Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 1
Customer Analytics Overview

Customer Analytics Overview

Consumers versus Customers

- A consumer is a person who consumes the goods, i.e. the user of the goods.
- A customer is a person who buys your product/service.
- However, the terms are virtually interchangeable, and are both frequently used to describe the purchaser of goods/services from a given company.
- Consumer: A person who identifies a need or desire, makes a purchase, and then disposes of the product.

What is Consumer Behavior?

- Consumer Behavior:
  - The study of the processes involved when individuals or groups select, purchase, use, or dispose of products, services ideas, or experiences to satisfy needs and desires

- Role Theory:
  - Identifies consumers as actors on the marketplace stage

- Consumer Behavior is a Process:
  - Exchange: A transaction in which two or more organizations give and receive something of value

Stages in the Consumption Process

	CONSUMER'S PERSPECTIVE	MARKETER'S PERSPECTIVE
PREPURCHASE ISSUES	How does a consumer decide that he/she needs a product? What are the best sources of information to learn more about alternative choices?	How are consumer attitudes toward products formed and/or changed? What cues do consumers use to infer which products are superior to others?
PURCHASE ISSUES	Is acquiring a product a stressful or pleasant experience? What does the purchase say about the consumer?	How do situational factors, such as time pressure or store displays, affect the consumer's purchase decision?
POSTPURCHASE ISSUES	Does the product provide pleasure or perform its intended function? How is the product eventually disposed of, and what are the environmental consequences of this act?	What determines whether a consumer will be satisfied with a product and whether he/she will buy it again? Does this person tell others about his/her experiences with the product and influence their purchase decisions?

[Activity] Consumption Process

- Ask your friend about his/her recent purchase
  - **Pre-purchase**: how did he/she decide that he/she needs a product? How did he/she find information and make a decision?
  - **Purchase**: how did he/she make a purchase? Off-line or on-line? How was the experience?
  - **Post-purchase**: how was the consumption process? Was it satisfied the need? Did he/she share the experience?

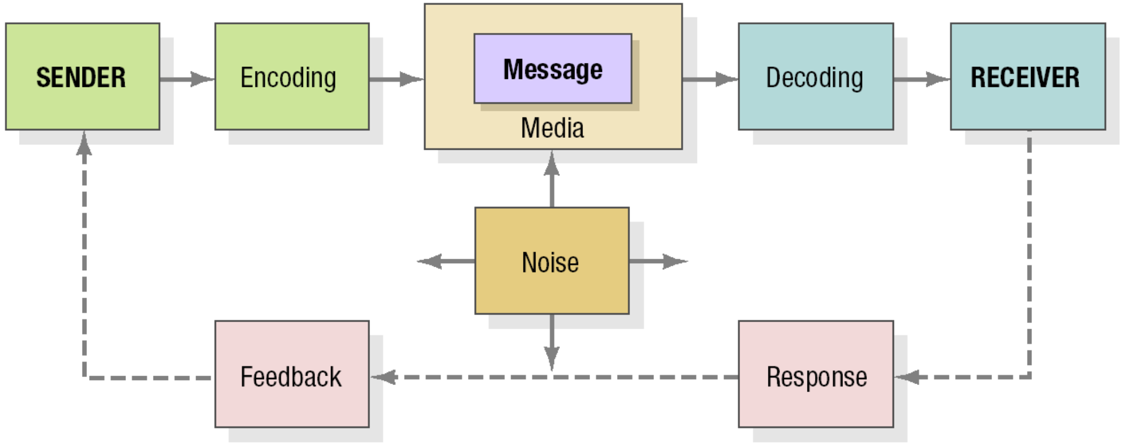
Marketing and Consumers

- Bringing together the needs and wants of the consumer with the products and services that match them.

- Marketing Mix (4Ps)
  - **Product** – the item or service you offer
  - **Price** – enough to make a profit and not too much for the market to bear
  - **Place** – distribution channels where a consumer can get access
  - **Promotion** - how you communicate the existence of your product or service and its benefits

Communicate to Consumers: Marketing Communications

- the means by which firms attempt to inform, persuade, and remind consumers, directly or indirectly, about the products and brands they sell.



Consumer Insight

- "Insight" combines several ideas.
- It includes "classic" areas, such as
  - knowing who consumers are,
  - what they do,
  - where they are,
  - what they buy,
  - what they would like to buy,
  - what media they are exposed to and what media they choose to view, listen to or read.

- It also includes more psychological areas –
  - what consumers think and feel,
  - what their objectives and strategies are,
  - and how these influence how they behave.

Above The Line vs Through The Line vs Below The Line

Above The Line (ATL)

- Above the line marketing also above the line advertising and above the line promotion is an advertising activity that helps companies to use mass marketing strategies.

Through The Line (TTL)

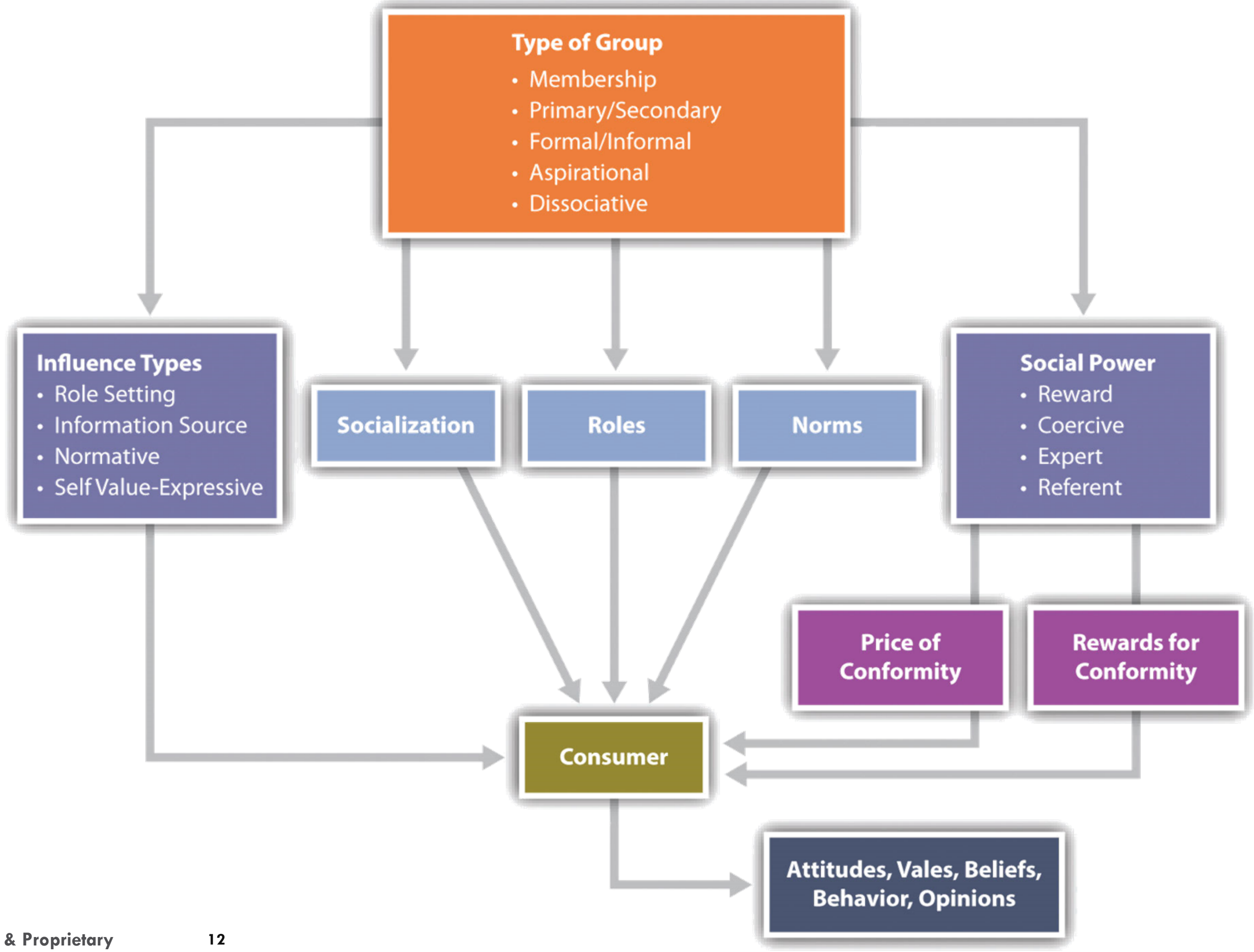
- Through the Line advertising also Through the Line promotion combine both Above the Line and Below the Line marketing strategies. Integration of both ATL and BTL strategies is crucial in era of technological advancement and competition. These days marketers really want to build brands visibility and brand recall at the same time that is why they use TTL advertising strategies.

Below The Line (BTL)

- Below the line marketing also below the line advertising and below the line promotion is a direct advertising strategy to reach the targeted customers. When companies follow this type of line marketing strategy, they focus on conversation rate rather than brand building and awareness.

Consumer Differences

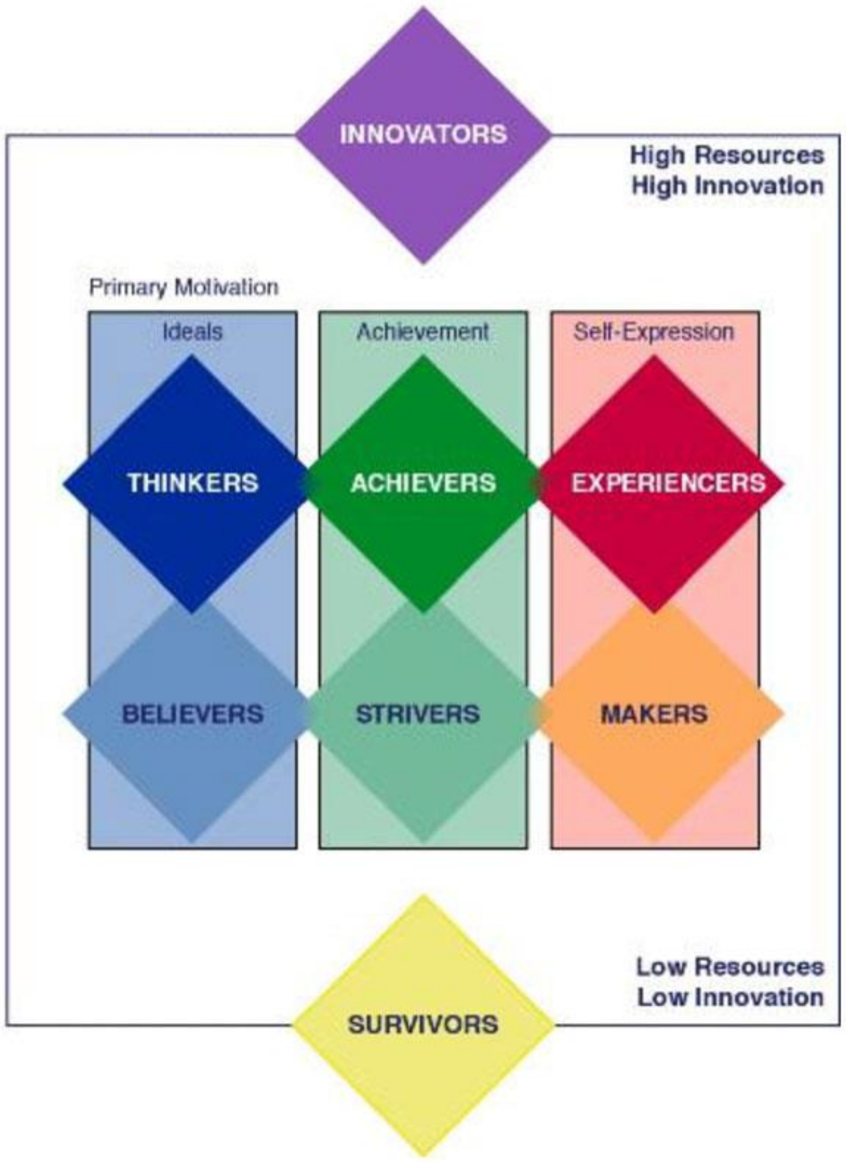
Individual preferences create different demands



[Activity] Consumer Type

- Take a survey

http://www.strategicbusinessinsights.com/vals/presurvey.shtml



Innovators

- As a consumer group, Innovators exhibit all three primary motivations in varying degrees
- Members of this group typically:
  - Are always taking in information (antennas up)
  - Are confident enough to experiment
  - Make the highest number of financial transactions
  - Are skeptical about advertising
  - Have international exposure
  - Are future oriented
  - Are self-directed consumers
  - Believe science and R&D are credible
  - Are most receptive to new ideas and technologies
  - Enjoy the challenge of problem solving
  - Have the widest variety of interests and activities.

Thinkers

- As a consumer group, Thinkers have high resources and an Ideals motivation.
- Members of this group typically:
  - Have "ought" and "should" benchmarks for social conduct
  - Have a tendency toward analysis paralysis
  - Plan, research, and consider before they act
  - Enjoy a historical perspective
  - Are financially established
  - Are not influenced by what's hot
  - Use technology in functional ways
  - Prefer traditional intellectual pursuits
  - Buy proven products.

Believers

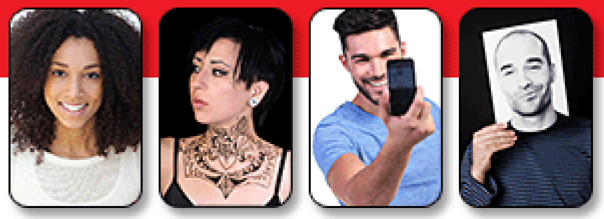
- As a consumer group, Believers have low resources and an Ideals motivation.
- Members of this group typically:
  - Believe in basic rights and wrongs to lead a good life
  - Rely on spirituality and faith to provide inspiration
  - Want friendly communities
  - Watch TV and read romance novels to find an escape
  - Want to know where things stand; have no tolerance for ambiguity
  - Are not looking to change society
  - Find advertising a legitimate source of information
  - Value constancy and stability (can appear to be loyal)
  - Have strong me-too fashion attitudes.

Achievers

- As a consumer group, Achievers have high resources and an Achievement motivation.
- Members of this group typically:
  - Have a "me first, my family first" attitude
  - Believe money is the source of authority
  - Are committed to family and job
  - Are fully scheduled
  - Are goal oriented
  - Are hardworking
  - Are moderate
  - Act as anchors of the status quo
  - Are peer conscious
  - Are private
  - Are professional
  - Value technology that provides a productivity boost.

Strivers

- As a consumer group, Strivers have low resources and an Achievement motivation.
- Members of this group typically:
  - Have revolving employment; high temporary unemployment
  - Use video and video games as a form of fantasy
  - Are fun loving
  - Are imitative
  - Rely heavily on public transportation
  - Are the center of low-status street culture
  - Desire to better their lives but have difficulty in realizing their desire
  - Wear their wealth.



Experiencers

- As a consumer group, Experiencers have high resources and a Self-Expression motivation.
- Members of this group typically:
  - Want everything
  - Are first in and first out of trend adoption
  - Go against the current mainstream
  - Are up on the latest fashions
  - Love physical activity (are sensation seeking)
  - See themselves as very sociable
  - Believe that friends are extremely important
  - Are spontaneous
  - Have a heightened sense of visual stimulation.

Makers

- As a consumer group, Makers have low resources and a Self-Expression motivation.
- Members of this group typically:
  - Are distrustful of government
  - Have a strong interest in all things automotive
  - Have strong outdoor interests (hunting and fishing)
  - Believe in sharp gender roles
  - Want to protect what they perceive to be theirs
  - See themselves as straightforward; appear to others as anti-intellectual
  - Want to own land.

Survivors

- As a consumer group, Survivors have the lowest resources; they exhibit no primary motivation.
- Members of this group typically:
  - Are cautious and risk averse
  - Are the oldest consumers
  - Are thrifty
  - Are not concerned about appearing traditional or trendy
  - Take comfort in routine, familiar people, and places
  - Are heavy TV viewers
  - Are loyal to brands and products
  - Spend most of their time alone
  - Are the least likely use the internet
  - Are the most likely to have a landline-only household.

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 2
Consumer Behaviors

Consumer Behavior

- Consumer behavior encompasses mental and physical activities that consumers engage in when searching for, evaluating, purchasing, and using products and services.

- Example of long-term care insurance

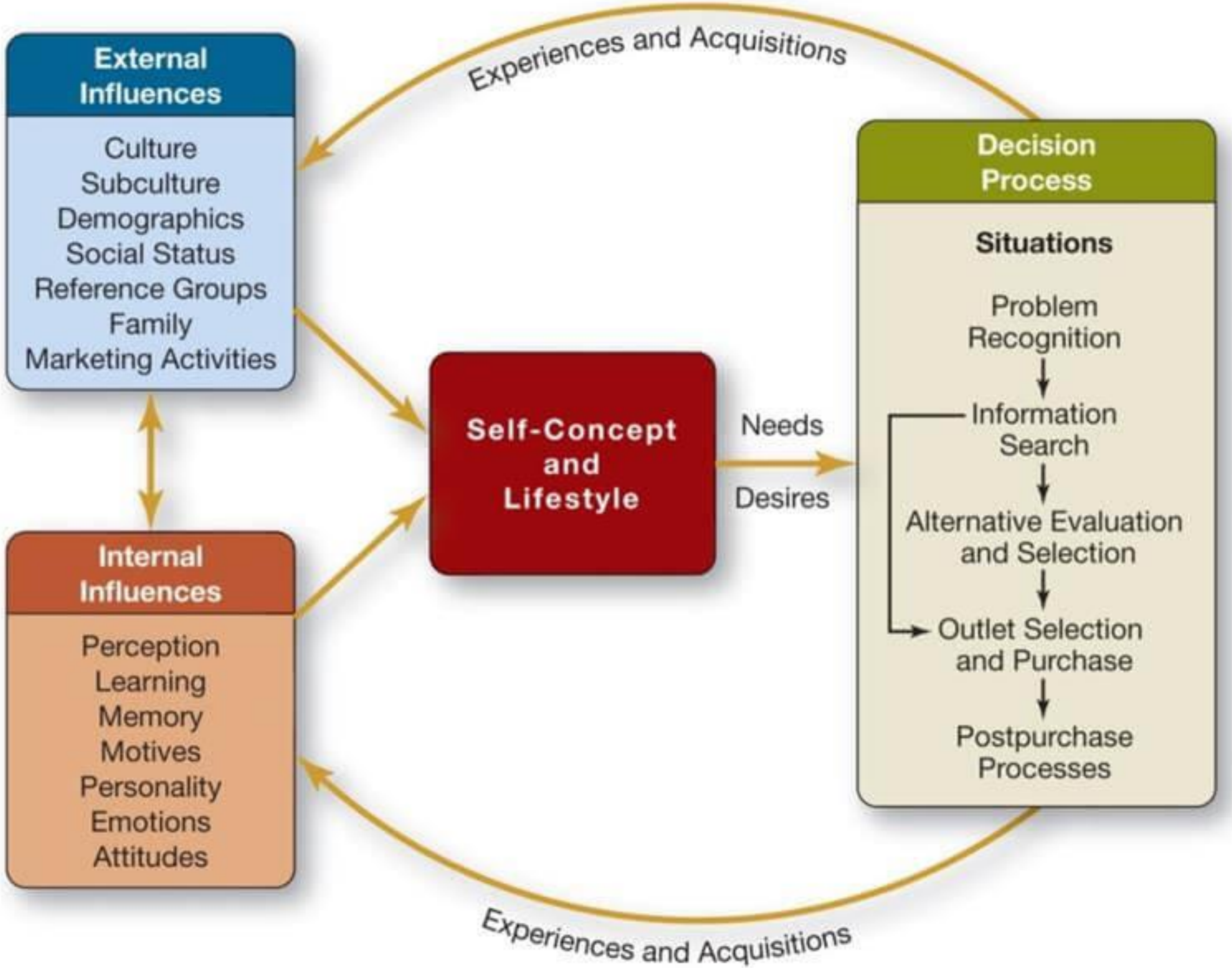
Key Questions	Details
Characteristics of consumers who buy it	Income, Age, Lifestyle
Where they buy it	Agent, Telesales, Online
When they buy it	After a critical event, after seeing an ad
How they buy it	Comparing many policies, select the same one that a friend has
Why they buy it	Fear of depleting life savings, desire for excellent care in old age
What happens after they buy it	Satisfaction with the decision and the company

Consumer Behavior Exercise

Product:

Key Questions	Details
Characteristics of consumers who buy it	
Where they buy it	
When they buy it	
How they buy it	
Why they buy it	
What happens after they buy it	

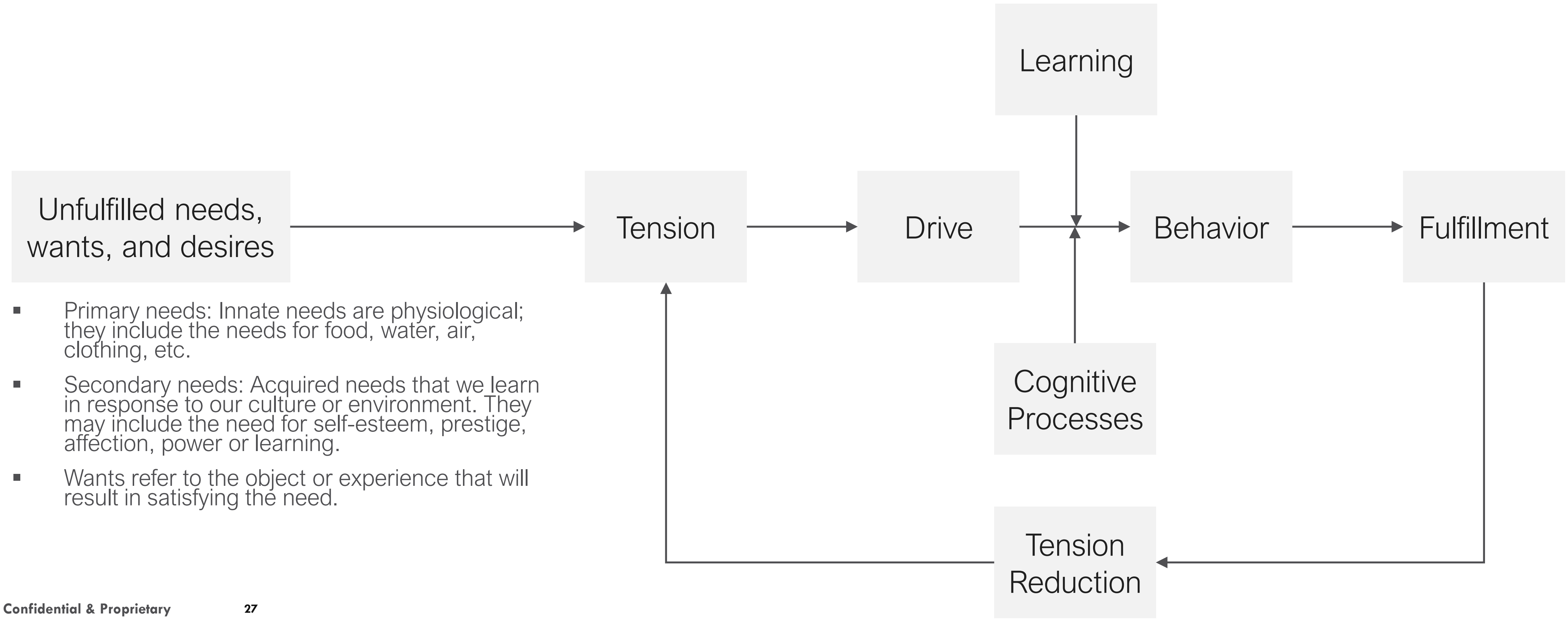
Consumer Decision Making Process



Consumer as an individual

Motivation

- The driving force within individuals that impels them to action
- It is produced by a state of tension, which exists as the result of an unfulfilled needs



- Primary needs: Innate needs are physiological; they include the needs for food, water, air, clothing, etc.
- Secondary needs: Acquired needs that we learn in response to our culture or environment. They may include the need for self-esteem, prestige, affection, power or learning.
- Wants refer to the object or experience that will result in satisfying the need.

Two types of motivation

- **Negative Motivation** — Choosing a course of action or behavior based on an undesirable experience that may follow that particular action. “I don’t want bad things to happen to me.”

- **Positive Motivation** — Choosing a course of action or behavior based on a desirable experience that may follow that particular action. “I want good things to happen to me.”

Needs vs. Wants

- Needs
  - Human needs are the basic requirements and include food, clothing and shelter. Without these humans cannot survive. An extended part of needs today has become education and healthcare. Generally, the products which fall under the needs category of products do not require a push.
  - Instead the customer buys it themselves. But in todays tough and competitive world, so many brands have come up with the same offering satisfying the needs of the customer, that even the "needs category product" has to be pushed in the customers mind.

- Wants
  - Wants are a step ahead of needs and are largely dependent on the needs of humans themselves. For example, you need to take a bath. But you take baths with the best soaps. Thus, Wants are not mandatory part of life. You DONT need a good smelling soap. But you will definitely use it because it is your want.

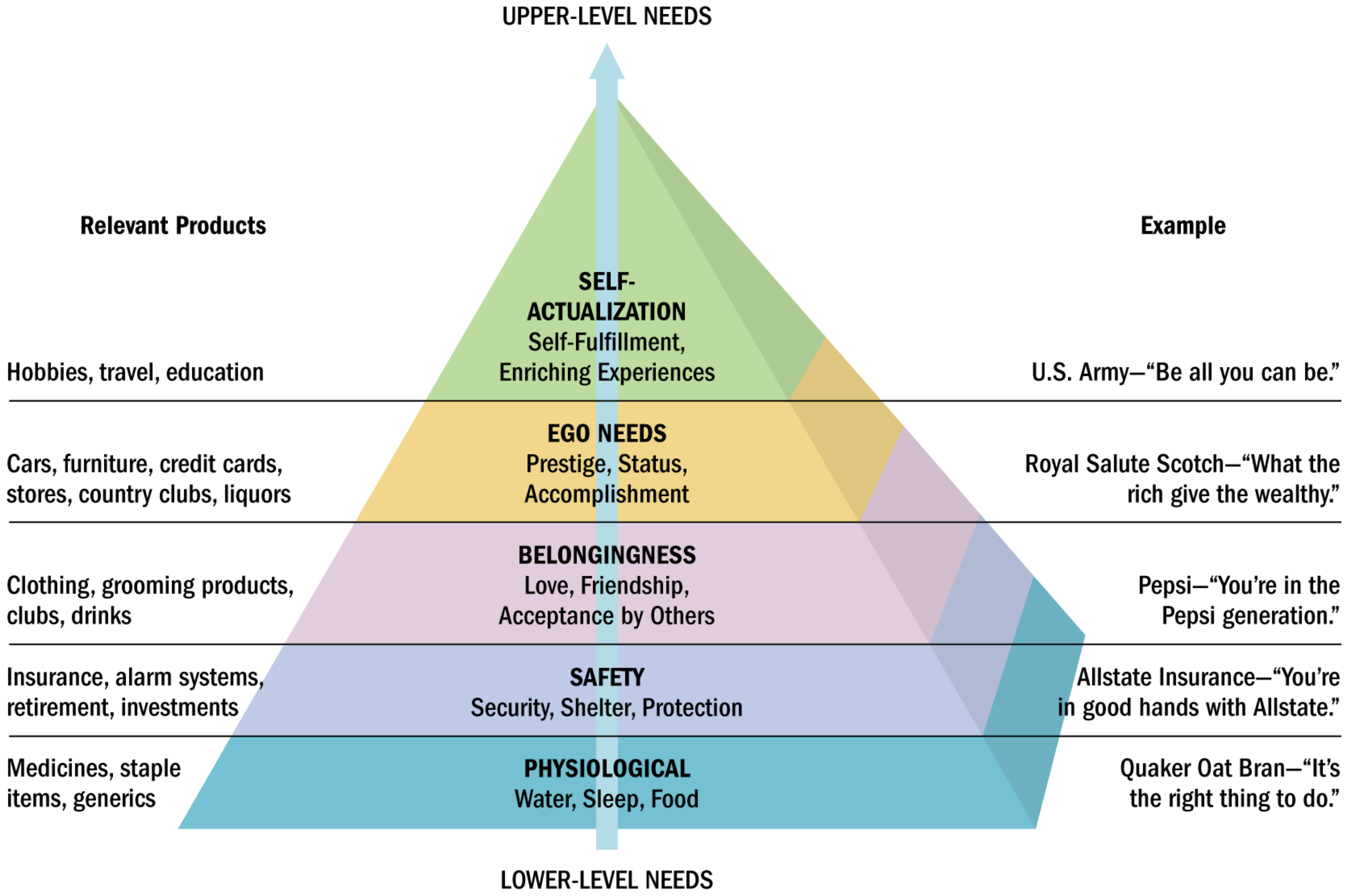
Motivational Direction

- Needs Versus Wants:
  Want: The particular form of consumption used to satisfy a need.

- Types of Needs
  Biogenic needs: Needs necessary to maintain life
  Psychogenic needs: Culture-related needs (e.g. need for status, power, affiliation, etc.)
  Utilitarian needs: Implies that consumers will emphasize the objective, tangible aspects of products
  Hedonic needs: Subjective and experiential needs (e.g. excitement, self-confidence, fantasy, etc.)

Classifying Consumer Needs

Maslow's Hierarchy of Needs



The Six Human Needs

- CERTAINTY: The need to feel safe, comfortable, secure, stable, protected, and have predictability in out lives.
- SIGNIFICANCE: The need to feel important, achievement, respect, special, needed, wanted and unique in our lives.
- VARIETY / UNCERTAINTY: The need to feel different, challenged, risk, change, excitement, surprise and entertained in our lives.
- CONECTION / LOVE: The need to feel togetherness, passion, unity, warmth, desire and love in our lives.
- GROWTH: The need to feel like we are developing, learning, strengthening, expanding, and cultivating ourselves.
- CONTRIBUTION: The need to feel like we are giving, donating, leaving our mark, serving, offering and contributing to others.

Jobs to be done

Clayton Christensen, the late author of The Innovator's Dilemma and former Harvard Business School professor, made the case that to understand what motivates people to act, you first must understand what it is they to need to get done.

"When people find themselves needing to get a job done, they essentially hire products to do that job for them ...If a [businessperson] can understand the job, design a product and associated experiences in purchase and use to do that job, and deliver it in a way that reinforces its intended use, then *when customers find themselves needing to get that job done they will hire that product*."

https://www.fullstory.com/blog/clayton-christensen-jobs-to-be-done-framework-product-development

Clayton Christensen's milkshake mystery

https://www.youtube.com/watch?t=41s&v=sfGtw2C95Ms&feature=youtu.be

- Standard market research was conducted
  - They brought in customers who fit the profile of the quintessential milkshake consumer. They'd give them samples and ask, "Could you tell us how we could improve our milkshakes so you'd buy more of them? Do you want it chocolate-ier, cheaper, chunkier, or chewier?"
  - They'd get very clear feedback and they'd improve the milkshake on those dimensions and it had no impact on sales or profits whatsoever.

Clayton Christensen's milkshake mystery

https://www.youtube.com/watch?t=41s&v=sfGtw2C95Ms&feature=youtu.be

- So one of our colleagues went in with a different question on his mind. And that was, “I wonder what job arises in people’s lives that cause them to come to this restaurant to hire a milkshake?”

- The team stood in a restaurant for 18 hours one day and just took very careful data. What time did they buy these milkshakes? What were they wearing? Were they alone? Did they buy other food with it? Did they eat it in the restaurant or drive off with it?

- It turned out that nearly half of the milkshakes were sold before 8 o’clock in the morning. The people who bought them were always alone. It was the only thing they bought and they all got in the car and drove off with it.

Clayton Christensen's milkshake mystery

https://www.youtube.com/watch?t=41s&v=sfGtw2C95Ms&feature=youtu.be

- To figure out what job they were trying to hire it to do, we came back the next day and stood outside the restaurant so we could confront these folks as they left milkshake-in-hand. And in language that they could understand we essentially asked, “Excuse me please but I gotta sort this puzzle out. What job were you trying to do for yourself that caused you to come here and hire that milkshake?”

- They’d struggle to answer so we then helped them by asking other questions like, “Well, think about the last time you were in the same situation needing to get the same job done but you didn’t come here to hire a milkshake. What did you hire?”

- And then as we put all their answers together it became clear that they all had the same job to be done in the morning. That is that they had a long and boring drive to work and they just needed something to do while they drove to keep the commute interesting. One hand had to be on the wheel but someone had given them another hand and there wasn’t anything in it. And they just needed something to do when they drove. They weren’t hungry yet but they knew they would be hungry by 10 o’clock so they also wanted something that would just plunk down there and stay for their morning.

Clayton Christensen's milkshake mystery

https://www.youtube.com/watch?t=41s&v=sfGtw2C95Ms&feature=youtu.be

- "Well, think about the last time you were in the same situation needing to get the same job done but you didn't come here to hire a milkshake. What did you hire?"

- "Good question. What do I hire when I do this job? You know, I've never framed the question that way before, but last Friday I hired a banana to do the job. Take my word for it. Never hire bananas. They're gone in three minutes—you're hungry by 7:30am. If you promise not to tell my wife I probably hire donuts twice a week, but they don't do it well either. They're gone fast. They crumb all over my clothes. They get my fingers gooey. Sometimes I hire bagels but as you know they're so dry and tasteless. Then I have to steer the car with my knees while I'm putting jam on it and if the phone rings we got a crisis. I remember I hired a Snickers bar once but I felt so guilty I've never hired Snickers again. Let me tell you when I hire this milkshake it is so viscous that it easily takes me 20 minutes to suck it up through that thin little straw. Who cares what the ingredients are—I don't. All I know is I'm full all morning and it fits right here in my cupholder."

Clayton Christensen's milkshake mystery

https://www.youtube.com/watch?t=41s&v=sfGtw2C95Ms&feature=youtu.be

Well it turns out that the milkshake does the job better than any of the competitors, which in the customer’s minds are not Burger King milkshakes but bananas, donuts, bagels, Snickers bars, coffee, and so on.

Jobs to be done

- Jobs to be Done is a theory of consumer action. It describes the mechanisms that cause a consumer to adopt an innovation.

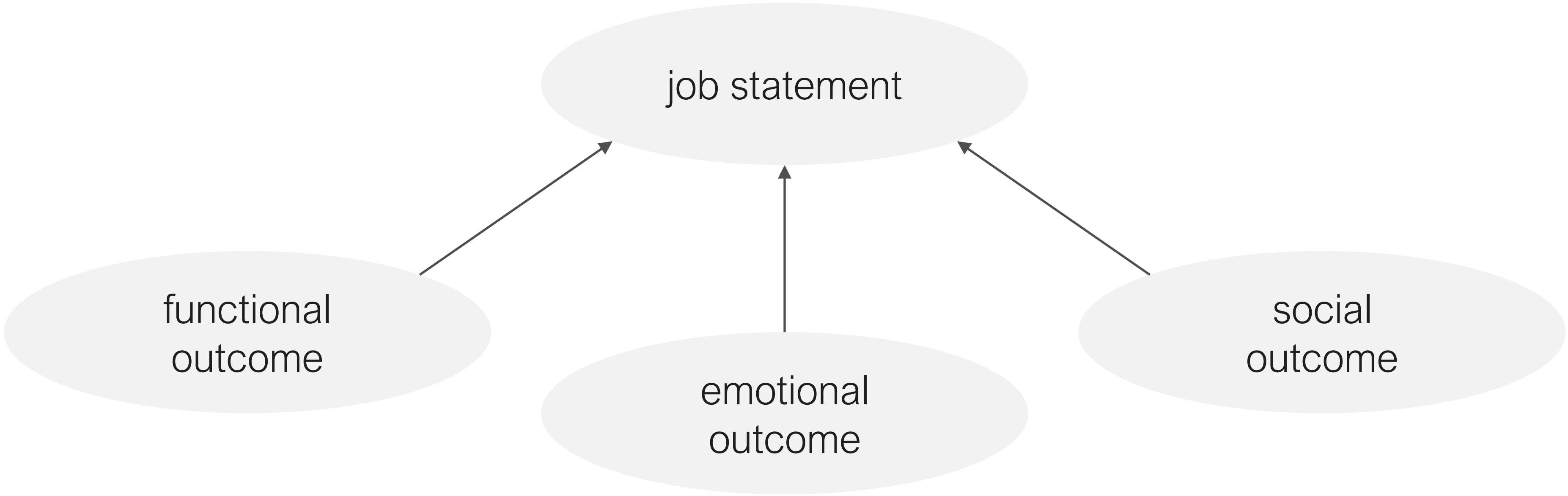
- The theory states that markets grow, evolve, and renew whenever customers have a Job to be Done, and then buy a product to complete it (get the Job Done). This makes a Job to be Done a process: it starts, it runs, and it ends. The key difference, however, is that a JTBD describes how a customer changes or wishes to change.

> A Job to be Done is the process a consumer goes through whenever she aims to change her existing life-situation into a preferred one, but cannot because there are constraints that stop her.

https://jtbd.info/2-what-is-jobs-to-be-done-jtbd-796b82081cca

So, What is jobs to be done?

- Jobs to be Done is about a customer making progress towards a desired future.



