Using the Power of Data Analytics and Predictive Models to Combat Sales Suppression and Refund Fraud

Ted London
Partner, Government Solutions
August 1, 2017

Introduction to the Power of Analytics: Flipping a Coin

- Two people asked to participate in a test
  - One will flip a coin 50 times
  - One will pretend to flip a coin 50 times and write down “H” or “T”

Person 1:

H T T H H H T T H H H T H H H H H H H H T T H T T
H H H T H T H H H H H T T T T T T T H H H T T T H

Person 2:

H T H T T T H H H T H T H T T T T H H T T H T T H
H T T H T H H T T T H H H T T T T H H T H T H T H

How should it look statistically?

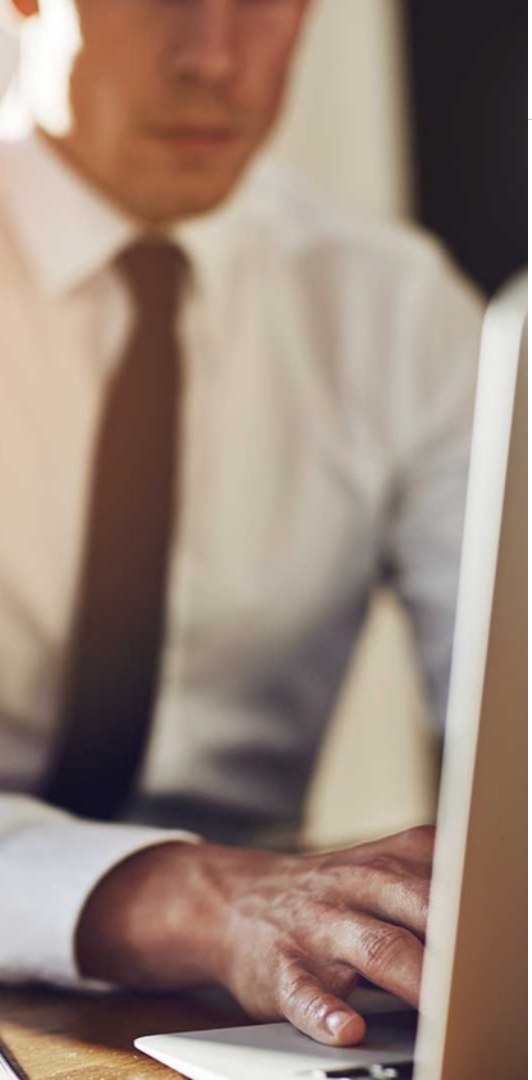
- Chance of getting 6 heads or tails in a row: 1 chance in 32
- Chance of getting 7 heads or tails in a row: 1 chance in 64
- A long run is therefore likely

- Person 1:

H T T H H H T T H H H T H H H H H H H H T T H T T
H H H T H T H H H H H T T T T T T H H H T T T H

- Person 2:

H T H T T T H H H T H T H T T T T H H T T H T T H
H T T H T H H T T T H H H T T T T H H T H T H T H



Agenda

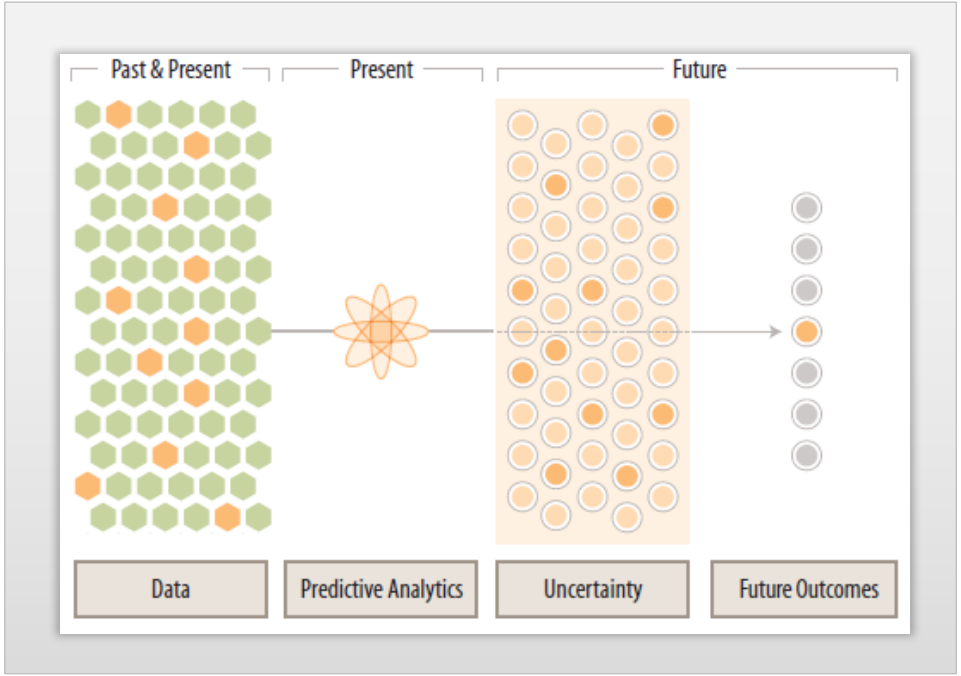
- Predictive Analytics: What is it and how can you use it?
- Using Predictive Analytics to Identify Sales Suppression
- Using Predictive Analytics to Identify Refund Fraud
- Questions

