Data Mining/Fraud Detection

April 28, 2014
Jonathan Meyer, CPA
KPMG, LLP
Agenda

• Overview of Data Analytics & Fraud
• Getting Started with Data Analytics
• Where to Look & Why?
• What is Possible?
D&A Business Value – 5 Points:

- Business has **fundamentally changed** – Harness data to drive growth and profitability now.

- A data-driven organization embeds the ability to harness data in **every aspect of its business**.

- Data is **raw material** that becomes valuable through the use of analytics.

- Not simply technology - need a **business problem** to zero in on the right data and apply the right analytics.

- Balance new opportunities vs new risks – while data can be a valuable asset, it pose new risks.
Data Analytics & Fraud

Effectively identify anomalies, patterns and trends to prevent, detect and respond to fraud risks

<table>
<thead>
<tr>
<th>Prevent</th>
<th>Detect</th>
<th>Respond</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Proactive forensic data analysis</td>
<td>• Retrospective forensic data analysis</td>
<td>• Internal investigation protocols</td>
</tr>
<tr>
<td>• Component of active compliance programs</td>
<td>• Due diligence that employees and business partners are not engaged in inappropriate activities</td>
<td>• Enforcement and accountability protocols</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Disclosure protocols</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Remedial action protocols</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Analytics Maturity Continuum

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactive</td>
<td>Ad-hoc</td>
</tr>
<tr>
<td></td>
<td>Largely Ad-hoc Application of rules to detect known patterns of fraudulent activity and anomaly detection for unknown patterns; includes queries, drill downs, alerts</td>
</tr>
<tr>
<td>Hindsight</td>
<td>Reactive Data Mining</td>
</tr>
<tr>
<td>Insight</td>
<td>Prospective</td>
</tr>
<tr>
<td>Foresight</td>
<td>Predictive</td>
</tr>
<tr>
<td></td>
<td>Sustainable Pattern recognition and prospective management, machine learning and neural networks</td>
</tr>
<tr>
<td>Compliance</td>
<td>Reactive</td>
</tr>
<tr>
<td>Strategic</td>
<td>Largely Ad-hoc Application of rules to detect known patterns of fraudulent activity and anomaly detection for unknown patterns; includes queries, drill downs, alerts</td>
</tr>
<tr>
<td></td>
<td>Repeatable into Continuous Management of known suspicious activities/claimants/vendors</td>
</tr>
<tr>
<td></td>
<td>Continuous into Sustainable Identification of unusual trends, false claims and statistical anomalies</td>
</tr>
<tr>
<td></td>
<td>Sustainable Pattern recognition and prospective management, machine learning and neural networks</td>
</tr>
</tbody>
</table>

We mean business in higher education
Data Analytics Methodology

PLANNING
- Preparation of Data Request for Data Extract
- Conduct Meetings with Key Stakeholders/IT
- Obtain Commitment on Receipt of Data
- Assist with Extract (If Necessary)

• Receive Data Extract
• Test Data Completeness & Accuracy
• Load Data into Processing Tool
• Adjust Data for Specific Analytics to be Performed
• Process Data & Refine Analytics

• Discussion of Analytic Capabilities
• Agreement on Budget, Scope and Delivery
• Preparation of Project Plan
• Schedule Process Owner Interviews

• Report on Findings
• Identify Additional Benefits of the Analysis
• Review Results with Project Stakeholders
• Review with Board or Investigation Team

• Kick-off Discussion
• Risk Assessment & Planning Discussion

• Preparation of Data Request for Data Extract
• Conduct Meetings with Key Stakeholders/IT
• Obtain Commitment on Receipt of Data
• Assist with Extract (If Necessary)
Sustainable D&A

- Insights and decisions become data-driven, enhancing the traditional gut instinct
- Actions may become automated based on models, increasing agility and repeatability
Why Use Data Analytics?

