

# SGI MineSet

Peter McGonigal  
Systems Engineer

©2015 SGI. All rights reserved.

# Agenda

- Introduce **MineSet** (5mins)
- Practical demo of **MineSet** (30mins)

# What is MineSet

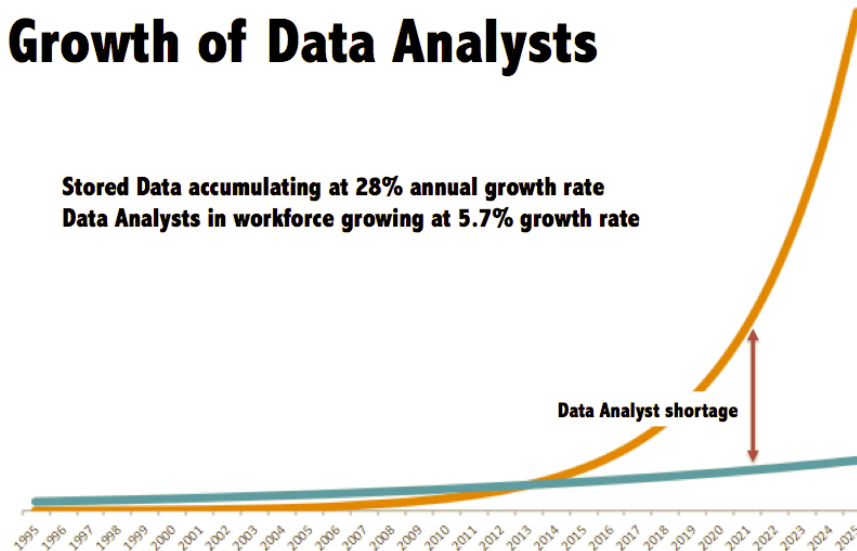
- **MineSet**: Is a Data Mining and Data Visualization tool developed by SGI
  - *Explore datasets*
  - *Uses Machine Learning*
  - *Finds patterns & relationships hidden in data*
- **Goal**: Make MineSet Easy & Intuitive. Allows Users to Explore/Discover Insights. Gain **VALUE** from Insights.
- **When**: SGI plans to launch in Q2 CY2015

# Why develop MineSet

- Saw organisations are collecting large amounts of data often didn't have the resources to make effective use of the data

## Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate  
Data Analysts in workforce growing at 5.7% growth rate



Source: <http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/>

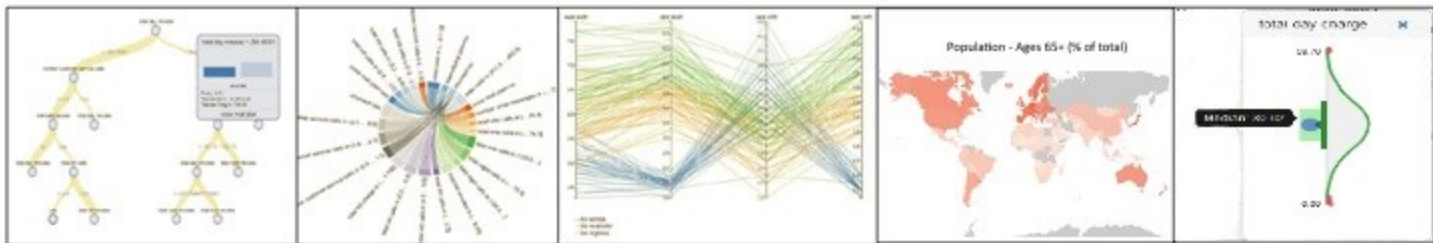
- There is a need for a good Visual-Analytics tool like MineSet

# About MineSet

**What:** Visual Analytics toolset  
to help users Discover Insights

**How:** Web based – easy to access & use.  
No software installed on client system.  
Zero Programming.

**Why:** Discover more. Predict better. Act faster



# MineSet – Markets

## Business

- **Market segmentation** - Identify the common characteristics of customers who buy the same products
- **Customer churn** - Predict which customers are likely to leave and go to a competitor
- **Fraud detection** - Identify which transactions are most likely to be fraudulent
- **Direct marketing** - Identify which prospects should be included in a mailing list to obtain the highest response rate
- **Market basket analysis** - Understand what products or services are commonly purchased together
- **Trend analysis** - Reveal the difference between a typical customer this month and last