Predictive Analytics

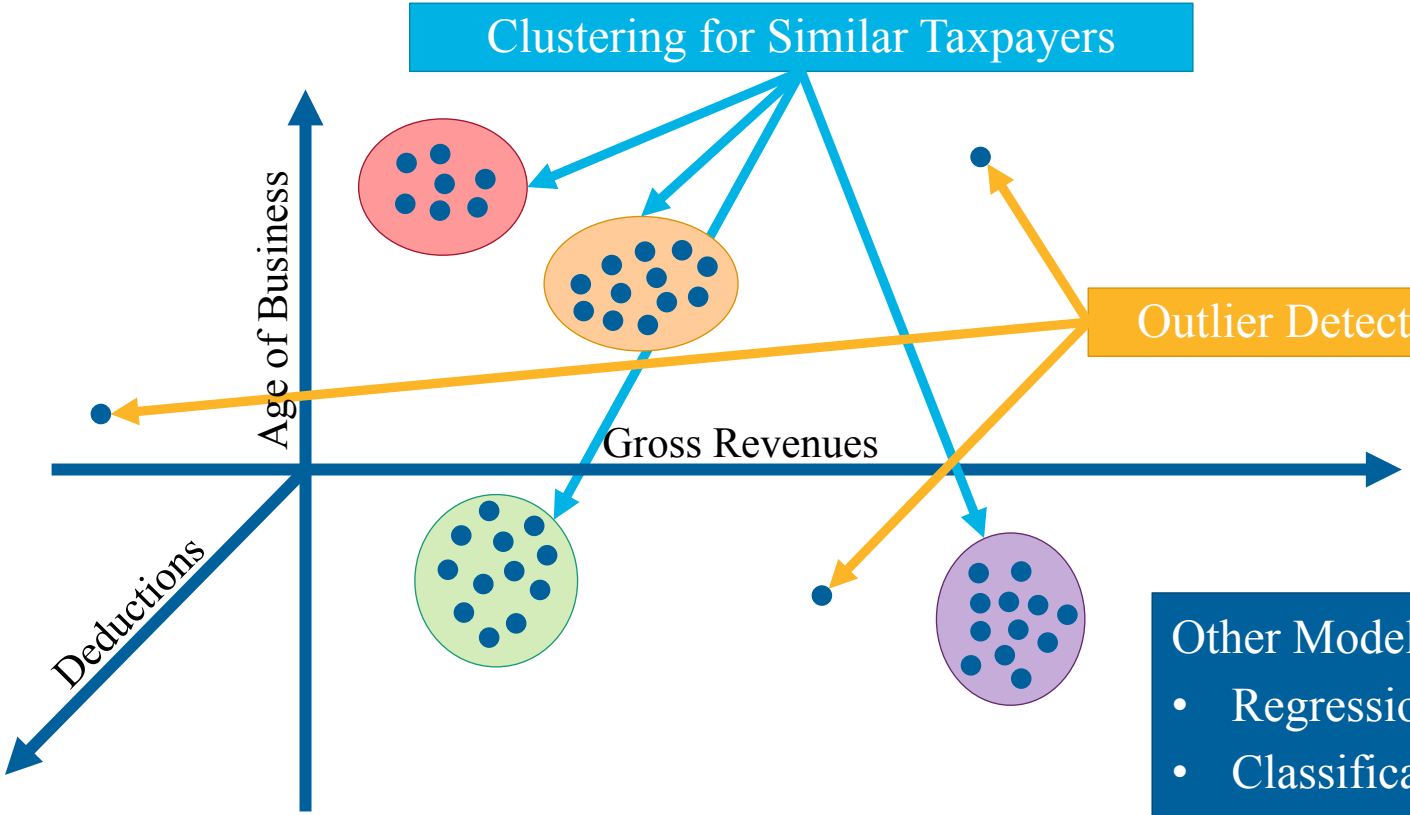
What it is

Predictive Analytics: What it is?

