



# Customer Analytics

By Olga Buchonina  
March 14, 2016  
[olga@prostolabs.com](mailto:olga@prostolabs.com)



- Pi Day is celebrated on March 14th (3/14) around the world. Pi (Greek letter “ $\pi$ ”) is the symbol used in mathematics to represent a constant — the ratio of the circumference of a circle to its diameter — which is approximately 3.14159.
- Pi has been calculated to over one trillion digits beyond its decimal point. As an irrational and transcendental number, it will continue infinitely without repetition or pattern. While only a handful of digits are needed for typical calculations, Pi’s infinite nature makes it a fun challenge to memorize, and to computationally calculate more and more digits.

# Wiki Definition

- o **Customer analytics** is a process by which data from customer behavior is used to help make key business decisions via market segmentation and predictive analytics. This information is used by businesses for direct marketing, site selection, and customer relationship management. Marketing provides services in order to satisfy customers. With that in mind, the productive system is considered from its beginning at the production level, to the end of the cycle at the consumer. Customer analytics plays a very important role in the prediction of customer behavior today



# Big Data Landscape 2016 (Version 2.0)

## Infrastructure

**Hadoop On-Premise**  
cloudera Hortonworks MAPR Pivotal IBM InfoSphere bluedata jethro

**Hadoop in the Cloud**  
amazon web services Microsoft Azure IBM InfoSphere CAZEN TREASURE DATA altiscale Du bale

**Spark**  
databricks GridGain TACHYON NEXUS

**Cluster Services**  
amazon web services kubernetes docker HPCC SYSTEMS MESOSPHERE Core OS pepperdata StackIQ

## Analytics

**Analyst Platforms**  
Palantir AYASDI Quid enigma Digital Reasoning ORBITAL INSIGHT

**Analytics Platforms**  
Microsoft guavus Datameer Bottlenose interana

**Data Science Platforms**  
context relevant DataRobot CONTINUUM Alpine MODE plotly ADATAQ dataiku Roambi DOMINO sense yhat ALGORITHMIA

**Visualization**  
tableau Google Cloud Platform Qlik Q looker SUSENSE CHARTIO datarman

## Applications

**Sales & Marketing**  
RADIUS Gainsight bloomreach Zeta EVERSTRING blueyonder Lattice @kahuna infer SAILTHRU persado AVISO sense QUANTIFIND ACTIONIQ fusemachines JENAGGIO

**Customer Service**  
MEDALLIA ATTENTIFY CLARABRIDGE CLICKFOX STELLA Service NGDATA Preact Digital Genius appuri wiseio

**Human Capital**  
gild Connectifier textico entelo hiIQ

**Legal**  
RAVEL JUDICATA Everlaw eBrevia PREMONITION

**NoSQL Databases**  
amazon DynamoDB Google Cloud Platform Microsoft Azure ORACLE mongoDB MarkLogic DATASIX Couchbase SequoiaDB redislabs influxdata

**NewsSQL Databases**  
SAP HANA Clustrix Pivotal paradig4 nuodb splice machine MariaDB citusdata deepdb Trafalgar Cockroach LABS

**BI Platforms**  
Power BI amazon web services Domo Wave Analytics birst GoodData Kyrus Insights platform atscale ARKADIA

**Statistical Computing**  
sas SPSS MATLAB

**Log Analytics**  
splunk sumologic kibana CLOUD PHYSICS loggly

**Social Analytics**  
Hootsuite NETBASE DATASIFT track bitly synthio simplereach

**Ad Optimization**  
AppNexus criteo MediaMath OpenX rocketfuel Ad Science theTradeDesk Algorithms dsterly DataXu Oppier TAPAD

**Security**  
CYCLANCE CounterTack cybereason ThreatMetrix AREA1 SentinelOne Recorded Future Guardian Analytics FORTSCALE siftscience Keybase feedzai SIGNIFYD

**Vertical AI Applications**  
facebook Clara KASIST lumiata

**Graph Databases**  
neo4j OrientDB InfiniteGraph

**MPP Databases**  
TERADATA VERTICA Netezza Action kognitio iSQL dremio

**Cloud EDW**  
amazon web services Google Cloud Platform Microsoft Azure Pivotal snowflake WATERLINE DATA Infoworks

**Data Transformation**  
alteryx talend TRIFACTA tamr Palata Alation

**Data Integration**  
informatica Put potential to work: MuleSoft snapLogic BedrockData xplenty

**Real-Time**  
amazon web services METAMARKETS stream confluent DATATONIGHT dataArtisans

**Machine Learning**  
Azure Machine Learning H2O.ai Amazon SageMaker SKYTREE rapidminer DATAFORM VIZENEE PredictionIO glowfish

**Speech & NLP**  
NarrativeScience ARRIA NUANCE semantic machines Cortana. VIV nervana Vocabular nora Sennheiser clarifai

**Horizontal AI**  
IBM Watson Cortana. VIV nervana Vocabular nora Sennheiser clarifai

**Publisher Tools**  
Outbrain Taboola quantcast Chartbeat yieldbot Yieldmo

**Govt / Regulation**  
Socrata OPENGOV FN FiscalNote enigma mark43 PREDPOL OpenDataSoft

**Finance**  
affirm LendingClub OnDeck Kreditech LendUp Kabbage tidemark INSIGHT uora Dataminr Lenddo KENSHO AIDYA ISENTIUM Quantipian

**Management / Monitoring**  
New Relic APPDYNAMICS amazon web services actifio Numerify splunk DATADOG Trocena Driven Anodot

**Security**  
TANUUM illumio CODE42 DataGravity CipherCloud VECTRA sqrrl BlueTalon

**Storage**  
amazon web services Google Cloud Platform Microsoft Azure panasas nimblestorage COHO Qumulo

**App Dev**  
apigee CRK ASKIO Typesafe DRIVEN

**Crowd-sourcing**  
amazon mechanical turk CrowdPower WorkFusion

**Search**  
hp ELASTICSEARCH ORACLE ENDICE Lucidworks elastic ThoughtSpot MAANA swifttype Algolia SINEQUA

**Data Services**  
UO OPERA Mta Sigma DATE SCIENCE ELISION VALLEY DATA SCIENCE kaggle datacscope DataKind

**For Business**  
OrigamiLogic ClearStory CIRRO import.io

**Web / Mobile / Commerce**  
Google Analytics mixpanel R.J.Metrics BLUECORE AMPLITUDE granify sumall Airtale retention custora

**Education / Learning**  
KNEWTON Clever Cleclara PANORAMA knowite

**Life Sciences**  
23andMe Counsyl Recombine KYRUUS FLATIRON zymergen HealthTap METABIOTA ZEPHYR HEALTH ovia GINGER.io transcripTIC Glow enlithic AICure Atomwise