## Research & Science

- **Bio-medical and Clinical data** - Identify trends in research data, assist in detections and predictions
- **Environmental data** - Environmental agencies assessing the impact of changing land-use patterns on climate change
- **Demographics** – Identify trends in population and human behavior

**And many other areas.....**



# MineSet – demo

Name	Type	N
state	text	5,000
account-length	number	5,000
area-code	number	5,000
phone-number	text	5,000
international-plan	text	5,000
voice-mail-plan	text	5,000
number-of-vmail-messages	number	5,000
total-day-minutes	number	5,000
total-day-calls	number	5,000
total-day-charge	number	5,000
total-eve-minutes	number	5,000
total-eve-calls	number	5,000
total-eve-charge	number	5,000
total-night-minutes	number	5,000
total-night-calls	number	5,000
total-night-charge	number	5,000
total-intl-minutes	number	5,000
total-intl-calls	number	5,000
total-intl-charge	number	5,000
number-of-service-calls	number	5,000
churned	text	5,000

**Column Importance**

**Machine Learning Options**

Predicted Column: churned

Weight: [dropdown]

Extra attributes: 5

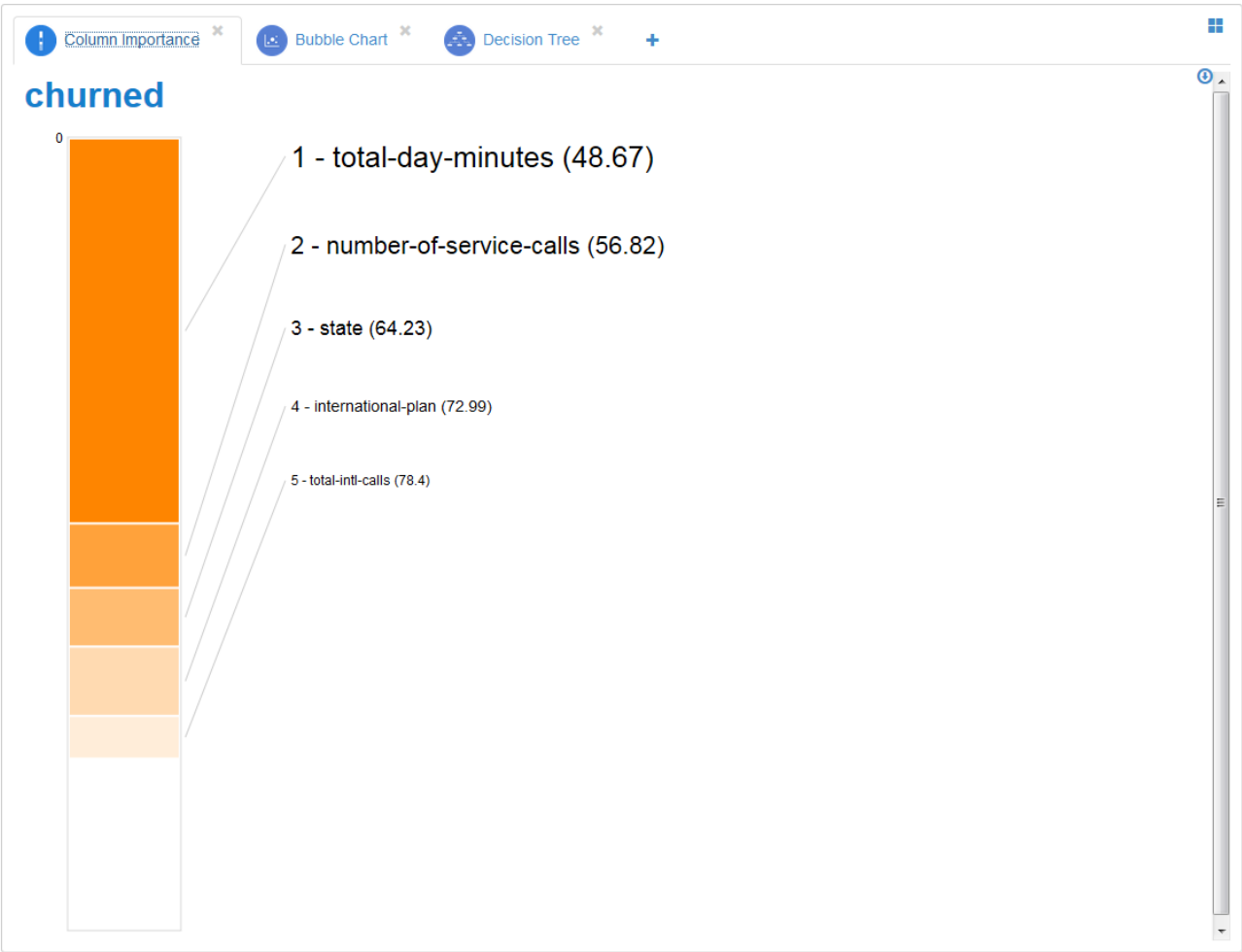
Columns to ignore: [drag column(s) here]