- Predictive Analytics uses **mathematical techniques deriving insight** from data to find the best action for a given situation
- Increase the **precision**, **consistency** and **speed** of decisions
- Identify **patterns** in data that might otherwise remain hidden
- Tax agencies are data rich, which is helpful to build and power models



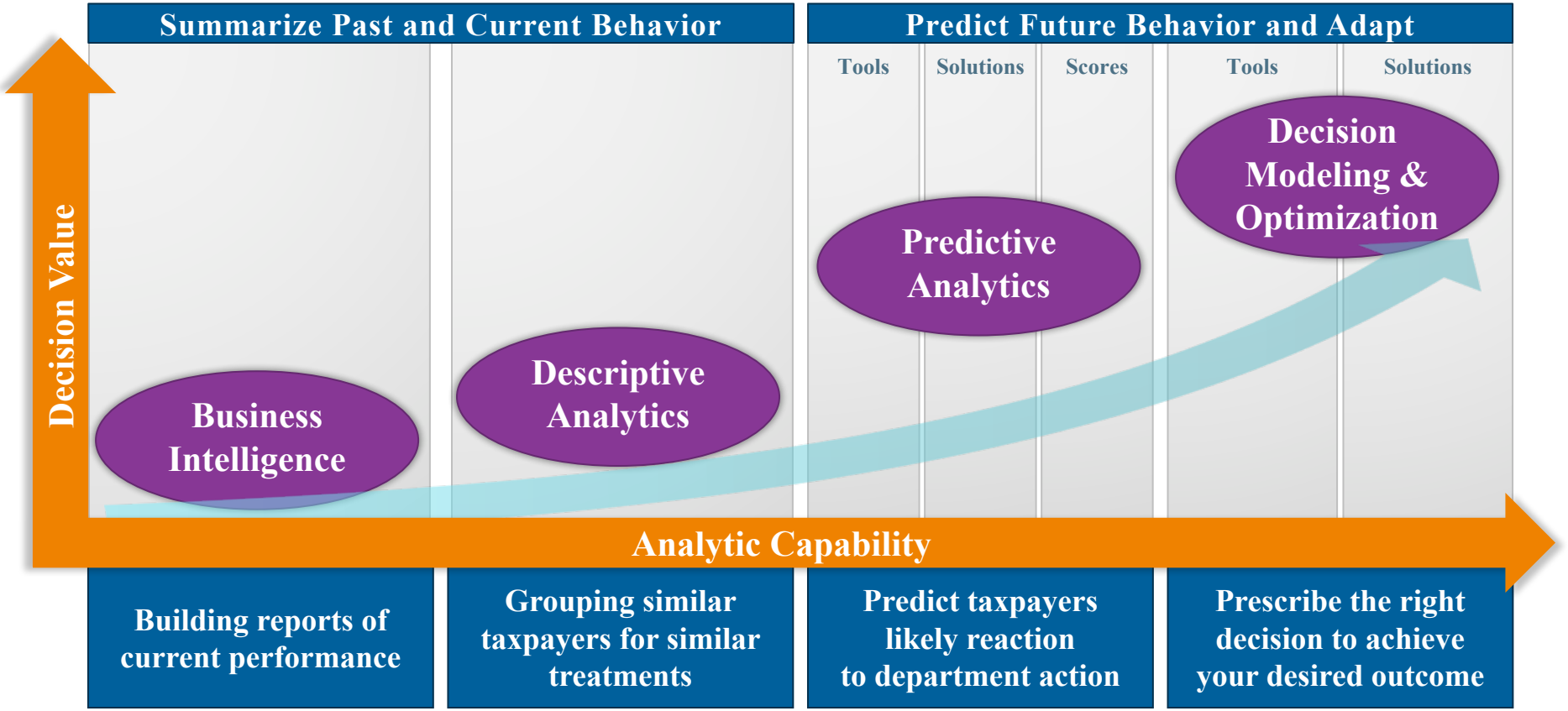
Many types of Predictive Models



Other Models Types:

- Regression Models
- Classification Models
- Association Models

Data Analytics Maturity Curve of Effectiveness



Using Predictive Analytics to Identify Sales Suppression

Sales Suppression is well documented in the News

April 14, 2017: Everett, WA	Everett Man Sent To Prison For Selling “Tax Zapper” Software
April 21, 2017: Long Island, NY	Current, former restaurant owners charged with tax fraud
April 25, 2017: West Bloomfield, MI	Sushi Restaurant Owners Charged with Tax Fraud, Using 'Zapper' to Hide Income
May 16, 2017: Wythe County, VA	Five Wythe residents plead guilty in Old Fort tax evasion scheme
June 19, 2017: Muncie, IN	Muncie restaurant owners sentenced in $1.8M tax evasion case
June 20, 2017: Greenville, SC	Bar owner arrested for sales tax evasion
July 2, 2017: Mankato, MN	Sushi restaurant investigated for tax fraud