- Evaluate all transactions, not just samples
- Cross reference datasets
  - Financial vs. Nonfinancial
- Uncover novel schemes or new risk areas
- Granular detail
Getting Started
## Getting Started with Data Analytics

### Get Organized
- What information is available?
- What software to use?

### Set Priorities
- Understand unique risks
- Select initial focus area

### Start Small
- Develop rules and routines
- Visualize results

### Build on Experience
- Advanced statistics
- Predictive or dynamic scoring

---

Manage the complexity
Get Organized

• Is information readily available in standard formats?
  – CSV extracts/ODBC connections

• Can information in non-financial systems be accessed?
  – HR records
  – Travel history

• Software
  – Excel?
  – ACL/IDEA
  – Should have procedure logging
  – Ability to connect various data sources
Set Priorities

• What are the potential risk areas based on your organization?
  – Travel & Entertainment
  – P-Card
  – Cash donations/payments
  – Procurement Fraud
  – Misuse of Grants or donations funds
  – Construction projects

• Select focus areas to start
  – Data is available, high quality and consistent
  – T&E is usually a good place
Start Small

• Develop rules and routines
  – Tests
  – Thresholds
  – If “X” and “Y > 500” = flagged transactions

• Understand your software and data capabilities

• Visualize the results
  – Interactive visualizations are analytics
  – Ability to summarize large & diverse data
Build on Experience

• Advanced data analytics requires significant understanding of your unique data and systems.
• Using Advanced statistics
  – Detect anomalies and outliers
  – Identify high-risk transactions without Pre-defined rules
• Predictive capabilities
  – Building models requires “training” data
  – Transaction scoring should adjust to reduce false positives
Biggest Challenges

• Data Availability
  – Getting info out of transaction systems and into analytical software
  – Permission to access data from other departments

• Too Much Data – Too Little Granularity?
  – Signal v Noise
  – Where to start?

• Data Quality
  – Significant challenges in data cleansing and normalization
  – Getting consistency from different systems

• Unstructured data
  – Email
  – Manual expense reports
  – Free Form descriptions
<table>
<thead>
<tr>
<th>Issue</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master data records distributed across multiple systems</td>
<td>Inconsistent and duplicate data leads to complex maintenance processes</td>
</tr>
<tr>
<td></td>
<td>increasing the total cost of ownership</td>
</tr>
<tr>
<td>Inconsistent business intelligence information</td>
<td>Missed cost savings opportunities in the absence of EIM</td>
</tr>
<tr>
<td>Quality of master data is poor</td>
<td>Inconsistent representation of data throughout the enterprise</td>
</tr>
<tr>
<td>• No clear set of standards</td>
<td>Significant resources expended making point updates</td>
</tr>
<tr>
<td>• Incomplete and inconsistent descriptions</td>
<td>Rework due to bad data</td>
</tr>
<tr>
<td>• High number of duplicate data elements</td>
<td></td>
</tr>
<tr>
<td>• Inaccurate data</td>
<td></td>
</tr>
<tr>
<td>Decentralized data management with no enterprise wide plan</td>
<td>Conflicting policies lead to inconsistent creation/updates of master data</td>
</tr>
<tr>
<td></td>
<td>elements complicating maintenance and increasing the total cost of</td>
</tr>
<tr>
<td></td>
<td>ownership</td>
</tr>
</tbody>
</table>

We mean business in higher education
Where to Start?
Where to Look

• Lots of areas including:
  – Travel & Entertainment (Expenses)
  – P-Card
  – Procurement
  – Manual journal entries
  – Construction Spend
  – Grant Compliance
  – Payroll
  – Anywhere money / benefits / materials / anything of value leaves your organization.
Travel and Entertainment (T&E)

Data Required:
- Travel Data
- Expense Reports

Key Considerations
- Expenses & travel often in hardcopy
- Lack of employee entered descriptions
- Need granular/itemized detail

Potential Fraud Risks and Schemes
- Fictitious vendor
- Fictitious receipt
- Inflated expenses/mileage
- Duplicated or split expenses
- Purchases for personal use or using personal credit card
- Use of unapproved travel agency
- Inappropriate entertainment venues
- Bribery
- Ticket exchange

