



How Machine Learning and Big Data Shaped the Modern Organization

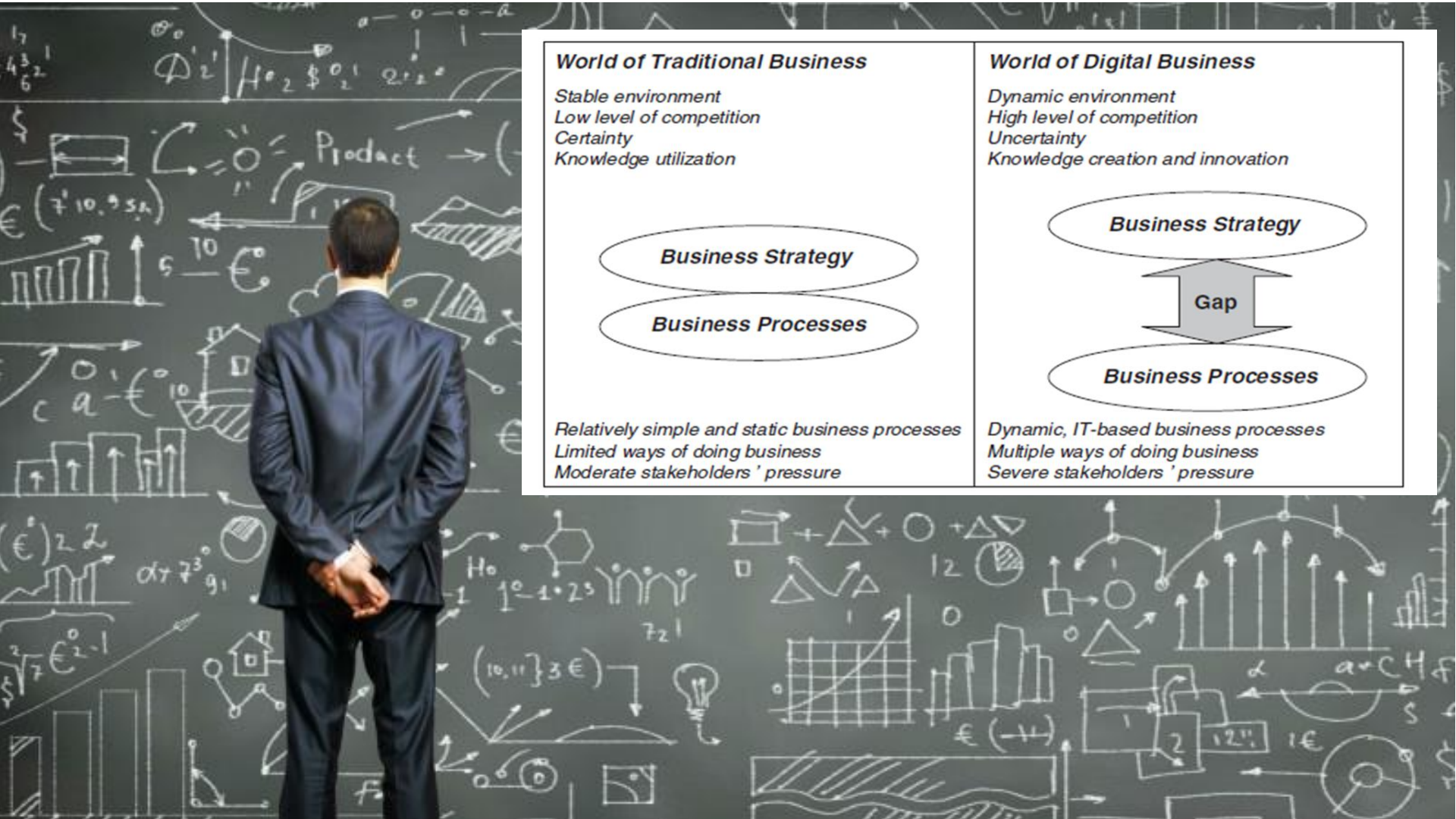
By :

Mehdi Habibzadeh (@NimaHM1980)

Hamid Bekamiri (@HamidBekamiri)

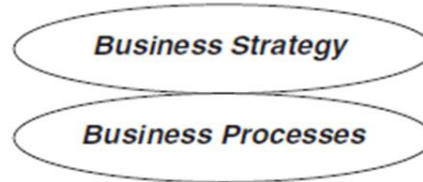
July 2017

How to Make the Right Decision?



World of Traditional Business

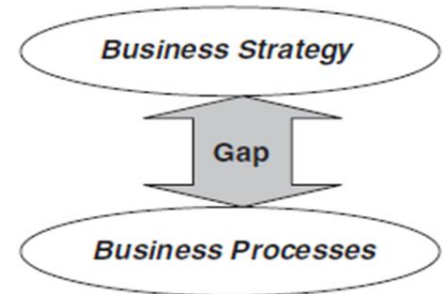
*Stable environment
Low level of competition
Certainty
Knowledge utilization*



*Relatively simple and static business processes
Limited ways of doing business
Moderate stakeholders' pressure*

