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 T T T T T T T H

Person 2:
H T H T T T H H H T H T T T T T T H H T T T T T H H H T T H T 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 H T T H T H H H H H H H H T T H T T H H H H H H H T T T
H H T T H T H H H H H H H H H T T T T T T T T T T T T T H

Person 2:
H H T T T T H H H H T T H T T H H H H H T T T T T T T T T T T T T H
H H H H H H H H H H H H H H H H H H H H T T T T T T T T T T T T 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

- Clustering for Similar Taxpayers
- Outlier Detection

Other Models Types:
- Regression Models
- Classification Models
- Association Models
Data Analytics Maturity Curve of Effectiveness

**Summarize Past and Current Behavior**
- Business Intelligence
- Descriptive Analytics

**Predict Future Behavior and Adapt**
- Predictive Analytics
- Decision Modeling & Optimization

**Analytic Capability**
- Building reports of current performance
- Grouping similar taxpayers for similar treatments
- Predict taxpayers likely reaction to department action
- Prescribe the right decision to achieve your desired outcome
Using Predictive Analytics to Identify Sales Suppression
Sales Suppression is well documented in the News

<table>
<thead>
<tr>
<th>Date</th>
<th>Location</th>
<th>News Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 14, 2017</td>
<td>Everett, WA</td>
<td>Everett Man Sent To Prison For Selling “Tax Zapper” Software</td>
</tr>
<tr>
<td>April 21, 2017</td>
<td>Long Island, NY</td>
<td>Current, former restaurant owners charged with tax fraud</td>
</tr>
<tr>
<td>April 25, 2017</td>
<td>West Bloomfield, MI</td>
<td>Sushi Restaurant Owners Charged with Tax Fraud, Using 'Zapper' to Hide Income</td>
</tr>
<tr>
<td>May 16, 2017</td>
<td>Wythe County, VA</td>
<td>Five Wythe residents plead guilty in Old Fort tax evasion scheme</td>
</tr>
<tr>
<td>June 19, 2017</td>
<td>Muncie, IN</td>
<td>Muncie restaurant owners sentenced in $1.8M tax evasion case</td>
</tr>
<tr>
<td>June 20, 2017</td>
<td>Greenville, SC</td>
<td>Bar owner arrested for sales tax evasion</td>
</tr>
<tr>
<td>July 2, 2017</td>
<td>Mankato, MN</td>
<td>Sushi restaurant investigated for tax fraud</td>
</tr>
</tbody>
</table>
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
Learn complex relations from identified fraud in historical data, then leverage this to predict fraud

Outlier Detection
Uncover new types of fraud by identifying outlier and aberrant behavior

SNA
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

<table>
<thead>
<tr>
<th>Form 1</th>
<th>Individual Wage Tax</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sweet Polly Purebred</strong></td>
<td></td>
</tr>
<tr>
<td>111 Never Tell a Lie Street</td>
<td></td>
</tr>
<tr>
<td>Safety Harbor, FL 34695</td>
<td></td>
</tr>
<tr>
<td>1 Wages, Salary and Pensions</td>
<td>$50,000</td>
</tr>
<tr>
<td>2 Personal allowance</td>
<td>$0</td>
</tr>
<tr>
<td>3 Number of dependents, not including spouse</td>
<td></td>
</tr>
<tr>
<td>4 Personal allowances for dependents (line 3 multiplied by $6,000)</td>
<td></td>
</tr>
<tr>
<td>5 Total personal allowances (line 2 plus line 4)</td>
<td>$0</td>
</tr>
<tr>
<td>6 Taxable wages (line 1 less line 5, if positive: otherwise zero)</td>
<td></td>
</tr>
<tr>
<td>7 Tax</td>
<td>$10,000</td>
</tr>
<tr>
<td>8 Tax already paid</td>
<td>$0</td>
</tr>
<tr>
<td>9 Tax due (line 7 less line 8, if positive)</td>
<td>$8,000</td>
</tr>
<tr>
<td>10 Refund due (line 8 less line 7, if positive)</td>
<td>$2,000</td>
</tr>
</tbody>
</table>

| **Simon Bar Sinister** |
| 666 I Am Evil Lane |
| Hell, MI 48169 |
| 1 Wages, Salary and Pensions | $50,000 |
| 2 Personal allowance | $0 |
| 3 Number of dependents, not including spouse | |
| 4 Personal allowances for dependents (line 3 multiplied by $6,000) | |
| 5 Total personal allowances (line 2 plus line 4) | $0 |
| 6 Taxable wages (line 1 less line 5, if positive: otherwise zero) | |
| 7 Tax | $0 |
| 8 Tax already paid | $1 |
| 9 Tax due (line 7 less line 8, if positive) | $800,000,000 |
| 10 Refund due (line 8 less line 7, if positive) | $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

- **Similarity Search and Linking Algorithms**: Algorithms reveal fuzzy matching and shared attributes within and across disparate data.
- **Identity and Entity Resolution**: Create single customer view of unique identities to determine if the tax filer is in fact the taxpayer.
- **Analyze totality of Data**: Fraudsters often reuse identity data -- e.g., names, phone numbers, addresses.
- **Network Analysis (SNA)**: Find related parties who share physical addresses, IP addresses, bank accounts, or other information.
- **Self-learning**: Use model results to become more accurate over time.
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”
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

<table>
<thead>
<tr>
<th>Match</th>
<th>Full Name</th>
<th>Address</th>
<th>DL #</th>
<th>Ref #</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>Jon Smythe</td>
<td>325 Mast Ct</td>
<td>981347823</td>
<td>345333</td>
</tr>
<tr>
<td>96%</td>
<td>John Smythe</td>
<td>335 Mast Apt 202</td>
<td>981347820</td>
<td>656966</td>
</tr>
<tr>
<td>95%</td>
<td>Jonathan Smith</td>
<td>202 – 325 Mast</td>
<td>981347820</td>
<td>965233</td>
</tr>
<tr>
<td>92%</td>
<td>John Smith</td>
<td>352 Mast Ct</td>
<td>981347832</td>
<td>966898</td>
</tr>
</tbody>
</table>
Social Network Analysis

Utilizing Network Level Analytics for Proactive Fraud Detection

Persisted Network

Network Variables

Predictable Insight

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of participants</td>
<td>5</td>
</tr>
<tr>
<td>Count of tax accounts</td>
<td>26</td>
</tr>
<tr>
<td>Sum in Collections</td>
<td>1</td>
</tr>
<tr>
<td>Tax Account velocity</td>
<td>7.56</td>
</tr>
<tr>
<td>Number of connections to</td>
<td>1</td>
</tr>
<tr>
<td>investigations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analytic Result</th>
<th>Risk Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score from Returns</td>
<td>+ 163</td>
</tr>
<tr>
<td>Adjustment from related individuals within network</td>
<td>+ 131</td>
</tr>
<tr>
<td>Adjustment for high velocity of new accounts</td>
<td>+ 220</td>
</tr>
<tr>
<td>Adjustment for open investigations within the network</td>
<td>+ 180</td>
</tr>
</tbody>
</table>
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 |

---

Thank you

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