BIG DATA
Big Data - what's the Big Deal?

and what it means to Paychex

Risk Analytics
Providing the Power of Knowing
Paychex Predictive Analytics

Speaking Appearances (Domestic and International)
what’s the Buzz?

In 2009, Google processed 24 petabytes of data each day
Or think of it this way:

1. 🌍 = 1.25 terabytes
What’s the Buzz?

In 2009, Google processed 24 petabytes of data each day.

Or think of it this way:

1. \( \text{person} = 1.25 \text{ terabytes} \)

So

24 petabytes = 20k

This is enough text to fill up 2 billion filing cabinet drawers daily!
Over One Year...

To the Moon and Back.... Lite
Over One Year...

To the Moon and Back.... Literally!

In 2012 alone, 2.5 quintillion (that's 18 zeros) bytes of data was generated.

Or the equivalent of 24 trillion minutes of music
Big Data Intelligence Competitive Advantage

Proactive Decision Making
- What's the best that can happen?
- What will happen next?
- What if these trends continue?
- Why is this happening?

Reactively Decision Making
- What actions are needed?
- Where exactly is the problem?
- How many, how often, where?
- What happened?

Degree of Intelligence

Competitive Advantage
Where You See Predictive Modeling

Ever wonder how Amazon can make recommendations to you?

Or how your credit score works?

Or how your bank or credit card company can flag certain transactions as being potentially fraudulent?

Many use the collected data and statistical analyses to take the ‘guess-work’ out of guessing.
Solving Problems with Data

Customer Behavior

Customer Profile

Predictive Model

Business Response
**LIFC CYCLE**

- **SaRAH**: Predicts sales units for quota purposes. Prior 2 Fiscal Year predictions within 0.5% of company total.

- **3B**: Territory optimization model. Assisted Core Sales in creating and placing an additional 35 head count for FY13.

- **TIM**: New business sales model consisting of a 4-week trial resulting in over 200 units to date.

- **SAM**: Sales Reps in bucket ‘F’ are ~2x more likely to turnover than the standard population and ~5x more likely than reps in bucket ‘A’.

- ‘A’ clients provide a 1.4x lift over a random selection.
3B

SaRAH

PST 1000

50% improvement in NSS dial:appt ratio. To date, has accounted for over $11.5M in revenue.

FY13 results show ‘A’ clients 7x more likely to buy than ‘C’ clients

‘A’ clients provide a 1.4x lift over a random selection

‘A’ client 15x more likely to purchase than ‘F’ clients

Price elasticity gauge used as a guide for annual price increase and understanding of PPS discounting

PAUL

PEG

Workers Comp

STEPH

PATRICK

A leads are 1.5x more likely to move to an approved quote than a random selection of clients
Combined with OLIVER, PAM drives the RTS which is used for client retention strategies throughout the country.

‘Bucket 10’ clients return 3x more frequently than the standard population. Used to auto decision payrolls based on liability and model score.

Price elasticity gauge used as a guide for annual price increase and understanding of PPS discounting.

OLIVER PEG Discounting

Key model in both discounting and client retention systems. Clustering changes should account for close to $2.3M in additional revenue.

A leads are 1.5x more likely to move to an approved quote than a random selection of clients.
Eliminated 30% of CCS workload by sacrificing only a small portion of the collectible dollars

Collection model used as a benchmark for AR collections activity

Key Lifetime Value model used in both discounting and client retention systems

Combined with OLIVER, PAM drives the RTS which is used for client retention strategies throughout the country.

Price elasticity gauge used as a guide for annual price increase and understanding of PPS discounting

‘Bucket 10’ clients return 3x more frequently than the standard population. Used to rate decision.
BIG DATA
Paychex Model Portfolio

**Up-Sell / Cross-Sell Models**

- **SAM**
  Predicts the number of units per zip code

- **TIM**
  Optimization model, representative of the inherent opportunity within a zip code.

- **Sarah**
  Employee Retention model

- **MMS Territory**
  MMS Sales Strategy model aimed to identify overall opportunity per territory

**Predictive Models**

- **OLIVER**
  Core Advanced lifetime value model predicts both revenue and retention. OLIVER 2.0 in development

- **PEG**
  Clustering model used to gauge the price sensitivity of each client. PEG 2.0 in development

- **PAM**
  Client retention model geared towards controllable losses

- **MMS Retention**
  MMS controllable retention model

**Sales Strategy**

- **Portfolio Credit Model**
  Identifies clients most likely to return within the next 3 months, for controllable reasons.