Unusual Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor</td>
<td>$1,800,000</td>
</tr>
<tr>
<td>Electronic</td>
<td>$1,600,000</td>
</tr>
<tr>
<td>Healthcare Related</td>
<td>$1,400,000</td>
</tr>
<tr>
<td>Automotive</td>
<td>$1,200,000</td>
</tr>
<tr>
<td>Jewelry</td>
<td>$1,000,000</td>
</tr>
<tr>
<td>Personal Expenses</td>
<td>$800,000</td>
</tr>
<tr>
<td>Adult Entertainment</td>
<td>$600,000</td>
</tr>
<tr>
<td></td>
<td>$400,000</td>
</tr>
<tr>
<td></td>
<td>$200,000</td>
</tr>
</tbody>
</table>
Common T&E Tests

• Travel non-standard locations
• Vendor name is unusual for business purpose
• Expenses incurred while on vacation/holiday (x-ref with HR data)
• Same meal claimed by multiple employees to circumvent policy spend (same vendor, same date, different employees)
• Airfare with no meals or hotel expense (possibly refunded for employee’s personal use)
• Mileage claimed along with gas or rental car
• High hotel spend vs. others in same geography (possible policy violation or items charged to hotel room)
• High volumes to common vendors = opportunity to negotiate better prices through standard procurement.
What Can I do Today in T&E

Start capturing the right data.

- **Who:** Attendees or Vendor
- **What:** Expense type
- **Where:** Geographic Location
- **Why:** Business purpose

Require all employees to enter this information
Where to Look – P-Card

Potential Fraud Risks and Schemes
• False reimbursement submissions
• Inflated reimbursement submissions
• Purchases for personal use
• Returns of purchases for cash
• Expense splitting/combining
• Rolling corporate card balance

Number of transactions used for Cash Advance

99.1%
0.9%

Number of Payments Recurring Each Month

98.4%
1.6%
Common P-Card Tests

- Reoccurring (direct debit) transactions
- Late night timestamps
- Expenses incurred while on vacation/holiday (x-ref with HR data)
- Merchant codes not customary for business
- Multiple transactions in short time frame (2 charges within 3 minutes)
- Employees with charges greater then reimbursements = using P-card for personal use
What Can I do Today in P-Card

Same at T&E -

• Who: Attendees or Vendor
• What: Expense type
• Where: Geographic Location
• Why: Business purpose

But

• Require itemized entries – not simply “hotel” expenses
Detecting Outliers

- Most “data mining” is outlier / pattern detection
- Items without a “neighbor”
- Scatter Plots / Box Charts
- Mathematics & Statistics
  - X std devs from mean
  - 1.5x Interquartile range
  - Machine Learning
  - Advanced algorithms
What’s Possible?
Geospatial Analysis

Combine transactional data with maps & demographic data
=> insightful conclusions
Generic Application Design

Share Folder

Tools
- Database load Scripts
- Data Pre-Processing Scripts

Initial

Data – Initial Period
- Expense data
- Hierarchies & Maps
- Procurement data
- Other

Data – Period X
- Incremental Expense data
- Replacement Hierarchies & Maps
- Incremental Procurement data
- Replacement/Incremental Other

Data Transformation and Load

Reformat Files
(Data Pre-Processing Scripts)

Load to raw_tables
(Database load Scripts)

Prepare dash_tables
(stored procedures)

Visualization Tool
Database

raw_1

dash_2

We mean business in higher education
Example: Business Purpose

EXPENSES BY REPORTED EXPENSE TYPE

<table>
<thead>
<tr>
<th>Meals</th>
<th>Conference</th>
<th>Gifts</th>
<th>Education</th>
<th>Miscellaneous</th>
<th>Other</th>
<th>Lodging</th>
<th>Airfare</th>
</tr>
</thead>
</table>

SPEND BY SPECIFIC MCC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$0</td>
<td>$50,000</td>
<td>$100,000</td>
<td>$150,000</td>
<td>$200,000</td>
<td>$250,000</td>
<td>$300,000</td>
<td>$350,000</td>
</tr>
</tbody>
</table>