Example of Job Statements

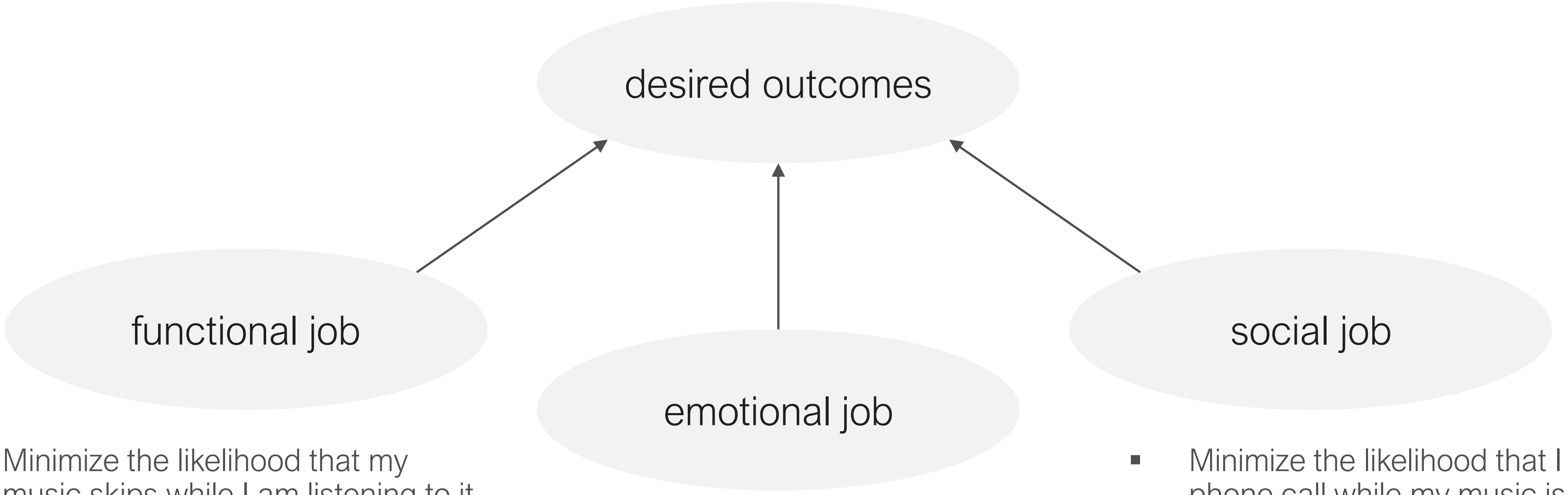
VERB	OBJECT OF CONTROL	CONTEXT
Transport	me and my belongings	between different cities
Teach	programming	to children aged 5 to 7
Generate	a work schedule	for my high-turnover small business
Listen	to music	in my home
Find	authentic experiences	when I travel to new places
Cut	wood	in a straight line
Find	information	on the internet
Spot	wildfires	before they are uncontrollable

Example of Desired Outcomes

IMPROVEMENT	OBJECT OF CONTROL	CONTEXT
Minimize the likelihood	that my music skips	while I am listening to it
Minimize the time it takes	to sort the songs	into the desired order for listening
Minimize the time it takes	to discover new songs	that are similar to songs that I like
Minimize the time it takes	to remove songs I don’t like	from my playlist
Minimize the likelihood	that I miss a phone call	while my music is playing loudly
Minimize the time it takes	to change the volume	while the music is playing
Minimize the likelihood	that I hear the same song twice	within 30 minutes
Minimize the likelihood	that I find a song distracting	when I listen to music while studying

Example

Listen to music at home



functional job

- Minimize the likelihood that my music skips while I am listening to it
- Minimize the time it takes to sort the songs into the desired order for listening

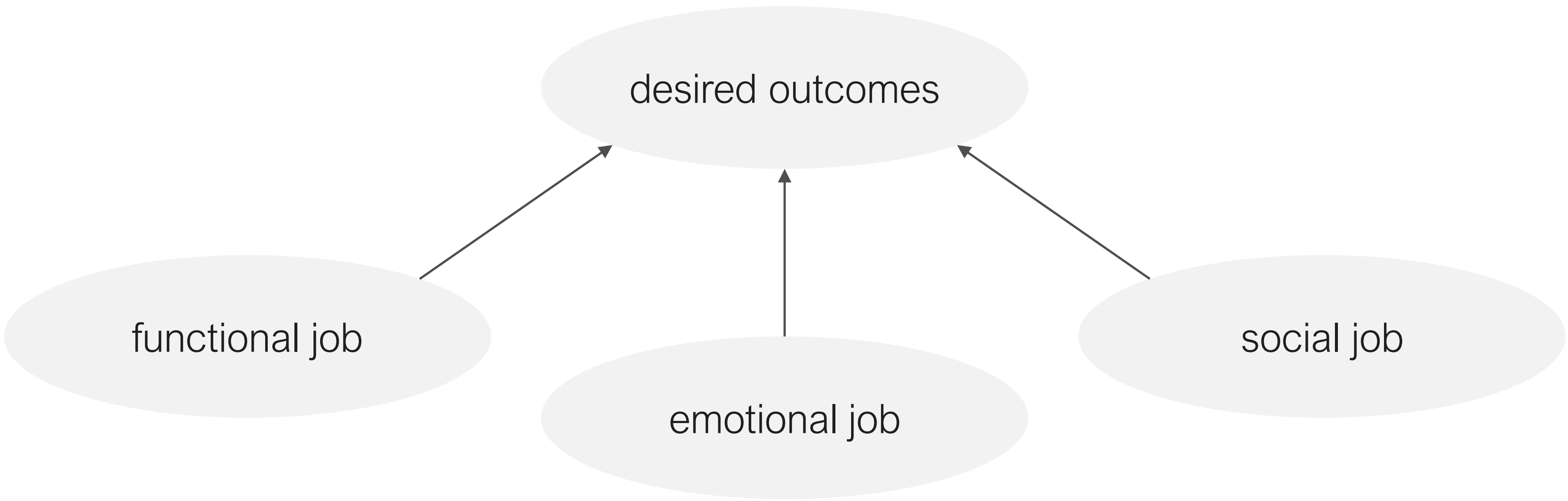
emotional job

- Minimize the time it takes to discover new songs that are similar to songs that I like

social job

- Minimize the likelihood that I miss a phone call while my music is playing loudly

Exercise



Consumers in their social and cultural settings

Reference Group

- Any person or group that serves as a point of comparison for an individual in forming either general or specific values, attitudes or behavior.

- Normative reference groups – influencing general values or behaviors
  - Ex. Family – which foods to select, appropriate ways to dress, how and where to shop

- Comparative reference groups – serving as benchmarks for specific or narrowly defined attitudes or behavior.
  - Ex. Neighbor family – the way they maintain their home, their choice of home furnishings and cars, where they take holidays

Classification of reference groups

	Regular face-to-face contact (membership)	Non-regular face-to-face contact (non-membership)
Positive influence (Adopting attitudes and behavior)	Contactual Group	Aspirational Group
Negative influence (Adopting opposite attitudes and behavior)	Disclaimant Group	Avoidance Group

Selected Consumer-Relevant Reference Groups

- Friendship groups
  - Consumers are more likely to seek information from the friends they believe have values or outlooks similar to their own
- Shopping groups
  - They are often offshoots of family or friendship groups. Shopping with others also provides an element of social fun and reduce the risk that a purchase decision will be socially unacceptable.
- Work groups
  - Consumers spend the sheer amount of time at their jobs providing opportunity to influence on the consumption behavior
- Virtual groups or communities
  - A new norm of social interaction provides great opportunity of influence.

Roles in the family decision-making process

- Influencers – providing information
- Gatekeepers – control the flow of information
- Deciders – with the power to make choices
- Buyers – make actual purchase
- Preparers – transform the product into a form suitable for consumption
- Users – use or consume
- Maintainers – service or repair
- Disposers – dispose or discontinue

Social Class

- Two-category
  - Blue collar, white collar
  - Lower, upper
- Three-category
  - Blue collar, grey collar, white collar
  - Lower, middle, upper
- Four-category
  - Lower, lower-middle, upper-middle, upper
- Five-category
  - Lower, lower-middle, middle, upper-middle, upper
- Six-category
  - Lower-lower, upper-lower, lower-middle, upper-middle, lower-upper, upper-upper

Social class profiles

- Upper-upper: country club establishment
  - Belong to best country clubs
  - Accustomed to wealth, so do not spend money conspicuously
- Lower-upper: new wealth
  - Represent ‘new money’
  - Conspicuous users of their new wealth
- Upper-middle: achieving professionals
  - Career-oriented and active in professional, community, and social activities
  - Consumption is often conspicuous
- Lower-middle: faithful followers
  - Want to achieve respectability and be accepted as good citizens
  - Do-it-yourself lifestyle
- Upper-lower: security-minded majority
  - Strive for security
  - View work as a means to buy enjoyment
- Lower-lower: rock bottom
  - Live day-to-day existence
  - Unskilled labors

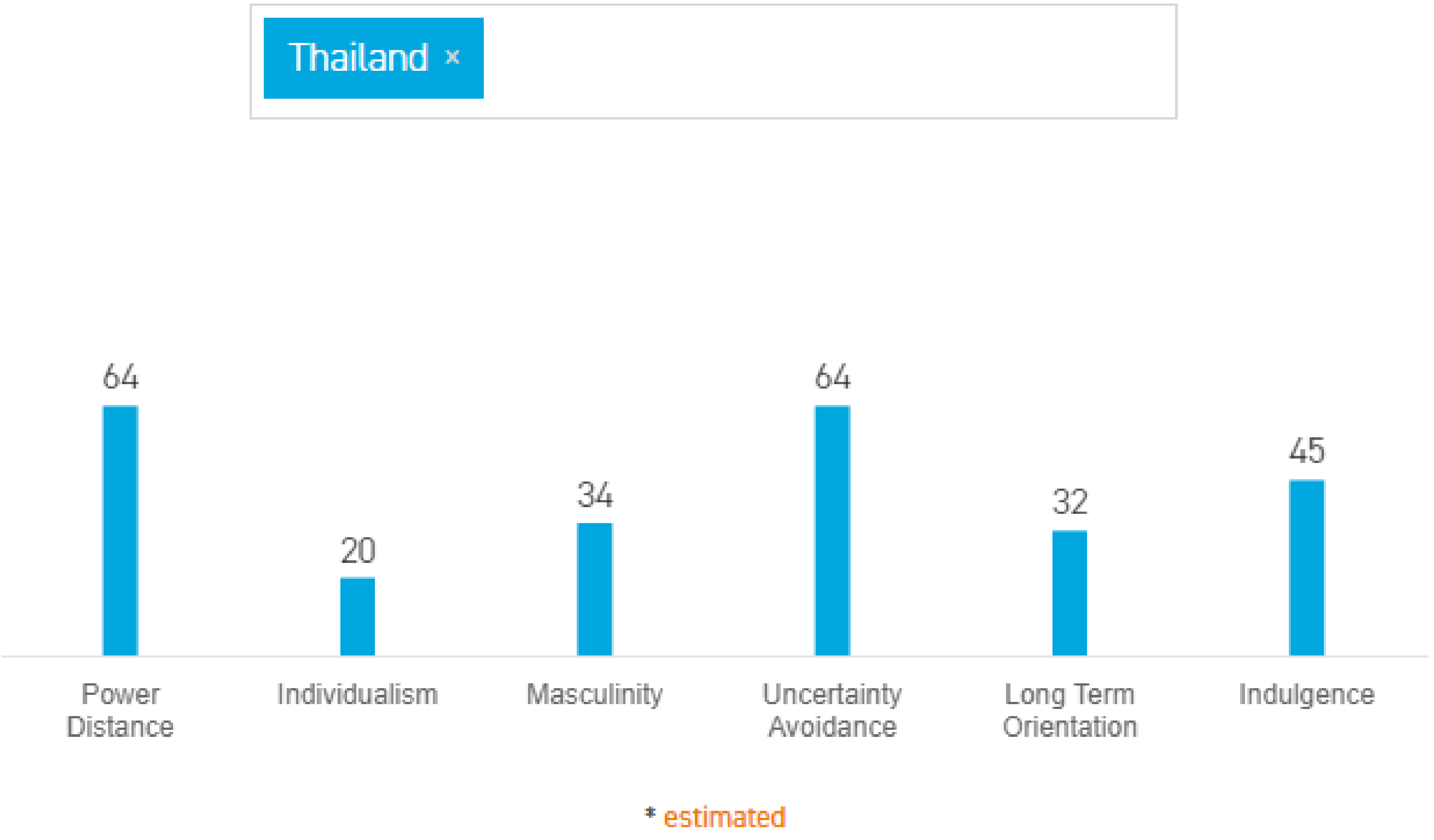
Influence of Culture

- Certain cultures are more youth-oriented than others hence are more liberal and individualistic, and are more likely to work harder and spend more.
- Many Islamic cultures and some Catholic cultures are much more religiously oriented as compared to Chinese culture where religion plays a very small role. This implies that in Islamic and some Catholic cultures people are more inclined to buy and consume religious artifacts and related material.
- Mexico adolescents are more likely to seek parental advice or respond positively to ads with parental figures in the purchase of items ranging from candy to movies to fashion clothing than United States.
- Japan and China people tend to consume products and avail services that everyone else is consuming, where as in United Kingdom and Unites States people are more inclined to make their own individual decisions bases on personal preferences and tastes.

Hofstede's Cultural Dimensions

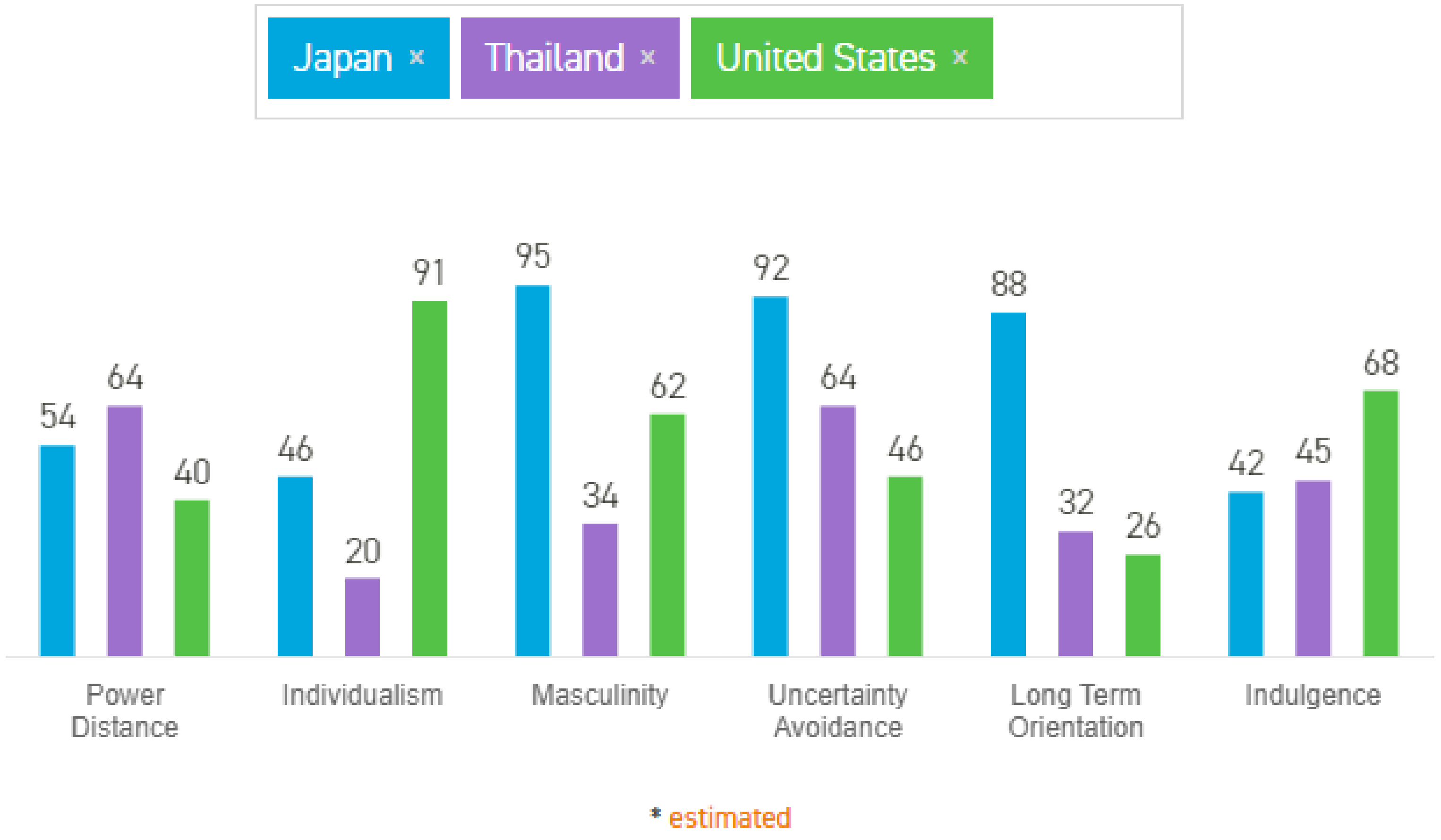
0 ←--------	HOFSTEDE'S CULTURAL DIMENSIONS	------→ 100	
LOW POWER DISTANCE	PDI	HIGH POWER DISTANCE	This dimension deals with the fact that all individuals in societies are not equal
COLLECTIVISTIC	INV	INDIVIDUALISTIC	It has to do with whether people´s self-image is defined in terms of "I" or "We".
FEMININE	MAS	MASCULINE	A high score (Masculine) indicates that the society will be driven by competition and success
LOW UNCERTAINTY AVOIDANCE	UAI	HIGH UNCERTAINTY AVOIDANCE	The way that a society deals with the fact that the future can never be known
SHORT TERM ORIENTATION	LTO	LONG TERM ORIENTATION	how every society dealing with the challenges of the present and future
RESTRAINT	IND	INDULGENCE	the degree to which small children are socialized

Thailand



Thailand ×
64
20
34
64
32
45
Power Distance
Individualism
Masculinity
Uncertainty Avoidance
Long Term Orientation
Indulgence
* estimated

Thailand – Japan - USA



Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 3
Customer Relationship Management

Roots of Customer Relationships and Experience

The Shifting Source of Competitive Advantage

WHAT ELSE CAN WE MAKE AND SELL?				WHAT ELSE CAN WE DO FOR OUR CUSTOMERS?		
UPSTREAM ACTIVITIES				**DOWNSTREAM ACTIVITIES**		
Sourcing	**Production**	**Logistics**	**Innovation**	**Shaping Customer Perception**	**Innovation**	**Building Accumulative Advantage**
Contract with lowest-cost suppliers	Reduce costs/ maximize scale and throughput	Optimize supply-chain and distribution efficiency	Build better products	Define competitive set Change purchase criteria Build trust	Tailor offering to consumption circumstances Reduce customer costs and risks	Harness network effects Accrue and deploy customer data

FIXED COSTS, CUSTOMER VALUE, AND COMPETITIVE ADVANTAGE ARE MOVING DOWNSTREAM

CRM as Source of Competitive Advantage

- What is CRM?
  - a technology or software solution that helps track data and information about customers to enable better customer service
  - a strategy or an ongoing process that helps transform the enterprise from a focus on traditional selling or manufacturing to a customer focus while increasing revenues and profits in the current period and the long term
  - the leadership and commitment necessary to cascade throughout the organization the thinking and decision-making capability that puts customer value and relationships first as the direct path to increasing shareholder value
  - a set of business practices designed to put an enterprise into closer and closer touch with its customers, in order to learn more about each one and to deliver greater and greater value to each one, with the overall goal of making each one more valuable to the firm to increase the value of the enterprise

Increasing the Value of the Customer Base

Get

- **Acquire** profitable **customers.**

Keep

- **Retain** profitable **customers longer.**
- **Win back** profitable **customers.**
- **Eliminate** unprofitable **customers.**

Grow

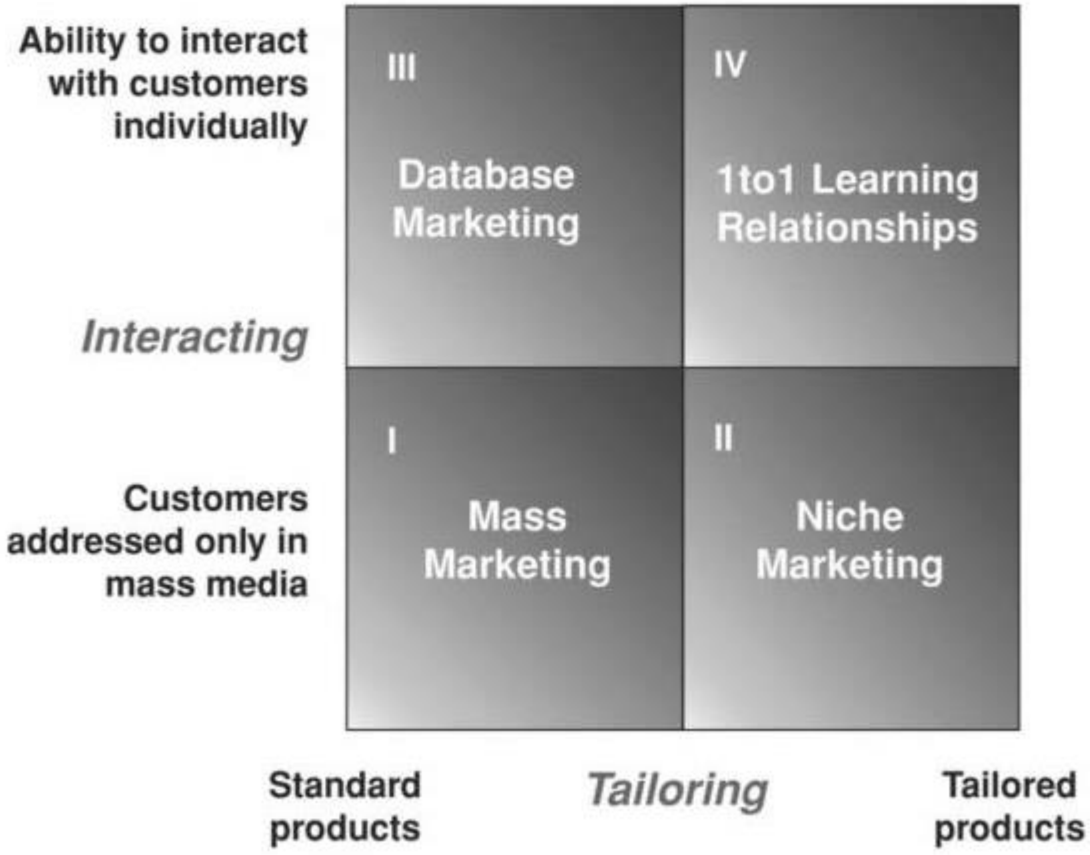
- Upsell **additional products in a solution.**
- Cross-sell **other products to customers.**
- Referral **and word-of-mouth benefits.**
- Reduce **service and operational costs.**

Operational CRM vs. Analytical CRM

- Operational CRM focuses on the software installations and the changes in process affecting the day-today operations of a firm-operations that will produce and delivery different treatments to different customers

- Analytical CRM focuses on the strategic planning needed to build customer value as well as the cultural, measurement, and organizational changes required to implement that strategy successfully

Enterprise Strategy Map



Comparison of Market-Share and Share-of-Customer Strategies

Market-Share Strategy	Share-of-Customer Strategy
Company sees products and brands as the source of all company value.	Company sees customers as—by definition—the only source of revenue.
Product (or brand) managers sell one product at a time to as many customers as possible.	Customer manager sells as many products as possible to one customer at a time.
Differentiate products from competitors.	Differentiate customers from each other.
Sell *to* customers.	Collaborate *with* customers.
Find a constant stream of new customers.	Find a constant stream of new business from established customers.
Company makes sure each product, and likely each transaction, is profitable, even at the cost of a customer's confidence.	Company makes sure each customer is profitable, even if that means losing money on an occasional product or transaction.
Use mass media to build brand and announce products.	Use interactive communication to determine individual needs and communicate with each individual.

What is relationship?

- Is a relationship possible if the company knows the customer and tailors offers and communications, remembers things for the customer, and deliberately builds customer experience-even if the customer is not aware of a relationship?
- Is it possible for a customer to have a relationship with a brand?
- Managing the customer relationship is all about what the company does, and customer experience is what the customer feels like a result
- The exchange between a customer and the enterprise becomes mutually beneficial, as customer gives information in return for personalized service that meets their individual needs
- The interaction forms the basis of the Learning Relationship, based on a collaborative dialogue between the enterprise and the customer that grows smarter and smarter with each successive interaction

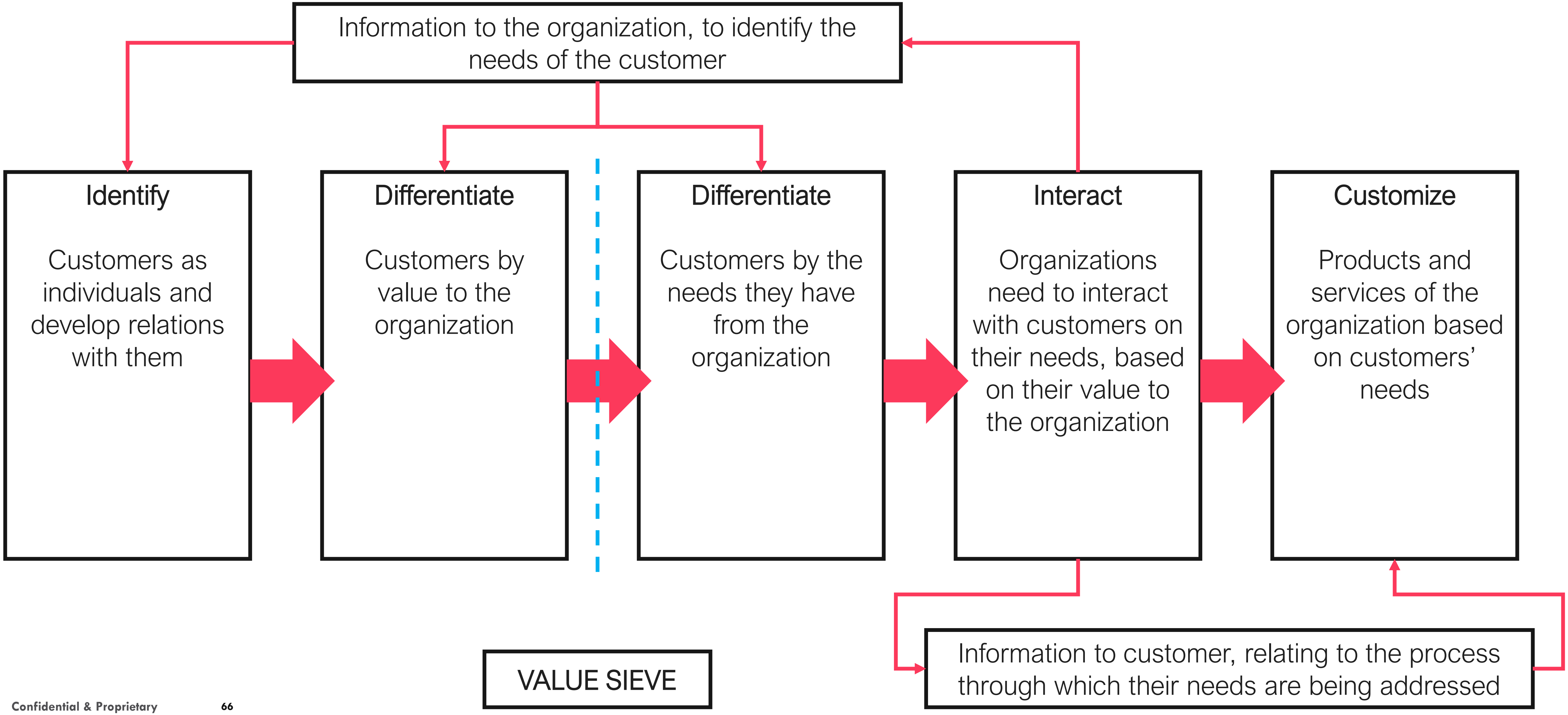
Characteristics of a Genuine Business Relationship

- Mutual – two-way in nature. A brand must be aware of the individual customer's existence
- Interactive – information exchange
- Iterative – gives future interactions greater efficiency
- Provides ongoing benefit to both parties – reducing cost of interaction
- Requires a change in behavior for both parties – history, current, and future actions
- Unique - individualized
- Trust – trust and affection and satisfaction are related feelings

4 Levels of Relationships

- **No-Contact:** relationships rarely or never require a customer to interact with an enterprise directly. Customers typically interact with a distributor or agent, as in the case of buying a favorite brand of soda at a supermarket
- **Distant:** relationships involve less frequent interactions and might occur over the telephone, online, or through videoconferencing
- **Face-to-face:** customer relationships may of may not require the customer to reveal personal information. Such relationships often occur in retail store
- **Intimate:** relationships are characterized as personal and friendly and generally involve the disclosure of personal information. Such relationships may involve physical touch, as in the relationship between doctors and patients or hairstylists and clients.

IDIC Methodology



Identify customer expectations

- A business working towards customer relationship management has to first identify its customer needs, wants and preferences. This helps to curate new products and strategies oriented towards customers.
- For instance, in terms of identification, Nestle conducted a market survey to identify the customer base for their coffee by assessing the nations where caffeine intake was high. It found that Western countries had a more positive outlook towards coffee drinks while Eastern nations preferred tea as their caffeine drink. The company then established its coffee brand, Nescafe, in the Western countries. The brand also offered coffee-flavored candies in Eastern countries to gauge reaction to the flavor.

Differentiate customer expectations

- To meet customer expectations, it is important to differentiate customer base into value and need. This differentiation helps to prepare strategies that meet company goals and customer expectations for strong customer relations.

- Differentiation by value helps to identify customers that are loyal and should be focused on. Their feedback can be valuable for product line extension, increased efficacy, and performance, as well as revenue generation.

- Differentiating existing customers on the basis of their needs helps to strategize promotions. For instance, apart from analyzing the customers who add value to the company’s brand image, DuPont made it a point to understand that other customers needed effective customer service to address their needs individually.

Customer interaction

- Customer interactions help to understand customer expectations and how they relate to business products and services. The interaction, based on customer value and needs, provides a chance to collect feedback, opinions and wants.
- Unilever always has its sales representatives and customer care executives ready to discuss possible improvements they can make on their product line or take feedback for improvements. This has helped them develop customer-oriented products.

Customizations to meet customer expectations

- After identifying, differentiating and interacting with the customers a fair idea can be drawn for new strategies and product customizations.
- For instance, market research firms like McKinsey realized the need for customizations from the feedbacks of customers. Their customers expected increased customizations on their requests. This will help them in taking their own strategic decisions based on the reports of McKinsey. McKinsey incorporated the clause of providing a percentage of customization leverage in its reports.

The Framework of Understanding: The 5E's of Customer Relationships

- Customer Environment

the company needs to understand the customer environment, in other words, it has to know what is happening in customers lives, what they aiming to achieve. Companies that understand the context in which the customer operates and can see the underlying issues he/she is trying to solve, will have more success in establishing true customer relationship.

- Customer Expectations

In order to build a genuine relationship a business should also understand what the customer is not expecting and surprise him/her by delivering value above the established expectations. Surprising the customer will bring up emotional element, and prompt storytelling and sharing of the experience. To make it clear, customer surprise should be of a positive nature.

The Framework of Understanding: The 5E's of Customer Relationships

- Customer Emotions

By definition, relationships are emotional connections that a person has with other people, companies or brands. Therefore, when building customer relationships business must operate in such a way as to reduce negative customer feelings and strengthen positive ones.

- Customer Experience

In order to establish a long-lasting customer relationship, there should be a consistency in delivering series of positive customer experiences.

- Customer Engagement

Customer engagement is achieved by involving customers in production and delivery of products and services, thus creating high levels of commitment. The value of the engaged customer is in spreading positive word of mouth, creating communities of devoted customers leading to co-dependency and establishing of a strong relationship.

Understanding Customers

The Value Proposition Canvas

Describes the features of your value proposition in a structured and detail way.

Value Map – What / Solution

The Value Map describes a specific offering in terms of specific product features and services, pain relievers, and gain creators.

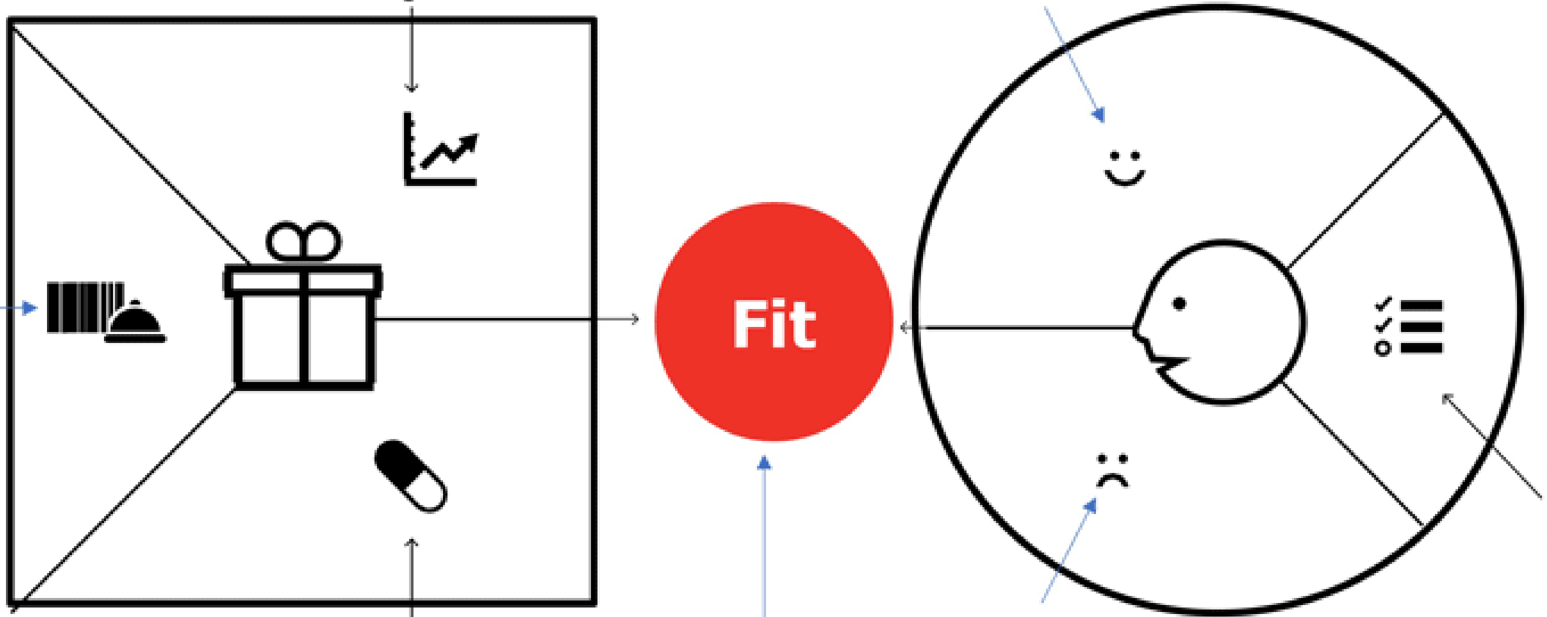
Customer Profile – Why / Problem

The Customer Profile describes a specific customer segment in terms of jobs to be done , points of pain, and desired gains.

GAIN CREATORS describe how your products and services create customer gains.

GAINS describe the outcomes customers want to achieve or the concrete benefits they are seeking

List of all the **PRODUCT FEATURES AND SERVICES** a value proposition is built around.



CUSTOMER JOBS describe what customers are trying to get done in their work and in their lives, as expressed in their own words.

PAIN RELIEVERS describe how your products and services alleviate customer pains.

PAINS describe bad outcomes, risks, and obstacles related to customer jobs.

You achieve FIT when your value map aligns with your customer profile — when your products and services produce pain relievers and gain creators that match one or more of the jobs, pains, and gains that are important to your customer.

Customer Profile

- ## Customer Jobs

Jobs describe the things your customers are trying to get done in their work or in their life. A customer job could be the tasks they are trying to perform and complete, the problems they are trying to solve, or the needs they are trying to satisfy.

- ## Customer Pains

Pains describe anything that annoys your customers before, during, and after trying to get a job done or simply prevents them from getting a job done. Pains also describe risks, that is, potential bad outcomes, related to getting a job done badly or not at all.

- ## Customer Gains

Gains describe the outcomes and benefits your customers want. Some gains are required, expected, or desired by customers, and some would surprise them. Gains include functional utility, social gains, positive emotions, and cost savings.

Value Map

- ## Products and Services

This is simply a list of what your offer. Think of it as all the items your customers can see in your shop window. A list of various types of products and services may include

- Physical/tangible

Goods, such as manufactured products

- Intangible

Products such as copyrights or services such as after-sales assistance

- Digital

Products such as music downloads or services such as online recommendations

- Financial

Products such as investment funds and insurances or services such as the financing of a purchase

- ## Pain Relievers

Pains relivers describe how exactly your products and services alleviate specific customer pains. You don’t need to come up with a pain reliver for every pain. Great value proposition focuses only on a few pains that they alleviate extremely well.

- ## Gain Creators

Gains creators describe how your products and services create customer gains. They explicitly outline how you intend to produce outcomes and benefits.

Customer Jobs – Trigger Questions

- What is the one thing that your customer couldn't live without accomplishing? What are the stepping stones that could help your customer achieve this key job?
- What are the different contexts that your customers might be in? How do their activities and goals change depending on these different contexts?
- What does your customer need to accomplish that involves interaction with others?
- What tasks are your customers trying to perform in their work or personal life? What functional problems are your customers trying to solve?
- Are there problems that you think customers have that they may not even be aware of?
- What emotional needs are your customers trying to satisfy? What jobs, if completed, would give the user a sense of self-satisfaction?
- How does your customer want to be perceived by others? What can your customer do to help themselves be perceived this way?
- How does your customer want to feel? What does your customer need to do to feel this way?
- Track your customer's interaction with a product or service throughout its lifespan. What supporting jobs surface throughout this life cycle? Does the user switch roles throughout this process?

Customer Pains – Trigger Questions

- How do your customers define too costly? Takes a lot of time, costs too much money, or requires substantial efforts?
- What makes your customers feel bad? What are their frustrations, annoyances, or things that give them a headache?
- How are current value propositions under performing for your customers? Which features are they missing? Are there performance issues that annoy them or malfunctions they cite?
- What are the main difficulties and challenges your customers encounter? Do they understand how things work, have difficulties getting certain things done, or resist particular jobs for specific reasons?
- What negative social consequences do your customers encounter or fear? Are they afraid of a loss of face, power, trust, or status?
- What risks do your customers fear? Are they afraid of financial, social, or technical risks, or are they asking themselves what could go wrong?
- What's keeping your customers awake at night? What are their big issues, concerns, and worries?
- What common mistakes do your customers make? Are they using a solution the wrong way?
- What barriers are keeping your customers from adopting a value proposition? Are there upfront investment costs, a steep learning curve, or other obstacles preventing adoption?

Customer Gains – Trigger Questions

- Which savings would make your customers happy? Which savings in terms of time, money, and effort would they value?
- What quality levels do they expect, and what would they wish for more or less of?
- How do current value propositions delight your customers? Which specific features do they enjoy? What performance and quality do they expect?
- What would make your customers’ jobs or lives easier? Could there be a flatter learning curve, more services, or lower costs of ownership?
- What positive social consequences do your customers desire? What makes them look good? What increases their power or their status?
- What are customers looking for most? Are they searching for good design, guarantees, specific or more features?
- What do customers dream about? What do they aspire to achieve, or what would be a big relief to them?
- How do your customers measure success and failure? How do they gauge performance or cost?
- What would increase your customers’ likelihood of adopting a value proposition? Do they desire lower cost, less investment, lower risk, or better quality?

Pain Relievers – Trigger Questions

- ... produce savings? In terms of time, money, or efforts.
- ... make your customers feel better? By killing frustrations, annoyances, and other things that give customers a headache.
- ... fix under-performing solutions? By introducing new features, better performance, or enhanced quality.
- ... put an end to difficulties and challenges your customers encounter? By making things easier or eliminating obstacles.
- ... wipe out negative social consequences your customers encounter or fear? In terms of loss of face or lost power, trust, or status.
- ... eliminate risks your customers fear? In terms of financial, social, technical risks, or things that could potentially go wrong.
- ... help your customers better sleep at night? By addressing significant issues, diminishing concerns, or eliminating worries.
- ... limit or eradicate common mistakes customers make? By helping them use a solution the right way.
- ... eliminate barriers that are keeping your customer from adopting value propositions? Introducing lower or no upfront investment costs, a flatter learning curve, or eliminating other obstacles preventing adoption.

Gain Creators – Trigger Questions

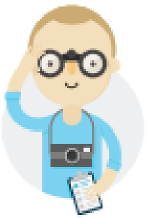
- ... create savings that please your customers? In terms of time, money, and effort.
- ... produce outcomes your customers expect or that exceed their expectations? By offering quality levels, more of something, or less of something.
- ... outperform current value propositions and delight your customers? Regarding specific features, performance, or quality.
- ... make your customers' work or life easier? Via better usability, accessibility, more services, or lower cost of ownership.
- ... create positive social consequences? By making them look good or producing an increase in power or status.
- ... do something specific that customers are looking for? In terms of good design, guarantees, or specific or more features.
- ... fulfill a desire customers dream about? By helping them achieve their aspirations or getting relief from a hardship?
- ... produce positive outcomes matching your customers' success and failure criteria? In terms of better performance or lower cost.
- ... help make adoption easier? Through lower cost, fewer investments, lower risk, better quality, improved performance, or better design.

Homework 3 – Value Proposition

- Fill in your activity in a day using the following template http://bit.ly/2Me0jsQ
- Save file and name the file as activities_studentID.xls
- Upload to the folder in MS Team
  Files / Homework 3 - A day in the life
- Analyze activities and choose an activity of your interest
- Develop The Value Proposition Canvas
- Save the canvas to your homework repository

	A	B	C	D
1	Time	Activity	Pain	Gain
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				

A Day in the Life Worksheet



Dive deep into your (potential) customers' worlds to gain insights about their jobs, pains, and gains. What customers do on a daily basis in their real settings often differs from what they believe they do or what they will tell you in an interview, survey, or focus group.

OBJECTIVE
Understand your customer's world in more detail

OUTCOME
Map of your customer's day

Capture the most important jobs, pains, and gains of the customer you shadowed

Time	Activity (what I see)		Notes (what I think)

Pull: Job Selection

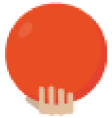
Imagine your customers are chief information officers (CIOs) and you have to understand which jobs matter most to them. Do this exercise to prioritize their jobs or apply it to one of your own customer profiles.