**Industries**  
OPower eHarmony RetailNext STITCH FIX WorkFusion BLUE@RIVER TACHYUS Seeq FarmLogs SwiftKey select NIGHT MACHINE HowGood statmuse B@XEVER

## Cross-Infrastructure/Analytics

amazon web services Google Microsoft IBM SAP SAS 1010 data hp Autonomy VERTICA vmware TIBCO TERADATA ORACLE NetApp

## Open Source

**Framework**  
hadoop HADOOP YARN Spark MESOS TEZ Flink CDAP

**Query / Data Flow**  
SLAMDATA APACHE DRILL Google Cloud Dataflow

**Data Access**  
cassandra HBASE mongoDB CouchDB riak OPENTSDB kafka nifi

**Coordination**  
talend Apache Zookeeper Apache Ambari

**Real-Time**  
STORM Spark APEX Flink TACHYON druid

**Stat Tools**  
R Scala NumPy SciPy

**Machine Learning**  
mllib Apache SINGA MAUIB Aerolve Caffe FeatureFu WEKA DIMSUM jupyter DL4J

**Search**  
elasticsearch Solr

**Security**  
Apache Ranger Visualization Zepplin

## Data Sources & APIs

**Health**  
JAWBONE GARMIN practicefusion fitbit Withings VALIDIC netatmo kinsa Human API

**IOT**  
UPTAKE ThingWorx helium samsara

**Financial & Economic Data**  
Bloomberg THOMSON REUTERS DOW JONES YODLEE PREMISE S&P CAPITAL IQ quandl xignite CB INSIGHTS mattermark estimate PLAID

**Air / Space / Sea**  
PLANET LABS spire WINDWARD CRUISE SKYCATCH Airware DroneDeploy

**Location / People / Entities**  
axiom Experian EPSILON GARMIN foursquare InsideView esri STREETLINE CARTOBS factual PlaceIQ Crism Hexagon placemeter BASIS Sense

**Other**  
qualtrics panjiva DATA.GOV

**Incubators & Schools**  
DataCamp INSIGHT DataElite METIS The Data Incubator

Last Updated 2/12/2016

© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)

FIRSTMARK

# Customer Analytics Models

	Descriptive	Predictive	Prescriptive
	What <b>HAS</b> happened?	What <b>COULD</b> happen?	What <b>SHOULD</b> happen?
What the user needs to <b>DO</b>	<ul style="list-style-type: none"> <li>• <b>Increase</b> asset reliability</li> <li>• <b>Reduce</b> labor and inventory costs</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Predict</b> infrastructure failures</li> <li>• <b>Forecast</b> facilities space demands</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Increase</b> asset utilization</li> <li>• <b>Optimize</b> resource schedules</li> </ul>
What the user needs to <b>KNOW</b>	<ul style="list-style-type: none"> <li>• The <b>number and types</b> of asset failures</li> <li>• Why <b>maintenance costs</b> are high</li> <li>• The value of the <b>materials inventory</b></li> </ul>	<ul style="list-style-type: none"> <li>• How to <b>anticipate failures</b> for specific asset types</li> <li>• When to <b>consolidate underutilized</b> facilities</li> <li>• How to <b>determine costs</b> to improve service levels</li> </ul>	<ul style="list-style-type: none"> <li>• How to <b>increase</b> asset production</li> <li>• Where to <b>optimally route</b> service technicians</li> <li>• Which strategic facilities plan provides the <b>highest long-term utilization</b></li> </ul>
How analytics gets <b>ANSWERS</b>	<ul style="list-style-type: none"> <li>• <b>Standard reporting</b> - What happened?</li> <li>• <b>Query/drill down</b> - Where exactly is the problem?</li> <li>• <b>Ad hoc reporting</b> - How many, how often, where?</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Predictive modeling</b> - What will happen next?</li> <li>• <b>Forecasting</b> - What if these trends continue?</li> <li>• <b>Simulation</b> - What could happen?</li> <li>• <b>Alerts</b> - What actions are needed?</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Optimization</b> - What is the best possible outcome?</li> <li>• <b>Random variable optimization</b> - What is the best outcome given the variability in specified areas?</li> </ul>
What makes this analysis <b>POSSIBLE</b>	<ul style="list-style-type: none"> <li>• Alerts, reports, dashboards, <b>business intelligence</b></li> </ul>	<ul style="list-style-type: none"> <li>• Predictive <b>models</b>, forecasts, statistical <b>analysis</b>, scoring</li> </ul>	<ul style="list-style-type: none"> <li>• Business rules, organization <b>models</b>, comparisons, <b>optimization</b></li> </ul>
Business value →			





View of the Customer Experience  
Analytics today





■ ■ ■ ■ ■

[https://www.youtube.com/watch?v=6M0TgEUbRBC&index=2&list=PLEPPAIRmgCSj34hDP7Jt4PGK\\_2p\\_i1m4G](https://www.youtube.com/watch?v=6M0TgEUbRBC&index=2&list=PLEPPAIRmgCSj34hDP7Jt4PGK_2p_i1m4G)





# Business Organization



"Our philosophy is delivering happiness to customers and employees. People may not remember exactly what you did or what you said, but they always remember how you made them feel".

- Tony Hsieh -CEO at Zappos.



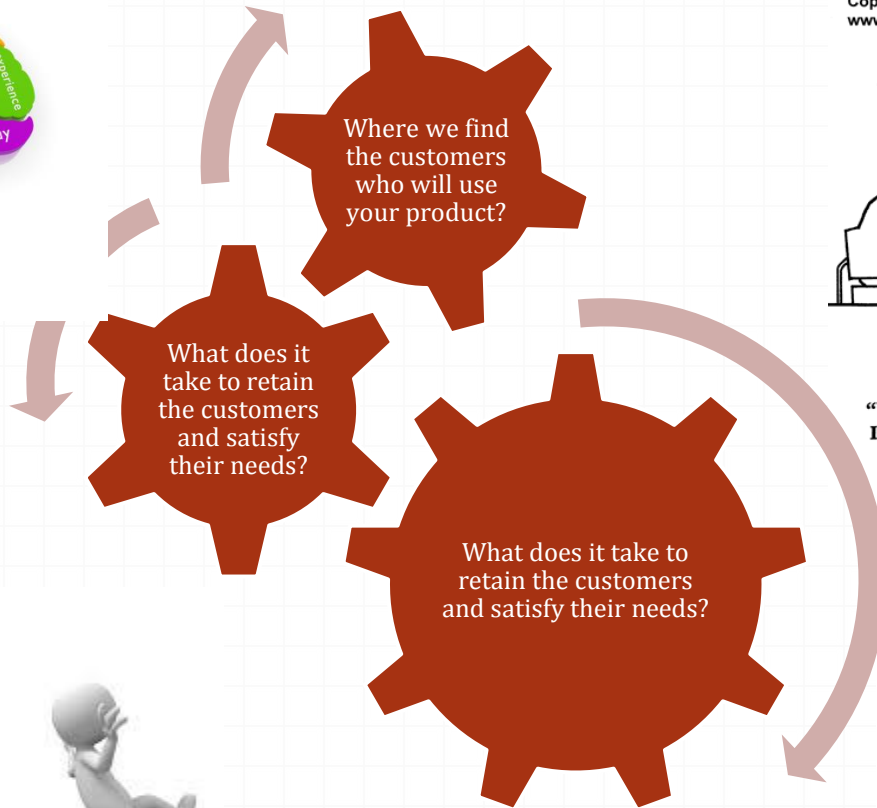


# Focus: What is the problem?

Copyright 2003 by Randy Glasbergen.  
www.glasbergen.com



**"You say this job takes you to Hell and back everyday  
I hear that Hell is inhabited by millions of lost souls.  
How many have you signed up as new customers?!"**



