Data Wrangling: The New Focus for Financial Institutions for Customer & Marketing Analytics

In banking, as with all industries, there are two ways to maximize profits—increase revenue or decrease costs. Using data and analytics to better understand your customers is one way to do both.

A decade ago financials institutions were at the vanguard of customer analytics. While other industries were limited, at best, to the biased data from customer loyalty programs, banks had it all—virtually everything their customers did within the four walls of the bank was at the fingertips of the banking analyst.

But now, all the low-hanging fruits have been picked. Any bank worth its salt has a decent customer profitability model and a Pareto chart to go with it. A/B testing for the effectiveness of single campaigns and targeting models for specific products have been optimized. And all banks have developed sophisticated attrition and churn models. (Although most would probably admit that they've struggled to successfully productionalize one into a real-time customer retention program.)

At the same time, other industries' customer analytics capabilities have caught up, and with them customers' expectations have increased. Customers are now used to the service that they receive from sophisticated data-driven companies like Amazon, Uber and Airbnb and they expect no less from their banks.

But catching up is not easy due to the "perfect storm" of challenges facing the financial services industry when it comes to data and analytics.

Financial institutions face unprecedented challenges in today's "new normal" environment

Challenge #1 - Old Data

What was once their advantage when it came to data, is now a big challenge—Banks' big legacy systems are often siloed by years of acquisition lineage and separately run lines of business. To make things more difficult, these systems are often run on disparate technology ranging from older mainframe systems to traditional data warehouses to 3rd party bureau data. In most cases, these systems use different customer and product identifiers, different aggregation hierarchies and different data formats which produce disparate results that are not easy to aggregate.

> " Research has shown that 85% of analysts' time is spent hunting for and gathering information, with only 15% of their time left to perform the kinds of analytical functions they were hired to do. "

https://www.clearpeak.com/data-driven-culture-let-analysts-analyze/

ClearPeak Consulting

April 28, 2016

Challenge #2 - New Data

And while still struggling with these cumbersome legacy systems, bank execs are having to provide leadership on new types of data—exploding volumes of data and unstructured data from new sources such as email, CRM systems, GPS, chat logs, blockchain and social media to name a few. Add to that the strict regulatory requirements for segregating, storing and securing the data, and the need for simplification and transparency into the data becomes paramount.

Challenge #3 - Regulations

While the financial services industry has always been a heavily regulated market, the meltdown of 2008 has, in particular, put banks' data and analytical models under the regulator's magnifying glass. Reforms such as the Volcker Rule and Basel capital requirement rules were enacted to prevent another crisis. However, it was the 2010 creation of the Consumer Financial Protection Bureau (CFPB) as part of the Dodd-Frank reforms that have had the biggest impact on customer and marketing analytics in the banking industry. Any new marketing initiative mandates to demonstrate that this respects some kind of regulation and can be audited.

Challenge #4 - The War for Talent

And lastly, whereas 10 years ago working at big bank after college was a desirable career path, these days most data science graduates from elite schools and programs are far more interested in landing their dream job at a tech company where they can have free lunch, take onsite yoga classes and wear jeans every day.

How to overcome these challenges?

Amidst all of these challenges, banks still need to drive sales, increase retention, improve service and identify future product needs. Data and analytics are the key to getting there.

The problem facing banks isn't lack of data; banks have plenty of data. The problem is being able to meaningfully connect the data to uncover insights and tell stories.

As any analyst can tell you, it is hard to join data across disparate legacy systems with inconsistent identifiers. You need to have a deep knowledge of what you are looking for and are often required to replicate the data onto your own environment. Not surprisingly, this is cumbersome, error prone, expensive and redundant. And duplicating data also infers a risk of data leaks and illicit usage.

In today's world of new data sources, new formats, and modern data management platforms such as Hadoop, SQL is not enough. Without a small army of data scientists to write Spark or MapReduce jobs, much of this data will be left untouched. Is it any wonder that analysts spend most of their time structuring and joining data?

And to complicate matters further, at any moment, a risk partner, auditor or legal counsel may stop by your desk to ask about a model or data source requiring you to trace data from system of record (SOR) to ultimate use.

How does Trifacta change the customer analytics game?

There are four key capabilities that can be solved by the Trifacta solution:

1	2	3	4
Data Quality Management	Data Blending Diversity	Reporting Response Timeliness	Data Lineage

> " In order to have a visual interaction with highly unstructured data, you'd have to be a person with thirty years of experience in data transformation technology. Trifacta data wrangling gives you that kind of a view on the problems out of the box. … It's not only that it has increased productivity, it's that we are now addressing things that we wouldn't even do before because it would just take too long. I can't put a value on how much it has improved how I work! "
>
> Spyros Marketos
> Data Scientist
>
> Royal Bank of Scotland

Data Quality Management

One reason why analysts spend so much time preparing data is that **data quality issues will undermine even the most sophisticated analysis**. If the data isn't right, you might as well throw the whole model out the window. This is exacerbated when teams are working with unfamiliar data from business lines where they are not deep subject matter experts.

Imagine an analyst in a large bank, trying to unlock the value contained within chat logs—petabytes of unstructured data. Where to start to find useful insights into customer service issues?

