Predicting Fraud with Oracle Analytics

Eric Probst, Manager Risk Analytics, Certegy

Dan Vlamis, President, Vlamis

June 13, 2022

Agenda

- **Who is Certegy?**
- **OAC Experience**
- **Use of Spatial Analytics**
- **Machine Learning**

Company Overview

WHEN YOU THINK OF THIS….

certegy

CERTEGY DECISIONED

2019

378M $160B

2020 US Treasury

1.9M $2.3B

certegy

Who Is Certegy

The most comprehensive DDA authorization and risk management system in the industry delivering unprecedented accuracy, convenience and simplicity.

At the heart of Certegy's system is a sophisticated risk management system including a **proprietary consumer database with approximate 38M unique consumers**, and extensive fraud management tools to minimize losses.

Retail: Point of Sale and ACH transaction authorizations (first party checks)

InstantFunds: Check cashing and deposit authorization for retailers, financial institutions, and mobile remote deposit capture (third party checks)

4,000 → **Strong client base with over 4,000 national and regional retailers**

Client Base → **23 of the top 50 retailers**

300,000 → **Over 300K locations across multiple industries including retail, auto and home improvement**

certegy

The powerhouse behind our solutions: Risk Management

certegy

Risk Management

Industry leading tools combined with 55+ years experience

Confirm Information

Data Science

- Develop Analytical Scoring Model.
- Creating Neural Networks and Random Forest Analysis

Industry Insight

Risk Analytics

- Use Risk Analytical tools to determine acceptable loss dollars and approve as much good volume as possible
- Targeted fraud cases
- Customer presentations and control recommendations

Supplemental Data

Fraud Investigations

- Work with Law Enforcement to track and arrest organized fraud rings
- Investigate and Collect large dollar returns.

certegy

The powerhouse behind our solutions: Customer Presentations

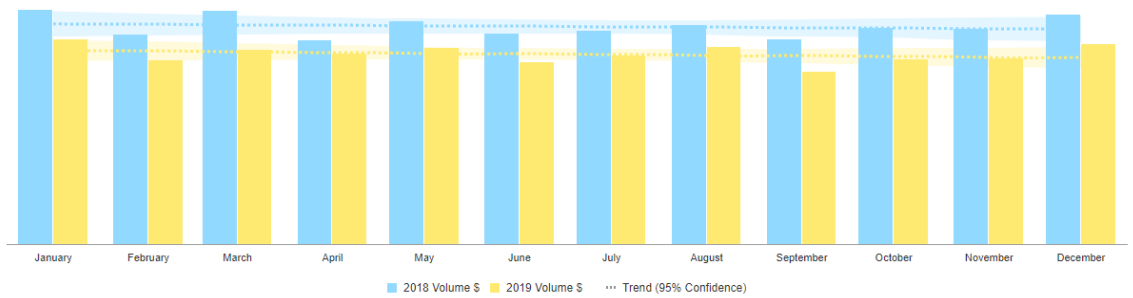
certegy

CUSTOMER REVIEWS

Hyperion Interactive Reporting vs Oracle Analytical Cloud

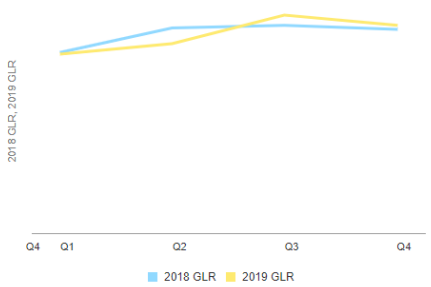
Month	Approved Amount		Approved Variance %	Gross Loss		GLR	
	2018	2019	2018	2018	2019	2018	2019
1	$ 46,614,068	$ 40,732,996	-13%	469,825	403,207	1.01%	0.99%
2	$ 41,709,875	$ 36,672,835	-12%	336,754	316,270	0.81%	0.86%
3	$ 46,349,110	$ 38,639,757	-17%	404,579	310,004	0.87%	0.80%
4	$ 40,604,480	$ 37,931,785	-7%	401,983	329,695	0.99%	0.87%
5	$ 44,438,560						
6	$ 41,933,243						
7	$ 42,488,974						
8	$ 43,504,594						
9	$ 40,742,988						
10	$ 42,960,362						
11	$ 42,962,217						
12	$ 45,579,039						
Total	$ 519,887,507						

2018-2019 Performance



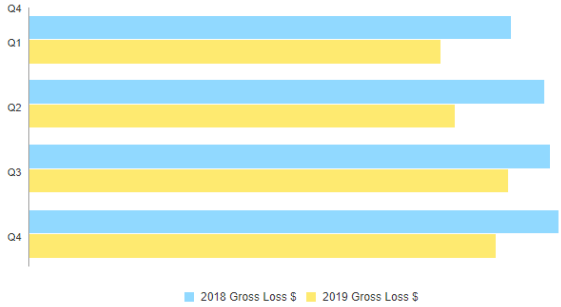
	2018 Volume $	2019 Volume $	19 v 18 Volume ▲ %
January	$46,610,570	$40,718,779	-13%
February	$41,698,406	$36,659,131	-12%
March	$46,339,647	$38,645,570	-17%
April	$40,578,225	$37,923,315	-7%
May	$44,402,901	$39,103,002	-12%
June	$41,902,936	$36,290,130	-13%
July	$42,467,102	$37,559,210	-12%
August	$43,493,420	$39,266,215	-10%
September	$40,726,449	$34,239,400	-16%
October	$42,923,200	$36,731,603	-14%
November	$42,826,366	$36,973,516	-14%
December	$45,574,427	$39,801,056	-13%
Grand Total	$519,543,648	$453,910,925	-13%

