



Big Data: From Transactions, To Interactions

Martin Willcox [[@willcoxmnk](#)], Director Big Data Centre of Excellence (Teradata International)

April 2016





Agenda

- Beyond transactions
- Riding the three waves: use-case examples
- Succeeding with Big Data
- Afterword



Beyond transactions

Big Data circa 1986



Big Data's three new waves

1

Analysis of web / clickstream data enables Google, Amazon, eBay and others to achieve "mass customisation".

People interacting with things

2

Clickstream / social / mobile interaction data enables Amazon, LinkedIn, Netflix, etc. to go social ("people who like what you like also like...")

People interacting with people

3

Increasing instrumentation is now leading to the emergence and optimisation of "the Internet of Things".

Things interacting with things

From transactions and events – to interactions and observations

Simple computing devices are now so inexpensive
that increasingly everything is instrumented;

Instead of capturing transactions and events and analysing them to
infer behaviour (for example, of customers and complex systems like
value chains), we can increasingly measure and analyse it directly.

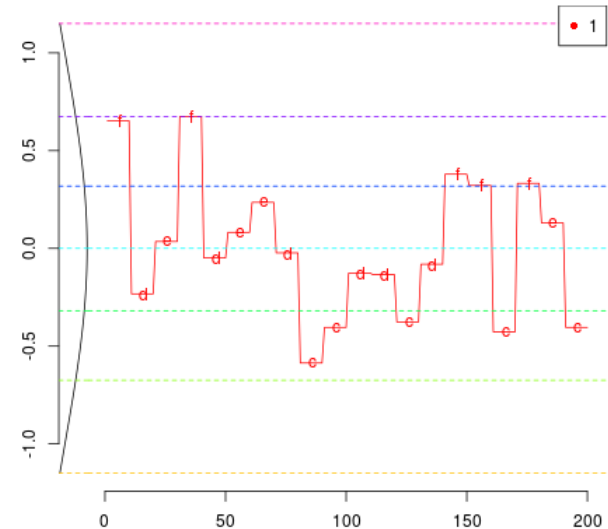
Big Data = “New” Data + “New” Analytics

“NEW DATA”

```
THORISATION FOR A CHALLENGED TRANSACTION BY ONE
STOMERS.XXXXXX ADV WE SHOULD HAVE WRITTEN TO ADVIS
STOMER WAS CHALLENGING TRANSACTION AS PREVIOUSLY
HAVE DONE. XXXXX ADV HAS THIS SLIP CUSTOMER SIGN
THORISE THE TRANSACTION. XXXXX WANTS THE $750 PLU
VERSED OR IS WRITING TO THEOMBUDSMAN.
LTS TAKEN BY XXXXX
.01.2005 (12:58) XXXXX (08 8424 8552) SLCRU INBOX
EA RESP changed from - CRSLIN - SLCRU INBOX
- CRULIN - ULCRU INBOX
T changed from - CRSLIN - SLCRU INBOX
- CRULIN - ULCRU INBOX
VESTIGATE AS CUSTOMER IS DISPUTING THE TRANSACTIO
OM HIS ACCOUNT.
```

*Clickstream, Social, Call Centre Agent
Notes, etc., etc., etc*

“NEW” ANALYTICS



*Path, Graph, Time-Series, Pattern
Matching, etc., etc.*

A full-page background image of a surfer riding a large, curling blue wave. The surfer is wearing a dark wetsuit and is positioned near the base of the wave's face. The water is a deep blue, and the wave's crest is white with foam. An orange horizontal bar is overlaid across the middle of the image, containing the title text.

Riding the three waves: use-case examples

Understanding what customers are *really* searching for online

1

Up to **88%** of purchasers on some ecommerce sites search for products.

75% view just the first page of results. **35%** view just the top 3 results.

Search is becoming more-and-more **important** as more-and-more online customer journeys are made using smartphones and mobile devices.



Understanding what customers are *really* searching for online

1

What Does Under Performing On-site Search Behavior Look Like?

