **Overview of BI Tools**
 - Excel, Looker, Tableau, TIBCO Spotfire, Stata
 - Attention of the consumer
- **Custom BI Tool #1: Google Webmaster Tools (GWT)**
 - GTW Python Library (`gwt.py`)
 - Custom SEO functions using Pandas (`df.py`)
- **Custom BI Tool #2: Google Adwords**
 - GA Python Library (`ga.py`)
 - Custom function using Pandas (`ga_models.py`)

Q & A

Introduction to Data Analysis and Business Intelligence (BI)

Santiago Paiva

paiva.santiago@gmail.com

Twitter: @stronnics

github.com/spaiva/confoo-2017