OBJECTIVE
Identify high-value customer jobs that you could focus on

OUTCOME
Ranking of customer jobs from your perspective

Scoring Scale:
• (Low) to • • • • (High)

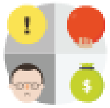
- Does failing the job lead to extreme pains?
- Does failing the job lead to missing out on essential gains?



- Can you feel the pain?
- Can you see the gain?

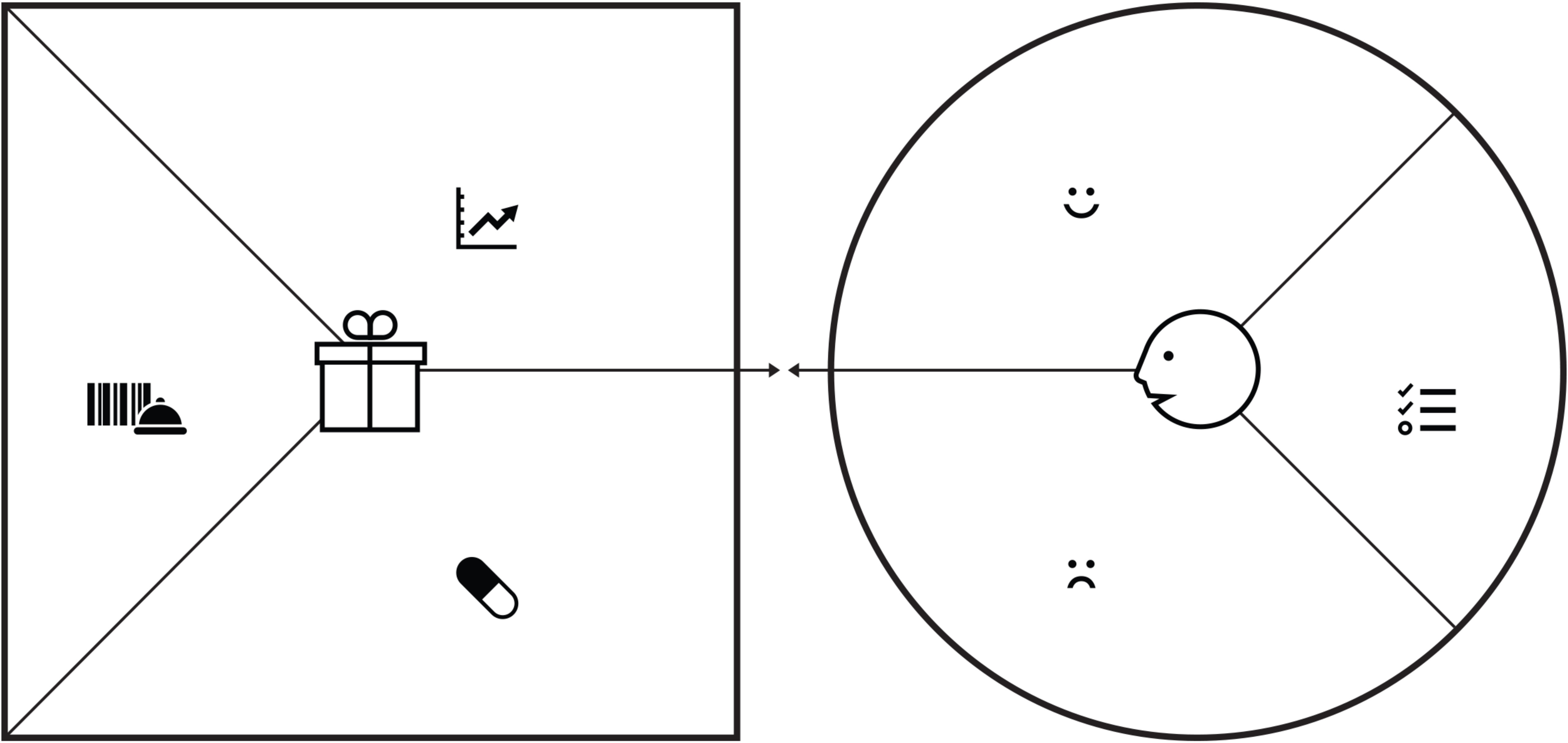
- Are there unresolved pains?
- Are there unrealized gains?

- Are there many with this job, pain, or gain?
- Are there few willing to pay a lot?



Focus on the highest value jobs and related pains and gains.

Jobs	Important	Tangible	Unsatisfied	Lucrative	High-Value jobs
→					
→					
→					
→					
→					
→					
→					
→					



Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 4
Customer Experience Management

Customer Experience

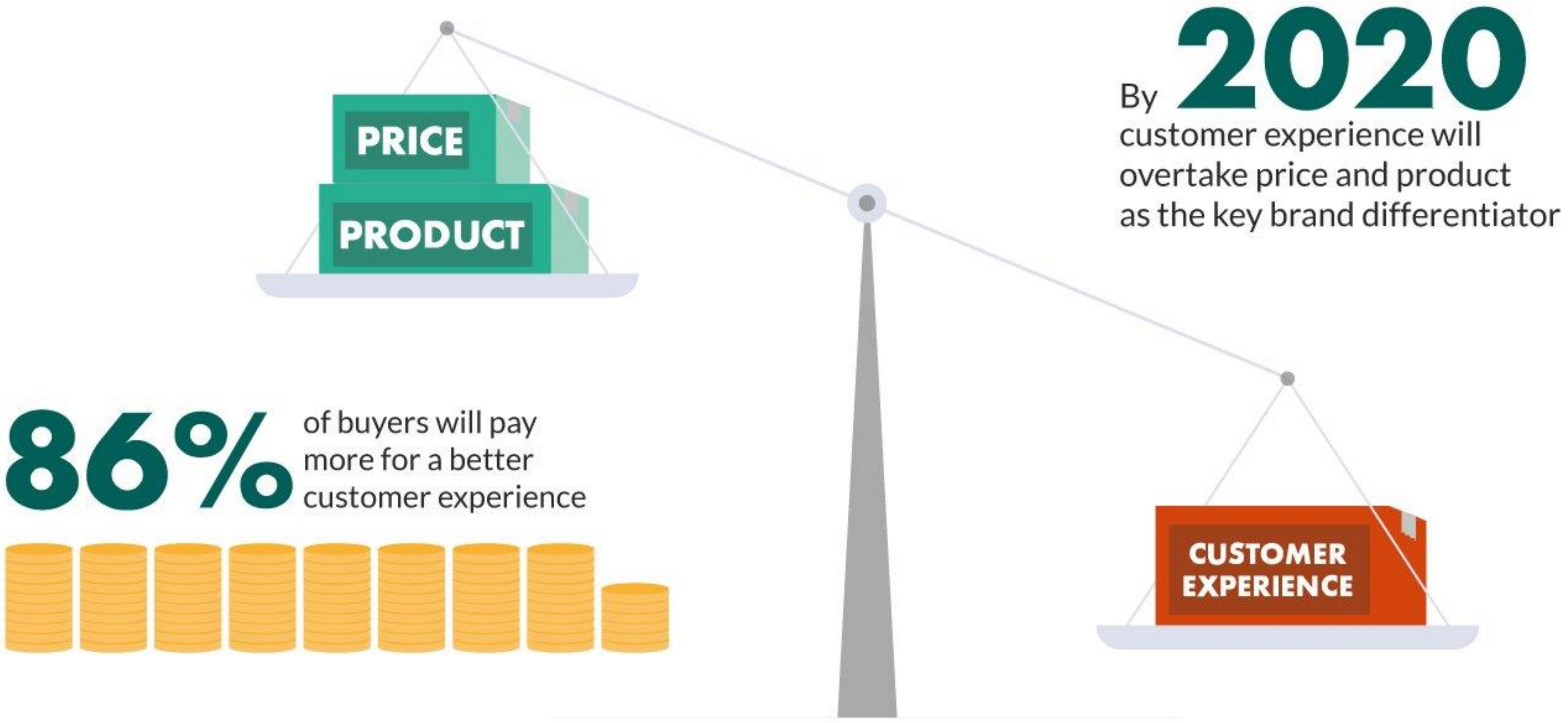
- Customer experience (also known as CX) is defined by the interactions and experiences your customer has with your business throughout the entire customer journey, from first contact to becoming a happy and loyal customer.

- CX is an integral part of Customer Relationship Management (CRM) and the reason why it’s important is because a customer who has a positive experience with a business is more likely to become a repeat and loyal customer.

- In fact, according to a global CX study by Oracle found that 74% of senior executives believe that customer experience impacts the willingness of a customer to be a loyal advocate. If you want your customers to stay loyal, you have to invest in their experience!

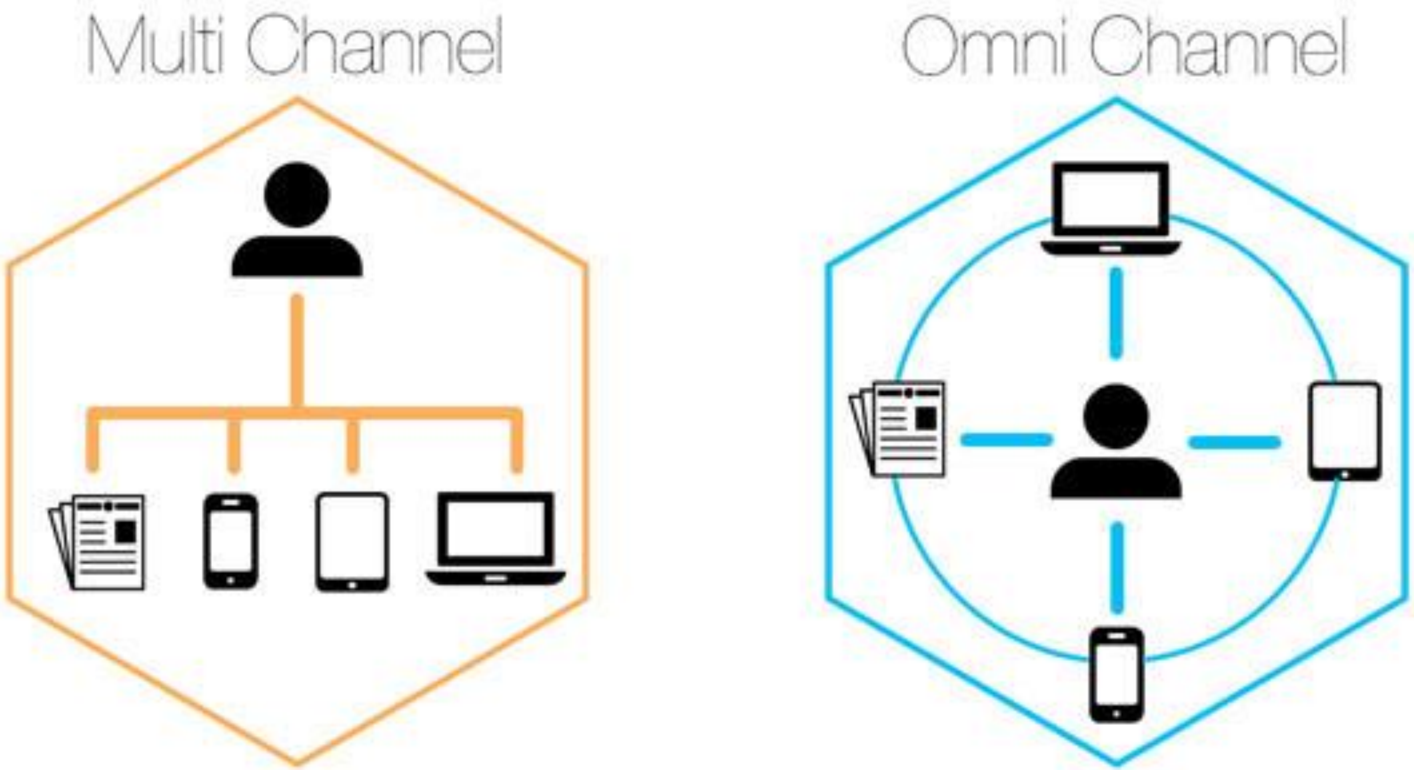
https://www.superoffice.com/blog/customer-experience-strategy/

SURVEY SAYS: CUSTOMERS HIGHLY VALUE GREAT CUSTOMER EXPERIENCES



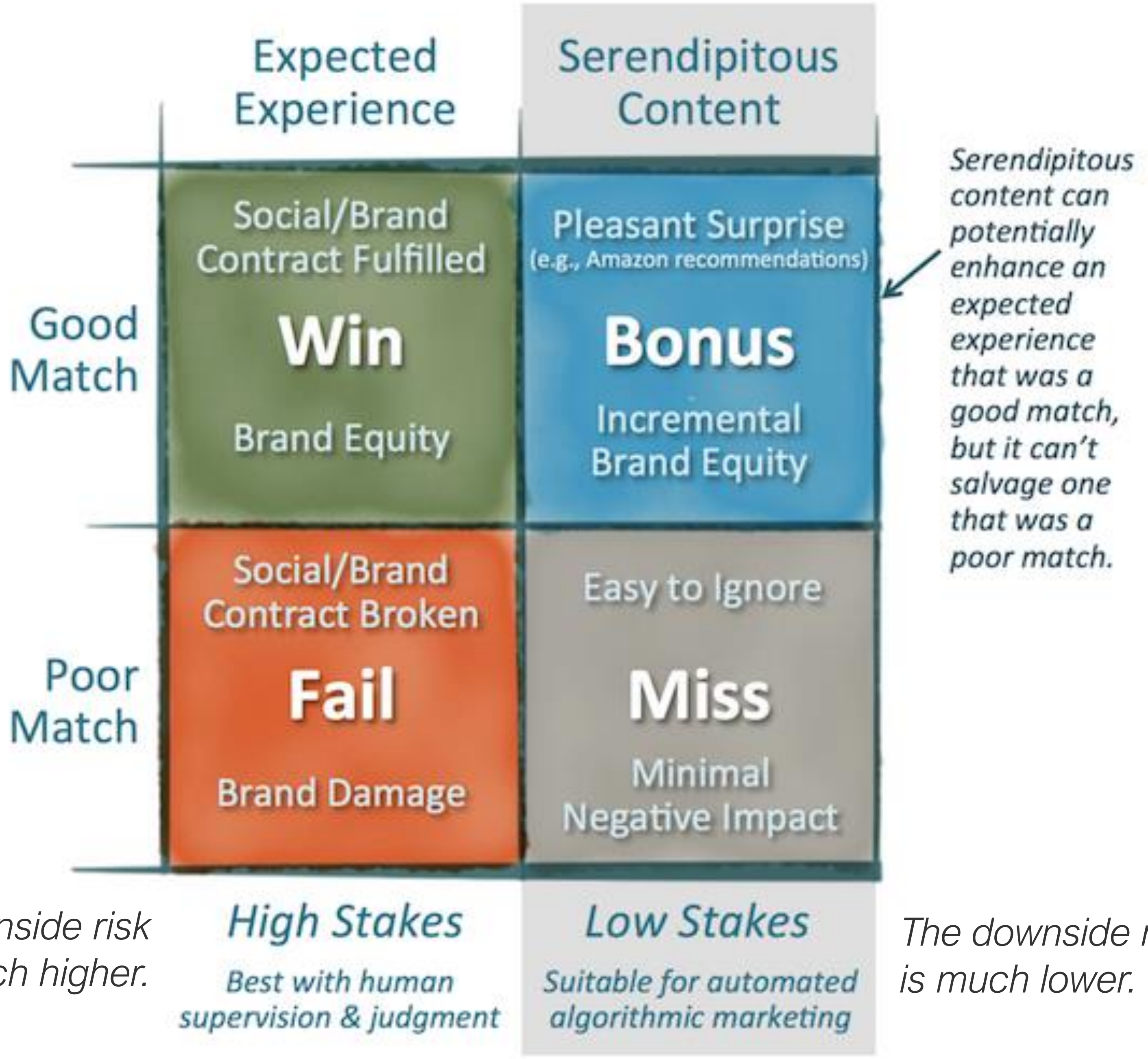
https://www.superoffice.com/blog/customer-experience-statistics/

Omni-channel Customer Experience



Customer Experience Matrix

Robots vs. marketers: the algorithmic marketing matrix

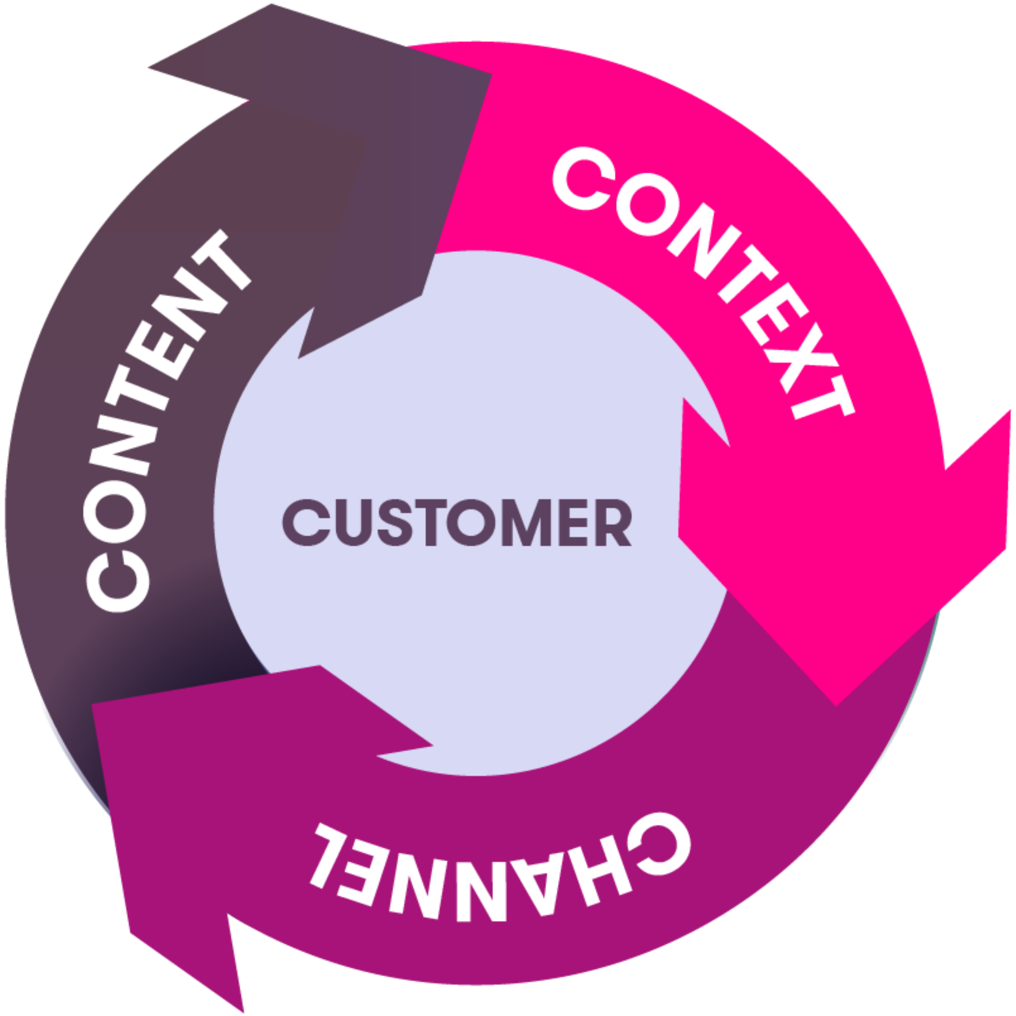


The downside risk is much higher.

The downside risk is much lower.

The Ultimate Business Competitive Advantage: The Frictionless Framework

Frictionless Customer Experience is the alignment of the 4 Cs which describe the mindset of a company that has a customer-centric approach to offer the right content in the right channel with the right context



https://medium.com/swlh/the-ultimate-business-competitive-advantage-the-frictionless-framework-94a6f482116e

Customer

When defining the customer we should think about 3 dimensions:

- Customer Demographic: who is our customer? Some key factors about the demographics: location, age, gender, business or consumer, etc.
- Customer Behavior: what need/problem does the customer have? Think carefully all the need that´s most relevant to your business and that you can solve best.
- Customer Relationship: how will you interact with the customer? Will you be interacting with the customer personally or it will be more self-service with a software platform?

Customer

- Choosing the right customer to work with is fundamental to gain an advantage to your competition. A very good example is the case of Nintendo Wii. Back in 2006 in the video game console market, the leaders are Sony (PlayStation) and Microsoft (Xbox). Both companies were in a fierce competition for getting the best graphics and processing power. However, Nintendo at that time couldn´t afford to compete in technical specifications. So they thought hard about how to innovate in other areas. After doing some research they realized that both Sony and Microsoft were mostly targeting hardcore gamers.

- With that insight what Nintendo did was design a new game console for the casual gamers, who don´t care so much about the game graphics and processing power. In exchange, Nintendo would offer a new experience that would entertain better for the customer target. As result, when Nintendo launched Wii it became an instant hit exceeding the sales of PlayStation and Xbox. In December 2009, the console broke the sales record for a single month in the United States

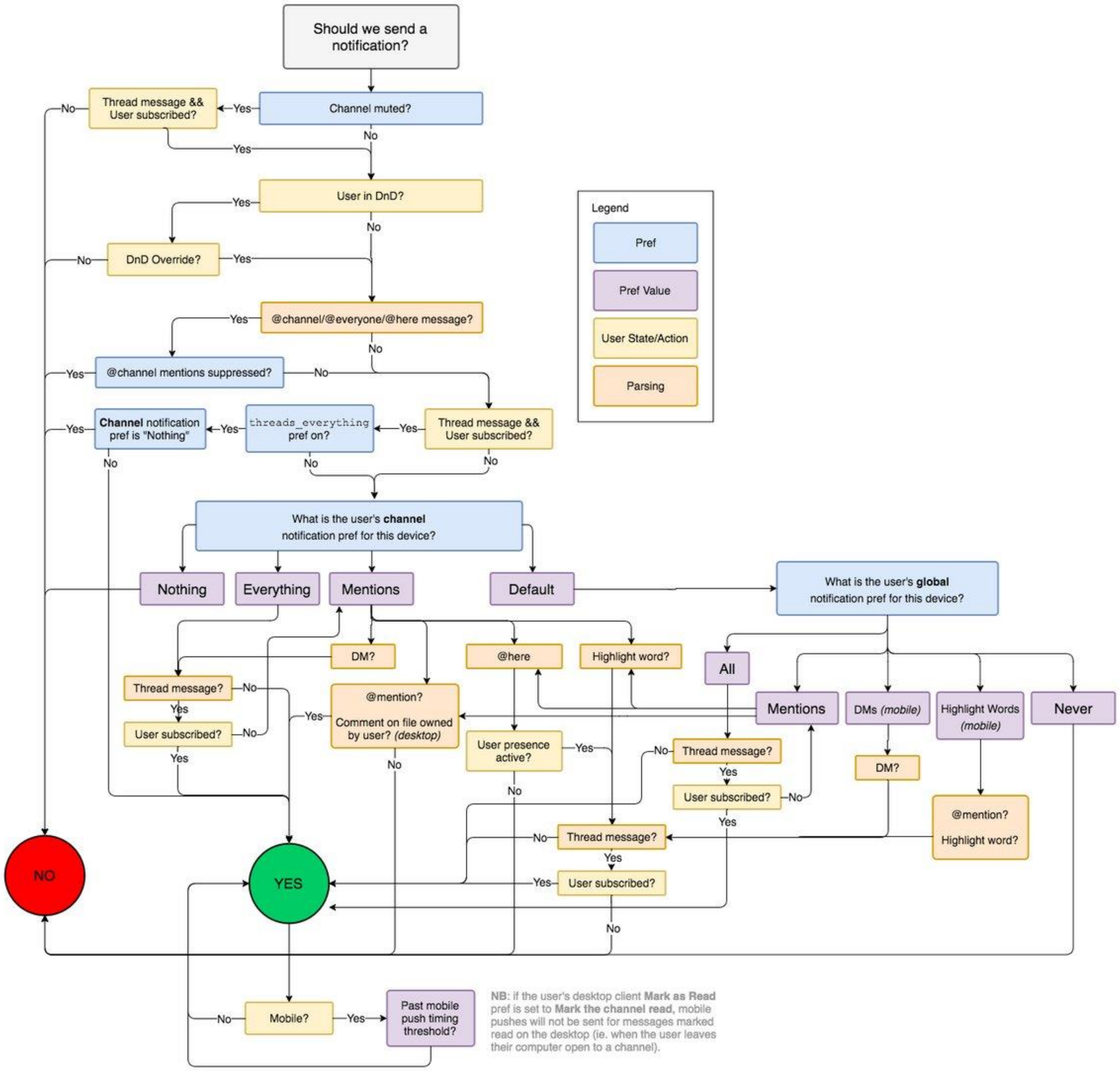
Context

- Once you have considered who is your customer, the next step is to understand the context. For me, context is all the surrounding events that the customer is living in each moment. It can be very dynamic and be changing all the time. Is the customer working, eating, talking with somebody, reading, playing a game, etc?

- Bill Buxton the principal researcher at Microsoft Research introduced a very interesting concept called Place-ona, adapting from persona and basically describes the limitation of interaction depending on the location of the customer at each moment.
- For example: if a user is driving a car, his hands and eyes are limited so the only way to interact with the user would be through his mouth and ears which would be the ideal scenario for a voice-based system.
- However, if the user is in a library, his mouth would be restricted but his hands and eyes are free. So, in this case, a visual interface based interaction would make more sense.

Context

- Understanding the customer´s context is critical to offer frictionless experience. To illustrate this case we can see how Slack has designed its notification system. Slack is a very popular messaging platform for internal team communication used by millions of companies. They are known for their excellent design and seamless user experience. Every product decision they make must be carefully studied to truly help customers become more efficient in their team communication. Here below is the workflow of the feature “message notification”



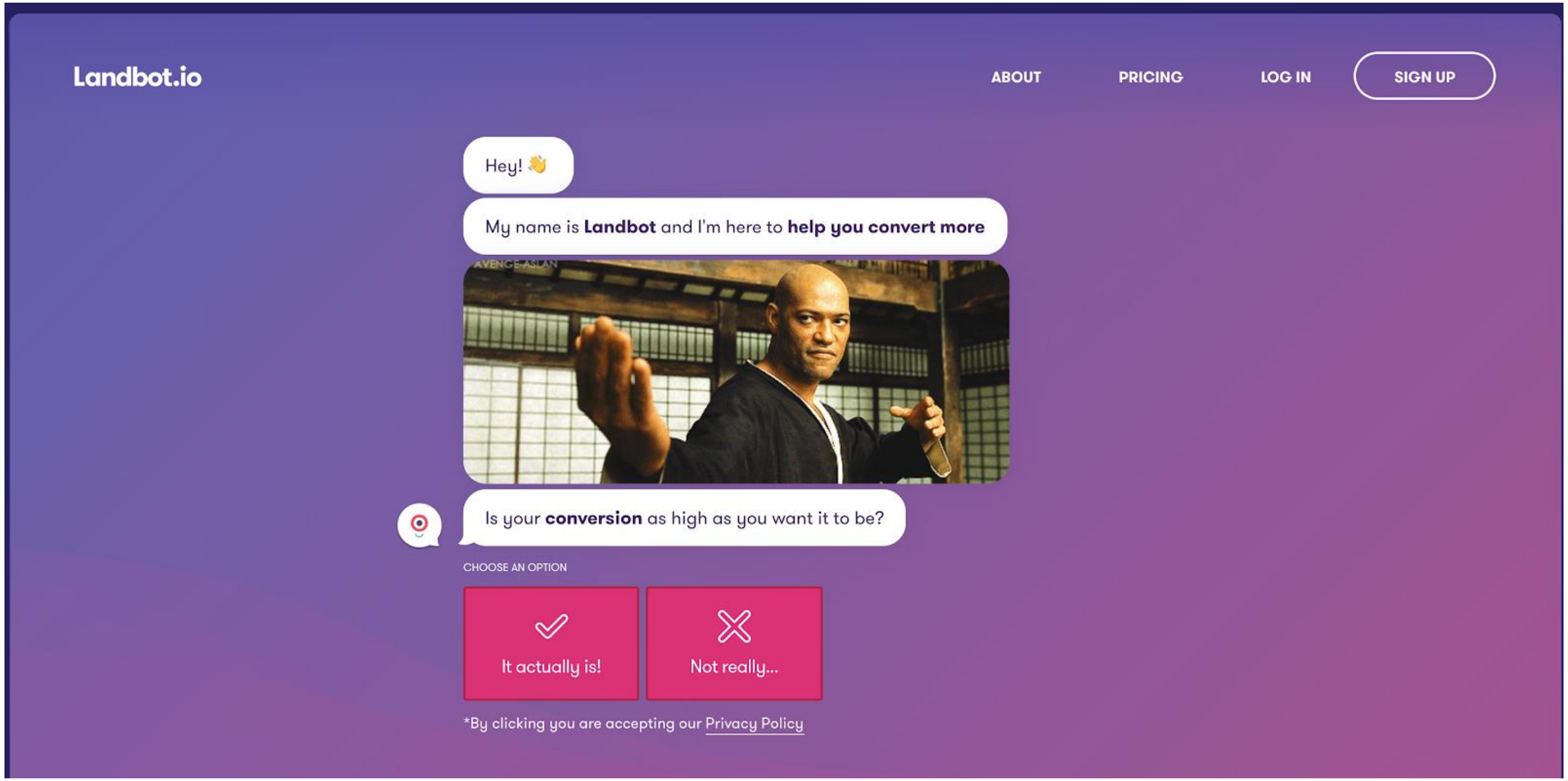
Content

- After visualizing the context of our customer, we should start working on the right content*. The mission of every business is to provide a solution that can solve a problem and satisfy customer's need.

- Content can be anything material or digital goods that you can offer to customers as a solution to their problem: a software application or consulting service both can be considered as a content.

Content

Most website on the market today offer the same content to every customer which is why the average conversion rate is around 2,35%. To solve this problem, the industry came up with the concept of conversational websites. Unlike a static web, with a conversational site (here is an example), we can offer personalized information in each customer interaction. As result companies can get between 50%-100% increase in conversion rate and customer engagement.



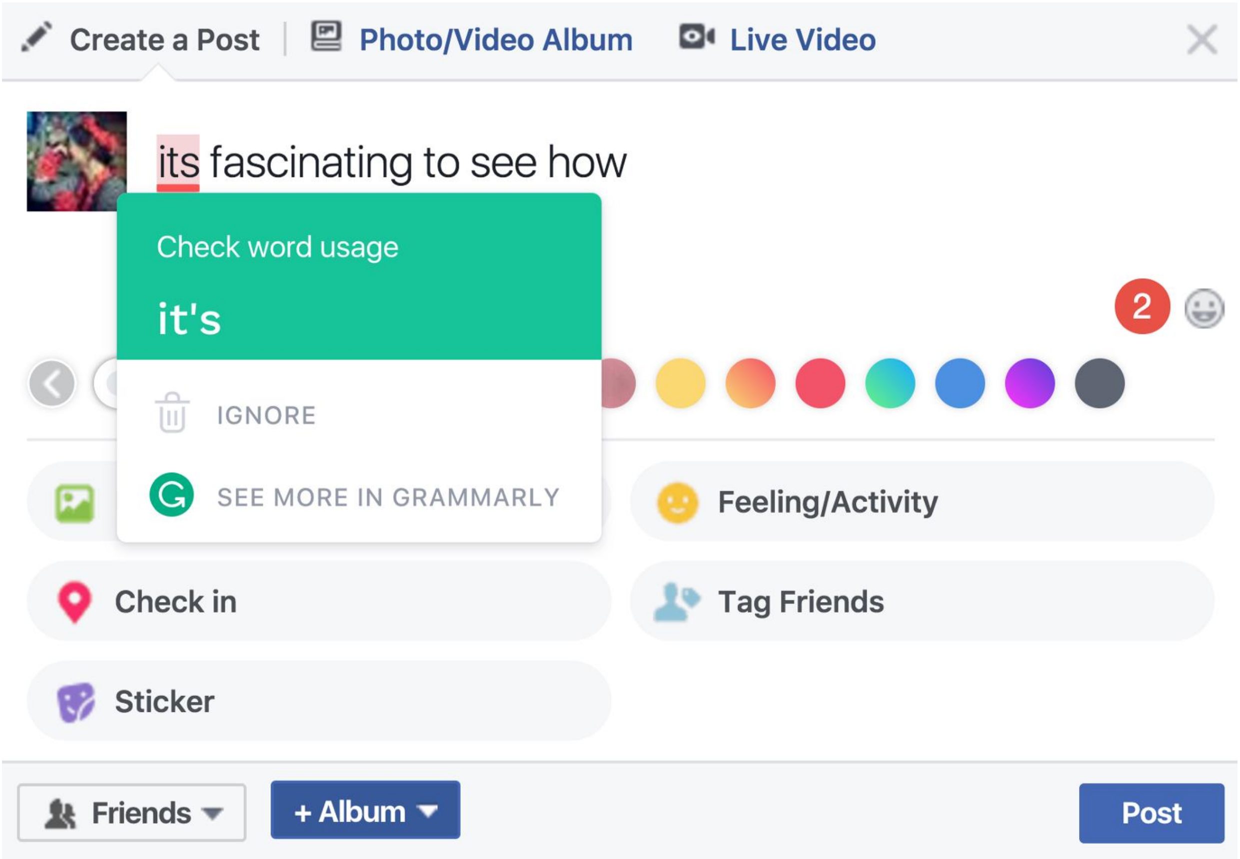
Channel

- What channel should you use? To choose the best channel we need to keep in mind some important factors:

- **Cost**: what is the cost involved to acquire and serve customers in the channel you choose? You should have a clear idea of the CAC and the LTV of the customer acquired from each channel.
- **Maturity**: Most channels tend to reach a saturation point from which it will be tough for new players coming in. Example: it will be quite difficult to start a new mobile app company right now due to the market share of some big corps. If you look at the ranking of top ten apps, most of them are from companies like Facebook, Google, Snapchat.
- **Control**: you should look for the flexibility and the scalability you can have with a particular channel. Does it offer tools for 3rd party developers like APIs; analyze possible risks of being cannibalized by native solutions from the channel. You don't want experiment with what happened to Meerkat when Twitter banned it from the social network.

Channel

- A fascinating case study is Grammarly, the best grammar checking tool of the planet. Hiten Shah wrote an in-deep review of how Grammarly grow into millions of users. One key point in their strategy was that they had designed the product to be where their customers are. Grammarly builds plugins for Microsoft Office and later as Chrome extensions so people can use it where they need the tool most: when writing a post, filling job forms, editing text documents, etc.



Customer Journey

- Customer journeys describe how to deliver the right experience to each individual customer.
- Journeys provide step-by-step descriptions of a customer's path as he or she interacts with the organization.
- They visualize how a customer might engage through a range of channels from the web to a retail environment.
- Customer journeys are well-known customer experiences of service design tools.
- A customer journey creates value when it helps an organization design how it should engage customers to create better performing services.