# How you can do it?



- o As unsexy and low-tech as it may sound, the telephone is one of the best branding devices out there. (Tony Hsieh, CEO of Zappo's)
- o By Sir Richard Branson:
  - o 1. Be visible
  - o 2. Express passionate commitment to serving the customer.
  - o 3. Your company's employees are its greatest assets
  - o 4. Hire people who have the "Virgin" attitude
  - o 5. Empower your employees to solve problems and to make every experience great.
  - o 6. Engage in social media with a genuine voice.
  - o 7. Have fun.

# Why do we like things Free?



Gary and Robert Intero  
Recognized as the Top 1%  
BRE# 00859303 & 01272788  
518 N. Santa Cruz Ave  
Los Gatos, CA 95030



<http://www.sunnyvalerealestate.com>

<https://www.cardcash.com>



# Customer Analytics Models

	Descriptive	Predictive	Prescriptive
	What <b>HAS</b> happened?	What <b>COULD</b> happen?	What <b>SHOULD</b> happen?
What the user needs to <b>DO</b>	<ul style="list-style-type: none"> <li>• <b>Increase</b> asset reliability</li> <li>• <b>Reduce</b> labor and inventory costs</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Predict</b> infrastructure failures</li> <li>• <b>Forecast</b> facilities space demands</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Increase</b> asset utilization</li> <li>• <b>Optimize</b> resource schedules</li> </ul>
What the user needs to <b>KNOW</b>	<ul style="list-style-type: none"> <li>• The <b>number and types</b> of asset failures</li> <li>• Why <b>maintenance costs</b> are high</li> <li>• The value of the <b>materials inventory</b></li> </ul>	<ul style="list-style-type: none"> <li>• How to <b>anticipate failures</b> for specific asset types</li> <li>• When to <b>consolidate underutilized</b> facilities</li> <li>• How to <b>determine costs</b> to improve service levels</li> </ul>	<ul style="list-style-type: none"> <li>• How to <b>increase</b> asset production</li> <li>• Where to <b>optimally route</b> service technicians</li> <li>• Which strategic facilities plan provides the <b>highest long-term utilization</b></li> </ul>
How analytics gets <b>ANSWERS</b>	<ul style="list-style-type: none"> <li>• <b>Standard reporting</b> - What happened?</li> <li>• <b>Query/drill down</b> - Where exactly is the problem?</li> <li>• <b>Ad hoc reporting</b> - How many, how often, where?</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Predictive modeling</b> - What will happen next?</li> <li>• <b>Forecasting</b> - What if these trends continue?</li> <li>• <b>Simulation</b> - What could happen?</li> <li>• <b>Alerts</b> - What actions are needed?</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Optimization</b> - What is the best possible outcome?</li> <li>• <b>Random variable optimization</b> - What is the best outcome given the variability in specified areas?</li> </ul>
What makes this analysis <b>POSSIBLE</b>	<ul style="list-style-type: none"> <li>• Alerts, reports, dashboards, <b>business intelligence</b></li> </ul>	<ul style="list-style-type: none"> <li>• Predictive <b>models</b>, forecasts, statistical <b>analysis</b>, scoring</li> </ul>	<ul style="list-style-type: none"> <li>• Business rules, organization <b>models</b>, comparisons, <b>optimization</b></li> </ul>
Business value →			

# CLV (calculating lifetime value)

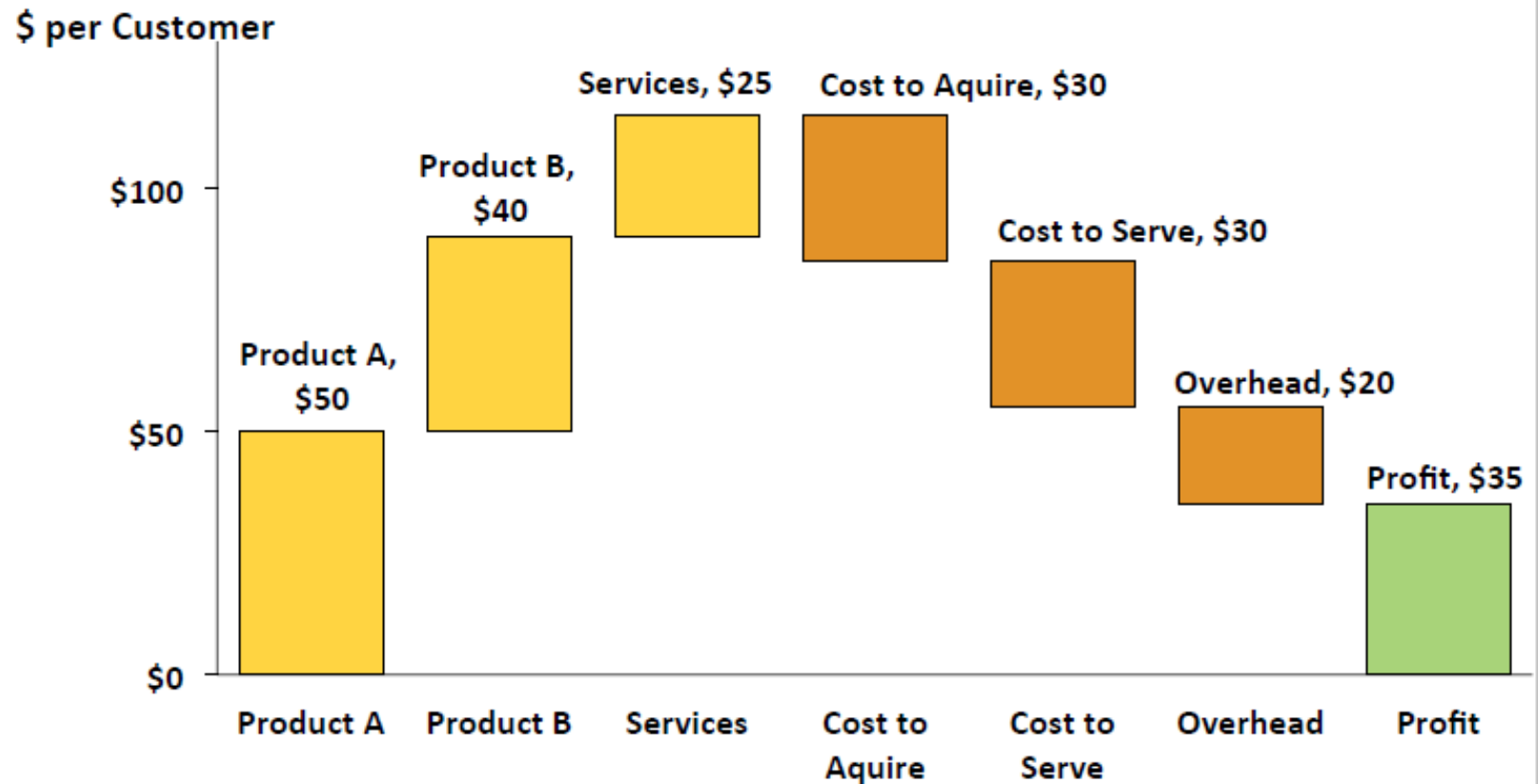
“There is only one valid definition of a business purpose: to create a customer”

