wePAY

Payments for platform businesses.

QCon New York 2014

Leveraging Big Data For Payment Risk Management

John Canfield,
VP Risk Management, WePay
@JCRisk

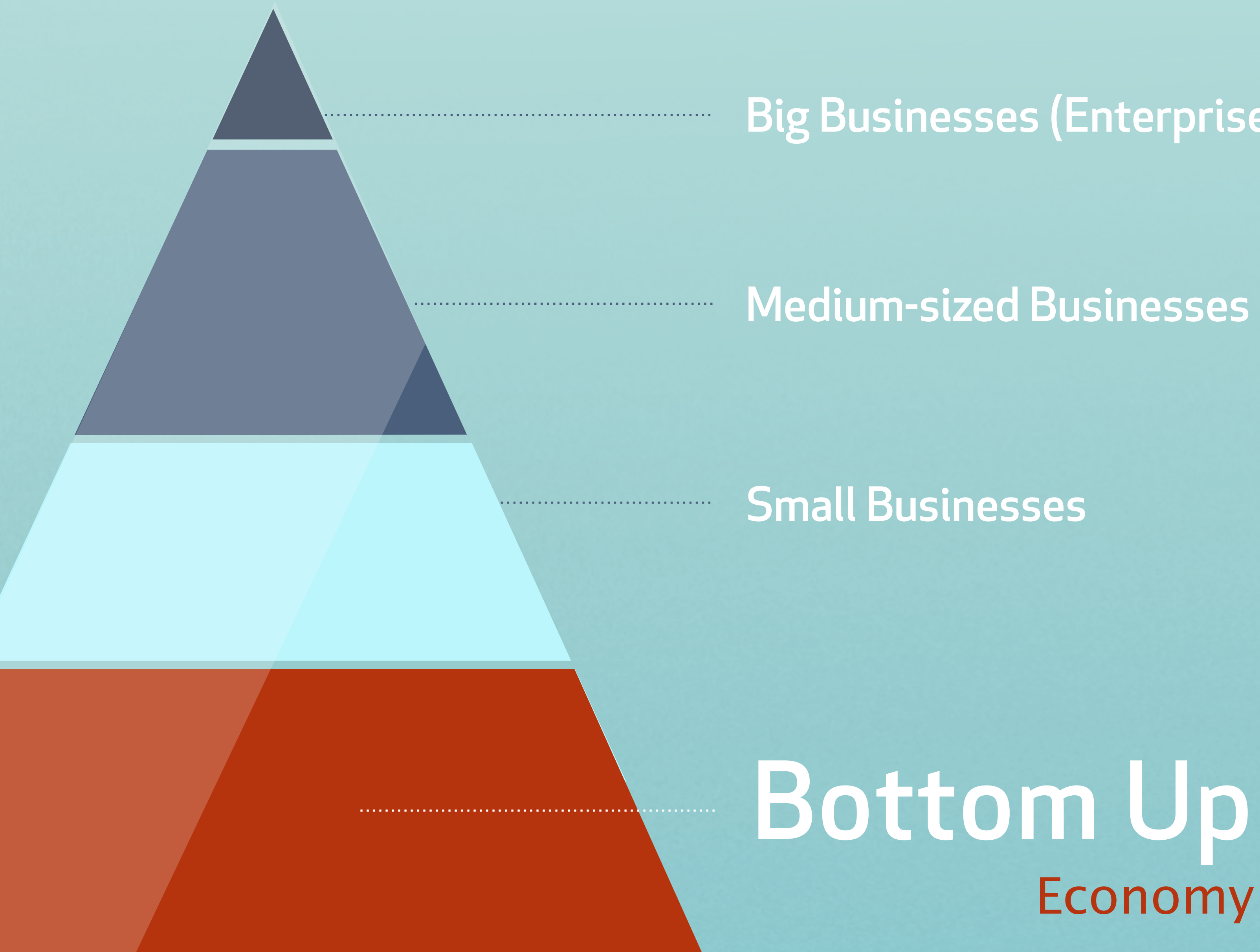
June 11, 2014

Outline

1. Payments opportunity for the bottoms-up economy
2. Requirements
3. Solution
   a) Big data collection
   b) Decisioning using machine learning, rules, and expert staff
   c) Metrics and feedback

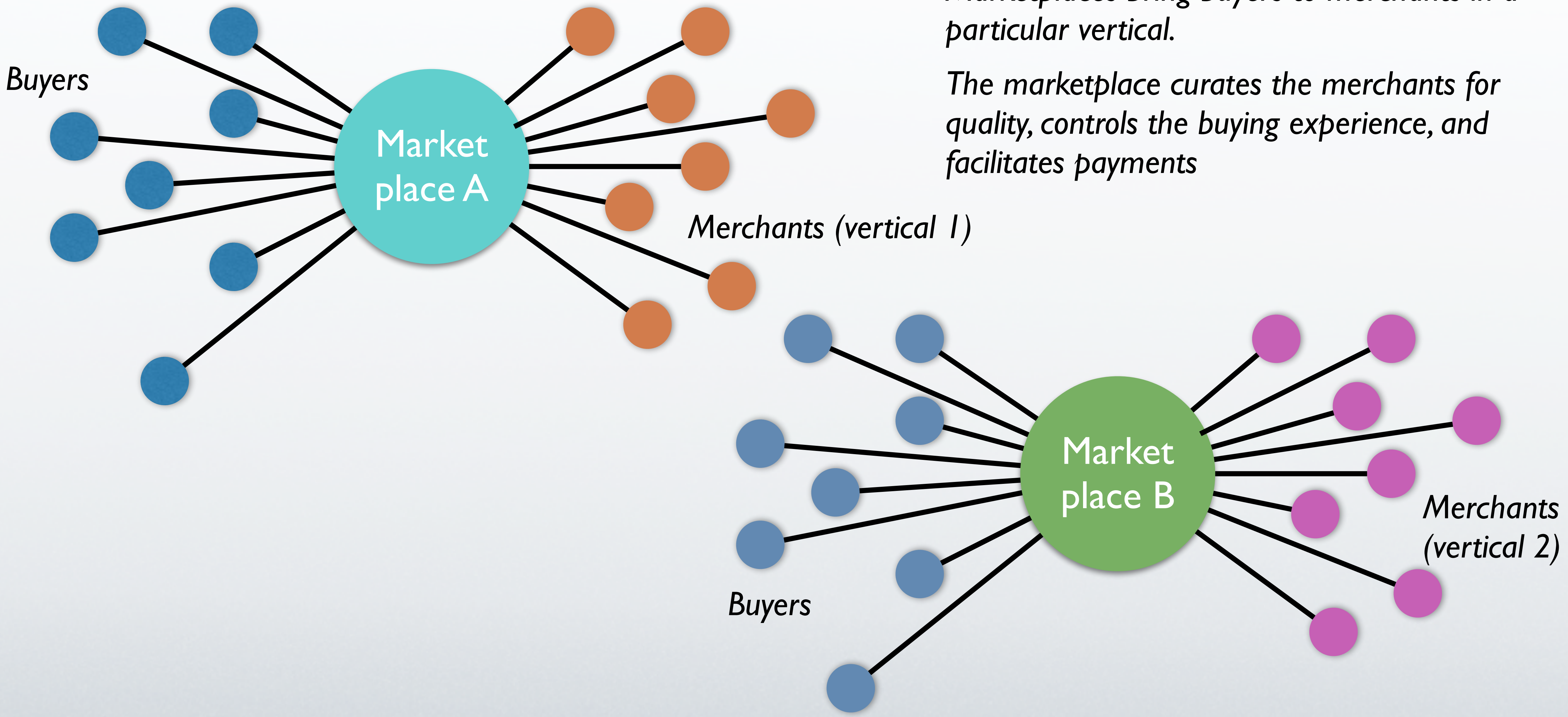

$5.00

What is the “Bottom Up” Economy?



- According to IRS data, there are over 25 million businesses in the United States with less than 5 employees.
- They collectively take in over $2 trillion annually - a tremendous opportunity for electronic payment conversion.
- Services (non-retail) is most of the market: 22m businesses & $1.7 trillion in income.

Marketplaces



Marketplaces bring buyers to merchants in a particular vertical.

The marketplace curates the merchants for quality, controls the buying experience, and facilitates payments


$10 Billion
valuation
airbnb


UBER
$18 Billion
valuation


$300M
valuation
Find great sitters and nannies near you!
Try it FREE
care.com

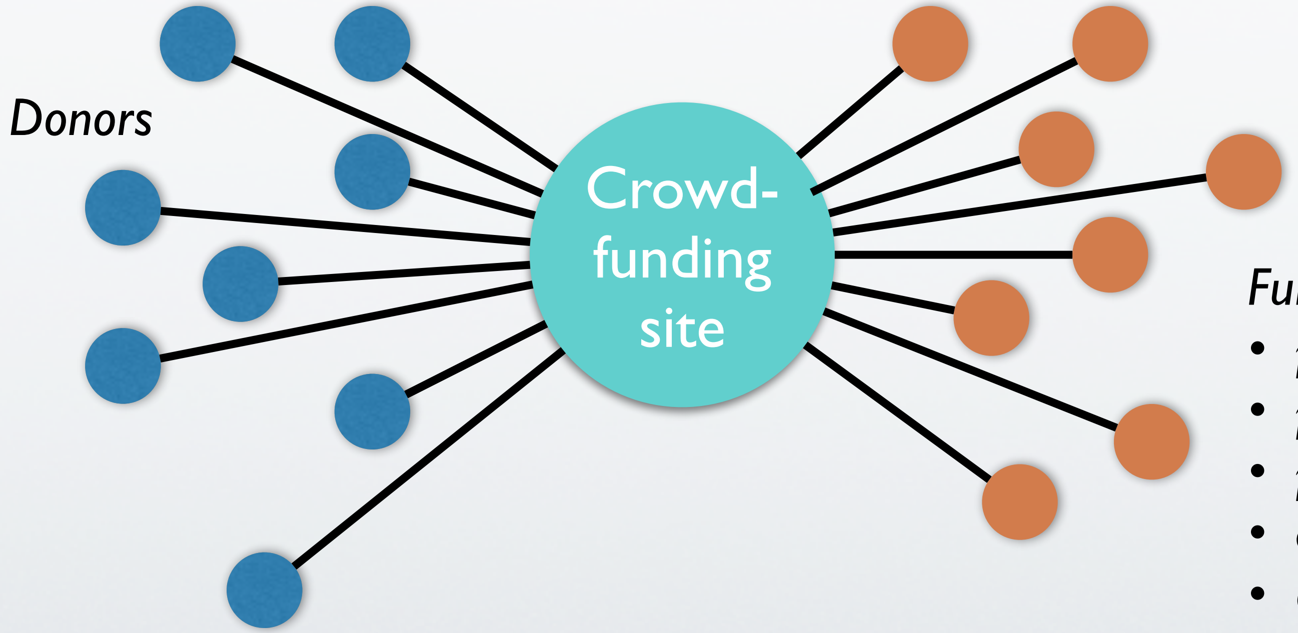

taskrabbit
GROCERY SHOPPING $25
1. Post a Task
2. TaskRabbits make offers
3. Pay when Task is done
GROCERY PICK UP!


SHARKMARX
BUY BETTER.

Crowdfunding sites

Crowdfunding sites are similar to marketplaces but instead of bringing buyers to sellers, they bring donors to fundraisers.

The crowdfunding site enables the fundraiser to market their campaign and accept payments