GLR %



	2018 Gross Loss $	2019 Gross Loss $	2018 GLR	2019 GLR
January	$468,108	$395,620	1.00%	0.97%
February	$334,397	$314,436	0.80%	0.86%
March	$387,393	$305,366	0.84%	0.79%
April	$397,066	$324,398	0.98%	0.86%
May	$445,107	$379,145	1.00%	0.97%
June	$430,609	$347,184	1.03%	0.96%
July	$434,881	$409,085	1.02%	1.09%
August	$433,750	$415,231	1.00%	1.06%
September	$418,298	$358,955	1.03%	1.05%
October	$446,003	$386,382	1.04%	1.05%
November	$415,103	$357,686	0.97%	0.97%
December	$446,282	$408,649	0.98%	1.03%
Grand Total	$5,056,995	$4,402,137	0.97%	0.97%

GL $



FRAUD PREVENTION

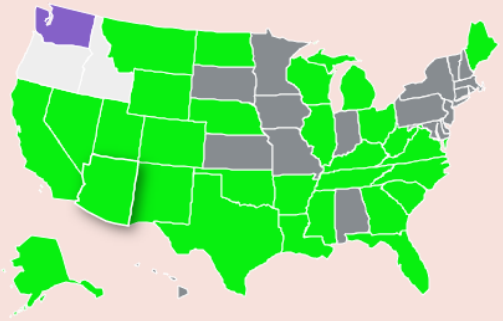
Working with the State of Washington to prevent fraud

Approved Transactions by Month

AUTH_IND	TRAN_DATE (Month of Year)	TRAN_COUNT	TRAN_AMOUNT
A	July	1,839	1,208,260.92
	August	1,469	1,096,157.95
	September	1,491	1,130,795.24

Map of States w/ Approved $'s

AUTH_IND: A



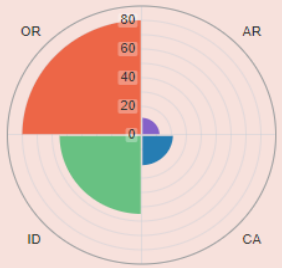
500 mi

Approved by State

MERCHANT_STATE	TRAN_COUNT	TRAN_AMOUNT
★ WA	4,482	3,233,422.95
OR	80	51,637.77
ID	55	37,411.32
OH	15	14,583.97
AR	29	13,685.11
CA	26	12,914.80
NV	9	10,567.24
AZ	13	9,649.99
CO	13	8,891.72
MT	7	8,586.96
NM	4	8,024.00
UT	6	5,550.65
TN	11	4,329.36
NC	3	2,426.24
LA	3	2,323.00
AK	3	2,318.46
TX	7	2,280.22
FL	4	1,977.42
KY	4	1,274.73
ND	3	600.00
OK	4	532.61
WI	1	481.29
WV	1	449.00
GA	2	294.25
ME	1	266.37
VA	7	258.89
MS	1	158.98
IL	1	97.04
NE	1	83.21
WY	1	61.15
MI	1	39.91

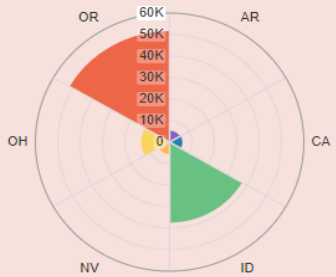
Approved # by State other than WA

TRAN_COUNT: 10 - 5,000 MERCHANT_STATE: AK, AR, AZ, CA, CO, FL, HI



Approved $'s by States other than WA

MERCHANT_STATE: AK, AR, AZ, CA, CO, FL, HI, ID, IL, IN, KY, LA... +17 TI



IMPLEMENTATION

OAC Implementation Challenges

User Acceptance
The average Certegy employee has over 16 years of service. Certegy installed Hyperion in 2002 (18 years of use). Risk processes over the years relied heavily on Hyperion reporting and queries.

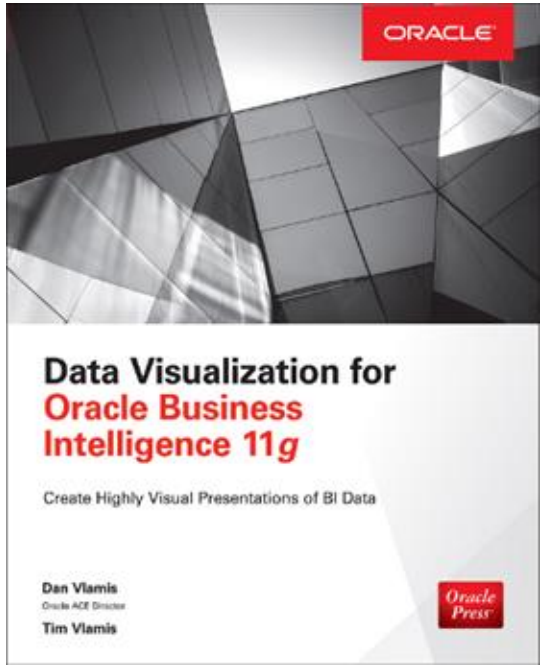
There was plenty of skepticism about changing systems.

Challenges
Training: including best practices.

Analysts were spending 90% of their time connecting data tables with Data Flows, which requires a certain technical skill set and was taking away from our mission; identifying and stopping fraudulent transactions.