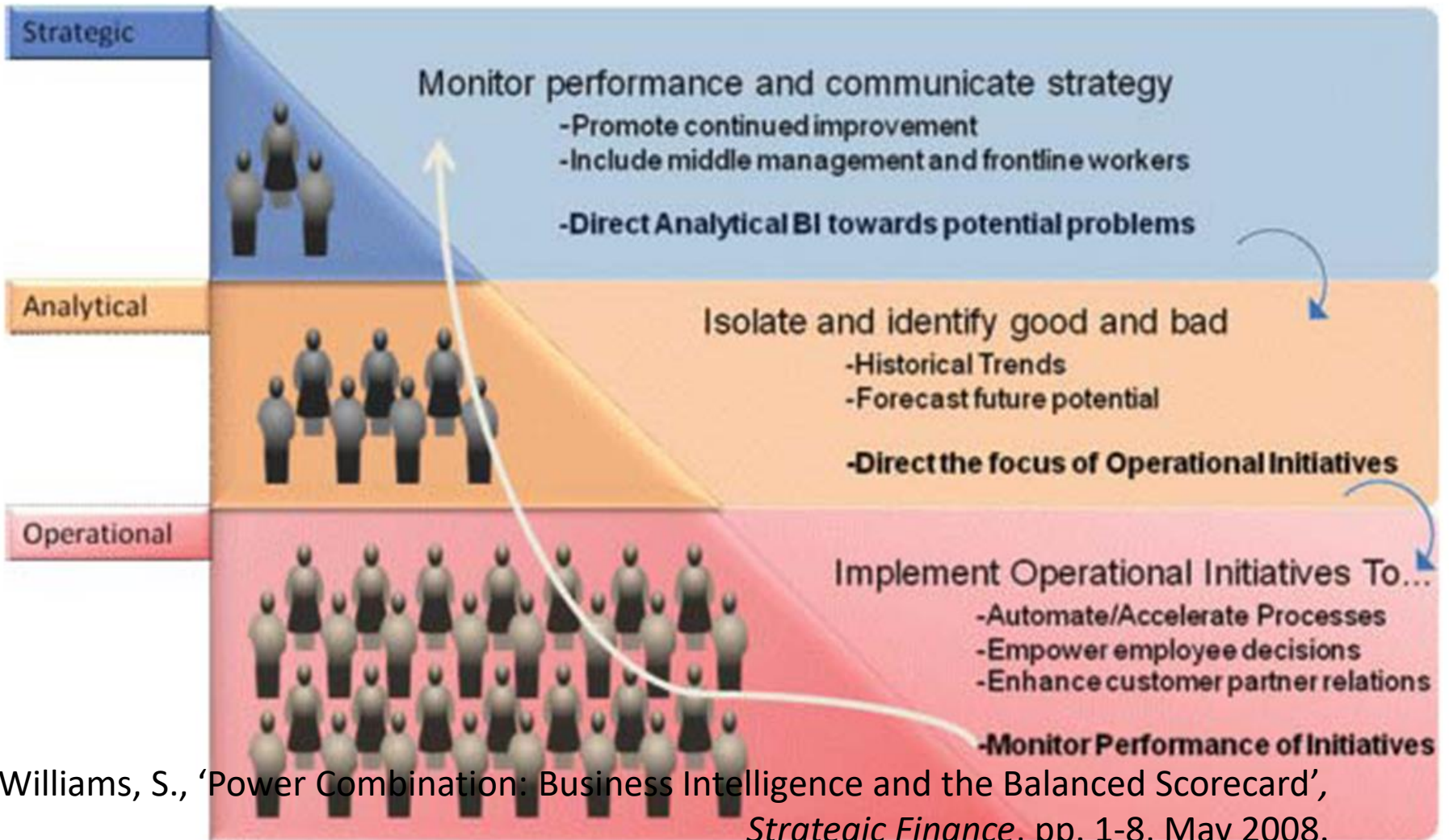
World of Digital Business

*Dynamic environment
High level of competition
Uncertainty
Knowledge creation and innovation*



*Dynamic, IT-based business processes
Multiple ways of doing business
Severe stakeholders' pressure*

What is the problem?



Williams, S., 'Power Combination: Business Intelligence and the Balanced Scorecard',
Strategic Finance, pp. 1-8, May 2008.



What season is it?



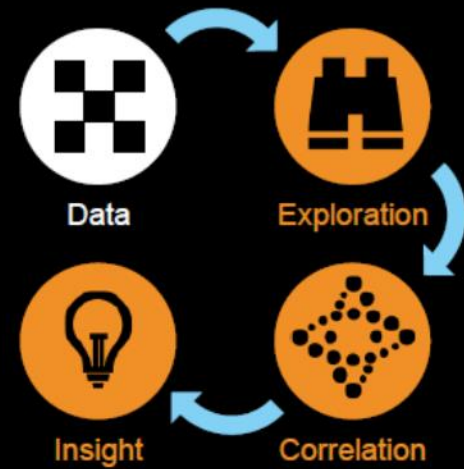
This way Data Mining helps organizations.

TRADITIONAL APPROACH



Start with hypothesis and test against selected data

BIG DATA APPROACH



Explore *all* data and identify correlations

So we know ...

Data will talk to
you if you're
willing to listen.

• JIM BERGESON •

Company marketing plans are not successful!!!!



امتیاز	کل سود
۱۲۷	۸۵.۵۵۹.۶۸۳
۹۷	۷۹.۲۴۳.۶۶۴
۱۰۰	۷۰.۷۵۳.۴۰۵
۱۰۷	۴۱.۸۴۶.۵۸۱
۶۹	۳۷.۵۴۷.۹۵۸
۷۳	۳۷.۰۰۹.۰۲۳
۱۲۷	۳۳.۷۷۵.۹۳۱
۱۶۹	۲۳.۴۹۰.۰۸۱۵
۳۹	۲۳.۴۶۸.۴۶۹
۱۵۱	۲۰.۵۳۰.۳۴۷
۱۰۷	۱۸.۲۴۹.۸۹۳
۱۳۰	۱۷.۵۴۱.۸۲۳
۵۵	۱۶.۱۲۲.۰۵۴
۱۰۱	۱۵.۸۴۵.۳۹۸
۱۰۷	۱۵.۳۰۱.۷۸۸
۷۹	۱۳.۹۳۹.۵۲۵
۱۲۷	۱۲.۳۵۳.۵۹۲
۹۷	۱۲.۲۲۹.۷۱۸
۱۷۷	۱۱.۱۹۴.۲۰۹
۱۰۸	۱۱.۰۶۵.۴۳۳
۲۶	۱۰.۹۰۸.۴۲۸
۲۰۲	۱۰.۶۳۳.۷۳۶
۷۷	۱۰.۱۱۹.۵۹۳
۸۷	۱۰.۰۱۰.۴۱۳۷
۱۱۷	۱۰.۰۱۰.۴۰۶۸