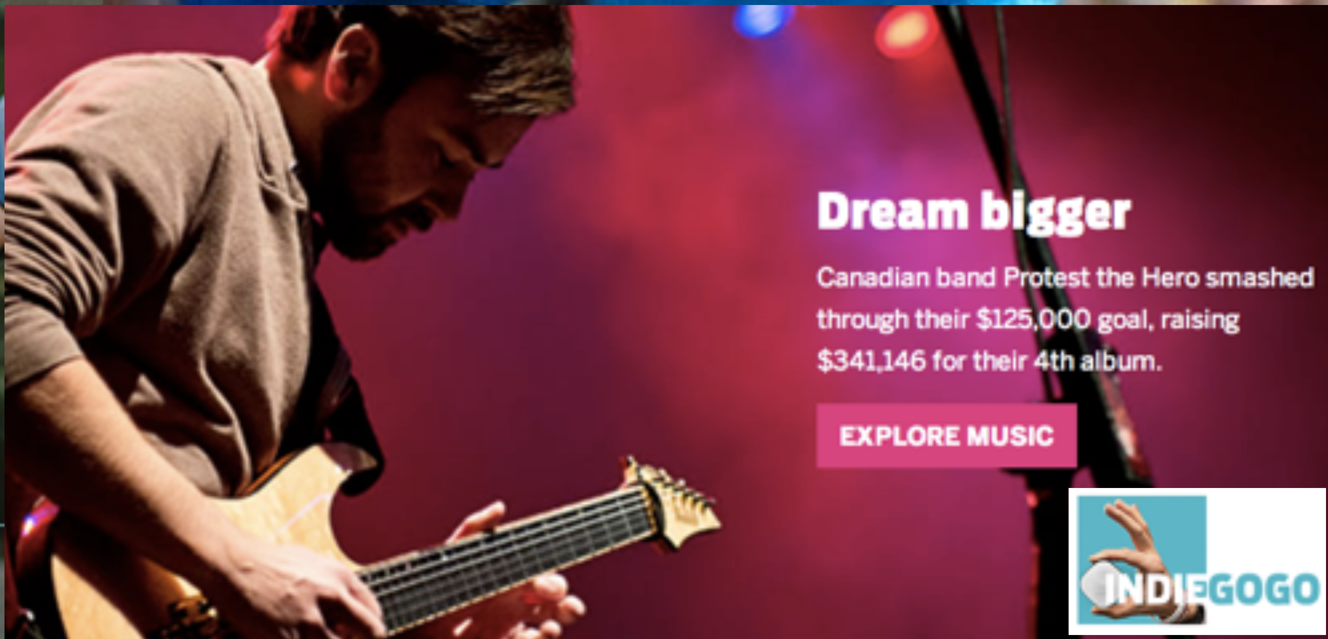
Fundraisers for

- *projects*
- *pre-orders*
- *personal*
- *charity*
- *equity - rule TBD*


KICKSTARTER
People everywhere are painting murals.
Explore public art

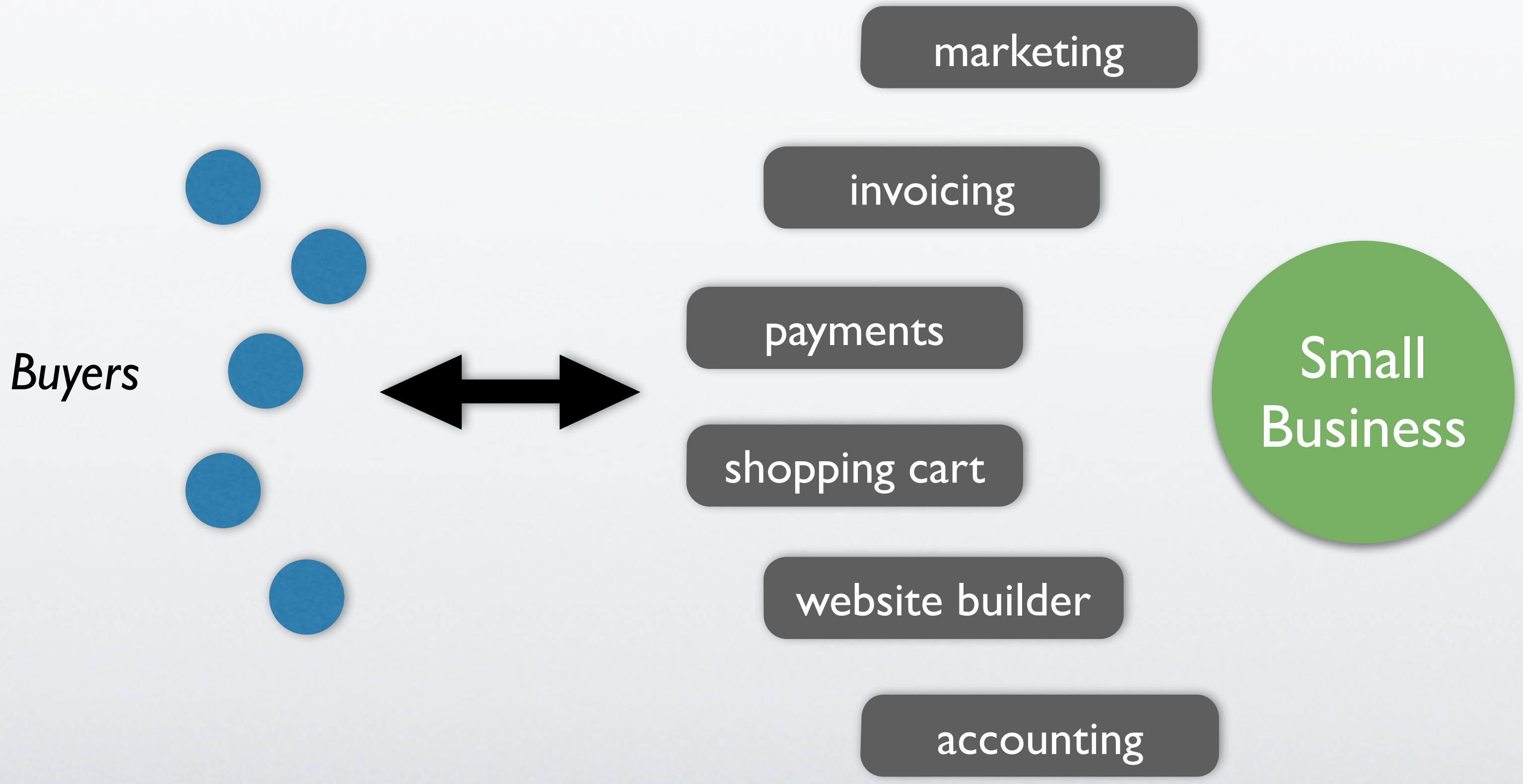

crowdrise
ROBIN HOOD


GiveForward has helped raise $103,238,000 towards medical ex
Team Ryan pulled together to raise money towards his battle with B-cell Non-Hodgkins lymphoma.
Raise more money on GiveForward.


Dream bigger
Canadian band Protest the Hero smashed through their $125,000 goal, raising $341,146 for their 4th album.
EXPLORE MUSIC
INDIEGOGO

Small Business Platforms

Small business platforms offer online services to small businesses like marketing, invoicing, shopping cart, accounting and payments

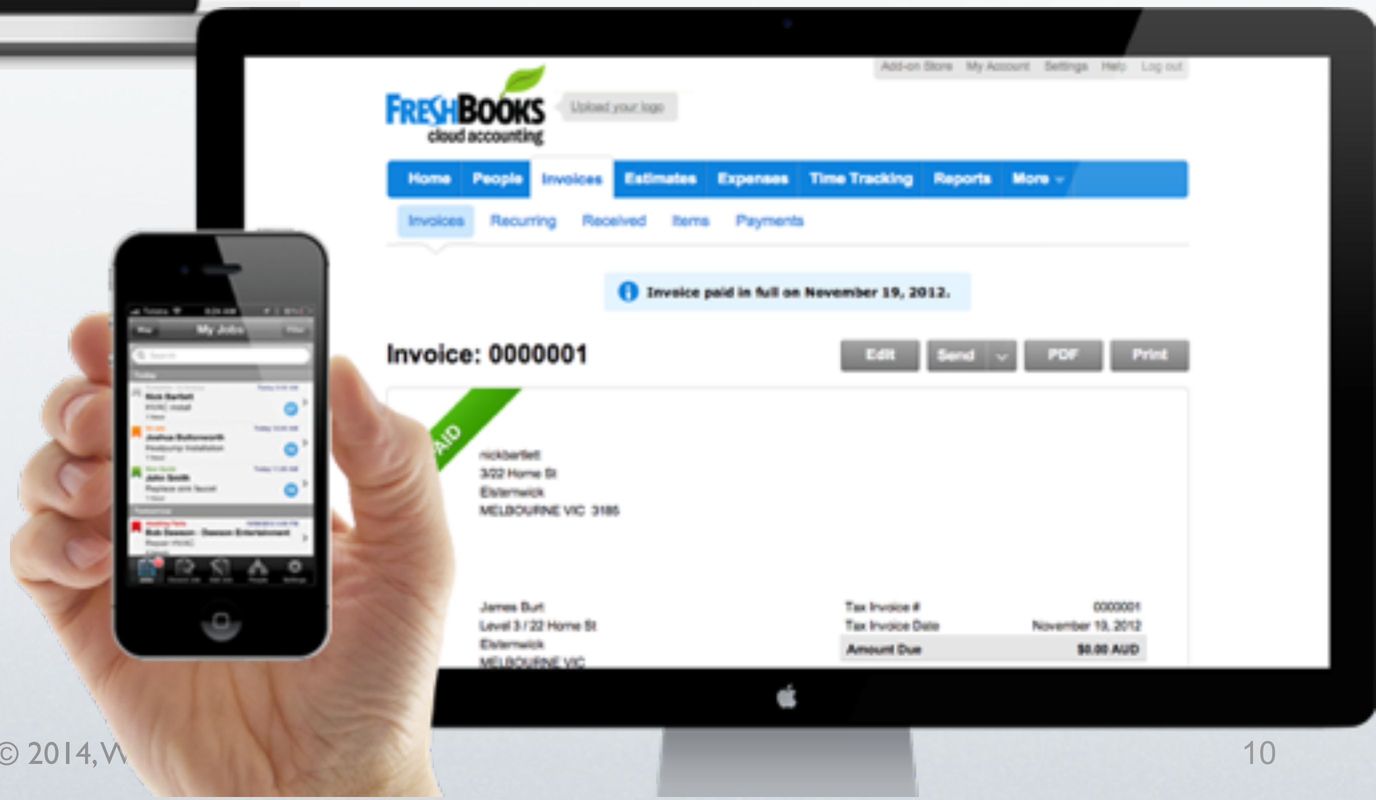


iPod
5:22 PM
invoiceASAP
Invoices
Estimates
Sales Receipts
Customers
Reports
Manage Account

YAHOO!
SMALL BUSINESS
Playable Art
$29.95
Tangle Jr. Chrome
$24.99
Original Koosh Ball
$4.99

SwimTopia

LabStar
Cloud-Based Dental Lab Software

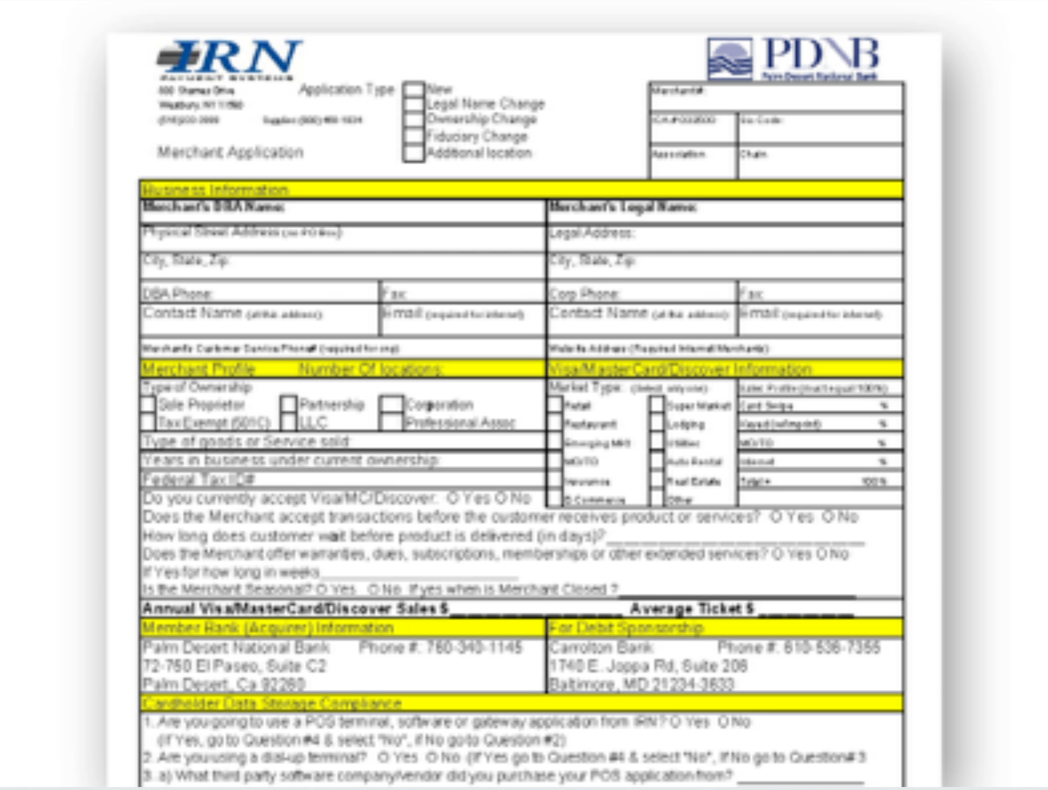


FRESHBOOKS
cloud accounting
Home
People
Invoices
Estimates
Expenses
Time Tracking
Reports
More
Invoices
Recurring
Received
Items
Payments
Invoice paid in full on November 19, 2012.
Invoice: 0000001
Edit
Send
PDF
Print

Requirements

Traditional payments do not work for bottoms-up economy

Long application form + day/weeks to approve + approval difficult for individuals



Poor conversion rate for micro-businesses and fundraisers

Payments requirements for bottoms-up economy

- Easy and fast merchant onboarding
- Underwrite individuals or micro-businesses with little or no traditional business history
- Prevent collusion and takeover fraud