*(Peter Drucker, The Practice of Management, 1954)*

- Revenue per purchase
- X Frequency of purchase
- Customer lifetime
- Profit margin
- CLV = the accumulated profit or loss from each customer over the course of that customer's relationship with you.

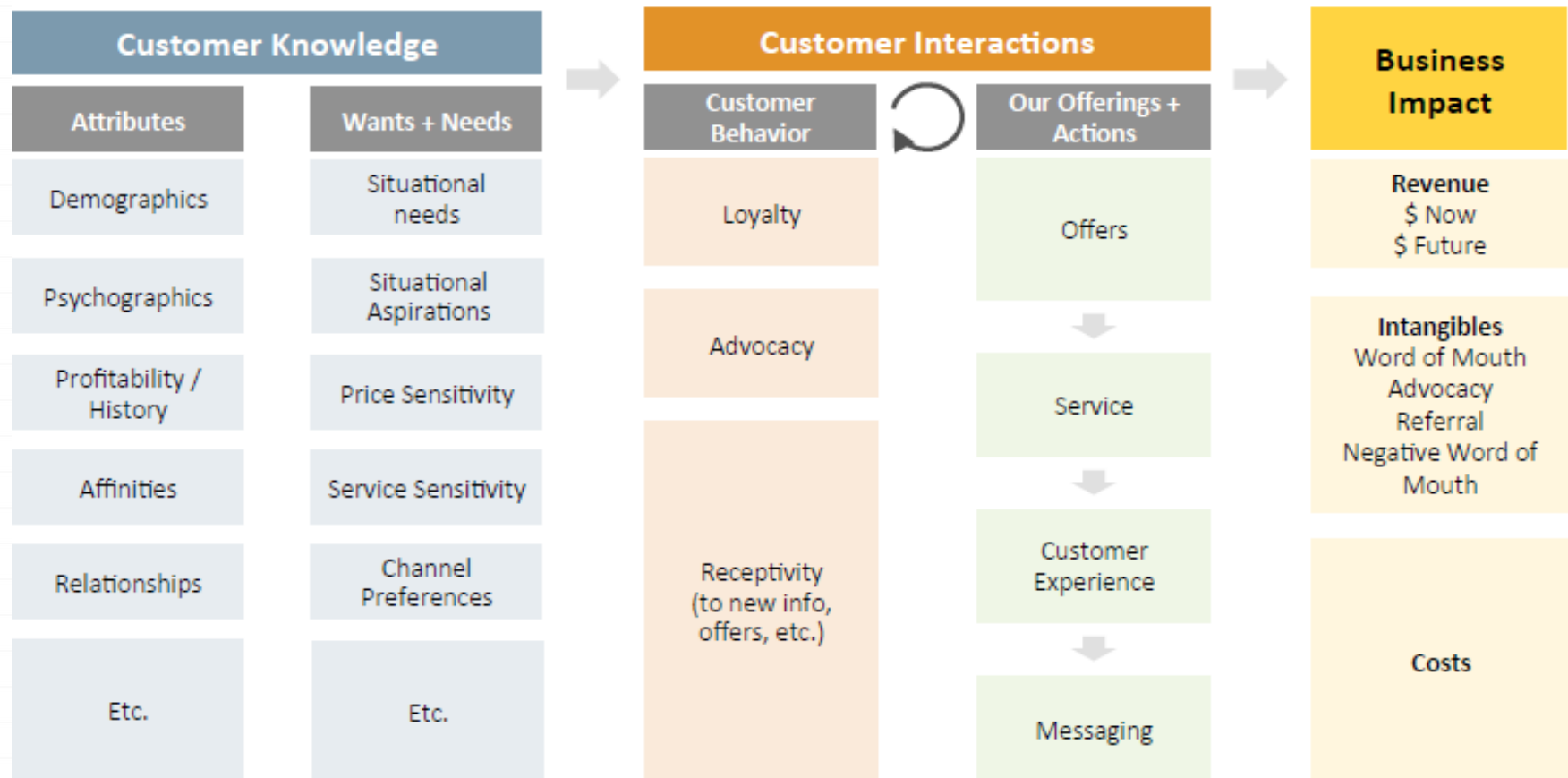
# Classic CPA Output: "Waterfall Chart"