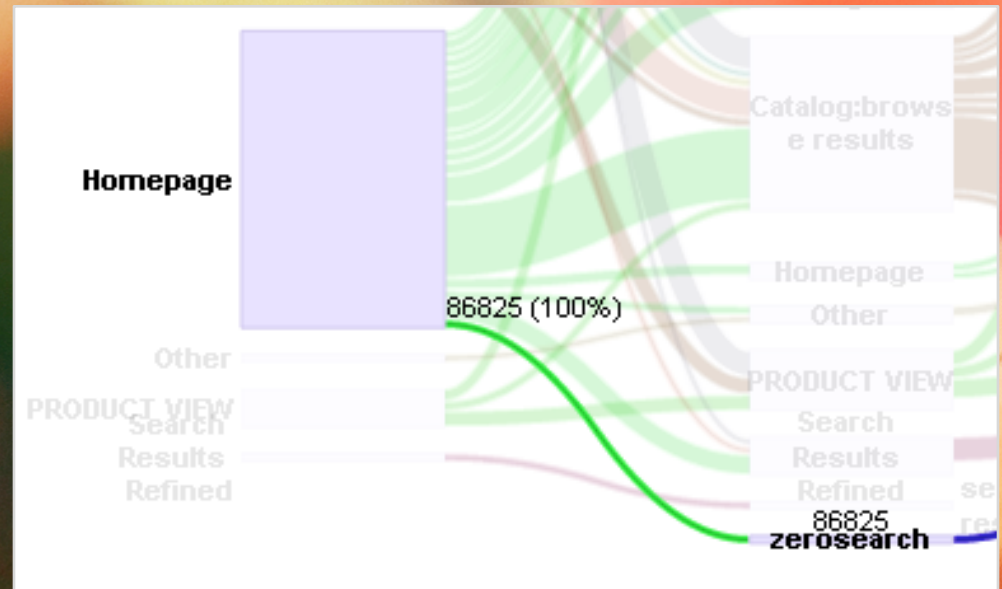
Queries that Return Zero Results

Immediate Exit
After Initial Search

Multiple Search Attempts

Search as the Last Event of Session

No Conversion



Understanding what customers are *really* searching for online

1

Search "toddler long sleeve top"

Toddler long sleeve top

x



And zero results returned... yet

Narrow your search

0 items

Search Results

toddler long sleeve top

Long-Sleeved

Back to top

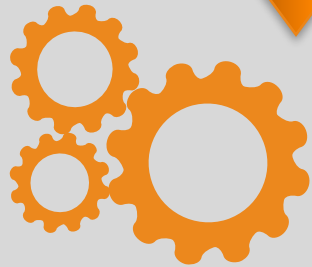


Understanding what customers are *really* searching for online

1

- Web logs with customer search words & queries
- Post search navigation and conversion


- Predict keywords by product
- Generate file with keywords by product
- Integrate keywords in content management



Text Analytics

1. Text-Parser (Tokenization)
2. TF-IDF
3. Cosine Similarity
4. Naïve Bayes Text Classifier
5. nPath
6. cFilter
7. Graph
8. SVM (Support Vector Machines)

Update
unlimited
products and
improve
unlimited
queries per day

mixer| 

Search Suggestions

- mixers in **Electrics**
- mixers in **Food**
- mixers in **Cooks' Tools**

- stand mixers
- hand mixers
- kitchenaid artisan stand mixer
- kitchenaid mixer
- gluten-free mixes
- kitchenaid stand mixer pasta roller attachment
- bread mixes

On-site search **optimisation** typically results in **2.0% or better improvement in search conversion** – which often represents millions of dollars, even for medium-sized eCommerce sites.

Realising cross-selling opportunities with smart recommender systems

2

Recommendations account for up to **30%** of sales at Amazon, **50%** of connections made on LinkedIn and **75%** of viewings on Netflix.