Fraud threats are everywhere

Information Week
Target Confirms Hackers Stole 40 Million Credit Cards

The New York Times
Neiman Marcus Data Breach Worse Than First Said

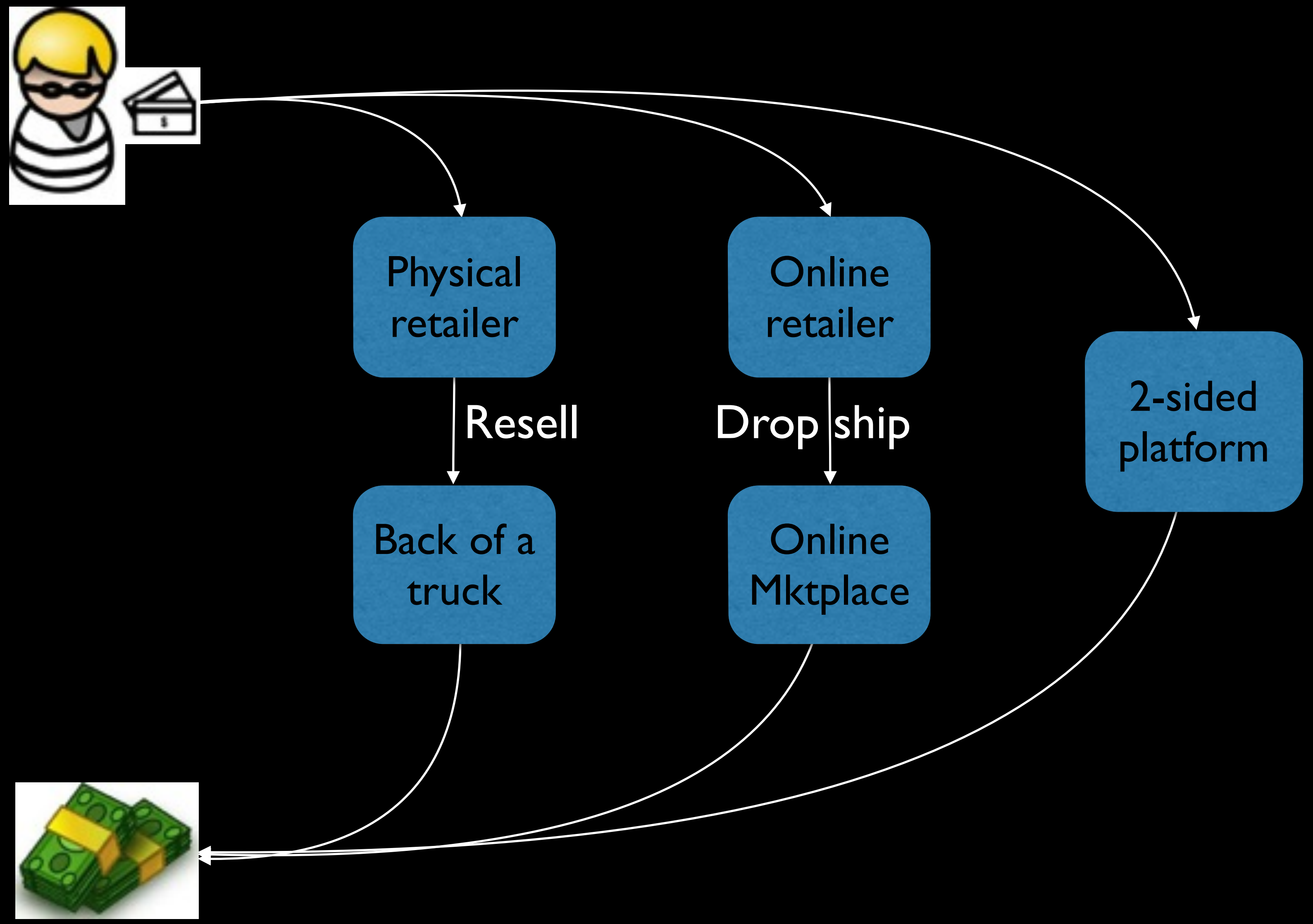
The New York Times
Michaels Stores' Breach Involved 3 Million

The Register
Krebs: Lexis-Nexis, D&B and Kroll hacked

CyberSource Online Fraud Report
Estimated $3.5 Billion Lost to Online Fraud

Entrepreneur, April 18 2014
Online Debit, Credit Fraud Will Soon Get Much Worse. Here's Why.

Fraudsters want to monetize their stolen cards



Solution outline

1. Lots of risk data from many sources
    a) What data
    b) How to collect
    c) What infrastructure
2. Multi-level risk decisioning
    a) Machine learning
    b) Rules
    c) Manual
3. Metrics & feedback

Data

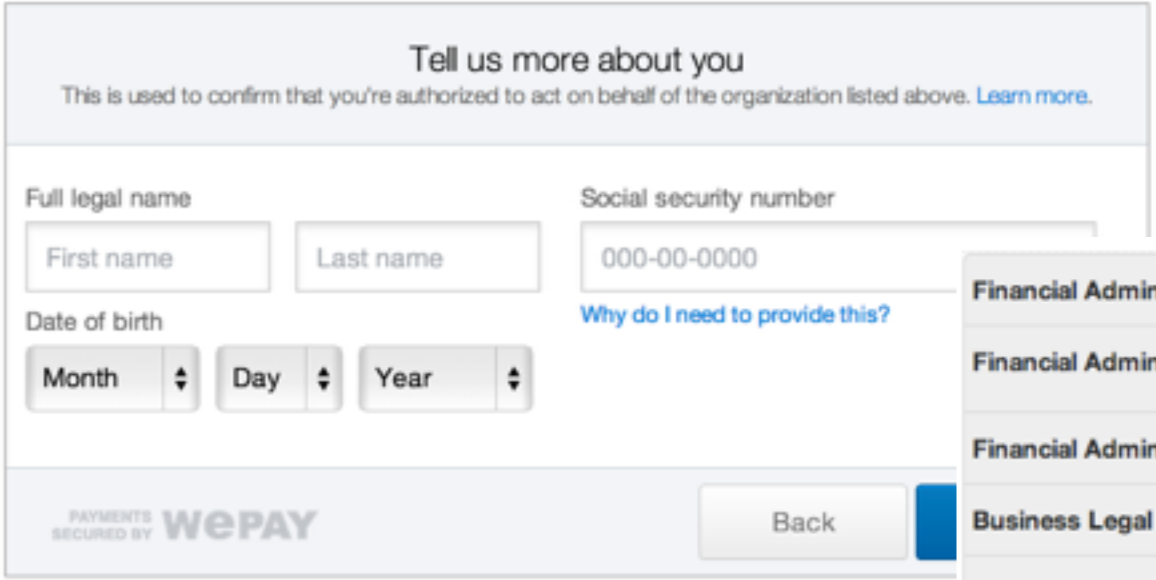
Data approach

data

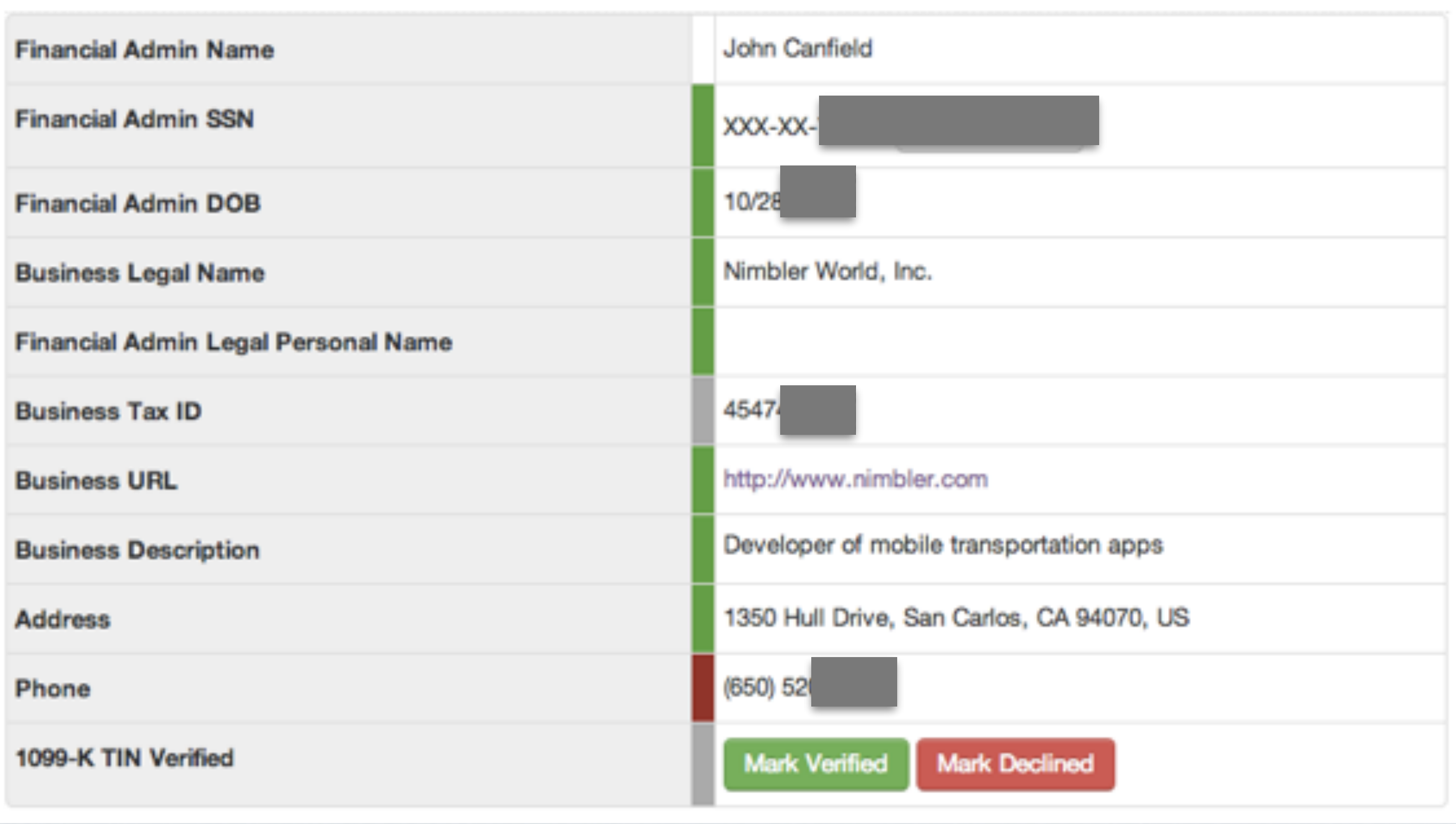
- Predictive of loss / fraud
- Compliance
- More data is better if you have scalable data arch and decisioning
- No silver bullet, so move towards big data

Know-Your-Customer (KYC) checks

data



User



Financial Admin Name	John Canfield
Financial Admin SSN	XXX-XX-[redacted]
Financial Admin DOB	10/2[redacted]
Business Legal Name	Nimbler World, Inc.
Financial Admin Legal Personal Name	
Business Tax ID	4547[redacted]
Business URL	http://www.nimbler.com
Business Description	Developer of mobile transportation apps
Address	1350 Hull Drive, San Carlos, CA 94070, US
Phone	(650) 52[redacted]
1099-K TIN Verified	Mark Verified Mark Declined

Typical vendors:

- Experian
- Equifax
- Lexis Nexis
- ID Analytics
- IDology

vendor → API

Traditional business credit reports

data

Risk Dashboard

Risk Scores and Credit Limit Recommendation		Days Beyond Terms
Intelliscore Plus	Financial Stability Risk	Company DBT
1 HIGH RISK	3 HIGH RISK	55
Score range: 1 - 100 percentile		Industry DBT: 16
Credit Limit Recommendation: $20,400		

Business Facts

Years on File:	5 (FILE ESTABLISHED 09/2008)	SIC Code:	C
State of Incorporation:	CA		
Date of Incorporation:	08/14/2008		SUBDI

Risk Dashboard

Risk Scores and Credit Limit Recommendation		Days Beyond Terms
Intelliscore Plus	Financial Stability Risk	Company DBT
75 LOW-MEDIUM RISK	83 LOW RISK	0
Score range: 1 - 100 percentile		Industry DBT: 4
Credit Limit Recommendation: $53,300		

Business Facts

Years on File:	23 (FILE ESTABLISHED 05/1991)	SIC Code:	
State of Incorporation:	CA		

Risk Dashboard

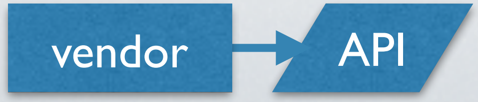
Risk Scores and Credit Limit Recommendation		Days Beyond Terms
Intelliscore Plus	Financial Stability Risk	Company DBT
Score unavailable. Information on file not proven to predict serious future risk.	Score unavailable. Information on file not proven to predict financial stability risk.	DBT Unavailable
Credit Limit Recommendation: N/A		

Business Facts

Years on File:	2 (FILE ESTABLISHED 04/2012)
State of Incorporation:	DE
Date of Incorporation:	03/12/2012
Business Type:	Institutions - Profit

Typical vendors:

- D&B
- Experian
- Equifax



Business Incorporation Docs

data

IRS DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
CINCINNATI OH 45999-0023

Date

Emplo
46-09

Form:

Numbe

IT ONLINE TRAINING LLC

1621 CENTRAL AVE
CHEYENNE, WY 82001

For a
1-800

IF YO
STUB

WE ASSIGNED YOU AN EMPLOYER IDENTIFICATION NU

Corp No. 583210

GOVERNMENT OF
THE VIRGIN ISLANDS OF THE UNITED STATES
— 0 —
CHARLOTTE AMALIE, ST. THOMAS, VI 00802

CERTIFICATE OF EXISTENCE

To All To Whom These Presents Shall Come:

I, GREGORY R. FRANCIS, Lieutenant Governor of the Virgin Islands do hereby certify that I am, by virtue of the laws of the Virgin Islands, the custodian of the corporate records and the proper officer to execute this certificate.