## *Key components of profit and loss per customer*

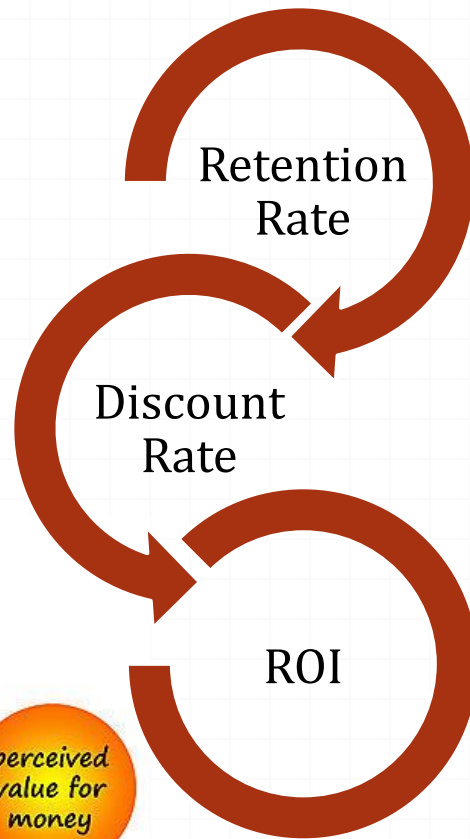




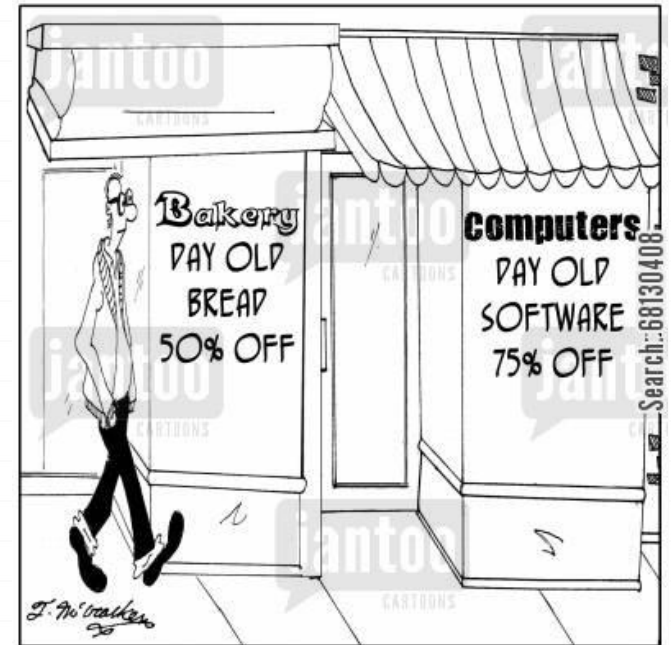
# Managing Customer Lifetime Value



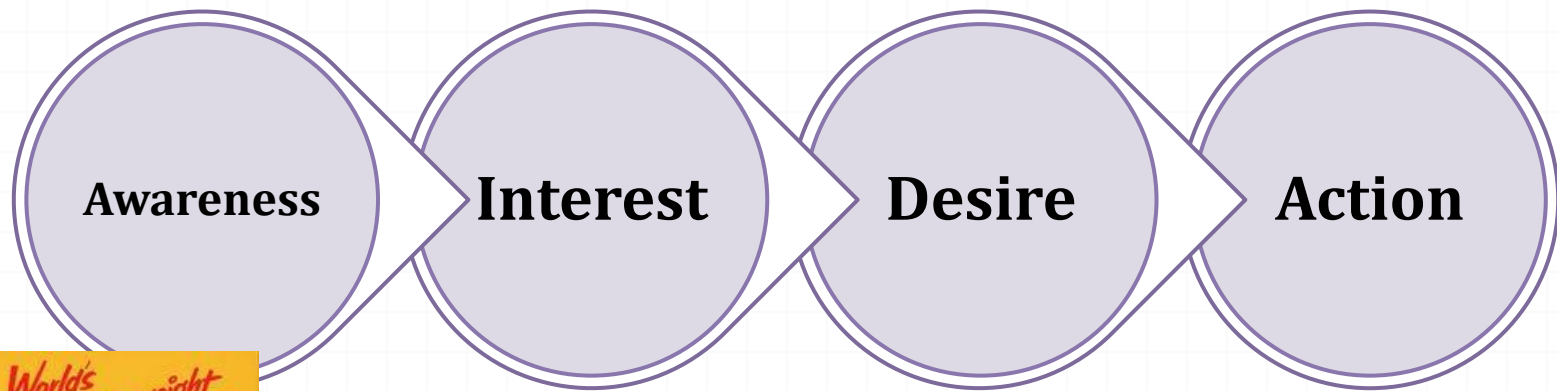
# CLV



"And this is where our ROI became an IOU."



# Customer Journey Analytics





# Customer Journey

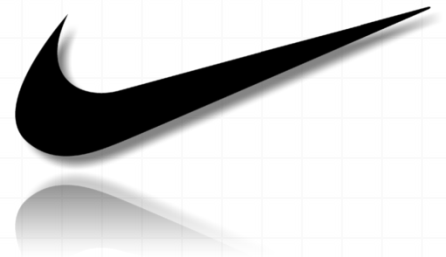
- Compare Odyssey by Homer Journey and Ernst Hemingway who was attributed with writing six-word story “ For Sale: baby shoes, never worn”.

# Follow me home

- o Awareness
- o Consideration
- o Preference
- o Action
- o Loyalty

# How often should a consumer brand change it's brand identity?

- As little as possible. Consumers are fickle and forget easily. Your most important goal in all of branding is awareness and recall, and that becomes hard to sustain when your identity is in flux.







# Brand Awareness

- ◊ List mobile phone manufacturers
- ◊ List three makers and models of family sedans
- ◊ Name Four rental companies

# Customer Satisfaction

- ◊ General satisfaction ( relational)
- ◊ Attribute Satisfaction ( transactional)
  - ◊ Theacsi.org – American Customer Satisfaction Index



Chart browser



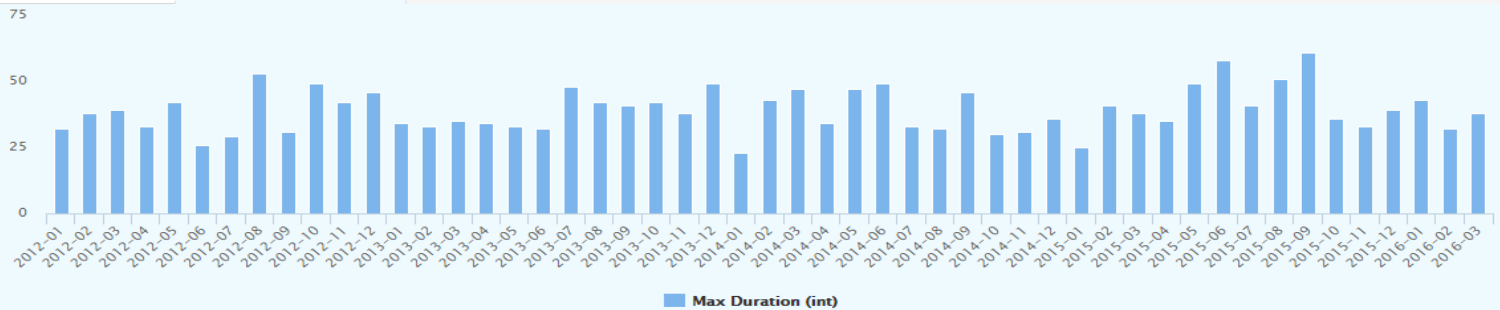
Data



Customize

CUSTOMER ...

EDIT WIDGET ✕



stingrayreports.com

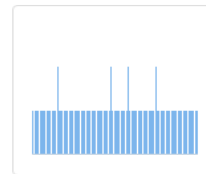
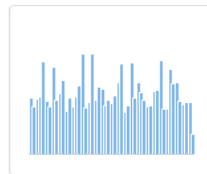
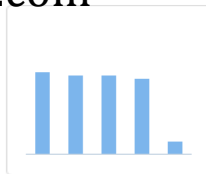


Chart browser



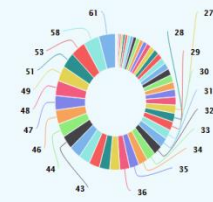
Data



Customize

CUSTOMER ...