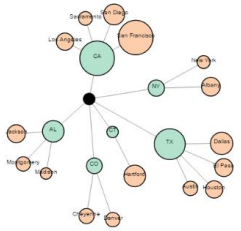
Game Changer
Vlamis Software Solutions evaluated our data and introduced us to Subject Areas

Vlamis Software Solutions



- Founded in 1992 in Kansas City, Missouri
- 400+ Enterprise Clients
- Consults in :
  - Enterprise Business Intelligence & Analytics
  - Analytic Warehousing
  - Machine Learning and Predictive Analytics
  - Data Visualization
  - ETL and data integration
- Vlamis consultants average 15+ years
- Creators of Force Directed Graph Plugin on Oracle Analytics Library
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book "Data Visualization for OBI 11g"

ORACLE ACE Director



vlamis™ SOFTWARE SOLUTIONS

Vlamis Contributions to Certegy

- Assisted with transition from Hyperion to OAC
- Assessed use of data sets and data flows at Certegy
- Suggested using RPD Subject Areas and trained on SA's
- Improved performance of some data sets and data flows
- Created common time dimension
- Implemented summary tables to improve performance

- Supported Certegy's implementation on adhoc basis

vlamis™ SOFTWARE SOLUTIONS

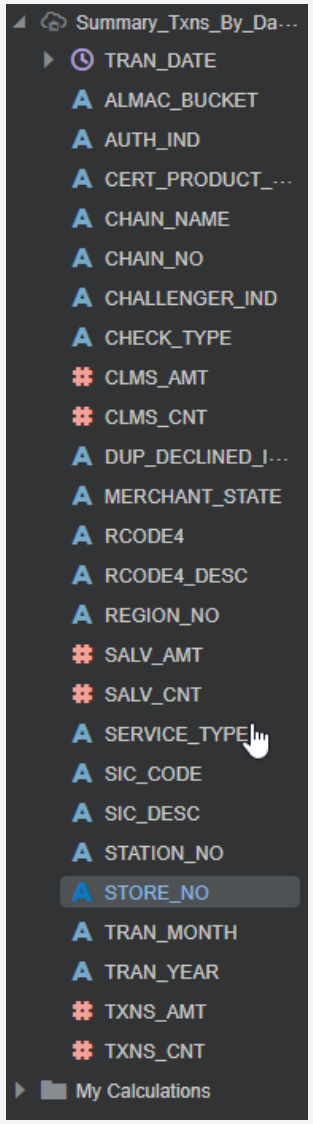
Risk Analytics – Subject Areas

certegy

BUILDING SUBJECT AREAS

Need

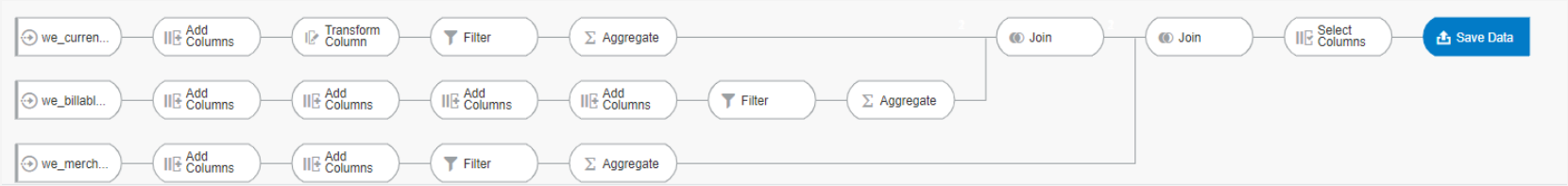
- Certegy team is experiencing unacceptable performance when building and running analytics in OAC.
  - Dashboards and DV projects are taking a while to display.
  - List of filter values are taking several seconds, sometimes over a minute, to populate.
  - Long running queries are impacting the database.



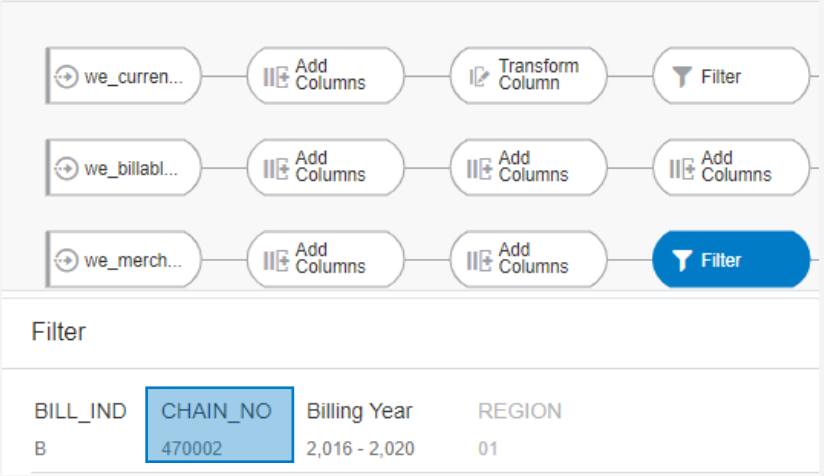
BUILDING SUBJECT AREAS

Need

- Complex data flows are created to merge data.



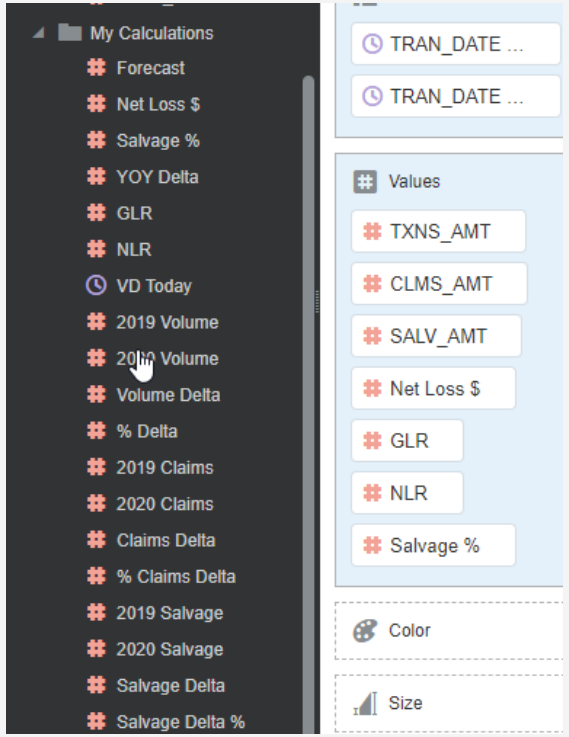
- Due to limited dataset sizes, data has to be separated into smaller sections.