Customer Journey

Channels	Aware	Interest	Purchase	Re-purchase
Google		*Googled to find information. Found several reviews on the internet.*		
Website		*Read several reviews and became interested to buy.*	*Spent way too long to complete the purchase.*	*Could not find what to buy.*
Facebook	*Saw a friend posting picture of the product. Kept seeing the posts.*			
E-mail				*Received the next purchase coupon.*
Store		*Visited the store to try out the product. Searched for the discount at the store.*		
Call Center			*Complaint about the late delivery.*	

Fixing Customer Journey with Data Analytics

Channels	Aware	Interest	Purchase	Re-purchase
Google		*Googled to find information. Found several reviews on the internet.*	Conversion Analysis	Product Recommendation
Website		*Read several reviews and became interested to buy.*	*Spent way too long to complete the purchase.*	*Could not find what to buy.*
Facebook	*Saw a friend posting picture of the product. Kept seeing the posts.*			
E-mail				*Received the next purchase coupon.*
Store		*Visited the store to try out the product. Searched for the discount at the store.*		
Call Center			*Complaint about the late delivery.*	

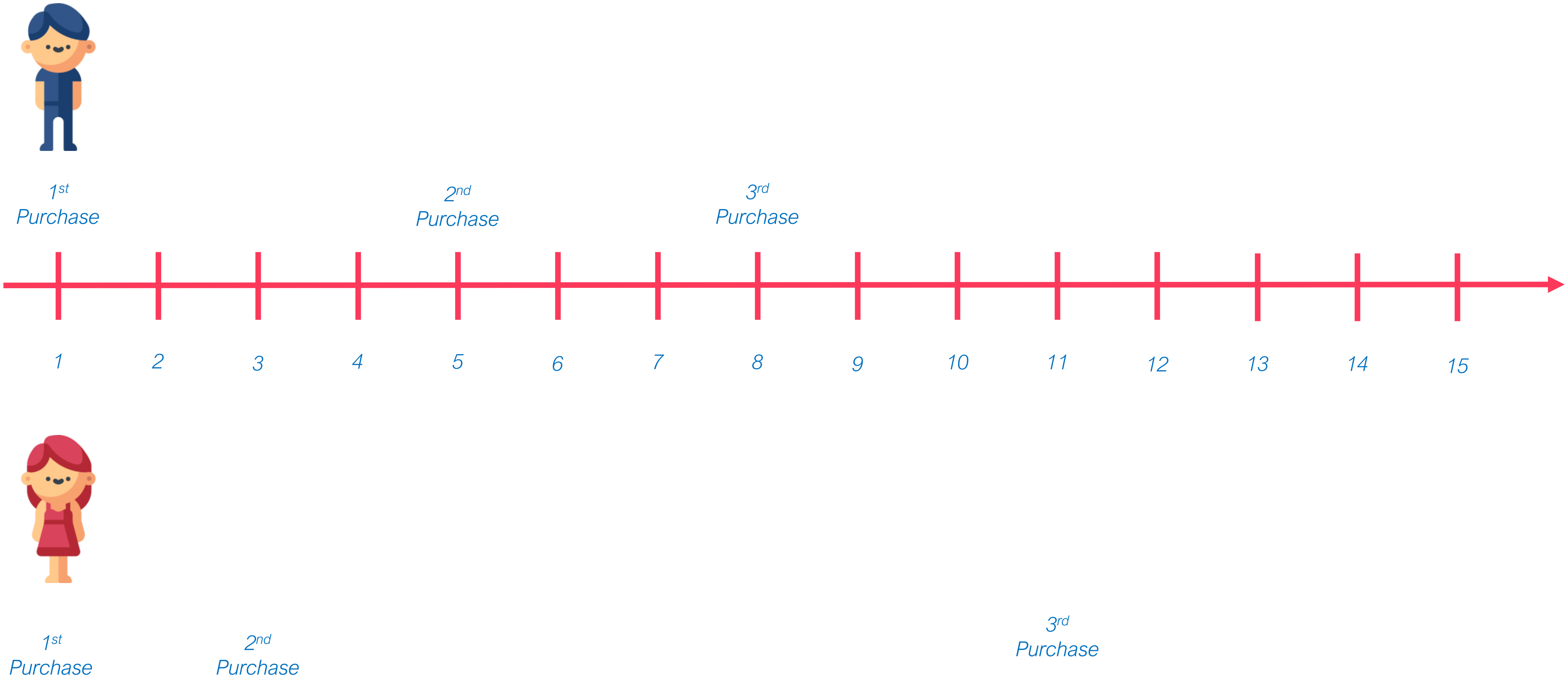
Route Optimization

Collecting Data throughout Customer Journey

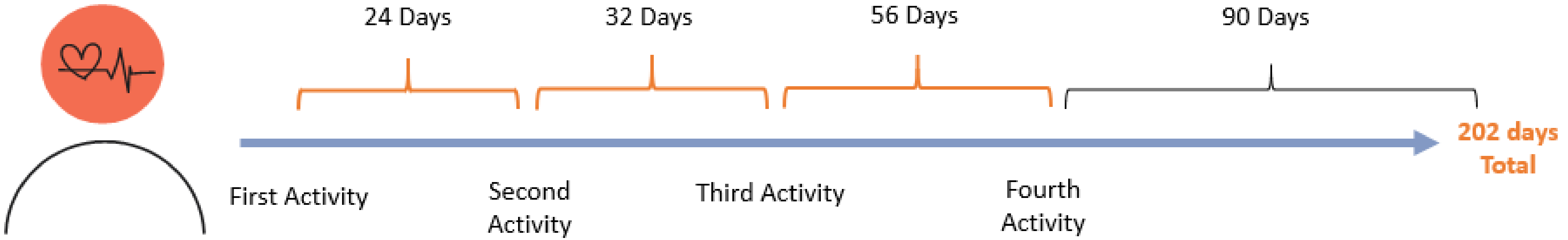
Channels	Aware	Interest	Purchase	Re-purchase
Google		GA *Googled to find information. Found several reviews on the internet.*	GA Payment System Membership	
Website		*Read several reviews and became interested to buy.*	*Spent way too long to complete the purchase.*	*Could not find what to buy.*
Facebook	*Saw a friend posting picture of the product. Kept seeing the posts.*			
E-mail				*Received the next purchase coupon.* EDM
Store		*Visited the store to try out the product. Searched for the discount at the store.*		
Call Center			*Complaint about the late delivery.* Call Center	

Customer Journey Analysis – Customer Lifeline

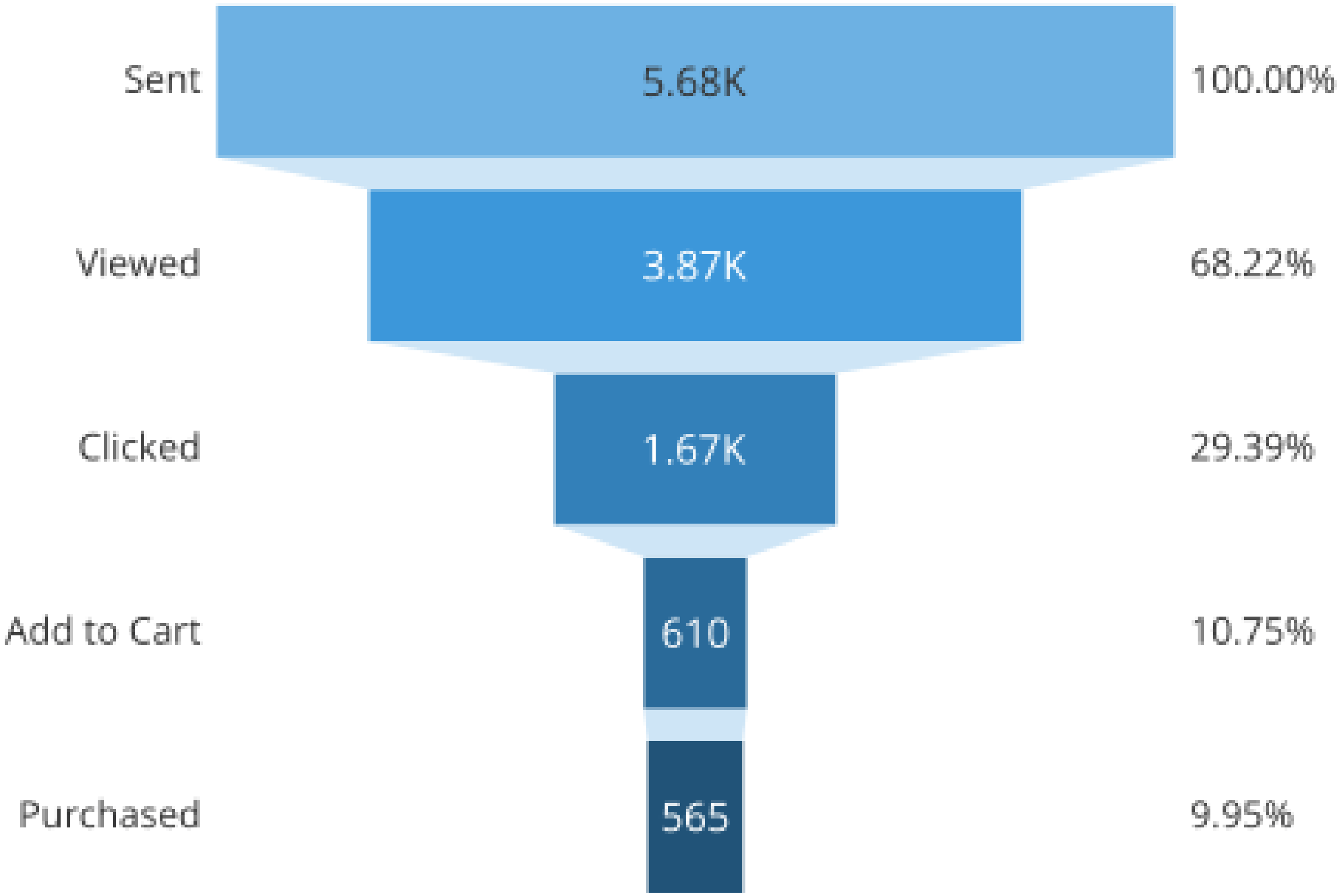
Customer Journey Analysis – Customer Lifeline Cohort



Customer Journey Analysis – Time to Event



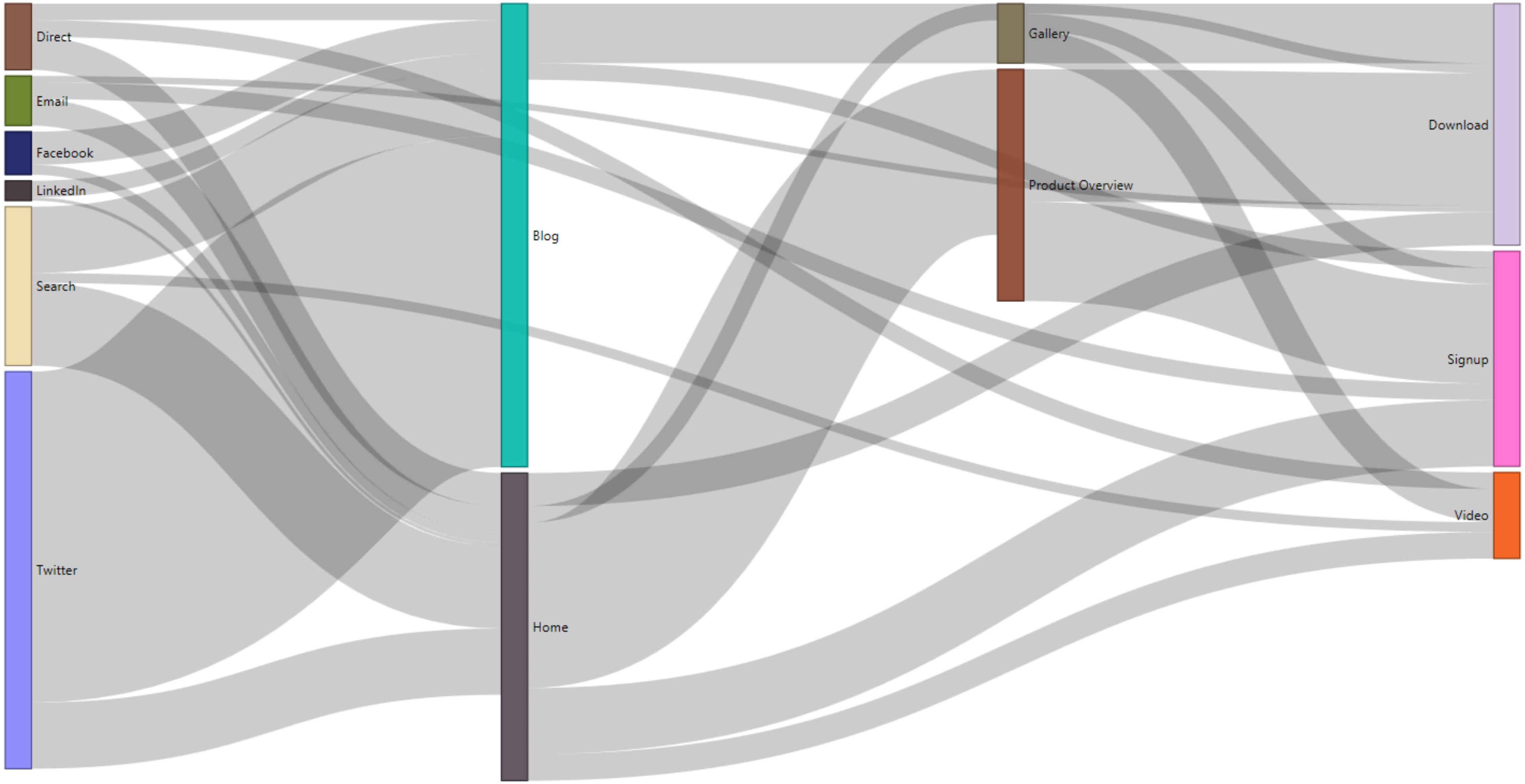
Customer Journey Analysis - Aggregated Journeys using Customer Funnel



Customer Journey Analysis - Aggregated Journeys using Cohort Analysis

Acquisition Date	Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Jan 25	1,098	100%	33.9%	23.5%	18.7%	15.9%	16.3%	14.2%	14.5%	13.3%	13.0%	12.1%
Jan 26	1,358	100%	31.1%	18.6%	14.3%	16.0%	14.9%	13.2%	12.9%	14.5%	11.3%	
Jan 27	1,257	100%	27.2%	19.6%	14.5%	12.9%	13.4%	13.0%	10.8%	11.4%		
Jan 28	1,587	100%	26.6%	17.9%	14.6%	14.8%	14.9%	13.7%	11.9%			
Jan 29	1,758	100%	26.2%	20.4%	16.9%	14.3%	12.7%	12.5%				
Jan 30	1,624	100%	26.4%	18.1%	13.7%	15.4%	11.8%					
Jan 31	1,541	100%	23.9%	19.6%	15.0%	14.8%						
Feb 01	868	100%	24.7%	16.9%	15.8%							
Feb 02	1,143	100%	25.8%	18.5%								
Feb 03	1,253	100%	24.1%									
All Users	**13,487**	**100%**	**27.0%**	**19.2%**	**15.4%**	**14.9%**	**14.0%**	**13.3%**	**12.5%**	**13.1%**	**12.2%**	**12.1%**

Customer Journey Analysis - Aggregated Journeys using Sankey



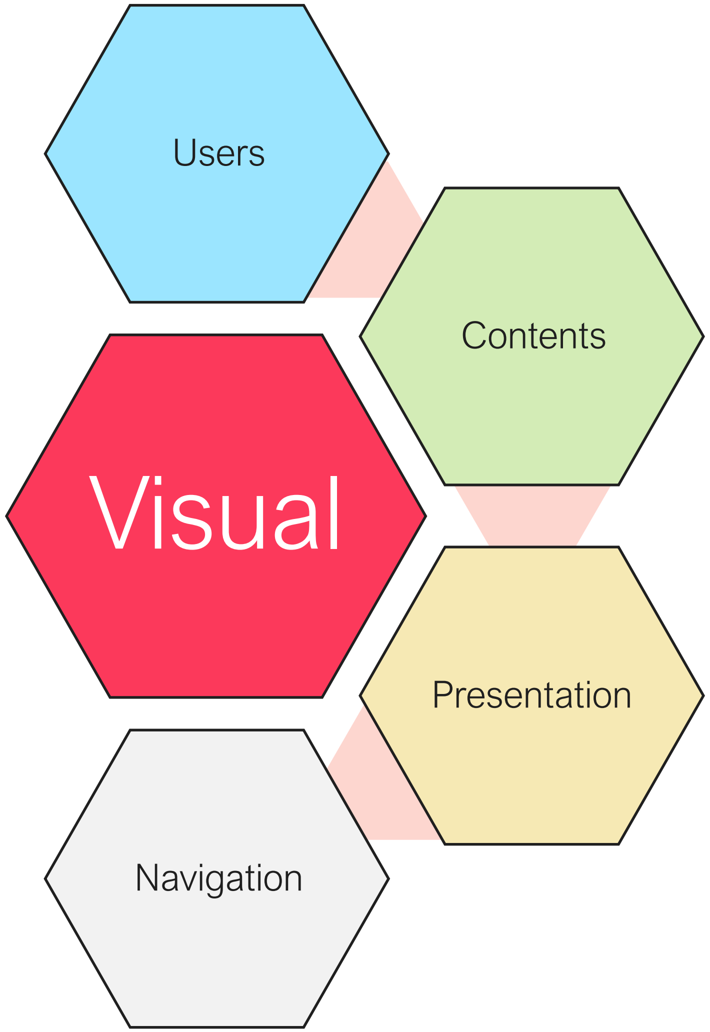
Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 5
Customer Insights with Multidimensional Analysis

Framework of Visual Analytics

Framework for Visual Design



Users:	know your users, understand what they need
Contents:	choose the right information, understand your business
Presentation:	use the appropriate charts and colors
Navigation:	place information into the right position

USERS

User Empathy Map

SEEING

What do users need to see? How often?

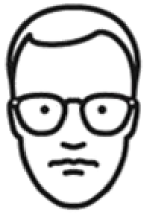
THINKING

What are users' mental tasks? Calculate growth? Making Comparison?

DOING

How do users use this visual? Make a decision? Monitor progress?

User Empathy Map Template



Name:
Position:
Main Screen:

DOING

decide monitor control alert plan order execute

THINKING

calculate compare classify forecast identify understand

SEEING

measures time locations dimensions

What should users see?

SEEING

What do users need to see? How often?

THINKING

What are users' mental tasks? Calculate growth? Making Comparison?

DOING

How do users use this visual? Make a decision? Monitor progress?

Problems

- Key results
- Lag indicators
- Desired outcomes

Reasons

- KPI drivers
- Lead indicators
- Breakdown key problems into sub-problems

Actions

- Show data at the level where you can take actions
- Highlight business opportunities
- Identify a root cause of a problem

Example

SEEING

What do users need to see? How often?

>

THINKING

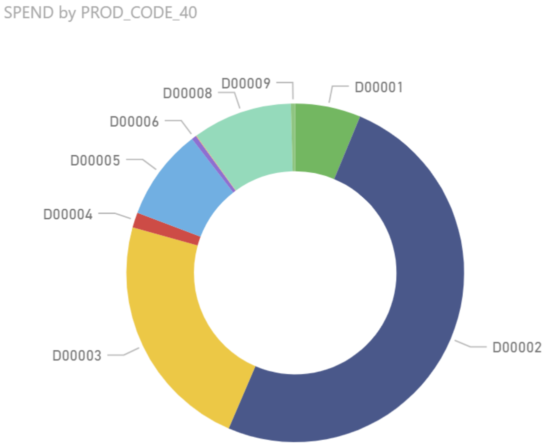
What are users' mental tasks? Calculate growth? Making Comparison?

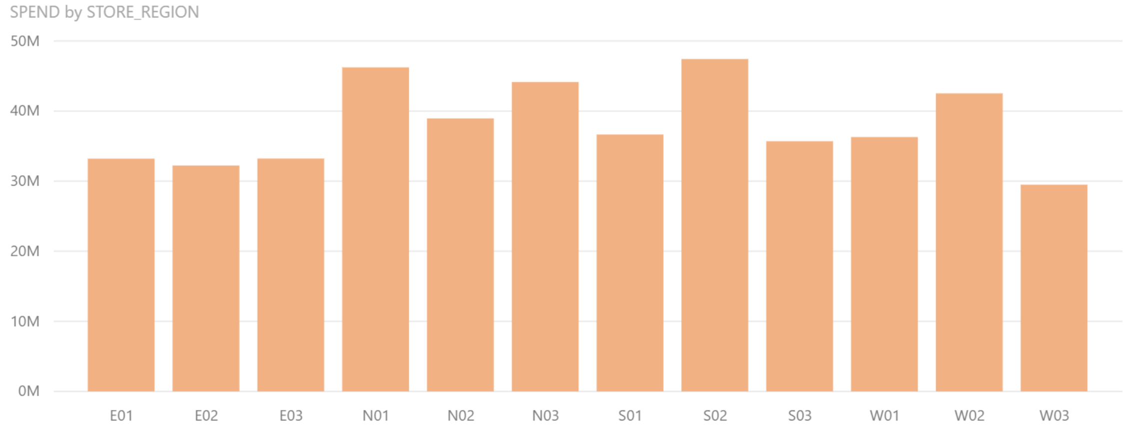
>

DOING

How do users use this visual? Make a decision? Monitor progress?

I need to decide whether I should offer discount





Top 10 SKUs

PROD_CODE	SPEND
PRD0904358	7,894,739.16
PRD0900121	4,767,591.22
PRD0903052	3,860,043.21
PRD0901265	3,141,754.31
PRD0904250	1,913,687.01
PRD0904976	1,886,013.69
PRD0903377	1,788,488.00
PRD0901923	1,741,934.97
PRD0900173	1,697,645.24

Bottom 10 SKUs

PROD_CODE	SPEND
PRD0902714	26.08
PRD0902990	63.25
PRD0900750	79.44
PRD0901273	86.96
PRD0900480	99.95
PRD0902373	106.17
PRD0901602	124.82
PRD0901686	130.27
PRD0903919	136.65

After

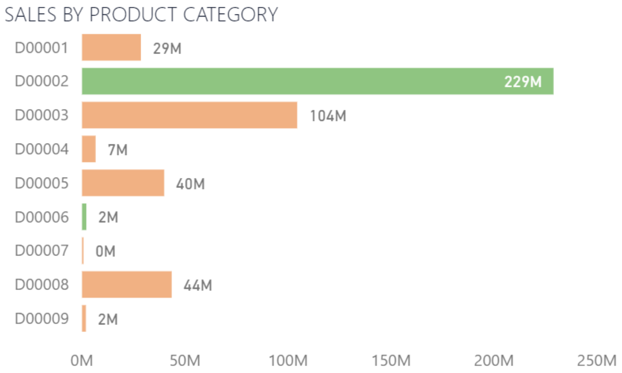
456.10M

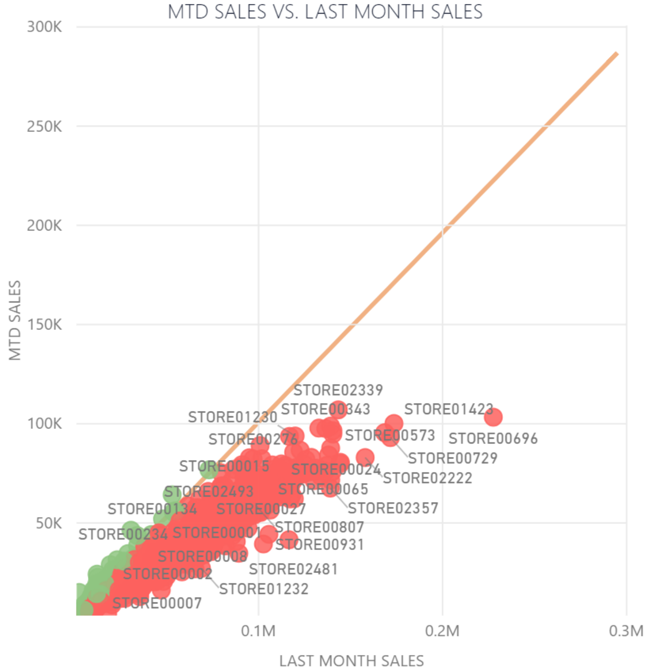
SALES

- 7.59 %

674.27M / 800M

SALES RUN RATE

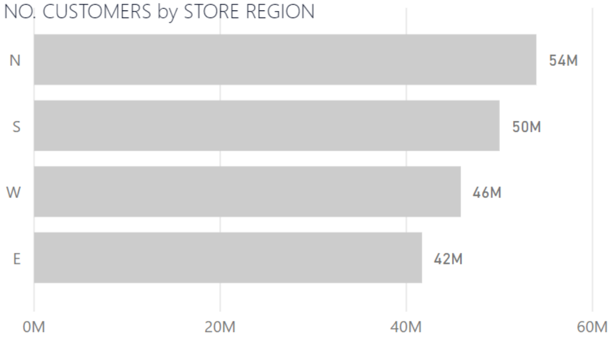




473.04K

NO. CUSTOMERS

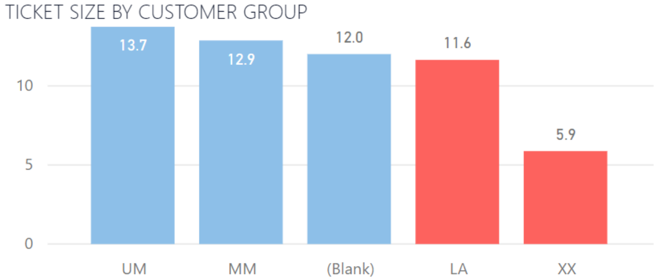
4.38 %



12.59

TICKET SIZE

- 12.93 %



STORE_FORMAT	TICKET SIZE	SALES
LS	13.86	289,691,869.52
MS	10.71	96,299,382.33
SS	8.26	30,510,974.79
XLS	14.94	39,595,757.01
Total	**12.59**	**456,097,983.65**

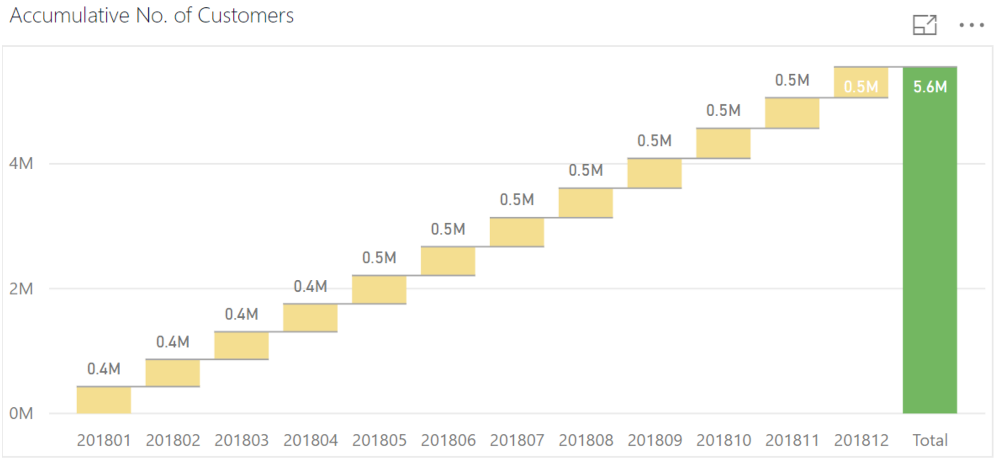
Example

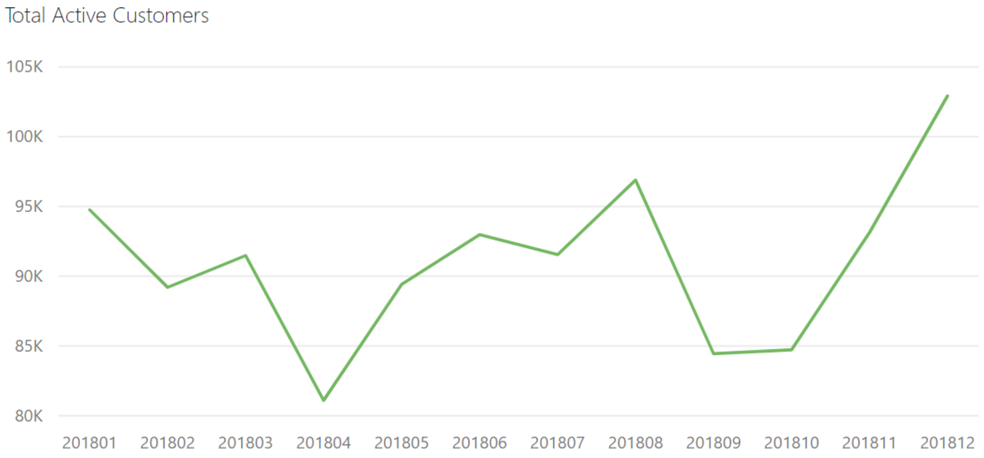
SEEING
What do users need to see? How often?

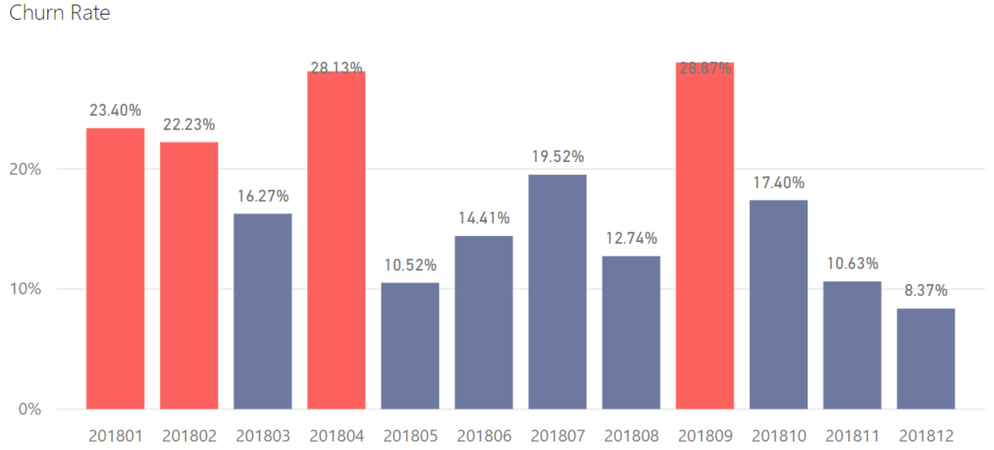
THINKING
What are users' mental tasks? Calculate growth? Making Comparison?

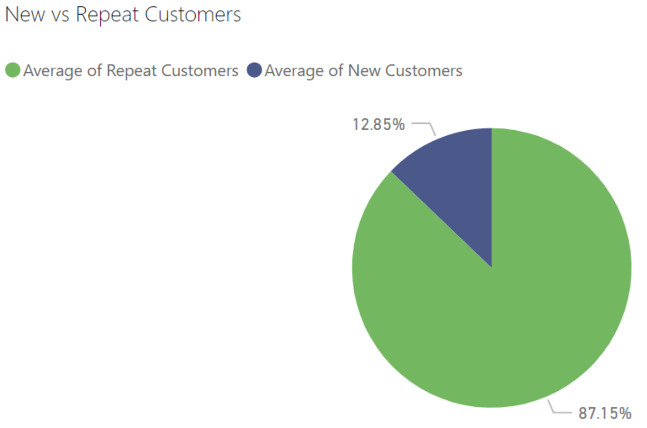
DOING
How do users use this visual? Make a decision? Monitor progress?

I need to monitor whether I will hit the target.
My KPI is no. of users.



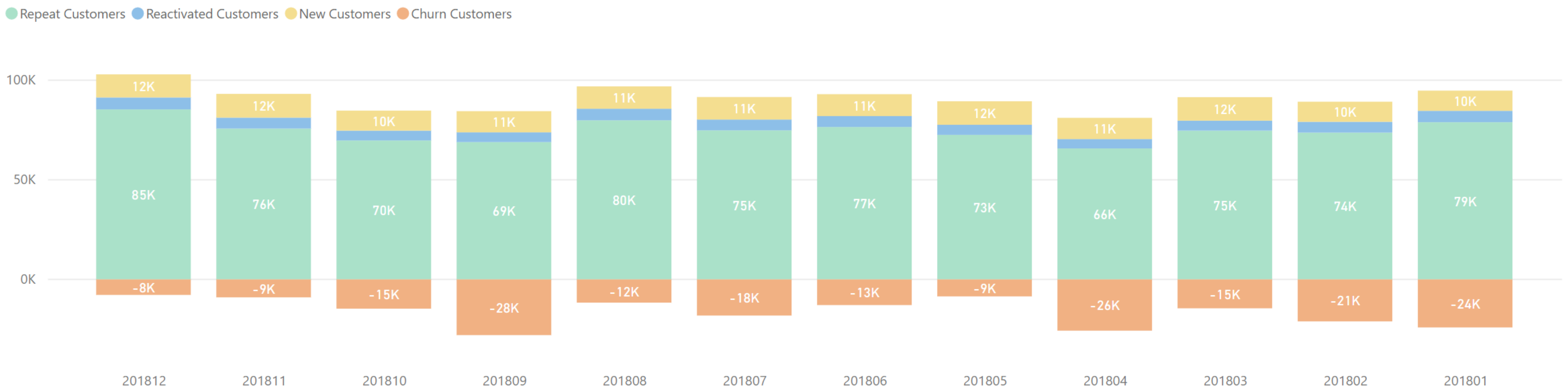


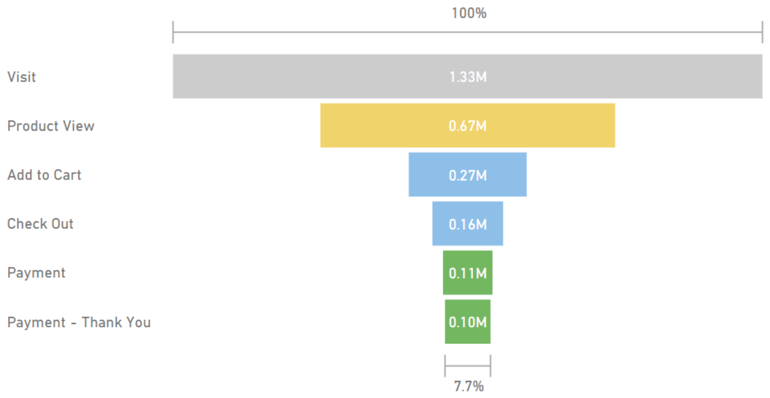


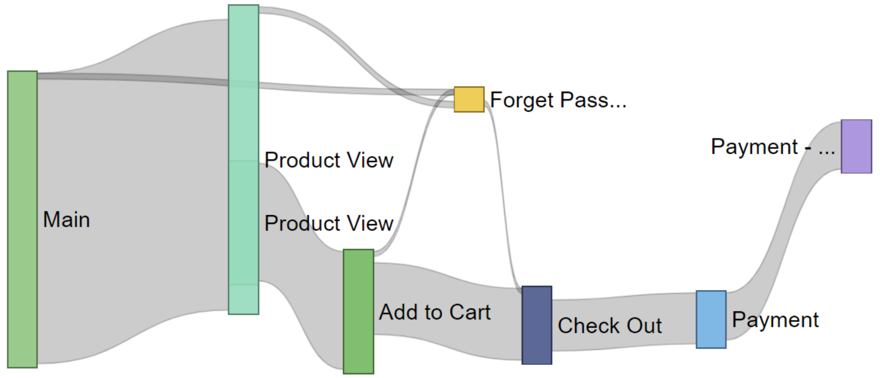


After

91.05K	17.71%	19.68%	11K
Average of Total Active Customers	Average of Churn Rate	% Active Customers	Average of New Customers

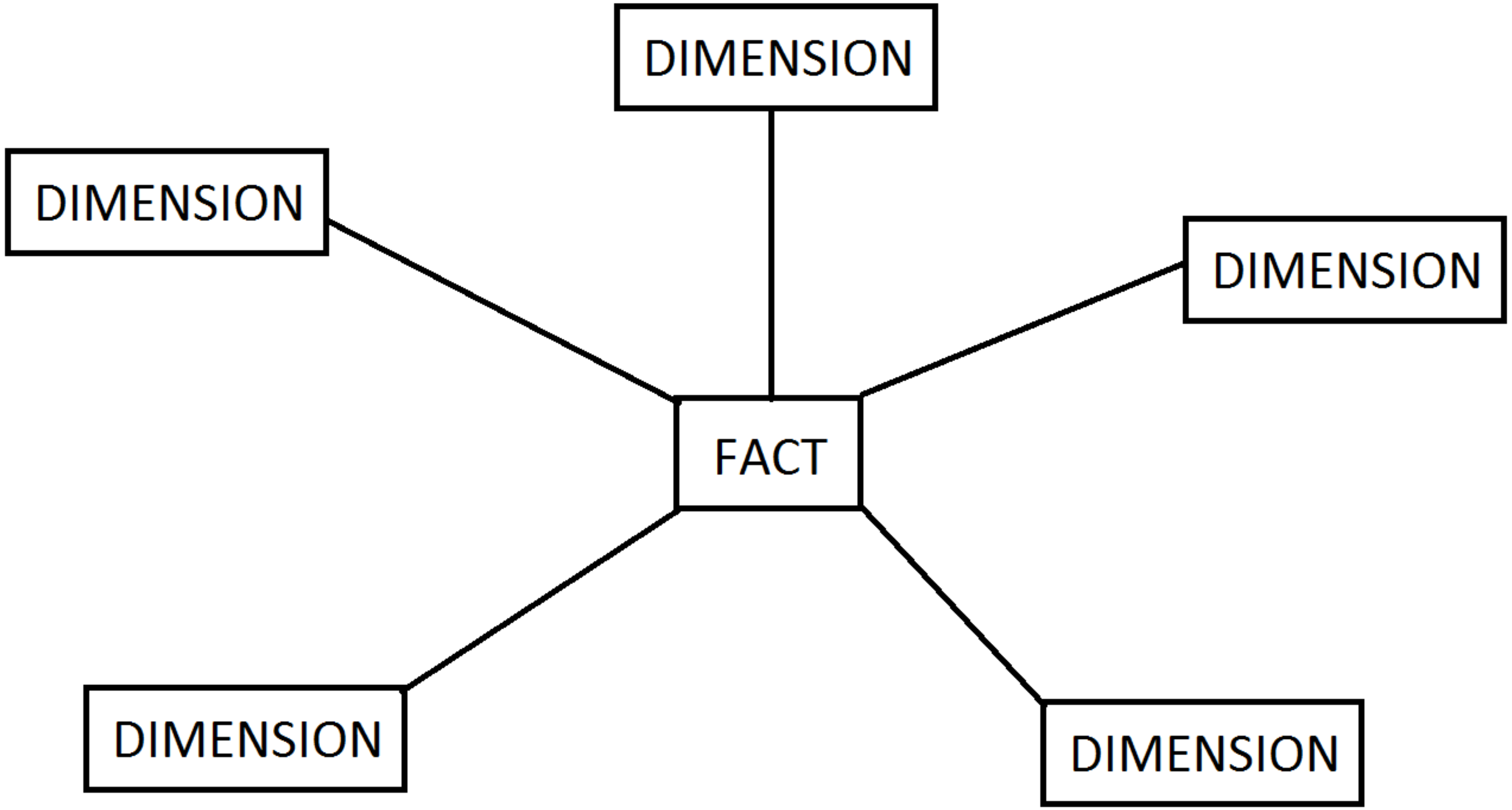






CONTENTS

Star Schema



A Star schema is a schema in which a fact is connected to multiple dimensions and dimension table doesn't have any parent table.