MOST FREQUENT BUSINESS PURPOSE TERMS

<table>
<thead>
<tr>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
<th>Term 4</th>
<th>Term 5</th>
<th>Term 6</th>
<th>Term 7</th>
<th>Term 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Fiscal period
Business Unit
Expense Type
Domestic / Foreign
Personal vs. Business
G/L Code
Supervisor
Range of Amount (USD)
Example: Splits & Duplicates

Filters
- Fiscal period
- Business Unit
- Expense Type
- Personal vs. Business
- G/L Code
- Supervisor
- Range of Amount (USD)
Example: Outliers

### T&E Spend by City by Type

- **Moscow**: $8,000,000
- **Tokyo**: $7,000,000
- **Singapore**: $6,000,000
- **Hong Kong**: $5,000,000
- **Geneva**: $4,000,000
- **Zurich**: $3,000,000
- **Bern**: $2,000,000
- **Sydney**: $1,000,000
- **London**: $0
- **Shanghai**: $0

### Lodging vs. Meals vs. Transport

- **Average**: Lodging: $2,000, Meals: $1,000, Transport: $500
- **Minimum**: Lodging: $0, Meals: $0, Transport: $0
- **Maximum**: Lodging: $5,000, Meals: $5,000, Transport: $5,000
- **TX**: Lodging: $1,000, Meals: $1,000, Transport: $1,000

### No. Employees with Meals > $100 USD

- Less than 10: 385
- 11 to 50: 149
- 51 to 100: 72
- More than 100: 26

---

**CASH VS CREDIT**

<table>
<thead>
<tr>
<th>No. T&amp;E Txn</th>
<th>% Cash</th>
<th>% Credit</th>
<th>T&amp;E in USD</th>
<th>% Cash</th>
<th>% Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith, C</td>
<td>258</td>
<td>24%</td>
<td>76%</td>
<td>$29,398</td>
<td>7%</td>
</tr>
<tr>
<td>Jordan, D</td>
<td>573</td>
<td>24%</td>
<td>76%</td>
<td>$32,398</td>
<td>5%</td>
</tr>
<tr>
<td>Chu, R</td>
<td>121</td>
<td>21%</td>
<td>79%</td>
<td>$3,298</td>
<td>1%</td>
</tr>
<tr>
<td>Lee, M</td>
<td>56</td>
<td>17%</td>
<td>83%</td>
<td>$62,583</td>
<td>23%</td>
</tr>
<tr>
<td>White, S</td>
<td>478</td>
<td>15%</td>
<td>85%</td>
<td>$7,387</td>
<td>3%</td>
</tr>
<tr>
<td>Parsons, T</td>
<td>1,132</td>
<td>15%</td>
<td>85%</td>
<td>$229,283</td>
<td>40%</td>
</tr>
<tr>
<td>Khahn, V</td>
<td>422</td>
<td>14%</td>
<td>86%</td>
<td>$54,348</td>
<td>18%</td>
</tr>
<tr>
<td>Summers, Y</td>
<td>98</td>
<td>12%</td>
<td>88%</td>
<td>$17,348</td>
<td>10%</td>
</tr>
<tr>
<td>Purdue, W</td>
<td>2,324</td>
<td>8%</td>
<td>92%</td>
<td>$153,348</td>
<td>33%</td>
</tr>
</tbody>
</table>

---

**Outlier View**

- **TOP 10 Hotel Cities – Nightly Rate Spread**
  - Moscow, Tokyo, Singapore, Hong Kong, Geneva, Zurich, Bern, Sydney, London, Shanghai

- **Filters**
  - Country
  - CPI Score
  - Fiscal period
  - Business Unit
  - Product
  - Expense Type
  - Domestic / Foreign
  - Personal vs. Business
  - G/L Code
  - Supervisor
  - Range of Amount (USD)
Thank You / Questions

Jonathan Meyer
jonathanmeyer@kpmg.com
(312) 665-1996