Apply

**Selections**

Selected records: none

Highlight Filter





Name	Type	N
state	text	5,000
account-length	number	5,000
area-code	number	5,000
phone-number	text	5,000
international-plan	text	5,000
voice-mail-plan	text	5,000
number-of-vmil-messages	number	5,000
total-day-minutes	number	5,000
total-day-calls	number	5,000
total-day-charge	number	5,000
total-eve-minutes	number	5,000
total-eve-calls	number	5,000
total-eve-charge	number	5,000
total-night-minutes	number	5,000
total-night-calls	number	5,000
total-night-charge	number	5,000
total-intl-minutes	number	5,000
total-intl-calls	number	5,000
total-intl-charge	number	5,000
number-of-service-calls	number	5,000
churned	text	5,000

**Plot**

X axis: total-day-minutes

Y axis: number-of-service-calls

Color by: churned

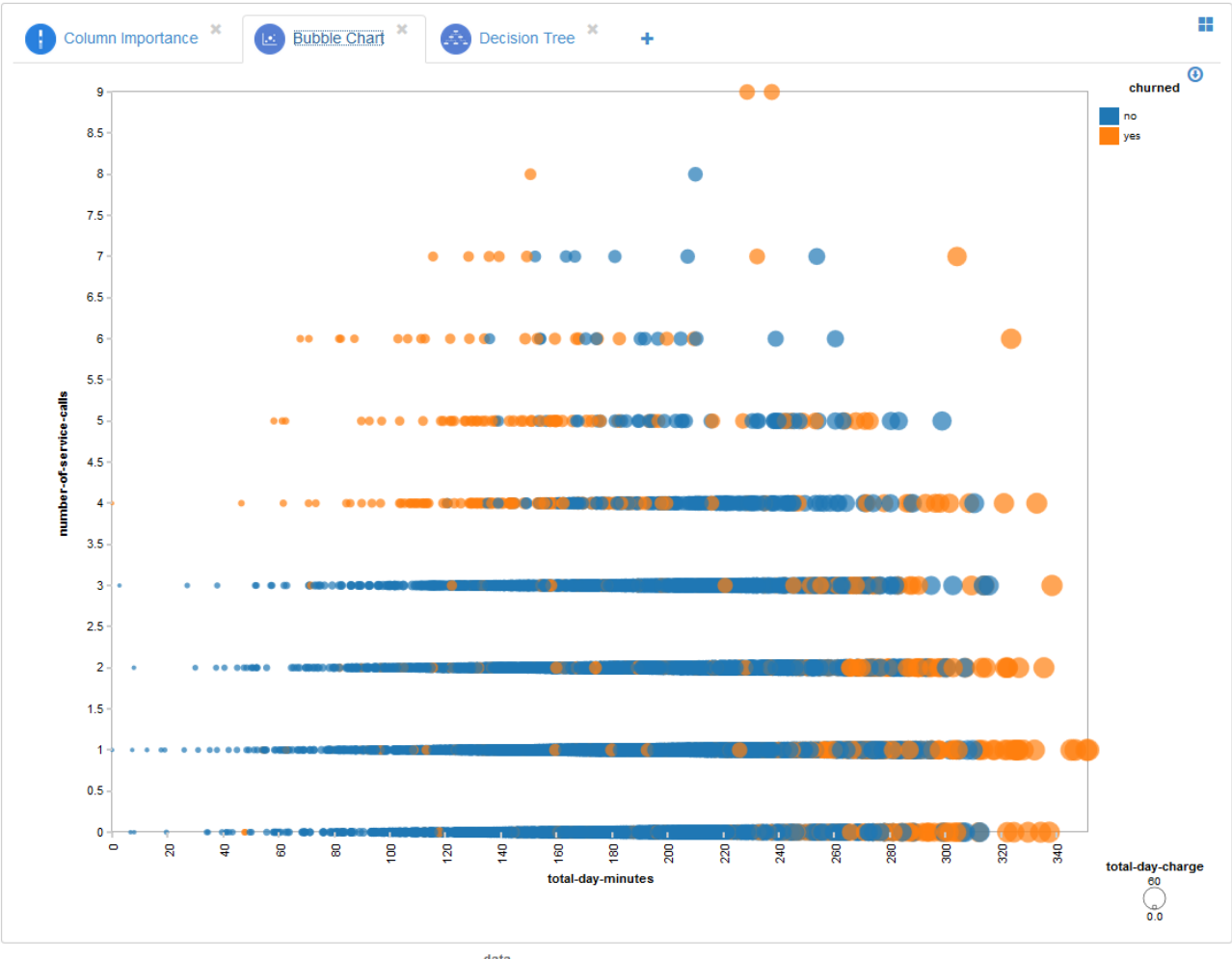
Size: total-day-charge

Quick - Precise [100]

Selections

Selected records: none

Highlight Filter



File Edit View History Bookmarks Tools Help

ra.engr.sgi.com:8080/mineset-web/dataset?id=281#

Most Visited Getting Started Latest Headlines Data Science 101: Mac...

MineSet® telco-customer-churn-data DATA EXPLORE Peter McGonigal

Columns	Name	Type	N
★	state	text	5,000
	account-length	number	5,000
	area-code	number	5,000
	phone-number	text	5,000
★	international-plan	text	5,000
	voice-mail-plan	text	5,000
	number-of-vmail-messages	number	5,000
★	total-day-minutes	number	5,000
	total-day-calls	number	5,000
	total-eve-minutes	number	5,000
	total-eve-calls	number	5,000
	total-eve-charge	number	5,000
	total-night-minutes	number	5,000
	total-night-calls	number	5,000
	total-night-charge	number	5,000
	total-intl-minutes	number	5,000
★	total-intl-calls	number	5,000
	total-intl-charge	number	5,000
★	number-of-service-calls	number	5,000
★	churned	text	5,000

**Decision Tree**

Orientation: vertical  
 Node distance: 200  
 Bars per line: 18  
 Autofit:   