Customer-Related Schema: Sales Results

GOALS	METRICS	DIMENSIONS	
New Customer Sales ($) Sales Growth (%) Sales Order ($)	Avg. Sales per Order ($) Avg. Units per Order (#) Credit Balance ($) Credit Limit ($) Customers (#) Lost Customer Count (#) New Customer Count (#) New Product Sales ($) Sales Order Count (#) Units Ordered (#)	Billing Customer – Industry Group – Industry – Category – Customer Name Customer Location – Region – State/Province – County – Postal Code/Zip Code Fiscal Week – Fiscal Year – Quarter – Month – Week Market Segment – Market Segment – Micro-Segment	Product SKU – Product Line – Brand – SKU Sales Channel Partners – Sales Channel Type – Sales Partner Sales Organization – Sales Region – Sales Territory – Org. Code Ship-To Location – Region – State/Province – County – City – Postal Code/Zip Code

Customer-Related Schema: Customer/Product Profitability

GOALS	METRICS	DIMENSIONS	
Average Customer Profit ($) Lifetime Profit ($) Net Profit ($)	Cost ($) Customer Acquisition Cost ($) Customer Retention Cost ($) Customers (#) Discount ($) Gross Profit ($/%) Net Sales ($) Sales Revenue ($) Units Sold (#)	Billing Customer Industry Group Industry Category Customer Name Customer Location Region State/Province County Postal Code/Zip Code Fiscal Month Year Quarter Month	Market Segment Market Segment Micro-Segment Organization Division Department Org. Code Product SKU Product Line Brand SKU Sales Channel / Partners Sales Channel Type Sales Partner

Customer-Related Schema: Sales Tactics

GOALS	METRICS	DIMENSIONS

Average Selling Price ($)
Direct Cost ($)
Discount (%)
Sales Calls (#)
	Avg. Sales Hrs/Inquiry (#)
	Close Days (#)
	Cost per Order ($)
	Customers (#)
	Discount (%)
	Inactive Customers (#)
	Inquiries ($)
	Inquiry Count (#)
	Inquiry S/O Conversion (%)
	Lost Business Count (#)
	Net Price ($)
	Quoted ($)
	Rep T&E ($)
	Sales Orders ($)
	Sales Order Count (#)
	Sales Prospect Rating Score
	Sales Rep Days (#)
	Units Quoted (#)
		Billing Customer
		  Industry Group
		  Industry
		  Category
		  Customer Name
		Credit Limit Range
		  Range
		Customer Location
		  Region
		  State/Province
		  County
		  Postal Code/Zip Code
		Fiscal Week
		  Fiscal Year
		  Quarter
		  Month
		  Week
		Market Segment
		  Market Segment
		  Micro-Segment
		Product Brand
		  Product Line
		  Brand
		Sales Organization
		  Sales Region
		  Sales Territory
		  Org. Code
		Sales Time Priority Rating
		  Priority Rating

Customer-Related Schema: Sales Pipeline

GOALS	METRICS	DIMENSIONS	
Pipeline Ratio (%)	Active Customers (#)	Billing Customer	Market Segment
Pipeline Revenue ($)	Avg. Sales per Order ($)	Industry Group	Market Segment
Sales Order Conversion (%)	Cancelled Order Count (#)	Industry	Micro-Segment
	Inactive Customers (#)	Category	Sales Channel / Partners
	Inquiries ($)	Customer Name	Sales Channel Type
	Inquiry Count (#)	Contracted Pay't Time Range	Sales Partner
	Inquiry/Quote Lead Days (#)	Fiscal Week	Sales Organization
	Lost Business Count (#)	Fiscal Year	Sales Region
	New/Lost Customer Ratio (%)	Quarter	Sales Territory
	New Customer Count (#)	Month	Org. Code
	S/O Quotes (#)	Week	Ship-To Location
	Sales Order ($)	Inquiry – S/O Status	Region
	Sales Order Count (#)	Inquiry S/O Status	State/Province
		Mfg. Product Component	County
		Product Line	City
		SKU	Postal Code/Zip Code
		Component	

Customer-Related Schema: Sales Plan Variance

GOALS	METRICS	DIMENSIONS

Sales Order ($)
Sales Plan ($/%)
	Avg. Sales per Order ($)
	New Customer Sales ($)
	New Product Sales ($)
	Sales Growth (%)
	Units Ordered (#)
	Units Sold (#)
		Fiscal Month
		    Year
		    Quarter
		    Month
		Forecast Scenario (Plan/Actual/Forecast)
		    Scenario
		Market Segment
		    Market Segment
		    Micro-Segment
		Product Line
		    Product Line
		Sales Channel Partners
		    Sales Channel Type
		    Sales Partner
		Sales Organization
		    Sales Region
		    Sales Territory
		    Org. Code

Customer-Related Schema: On-Time Delivery

GOALS	METRICS	DIMENSIONS

Average Lead-Time Days (#)
Order Fill Rate (%)
On-Time Unit Delivery (%)
	Avg. Quoted Lead Days (#)
	Avg. Sales per Order ($)
	Avg. Shipment Miles (#)
	Sales Order Count (#)
	Shipments On-Time (#)
	Units Delivered On-Time (#)
	Units Shipped (#)
		Billing Customer
		    Industry Group
		    Industry
		    Category
		    Customer Name
		Carrier/Distributor
		    Distributor/Carrier Type
		    Carrier
		Cust. Delivery On-Target Range
		    Range
		Fiscal Week
		    Fiscal Year
		    Quarter
		    Month
		    Week
		Lead-Time Range
		    Range
		On-Time Shipment Range
		    Range
		Plants
		    Plant
		Product SKU
		    Product Line
		    Brand
		    SKU
		Shipment Type/Bill of Lading (#)
		    Shipment Type
		    Shipment Bill of Lading (#)
		Ship-To Location
		    Region
		    State/Province
		    County
		    City
		    Zip Code/Postal Code

Customer-Related Schema: Information, Complaints, and Claims

GOALS	METRICS	DIMENSIONS	
Complaint Count (#) Failed Orders (#) Returned Units (#)	Canceled Order Count (#) Claim Payments ($) Claim Payments (#) Claim Settlement ($) Claims (#) Claims ($) Customer Recommendations (#) Damaged Units (#) Failed Orders ($) Returned Product ($/%) Service Call Count (#)	Billing Customer Industry Group Industry Category Customer Name Carrier/Distributor Distributor/Carrier Type Carrier Claim Status Claims Received Claim Type Type Identification (#) Complaint Type Identification (#)	Complaint Status Complaint Received Customer Location Region State/Province County Zip Code/Postal Code End-Customer by Type Type Group Customer ID Fiscal Month Year Quarter Month Product Product Line Brand SKU

Customer-Related Schema: Service Benchmark

GOALS	METRICS	DIMENSIONS

Average Resolution Response Time (#)
Customer Satisfaction Scorecard
Service Effectiveness Index
	Damaged Units (#)
	Failed Orders (#)
	Lost Customer Count (#)
	Outstanding Service Issues (#)
	Returned Product ($/%)
	Service Call Count (#)
		Billing Customer
		    Industry Group
		    Industry
		    Category
		    Customer Name
		Customer Location
		    Region
		    State/Province
		    County
		    Zip Code/Postal Code
		End-Customer Location
		    Region
		    State/Province
		    County
		    Zip Code/Postal Code
		Fiscal Month
		    Year
		    Quarter
		    Month
		Product Brand
		    Product Line
		    Brand
		Service Relationship Perspective
		    Service Relationship Perspective

Customer-Related Schema: Service Value

GOALS	METRICS	DIMENSIONS

Lifetime Profit ($)
Service Cost (%)
Service Effectiveness Index
	Claims (#)
	Claims ($)
	Complaint Count (#)
	Customer Retention Cost ($)
	Customer Service Cost ($)
	Customer Visits (#)
	Customers (#)
	Lost Customer Count (#)
	Net Profit ($/%)
	Outstanding Service Issues (#)
	Receivables ($)
		Aging Brackets
		    Range
		Billing Customer
		    Industry Group
		    Industry
		    Category
		    Customer Name
		Contracted Pay't Time
		    Range
		End-Customer by Type
		    Type
		    Group
		    Customer ID
		Fiscal Month
		    Year
		    Quarter
		    Month
		Product SKU
		    Product Line
		    Brand
		    SKU

PRESENTATION

Grammar of Graphics

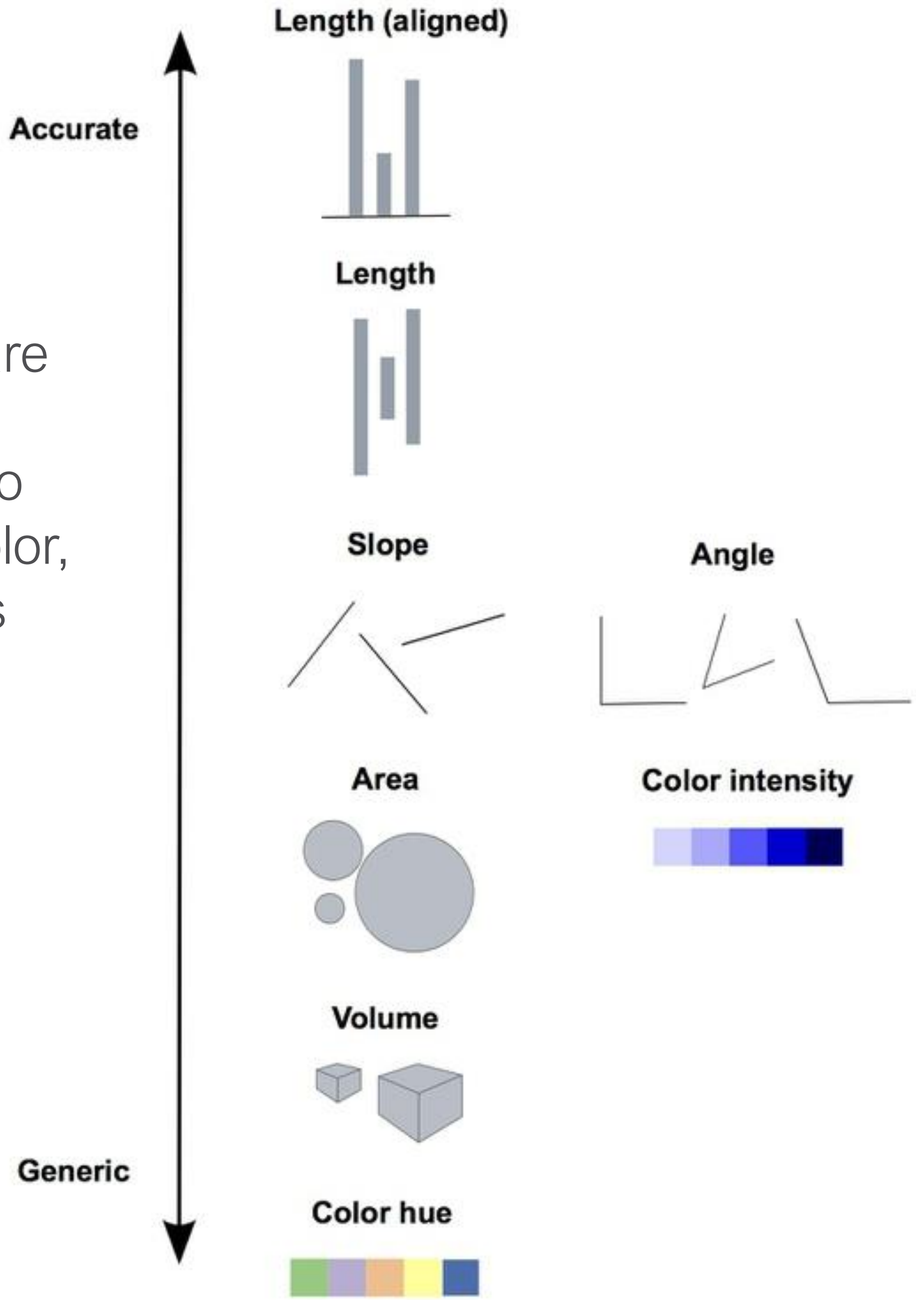
- **Data**: Always start with the data, identify the dimensions you want to visualize.
- **Aesthetics**: Confirm the axes based on the data dimensions, positions of various data points in the plot. Also check if any form of encoding is needed including size, shape, color and so on which are useful for plotting multiple data dimensions.
- **Scale:** Do we need to scale the potential values, use a specific scale to represent multiple values or a range?
- **Geometric objects:** These are popularly known as 'geoms'. This would cover the way we would depict the data points on the visualization. Should it be points, bars, lines and so on?
- **Statistics:** Do we need to show some statistical measures in the visualization like measures of central tendency, spread, confidence intervals?
- **Facets:** Do we need to create subplots based on specific data dimensions?
- **Coordinate system:** What kind of a coordinate system should the visualization be based on — should it be cartesian or polar?

Major Components of the Grammar of Graphics

Coordinate system	Cartesian, Polar?
Facets	Create subplots based on multiple dimensions
Statistics	Mean, Quantile, Confidence Intervals?
Geometric objects	Line, Bar, Points?
Scale	Scale values, represent multiple values?
Aesthetics	Axes, plot positions, encodings?
Data	Our datasets

Encoding

Whenever we visualize, we are encoding data using visual cues, or “mapping” data onto variation in size, shape or color, and so on. There are various ways of doing this, as this primer illustrates:



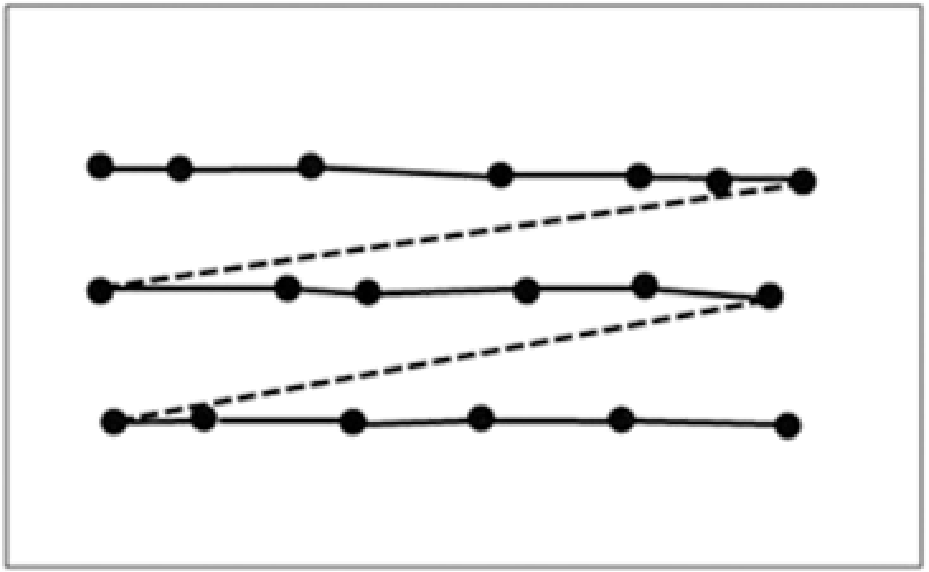
NAVIGATION

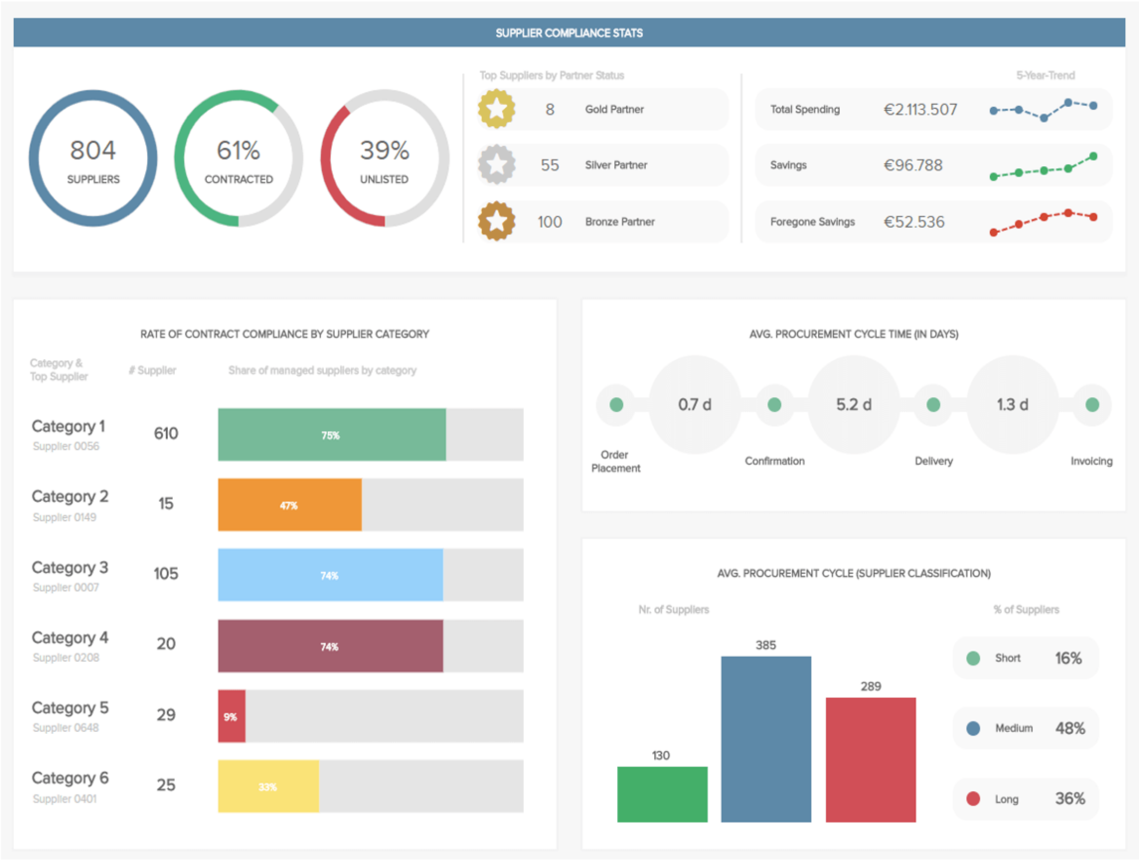
Navigation

- Understand user attention
- Numbers set in large type are consistently one of the first things on the page that received that attention.
- Colors that highly contrast the background and other elements on the dashboard also draw the eye.
- Place similar content together. It will help smooth the flow of reading.

Eye Movement

- Graphs should be placed to be easily navigated. Eye movement paths should not be crossed.





Multidimensional Customer Analytics

- Define customer behaviors for customer analytics
- Define customer metrics and dimensions
- Design star schema
- Design dashboard

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 6
Customer Lifetime Value

Basic CLV

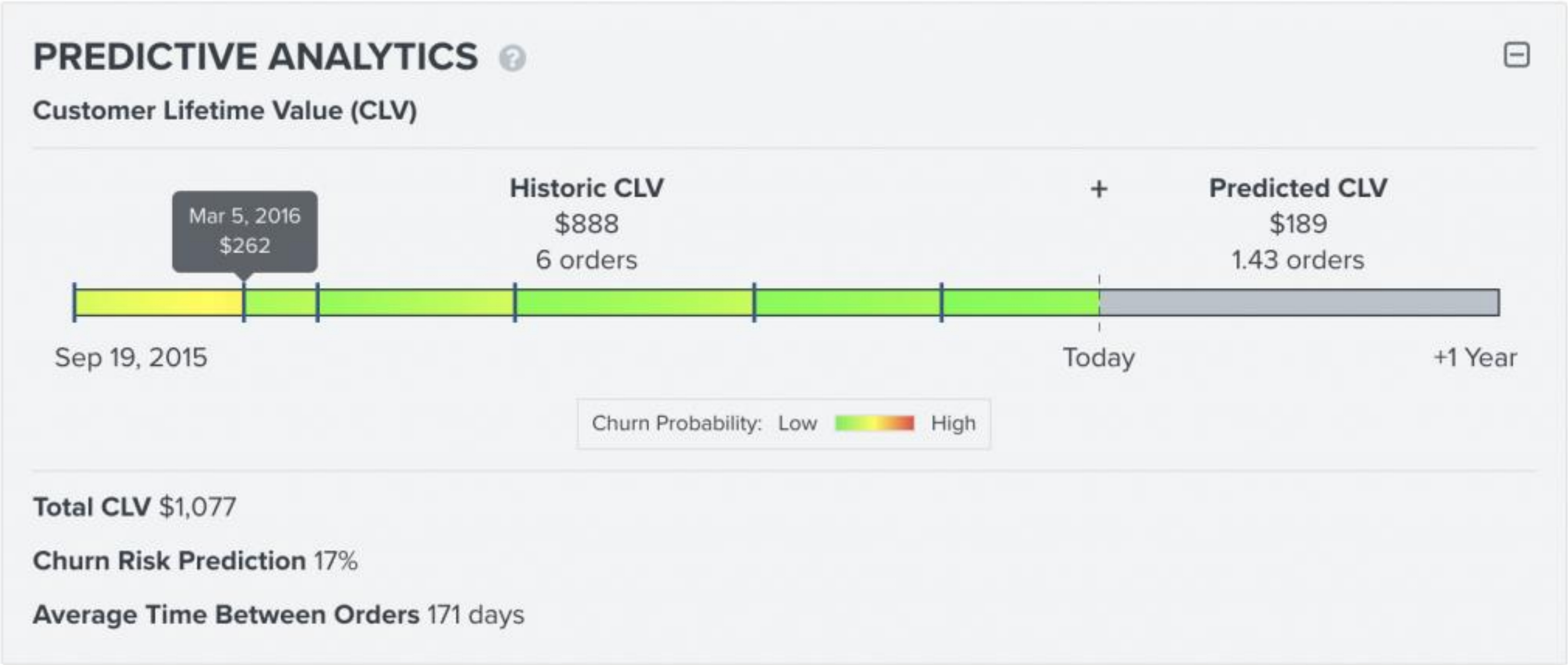
Business Case Example

- In October 2009, Groupon offered its first deal. For $13 (one half the normal price of $26) a customer could purchase two pizzas from the Chicago Motel Bar pub.
- Groupon took one half of the $13 and gave Motel Bar one half of the $13. On average the variable cost of a pizza is approximately 35 percent of the pizza's retail price.
- Motel Bar received $6.50 for pizza that had a variable cost of $9.10.

- Did Motel Bar make the right decision?

Customer Lifetime Value

Customer lifetime value is the dollar value of a customer relationship based on the present value of the projected future cash flows from the relationship



Historic CLV: the revenue generated from all orders placed to date.

Predicted CLV: the total dollars the customer is expected to spend in the next year(s?).

https://www.klaviyo.com/blog/ai-powered-customer-lifetime-value

Key Points of the CLV Definitions

- Customer lifetime value is calculated as a single dollar number,
- CLV summarizes total revenue and costs related to a customer over time,
- CLV provides a net profit/loss summary of the customer's total relationship with the firm,
- It is calculated on per customer basis, or more usually on the average value for a customer within a particular market segment,
- CLV is an important measure of customer profitability and
- Customer lifetime value is usually considered to be a very important marketing metric, because of the range of the marketing objectives it measures within a single number.

https://www.clv-calculator.com/clv/definition-customer-lifetime-value/

Simple Calculation Related to CLV

$$APRU = \frac{TR}{CQ}$$

TR - Total revenue for a chosen period

CQ - Number of customers for a chosen period

https://medium.com/swlh/5-simple-ways-to-calculate-customer-lifetime-value-5f49b1a12723

Example of ARPU

Let's suppose 20 customers brought $1,240 in profit over a three-month period.

ARPU (3 months) = $1240 / 20 = $62

Let's see what these customers will bring us in one year.

ARPU (12 months) = ARPU (3 months) × 4 = $62 × 4 = $248 per year per customer

The historical CLV equals the ARPU for one year, which is $248

Simple Calculation Related to CLV

$$CLV = \frac{T \times AOV \times AGM \times ALT}{\text{Number of clients for the period}}$$

T - Average number of transactions per month

AOV - Average order value

AGM - Average gross margin

ALT - Average customer lifespan in months

Example of CLV

1. First, we'll calculate the average number of transactions (T):

- Period: 6 months
  Total transactions: 120
  T = 120 / 6 = 20

2. AOV is the average value of an order, or the average revenue from each order.

- Total revenue (November): $12,000
  Number of orders: 20
  AOV = $12,000 / 20 = $600

Example of CLV

3. AGM is the average gross margin, which tells you what part of each sale is your actual profit and what part is the cost (expressed as a percentage).

Example:

Total revenue (November): $12,000

Cost of sales: $8,000

Gross margin (%) = (($12,000 — $8,000) / $12,000) × 100 = 33%

Example of CLV

Month	Total revenue	Cost of Sales	Gross Margin
June 2018	$10,500	$8,000	0.24
July 2018	$12,500	$9,500	0.24
August 2018	$11,200	$7,900	0.29
September 2018	$12,000	$8,000	0.33
October 2018	$12,500	$9,000	0.28
November 2018	$12,000	$8,900	0.26
Total gross margin			1.64

- AGV = 1.64 / 6 = 0.27, or 27%

Example of CLV

4. ALT represents the average lifespan of a customer, which tells you how long the average customer has been with your company.

$$ALT = \frac{1}{\text{Churn Rate \%}}$$

$$\text{Churn Rate} = \left(\frac{CB - CE}{CB}\right) \times 100$$

CB - Clients at the beginning of a month

CE - Clients at the end of a month

Example of CLV

- Suppose you had 200 customers at the beginning of November and 150 customers at the end of November.
- Churn rate (%) = (200–150) / 200 = 50 / 200 = 0.25, or 25%
- ALT = 1 / 25% = 1 / 0.25 = 4 months

Example of CLV

- Average number of transactions per month (T) = 20
- Average order value (AOV) = $600
- Average gross margin (AGM) = 27%
- Average customer lifespan in months (ALT) = 4 months

- CLV (total) = 20 × $600 × 27% × 4 = $1,296,000

Simple Customer Lifetime Value Formula

[Annual profit contribution per customer x Average number of years that they remain a customer] - the initial cost of customer acquisition

An example of the simple customer lifetime value formula

Let's assume the following:

- Profit generated by the customer each year = $1,000
- Number of years that they are a customer of the brand = 5 years
- Cost to acquire the customer = $2,000

- What is the customer lifetime value of this customer

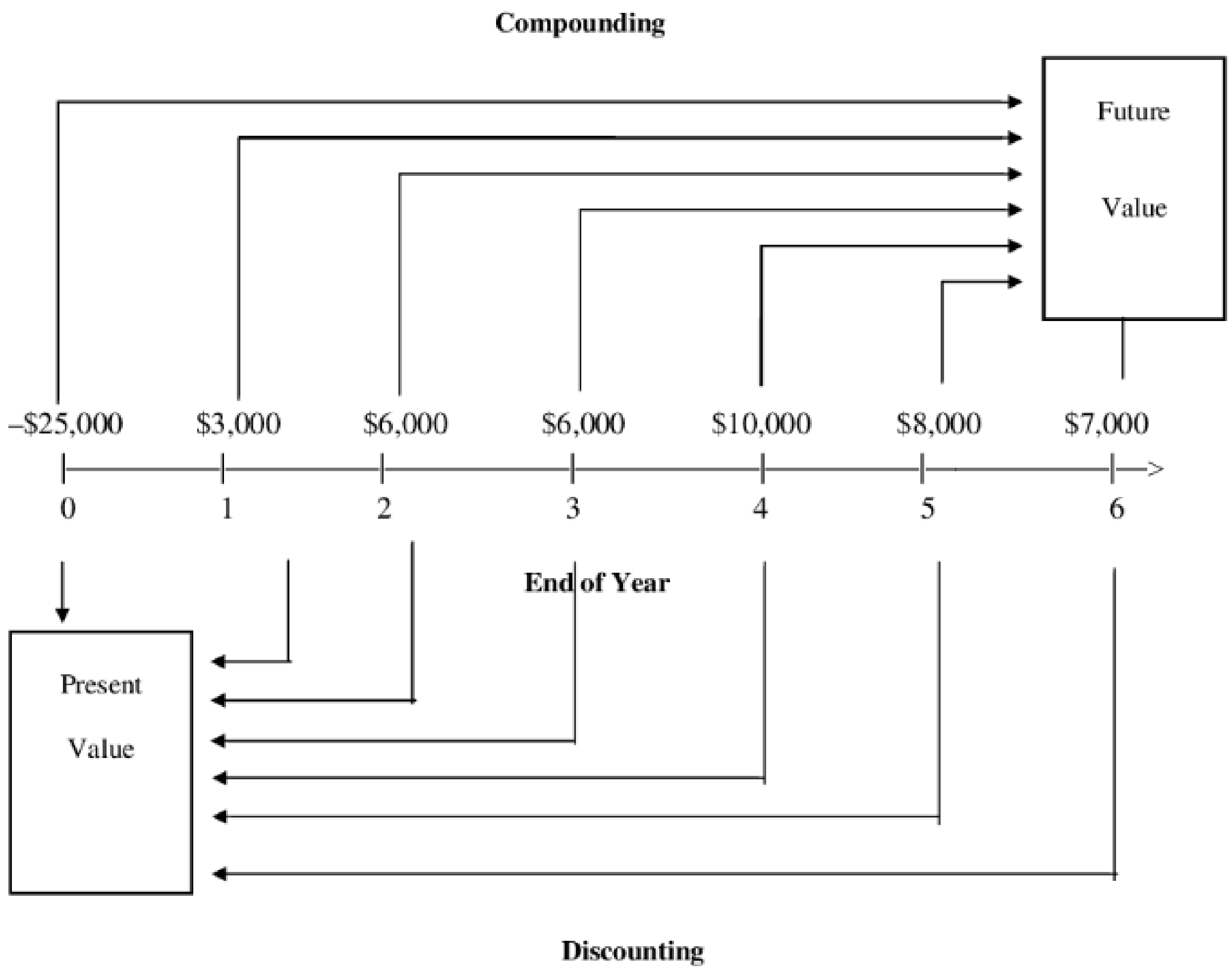
A more detailed example of the simple CLV formula

- annual revenue per average customer is $2,000 per annum
- product costs associated with the average customer’s purchases is $500 per year
- the firm also spends $100 a year per customer to provide customer service
- annual retention rate (loyalty rate) is 80%
- average costs to acquire a new customer are $1,000

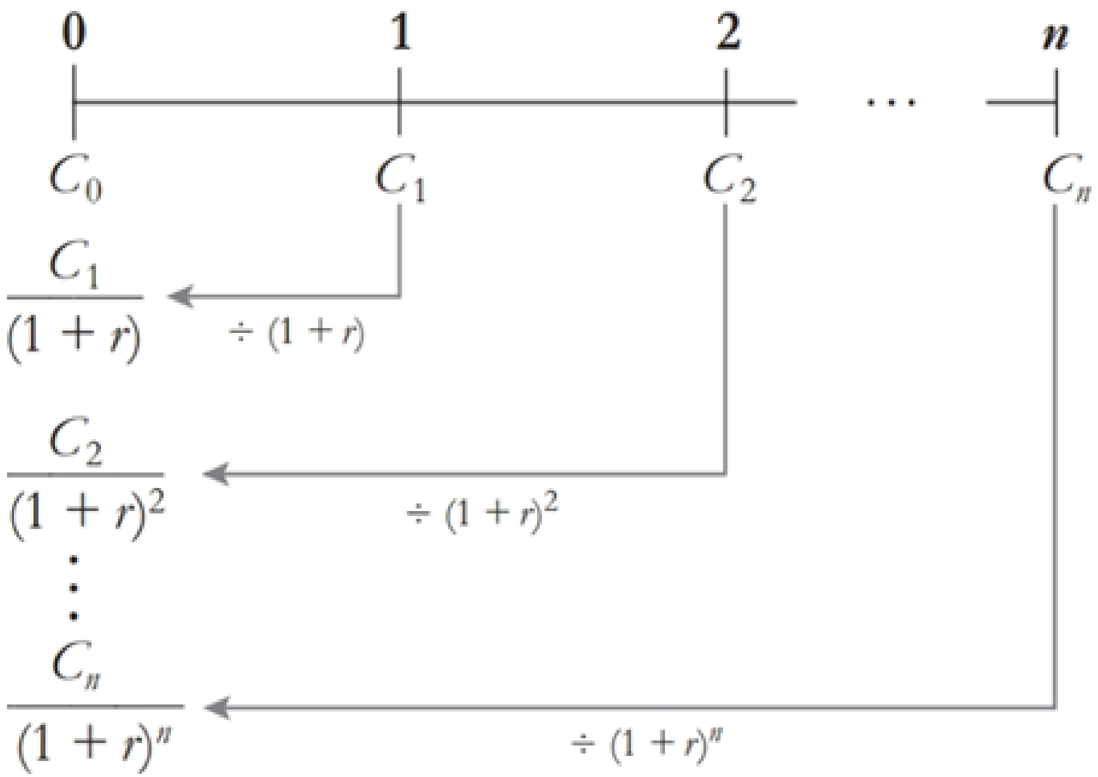
customer lifetime in years = (1 / 1- annual retention rate)

CLV with Time Value of Money

Time Value of Money



Present Value and NPV



$$NPV = -C_0 + \frac{C_1}{1+r} + \frac{C_2}{(1+r)^2} + \ldots + \frac{C_T}{(1+r)^T}$$

$-C_0 = Initial\ Investment$

$C = Cash\ Flow$

$r = Discount\ Rate$

$T = Time$

CLV Formula

$$\left[\text{Annual profit contribution per customer (for each year under consideration)} \times \text{The cumulative customer retention rate} \right] - \text{the initial cost of customer acquisition}$$

With each yearly figure adjusted by an appropriate discount rate

Main differences between the two CLV formulas

- Changing customer revenues each year,
- Changing customer retention and up selling costs over time,
- Changing retention (and therefore churn rates) over time, and
- It applies a discount rate (to determine the present value)

CLV Calculation

Discount Rate	10%

	1	2	3	4	5
Average Revenue per Customer	300	500	700	900	1,100
Average Customer Product Cost	150.00	250.00	350.00	450.00	550.00
Average Gross Profit Contribution	*150.00*	*250.00*	*350.00*	*450.00*	*550.00*
Up-selling and retention cost per Retained Customer	50.00	50.00	50.00	50.00	50.00
Average Net Profit Contribution per Customer	*100.00*	*200.00*	*300.00*	*400.00*	*500.00*
Expected Retention Rate (%)	N/A	60%	62%	64%	66%
% of total customers retained	100.0%	60.0%	37.2%	23.8%	15.7%
Average Net Profit Contribution on Retained Customers	*100.00*	*120.00*	*111.60*	*95.23*	*78.57*
Discounted Average Net Profit Contribution on Retained Customers	*90.91*	*99.17*	*83.85*	*65.04*	*48.78*

Exercise

- Profit (customer revenues less costs) generated by the customer in year one = $1,000
- This increases to $1,500 in year two and then to a maximum of $2,000 for year three onwards
- The retention rate of customers is 75% (and therefore the churn rate is 25%)
- Cost to acquire the customer = $1,000
- A discount rate of 10% is considered appropriate

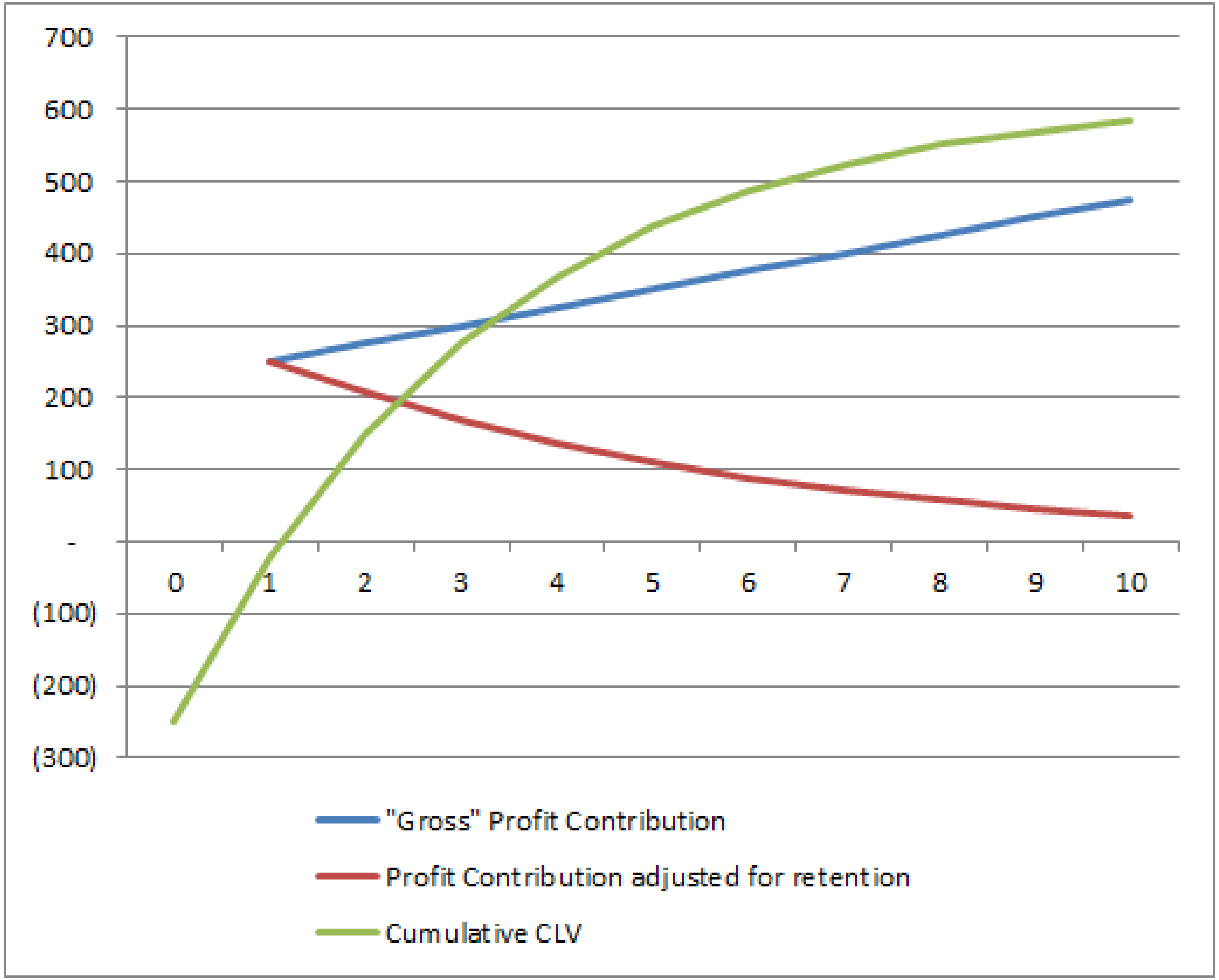
Another Example of a Customer Lifetime Value Calculation

- the initial acquisition cost per customer is $250
- the average customer revenue starts at $500 in year one and increases by $50 per year – rising to $950 by year 10
- customer costs have been set at 50% of revenue – therefore, it costs the firm in (in year one) $250 in product and service costs to generate $500 revenue
- only customer revenues and costs are considered up to year 10, with any subsequent revenues being disregarded in the calculation
- a retention rate of 75% has been used
- a discount rate of 10% has been applied

CLV Calculation

EXAMPLE CUSTOMER LIFETIME VALUE CALCULATION	*Year 0*	*Year 1*	*Year 2*	*Year 3*	*Year 4*	*Year 5*	*Year 6*	*Year 7*	*Year 8*	*Year 9*	*Year 10*
Initial Acquisition Cost per Customer	-250										
Average Customer **Revenue** (of those customers who are retained)		500	550	600	650	700	750	800	850	900	950
Average Customer **Cost**		250	275	300	325	350	375	400	425	450	475
Average **Profit** Contribution		250	275	300	325	350	375	400	425	450	475
% of Customers Retained each year at a 75% retention rate		100%	75%	56%	42%	32%	24%	18%	13%	10%	8%
Average Profit Contribution of RETAINED CUSTOMERS ONLY		250	206	169	137	111	89	71	57	45	36
Discount Rate Applied Each Year at a 10% Discount Rate	1.00	1.10	1.21	1.33	1.46	1.61	1.77	1.95	2.14	2.36	2.59
DISCOUNTED Average **Profit** Contribution PER RETAINED CUSTOMER		227	170	127	94	69	50	37	26	19	14
Running TOTAL of Average Profit Contribution, with Year 10 = CLV	-250	-23	148	275	368	437	487	524	550	569	583

CLV Plot


700
600
500
400
300
200
100
-
(100)
(200)
(300)
0
1
2
3
4
5
6
7
8
9
10
"Gross" Profit Contribution
Profit Contribution adjusted for retention
Cumulative CLV

Standard CLV Formula

Terminology

- Contribution Margin (M) = Revenue earned from serving a customer - VC (Variable Cost) spent in serving the customer = ARPU (Average Revenue Per User) - VC
- Discount Rate (d) = $100 in 2050 is less valuable than $100 today, so, we will discount the profit earned to get the present value of the money.
- Churn Rate(c) = Percentage of customers that leave the platform.
- Retention Rate (r) = 1 - The churn rate = Percentage of customers who stay on the platform. It can also be calculated as, ((Ce-Cn)/Cs))x100.
- where:
- Ce = Number of customers at the end of a given period,
- Cn = Number of new customers acquired during a given period,
- Cs = Number of customers at the beginning of a given period.

https://medium.com/@vinaysays/how-to-calculate-the-customer-lifetime-value-cltv-ecfe2b1d046f

Case I: Customer pays at the beginning of the period

$$CLTV = M + (M*r)/(1+d) + (M*r^2)/(1+d)^2 + \ldots \quad \text{(Eq. 1)}$$

$$CLTV * r/(1+d) = (M*r)/(1+d) + (M*r^2)/(1+d)^2 + (M*r^3)/(1+d)^3 + \ldots \quad \text{(Eq. 2)}$$

(Eq. 3) - (Eq. 4) => $CLTV\ (1 - r/(1+d)) = M/(1+d)$ or $CLTV(1+d-r)/(1+d) = M/(1+d)$ or $CLTV = M / (1+d-r)$

Thus, $CLTV = M/(1+d-r)$

Case II: Customer pays at the end of the period

$$CLTV = (M)/(1+d) + (M*r)/(1+d)^2 + (M*r^2)/(1+d)^3 + \ldots \text{ (Eq. 3)}$$

$$CLTV * r/(1+d) = (M*r)/(1+d)^2 + (M*r^2)/(1+d)^3 + (M*r^3)/(1+d)^4 + \ldots \text{ (Eq. 4)}$$

(Eq. 1) - (Eq. 2) => $CLTV (1 - r/(1+d)) = M$ or $CLTV(1+d-r)/(1+d) = M$

Thus, $CLTV = M * (1+d)/(1+d-r)$

Formula Generalization

From the above two cases, we see that CLTV can be calculated as

Payment at the Beginning of the period	Payment at the End of the period
$CLTV = \frac{M\,(1+d)}{(1+d-r)}$	$CLTV = \frac{M}{(1+d-r)}$

If we ignore the discount rate, then both formulae converge to

Payment at the Beginning of the period	Payment at the End of the period
$CLTV = \frac{M}{(1-r)}$	$CLTV = \frac{M}{(1-r)}$

(1-retention rate) = (1 - r) = Churn Rate = c

CLTV = M/c

Case III: Customer pays at the beginning of the period with Churn at the 1st Year

$$\text{CLTV} = M*r + (M*r^2)/(1+d) + (M*r^3)/(1+d)^2 + \ldots \text{ (Eq. 5)}$$

$$\text{CLTV} * r/(1+d) = (M*r^2)/(1+d) + (M*r^3)/(1+d)^2 + (M*r^4)/(1+d)^3 + \ldots \text{ (Eq. 6)}$$

$$\text{(Eq. 5) - (Eq. 6)} \Rightarrow \text{CLTV}\,(1 - r/(1+d)) = M*r \text{ or } \text{CLTV}(1+d-r)/(1+d) = M$$

Thus, $\text{CLTV} = M * r * (1+d)/(1+d-r)$

Case IV: Customer pays at the end of the period with Churn at the 1st Year

$$CLTV = M*r/(1+d) + (M*r^2)/(1+d)^2 + (M*r^3)/(1+d)^3 + \ldots \quad (Eq.\ 7)$$

$$CLTV * r/(1+d) = (M*r^2)/(1+d)^2 + (M*r^3)/(1+d)^3 + \ldots \quad (Eq.\ 8)$$

(Eq. 7) - (Eq. 8) => $CLTV\ (1 - r/(1+d)) = M*r/(1+d)$ or $CLTV(1+d-r)/(1+d) = M*r/(1+d)$ or $CLTV = M*r\ /\ (1+d-r)$

Thus, $CLTV = M*r/(1+d-r)$

Standard CLV Formula

When does the customer pay \ When does the customer Leave	After Next Payment	Before Next Payment
End of period	$\frac{M\,(1+d)}{(1+d\ -r)}$	$\frac{M\,(1+d)*r}{(1+d\ -r)}$
Beginning of period	$\frac{M}{(1+d\ -r)}$	$\frac{M\ *r}{(1+d\ -r)}$

Predicting Customer Lifetime Value with “Buy ‘Til You Die” probabilistic models in Python

https://towardsdatascience.com/predicting-customer-lifetime-value-with-buy-til-you-die-probabilistic-models-in-python-f5cac78758d9

Buy 'Til You Die

- "Buy 'Til You Die" probabilistic models help us in quantifying the lifetime value of a customer by assessing the expected number of his future transactions and his probability of being "alive".