I further certify that the records of this office disclose that

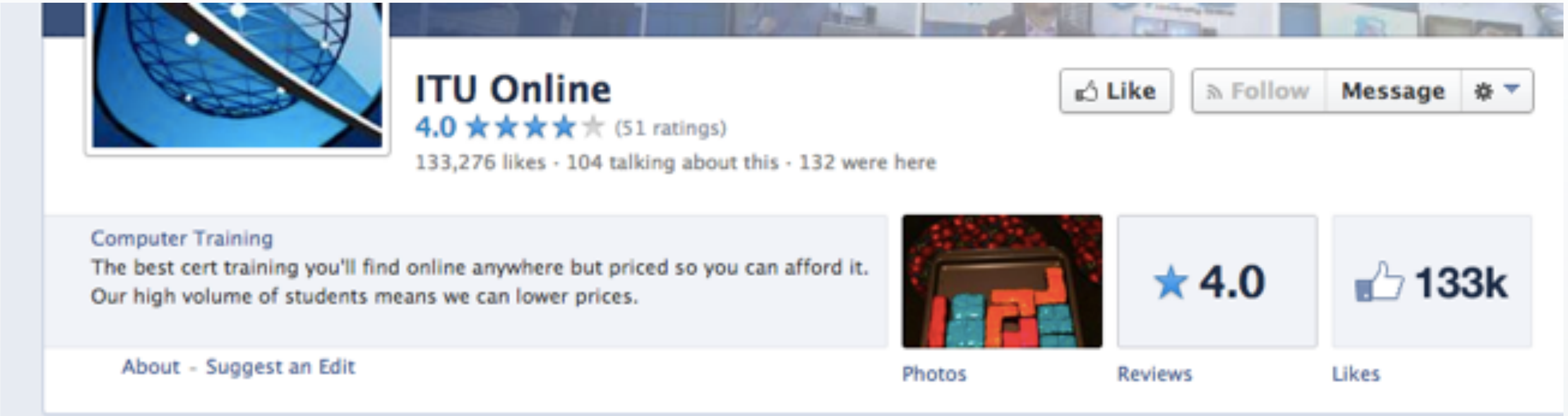
vendor → API

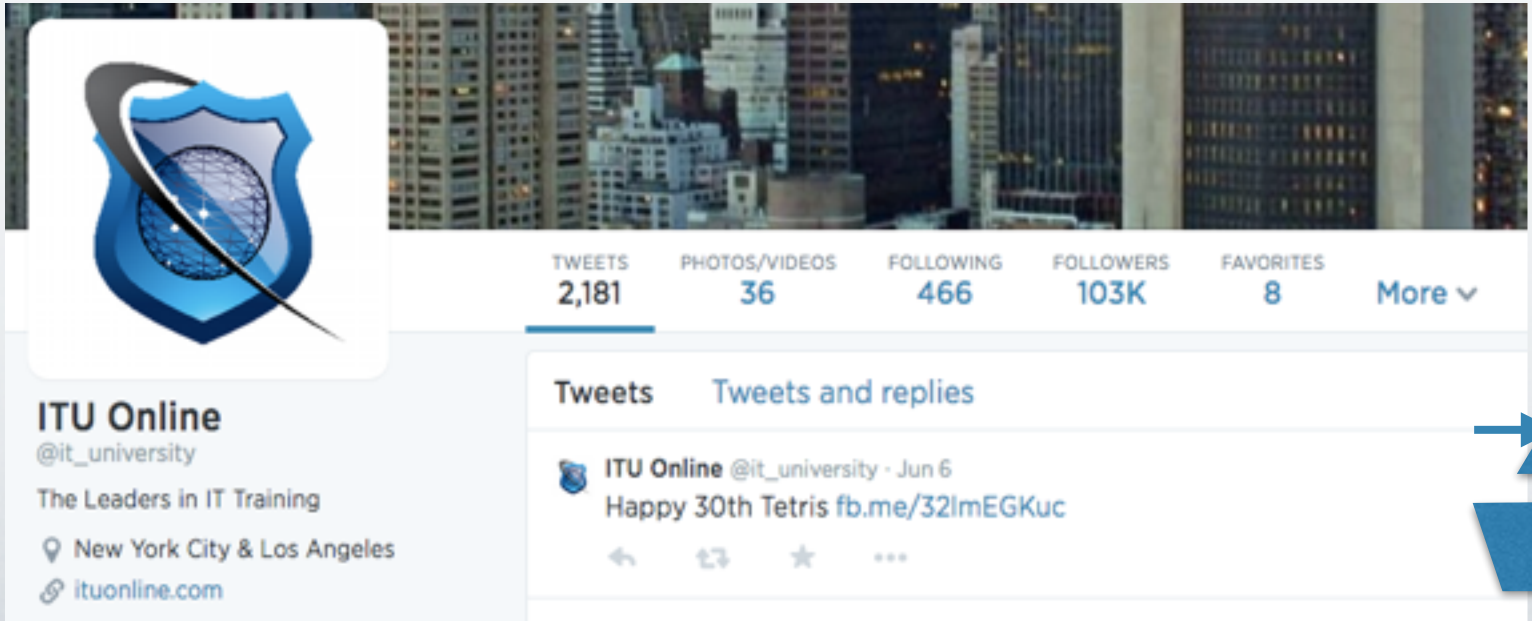
User

manual

Business social media

data



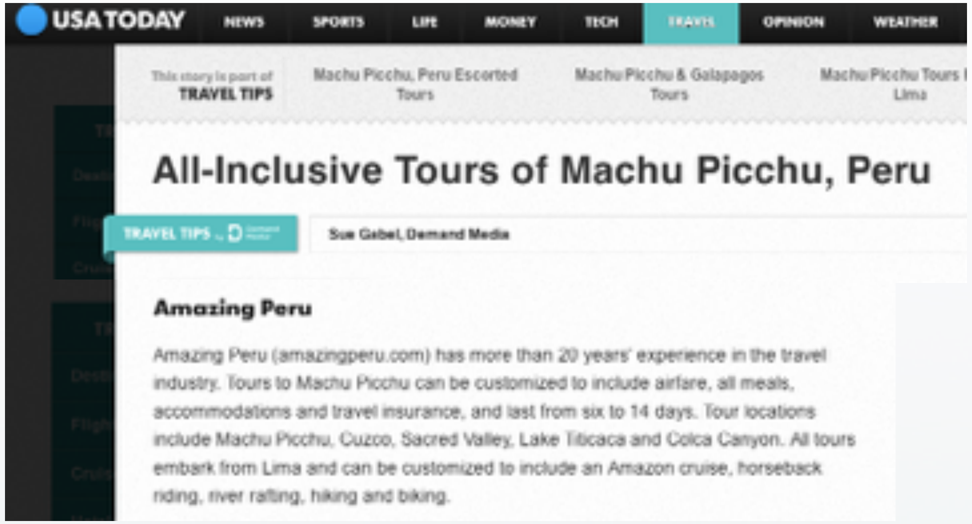


API

manual

Editorial Reviews and Ratings

data

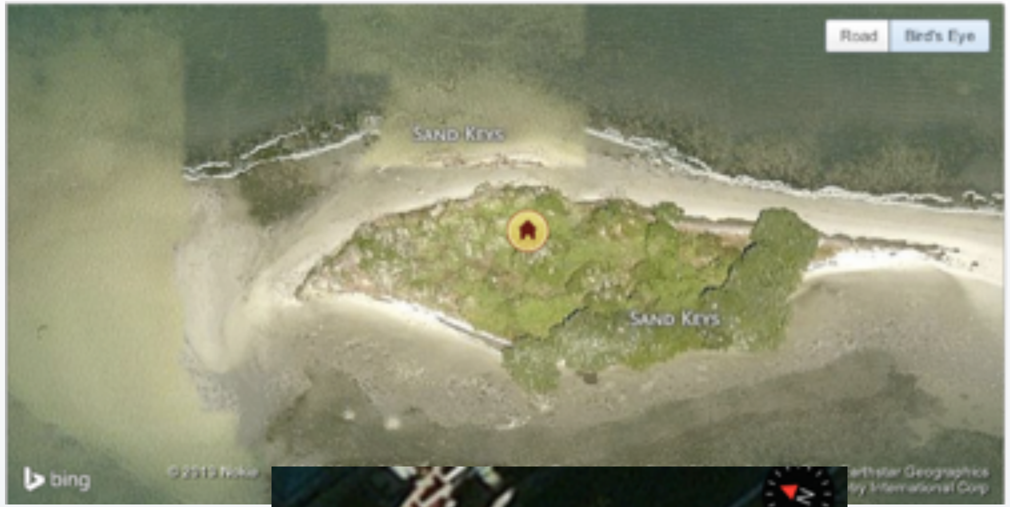


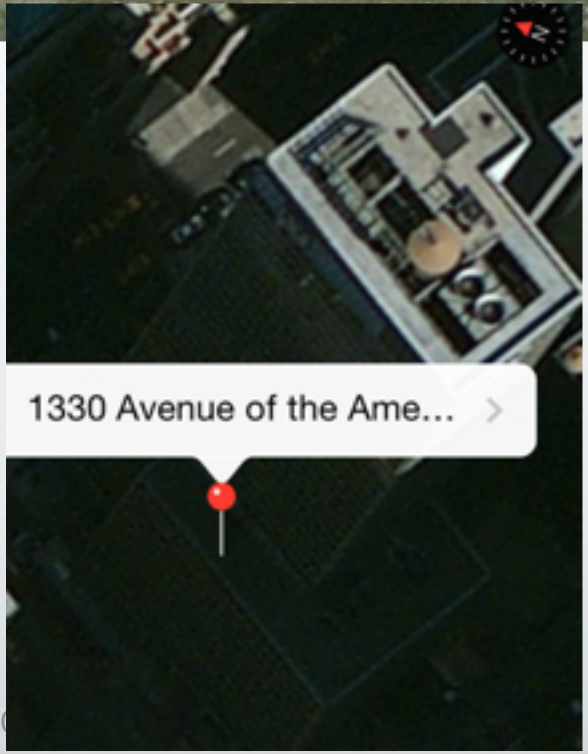
manual

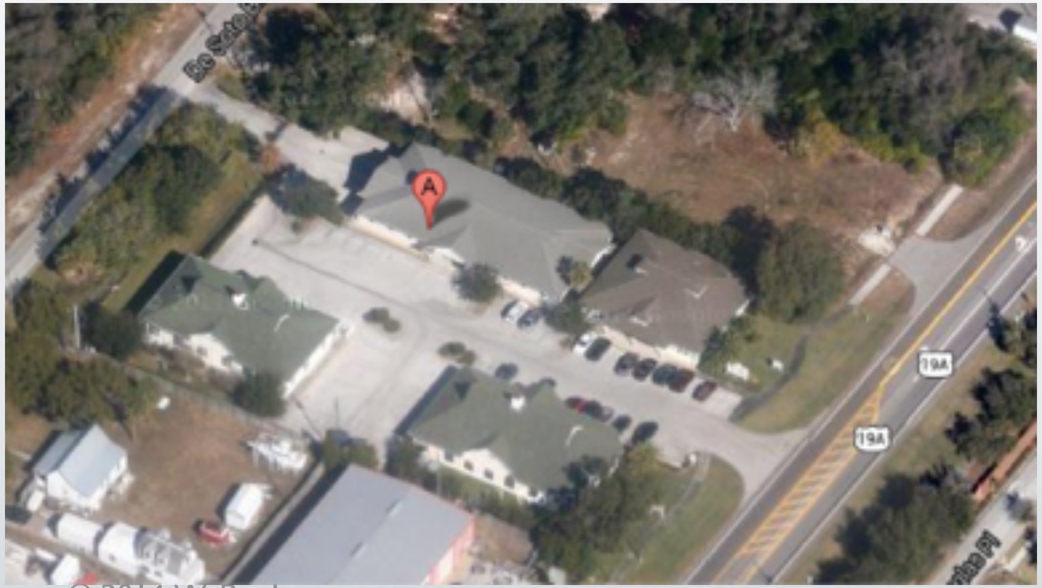
spider

Maps / Street View

data







API

manual

Facebook profile

data

Payee FB data (score: 70 / 100)

Type	Facebook	Score	70
3rd Party ID	1177872345	Name	Louis Sirico
Age (months)	83.931027777778	Update Ts	03/17/2013 | 23:06 UTC