Mistake #

1

Selling to the

WRONG CUSTOMER



امتیاز	کل سود
۳۳۷	۵۴۶.۸۲۷
۲۸۹	۳۵۱.۴۹۰
۲۸۱	۴۴۴.۲۷۰
۲۷۹	۲.۷۶۵.۸۸۱
۲۶۹	۵۳۱.۷۴۳
۲۶۳	۱۷۲.۴۵۲
۲۵۹	۱۶۴.۵۹۵
۲۵۸	۲.۷۷۷.۱۵۷
۲۵۸	۴۵۸.۶۱۲
۲۵۸	۵۴۴.۶۰۲
۲۵۵	۳۰۶.۴۶۱
۲۴۸	۶۳.۶۶۷
۲۴۶	۵۷۵.۱۶۴
۲۴۲	۴۰۲.۱۸۸
۲۴۱	۲۰۳.۴۴۳
۲۳۷	۱۰.۷۶.۹۷۶
۲۳۴	۲۹۹.۶۹۴
۲۲۹	۵۱۲.۱۰۴
۲۲۷	۲۸۷.۶۵۲
۲۲۶	۹۵۸.۹۱۸
۲۲۶	۴۴۲.۶۷۲
۲۲۴	۹۴۵.۶۷۹
۲۲۰	۵۸۲.۷۲۷
۲۱۹	۷۹.۴۵۶
۲۱۷	۱.۵۸۲.۲۹۰
۲۱۱	۸.۵۸۷.۱۰۹
۲۱۰	۲۲۳.۸۰۳
۲۰۸	۱۸۵.۳۴۸
۲۰۷	۱۷۶.۱۵۴
۲۰۳	۱۹۴.۳۸۰
۲۰۲	۱۰.۶۳۳.۷۳۶
۲۰۲	۱۰.۶۱۹.۰۳

What is problem?



The problem is we don't understand
the problem.

— *Paul MacCready* —



DATA

is the new **OIL!**

* Name started by Orin Hanby AMA

Corporate investment in artificial intelligence is predicted to triple in 2017, becoming a \$100 billion market by 2025.

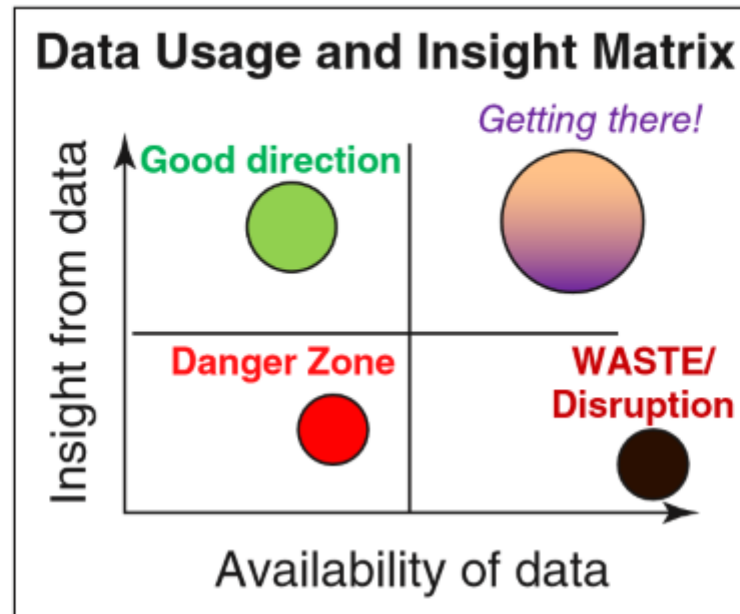


Figure 5.3 Data Usage and Insight Matrix

Availability is no guarantee of insight. The goal is to have insight available at the time of need, which means getting data accessible.

SOURCE: John F. Tanner Jr., "Building an Effective Data Strategy in Today's Information Overload World" (2014), white paper available at Teradata (www.teradata.com).