- **MARCO**
  Logistic model identifies which accounts are most likely to liquidate within 90 days

- **M aria**
  Scores past due invoices based on probability of turning 60 days past due

**Risk Models**

- **Client Referral**
  Prioritizes NSS lead call queues based on propensity for booking an appointment

- **NSS Target**
  Prioritizes NSS lead call queues based on propensity for booking an appointment

- **3B Refresh**
  Captures newly-formed business information and predicts which are most likely to buy +50 sales

- **Next Best Offer**
  Determines which product a new Core client is most likely to purchase next

**Operations Models**

- **Usage**
  Field Ops

- **HRS Sales**

- **Marketing**

- **National Sales Support**

- **Risk Management**

- **Sales**

- **Human Resources**
Advanced cross-sell model aimed to identify clients most likely to purchase H&B, up to 16x more likely to churn.

OLIVER
Core Advanced life time value model predicts both revenue and retention. OLIVER 2.0 in development.

PEG
Clustering model used to gauge the price sensitivity of each client. PEG 2.0 in development.

PAM
Client retention model geared towards controllable losses.

MMS Retention
MMS controllable retention model.

Usage:
- Field Ops
- HRS Sales
- Marketing
- National Sales Support
- Risk Management
- Sales
- Human Resources
Model Development
Development

Today

Sample set snapshot of

Evaluate client behavior at some time interval after snapshot

Model predictive behavior sample
Sample set snapshot of client base; gather MANY variables
Model Development

Today

Evaluate client behavior at some time interval after snapshot

Sample set snapshot of client base; gather MANY variables

Model allows us to predict future client behavior based on how sample set behaved
BIG DATA
OLIVER Validation

OLIVER Loss Rates

Lower expected loss rate identification in ‘A’ and ‘B’ buckets

Higher expected loss rate in critical ‘D’ and ‘F’ buckets

AVERAGE LOSS RATE
Putting this intelligence in your hands - MDDS
Putting this intelligence in your hands - MDDS
Retirement Tracking System

Even more info

---

### ABC FROYO SHOPPE - 11111111

<table>
<thead>
<tr>
<th>CLIENT CONTACT INFORMATION</th>
<th>Contact Name:</th>
<th>Tim Jones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phone # &amp; Email:</td>
<td>(585) 111-5555</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:tjones@froyo.com">tjones@froyo.com</a></td>
</tr>
</tbody>
</table>

### QUICK SUMMARY

- **Length of Service**: 8 month(s)
- **Average Employee Count**: 15
- **Payroll Frequency**: Weekly

- **3 months revenue (actual)**: $330.00
- **Estimated annualized revenue** *: $1,300.00

### CURRENT OLIVER SCORE

- 10: Low Revenue, High Probability of Leaving
- 9: 10D (Lowest revenue generators, Bottom 10%)
- 8: D (High probability of leaving for a controllable or uncontrollable reason)
- 7: C (High probability of leaving)
- 6: B (Low Probability of Leaving)
- 5: A (Low Probability of Leaving)

### CURRENT PEG SCORE

- 10: High Sensitivity to Price
- 9: 5 (Very sensitive to a change in price)
- 8: 4 (Moderate sensitivity)
- 7: 3 (Low sensitivity to price)
- 6: 2 (Low sensitivity to price)

### RETENTION STRATEGY NOTES

| Load Date: | 8/29/2013 |

This client has no history in the Retention Tracking System.

- **Strategy (Primary)**:
- **Strategy (Secondary)**:
- **Notes**:

*Revenue figures are estimated using the OLIVER model.*
ABC FROYO SHOPPE - 11111111

Length of Service - 8 month(s)  Average Employee Count - 15  Payroll Frequency - Weekly

3 months revenue (actual) - $330.00  Estimated annualized revenue* - $1,300.00

**CURRENT OLIVER SCORE**

<table>
<thead>
<tr>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Revenue</td>
<td>D</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td>Low Probability of Leaving</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CURRENT PEG SCORE**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Sensitivity to Price</td>
<td>The Original OLIVER Score - LDR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What OLIVER means for this client
OLIVER Score = 10D
10 = Lowest revenue generators (Bottom 10%)
D = High probability of leaving for a controllable or uncontrollable reason

What PEG means for this client
PEG Cluster = 5
Very sensitive to a change in price

The Original OLIVER Score - LDR
This is the score the client had during the original load. It is possible that at the time the client's score has shifted.

Strategy (Primary):  Strategy (Secondary):  Notes:

*Revenue figures are estimated using the OLIVER model

Completed (Strategies Applied)
BIG DATA
**Where We Can Go**

**What we’re doing now**

- Likely to Stay
- Likely to Leave

**With Increased Usage and Data Input**

Not only will we be able to provide analysis on the most effective strategies, we’ll also be able to identify the target clients better to optimize efforts.
Where We Can Go

What we’re doing now

Likely to Stay
Likely to Leave

Not only will we be able to provide analysis on the most effective strategies, we’ll also be able to identify the target clients better to optimize efforts.