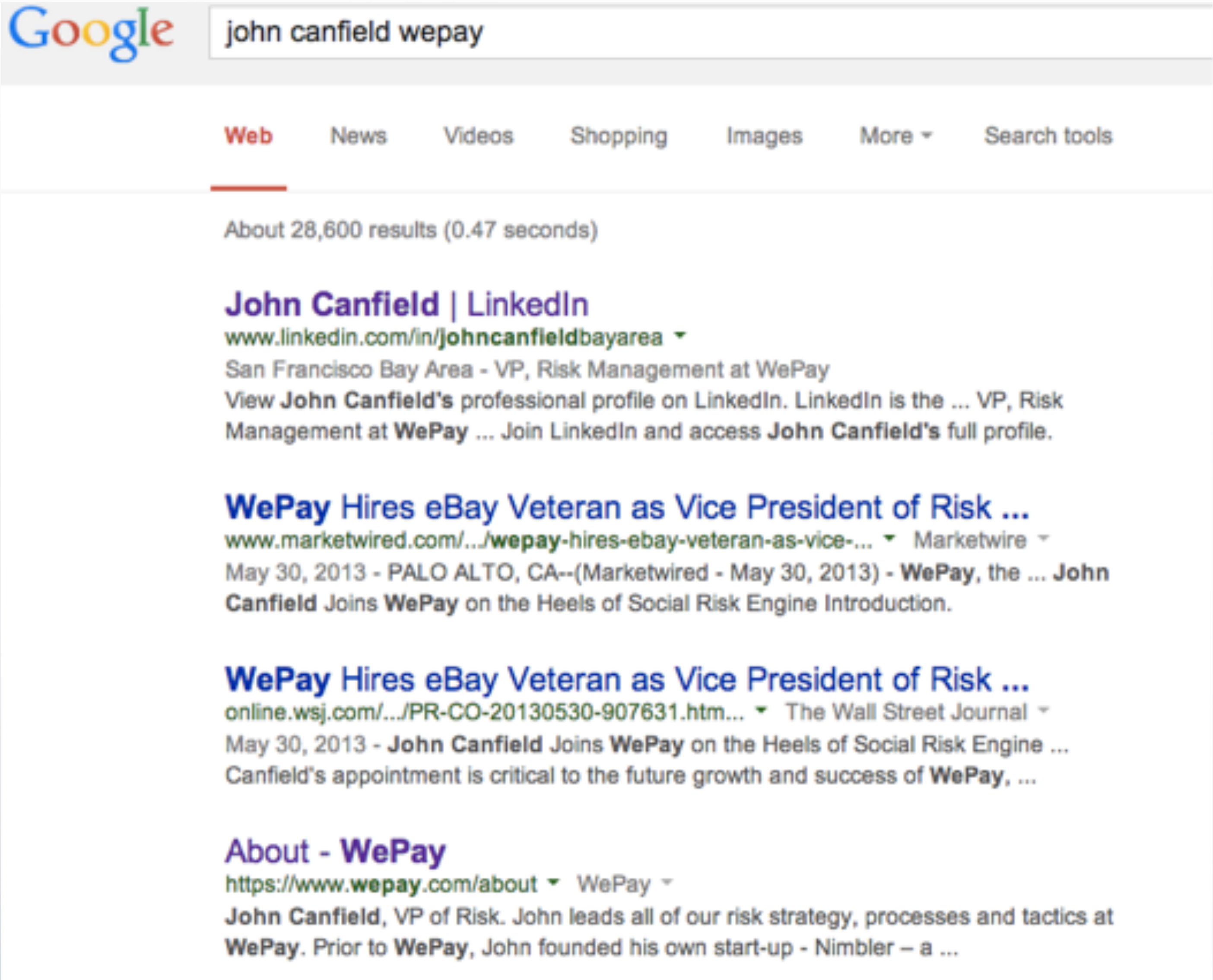
Raw Data

API

manual

Google

data



API

manual

Device ID

data

Device	
Exact ID	a2d51953ab82465187038dd13a6a012e
Exact ID Match Result	success
Smart ID	6e638addb0c94bd78f342ab45531e350
Smart ID Confidence	100
Local Time Offset	0
Local Time Offset Range	0
Offset Measure Time	
OS	Windows NT
OS Anomaly	

Typical vendors:

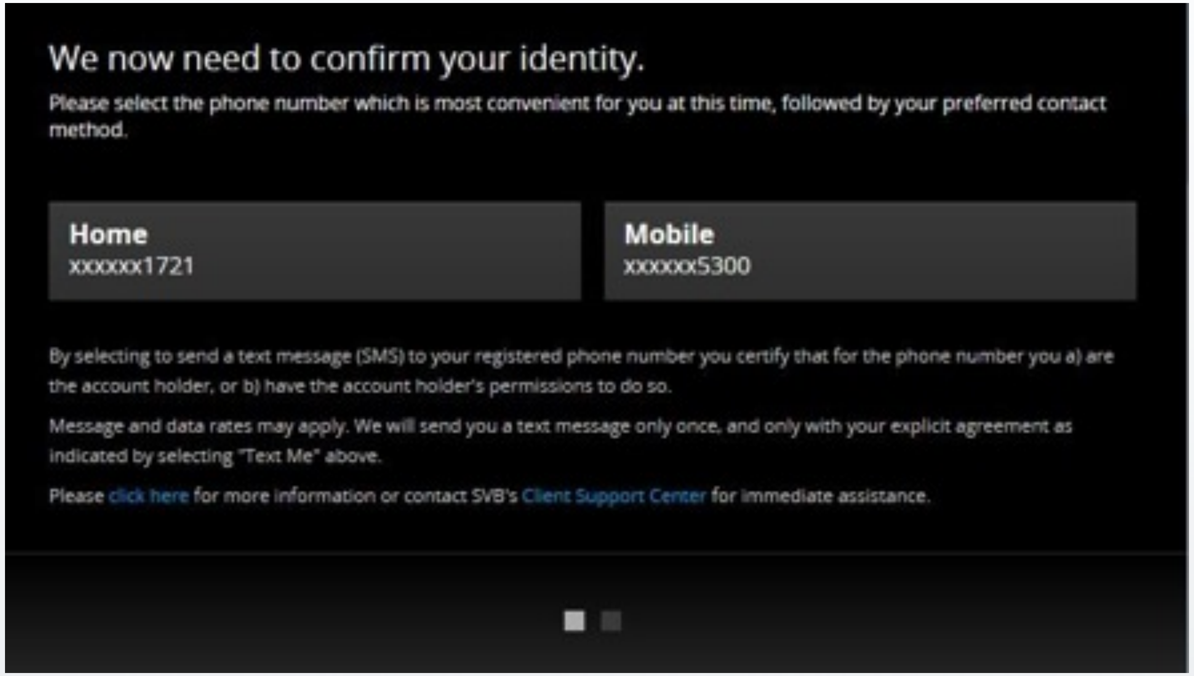
- ThreatMetrix
- iovation
- Experian / 41st Parameter

	For Group Admin	Total Users	Blacklisted Users	Blacklisted Groups
By IP:	26	10	7	10
By True IP:	23	8	5	6
By Device ID:	2	4	3	3
By Fuzzy Device ID:	2	4	3	3

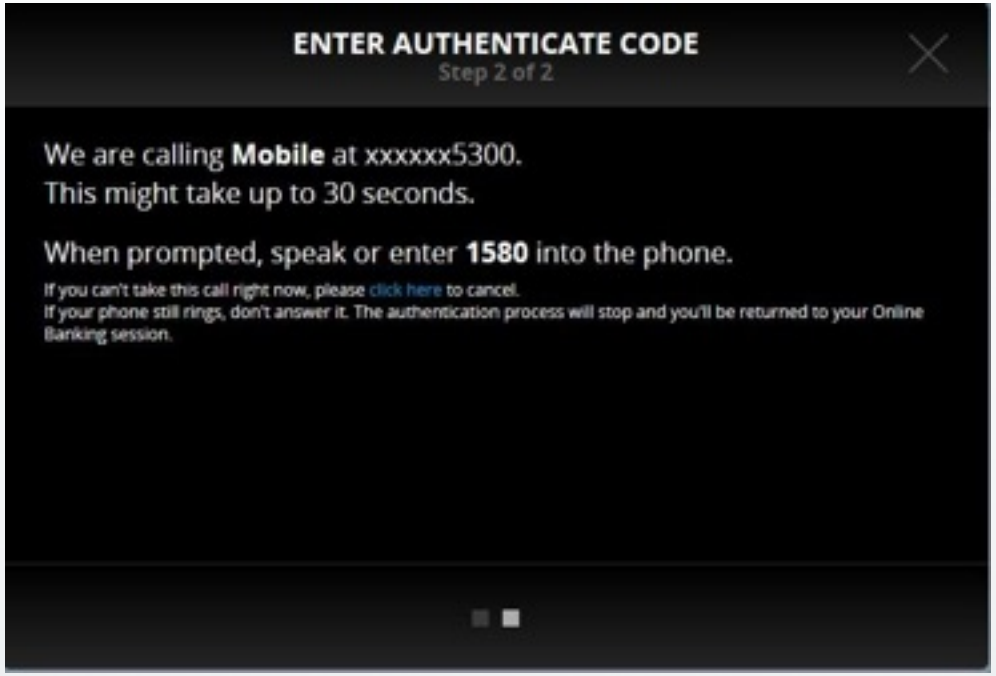
vendor → API

Control Verification

data



+



Typical vendors:

- Authentify
- Telesign
- Twilio

vendor → API

Transaction History

data

Payment Stats To Account (60 days)

State	Count	Count %	TPV	TPV %
Captured	53	100%	89559.73	100%
Chargebacks	0	0%		0%
Refunds	1	1.89%	215.68	0.24%
Unauthorized	0	0%		0%
Total:	**53**		**89559.73**	

Unauthorized Payment Stats To Account (60 days)

State	Count	Count %	TPV	TPV %
Total:	**0**		**0**	

Payment Stats To Account (60 days)

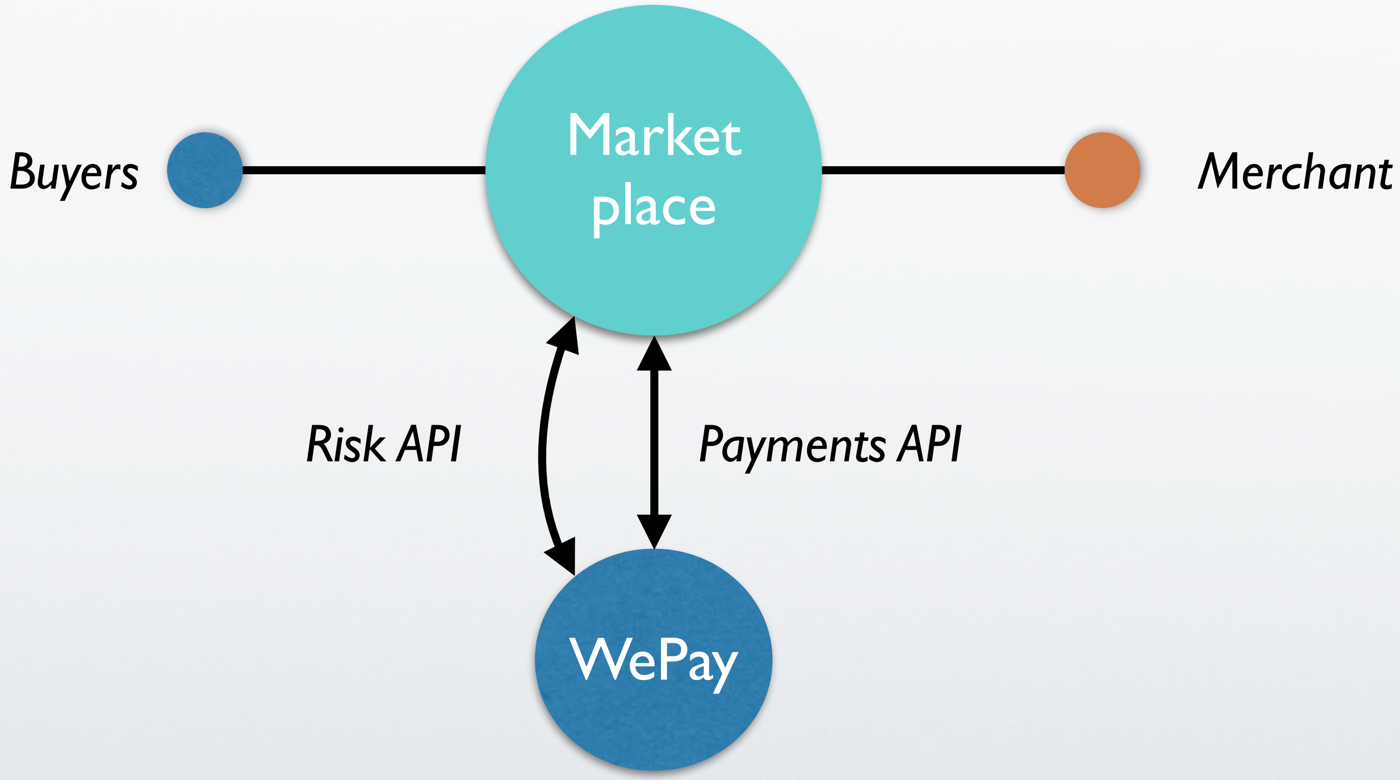
State	Count	Count %	TPV	TPV %
Captured	220	97.78%	499089.87	99.06%
Failed	4	1.78%	2029.10	0.4%
InReview	1	0.44%	2697.00	0.54%
Chargebacks	0	0%		0%
Refunds	2	0.89%	1368.83	0.27%
Unauthorized	10	4.44%	35984.00	7.14%
Total:	**225**		**503815.97**	

Unauthorized Payment Stats To Account (60 days)

State	Count	Count %	TPV	TPV %
User_CID	2	6.06%	9116.00	7.67%
Decline_InsufficientFunds	7	21.21%	27526.40	23.16%
Decline_General	20	60.61%	73656.40	61.97%
ExcessiveAuthorizations	1	3.03%	2448.00	2.06%
UndefinedResponseCode	3	9.09%	6106.00	5.14%
Total:	**33**		**118852.8**	