Data Accessibility Across the Organization



61%

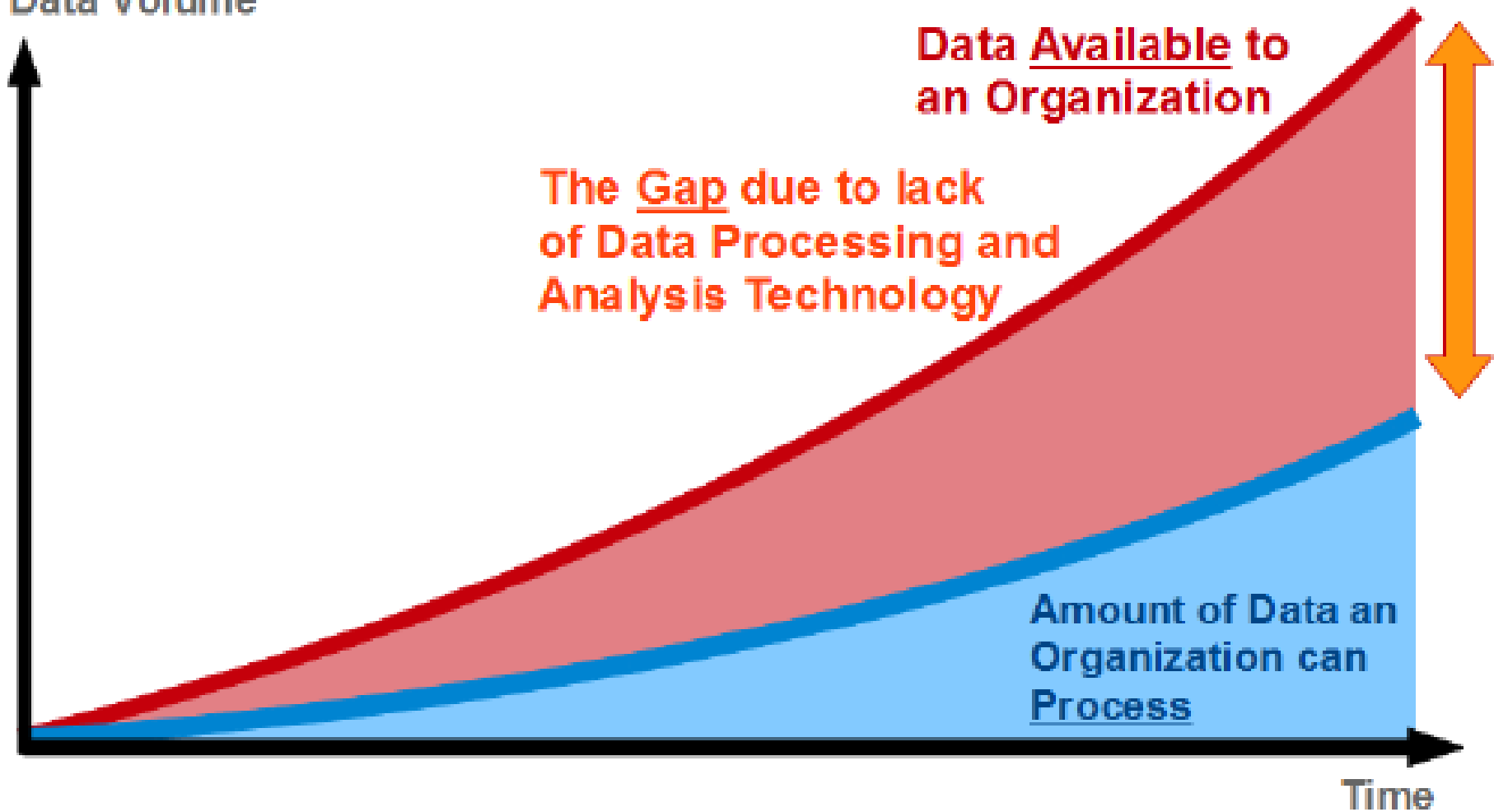
Data flow into enterprise software and are accessible to decision-makers

Data exist within departments and are not yet integrated across the organization **38%**