With Increased Usage and Data Input

Likely to Leave
Likely to Stay

PERSUADABLES

“Lost Causes”

“Sure Things”

“Sleeping dogs”
Is the RTS Working?

In Region 3, an analysis of their FY13 Retention program shows that clients in the RTS that had a strategy applied had a much lower loss rate than clients without a strategy applied.
Is the RTS Working?

In Region 3, an analysis of their FY13 Retention program shows that clients in the RTS that had a strategy applied had a much lower loss rate than clients without a strategy applied.

<table>
<thead>
<tr>
<th>Strategy Applied</th>
<th>No Strategy Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.7%</td>
<td>25.2%</td>
</tr>
</tbody>
</table>
# Region 3 Attrition Rates

(Client Base @11/30/2012)  
(Four months ended 3/31/13)

<table>
<thead>
<tr>
<th></th>
<th>R3 Retention/RTS Program</th>
<th>RTS Non Worked</th>
<th>Non RTS Payroll Base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loss Rate</td>
<td>Loss Rate</td>
<td>Loss Rate</td>
</tr>
<tr>
<td>Metro NY</td>
<td>6.1%</td>
<td>19.9%</td>
<td>6.1%</td>
</tr>
<tr>
<td>NNJ</td>
<td>8.2%</td>
<td>15.5%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Mid Hudson</td>
<td>6.2%</td>
<td>11.9%</td>
<td>5.5%</td>
</tr>
<tr>
<td>SNJ</td>
<td>2.3%</td>
<td>21.2%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Philly</td>
<td>6.7%</td>
<td>15.8%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Lehigh</td>
<td>2.7%</td>
<td>26.3%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Harrisburg</td>
<td>8.7%</td>
<td>16.2%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Pittsburg</td>
<td>6.2%</td>
<td>11.0%</td>
<td>5.9%</td>
</tr>
<tr>
<td><strong>Total Region 3</strong></td>
<td><strong>5.8%</strong></td>
<td><strong>17.8%</strong></td>
<td><strong>6.6%</strong></td>
</tr>
</tbody>
</table>

Note: Difference in April between the total worked RTS and Non Worked moved to 18.5% (6.7% vs 25.2%)
Region 3 Attrition Rates
(Client Base @11/30/2012)
(Four months ended 3/31/13)

<table>
<thead>
<tr>
<th></th>
<th>R3 Retention/RTS Program</th>
<th>RTS Non Worked</th>
<th>Non RTS Payroll Base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loss Rate</td>
<td>Loss Rate</td>
<td>Loss Rate</td>
</tr>
<tr>
<td>Metro NY</td>
<td>6.1%</td>
<td>19.9%</td>
<td>6.1%</td>
</tr>
<tr>
<td>NNU</td>
<td>8.2%</td>
<td>15.5%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Mid Hudson</td>
<td>6.2%</td>
<td>11.9%</td>
<td>5.5%</td>
</tr>
<tr>
<td>SNI</td>
<td>2.3%</td>
<td>21.2%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Philly</td>
<td>6.7%</td>
<td>12.8%</td>
<td>6.4%</td>
</tr>
<tr>
<td>Lehigh</td>
<td>2.7%</td>
<td>26.3%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Harrisburg</td>
<td>8.7%</td>
<td>16.2%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Pittsburg</td>
<td>6.2%</td>
<td>11.0%</td>
<td>5.9%</td>
</tr>
<tr>
<td><strong>Total Region 3</strong></td>
<td><strong>5.9%</strong></td>
<td><strong>17.8%</strong></td>
<td><strong>6.6%</strong></td>
</tr>
</tbody>
</table>

Note: Difference in April between the total worked RTS and Non Worked moved to 18.5% (6.7% vs 25.2%)

R3 Return on Investment and Potential Revenue Save

- Estimated Saved Unit revenue $2,316,177
  - (Based on difference of worked (5.8%) and non worked (17.8%) attrition rates)
- Program cost $681,891
- Return on Investment $1,634,286 or 339.7%
- Potential Revenue saved if worked all RTS at same success rate $5,797,945

(Includes HRS and Payroll Revenue)
PEG Validation

PEG Loss Rates

- Higher expected loss rate in the most sensitive buckets [4 & 5]

AVERAGE LOSS RATE AFTER PRICE INCREASE
Discounted Client Mix

52% Utilization

- A: 21%
- B: 20%
- C: 19%
- D: 21%
- E: 19%

Regions that did not follow the Discount Model recommendations were discounting more evenly across all buckets, including clients most likely to stay with Paychex.