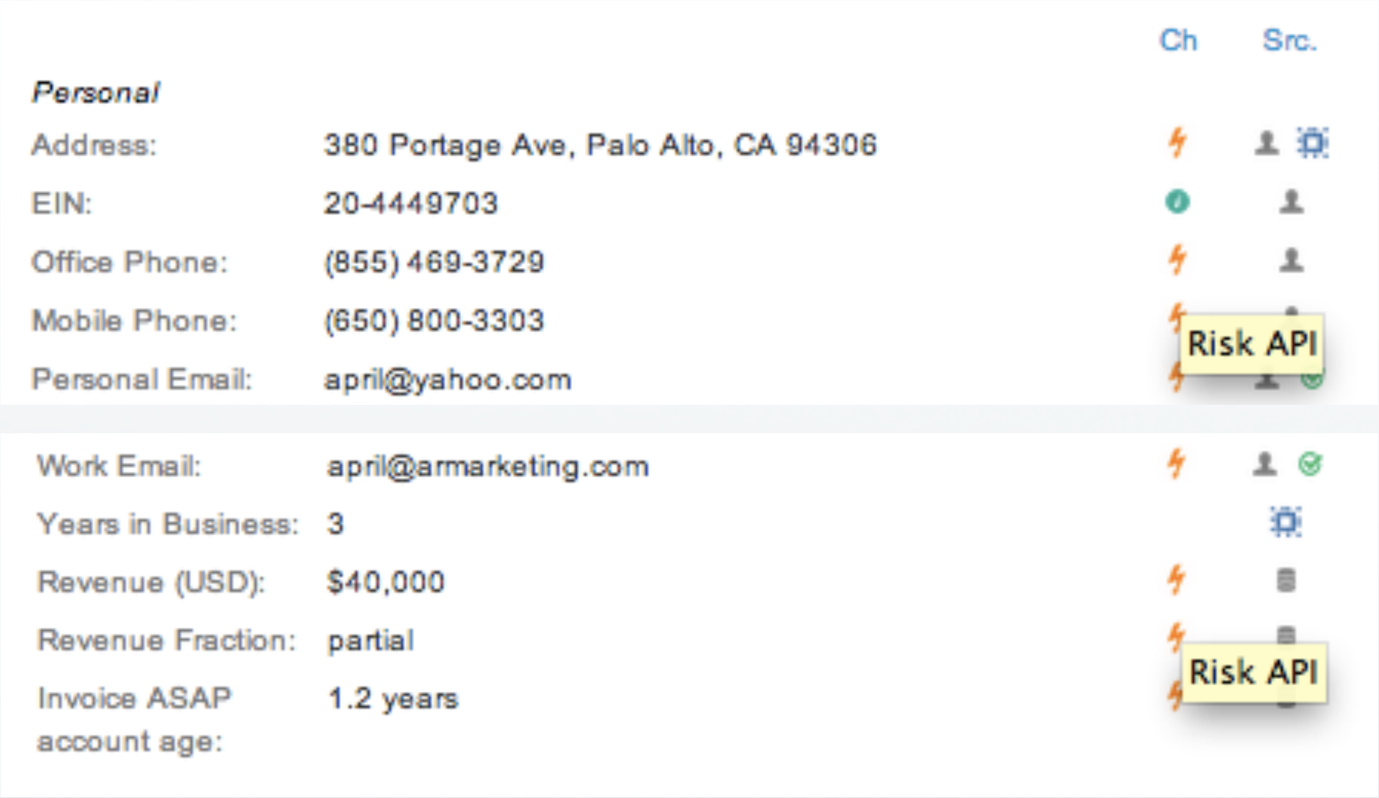
Data from Partners

data



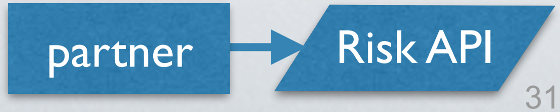
Risk API Account Information

data



Example Risk API data types

- person
- email
- business_name
- address
- phone
- tax_id
- website_uri
- employment
- industry_code
- business_description
- risk_score
- comment
- project
- fundraising_event
- fundraising_team
- acquisition_channel
- partner_service
- member_to_member_message
- external_account
- editorial_review
- other_web_content
- revenue
- conversation
- business_legal
- business_report
- other_document
- device_info
- control_verification
- risk_review
- risk_review_steps
- transaction_details



Risk API Transaction Information

data

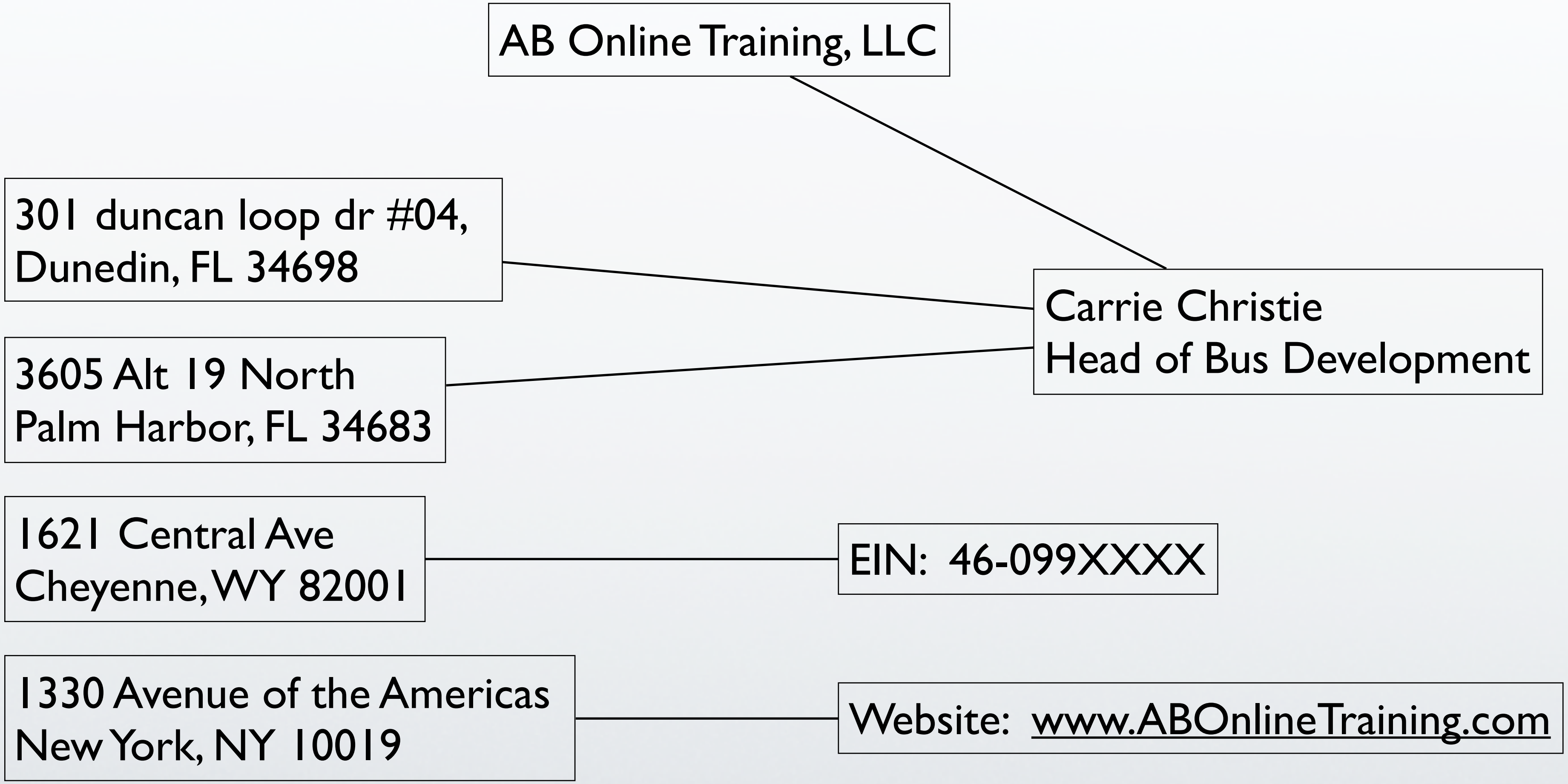
		Ch	Src.
Transaction Type:	Services		
Delivery Type:	Fully Delivered		
Amount:	$5000		
Receipt URL:	https://view.invoiceasap.com/view.php?invoi... [show full]		
Terms URL:	https://view.invoiceasap.com/view.php?invoi... [show full]		
Shipping Addr:			
Shipping Info:			

		Ch	Src.
Line Item 1			
Description:	Search Engine Optimization - hourly Search Engine Optimization - home page and product page		
Qty:	20.00		
Price:	$100		
Total:	**$2,000**		
Line Item 2			
Description:	Social media marketing - February Twitter, Facebook, Google+ social marketing of Nimbler app		
Qty:	1		
Price:	$3,000		
Total:	**$3,000**		

partner → Risk API

How to effectively organize this data?

data



Data storage systems

data

SQL databases:

- MySQL
- Oracle
- …

No-SQL databases:

- Document (MongoDB…)
- Key-value (Redis…)
- Graph (Neo4J…)
- Column (Cassandra…)

Other:

- Hadoop
- …

Decisioning

What is Machine Learning?

decisioning

1) Supervised learning

Txn #	BL accts by IP	BL accts by True IP	BL accts by Device ID	BL accts by fuzzy Device ID	Txn Amount	IP to Address miles
100000001	0	0	0	0	20.00	22
100000002	8	5	3	3	499.00	15
100000003	0	0	0	0	35.00	539
100000004	0	0	0	0	85.00	0
100000005	0	0	0	0	90.23	2
100000016	0	0	0	0	400.00	13
100000010	0	0	0	0	20.00	1
100000011	0	0	0	0	75.00	5
100000012	0	0	0	0	80.00	8
100000013	0	0	0	0	50.00	0
100000014	1	0	0	0	25.00	2
100000015	0	0	0	0	30.00	1

+

Label: is Fraud?
0
1
0
0
0
0
0
0
0
0
0
0
0
0
0

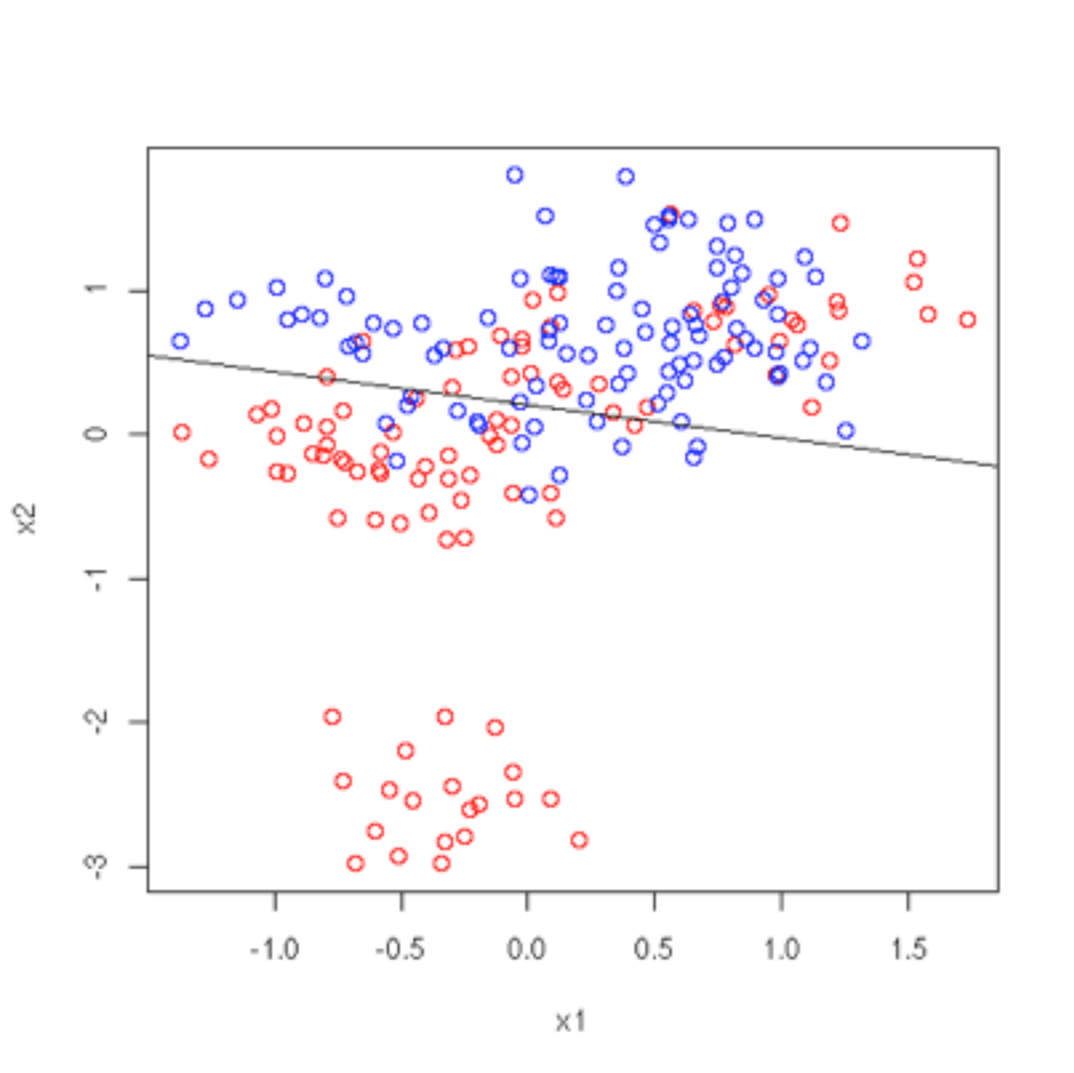
Training →

Machine Learning Model

How does machine learning work?