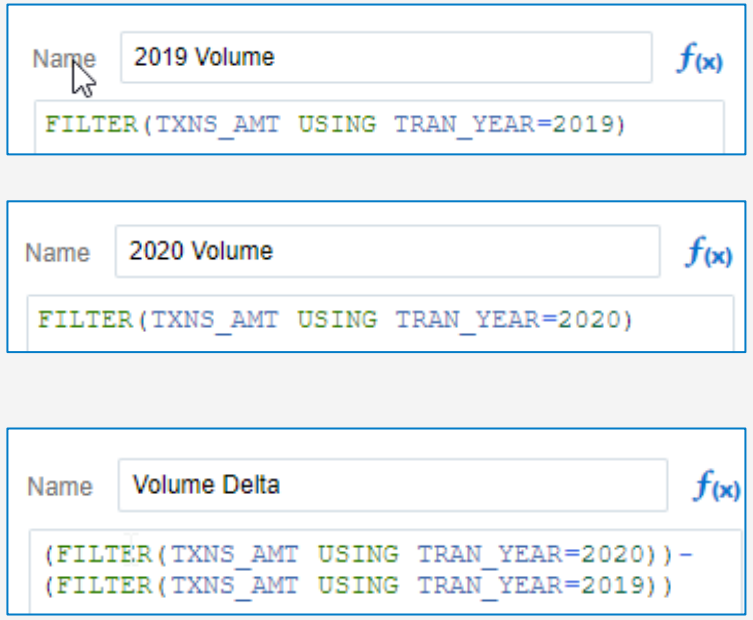


BUILDING SUBJECT AREAS

Need

- Calculations within projects with limited sharing capabilities and hardcoding of years within calculations will require updating each year.

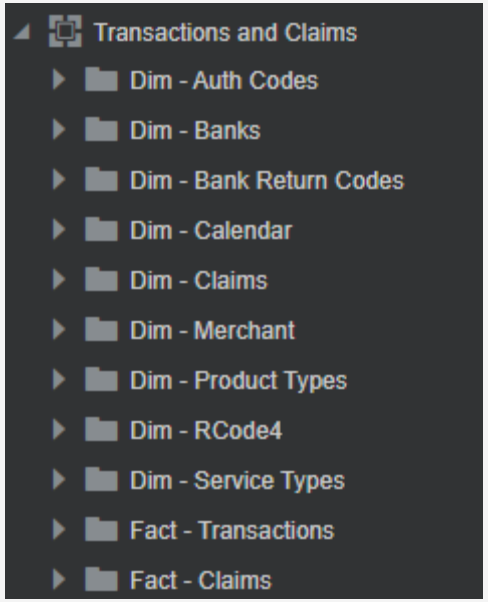




BUILDING SUBJECT AREAS

Solution – Proof of Concept

- Build a star schema model in the OAC repository, taking advantage of the following:
  - Dimensions for hierarchical drilling and aggregation
  - Dimension tables for faster filtering and reporting on dimension attributes
  - Fact tables with only facts
  - Summary table for query performance on higher level analytics
  - Calculations defined once
  - Time series functions available in the RPD (repository)



BUILDING SUBJECT AREAS

Results

- Certegy will easily share common calculations across the organization.
- Future maintenance of current projects and analyses will be reduced significantly.
- Analytics display will return in seconds with minimal wait for the filter list to populate.
- DBA will make fewer calls to the analytics team.

Happy Users,

Happy DBAs,

Happy Life!

BUILDING SUBJECT AREAS

Next Steps

- Complete the build of the current subject area.
- Work with the team to identify where the organization can take advantage of the use of subject areas.
- Design and build additional subject areas.