Realising cross-selling opportunities with smart recommender systems

2

Customers

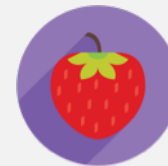
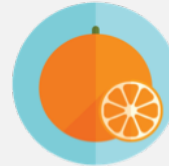
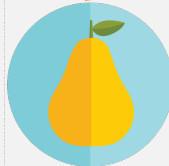
Customer 1

Customer 2

Customer 3



Products



Recommended to Customer 2

Recommended to Customer 1

Understanding customer sentiment

2

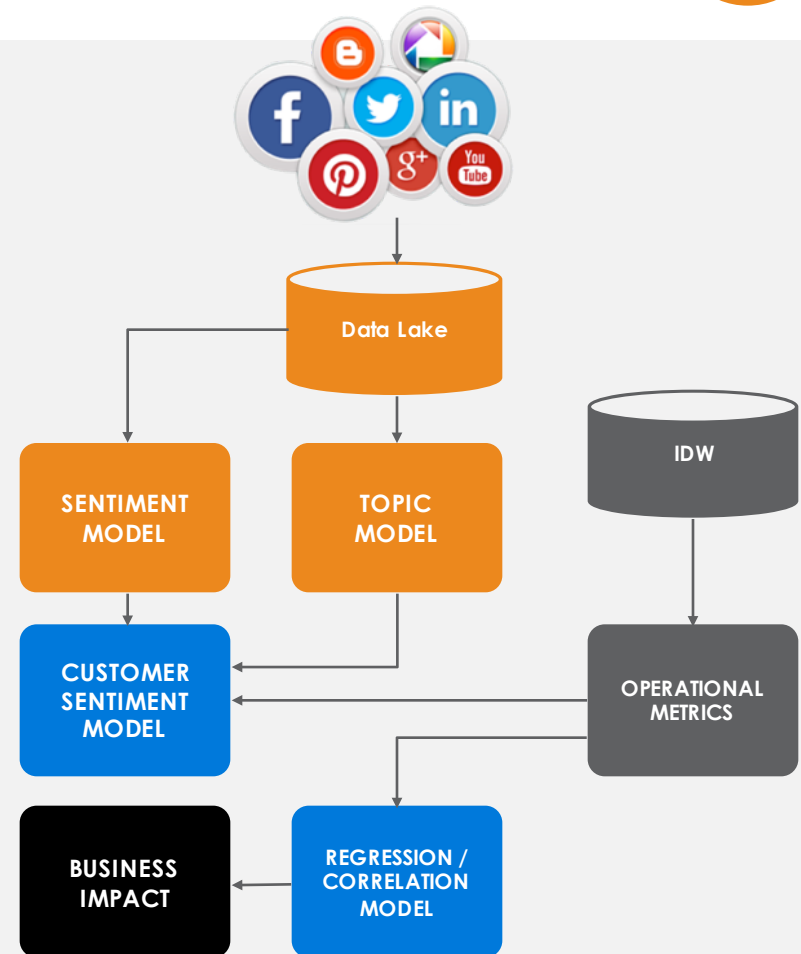
Sentiment Analysis can provide **more accurate** and **more timely** assessments of store and colleague performance than “mystery shoppers” – and statistically significant correlations with store **operational and financial measures** that enable the financial impact of these issues to be understood.



Understanding customer sentiment

2

- **Social feeds** with relevant hashtags and/or from particular geographic locations, restaurant reviews, etc., etc., **are prepared for analysis** by parsing them into individual sentences, removing stop words, etc.;
- A set of **randomly selected sentences** are labelled by a human expert;
- Using **sophisticated text analysis** and classification techniques, the labelled sentences are used to train both a Topic model and a Sentiment model;
- The **trained models** are used to **score** the remaining social feeds and reviews



Understanding customer sentiment

2

Visualising sentiment scores by review topics as a heat map enables the rapid identification of problem areas