 Expanded nodes:   
 Dashed branches:   
 Thickness/weight:

**Machine Learning Options**

Target: churned  
 Maximum Levels: 5  
 Splitting Method: Mutual Info  
 Pruning Method: Confidence  
 Pruning Factor: 0.7  
 Allow one-off splits:   
 Columns to ignore: [drag column(s) here]  
 Apply

**Selections**

Selected records: none  
 Highlight Filter

Column Importance x Bubble Chart x Decision Tree x +

```

graph TD
    Root["churned  
total-day-minutes"]
    Root -- "<= 264.45" --> Node1["'total-day-minutes' <= 264.45  
churned  
number-of-service-calls"]
    Root -- "> 264.45" --> Node2["'total-day-minutes' > 264.45  
churned  
voice-mail-plan"]
    Node1 -- "<= 3.5" --> Node1_1["number-of-service-calls' <= 3.5  
churned  
international-plan"]
    Node1 -- "> 3.5" --> Node1_2["number-of-service-calls' > 3.5  
churned  
total-day-minutes"]
    Node2 -- "no" --> Node2_1["'voice-mail-plan' == 'no'  
churned  
yes"]
    Node2 -- "yes" --> Node2_2["'voice-mail-plan' == 'yes'  
churned  
no"]
    Node1_2 -- "<= 160.2" --> Node1_2_1["'total-day-minutes' <= 160.2  
churned  
yes"]
    Node1_2 -- "> 160.2" --> Node1_2_2["'total-day-minutes' > 160.2  
churned  
total-eve-minutes"]
  
```

data

Name	Type	N
Country Name	text	2,354
Date	number	2,354
Transit-Railways, (million passenger-km)	number	854
Transit-Passenger cars (per 1,000 people)	number	776
Business-Mobile phone subscribers	number	2,214
Business-Internet users (per 100 people)	number	2,156
Health-Mortality, number under-5 (per 1,000 live births)	number	2,112
Health-Health expenditure per capita (current US\$)	number	2,059
Health-Health expenditure, total (% GDP)	number	-
Population-Total (count)	number	2,354
Population-Urban (count)	number	2,299
Population-Birth rate, crude (per 1,000)	number	2,176
Health-Life expectancy at birth, female (years)	number	2,142
Health-Life expectancy at birth, male (years)	number	2,142