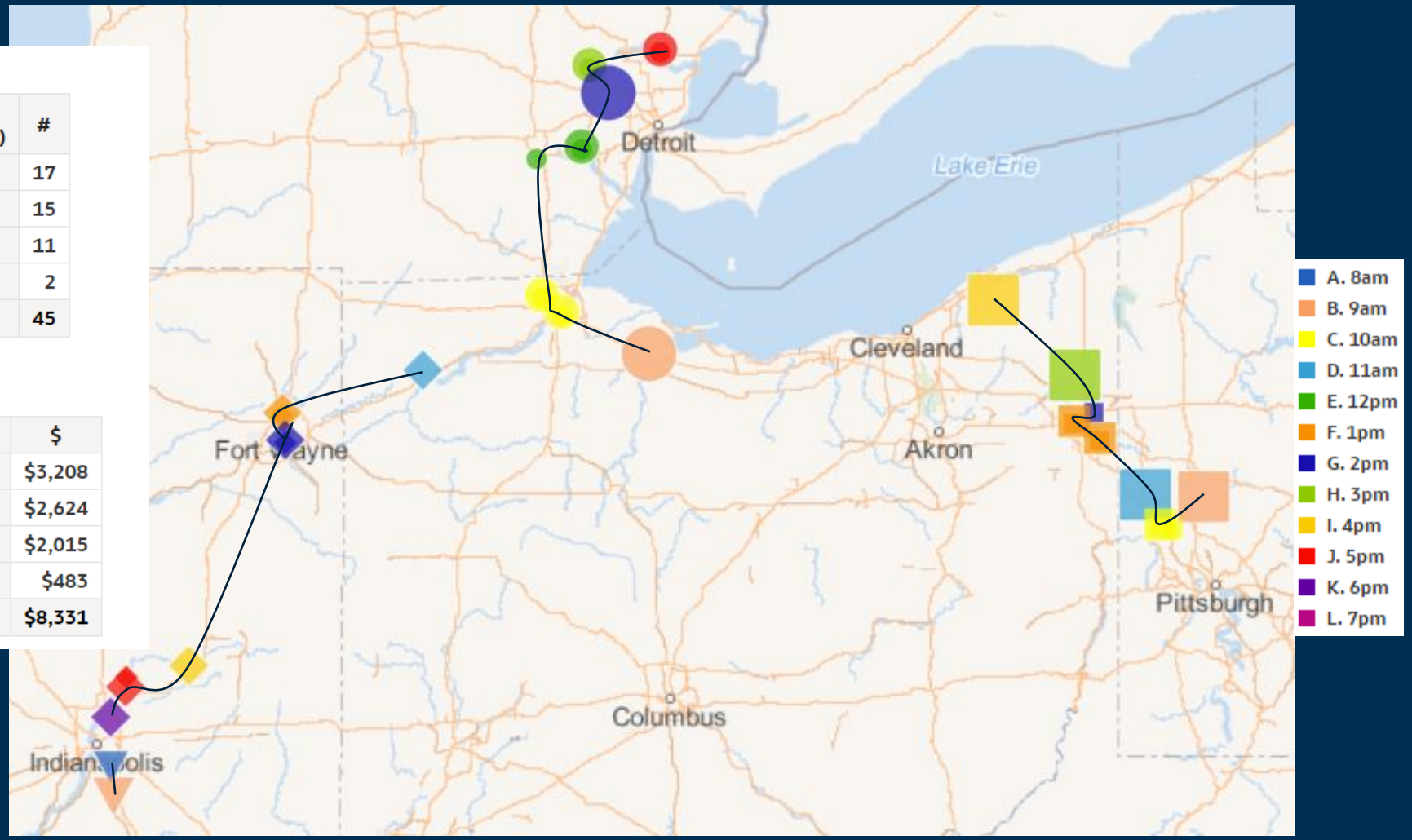
Risk Analytics Future

certegy

July 14th - July 17th 2020

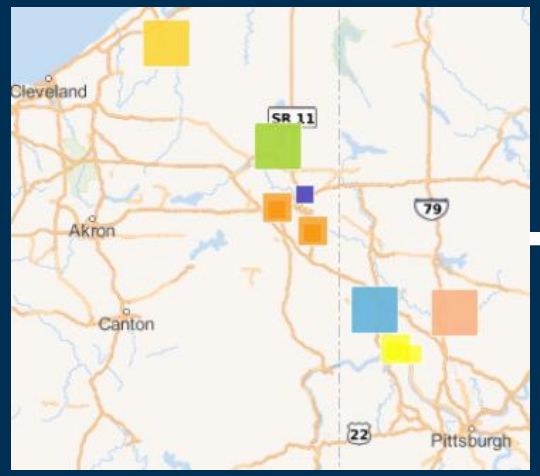
TRAN_DATE (Day)	TRAN_DATE (Weekday)	#
07/14/2020	Tuesday	17
07/15/2020	Wednesday	15
07/16/2020	Thursday	11
07/17/2020	Friday	2
Grand Total		**45**

TRAN_DATE (Day)	TRAN_DATE (Weekday)	$
07/14/2020	Tuesday	$3,208
07/15/2020	Wednesday	$2,624
07/16/2020	Thursday	$2,015
07/17/2020	Friday	$483
Grand Total		**$8,331**

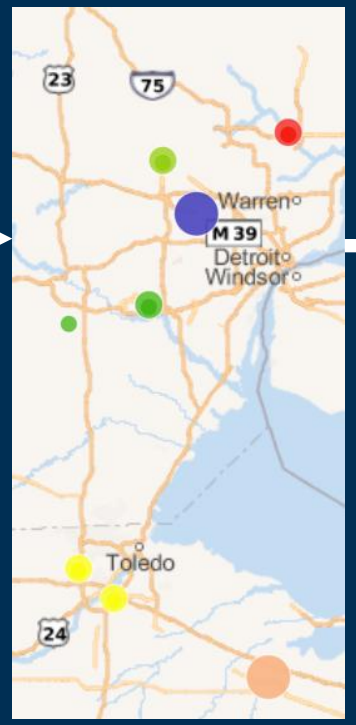


certegy

Tuesday $3.2K



Wednesday $2.6K



Thursday $2K

Friday $0.4K

- A. 8am
- B. 9am
- C. 10am
- D. 11am
- E. 12pm
- F. 1pm
- G. 2pm
- H. 3pm
- I. 4pm
- J. 5pm
- K. 6pm
- L. 7pm

certegy

SPEEDY GONZALEZ CASE

True life example

Speedy Gonzalez — who was wanted on forgery charges — is now in jail in Gwinnett County

By Curt Yeomans curt.yeomans@gwinnettdailypost.com Apr 27, 2020 0

Speedy Gonzalez — the man wanted on forgery and other charges, not the cartoon mouse — has been arrested and is sitting in the Gwinnett County Jail.

Gwinnett police said Gonzalez, 35, a resident of Buford, was arrested Saturday at the Embassy Suites located at 2029 Satellite Boulevard on Saturday. A Flock camera at Sugarloaf Parkway and Meadow Church Road had alerted police that a vehicle driven by a wanted suspect had passed by it.

Police responded to the area saw the vehicle near the intersection of Sugarloaf Parkway and Satellite Boulevard and witnessed it driving behind the Embassy Suites, where he got out of the car and attempted to enter the hotel through a back door.