Challenges Identifying Sales Suppression

- Harder to find issues that aren't there than analyzing issues in front of you
- By the time the Audit starts, sales have been suppressed
  - Deleted
  - Zapped
  - Never Entered
  - Destroyed (SSaaS)
- Cash businesses make it harder to follow the money
- Systemic under-reporting can be part of the merchant business model
- **A 2013 Estimate pegged losses to States at $21 Billion/Year**

Analytics can help Identify Suppression

- Whether using skimming, non-reporting or zapping, businesses leave a digital footprint detectable by advanced analytics
- Predictive models can identify:
  - Irregularities in taxpayer data
  - Irregularities in peer-to-peer comparisons of similar businesses
- A statistical analysis can virtually prove an under-reporting condition has occurred
- An agency can identify sales suppression using modeling:
  - Increasing auditor efficiency
  - Increasing audit accuracy
  - Increasing defendability of the assessment

Finding sales suppression using analytics

ECR/POS Systems

Sales Suppression Statistics

- Supervised and Unsupervised Analytics
- Machine Learning and Artificial Intelligence
- Statistical Anomaly Detection

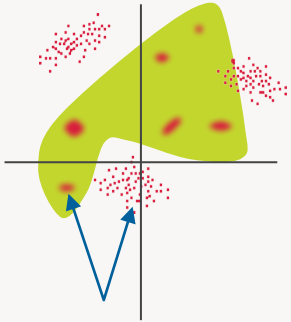
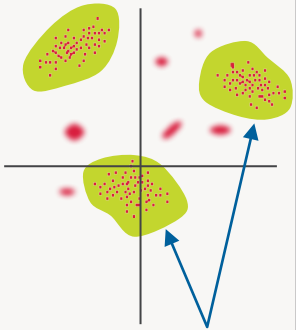
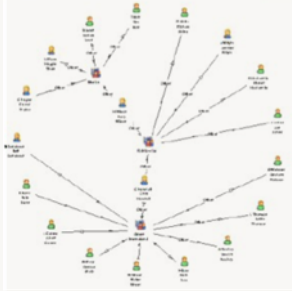
Statistical Results

- Exceptions identified with associated reason codes
- Estimate of the amount of sales suppression
- Comparisons to statistically derived industry norms
- Details on business analyzed

Statistical Analyses can be used

- Outlier models
- Scorecards
- Clustering techniques
- Decision Trees
- Neural networks

Data is Key
Modeling uses data obtained before and during the audit to perform analysis

Fraud-trained	Outlier Detection	SNA
Learn complex relations from identified fraud in historical data, then leverage this to predict fraud	Uncover new types of fraud by identifying outlier and aberrant behavior	Network Analysis for identity and entity resolution
		
Use tags to differentiate the innocent from the guilty	Learn patterns and identify aberrance	Connects fraud rings across seemingly unconnected data by detecting and linking identity information

Using Analytics to Identify Refund Fraud

Fraud detection is rarely 100% certain

Form 1 **Individual Wage Tax**

Your first name and initial (if joint return, also give spouse's name and initial) Last name	Your social security number
Sweet Polly Purebred	
Home address (number and street including apartment number or rural route)	Spouse's social security number
111 Never Tell a Lie Street	
City, town, or post office, state and ZIP code	Your occupation / Spouse's occupation
Safety Harbor, FL 34695	

1	Wages, Salary and Pensions	1	$50,000
2	Personal allowance	2	$0
3	Number of dependents, not including spouse	3	
4	Personal allowances for dependents *(line 3 multiplied by $6,000)*	4	
5	Total personal allowances *(line 2 plus line 4)*	5	$0
6	Taxable wages *(line 1 less line 5, if positive: otherwise zero)*	6	
7	Tax	7	$10,000
8	Tax already paid	8	$8,000
9	Tax due *(line 7 less line 8, if positive)*	9	$2,000
10	Refund due *(line 8 less line 7, if positive)*	10	

Form 1 **Individual Wage Tax**

Your first name and initial (if joint return, also give spouse's name and initial) Last name	Your social security number
Simon Bar Sinister	
Home address (number and street including apartment number or rural route)	Spouse's social security number
666 I Am Evil Lane	
City, town, or post office, state and ZIP code	Your occupation / Spouse's occupation
Hell, MI 48169	