12% Outsourced to a third party



Data Volume



Data Available to an Organization

The Gap due to lack of Data Processing and Analysis Technology

Amount of Data an Organization can Process

Time

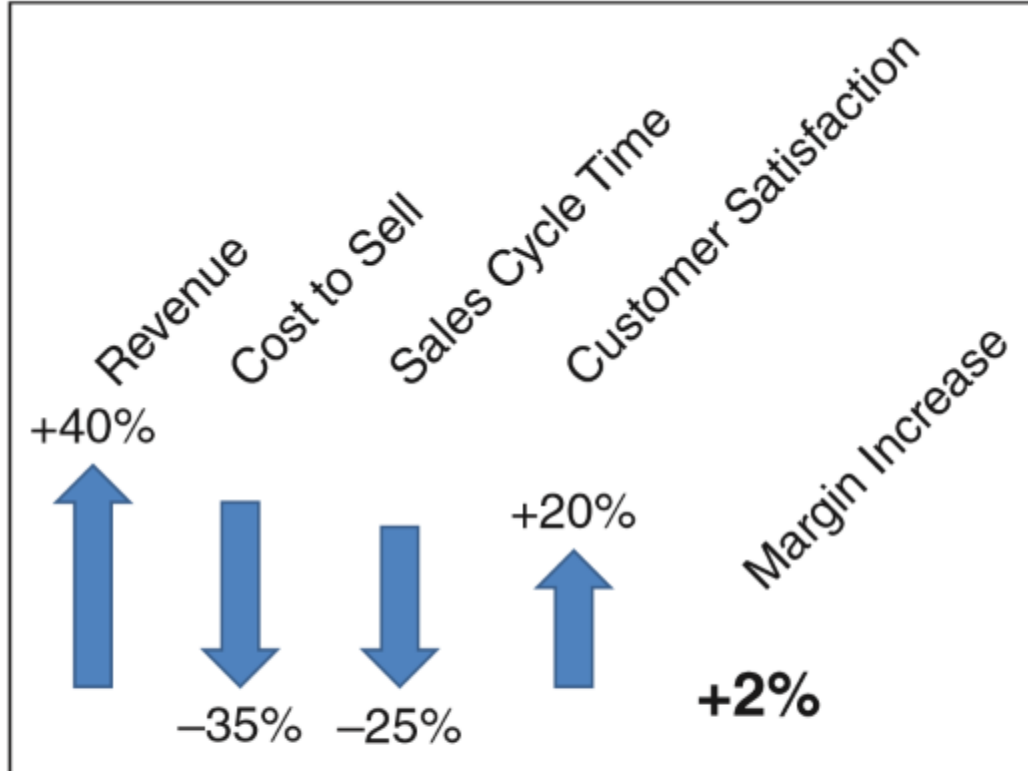
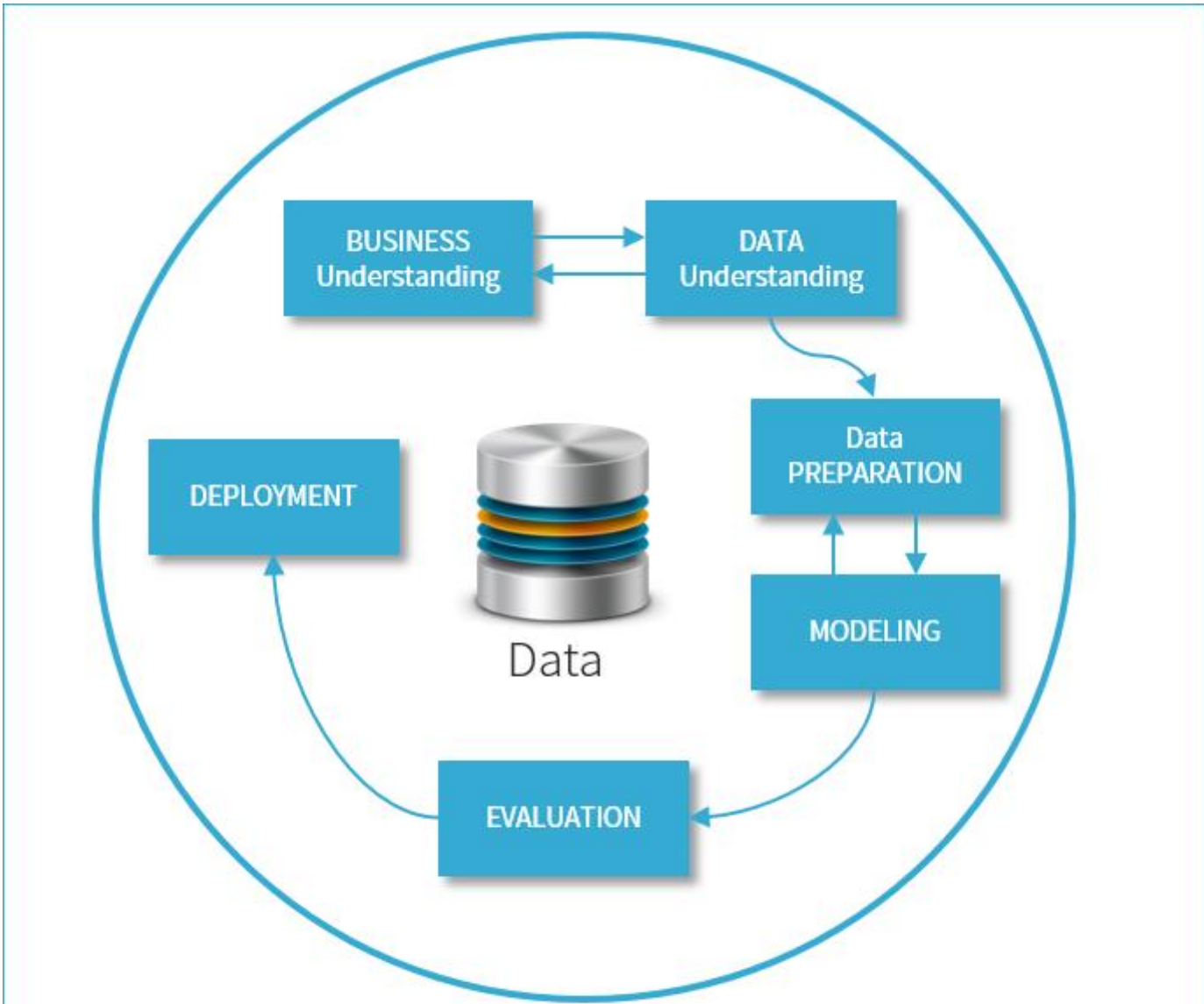


Figure 1.1 Maximum Gains Observed for Customer-Facing Big Data Solution

SOURCE: Data cited from CBP Research, *The Case for a New CRM Solution* (2013).

What Do You See?