- To understand how Buy 'Til You Die models work, we focus on our best choice to predict real life data: the BG/NBD model.
- The Beta Geometric/Negative Binomial Distribution model was introduced in 2004 by P. Fader's Paper as an improvement of the Pareto/NBD model (the first BTYD) developed by Schmittlein et al. in 1987.

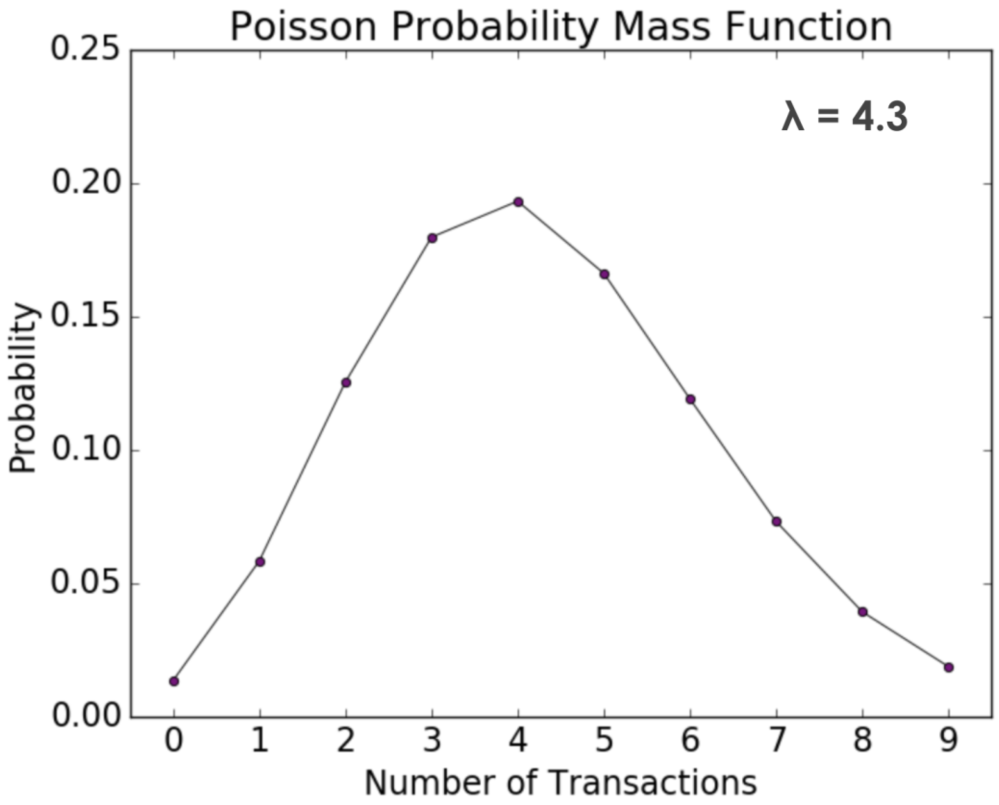
- In particular, to predict future transactions the model treats the customer purchasing behavior as a coin tossing game.
- Each customer has 2 coins: a buy coin that controls the probability of a customer to purchase, and a die coin that controls the probability of a customer to quit and never purchase again.

Assumption 1: while active, the number of transactions made by a customer follows a Poisson Process with transaction rate λ (=expected number of transactions in a time interval).

- At every sub-period (1 month) of a specific time interval (12 months) each customer tosses his buy coin and, depending on the result, he purchases or not.
- The number of transactions (heads) we observe in the period depends on each customer's probability distribution around λ.
- Let's plot below a customer's Poisson probability distribution to visualize what we just said.

Poisson Plot

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import poisson,expon,nbinom

poisson_lambda = 4.3
p_arr = []

distribution = poisson(poisson_lambda)
for transactions in range(0,10):
    p_arr.append(distribution.pmf(transactions))

plt.ylabel('Probability')
plt.xlabel('Number of Transactions')
plt.xticks(range(0, 10))
plt.title('Poisson Probability Mass Function')
plt.plot(p_arr, color='black', linewidth=0.7, zorder=1)
plt.scatter(range(0, 10), p_arr, color='purple', edgecolor='black', linewidth=0.7, zorder=2)
display(plt.show())
```



Here we assume that our random customer has a transaction rate λ = 4.3. As a consequence he will have a 19% probability of purchasing 4 times in a random 12 month period and a 4% probability of purchasing 8 times, and so on.

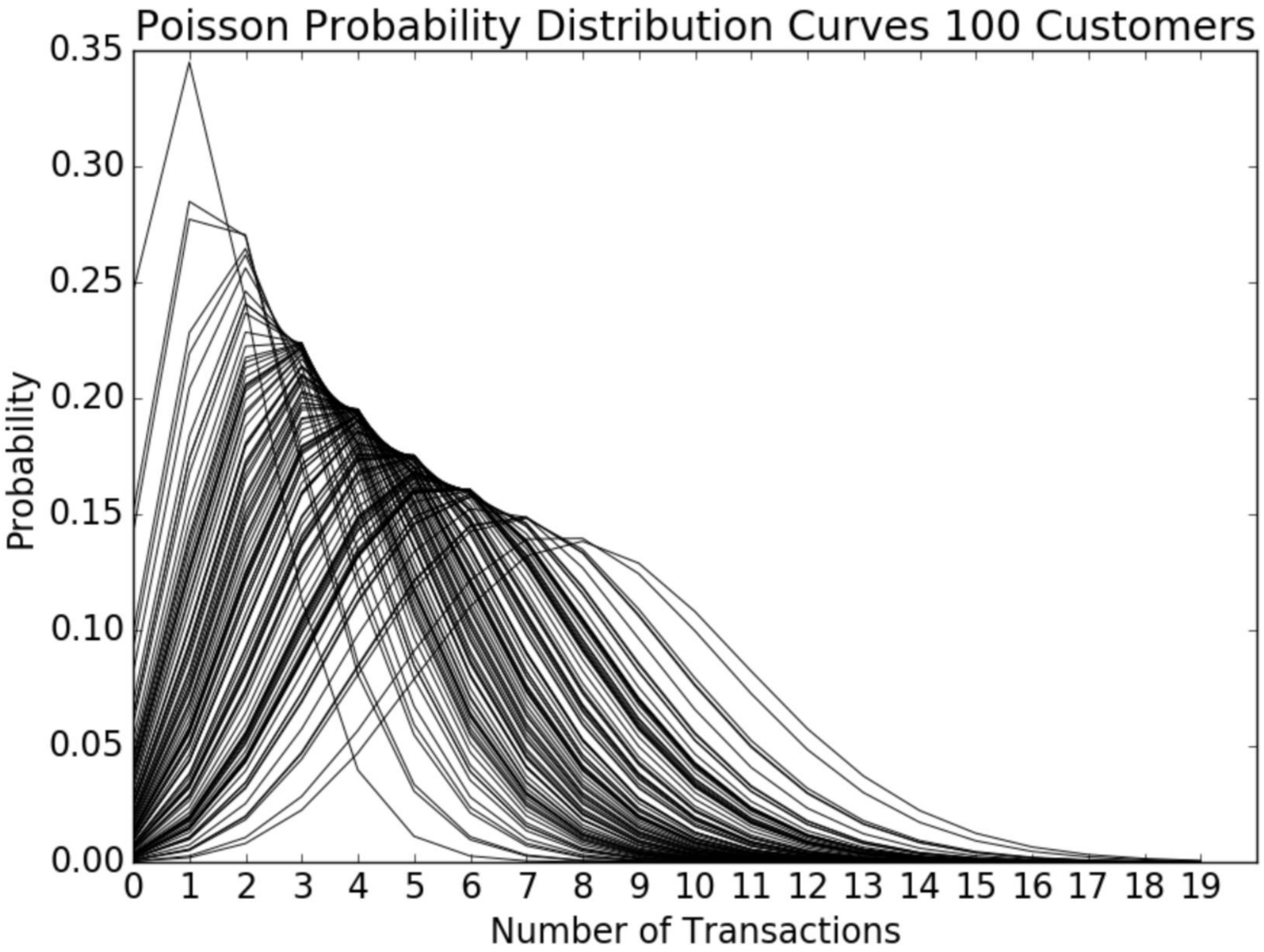
Assumption 2: heterogeneity in transaction rates among customers follows a Gamma distribution.

- This is equivalent to saying that each customer has its own buy coin (with its very own probability of head and tail).
- To better understand the assumption, we simulate the Poisson distribution of 100 customers where each λ is modelled with a Gamma distribution with parameters: shape=9 and scale=0.5.

Gamma Plot

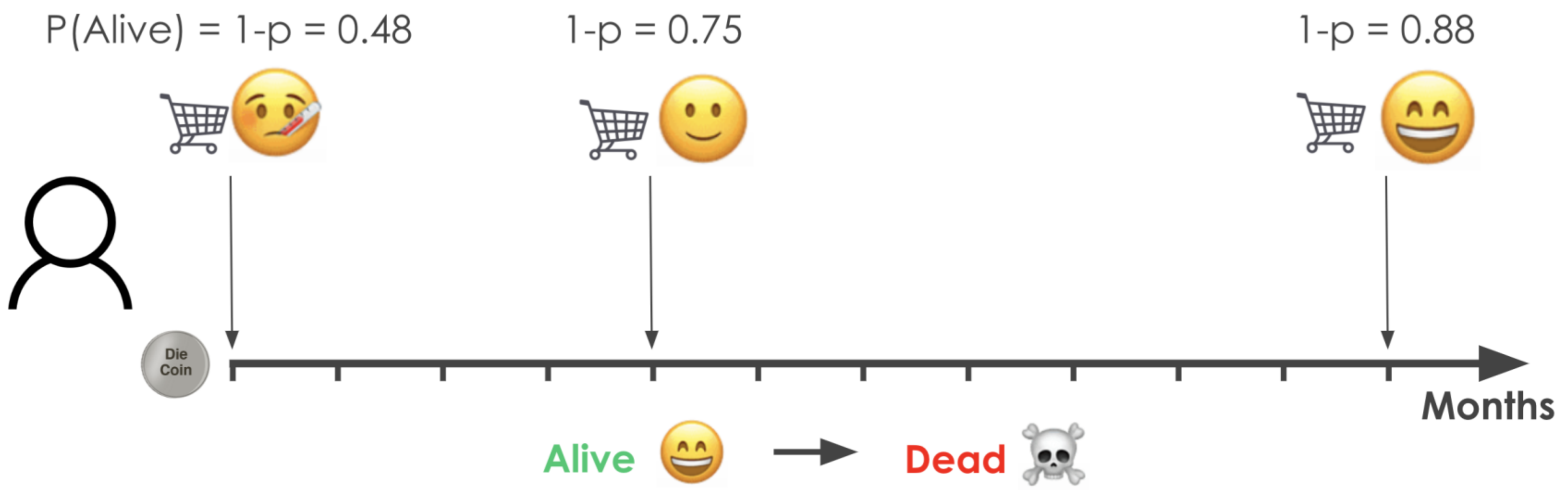
```
gamma_shape = 9
gamma_scale = 0.5

for customer in range(0, 100):
    distribution = poisson(np.random.gamma(shape=gamma_shape, scale=gamma_scale))
    p_arr = []
    for transactions in range(0,9):
        p_arr.append(distribution.pmf(transactions))
    plt.plot(p_arr, color='black', linewidth=0.7, zorder=1)

plt.ylabel('Probability')
plt.xlabel('Number of Transactions')
plt.xticks(range(0,9))
plt.title('Poisson Probability Distribution Curves 100 Customers')
plt.show()
```

Gamma Plot

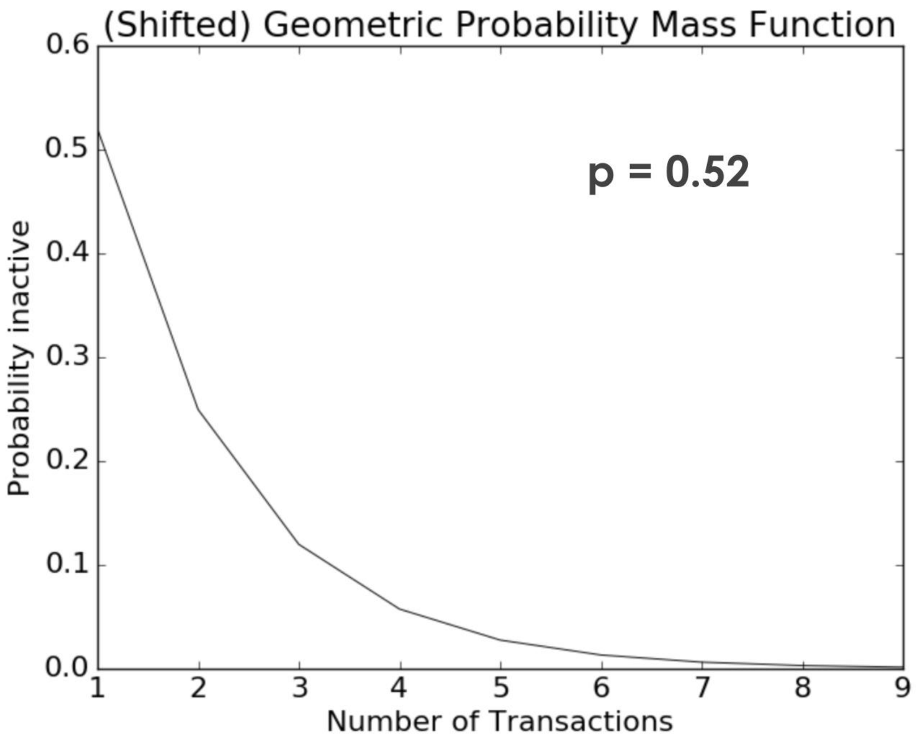

Poisson Probability Distribution Curves 100 Customers
Probability
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Number of Transactions

Assumption 3: after any transaction, a customer becomes inactive with probability p.

- Therefore the point at which the customer “drops out” is distributed across transactions according to a (shifted) Geometric distribution.
- After every transaction, each customer will toss the second coin, the die coin.
- Given that p is the probability of “dying”, then we can define P(Alive) = 1-p.
- Once again, let’s plot a random customer probability distribution to better grasp the meaning of this assumption.



(shifted) Geometric distribution Plot



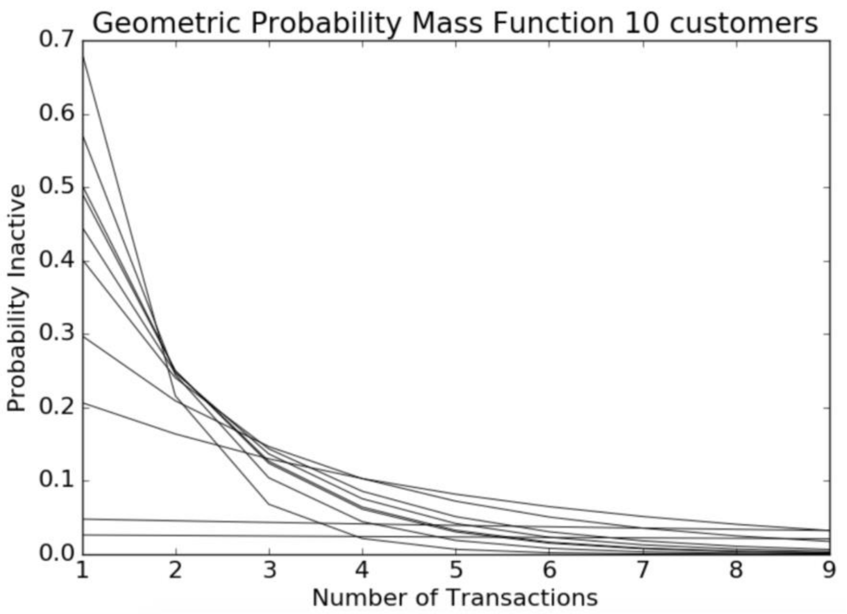
- Assuming that our customer becomes inactive with probability p = 0.52, then the probability that he becomes inactive after the 2nd transaction is 25%, and the probability that he becomes inactive after the 3rd transaction is 12%.
- As you see the more the customer purchases the higher his probability of being alive.

Assumption 4: heterogeneity in p follows a Beta distribution.

- As for the buy coin, each customer has his own die coin with its own probability of being alive after a specific amount of transactions.
- We can see below how that would look for a simulation of 10 customers where p follows a Beta distribution with $\alpha = 2$ and $\beta = 3$.

Geometric Probability Mass Function

```
beta_a = 2
beta_b = 3

for customer in range(0, 10):
  p_arr = []
  beta = np.random.beta(a=beta_a, b=beta_b)
  for transaction in range(1,10):
    proba_inactive = beta*(1-beta)**(transaction-1)
    p_arr.append(proba_inactive)
  p_arr = np.array(p_arr)
  plt.plot(p_arr, color='black', linewidth=0.7, zorder=1)

plt.ylabel('Probability Inactive')
plt.xlabel('Number of Transactions')
plt.xticks(range(1, 10))
plt.title('Geometric Probability Mass Function 10 customers')
display(plt.show())
```



Assumption 5: the transaction rate λ and the dropout probability p vary independently across customers.

- Eventually, by fitting the previously mentioned distributions on the historical customers data we are able to derive a model that for each customer provides:
- $P(X(t) = x \mid \lambda, p)$- the probability of observing x transactions in a time period of length t
- $E(X(t) \mid \lambda, p)$- the expected number of transactions in a time period of length t
- $P(\tau>t)$ - the probability of a customer becoming inactive at period τ
- The fitted distributions parameters are then used in the forward-looking customer-base analysis to find the expected number of transactions in a future period of length t for an individual with past observed behavior defined by x, t_x, T — where x = number of historical transactions, t_x = time of last purchase and T = Age of a customer.

the expected number of transactions in a future period of length t for an individual with past observed behavior

- And here below for the final formula (careful derivation is provided in the Appendix of P. Fader's Paper):

$$E(Y(t)\,|\,X = x, t_x, T, r, \alpha, a, b) =$$

$$\frac{\dfrac{a+b+x-1}{a-1}\left[1-\left(\dfrac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\right)\right]}{1+\delta_{x>0}\dfrac{a}{b+x-1}\left(\dfrac{\alpha+T}{\alpha+t_x}\right)^{r+x}}$$

“Buy ‘Til You Die” in Python

https://towardsdatascience.com/predicting-customer-lifetime-value-with-buy-til-you-die-probabilistic-models-in-python-f5cac78758d9

The Shape of Data

- **Recency (derived from t_x)**: the age of the customer at the moment of his last purchase, which is equal to the duration between a customer's first purchase and their last purchase.
- **Frequency (x)**: the number of periods in which the customer has made a repeat purchase.
- **Age of the customer (T)**: the age of the customer at the end of the period under study, which is equal to the duration between a customer's first purchase and the last day in the dataset.

customer_id	frequency	recency	T
16	3.0	7.983737	7.983737
65	1.0	2.004148	22.998419
875	1.0	6.045299	20.994271
975	0.0	0.000000	17.971622
1777	11.0	22.998419	26.973860

Dataset : http://archive.ics.uci.edu/ml/datasets/online+retail

Lifetimes package in Python

```
import pandas as pd
import numpy as np
from lifetimes.utils import *
from lifetimes import BetaGeoFitter
from lifetimes.plotting import plot_probability_alive_matrix, plot_frequency_recency_matrix
from lifetimes.generate_data import beta_geometric_nbd_model
import matplotlib.pyplot as plt
from lifetimes.plotting import plot_calibration_purchases_vs_holdout_purchases,
plot_period_transactions,plot_history_alive

bgf = BetaGeoFitter()
bgf.fit(data['frequency'], data['recency'], data['T'])

mbgf.summary
```

Model Results

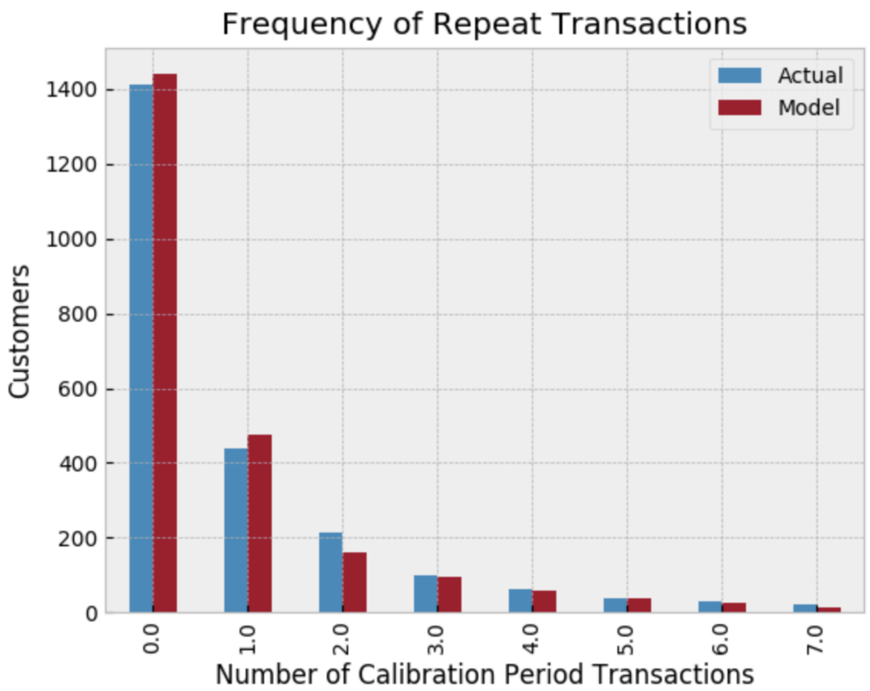
- We fitted the distributions from our assumptions into historical data and derived the model parameters: alpha and r are for the Gamma distribution (Assumption 2), and a and b for the Beta distribution (Assumption 4).
- In the summary we also have a confidence interval for each parameter that we can use to compute a confidence interval of the expected future transactions for each customer.

```
<lifetimes.BetaGeoFitter: fitted with 2357 subjects, a: 0.79, alpha: 4.41, b: 2.43, r: 0.24>

           coef  se(coef)  lower 95% bound  upper 95% bound
r      0.242593  0.012557         0.217981         0.267205
alpha  4.413532  0.378221         3.672218         5.154846
a      0.792886  0.185719         0.428877         1.156895
b      2.425752  0.705345         1.043276         3.808229
```

Assessing the Model Fit

- Now that we built a model, we can check if it really makes sense. A first way to do this is to artificially generate customers with expected purchasing behavior dependent on the fitted model parameters, and comparing it to the real data.

plot_period_transactions(bgf)



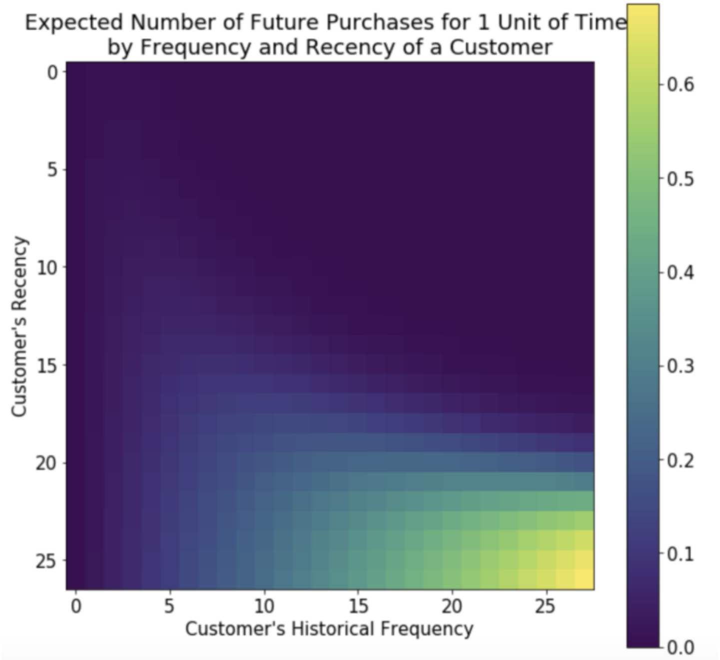
the artificial customers distribution resembles very closely the real data.

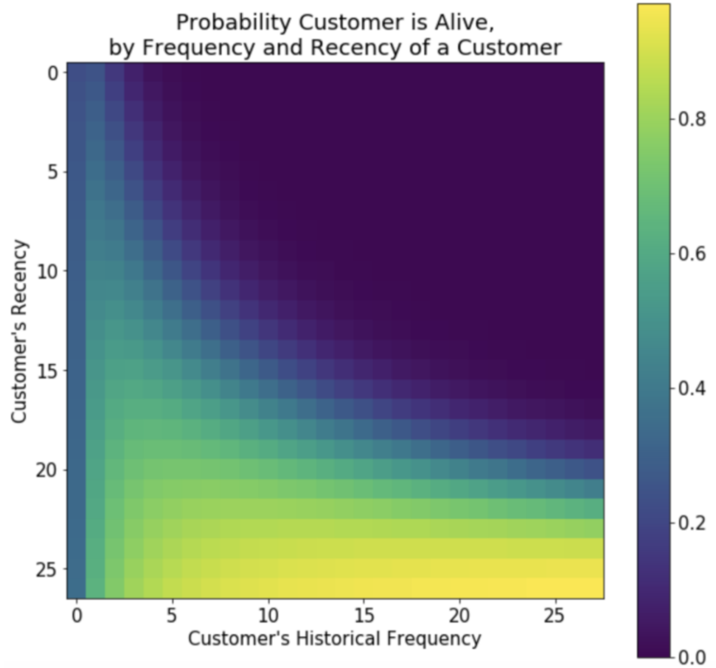
Visualizing the Model Frequency/Recency Matrix

- Now that we have a fitted model we can look at its Frequency/Recency Matrix to inspect the expected relationship (based on our fitted model paramaters) between a customer's recency (age at last purchase), frequency (the number of repeat transactions made) and the expected number of transactions in the next time period (left graph below).
- We can also visualize the expected probability of a customer to be alive depending on her recency and frequency (right graph below).

plot_frequency_recency_matrix(bgf)

plot_probability_alive_matrix(bgf)





Cross Validation

- Once you have verified that the model is close enough to actual data, we can see how good it is in predicting future purchases.

- We can quickly split a simple transactions dataset into calibration and holdout periods using Lifetimes' calibration_and_holdout_data() function. We will first fit the model to a calibration period of 2 years, then predict the next year transactions, and finally compare predicted vs holdout transactions.

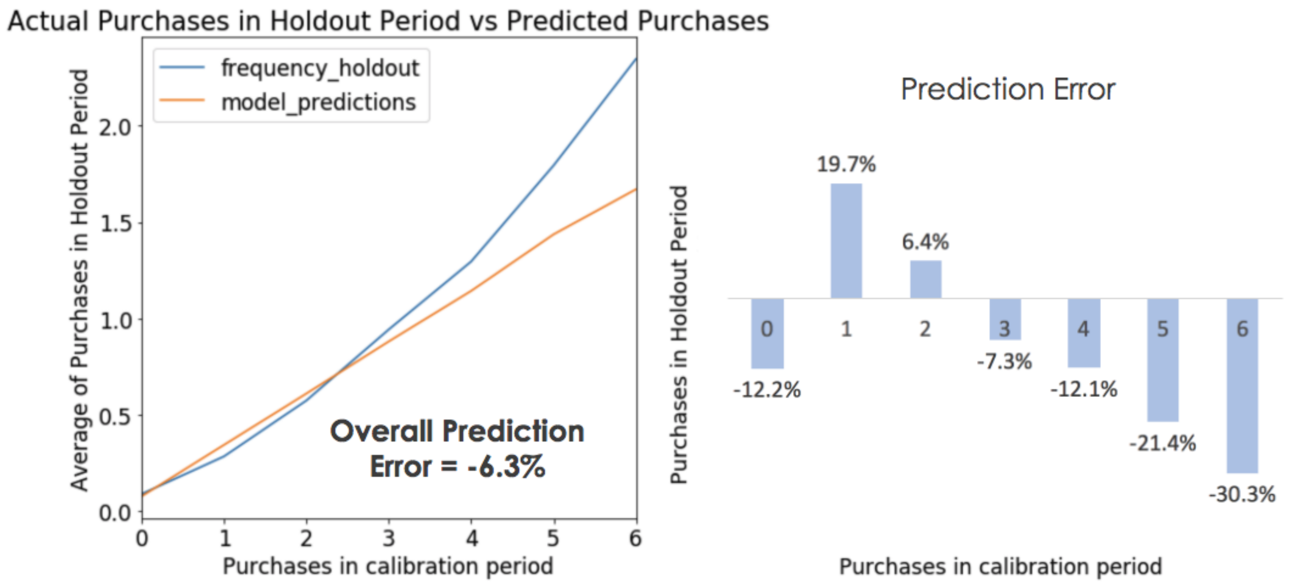
Cross Validation

```
cal_hold = calibration_and_holdout_data(trans_dataset,
                                        'customer_id',
                                        'date',
                                        calibration_period_end='2018-04-30', #2 years calibration
                                        observation_period_end='2019-04-30', #1 year holdout
                                        freq = 'M')
cal_hold.head()

bgf = BetaGeoFitter()
bgf.fit(cal_hold['frequency'], cal_hold['recency'], cal_hold['T'])

plot_calibration_purchases_vs_holdout_purchases(bgf, cal_hold)
```

```
             frequency_cal  recency_cal      T_cal  frequency_holdout  duration_holdout
customer_id
16                     3.0     7.983737   7.983737                2.0                12
65                     1.0     2.004148  22.998419                1.0                12
875                    1.0     6.045299  20.994271                0.0                12
975                    0.0     0.000000  17.971622                0.0                12
1777                  11.0    22.998419  26.973860                5.0                12
```

Prediction Error

By comparing the average actual and predicted purchases in the plot below, we can notice that the prediction and the actuals are very close for customers with 0 to 3 repeat transactions in the calibration period, while they increasingly diverge for customers with more repeats.



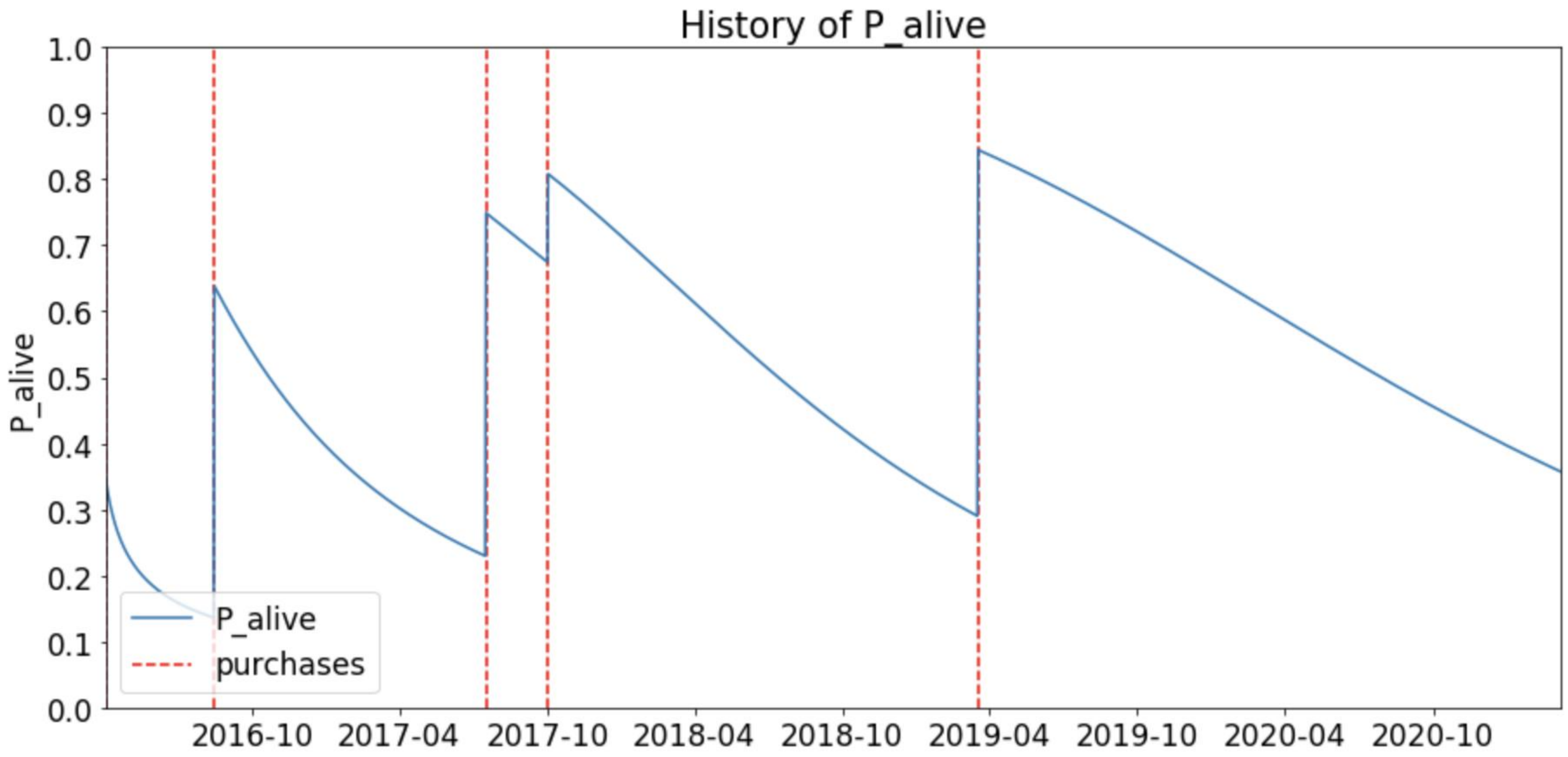
Customer Predictions and Probability Histories

- Once you have built the model and verified its validity you can easily look at single customer predictions and their probability of being alive.

- This is incredibly valuable because you can then use the CLV prediction for marketing activities, forecasting or more generally churn prevention.

Predicting Customer