1	Wages, Salary and Pensions	1	$50,000
2	Personal allowance	2	$49,999
3	Number of dependents, not including spouse	3	
4	Personal allowances for dependents *(line 3 multiplied by $6,000)*	4	
5	Total personal allowances *(line 2 plus line 4)*	5	$49,999
6	Taxable wages *(line 1 less line 5, if positive: otherwise zero)*	6	
7	Tax	7	$1
8	Tax already paid	8	$800,000,000
9	Tax due *(line 7 less line 8, if positive)*	9	
10	Refund due *(line 8 less line 7, if positive)*	10	$799,999,999

Fraud Detection: A Careful Balancing Act

Pick too many returns as potential fraud:

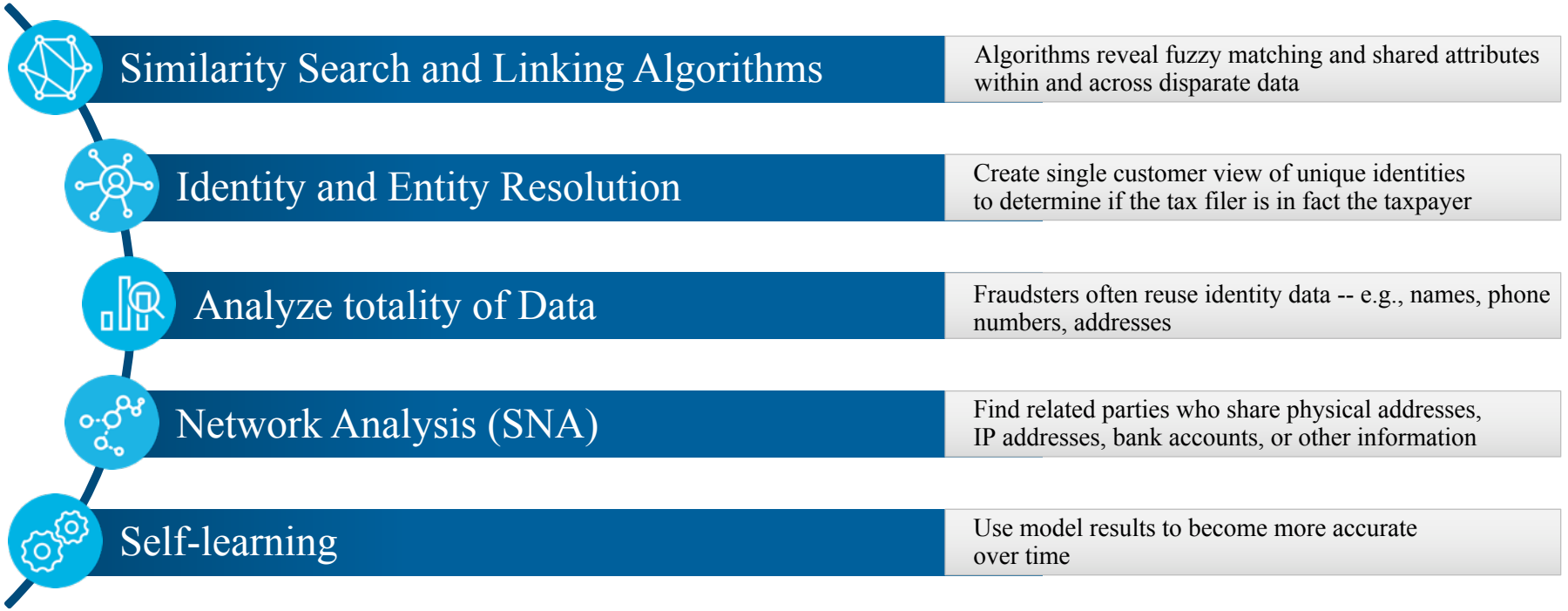
- Delay refunds
- Impact Customer Service
- Impact Department Staff load

Pick too few returns as potential fraud:

- Too much money lost

Not all returns created equal – requires a careful balancing act

Statistical approaches to identify refund fraud



Characteristics of Data Used in Identity Resolution

Personally Identifying Information (PII)