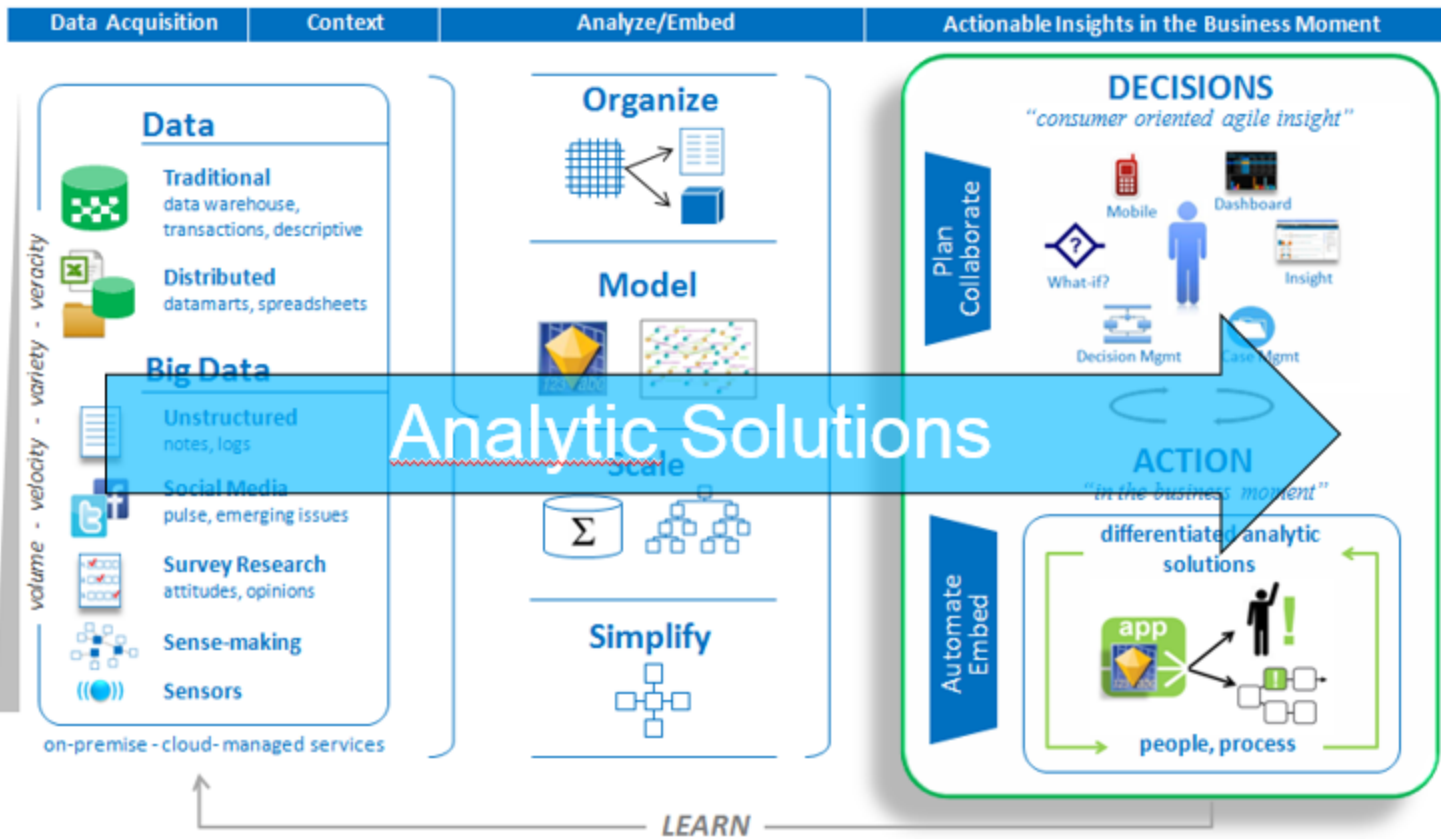


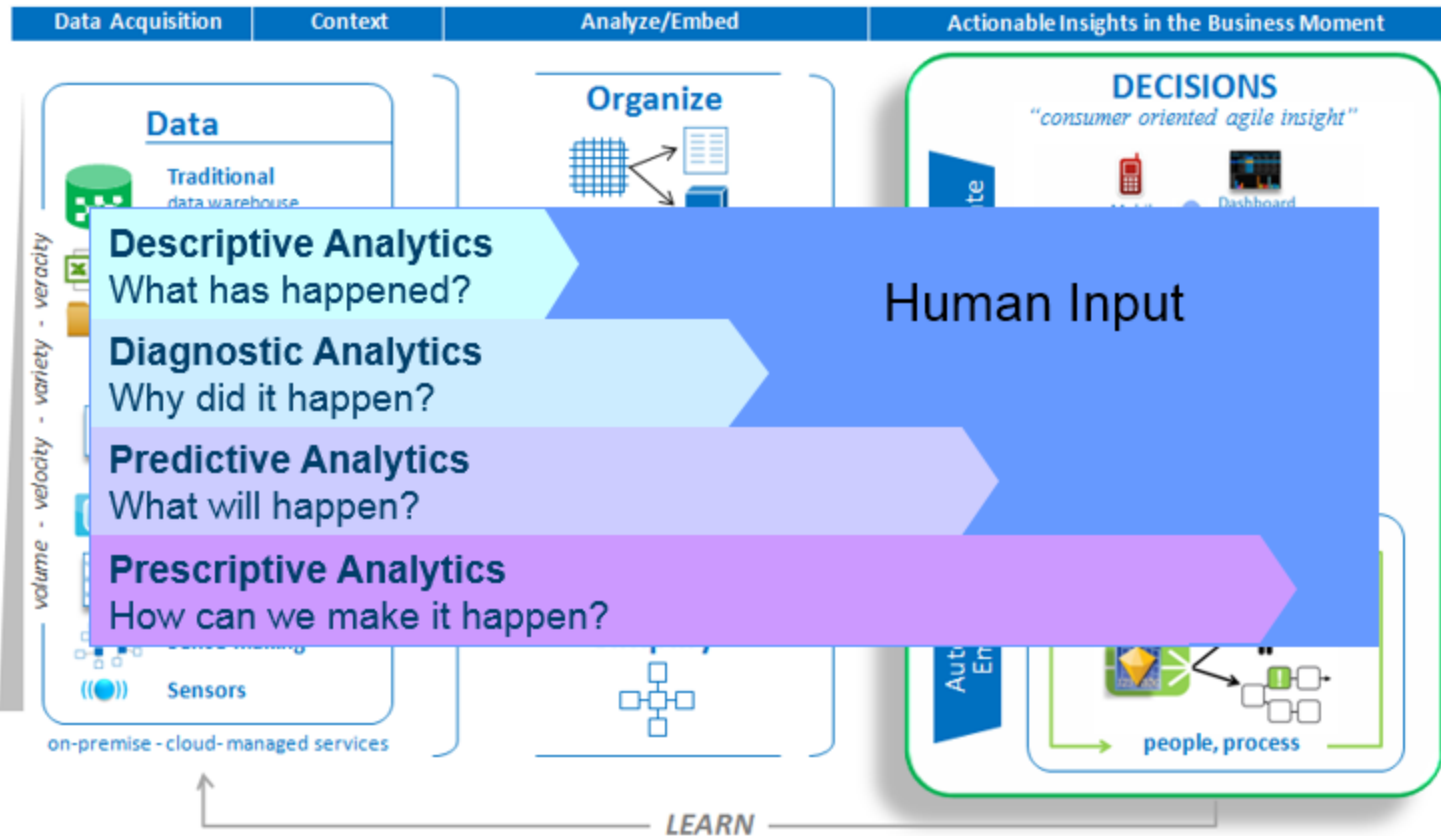


CRISP-DM Methodology









https://www.ibm.com/developerworks/community/blogs/jfp/entry/Cognitive_Computing_vs_Analytics?lang=en

predicting churn

churn prediction is a common application where the number of churners is typically small compared to the number of customers that stay