Average Sentiment

-2.000

2.000

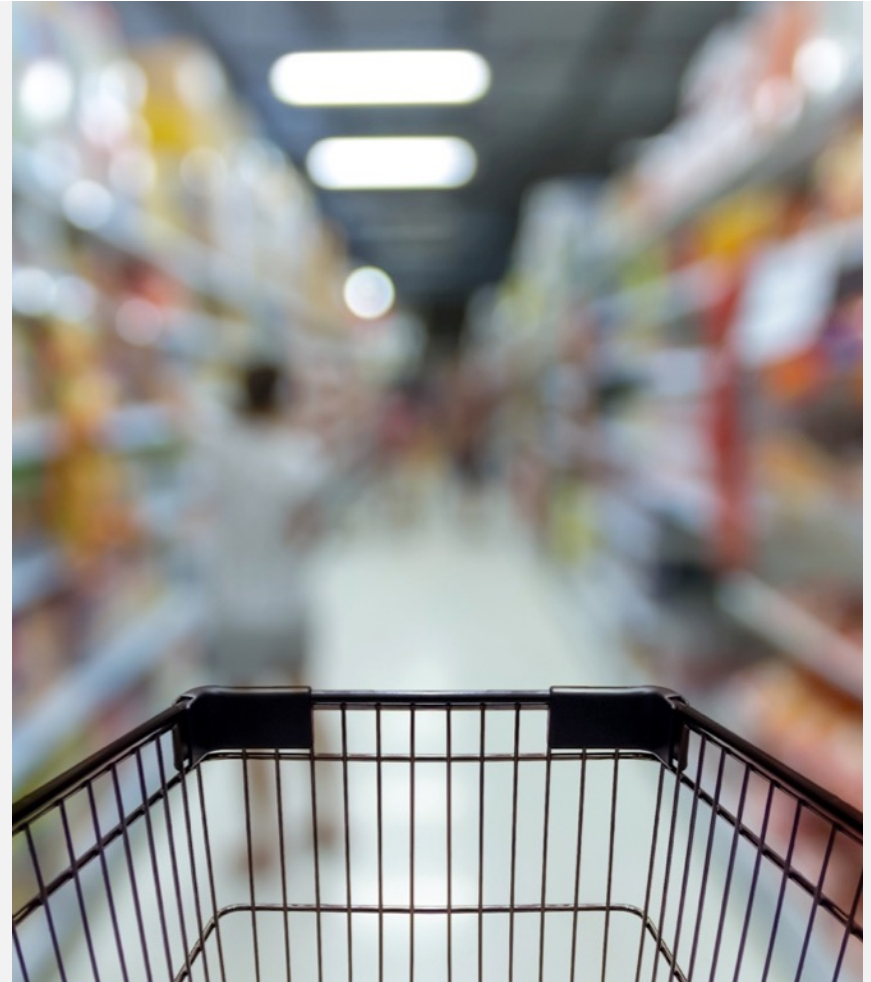
BUSINESS LOCATION	ASSORTIMENT	LOCATION	OVERALL	PRICE	Quality	Service
Trumpton	Blue	Blue	Blue	White	Blue	Orange
Camberwick Green	White	Blue	Light Blue	Blue	Blue	Orange
Chigley	Blue	Blue	Grey	White	White	Light Blue
Sodor	White	Orange	Blue	White	Orange	Orange
Beauchief	Blue	Blue	Blue	Orange	Grey	Light Blue
Broomhill	White	Light Blue	Grey	Grey	Orange	Light Blue
Burngreave	White	Light Blue	Blue	White	Blue	Orange
Crookes	White	Light Blue	Blue	White	Blue	Orange
Ecclesall	White	Blue	Blue	Light Blue	White	Orange
Fulwood	White	Light Blue	Grey	Blue	White	Light Blue
Ranmoor	Blue	Light Blue	Light Blue	White	White	Light Blue
Lodge Moor	White	Light Blue	Light Blue	Blue	White	Orange
Southey	White	Orange	Light Blue	Grey	Light Blue	Orange
Ranmoor	White	Light Blue	Light Blue	White	Blue	Grey
Walkley	White	Orange	Blue	White	Orange	Grey