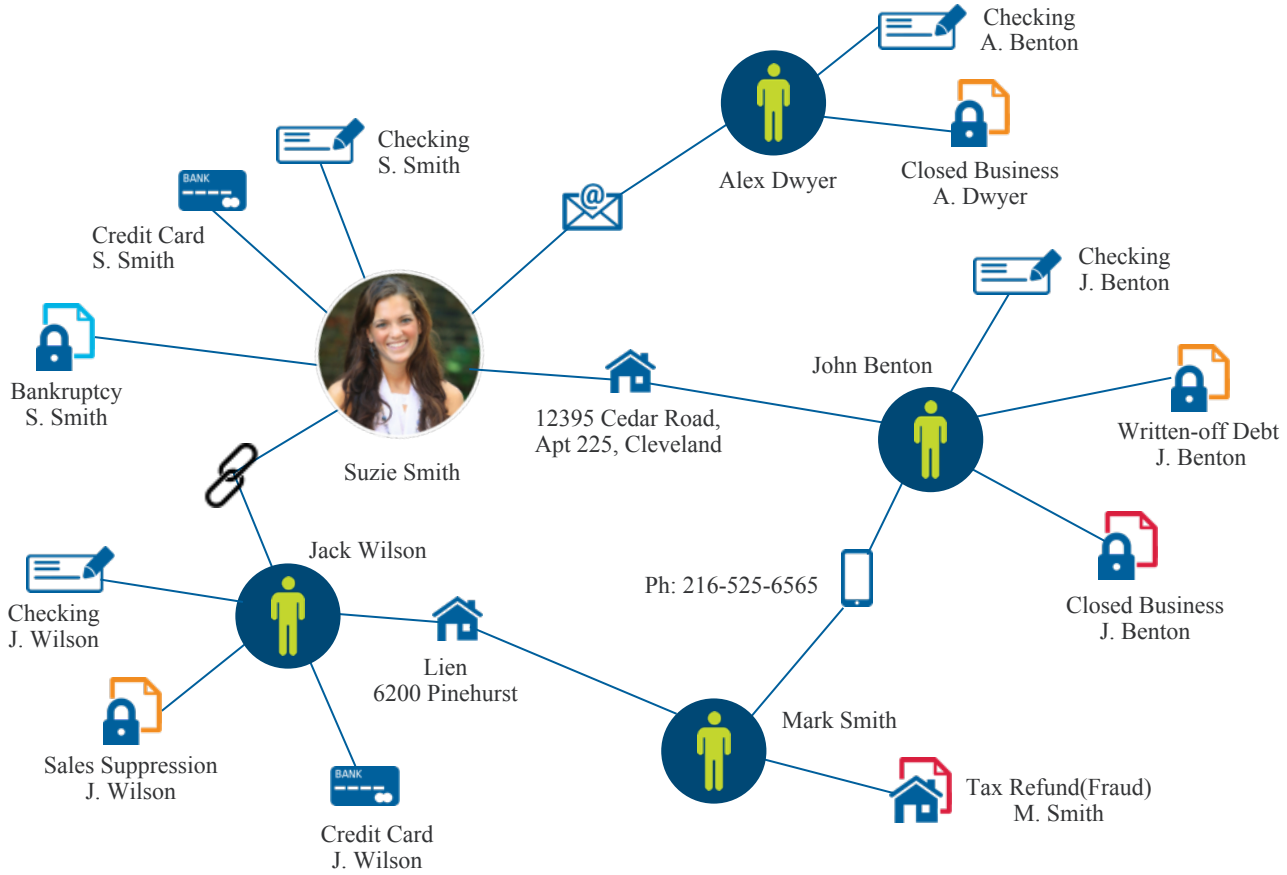
- Name, DOB, SSN, address, phone number, email, online logins, device ID, IP address account numbers and more
  - Structured format
  - Un-hashed to maintain the integrity of linking across shared attributes

Current and Historical Data Across Disparate Sources

- Internal
  - Registration Information
  - Hotlists / Watchlists
  - Known Fraud Lists
- Third Party
  - Subscription Based
  - External information managed in-house; e.g., GSA Excluded Parties List System (EPLS)

- Establish true identities: “who’s who”
- Discover hidden relationships/networks: “who knows whom”

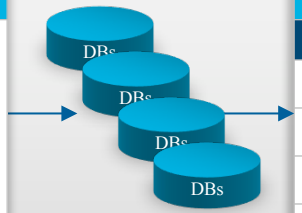
Link Analysis



Building a Model Score for Identity Resolution

- Searches across and within disparate data sources
- Search algorithms can be tuned to the nuances of your data and your appetite for fuzzy matching
  - Fraud detection is never 100% accurate: how many false positives can you tolerate?
  - Models tuned up or down during the tax year as you review results
- Would not require data cleansing, ETL or warehousing, eliminating problems associated with data normalization

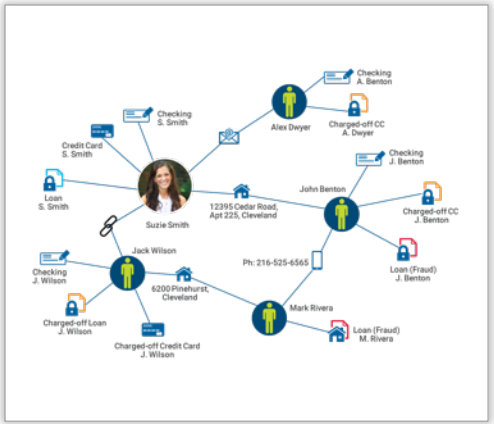
Search Request	
Full Name	John Smythe
Address	325 Mast Ct
City	Midland
State	MI
Country	USA
DOB	9/22/1966
DL #	981347823