RFM METRICS

RFM Metrics:



REGENCY

The *freshness* of customer activity.

e.g. time since last activity



FREQUENCY

The *frequency* of customer transactions.

e.g. the total number of recorded transactions



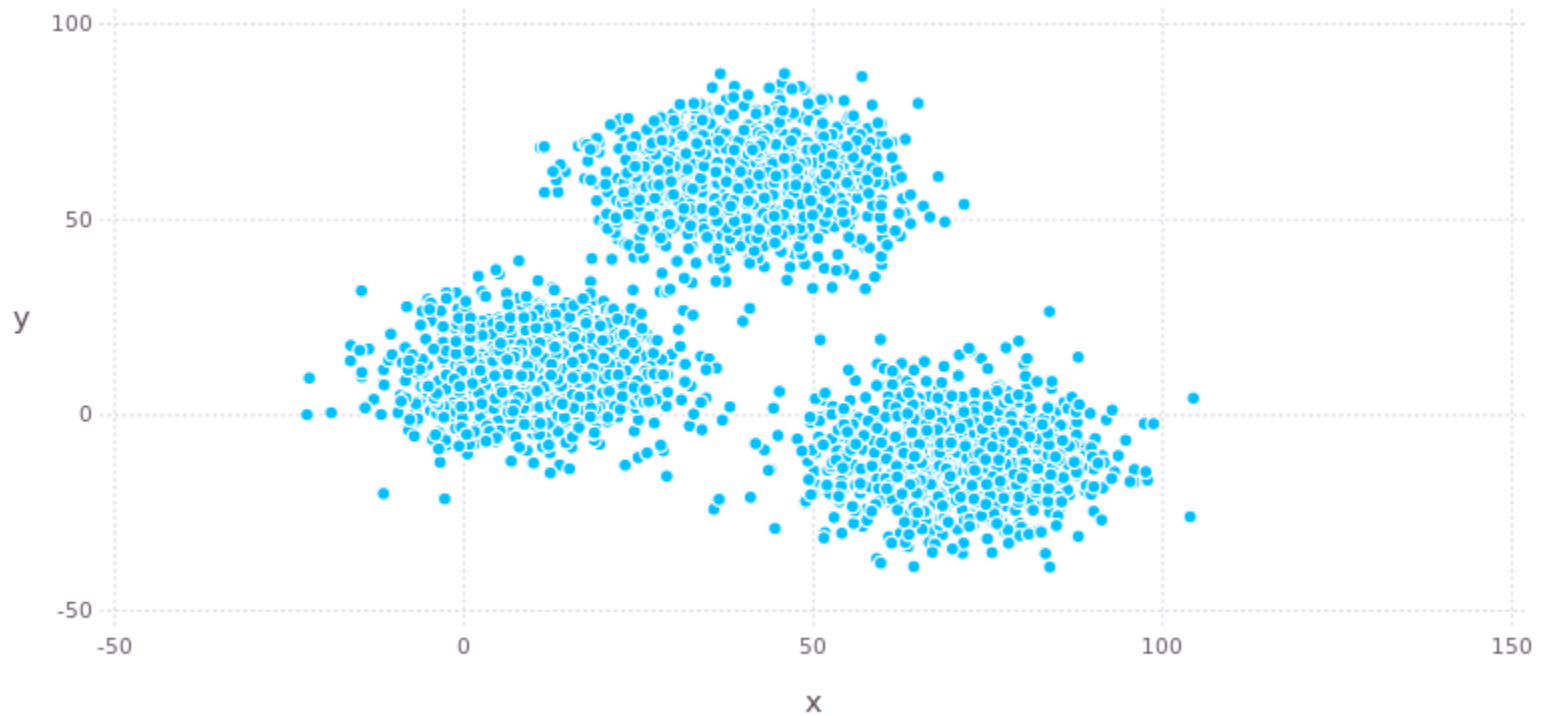
MONETARY

The *willingness* to spend.

e.g. the total transaction value

Levels Calculated for the Analysis Period

					Definitions
Recency	544 days	408 days	272 days	136 days	The number of days since the last order.
Frequency	1 Order(s)	2 Order(s)	5 Order(s)	6 Order(s)	The number of orders during the analysis period: 18 month(s)
Monetary	£ 52	£ 381	£ 1,367	£ 2,025	The total amount spent during the analysis period: 18 month(s).



جدول ۴- وضعیت متغیرهای RFM در هر خوشه

شماره خوشه	تعداد مشتری	تازگی خرید	تکرار خرید	مبلغ خرید	نام خوشه
خوشه ۱	۸۰۴	۰,۶۰۵	۳,۲۳۳	۰,۳۲۴	مشتریان رویگردان
خوشه ۲	۹۳۴۴	۰,۰۴۳	۰,۵۲۲	۰,۰۱	مشتریان خدماتی
خوشه ۳	۱۱۸۴۵	۱,۲۰۲	۱,۳۰۴	۰,۰۴۲	مشتریان ارزشمند
خوشه ۴	۱۴۷۸	۱,۲۵۲	۴,۴۰۴	۰,۵۶۴	مشتریان طلایی

$$DI_m = \min_{1 \leq i \leq m} \left\{ \min_{1 \leq j \leq m, j \neq i} \left\{ \frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq m} \Delta_k} \right\} \right\} \quad (4)$$