Average Sentiment Score broken down by Sentiment Category vs. Business Location

Understanding *offline* customer journeys

3

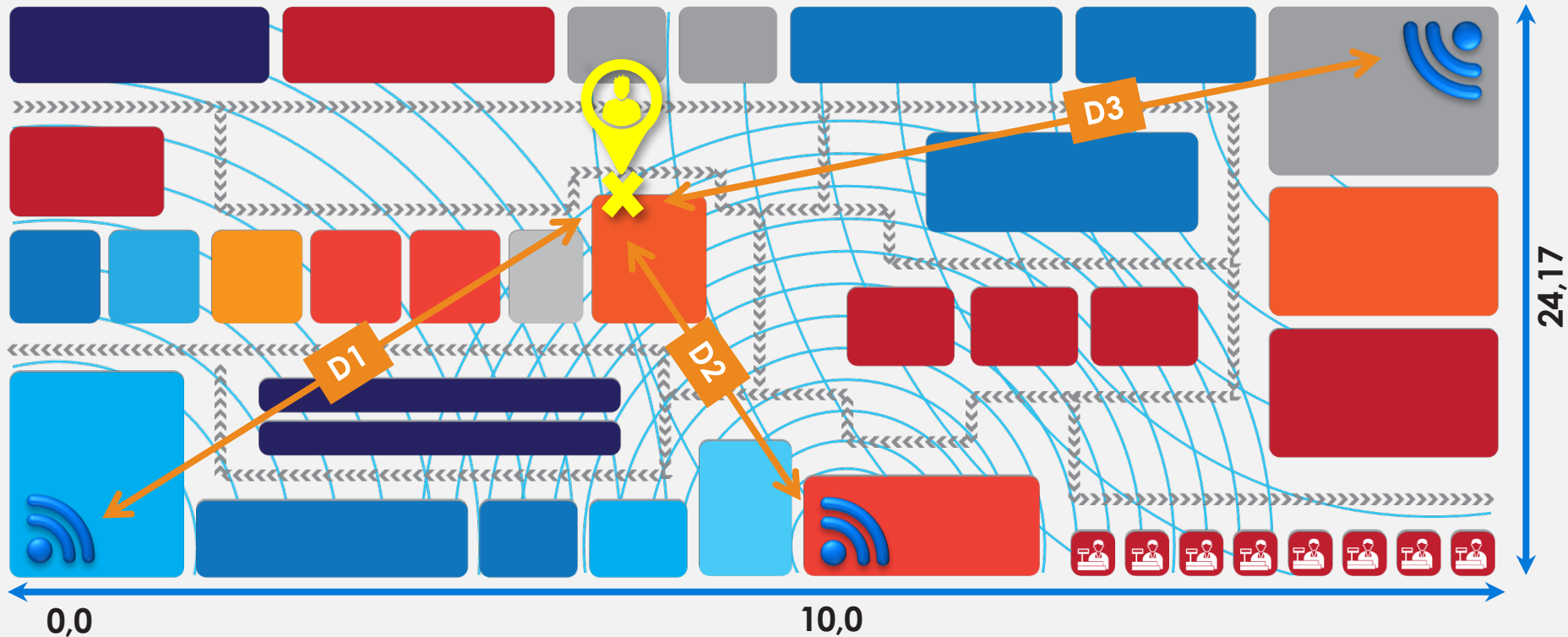
- Which paths do **customers who convert** take through the store?
- Which paths do **customers who don't buy** take through the store?
- Which departments do customers **visit multiple times**?
- How many customers are in **different parts** of the store at different times of day?
- And how many **staff**?



Understanding customer journeys

How do customers shop (physical) stores?

3





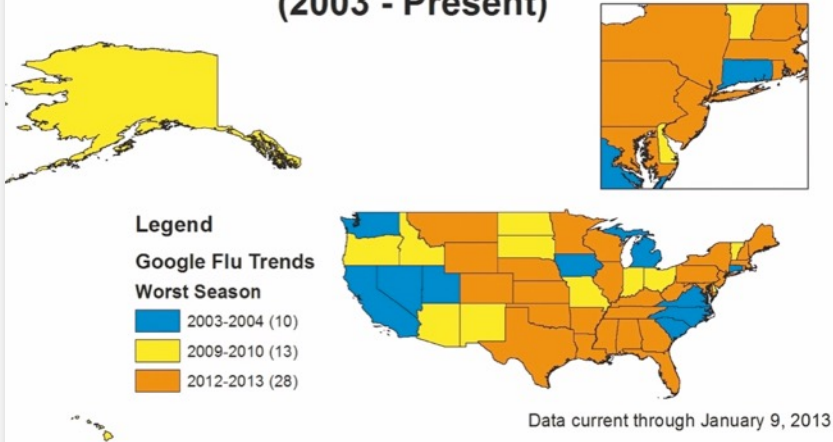
Succeeding with Big Data



Embrace “new” data-sets and Analytics without just discarding the “old”

Either / Or

Google Flu Trends Data (2003 - Present)



Ignored existing data-sets and models; consistently over-estimates number of cases and predicts winter, rather than flu.