Match Results				
Match	Full Name	Address	DL #	Ref #
99%	Jon Smythe	325 Mast Ct	981347823	345333
96%	John Smythe	335 Mast Apt 202	981347820	656966
95%	Jonathan Smith	202 – 325 Mast	981347820	965233
92%	John Smith	352 Mast Ct	981347832	966898

Social Network Analysis

Utilizing Network Level Analytics for Proactive Fraud Detection



Persisted Network

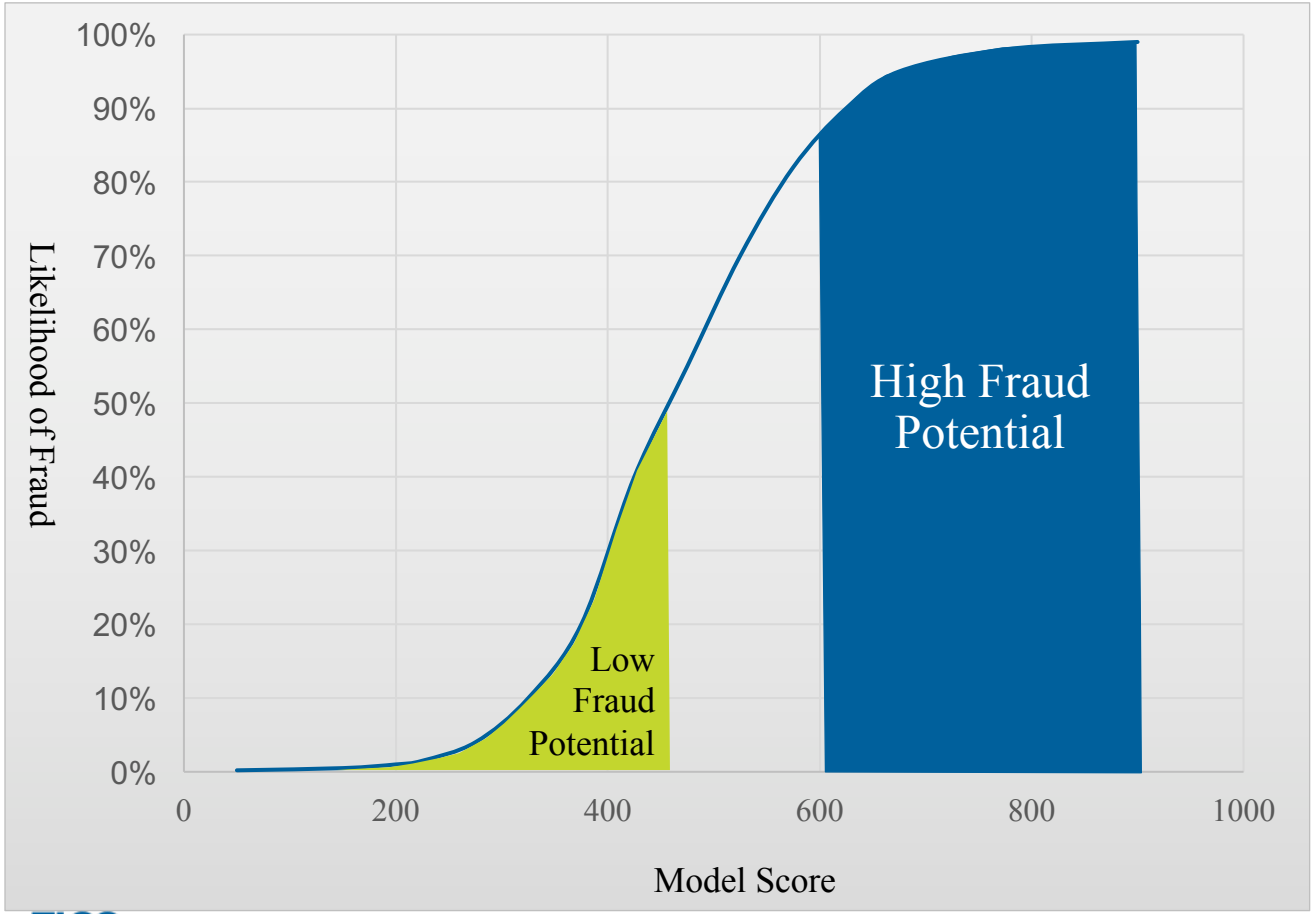
Variable	Value
Count of participants	5
Count of tax accounts	26
Sum in Collections	1
Tax Account velocity	7.56
Number of connections to investigations	1
	…