decisioning

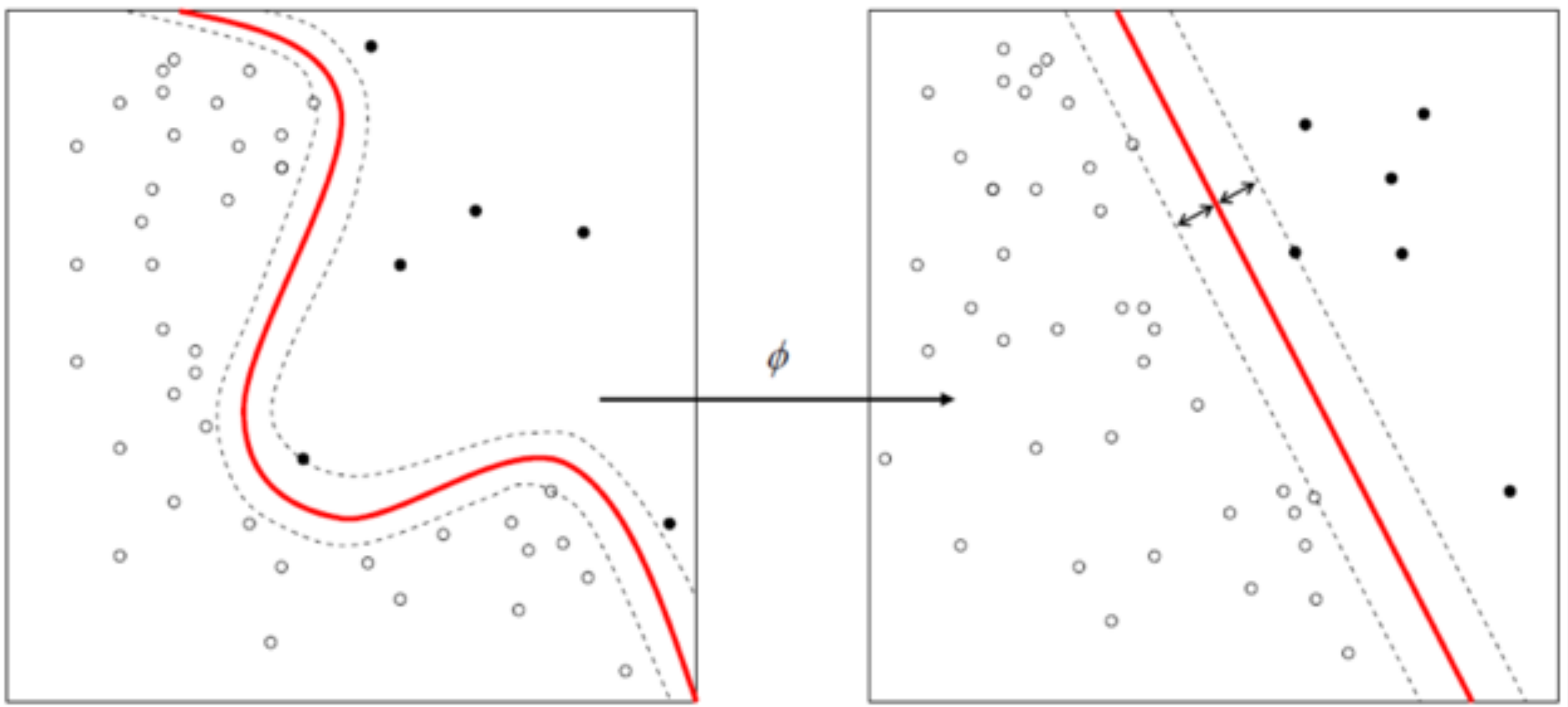
Linear regression classifier



$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right)$$

Support Vector Machine

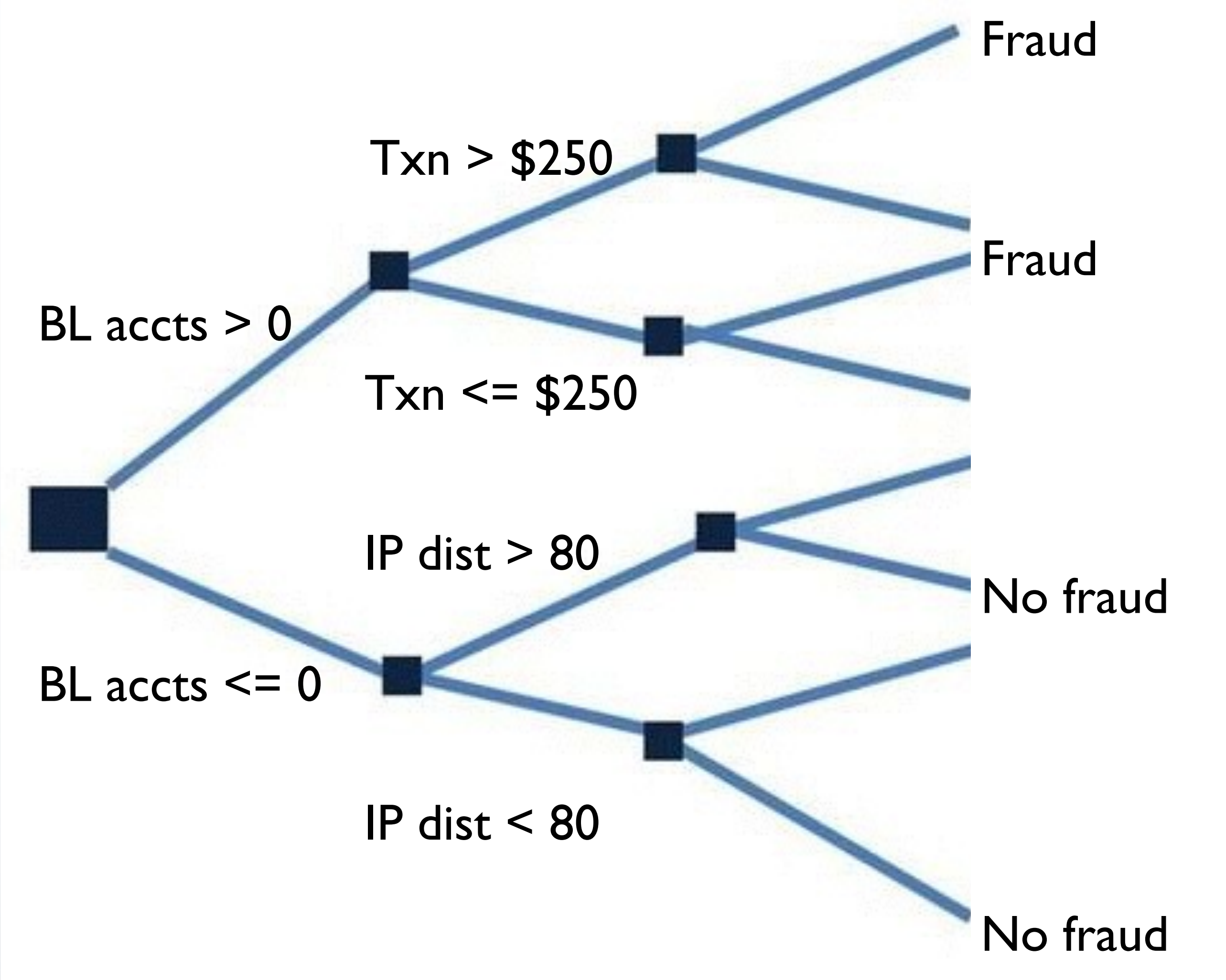
decisioning



- Polynomial (homogeneous): $k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j})^d$
- Polynomial (inhomogeneous): $k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + 1)^d$
- Gaussian radial basis function: $k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma \|\mathbf{x_i} - \mathbf{x_j}\|^2)$, for $\gamma > 0$. Sometimes parametrized using $\gamma = 1/2\sigma^2$
- Hyperbolic tangent: $k(\mathbf{x_i}, \mathbf{x_j}) = \tanh(\kappa \mathbf{x_i} \cdot \mathbf{x_j} + c)$, for some (not every) $\kappa > 0$ and $c < 0$

Decision Trees

decisioning



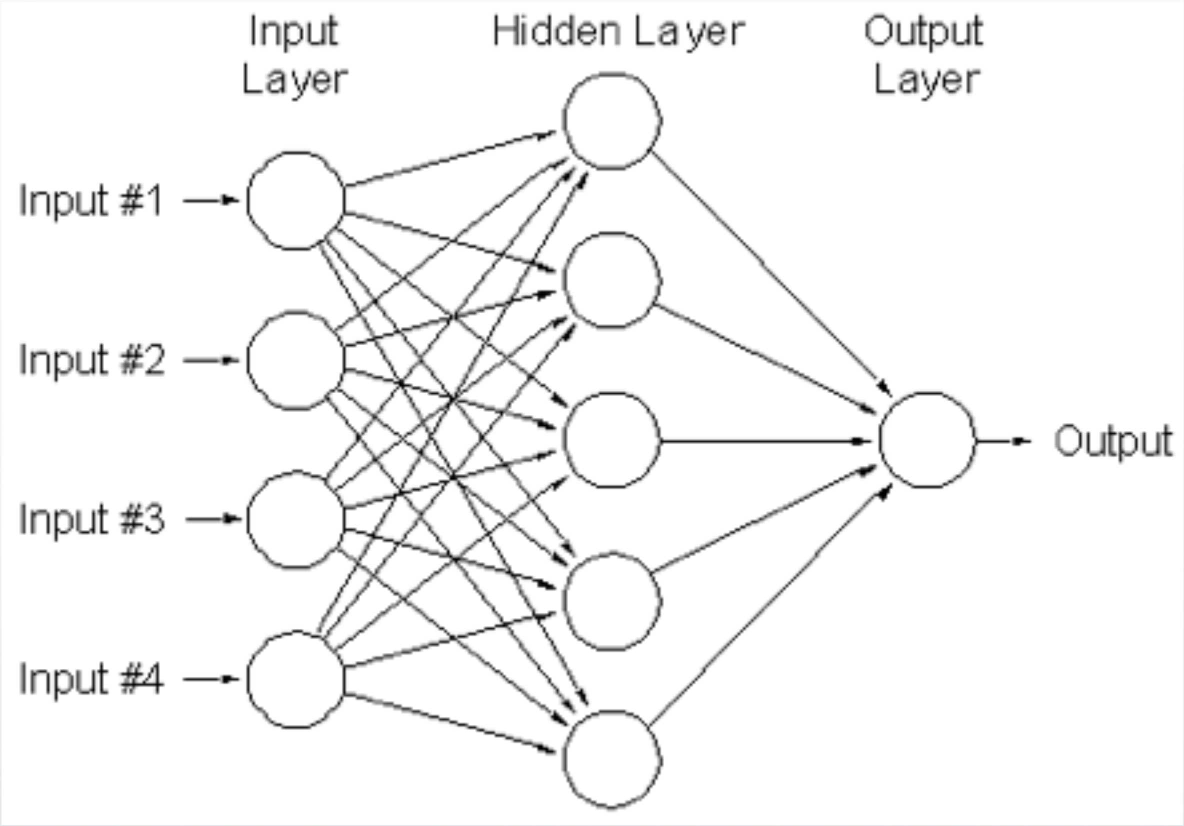
<u>ID3 Training algorithm</u>

Recurse and choose attribute that minimizes entropy:

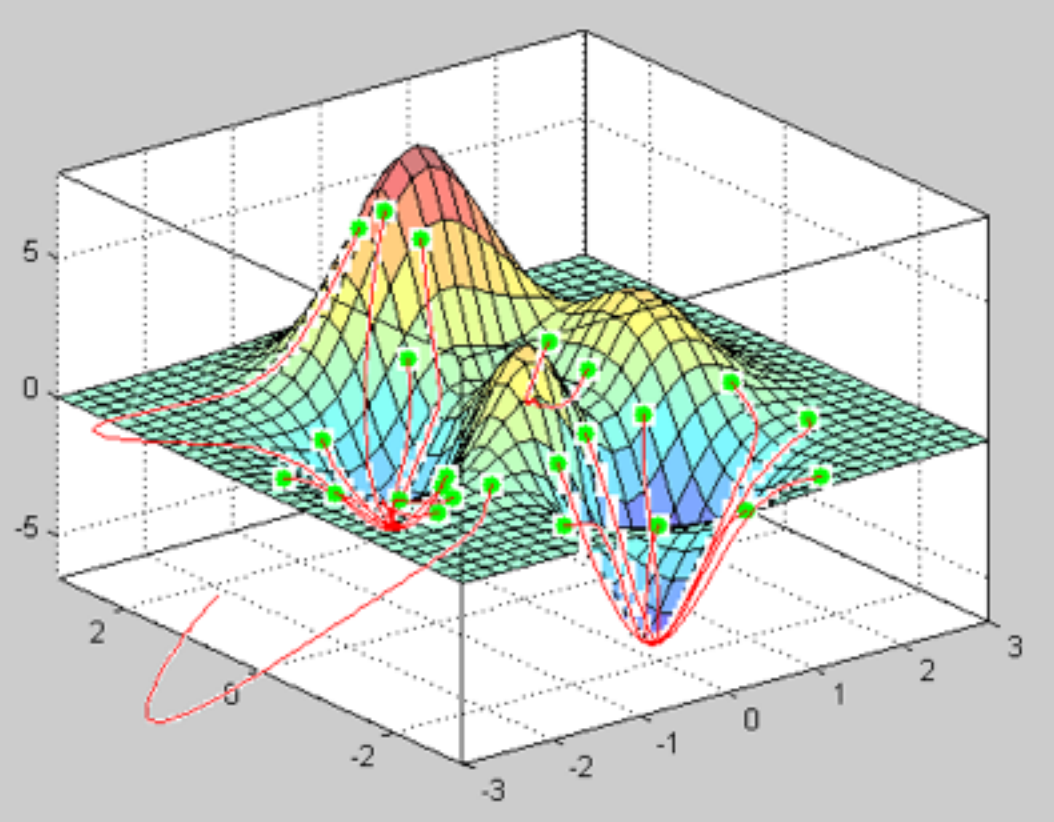
$$H(S) = -\sum_{x \in X} p(x) \log_2 p(x)$$