Trifacta provides a way to scour through that data—easily performing **data discovery** (understanding data types, trends and anomalies), **data structuring** (data formatting and pivoting), and **data cleansing** (uncovering missing data pieces, errors and defining rules to correct it).

Instead of relying on an army of expensive and hard to hire data scientists to write endless lines of code that may run for days—the analyst can visually pore through vast swathes of data to quickly discover patterns, uncover issues and put together actionable recommendations. **Trifacta puts that power at the tips of their fingers saving potentially days or months worth of work.**

Data Blending Diversity

As banks' analytics teams reach to the next level of customer insights, they need to push the data aggregation envelop over many diverse categories. For example:

Siloed Data

It's quite likely that at some point or another you have been involved in building a customer Financial **Lifetime Journey or Customer Lifetime Value Model**. However, the data you can access is likely limited to a high-level aggregation; for example, the number of credit and debit card transactions per household—virtually useless for any rigorous analysis. However, imagine the goldmine of insights waiting to be uncovered if you could combine, for example, full census **transactional card data** to other customer information—tenure, profitability, ACH transactions, deposit products, wealth management, mortgage status and other 3rd party data. By connecting this traditionally siloed and inaccessible data, imagine the triggers you will be able to identify as part of the customer journey and the next-best-product recommendations you will be able to provide.

Locked-In Big Data Sources

Banks have to extend their analytics to personalize the online and mobile banking experience. How can you possibly access and blend **log-level click stream data** from sources like Adobe Site Catalyst or your bank's Hadoop environment with your bank's customer data? And how can you crunch all that data in near real-time to help inform and guide the agile development process in a meaningful way? Most of the time this data would be in a shape and a location not accessible by the ordinary user. What if you could access, explore, and blend that massive amount of data yourself?

Unstructured Data

You might have voice-to-text data from call centers that is coming in as pulses, or image-to-ocr data from documents scanned in through mobile. You might have social data streaming in from multiple sources, with disparate features and undefined formats. What you need is a visual dimension into all of these data sources, allowing you to quickly discover and act on the information hidden within.

Trifacta's **data enriching features** (combining multiple data sources and types) solve the challenges of siloed data, big data and unstructured data. It replaces the manual integration processes with a seamless connection across multiple platforms ranging from traditional data warehouses to Hadoop clusters to excel files. Trifacta leverages Artificial Intelligence to guide users through the process of joining disparate sources and delivers an intuitive drag-

> " The dashboard is transforming the way I run my business. It is improving the customer centric approach in our chats and it is showing in the output. We are now able to coach our staff much more effectively and it is based on the themes that are coming from the application. "
>
> Akshay Vats
> Head of Web Chat Operations
>
> RBS

and-drop interface, mitigating the need for manual queries or data replication. This solves one of the biggest challenges that analytics team face—joining large volumes of data across siloed systems in disparate formats. Now the analytics team can spend less time munging their data and more time creating compelling visualizations, building predictive models, and identifying key insights.

Reporting Response Timeliness

As any analyst knows, **time is of the essence** when it comes to actionable insights. Whether it be a fire-drill from your bank's CEO or a recurring report, analysts need to quickly respond and collaborate to maintain the confidence of their executive team and stay relevant in the organization.

Trifacta helps analysts by: 1. **Building repeatable workflows** of what is typically a manual and time-consuming process of data preparation 2. Allowing for **collaboration across teams** to share context and knowledge and 3. Quickly **publishing to common data visualization tools** such as Tableau and Qlik.

The last thing any banking executive wants to hear is that they must wait days or weeks to get answers to their pressing questions.

Trifacta ensures that questions can be quickly answered and problems can be quickly solved with great confidence in the fidelity of the data.

Data Lineage

Thanks to the increased regulatory and compliance environment since the 2008 financial crisis, the CFPB (Consumer Financial Protection Bureau) is playing an increasingly large role in even seemingly innocuous analyses. At any moment, an analyst could be called upon to explain a model or report to ensure that, for example, no protected variables have been used.

The truth is that today, most analysts with their files of excel spreadsheets are simply not equipped to answer such data audit questions. This leaves banks exposed to CFPB violations and penalties.

Trifacta's interactive data lineage visualizations provides a remedy to this burdensome problem. Trifacta gives analysts the ability to trace data across the lifecycle from data sourcing, through blending, aggregation and finally reporting. The feature allows users to see downstream dependencies as well as filter on specific types of dependencies.

Conclusion

At Trifacta, we know that deeply understanding banking customers is the only way for financial services companies to optimize revenue and expenses to maximize profitability.

All banks in today's environment are facing daunting challenges including how to manage old and new data, the complex regulatory environment, and the war for talent. However, only those who embrace data and analytics hold the key to succeeding and thriving in this new normal.

Successful banks will not only drive sales, increase retention, improve service and identify future product needs but by using a data wrangling solution like Trifacta they will empower their analytics teams to spend less time manually massaging their data and more time delivering analysis and insights.

Do you want to spend less time preparing your data and more time analyzing it?

Go to trifacta.com to learn more.