Network Variables

Analytic Result	Risk Score
Score from Returns	+ 163
Adjustment from related individuals within network	+ 131
Adjustment for high velocity of new accounts	+ 220
Adjustment for open investigations within the network	+ 180

Predictable Insight

Goals for Scoring Returns



Goals:

- Flexibility to redefine scores that will get extra attention as workloads are reviewed
- Machine learning to identify new schemes during the filing season
- Automated self-service capabilities for suspect taxpayers to self-verify their identity

Summary

- Data Analytics and predictive modeling: proven tools that identify fraud and save money
- Predictive models can support tax agencies many ways:
  - Sales Suppression: Help auditors identify missing data, resulting in more accurate and defensible audits
  - Refund Fraud: Review and compare data to identify likely fraud situations

About the Presenter

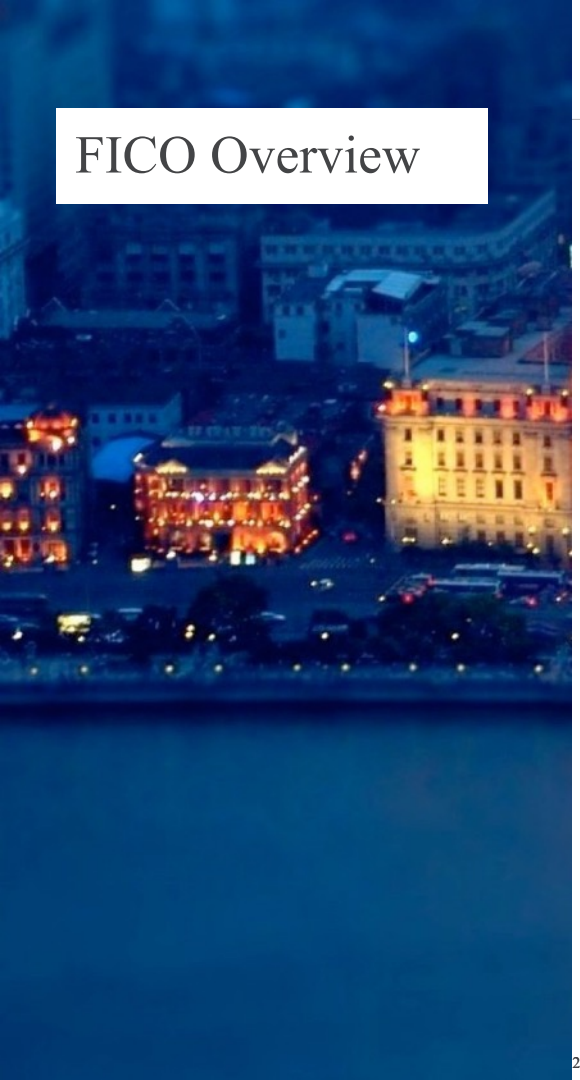
Ted London

Partner, Government Solutions

(916) 634-5179

tedlondon@fico.com

- 24 Years working with Federal, State and Local government agencies
- Experience with more than 20 different tax agencies worldwide
- Skilled in enhancing collections, audit and fraud systems and business processes
- Experience with predictive modeling and behavioral science techniques to enhance collections



FICO Overview

Profile	The leader in analytic solutions for risk management, fraud, and customer engagement Founded: 1956 NYSE: FICO Revenues: $839 million (fiscal 2015)
Products and Services	Pioneers at transforming Data into insights to help organizations achieve their mission FICO® Score and other models for making decisions 130+ patents in analytic and decision management technology, with an additional 90+ patents pending Analytic applications for collections, fraud, customer service and cybersecurity
Clients and Markets	10,000+ clients in 90+ countries Industry focus: Banking, government, insurance, retail, health care
Recent Rankings	#1 in services operations analytics (IDC)* #4 in worldwide analytics software (IDC)* #8 in Business Intelligence, CPM and Analytic Applications (Gartner)** #26 in the FinTech 100 (*American Banker*)
Offices	20+ offices worldwide, HQ in San Jose, California 2,900 employees Regional Hubs: New York, San Diego, Fairfax, London, Birmingham (UK), Johannesburg, Milan, Moscow, Munich, Madrid, Istanbul, Sao Paulo, Bangalore, Beijing, Singapore

*IDC, *Worldwide Business Analytics Software 2013-2017 Forecast and Vendor Shares*, June 2013.
**Gartner, *Market Share Analysis: Business intelligence, Analytics and Performance Management, 2012,Dan* Sommer & Bhavish Sood, May 7, 2013.

Thank you

Ted London
Partner, Government Solutions
tedlondon@fico.com
(916) 634-5179