87% Utilization

- D: 24%
- F: 47%
- C: 13%
- A: 8%
- B: 8%

Regions that followed the Discount Model recommendations are discounting the clients identified as being the most likely to leave.
Discounting Validation

Loss Rates for Discount Clusters

Statistically significant difference between the loss rates of the clusters most likely to leave versus the clusters most likely to stay.
Paychex Model P

**SAM**
Predicts the number of units per zip code

**TIM**
Optimization model, representative of the inherent opportunity within a zip code.

**SaRAH**
Employee Retention model

**MMS Territory**
MMS Sales Strategy model aimed to identify overall opportunity per territory

Sales Strategy

Up-Sell /
Territory Identification & Mapping

TIM

- A purple color indicates a high opportunity for that unassigned zip code. Deeper colors signify stronger opportunities.

- White areas represent unassigned zip codes that do not fall within the top 2,000 areas of opportunity.

- A dark green area represents zip codes currently assigned to a rep.
Territory Expansion
Territory Expansion

Rich zip codes just outside current footprint represent prime opportunities to expand existing territories.

Clusters of rich zip codes represent optimal areas to establish new territories.
PATRICK
Identifies Core Advanced clients most likely to purchase 401(k). 50% improvement in dial:appt ratio

WC Cross-Sell
Cross-sell model aimed to identify clients most likely to become an approved quote.

STEPH
Core Advanced cross-sell model aimed to identify clients most likely to purchase H&B. 'A' Clients up to 16x more likely to purchase

PAUL
Core Advanced up-sell model looking for clients best suited for PEO.

PST 1000
Core Advanced cross-sell model aimed to identify clients most likely to purchase PST 1000.

SAM
Predicts the number of units per zip code

TIM
Optimization model, representative of the inherent opportunity within a zip code.

SaRAH
Employee Retention model

MMS Territory
MMS Sales Strategy model aimed to identify overall opportunity per territory

Up-Sell / Cross-Sell Models

OLIVER
Core Advanced life time value model predicts both revenue and retention. OLIVER 2.0 in development

PEG
Clustering model used to gauge the price sensitivity of each client. PEG 2.0 in development

PAM
Client retention model
PATRICK Revenue Contributions

Over $6 million in incremental revenue.
Models in Planning and Development

Cross-sell
Up-sell

Retention
Valuation

Operations

Sales
Strategy

**Next Best Offer**
Determine which product a new Core client is most likely to purchase next.

**OLIVER 3.0**
Core Advanced life-time value model predicts both revenue and retention.

**Fulfillment Optimization**
Headcount optimization model for fulfillment center operations

**Client Referral**
Identifies which clients are most likely to produce referrals

**STEPH 2.0**
Core Advanced cross-sell model aimed to identify clients most likely to purchase H&B. 'A' Clients up to 16x more likely to purchase with STEPH 1.0

**PAM 2.0**
Client retention model geared towards controllable losses.

**Management Turnover**
Employee retention model for management level staff

**NSS Target**
Prioritizes NSS lead call queues based on propensity for booking an appointment

**PAUL 3.0**
Core Advanced up-sell model looking for clients best suited for PEO.

**ASO LTV Model**
ASO life-time value model predicts both revenue and retention.

**PRS Turnover**
Employee retention model for client-facing field operations staff

**Infogroup 3B**
Captures newly-formed business information and predicts which are most likely to buy

**ASO Up-sell**
Core Advanced up-sell model looking for clients best suited for ASO.

**MARIA 2.0**
Scores past due invoices based on probability of turning 60 days past due.

**CPA Referral**
Identifies which CPA's are most likely to produce referrals and/or transfer clients to Paychex service
<table>
<thead>
<tr>
<th>Match</th>
<th>Result</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A’s vs Blue Jays</td>
<td>LOST</td>
<td>68-51</td>
</tr>
<tr>
<td>A’s vs Blue Jays</td>
<td>WON</td>
<td>69-51</td>
</tr>
<tr>
<td>A’s vs Blue Jays</td>
<td>WON</td>
<td>70-51</td>
</tr>
<tr>
<td>A’s vs White Sox</td>
<td>WON</td>
<td>71-51</td>
</tr>
<tr>
<td>A’s vs White Sox</td>
<td>WON</td>
<td>72-51</td>
</tr>
<tr>
<td>A’s vs White Sox</td>
<td>WON</td>
<td>73-51</td>
</tr>
<tr>
<td>A’s vs Indians</td>
<td>WON</td>
<td>74-51</td>
</tr>
</tbody>
</table>
Analytics is Changing the World...

Let's not be left behind.

Risk Management

Frank Fiorille
ffiorille@paychex.com
585-336-7310
Follow @ffiorille

Erika McBride
emcbride@paychex.com
585-336-7313
Follow @Mcbride_erika