Latest

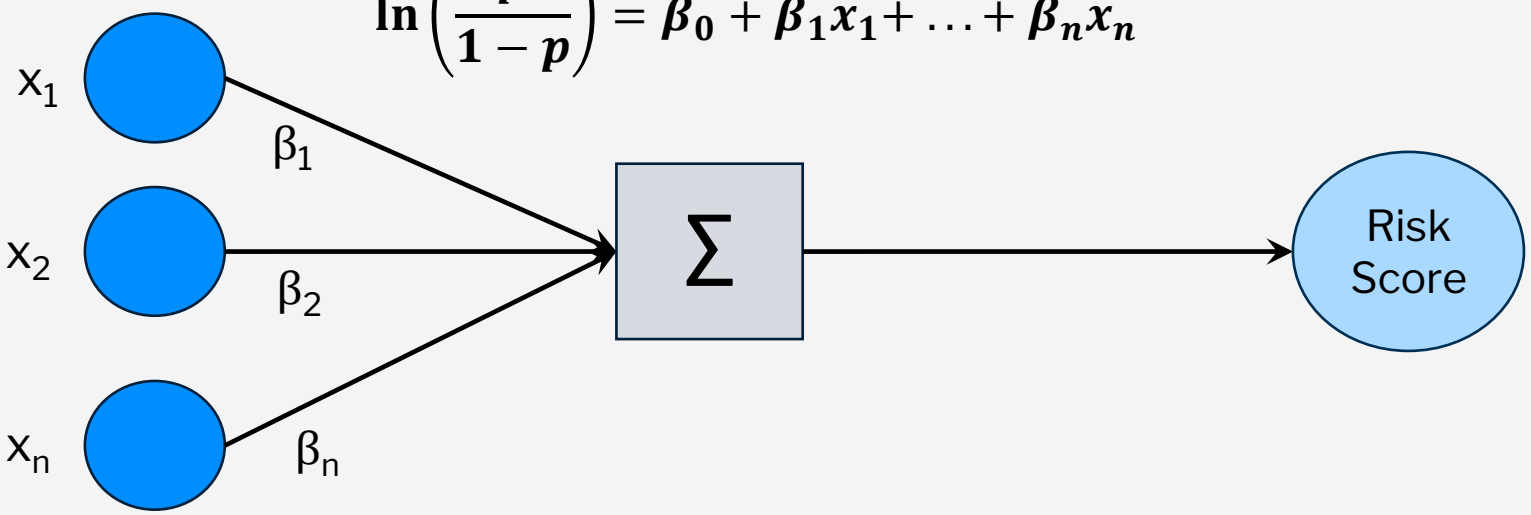
- Lawrencev Gwinnett C 2021
- Felony rioti Black fema
- Ongoing sl numbers c
- Trump calls until after E

Machine Learning

TRADITIONAL RISK MODELS

- **Explainable logistic regression model generates score reflecting probability transaction results in financial loss**
- **Score generated real time and used to decision transactions**
- **Inputs include consumer history, account history, current time characteristics (ie time of day, transaction amount, etc.)**

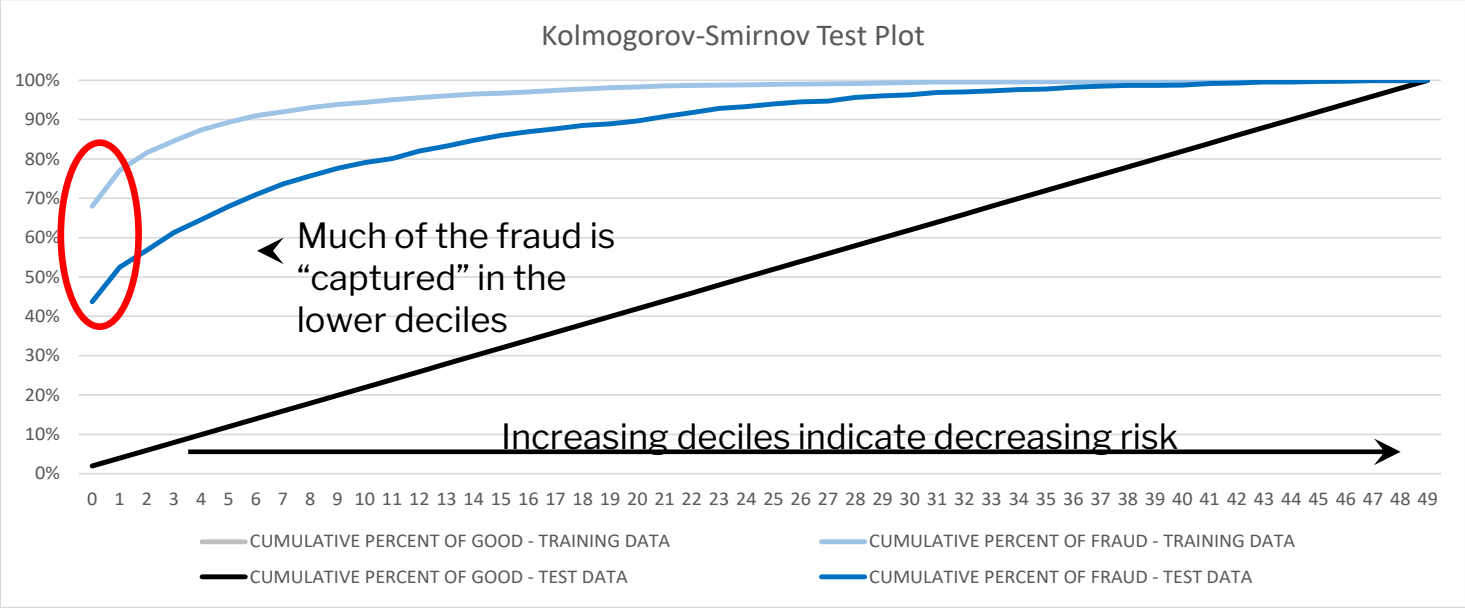
$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$