EDIT WIDGET ✕



SELECT IMPORT

test123 (March 11, 2016 5:25 PM)

SELECT DIMENSION, AXIS X

Duration (int)

SELECT METRICS, AXIS Y

Avg

Duration (int)

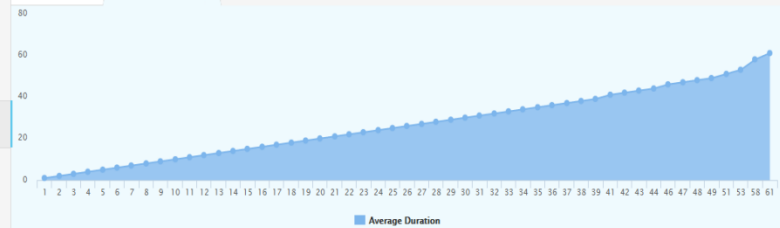
New report

+ Add new dimension

+ Add new metric

CUSTOMER ...

EDIT WIDGET ✕



CHANGE CHART TYPE ^

Area chart



# Customer Analytics Tool



Chart browser



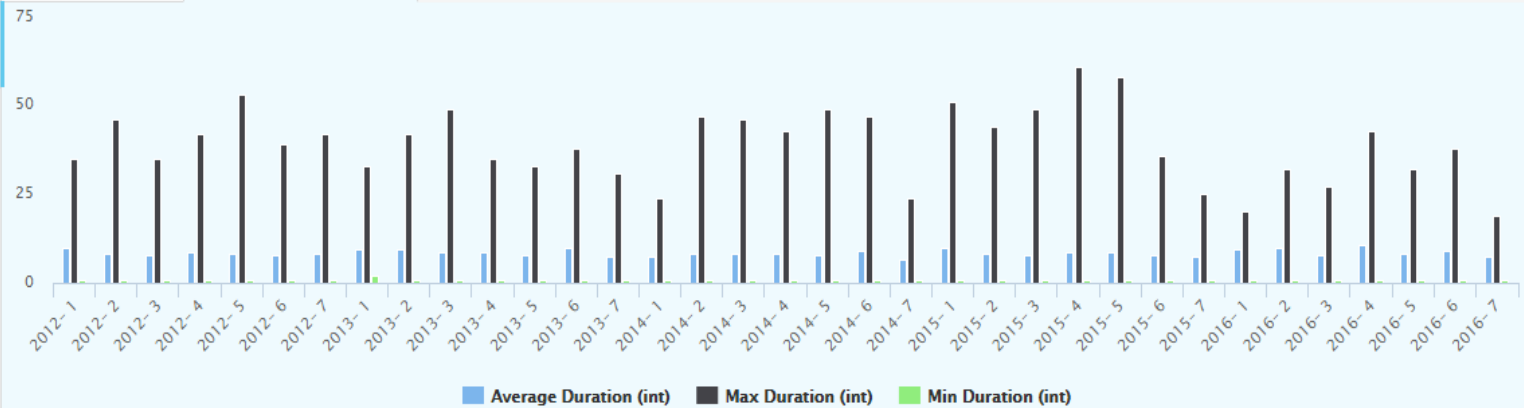
Data



Customize

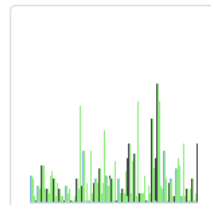
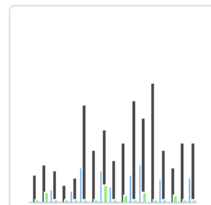
CUSTOMER ...

EDIT WIDGET ✕



Average Duration, Max Duration, Min Duration by CaseOpened (datetime):