Neural Network models

decisioning



Gradient descent training



Meta-algorithms: grid search, bagging, etc

decisioning

You can do:

Decision Tree

Or you can do:

Decision Tree 1	Decision Tree 2	Decision Tree 3	Decision Tree 4	Decision Tree 5
SVM 1	SVM 2	SVM 3	SVM 4	SVM 5
Neural Network 1	Neural Network 2	Neural Network 3	Neural Network 4	Neural Network 5
Logistic Regression 1	Logistic Regression 2	Logistic Regression 3	Logistic Regression 4	Logistic Regression 5

Rules Engine

decisioning

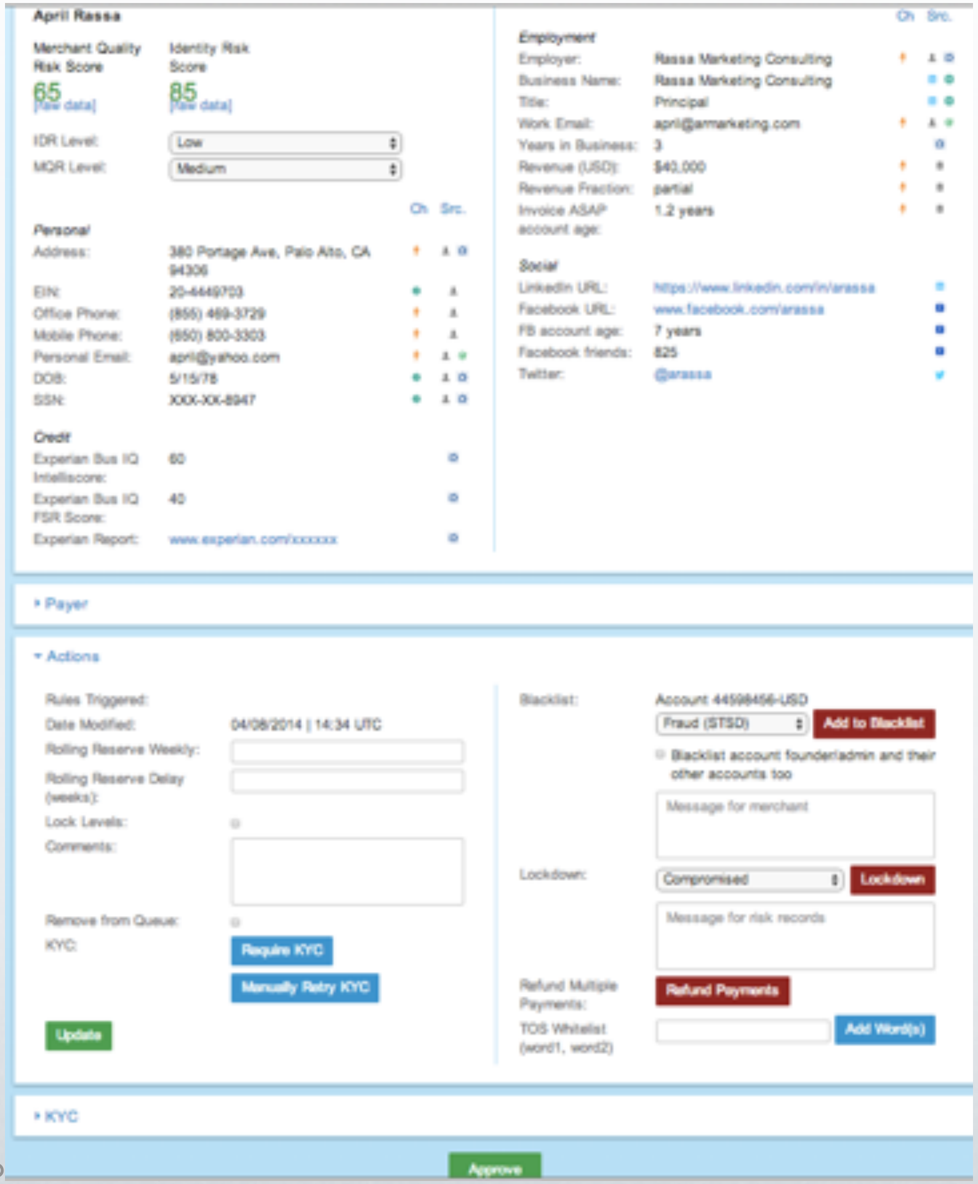
Rule 1 (Transaction threshold):
  If (signal.Group_30Day_txn_volume > account.tpv_review_threshold_t30d and
    fact.tpv_review_threshold_t30d > 0)
  then Refer_for_Review

Rule 2 (Model Score):
  If (model.fraud_score > 0.85)
  then Refer_for_Review

Rule 3 (Re-require KYC):
  If (signal.IDR <= 0) and (signal.MQR <= 0) and (signal.Group_30Day_txn_volume > 1000.00) and
    (signal.name_address_match = false)
  then Request_Address_Remedy

Manual Review

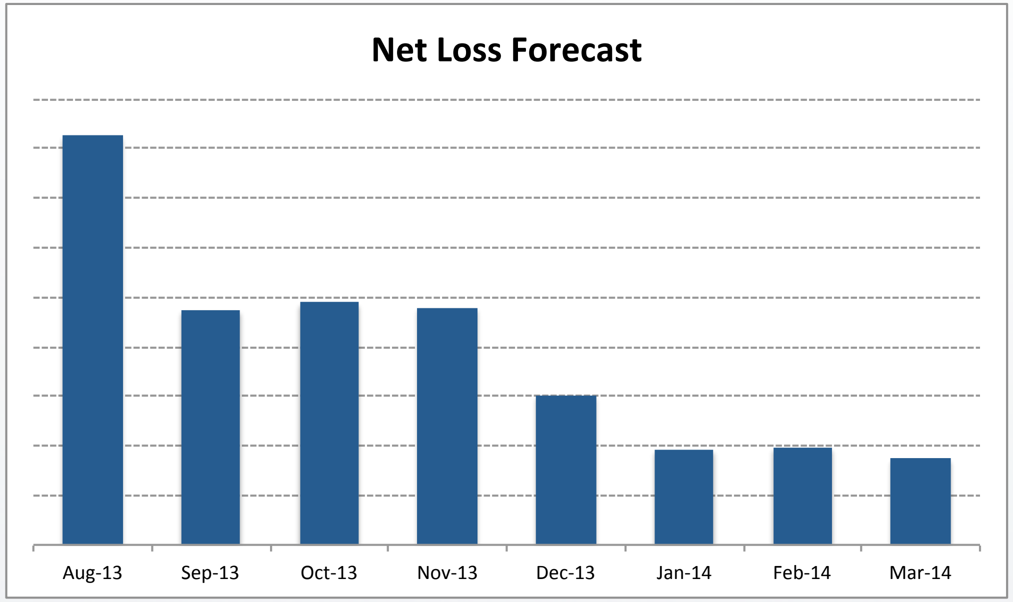
decisioning

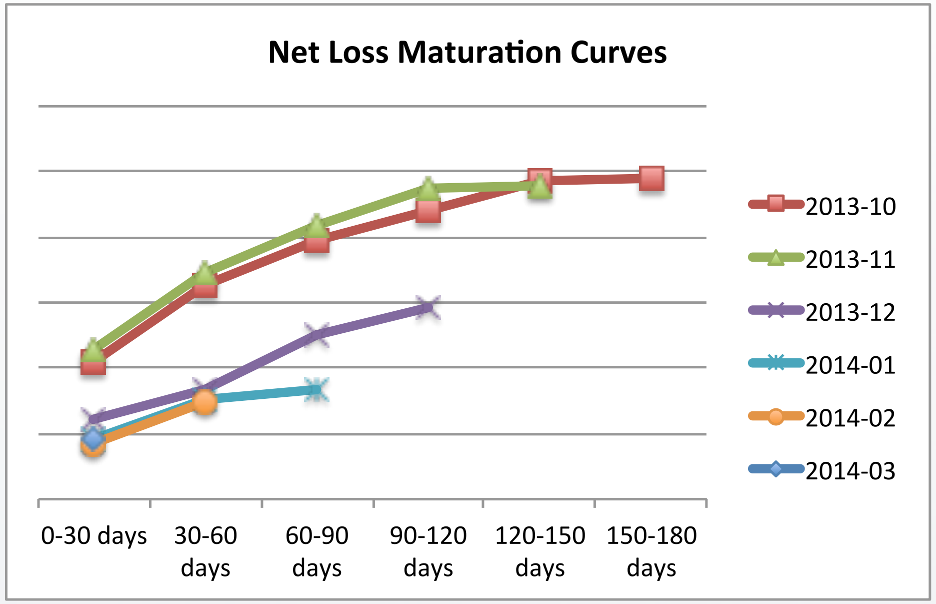


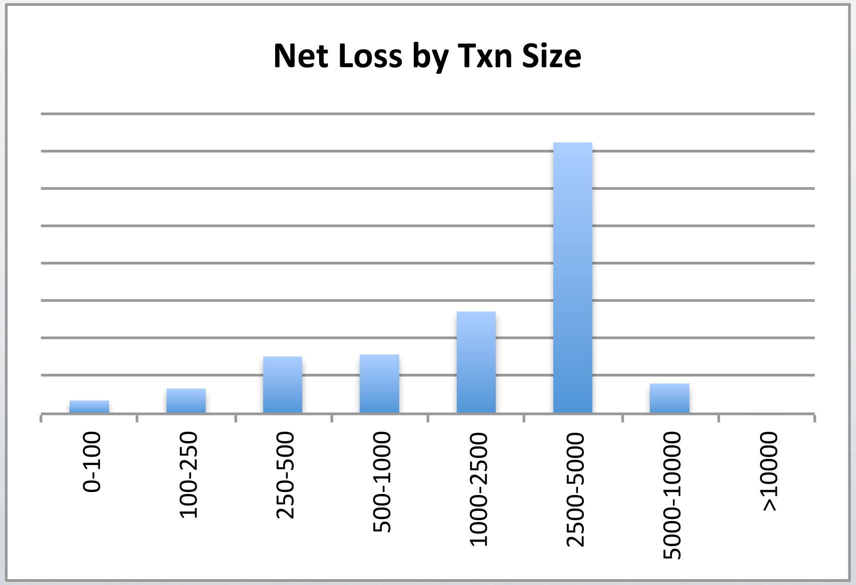
Metrics & Feedback

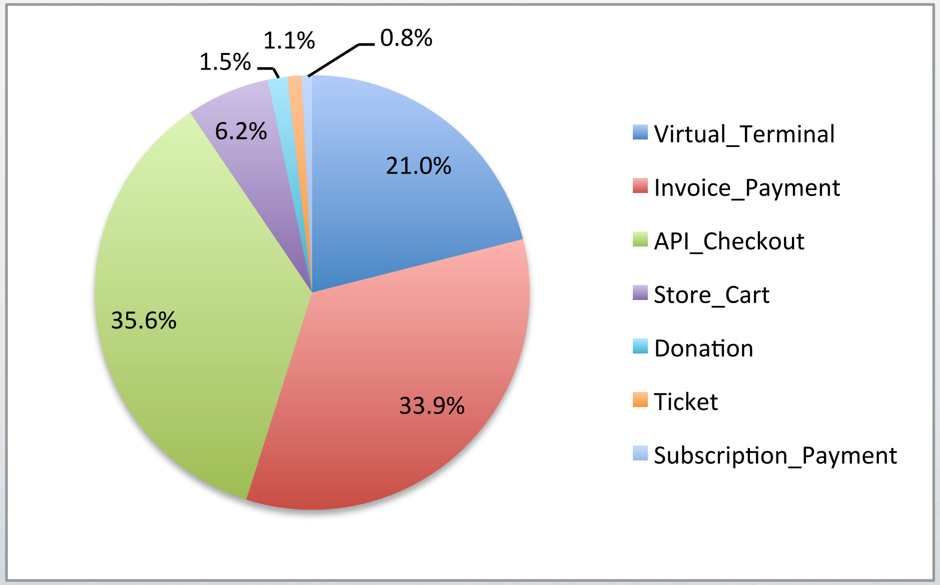
Metrics & Feedback

metrics









Top Net Loss Accounts

Loss Week	2014-17
Orig Month	(All)

Net Loss Row Labels	Total
Store_Cart	$187
988052845	$115
78252	$72
Invoice_Payment	$1,820
1426156	$1,166
1953717721	$505
1373219397	$134
190996	$15
API_Checkout	$10,156
846171008	$1,698
1692429545	$1,208
1570348118	$674
858508144	$663

Summary

1. Bottoms-up economy is growing and is the high-growth frontier for electronic payments

2. Big data needs to be pieced together bit by bit using flexible infrastructure

3. Machine learning, rules, and manual review when combined properly gives enables correct decisions on big data