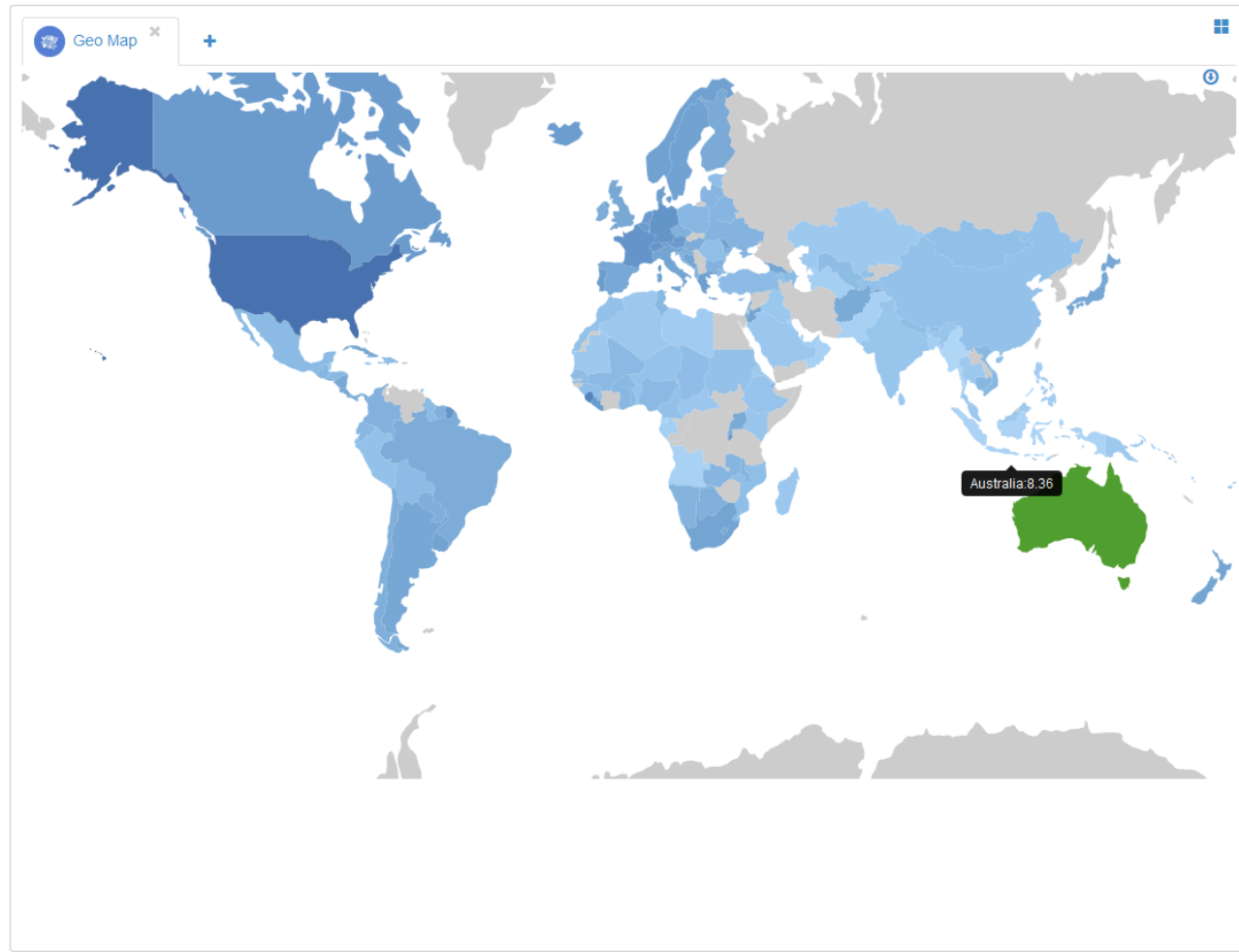
**Geo Map**

Geo Type: World  
Geo Column: Country Name  
Color: Health- Health expenditure per capita (current US\$)  
Value: average

**Selections**

Selected records: none

Highlight Filter



data

# MineSet

Thank-you !

Q & A

*Peter McGonigal*  
*[petermc@sgi.com](mailto:petermc@sgi.com)*