```
customer_id = ..
days_since_birth = ....
sp_trans = transaction_data.loc[transaction_data['customer_id'] == customer_id]
plot_history_alive(bgf, days_since_birth, sp_trans, 'date')
```



Suggested Readings

- https://www.clv-calculator.com/
- https://medium.com/get-engaged/calculating-customer-lifetime-value-for-startups-with-limited-data-cabce79d4121
- https://medium.com/swlh/5-simple-ways-to-calculate-customer-lifetime-value-5f49b1a12723
- https://srepho.github.io/CLV/CLV

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 7
Customer Acquisition and On-boarding

Customer Acquisition

Customer Acquisition

- The process by which you bring new clients or customers to your business is customer acquisition. The goal, for any company, is to create sustainable and systematic customer acquisition strategies that keep up with industry trends.

- As part of their journey through the funnel consumers:

1. Gain awareness about your brand and interest in your offerings
2. Add your products or services to the list of alternatives they're considering, and
3. Decide to become a paying customer of your business

https://medium.com/hackernoon/how-to-win-new-customers-with-proven-customer-acquisition-strategies-ea7203897354

Acquisition Channel

1. Mass media (TV, radio, and print)
2. Direct marketing (direct mail and outbound telephone)
3. Personal selling (door-to-door and networks)
4. Intermediaries (agents, dealers, and retail chains)
5. Word of mouth (telling friends about the firm)
6. The Internet (e-mail and Web sites)

Verhoef, P. C., & Donkers, B. (2005). The effect of acquisition channels on customer loyalty and cross-buying. *Journal of Interactive Marketing*, *19*(2), 31-43.

Outbound vs. Inbound Marketing

- **Outbound marketing** is just another acceptation for traditional advertising methods. Its counterpart is, of course, inbound marketing. It includes TV, radio, print, telemarketing and all kinds of outdoor advertising. The whole idea behind outbound marketing is that advertisers resort to mass media tools to push and convey their messages out to their audiences in hopes of reaching out to as many people as they can within their targets.

- **Inbound marketing** refers to all those marketing strategies that focus on attracting audiences instead of recklessly going out reaching out to prospects in hopes of getting their attention. Inbound marketing helps marketers to pull visitors and people in, increase levels of brand exposure across the Internet and create, to some extent, brand authority through the creation of genuine and target-tailored content.

- In a nutshell, when it comes to online marketing, an outbound approach is about constant product pushing and sales pitches. An inbound approach is about interactive, engaging content to a narrower audience.

https://medium.com/@ReputationDefender/outbound-vs-inbound-marketing-evident-differences-234ebd98a01d
https://medium.com/@ReadeMilner/inbound-vs-outbound-marketing-db202f8b5772

SEM – Search Engine Marketing

- SEM stands for Search Engine Marketing, and as the name suggests this is one of the many forms of digital marketing that is used nowadays.

- Although it is often confused with other marketing terms that are frequently used in this industry, SEM is not the same as PPC (Pay-per-click) marketing, nor is the same thing as SEO (Search Engine Optimization).

https://medium.com/the-logician/search-engine-marketing-7007e8a719e2

Pay-Per-Click

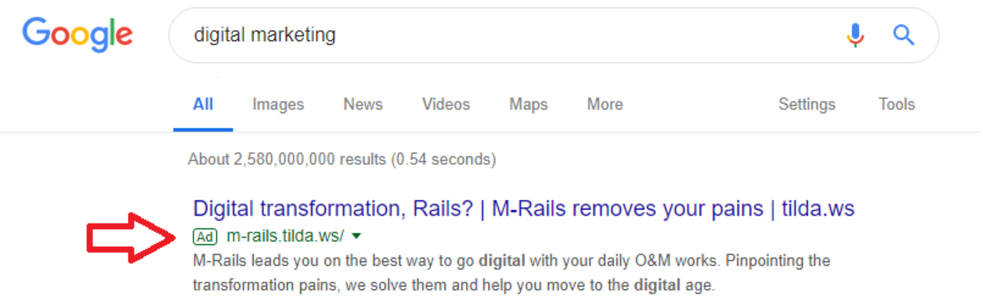
- Paid search engine marketing is often related to Google Ads. The confusing emerges from the fact that Google is by far the most used search engine nowadays, making it also the best option for paid ads. But paid SEM is a wider term, one that people know as PPC or Pay-Per-Click.

- PPC Models
  - Flat-rate PPC: With the flat-rate PPC, the publisher and the business that wants to be advertised agree on a fixed price for a single click.
  - Bid-Based PPC: The auction takes place every time someone searches for that certain keyword you are aiming for. The advertisers that won the bid pay for every click in their ad, with the price being their amount of bid.

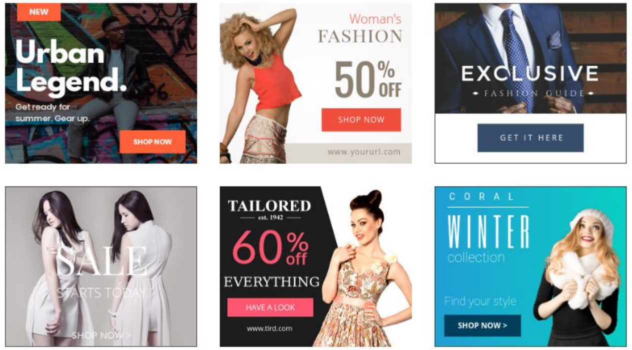
$$\text{Cost-per-click} = \text{Advertising cost} / \text{Ads clicked}$$

Different Types of PPC

- Search Advertising



- Display Advertising



- Social Advertising

SEO – Search Engine Optimization

- SEO (Search Engine Optimization) is the process of affecting the visibility of a website/webpage in a web search engine's unpaid results.
- It's about understanding what people are searching for online, the answers they are seeking, the words they're using, and the type of content they wish to consume.
- Organic search results are the ones that are earned through effective SEO, not paid for (i.e. not advertising).

https://medium.com/free-code-camp/seo-secrets-reverse-engineering-googles-algorithm-92fad4f5a39

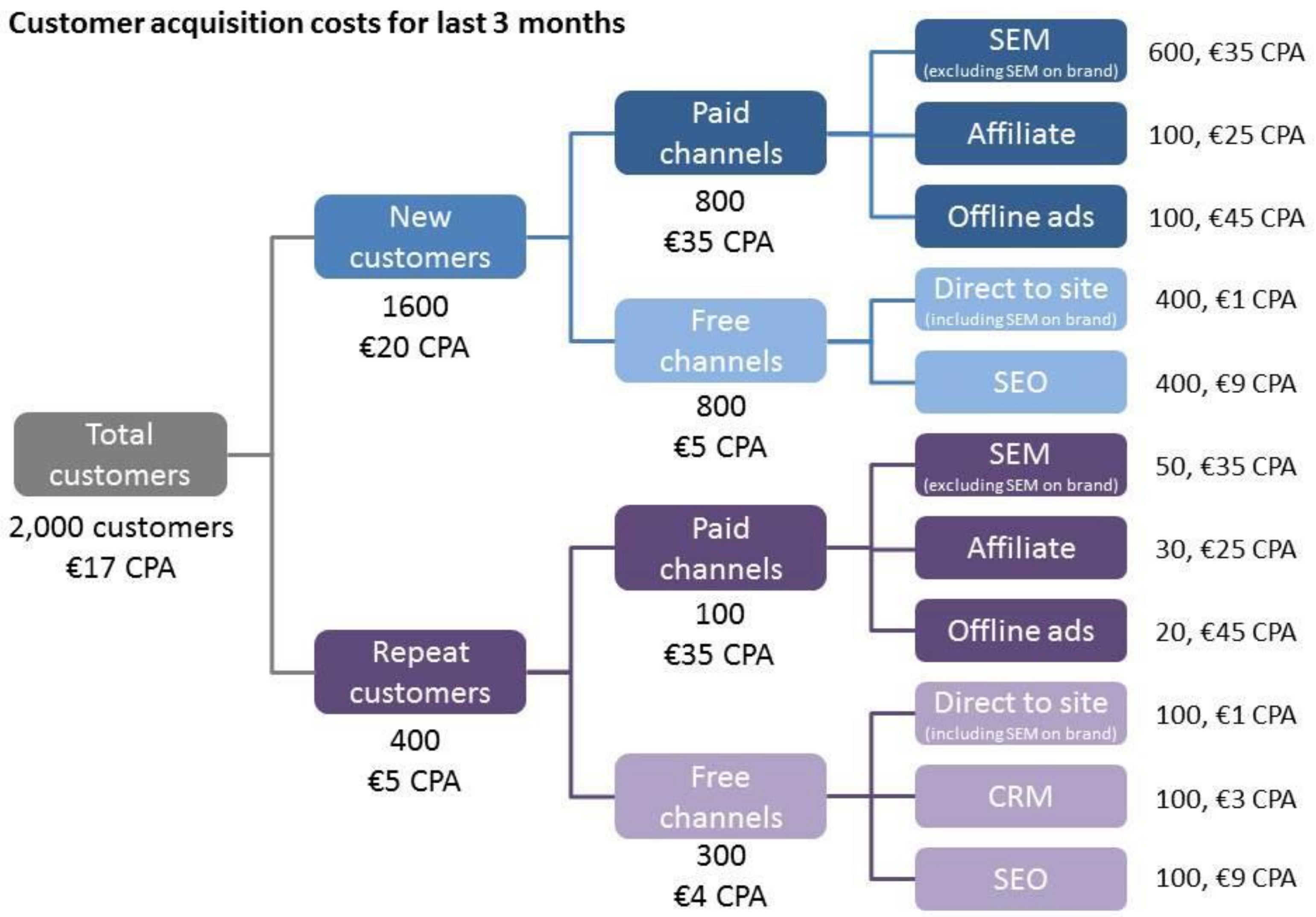
SEO Secrets: Reverse-Engineering Google's Algorithm

- Quality is King: Google likes to provide search results that have value. Whether it's a tutorial video, a funny article, or an amazing photo series, it's possible to create something of quality that benefits your brand's bottom line, and, benefits your audience.

- Who is sharing?: If Google sees that a lot people are sharing your content, it considers it as potential QUALITY. And, if Google sees that related sites are sharing your content, it considers it as likely RELEVANT.

- Keywords are Queen: Keywords, or search terms, are what Google uses to index your site. While keywords do not hold the same weight as they used to, they are still necessary.

Breaking Down Customer Acquisition Cost

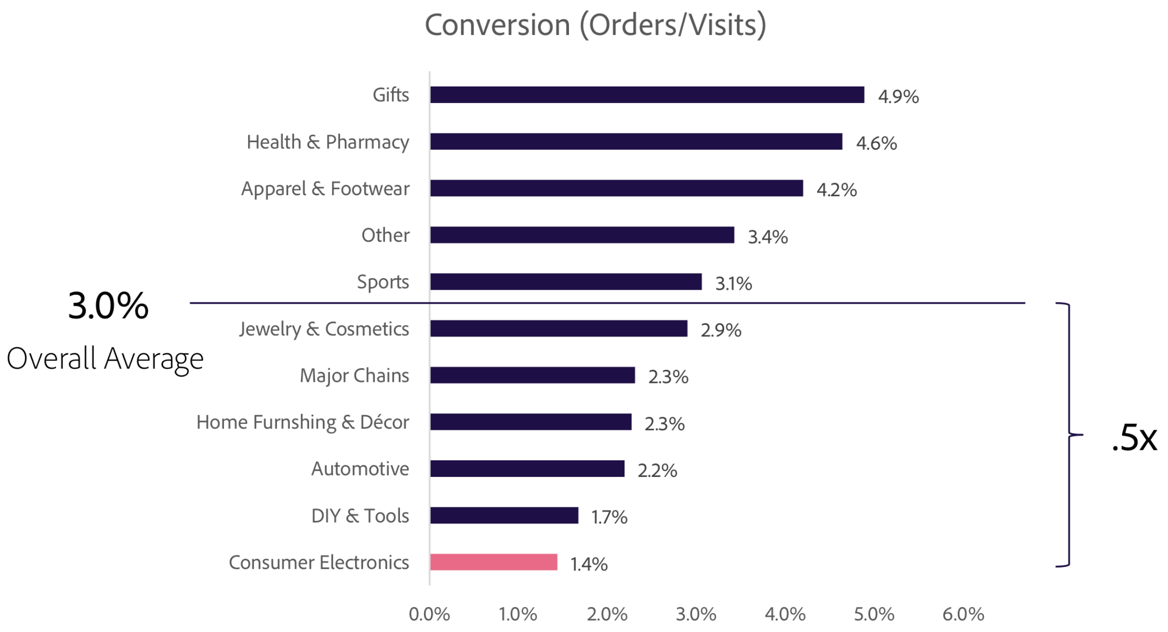


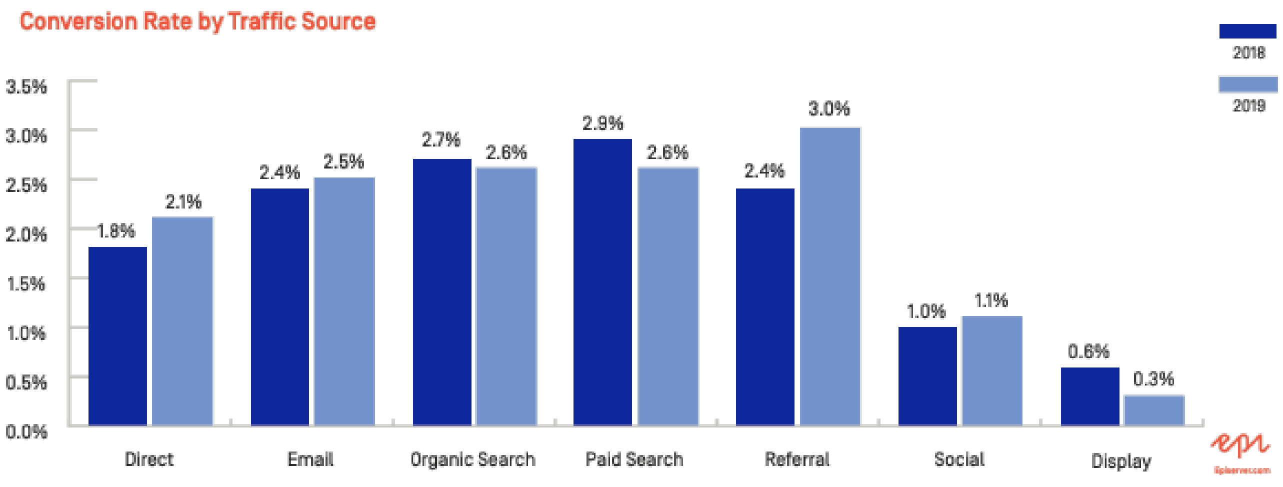
https://medium.com/venture-capital-growth-hacking/understanding-customer-acquisition-costs-74aec7538b4d

Conversion Rate

$$\text{CONVERSION RATE} = \frac{\text{NUMBER OF ORDERS ON A WEBSITE}}{\text{NUMBER OF SESSIONS}} * 100\%$$

E-Commerce Conversion Rate





https://www.smartinsights.com/ecommerce/ecommerce-analytics/ecommerce-conversion-rates/

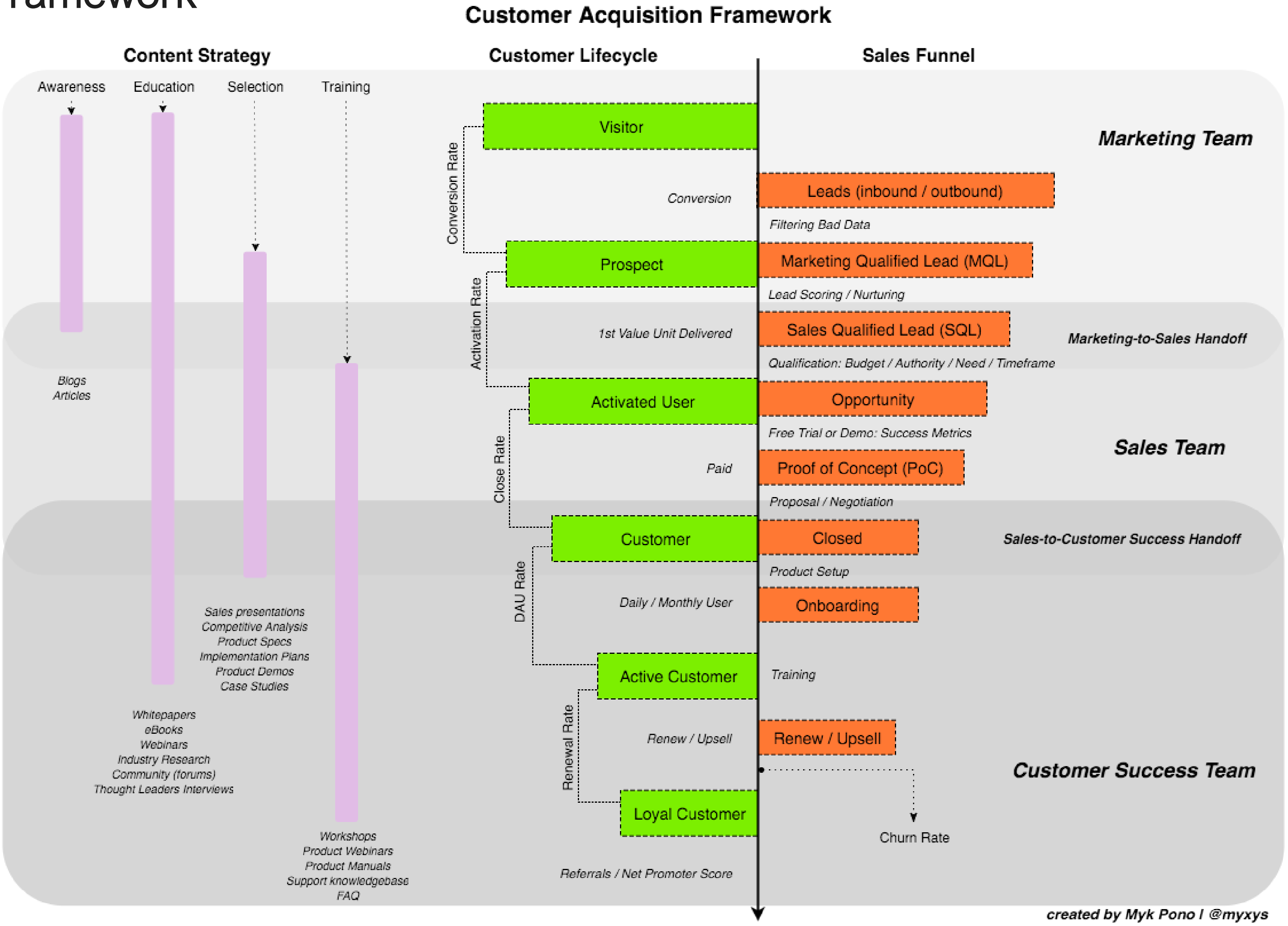
Customer On-boarding

Customer Onboarding

- It’s almost impossible to move ahead without clear onboarding, which helps the user understand how the product works. So, to grow sustainably and scalability.
- The fact is that after day one, your product retention will fall absurdly if nothing is done. A product that does not help the user to understand what it is for has very low engagement.
- And that goes for any product, be it physical or digital. Try using something that is not designed for ease of use and self-explanatory. Remember when you bought something new and read the instruction manual before using the product.

https://uxdesign.cc/user-onboarding-principles-and-best-practices-bb1ad848c553

Customer Acquisition Framework



https://medium.com/the-marketing-playbook/how-to-track-customer-acquisition-9d04b903535

Lead Scoring

https://towardsdatascience.com/a-true-end-to-end-ml-example-lead-scoring-f5b52e9a3c80

Predictive Lead Scoring

- Lead scoring is one of the key marketing automation tasks for targeting the right customers and prospects and improve the productivity and efficiency of marketing and sales teams.

- Predictive lead scoring takes the traditional lead scoring approach to the next level by applying big data and machine learning algorithms to evaluate the key behaviors of existing customers and prospects and rank them against a scale that can distinguish customers and prospects who are more likely to convert, retain, or buy from the company’s products and services.

https://medium.com/hackernoon/how-to-win-new-customers-with-proven-customer-acquisition-strategies-ea7203897354

Case Study: Lead Scoring with Logistic Regression

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

- The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.

- When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals.

- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Case Study: Lead Scoring with Logistic Regression

- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Case Study: Lead Scoring with Logistic Regression

Variables

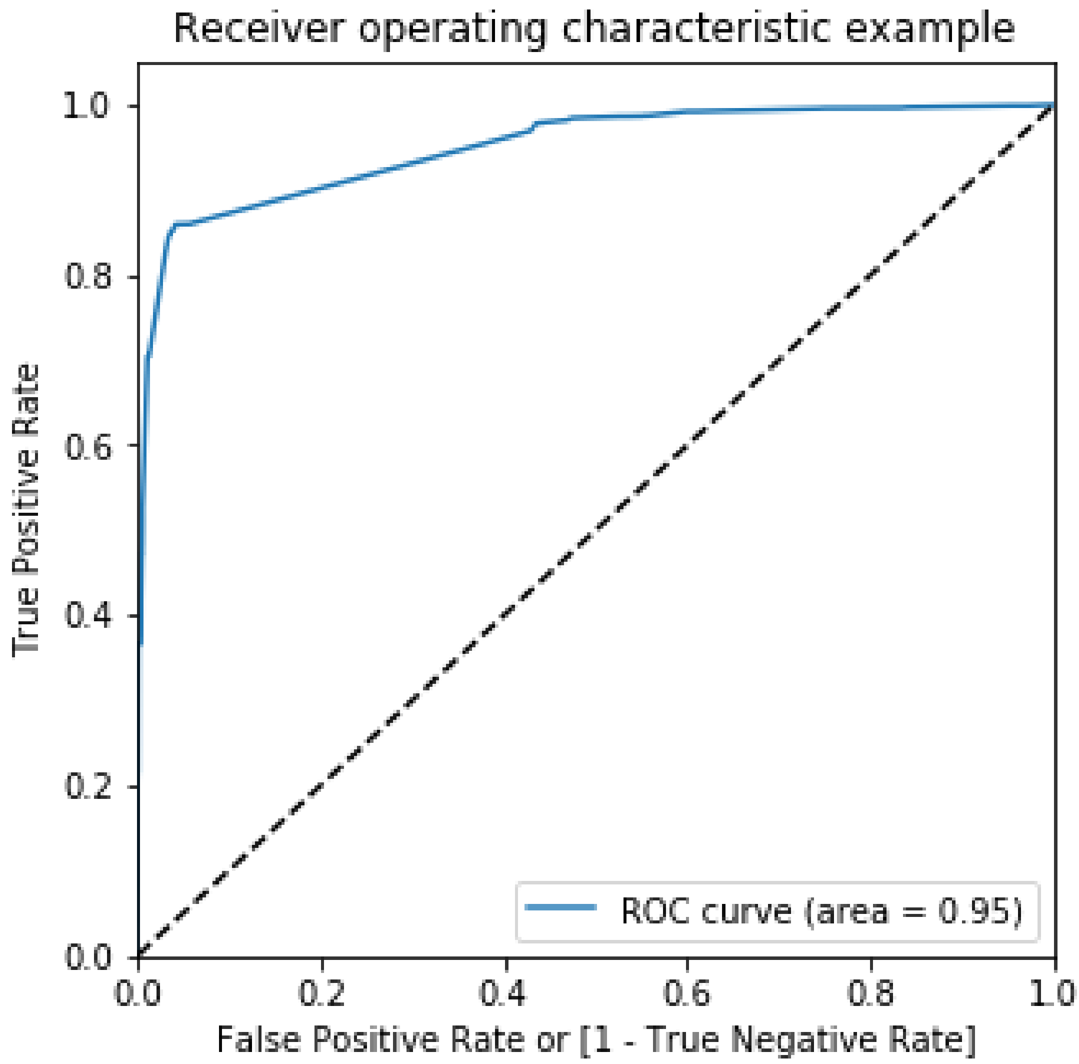
- Prospect ID
- Lead Number
- Lead Origin
- Lead Source
- Do Not Email
- Do Not Call
- Converted
- TotalVisits
- Total Time Spent on Website
- Page Views Per Visit
- Last Activity
- Country
- Specialization
- How did you hear about X Education
- What is your current occupation
- What matters most to you in choosing this course
- Search
- Magazine
- Newspaper Article
- X Education Forums
- Newspaper
- Digital Advertisement
- Through Recommendations
- Receive More Updates About Our Courses
- Tags
- Lead Quality
- Update me on Supply Chain Content
- Get updates on DM Content
- Lead Profile
- City
- Asymmetrique Activity Index
- Asymmetrique Profile Index
- Asymmetrique Activity Score
- Asymmetrique Profile Score
- I agree to pay the amount through cheque
- a free copy of Mastering The Interview
- Last Notable Activity

Results

```
# Predicted      not_churn      churn
# Actual
# not_churn          3270        365
# churn               579        708
```

<u>Training Set</u>
Recall = 0.8516
Precision = 0.9332

<u>Test Set</u>
Recall = 0.8433
Precision = 0.9429



Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 8
Customer Segmentation

BigQuery ML for Customer Segmentation

https://cloud.google.com/bigquery-ml/docs/bigqueryml-intro

BigQuery ML

- BigQuery ML lets you create and execute machine learning models in BigQuery using standard SQL queries. BigQuery ML democratizes machine learning by letting SQL practitioners build models using existing SQL tools and skills. BigQuery ML increases development speed by eliminating the need to move data.

Supported models in BigQuery ML

- Linear regression for forecasting; for example, the sales of an item on a given day. Labels are real-valued (they cannot be +/- infinity or NaN).
- Binary logistic regression for classification; for example, determining whether a customer will make a purchase. Labels must only have two possible values.
- Multiclass logistic regression for classification. These models can be used to predict multiple possible values such as whether an input is "low-value," "medium-value," or "high-value." Labels can have up to 50 unique values. In BigQuery ML, multiclass logistic regression training uses a multinomial classifier with a cross-entropy loss function.
- K-means clustering for data segmentation; for example, identifying customer segments. K-means is an unsupervised learning technique, so model training does not require labels nor split data for training or evaluation.
- Matrix Factorization for creating product recommendation systems. You can create product recommendations using historical customer behavior, transactions, and product ratings and then use those recommendations for personalized customer experiences.
- Time series for performing time-series forecasts. You can use this feature to create millions of time series models and use them for forecasting. The model automatically handles anomalies, seasonality, and holidays.
- Boosted Tree for creating XGBoost based classification and regression models.
- Deep Neural Network (DNN) for creating TensorFlow based Deep Neural Networks for classification and regression models.
- AutoML Tables to create best-in-class models without feature engineering or model selection. AutoML Tables searches through a variety of model architectures to decide the best model.
- TensorFlow model importing. This feature lets you create BigQuery ML models from previously trained TensorFlow models, then perform prediction in BigQuery ML.

k-means clustering model

```
CREATE MODEL
  `mydataset.mymodel`
OPTIONS
  ( MODEL_TYPE='KMEANS',
    NUM_CLUSTERS=3,
    KMEANS_INIT_METHOD='RANDOM') AS
SELECT
  *
FROM
  `mydataset.mytable`
```

https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create#training_a_k-means_model

Examples

```
CREATE OR REPLACE MODEL
 `nida-workshop.SUPERMARKET.THANACHART_CLUSTERS`
OPTIONS(model_type='kmeans', num_clusters=7)
AS (
  SELECT
  COUNT(DISTINCT BASKET_ID) AS TOTAL_VISIT,
  SUM(SPEND) AS TOTAL_SPEND
  FROM `nida-workshop.SUPERMARKET.TRANSACTIONS_2STORES`
  WHERE CUST_CODE IS NOT NULL
  GROUP BY CUST_CODE
)
```

Results

Centroid Id	Count	TOTAL_VISIT	TOTAL_SPEND
1	173	161.2254	2,561.9177
2	483	22.7847	263.8769
3	3822	1.9223	12.8228
4	66	340.9091	4,340.9502
5	271	106.2620	1,142.7635
6	957	9.1860	86.2810
7	328	48.3110	617.0677

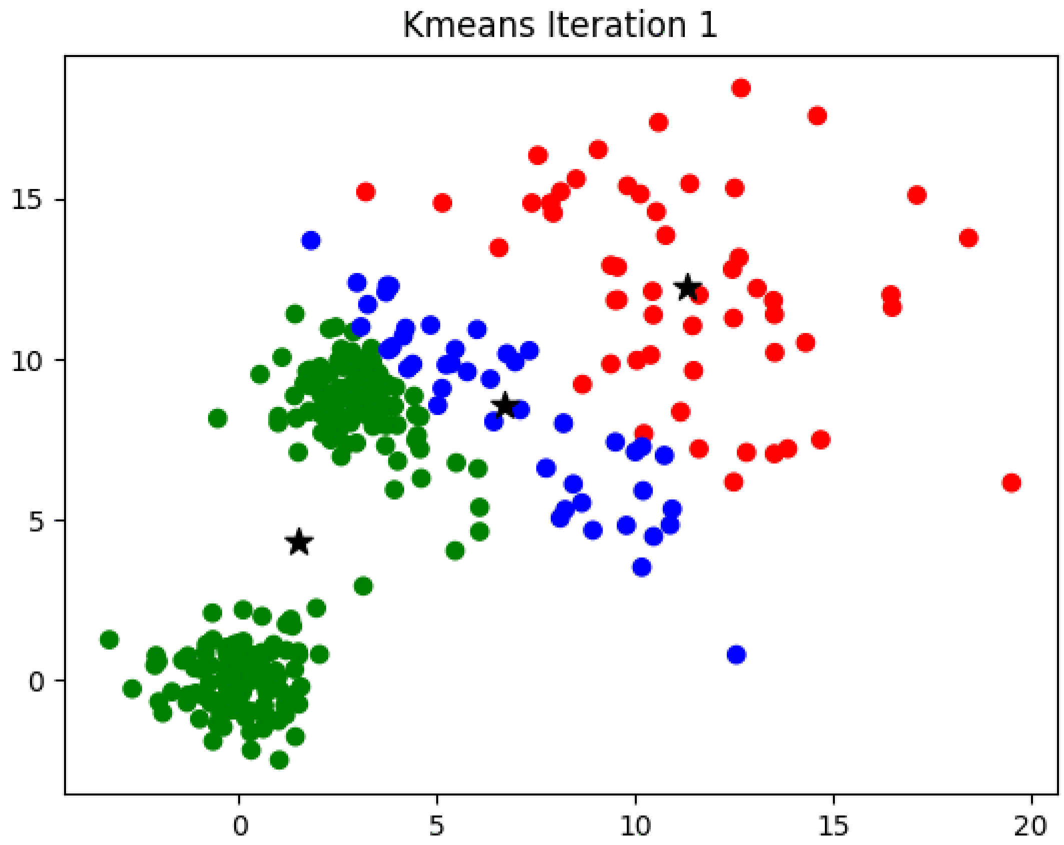
k-means clustering model

```
SELECT
  * EXCEPT(nearest_centroids_distance)
FROM
  ML.PREDICT( MODEL `project-name.dataset-name.model-name` ,
    (
  SELECT STATEMENT
))
```

Examples

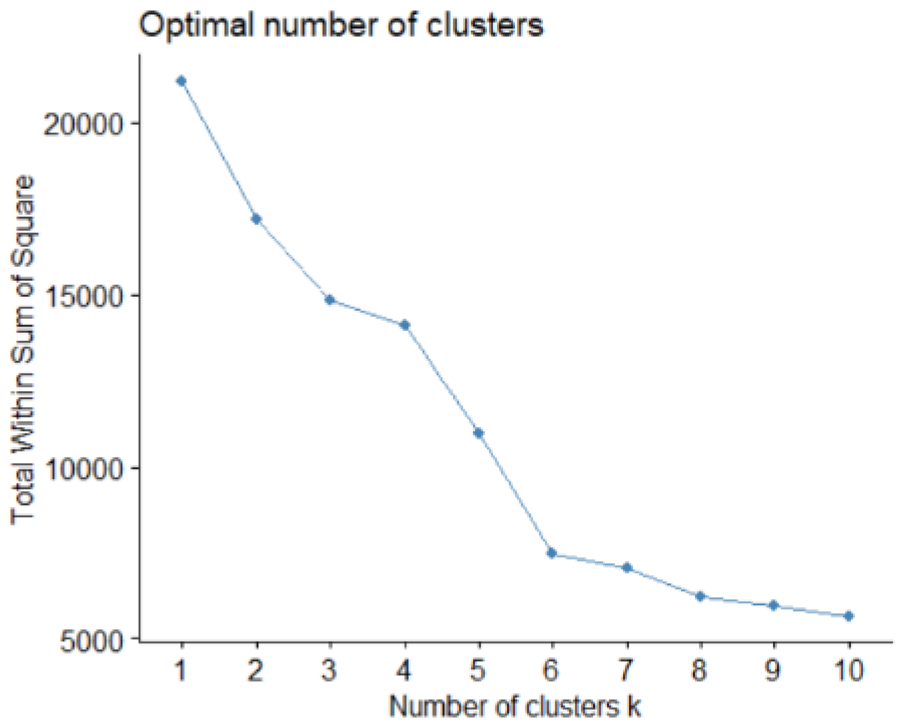
```
SELECT
 * EXCEPT(nearest_centroids_distance)
FROM
 ML.PREDICT( MODEL `nida-workshop.SUPERMARKET.THANACHART_CLUSTERS`,
  (
   SELECT
   CUST_CODE,
   COUNT(DISTINCT BASKET_ID) AS TOTAL_VISIT,
   SUM(SPEND) AS TOTAL_SPEND
   FROM `nida-workshop.SUPERMARKET.TRANSACTIONS_2STORES`
   WHERE CUST_CODE IS NOT NULL
   GROUP BY CUST_CODE
  ))
```

K-Means Clustering



https://sandipanweb.wordpress.com/2017/03/19/hard-soft-clustering-with-k-means-weighted-k-means-and-gmm-em/

How many K?



Elbow Method

Choose the optimal k using WSS, from the plot we choose 6 clusters as that's where the elbow sort of breaks.

Silhouette Index

$S(i) = (b(i) – a(i)) / (\max \{ (a(i), b(i)) \}$

- a(i) is the average dissimilarity of i^{th} object to all other objects in the same cluster
- b(i) is the average dissimilarity of i^{th} object with all objects in the closest cluster.

If the Silhouette index value is high (ranging from -1 to 1), the object is well-matched to its own cluster and poorly matched to neighboring clusters.

Other clustering validity index

1. Dunn Index
2. DB Index
3. CS Index
4. I- Index
5. XB or Xie Beni Index

Finding importance variables differentiating clusters

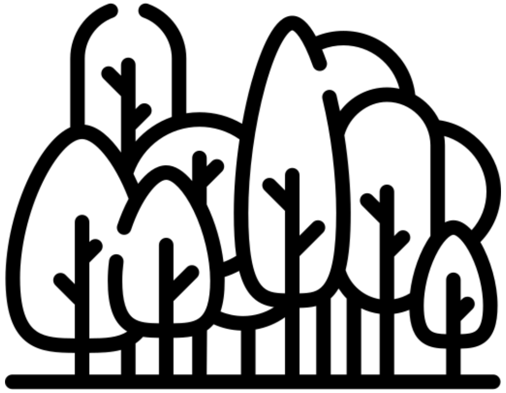
Input of K-Means

Cust_ID	X1	X2	X3	...	Xn

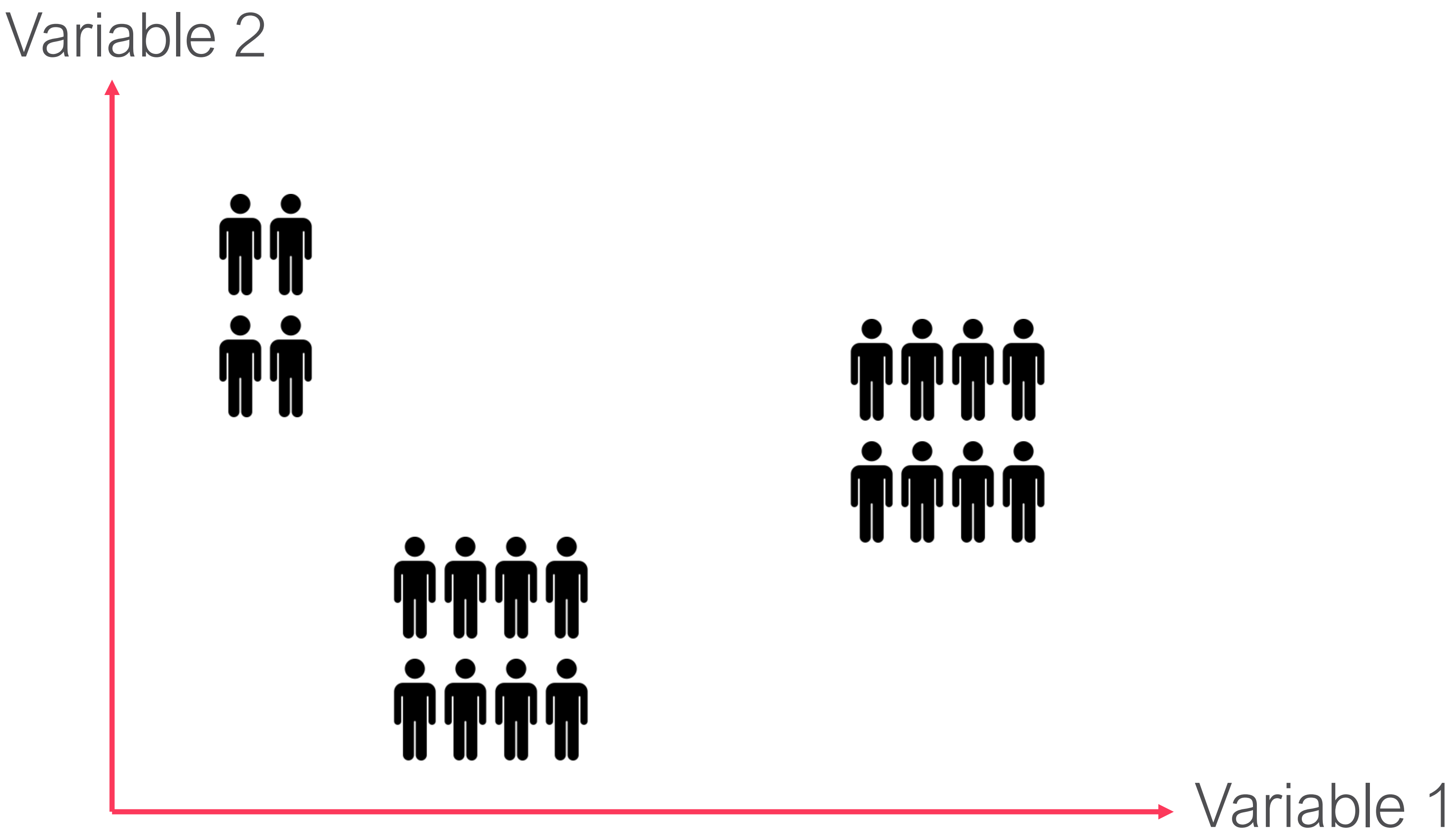
+

Output of K-Means

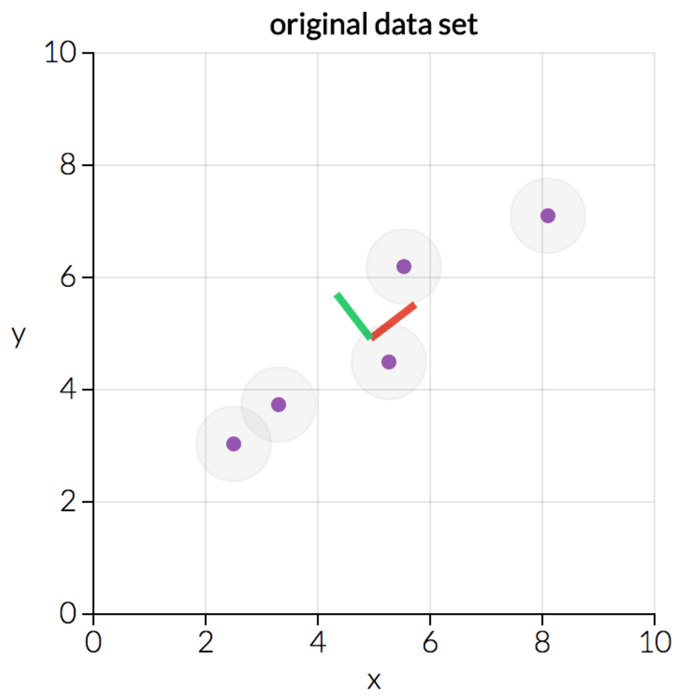
Cluster

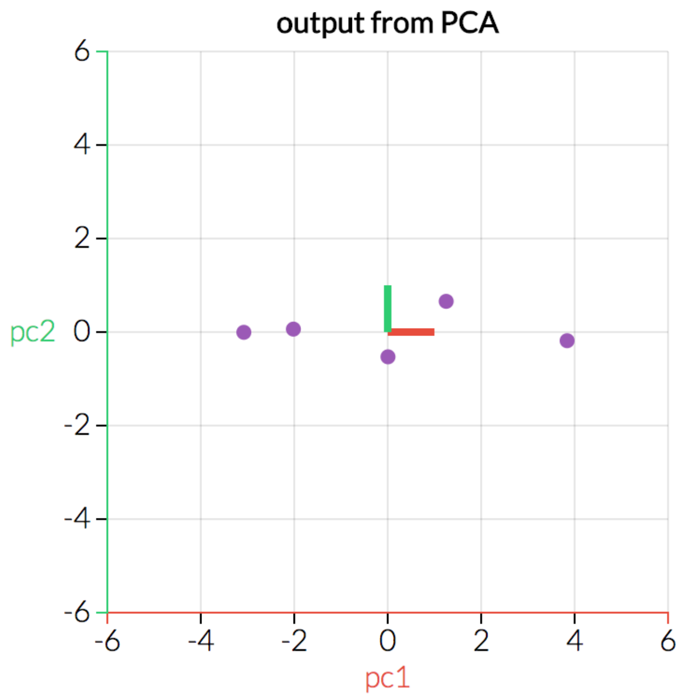


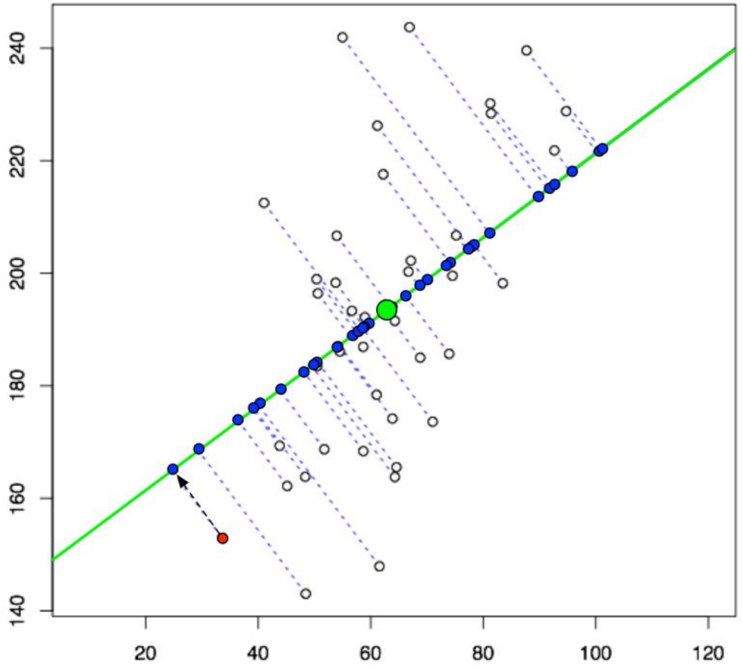
Visualizing K-Means Results



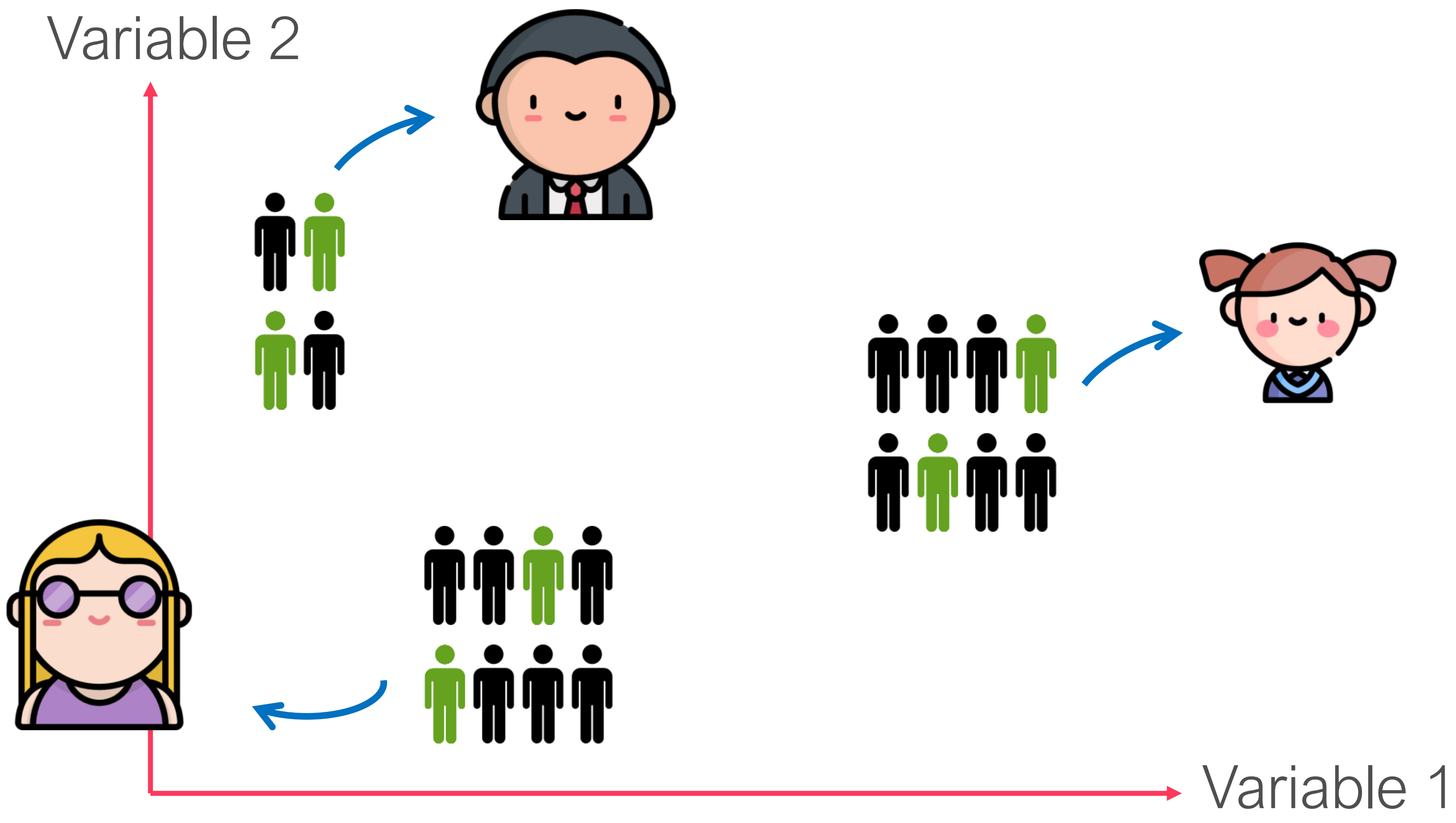
Using PCA to visualize clustering results







Interpreting K-Means Results



Interpreting K-Means Results

NEED TO WIN BACK

- <1k customers
  22% of the base
- Lowest Total Revenue
  6% revenue contribution
- Lowest Recency
- 1.67% Return Transactions
- Avg 7 days between each transaction

RECOMMENDATIONS

- Coupon discounts on products frequently bought by segment
- Identify cause for churn. Is there a new competitor in town?
  - Run analysis on the cost vs revenue of offering promotions.

SPORADIC VISITORS

- ~700 customers
  16% of the base
- 2nd Highest Total Revenue
  12% revenue contribution
- Average Recency
- 1.31% Return Transactions
- Avg 87 days between each transaction (longest)

RECOMMENDATIONS

- Increase engagement to encourage more visits:
  - Daily promotional specials
  - Coupons issued post purchase that expire in a short time frame
  - Test marketing efforts with a control group, to confirm whether more frequent visits will result in higher total revenue, or rather simply maintain the total revenue level (higher frequency, lower value per transaction)

HIGH-FLYERS

- ~1,500 customers
  30% of the base
- Highest Total Revenue
  70% revenue contribution
- High Recency
- 2.54% Return Transactions
- Avg 34 days between each transaction

RECOMMENDATIONS

- Marketing should be targeted to maintain their loyalty.
- Communications on new products can be rolled out to keep the retailer brand top of mind.

RUN-OF-THE-MILL

- ~1,200 customers
  30% customer base
- 2nd Lowest Total Revenue
  12% revenue contribution
- Average Recency
- 1.39% Return Transactions
- Avg 6 days between each transaction (shortest)

RECOMMENDATIONS

- Upsell strategies to encourage higher spends for each visit.
- Additional product association analysis can also be run to create an algorithm to recommend additional products.
- Spend stretch discounts eg. “Buy for amount X to get discount Y”.

Creating value from clustering results

	Action 1	Action 2	Action 3	Action 4
Cluster 1	✓	✓		
Cluster 2		✓	✓	
Cluster 3	✓			
Cluster 4		✓		✓
Cluster 5			✓	✓
Cluster 6	✓		✓	
Cluster 7	✓			✓

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 9
Cross-selling Models and Recommendation Systems

Product Recommendation using Market Basket Analysis

Business Case Overview

- The data belongs to a bakery called "The Bread Basket", located in the historic center of Edinburgh. This bakery presents a refreshing offer of Argentine and Spanish products.

- Data set containing 15.010 observations and more than 6.000 transactions from a bakery. The data set contains the following columns:

  Date. Categorical variable that tells us the date of the transactions (YYYY-MM-DD format). The column includes dates from 30/10/2016 to 09/04/2017.

  Time. Categorical variable that tells us the time of the transactions (HH:MM:SS format).

  Transaction. Quantitative variable that allows us to differentiate the transactions. The rows that share the same value in this field belong to the same transaction, that's why the data set has less transactions than observations.

  Item. Categorical variable with the products.

Apriori Algorithm - mlxtend

Apriori is a popular algorithm for extracting frequent itemsets with applications in association rule learning. The apriori algorithm has been designed to operate on databases containing transactions, such as purchases by customers of a store. An itemset is considered as "frequent" if it meets a user-specified support threshold. For instance, if the support threshold is set to 0.5 (50%), a frequent itemset is defined as a set of items that occur together in at least 50% of all transactions in the database.