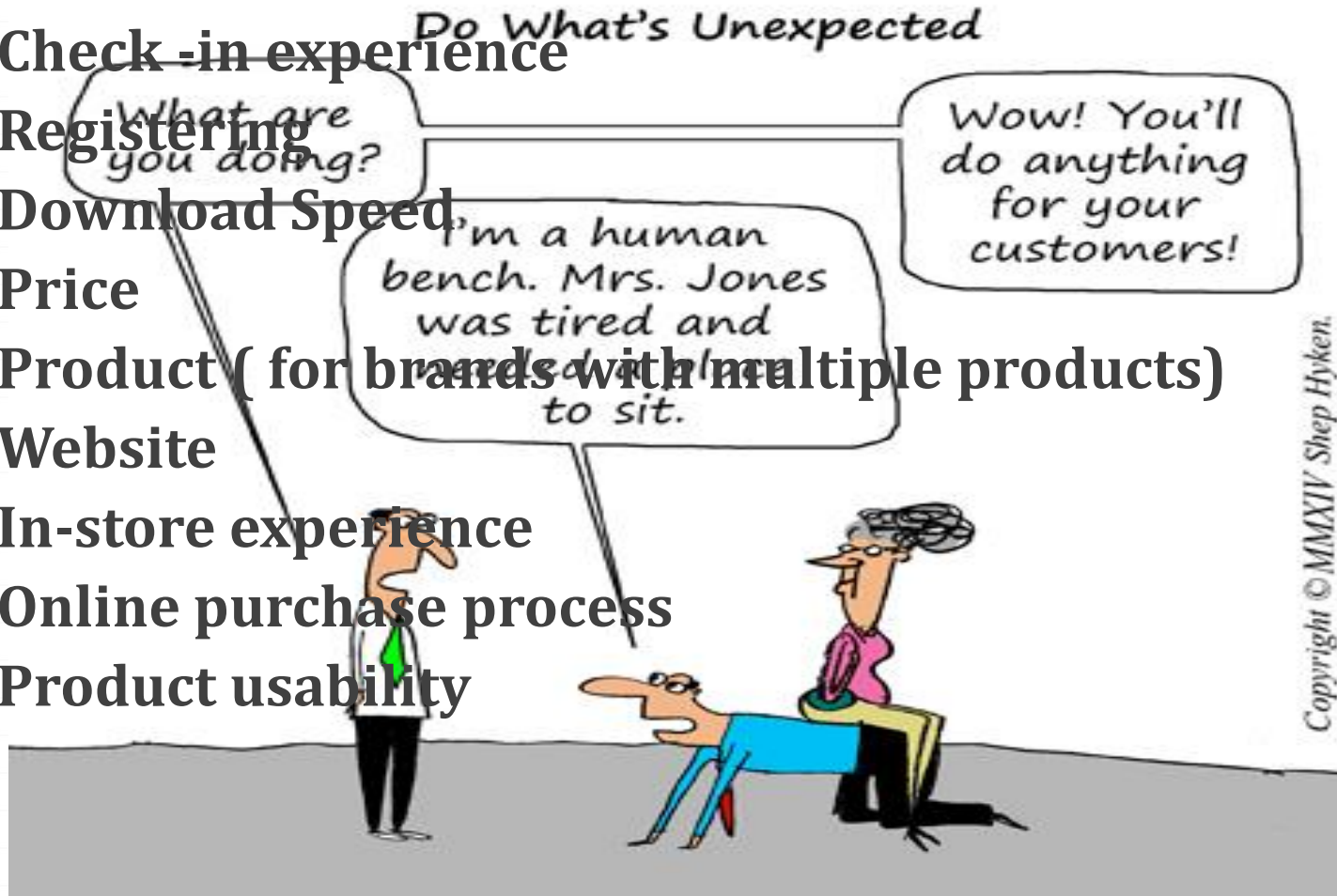
Set 2 of 8



- Predictive
- Descriptive
- Prescriptive
- Feedback

# What makes the product “the IT”?

- Check-in experience
- Registering
- Download Speed
- Price
- Product (for brands with multiple products)
- Website
- In-store experience
- Online purchase process
- Product usability



# Delighting Customers

# Wow!

- Measure attitude lift
- National and enterprise
- [www.measuringu.com](http://www.measuringu.com)



"Someone calling themselves a customer says they want something called service."

# Multiple Regression Analytics

- Independent and dependent variables
- Magic quadrant is not the “IT” but it helps to see where you are based on set of variables

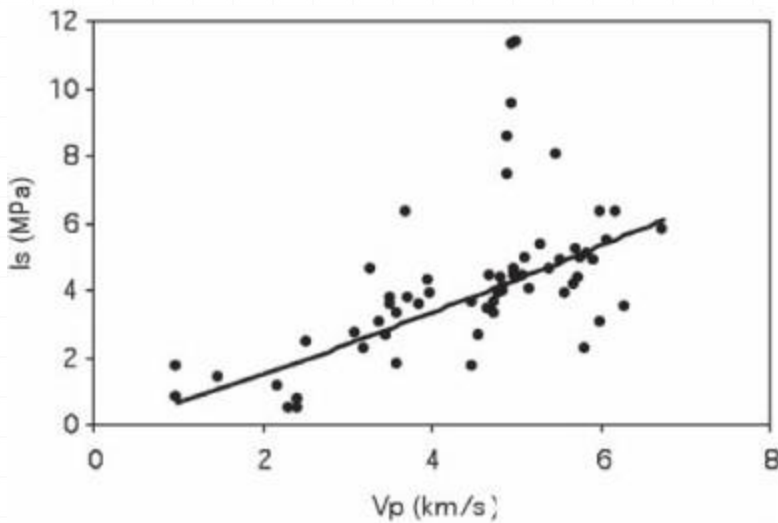
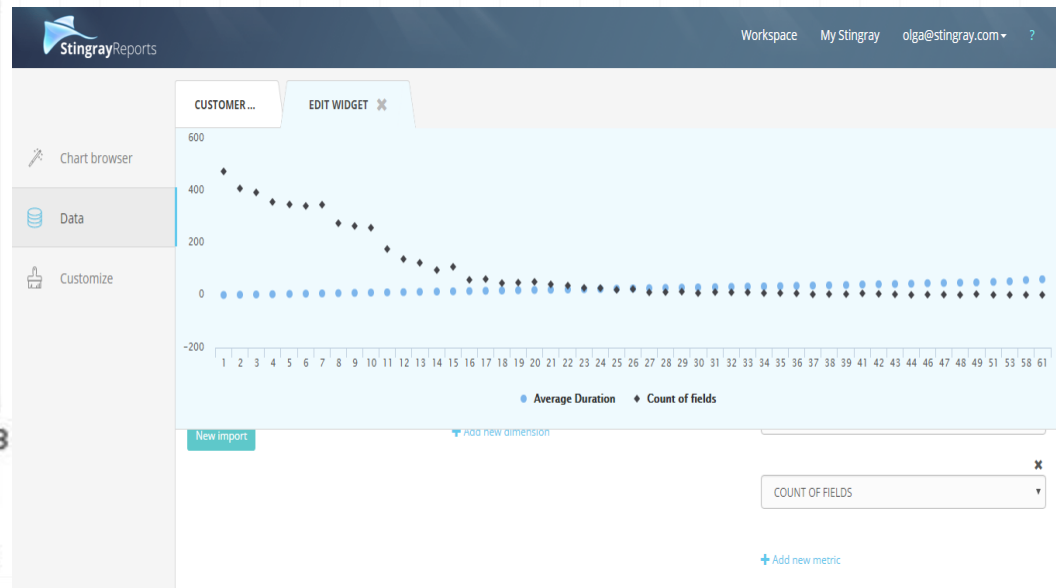
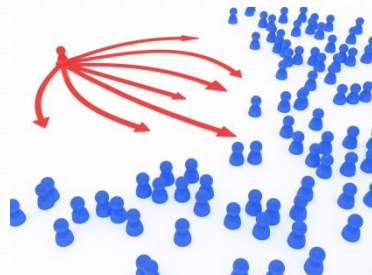


Figure 4—The relationship between the  $P$ -wave velocity and  $I_s$  of rocks





# Three R's



# Your Goal?



# Post-Sale Behavior

MERELY SATISFYING CUSTOMERS

WILL NOT BE ENOUGH TO EARN  
**THEIR LOYALTY.**

INSTEAD, THEY MUST EXPERIENCE



**EXCEPTIONAL  
SERVICE**



WORTHY OF THEIR REPEAT BUSINESS AND REFERRAL.