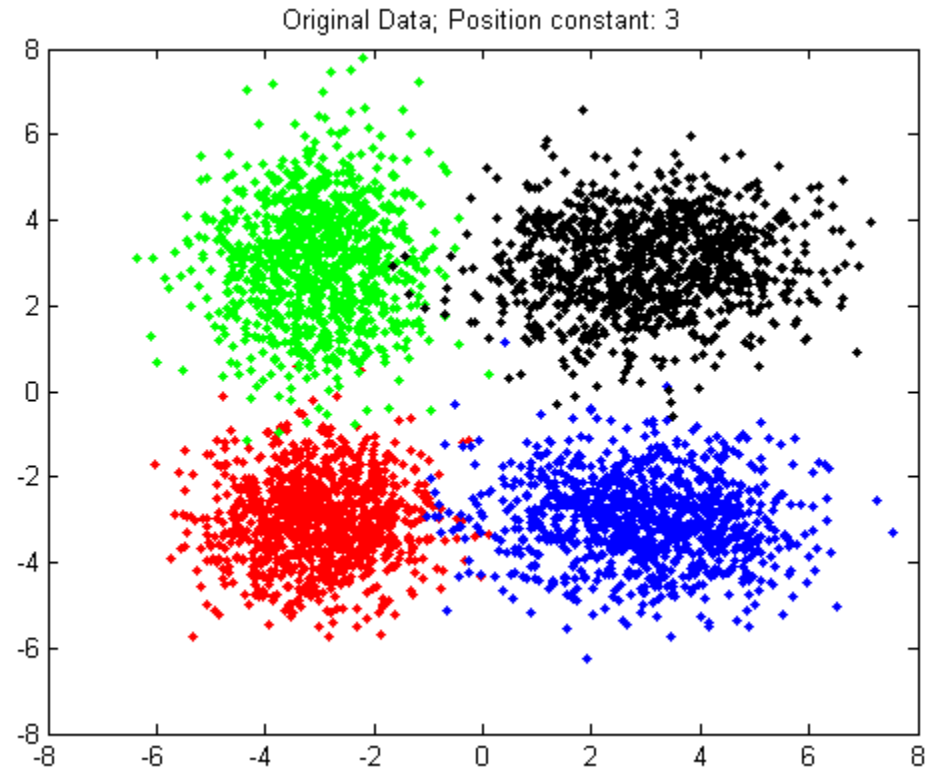
$$\delta(c_i, c_j) = \min_{x \in c_i, y \in c_j} \{d(x, y)\} \quad (5)$$

$$\Delta(c_i) = \max_{x, y \in c_i} \{d(x, y)\} \quad (6)$$

با توجه به شاخص دون، در وضعیتی که شاخص در بالاترین امتیاز می باشد تعداد خوشه‌ها متناسب با داده‌های تحقیق به عنوان بهترین حالت خوشه بندی انتخاب می شود. بر اساس محاسبات صورت گرفته در جدول شماره ۱ خوشه بندی با ۴ گروه بهترین حالت دسته‌بندی مشتریان می باشد.

جدول ۳- شاخص دون برای تعیین تعداد بهترین خوشه

تعداد خوشه	DI_m
۳	۰,۵۳۹۲
۴	۰,۵۷۳۵
۵	۰,۳۱۶۳



جدول ۵- وضعیت متغیرهای فرعی در هر خوشه

+

خوشه ۴	خوشه ۳	خوشه ۲	خوشه ۱		
بالا	بالا	کم	کم	تازگی خرید	متغیرهای اصلی
بالا	بالا	کم	بالا	تکرار خرید	
بالا	متوسط	کم	متوسط	مبلغ خرید	
مشتریان طلایی	مشتریان ارزشمند	مشتریان خدماتی	مشتریان رویگردان	نام خوشه	متغیرهای فرعی
٪۶	٪۵۱	٪۴۰	٪۳	تعداد مشتریان	
٪۴۹	٪۲۴	٪۷	٪۲۰	مبلغ مطالبات	
۹۴۱	۵۶	۲۵	۴۳۱	میانگین سپرده یکساله	
۷۱۹	۳۹	۱۸	۲۳۱	میانگین سپرده ۶ ماه	

جدول ۷- ارائه استراتژی‌های خوشه یک



خوشه اول - مشتریان رویگردان	
توضیحات	نوع تحلیل
این گروه از مشتریان در ماههای اخیر تعاملات مالی خود را با بانک کمتر کرده اند و خدمات الکترونیکی و شعبه ای کمتری از بانک دریافت نموده اند و تبع آن سپرده‌های خود را نزد بانک کاهش داده اند و خروج این سپرده‌ها برای بانک بسیار هزینه بر می باشد.	متغیرهای اصلی
در بررسی متغیرهای فرعی میانگین شش ماهه و یکساله مشتری به خوبی مشخص میشود که این مشتریان سپرده‌های خود را نزد بانک کاهش داده اند نکته حائز اهمیت در این خصوص مبالغ سپرده‌گذاری توسط این مشتریان می باشد که ارقامی قابل توجهی را در بر میگیرد.	متغیرهای فرعی
یکی از مهمترین دغدغه‌های کارشناسان بانکی، شناسایی مشتریانی رویگردان سپرده‌های گذاری بالا می باشد. از طرفی این نکته بسیار حائز اهمیت می باشد این گروه از مشتریان حدود ۲۰ درصد از مطالبات را به خود اختصاص داده اند از این رو ضروری است تا با توجه به اینکه فقط تعداد ۳ درصد از مشتریان را تشکیل می دهند راهکارهای به منظور جذب دوباره این مشتریان در سطح عملیاتی ارائه شود.	کارشناسان بانکی

Target Marketing



6. Evolution and Development of Knowledge Management:

KM 1: Collecting



KM 2: Sharing



KM 3: Using



KM 1.0	KM 2.0	KM 3.0
techno-centric	people-centric	productivity-centric
command and control	Social	Practical & Individual
"KM is extra work"	"KM is part of my work"	"KM is helping me do my work"

KM 1.0 focused on **collecting** knowledge "before it walked out the door".

KM 2.0 focused on **sharing** knowledge using web-enabled and social media tools.

KM 3.0 focus on **using** existing knowledge to help people get their job done.

*“Vision without action is a daydream.
Action without vision is a **nightmare.**”*