MODEL PERFORMANCE

Kolmogorov-Smirnov Test

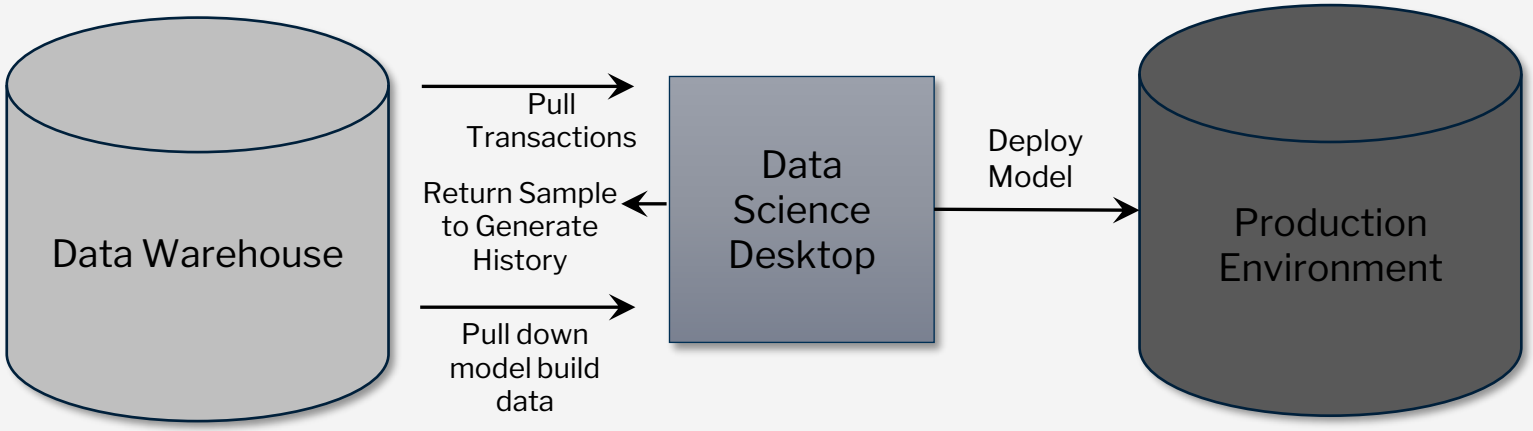
- **Model scores grouped in deciles**
- **Decile 0, bottom decile, represents the riskiest 2% of scores**
- **Goal it to capture 100% of fraud in bottom decile**
- **Will see this performance degrade over time**



BUILDING A MODEL

Data movement not ideal

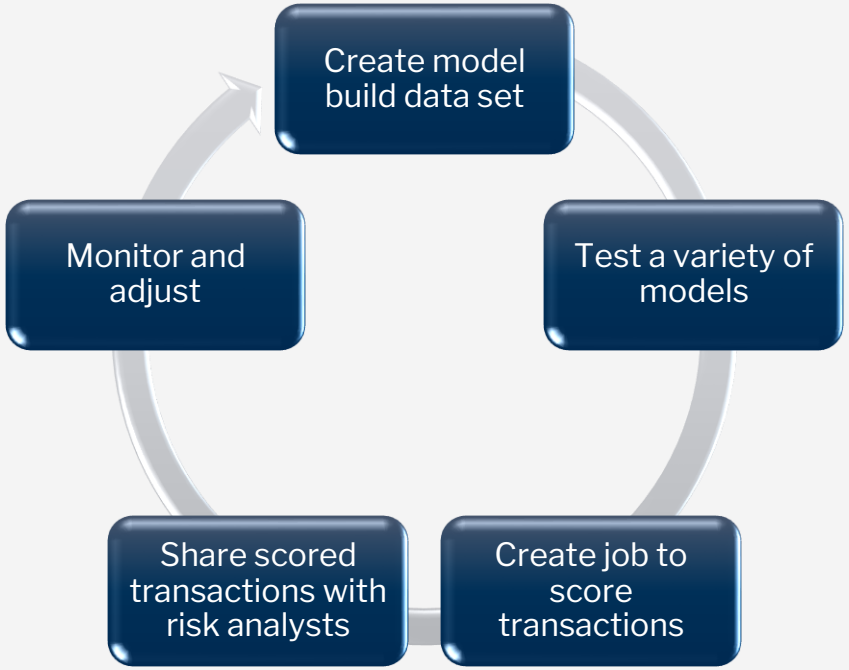
- **Transactions are pulled from the data warehouse and a sample is made**
- **To generate model attributes sampled data is returned to the data warehouse to query history**
- **The final data set is used to build a model on the data science desktop**
- **GUI used to put model deployment files in production**