Both And

The Future of Prediction:

How Google Searches Foreshadow Housing Prices and Sales

Lynn Wu
 The Wharton School
 3730 Walnut St. JMH 561
 Philadelphia, PA 19104
wulynn@wharton.upenn.edu

Erik Brynjolfsson
 MIT Sloan School of Management
 50 Memorial Drive, E53-313
 Cambridge, MA 02142
erikb@mit.edu

First draft: March, 2009; This Draft: August, 2013

Comments Welcome

Abstract

We demonstrate how data from search engines such as Google provide an accurate but simple way to predict future business activities. Applying our methodology to predict housing market trends, we find that a housing search index is strongly predictive of future housing market sales and prices. For state-level predictions in the US, the use of search

Extended an existing model based on transactional data with new, Google search based features; significant uplift.

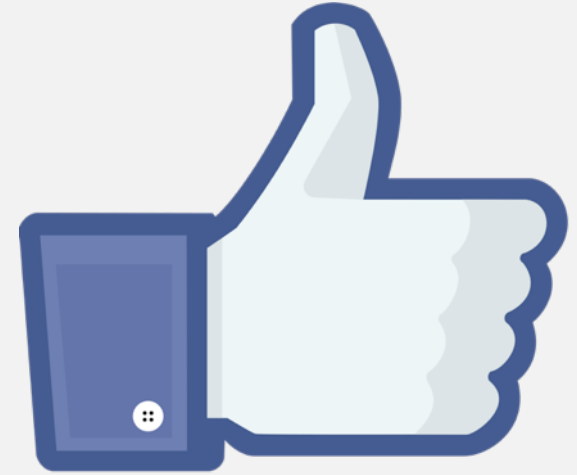
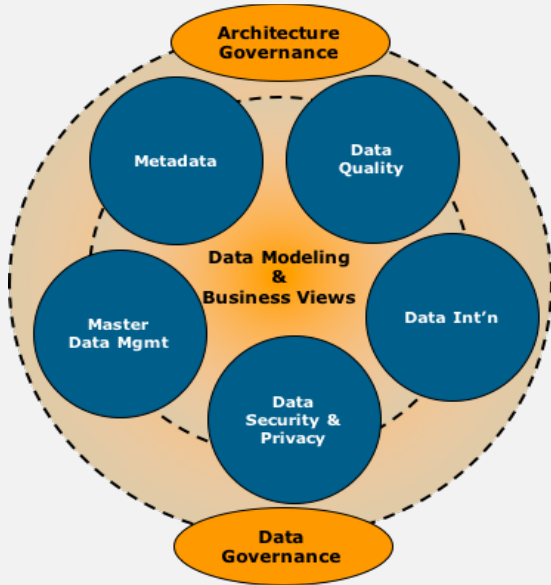
#2

Be willing to experiment constantly and adopt a “fail-fast-to-succeed-sooner” mentality



#3

Go beyond traditional Enterprise Information Management (EIM) and embrace social data curation*



* At least some of the time, for some data and Analytics and their interpretation.



Afterword

Get in the game!

1

2

3

*Digitization, in short, is **not a great equalizer** that drives all companies toward similar processes and outcomes. Instead, it's driving the leaders and laggards further apart.*

Andrew McAfee & Erik Brynjolfsson

What *business* problems could you solve with the “new” data and Analytics?

Clickstream / Path Analytics

Who navigates to the website, what do they do in each session – and then afterwards within other channels?

Social / Sentiment Analysis

What are customers saying about our products and services on social media sites?

Search / Text Analytics

What are customers really looking for online?

Sales / Graph Analytics

How are customers related? How can we make smarter recommendations?

Process / Path Analytics

What's the optimal process for collection of Supplier Funding?

TERADATA®