UNDERSTAND THE FACTORS THAT DRIVE  
**THIS CUSTOMER REVOLUTION.**

RICK



TATE

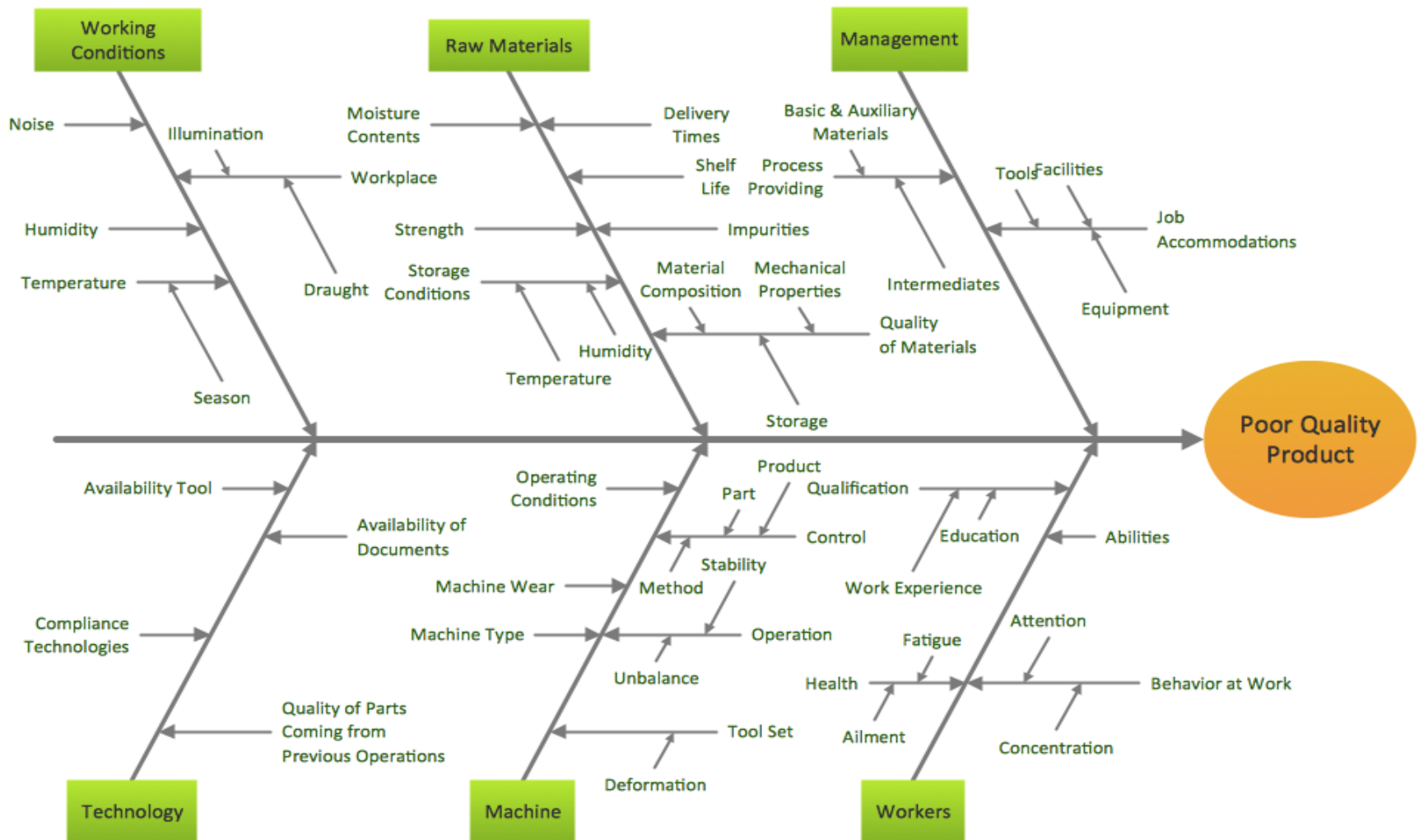
# Comcast transition experience

## o Story

- o Door to door sales and good sales representative provide good incentive to switch
  - o Gotcha's : hidden costs
- o Setup the installation time. Technician arrived and installed cable by routing it across the outside wall and roof: Result: unsightly cable in view and danger of cable being damaged. Plus installation did not work.
- o Outcome: 3 calls to customer support and unavailability to resolve the issue for a week ( lack of resources)
- o Reservations with switch: customer support were mitigated by assurance and number of publications about Comcast actions to resolve the support team issues.
- o Notes: call was transferred 4 times, on-hold for over one hour and switching between while trying to resolve the issue. Call directly to Sales Rep to resolve the issue since Support was unable to resolve it.
- o Observation: Bill time - 5 calls about the bill which was due and requests to pay immediately.
  
- o Moral of the story: Customer or Revenue
- o Customer notes: Looking for an alternative provider
- o Comcast investment: Sales representative investment, marketing, accounting. What is missing : Customer satisfaction. With post-sales experience -> customer retention, bad publicity and free marketing
  - o Compare with Zappos



# Cause and Effect diagram



# Emotions

- Recommendations for house repairs:
  - 5 contractors, willing to recommend only one



# View on Apache Spark and new trends in Data Analysis

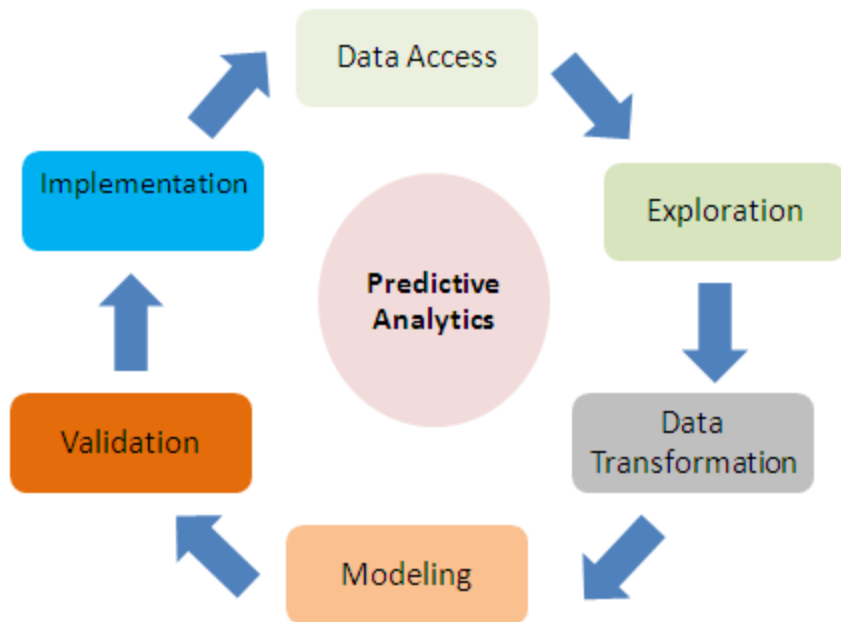
- o 2015 was without a doubt the year of Apache Spark, an open source framework leveraging in-memory processing, which was starting to get a lot of buzz when we published the previous version of our landscape. Since then, Spark has been embraced by a variety of players, from IBM to Cloudera, giving it considerable credibility. Spark is meaningful because it effectively addresses some of the key issues that were slowing down the adoption of Hadoop: it is much faster (benchmarks have shown Spark is 10 to 100 times faster than Hadoop's MapReduce), easier to program, and lends itself well to machine learning. (by Matt Turck VC at FirstMark)

[olga@prostolabs.com](mailto:olga@prostolabs.com)

<https://stingrayreports.com>



# Predictive Modeling



- 1. C4.5
- 2. k-means
- 3. Support vector machines
- 4. Apriori
- 5. EM
- 6. PageRank
- 7. AdaBoost
- 8. kNN
- 9. Naive Bayes
- 10. CAR