NEW TOOLS FOR BUSINESS

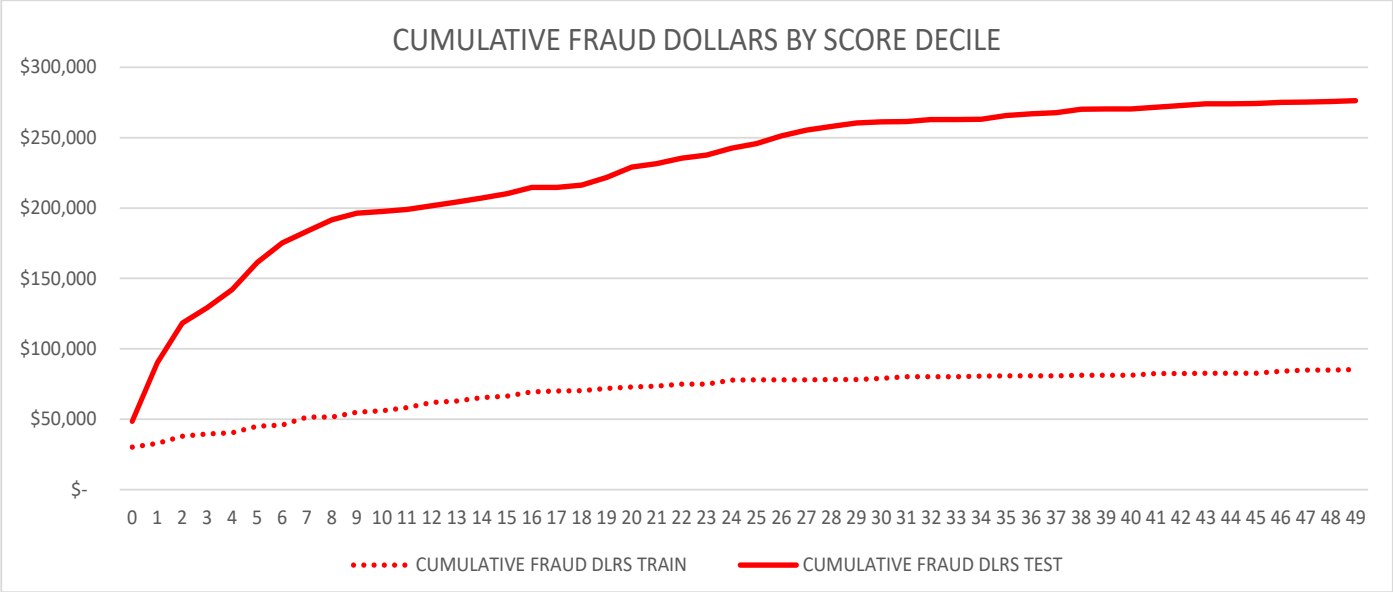
Oracle Machine Learning

- **Now the data scientists can work in the cloud**
  - Data resides on the cloud
  - Machine learning tools also available on the cloud
- **Eliminate need to move data to a desktop computer**
- **Can create automated jobs on the cloud**



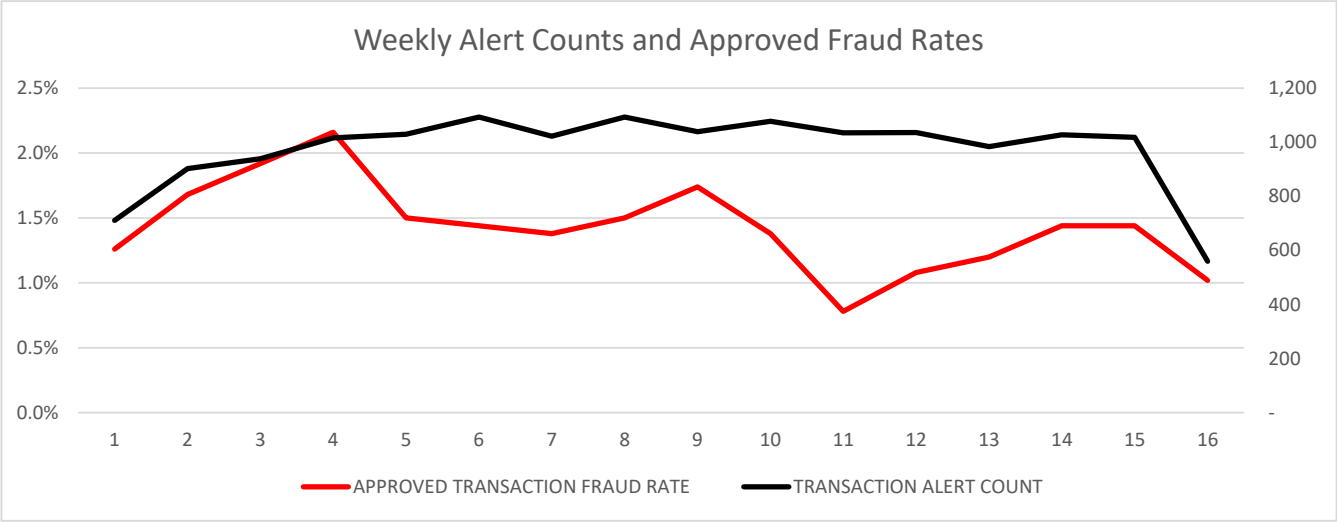
ADDITIONAL MEASURE OF PERFORMANCE

- **Performance in managing risk is based on fraud losses avoided**
- **Another way to show model performance by looking at fraud dollars captured by the neural network score deciles**
- **Notable for this model is that the train data set was a low loss month, the test data set was a month with a very active fraud ring**



HOW ALL THIS WORKS

- **Queries written in OML to pull yesterdays transactions and create model features**
- **Neural network scores transactions and inserts high probability of fraud transactions into table accessed by risk analysts**
- **Risk analysts review fraud alerts and create “fraud cases” to block future similar activity**
- **Performance measured based on the approved volume resulting in a net loss**



EXPLAINING NEURAL NETWORK RISK FACTORS

- **A con to neural networks is that they are very black box**
  - Regulators like risk decisions that can be explained
  - Risk analysts want to understand what makes a transaction high risk
- **OML offers the function PREDICTION_DETAILS**
  - Extracts top N features contributing to a high or low risk scores
  - Also provides values of those features
- **Risk analysts can use this to better understand the fraud**

Probability of Fraud	Risk Factor 1	Risk Factor 1 Value	Risk Factor 2	Risk Factor 2 Value	Risk Factor 3	Risk Factor 3 Value
0.12	FEAT_21	0.115	FEAT_66	5	FEAT_24	‘N’
0.31	FEAT_44	2100	FEAT_01	‘Y’	FEAT_12	16
0.05	FEAT_15	16	FEAT_36	0.001	FEAT_22	5241

Questions?