```
On Anaconda
pip install mlxtend

On Google Colab
!pip install mlxtend
```

Prepare Data

```
import pandas as pd

df = pd.read_csv('BreadBasket_DMS.csv')
```

Read CSV (comma-separated) file into DataFrame

Check Missing Data

```
missing = df.isnull().sum()
print(missing)
```

List unique items sold

```
df.Item.unique()
```

```
array(['Bread', 'Scandinavian', 'Hot chocolate', 'Jam', 'Cookies',
       'Muffin', 'Coffee', 'Pastry', 'Medialuna', 'Tea', 'NONE', 'Tartine',
       'Basket', 'Mineral water', 'Farm House', 'Fudge', 'Juice',
       "Ella's Kitchen Pouches", 'Victorian Sponge', 'Frittata',
       'Hearty & Seasonal', 'Soup', 'Pick and Mix Bowls', 'Smoothies',
       'Cake', 'Mighty Protein', 'Chicken sand', 'Coke',
       'My-5 Fruit Shoot', 'Focaccia', 'Sandwich', 'Alfajores', 'Eggs',
       'Brownie', 'Dulce de Leche', 'Honey', 'The BART', 'Granola',
       'Fairy Doors', 'Empanadas', 'Keeping It Local', 'Art Tray',
       'Bowl Nic Pitt', 'Bread Pudding', 'Adjustment', 'Truffles',
       'Chimichurri Oil', 'Bacon', 'Spread', 'Kids biscuit', 'Siblings',
       'Caramel bites', 'Jammie Dodgers', 'Tiffin', 'Olum & polenta',
       'Polenta', 'The Nomad', 'Hack the stack', 'Bakewell',
       'Lemon and coconut', 'Toast', 'Scone', 'Crepes', 'Vegan mincepie',
       'Bare Popcorn', 'Muesli', 'Crisps', 'Pintxos', 'Gingerbread syrup',
       'Panatone', 'Brioche and salami', 'Afternoon with the baker',
       'Salad', 'Chicken Stew', 'Spanish Brunch',
       'Raspberry shortbread sandwich', 'Extra Salami or Feta', 'Duck egg',
       'Baguette', "Valentine's card", 'Tshirt', 'Vegan Feast', 'Postcard',
       'Nomad bag', 'Chocolates', 'Coffee granules ',
       'Drinking chocolate spoons ', 'Christmas common', 'Argentina Night',
       'Half slice Monster ', 'Gift voucher', 'Cherry me Dried fruit',
       'Mortimer', 'Raw bars', 'Tacos/Fajita'], dtype=object)
```

Cleansing NONE items

```
df.loc[df["Item"]=='NONE'].shape[0]

df = df[df["Item"] != 'NONE']
```

Plot top 10 products

```
import numpy as np
import matplotlib.pyplot as plt

df_for_top10_Items=df['Item'].value_counts().head(10)
Item_array= np.arange(len(df_for_top10_Items))

plt.figure(figsize=(15,5))
Items_name=['coffee','bread','tea','cake','pastry','sandwich','me
dialuna','hot chocolate','cookies','brownie']
plt.bar(Item_array,df_for_top10_Items.iloc[:])
plt.xticks(Item_array,Items_name)
plt.title('Top 10 most selling items')
plt.show()
```

Prepare data for Apriori using one-hot encoding

```
hot_encoded_df=df.groupby(['Transaction','Item'
])['Item'].count().unstack().reset_index().fill
na(0).set_index('Transaction')
```

Transform transaction items into matrix of transaction-item

Turn float into integer

```
def encode_units(x):
    if x <= 0:
        return 0
    if x >= 1:
        return 1
hot_encoded_df = hot_encoded_df.applymap(encode_units)
```

Determine product association using apriori

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

frequent_itemsets = apriori(hot_encoded_df,
min_support=0.01, use_colnames=True)
```

List out rules

```
rules = association_rules(frequent_itemsets,
metric="lift", min_threshold=1)
rules.head(10)
```

Filter rules

```
rules[ (rules['lift'] >= 1) &
       (rules['confidence'] >= 0.5) ]
```

Visualize rules

```
import matplotlib.pyplot as plt
import networkx as nx

fig, ax=plt.subplots(figsize=(10,4))
GA=nx.from_pandas_edgelist(rules,source='antece
dents',target='consequents')
nx.draw(GA,with_labels=True)
plt.show()
```

networkx 2.0

nx.from_pandas_edgelist

Earlier versions

nx.from_pandas_dataframe

Product Recommendation using Collaborative Filtering

Prepare Data

```
item_item_matrix = pd.DataFrame(index=hot_encoded_df.columns,columns=hot_encoded_df.columns)
```

Create item-to-item matrix

Item	Adjustment	Afternoon with the baker	Alfajores	Argentina Night	Art Tray	Bacon	Baguette	Bakewell	Bare Popcorn	Basket	...	The BART	The Nomad	Tiffin	Toast	Truffles
Item																
Adjustment	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
Afternoon with the baker	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
Alfajores	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
Argentina Night	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN

Calculate Cosine Similarity

```
from scipy.spatial.distance import cosine

for i in range(0,len(item_item_matrix.columns)) :
    # Loop through the columns for each column
    for j in range(0,len(item_item_matrix.columns)) :
      # Fill in placeholder with cosine similarities
      item_item_matrix.iloc[i,j] = 1 -cosine(hot_encoded_df.iloc[:,i],hot_encoded_df.iloc[:,j])
```

Convert results into list of rules

```
links = item_item_matrix.rename_axis('related item',
                                     axis='columns').stack().reset_index()
links.columns = ['item', 'related item','value']
```

Filter rules

```
links_filtered=links.loc[ (links['value'] > 0.1) &
                          (links['item'] != links['related item']) ]
```

Visualize rules

```
import matplotlib.pyplot as plt
import networkx as nx

fig, ax=plt.subplots(figsize=(10,4))
GA=nx.from_pandas_edgelist(links_filtered,source='item',target='related item')
nx.draw(GA,with_labels=True)
plt.show()
```

Clustering for Customer Segmentation

Prepare Data

```
import pandas as pd

from google.colab import auth
auth.authenticate_user()
```

Prepare Data

```
project_id = 'nida-workshop'
sql ='''
SELECT CUST_CODE,
SUM(SPEND) AS TOTAL_SALES,
COUNT(DISTINCT BASKET_ID) AS TOTAL_VISIT
FROM `nida-workshop.SUPERMARKET.TRANSACTIONS`
WHERE CUST_CODE IS NOT NULL
GROUP BY CUST_CODE'''
df = pd.io.gbq.read_gbq(sql , project_id=project_id,
verbose=False, dialect="standard")
```

Visualize distribution of variables

```
import matplotlib.pyplot as plt
import seaborn as sns

for i, col in enumerate(df.columns[1:]):
    plt.figure(i)
    sns.distplot(df[col])

plt.show()
```

Standardize data

```
from sklearn.preprocessing import StandardScaler

df_std = pd.DataFrame(StandardScaler().fit_transform(df[df.columns[1:]]))
df_std.head(5)
```

Find optimum k using elbow method

```
from sklearn.cluster import KMeans
# Use the Elbow method to find a good number of clusters using Within-
Cluster-Sum-of-Squares (WCSS)
wcss = []
N = range(1, 15)
for i in N:
    kmeans = KMeans(n_clusters=i)
    kmeans.fit_predict(df_std)
    wcss.append(kmeans.inertia_)

plt.plot(N, wcss, 'bo-')
plt.title("Computing Within-Cluster-Sum-of-Squares (WCSS) for KMeans++")
plt.xlabel("Number of clusters")
plt.ylabel("WCSS")
plt.show()
```

K-means clustering

```
#let's choose k = 7
cluster = KMeans(n_clusters=7)
df['cluster'] = cluster.fit_predict(df_std)
```

Determine average behaviors of each cluster

```
dfCluster = df.groupby('cluster', as_index=False).mean()
dfCluster['NO_CUST'] = df[['cluster','CUST_ID']].groupby('cluster').count()

dfCluster
```

Visualize Clustering Results – Scatter Plot

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure()
fig, ax = plt.subplots()
ax.scatter(dfCluster['TOTAL_SALES'], dfCluster['TOTAL_VISIT'],
s=dfCluster['NO_CUST'], alpha = 0.5)
ax.set_xlabel("TOTAL_SALES")
ax.set_ylabel("TOTAL_VISIT")

for i, txt in enumerate(dfCluster['cluster']):
  ax.annotate(txt, (dfCluster['TOTAL_SALES'][i],
dfCluster["TOTAL_VISIT"][i]), horizontalalignment='middle',
verticalalignment='middle')

plt.show()
```

Visualize Clustering Results – Box Plot

```
for i, col in enumerate(df.columns[1:-1]):
    sns.boxplot(x="cluster", y =col, data=df)
    plt.show()
```

Employ RandomForest to determine important variables

```
from sklearn.ensemble import RandomForestClassifier

predictors = df.iloc[:,1:-1]
targets = df.iloc[:,-1:]

scaler = StandardScaler()
predictors_scaled = scaler.fit_transform(predictors)

classifier = RandomForestClassifier(n_estimators = 10, criterion = 'entropy', random_state = 42)
classifier.fit(predictors_scaled, targets.values.ravel())
```

List important variables

```
importances = list(classifier.feature_importances_)

feature_list = list(predictors.columns)
feature_importances = [(feature, round(importance, 2)) for feature,
importance in zip(feature_list, importances)]
feature_importances = sorted(feature_importances, key = lambda x: x[1],
reverse = True)
[print('Variable: {:20} Importance: {}'.format(*pair)) for pair in
feature_importances];
```

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 10

Up-selling Analysis and Customer Response Models


CASHBOOK
Pay Invoice
Card Type
Name on card
Card number
Expiry date
Zip / Poctal code

Up-selling

Up-selling

- Up-selling is a sales technique where a customer is induced to purchase more expensive items, upgrades to a premium option, buy add-ons in order to make a more profitable sale.

Choose the RIGHT Upsell

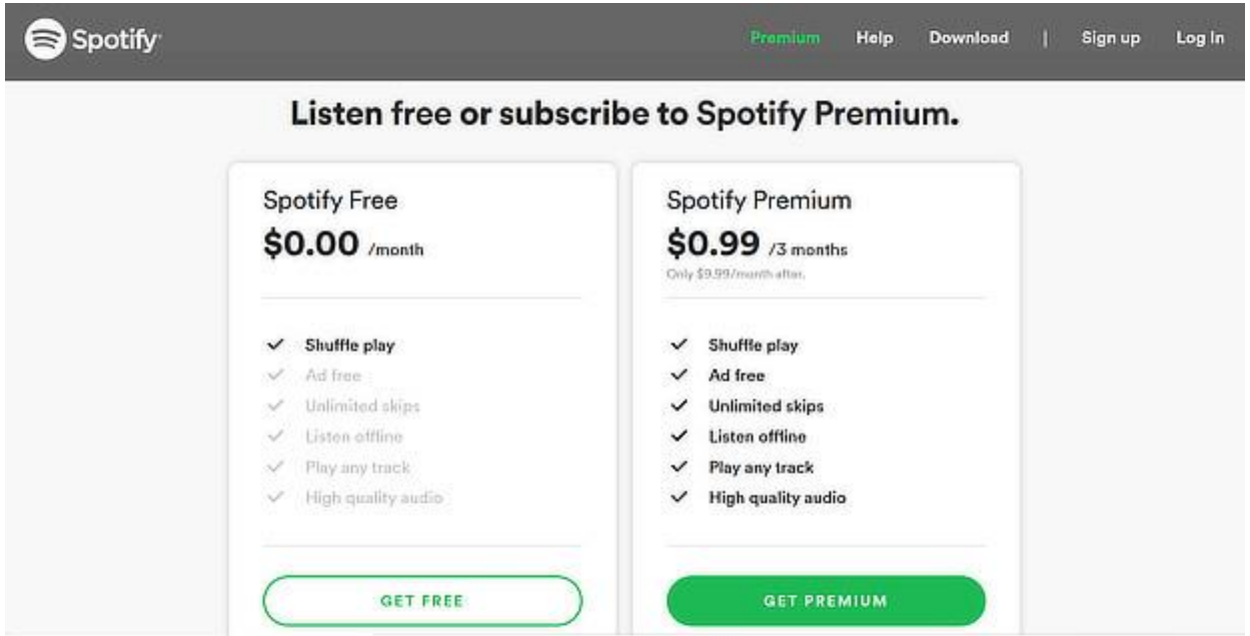
- There might be several options for up-selling for example, version upgrade, product protection/insurance, extended service period, or customization.

https://optinmonster.com/upselling-tips-and-examples/

Get the Language Right

- Language is an important part of all successful marketing, as we've seen before in our article on conversion-boosting power words. But it's even more important to persuade your customers when you're aiming for the upsell.

- Some of the best upselling examples get this right by making visitors imagine how they could make use of the upgrade, or by triggering the fear of missing out (FOMO). This is psychologically proven to help make more sales.

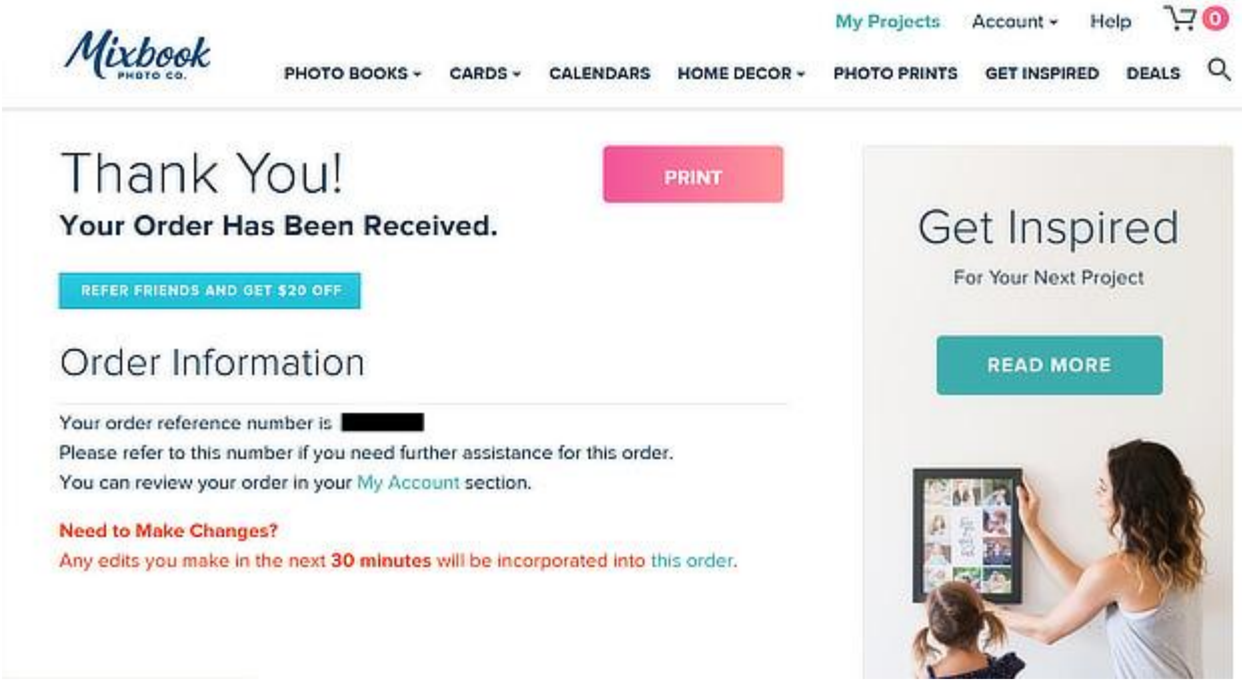


https://optinmonster.com/upselling-tips-and-examples/

Upsell After the Purchase

- While many upsells are offered on product or pricing pages, these aren't the only places to show off your upsell offer. In fact, you can offer an upsell even after customers have completed their purchase.

- For example, when placing an order with Mixbook, you get a limited time to edit or upgrade your order before it ships.



https://optinmonster.com/upselling-tips-and-examples/

How to calculate the level of spending for up-selling

Campaign Response Model

https://www.kaggle.com/regivm/retailtransactiondata

Importing libraries and datasets

```
[1] import numpy as np
    import pandas as pd
    import datetime as dt
    import numpy as np
    import matplotlib.pyplot as plt
    import xgboost as xgb
    import seaborn as sns
    from sklearn.model_selection import train_test_split
    from sklearn.metrics import precision_score, recall_score, f1_score, roc_auc_score, accuracy_score, classification_report, roc_curve, auc
    from imblearn.under_sampling import RandomUnderSampler
    from imblearn.over_sampling import RandomOverSampler
    from imblearn.over_sampling import SMOTE
    from xgboost import plot_importance
    from sklearn.linear_model import LogisticRegression
    from sklearn.metrics import confusion_matrix
```

```
[2] df_response = pd.read_csv('Retail_Data_Response.csv')
    df_transactions = pd.read_csv('Retail_Data_Transactions.csv', parse_dates=['trans_date'])
```

```
[3] df_response.head()
```

	customer_id	response
0	CS1112	0
1	CS1113	0
2	CS1114	1
3	CS1115	1
4	CS1116	1

```
[4] df_transactions.head()
```

	customer_id	trans_date	tran_amount
0	CS5295	2013-02-11	35
1	CS4768	2015-03-15	39
2	CS2122	2013-02-26	52
3	CS1217	2011-11-16	99
4	CS1850	2013-11-20	78

```
[5] print(df_transactions['trans_date'].min())
    print(df_transactions['trans_date'].max())
```

```
2011-05-16 00:00:00
2015-03-16 00:00:00
```

Data Preparation

```
[6] ## since the last date of the data is 16 March 2015, the campaign date is assumed to be 17 March 2015
    ## RFM model will be used to predict campaign response. Recency is calculated

    campaign_date = dt.datetime(2015,3,17)
    df_transactions['recent']= campaign_date - df_transactions['trans_date']
    df_transactions['recent'].astype('timedelta64[D]')
    df_transactions['recent']=df_transactions['recent'] / np.timedelta64(1, 'D')
    df_transactions.head()
```

	customer_id	trans_date	tran_amount	recent
0	CS5295	2013-02-11	35	764.0
1	CS4768	2015-03-15	39	2.0
2	CS2122	2013-02-26	52	749.0
3	CS1217	2011-11-16	99	1217.0
4	CS1850	2013-11-20	78	482.0

```
[7] ## create data set with RFM variables

    df_rfm = df_transactions.groupby('customer_id').agg({'recent': lambda x:x.min(),                  # Recency
                                                          'customer_id': lambda x: len(x),             # Frequency
                                                          'tran_amount': lambda x: x.sum()}) # Monetary Value

    df_rfm.rename(columns={'recent': 'recency',
                           'customer_id': 'frequency',
                           'tran_amount': 'monetary_value'}, inplace=True)
```

```
[8] df_rfm = df_rfm.reset_index()
    df_rfm.head()
```

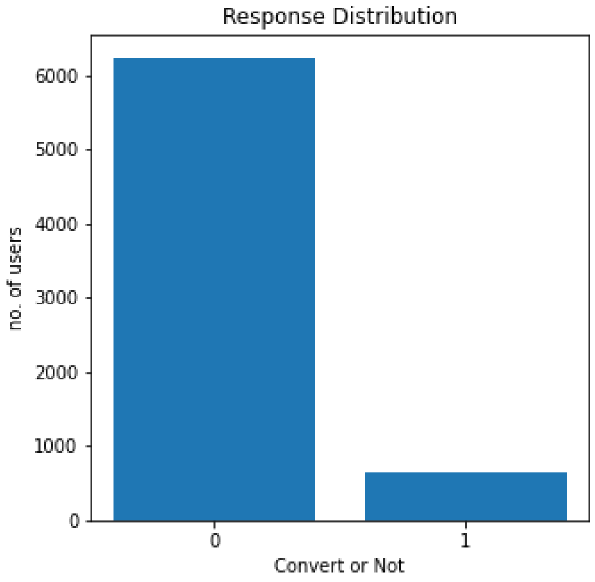
	customer_id	recency	frequency	monetary_value
0	CS1112	62.0	15	1012
1	CS1113	36.0	20	1490

Calculating response rate

```
[9] response_rate = df_response.groupby('response').agg({'customer_id': lambda x: len(x)}).reset_index()
    response_rate.head()
```

	response	customer_id
0	0	6237
1	1	647

```
[10] plt.figure(figsize=(5,5))
     x=range(2)
     plt.bar(x,response_rate['customer_id'])
     plt.xticks(response_rate.index)
     plt.title('Response Distribution')
     plt.xlabel('Convert or Not')
     plt.ylabel('no. of users')
     plt.show()

     ## data is imbalanced
```



[11]
```
## merging two data sets

df_modeling = pd.merge(df_response,df_rfm)
df_modeling.head()
```

	customer_id	response	recency	frequency	monetary_value
0	CS1112	0	62.0	15	1012
1	CS1113	0	36.0	20	1490
2	CS1114	1	33.0	19	1432
3	CS1115	1	12.0	22	1659
4	CS1116	1	204.0	13	857

Creating train and test dataset

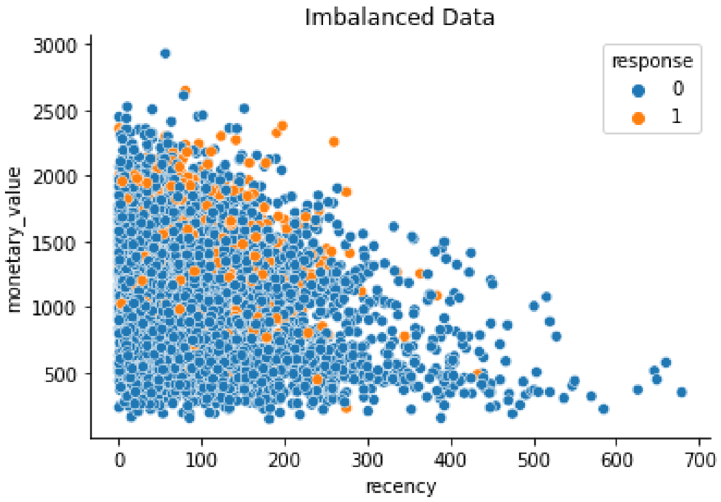
[12]
```
## spliting dataframe into X and y

X = df_modeling.drop(columns=['response','customer_id'])
y = df_modeling['response']
```

[13]
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)

print("Number transactions X_train dataset: ", X_train.shape)
print("Number transactions y_train dataset: ", y_train.shape)
print("Number transactions X_test dataset: ", X_test.shape)
print("Number transactions y_test dataset: ", y_test.shape)
```

```
Number transactions X_train dataset:  (4818, 3)
Number transactions y_train dataset:  (4818,)
Number transactions X_test dataset:  (2066, 3)
Number transactions y_test dataset:  (2066,)
```

[14]
```
sns.scatterplot(data=df_modeling, x='recency', y='monetary_value', hue='response')
sns.despine()
plt.title("Imbalanced Data")
```

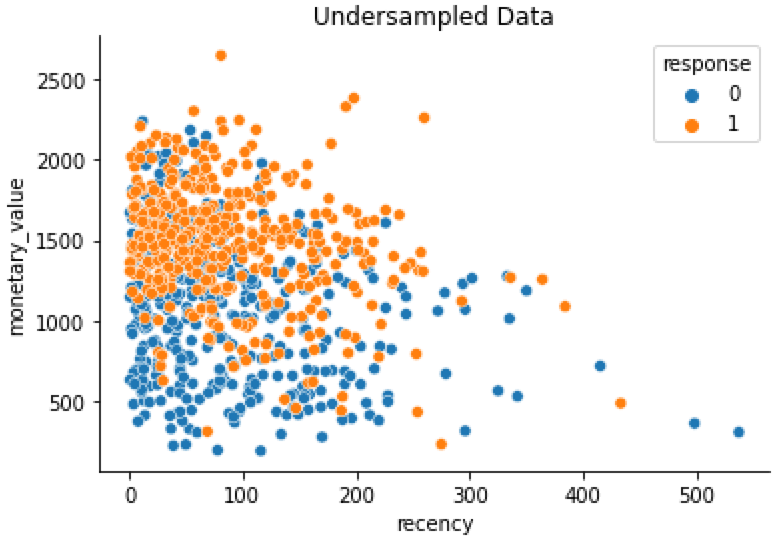
```
Text(0.5, 1.0, 'Imbalanced Data')
```



Fixing imbalanced with Undersampling

```
[15] rus = RandomUnderSampler(random_state=0)
     rus.fit(X_train, y_train)
     X_undersampled, y_undersampled = rus.fit_sample(X_train, y_train)
     df_undersampled = pd.concat([pd.DataFrame(data=X_undersampled),pd.DataFrame(data=y_undersampled)], axis=1, sort=False)
     df_undersampled.columns= ['recency', 'frequency', 'monetary_value', 'response']

     sns.scatterplot(data=df_undersampled, x='recency', y='monetary_value', hue='response')
     sns.despine()
     plt.title("Undersampled Data")
```

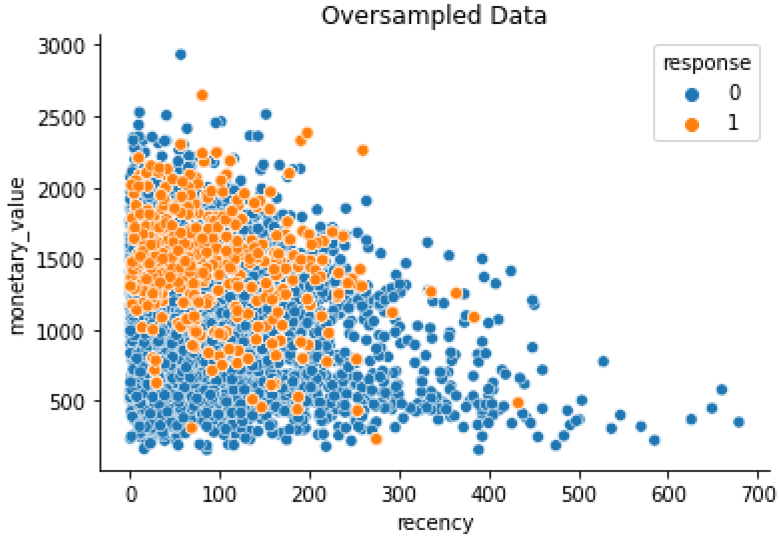
```
/usr/local/lib/python3.6/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function safe_indexing is deprecated;
  warnings.warn(msg, category=FutureWarning)
Text(0.5, 1.0, 'Undersampled Data')
```



Fixing imbalanced with Oversampling

```
[16] ros = RandomOverSampler(random_state=0)
     ros.fit(X_train, y_train)
     X_oversampled, y_oversampled = ros.fit_sample(X_train, y_train)
     df_oversampled = pd.concat([pd.DataFrame(data=X_oversampled),pd.DataFrame(data=y_oversampled)], axis=1, sort=False)
     df_oversampled.columns= ['recency', 'frequency', 'monetary_value', 'response']

     sns.scatterplot(data=df_oversampled, x='recency', y='monetary_value', hue='response')
     sns.despine()
     plt.title("Oversampled Data")
```

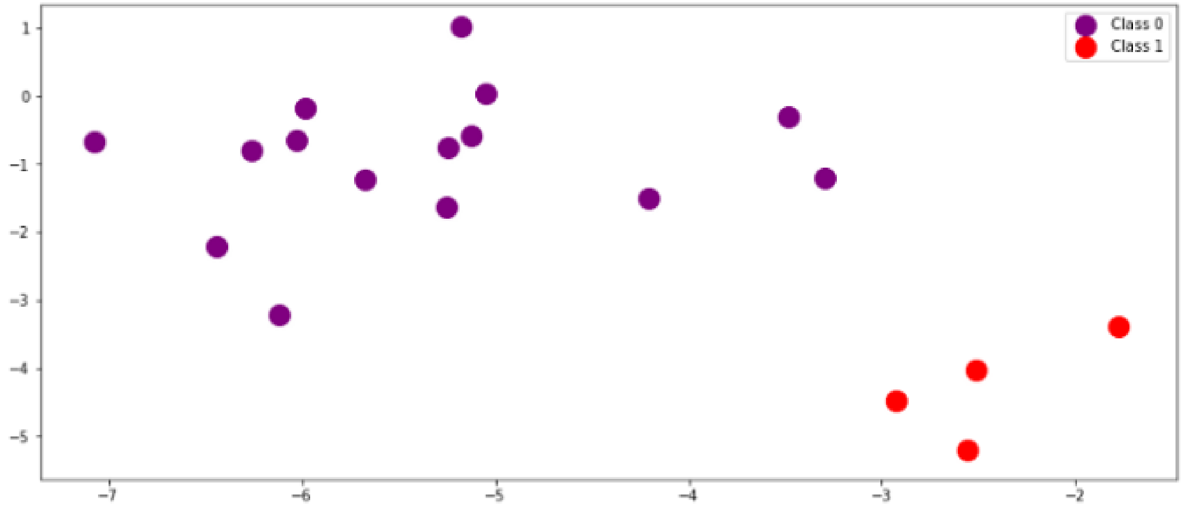
```
/usr/local/lib/python3.6/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function safe_indexing is deprecated; :
  warnings.warn(msg, category=FutureWarning)
Text(0.5, 1.0, 'Oversampled Data')
```

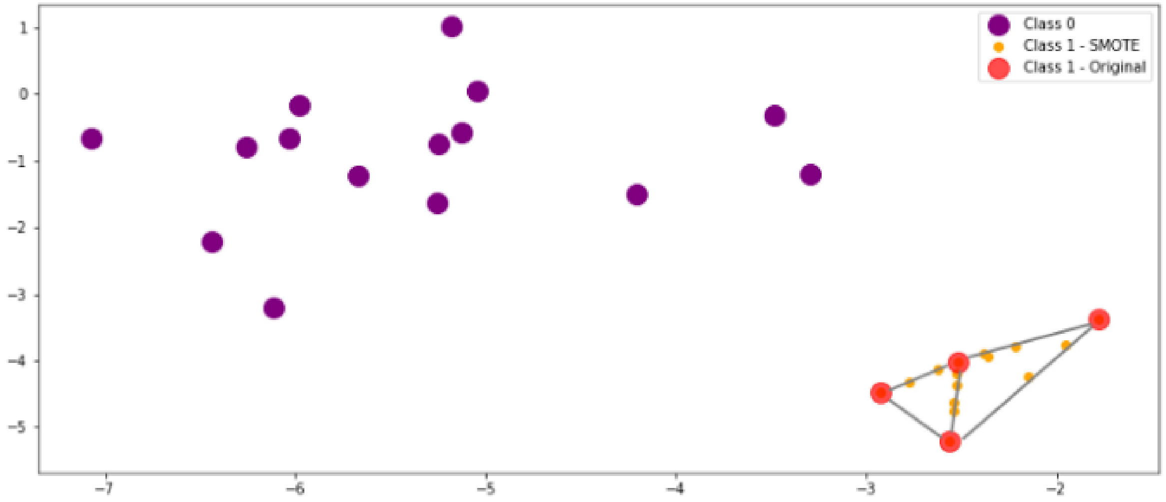


Synthetic Minority Over-sampling Technique (SMOTE)

- SMOTE is a technique based on nearest neighbors judged by Euclidean Distance between data points in feature space.
- SMOTE-NC is for mixed nominal and continuous features
- Borderline-SMOTE only makes synthetic data along the decision boundary between the two classes.

SMOTE proceeds by joining the points of the minority class with line segments and then places artificial points on these lines.

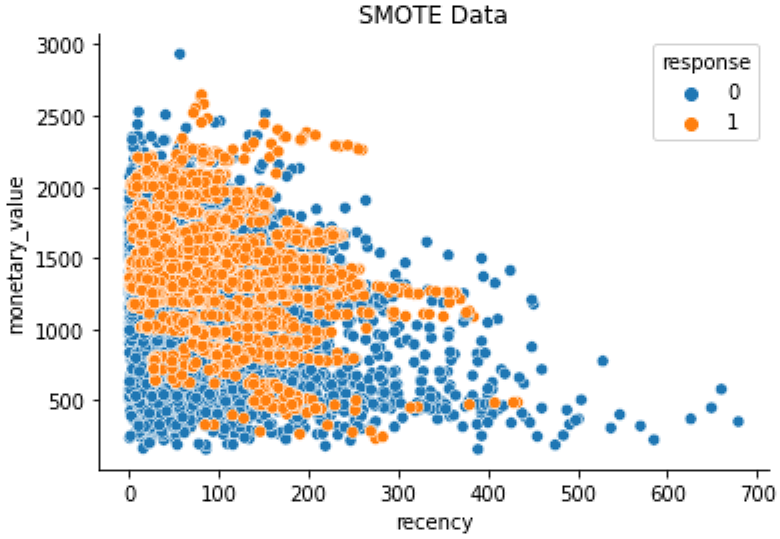




Fixing imbalanced with SMOTE

```
[17] sm = SMOTE(random_state=0)
     sm.fit(X_train, y_train)
     X_SMOTE, y_SMOTE = sm.fit_sample(X_train, y_train)
     df_SMOTE = pd.concat([pd.DataFrame(data=X_SMOTE),pd.DataFrame(data=y_SMOTE)], axis=1, sort=False)
     df_SMOTE.columns= ['recency', 'frequency', 'monetary_value', 'response']

     sns.scatterplot(data=df_SMOTE, x='recency', y='monetary_value', hue='response')
     sns.despine()
     plt.title("SMOTE Data")
```

```
/usr/local/lib/python3.6/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function safe_indexing is deprecated;
  warnings.warn(msg, category=FutureWarning)
Text(0.5, 1.0, 'SMOTE Data')
```



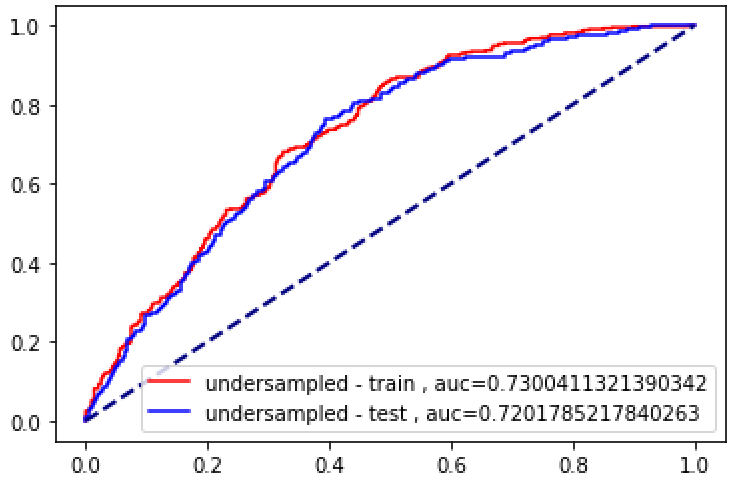
```python
[19] print('logistic regression model - undersampled')
     logreg = LogisticRegression(solver='liblinear', class_weight='balanced')

     predicted_y = []
     expected_y = []

     logreg_model_under = logreg.fit(X_undersampled, y_undersampled)
     predictions = logreg_model_under.predict(X_undersampled)
     predicted_y.extend(predictions)
     expected_y.extend(y_undersampled)
     report_train = classification_report(expected_y, predicted_y)
     print('training set')
     print(report_train)

     predicted_y = []
     expected_y = []
     predictions = logreg_model_under.predict(X_test)
     predicted_y.extend(predictions)
     expected_y.extend(y_test)
     report_test = classification_report(expected_y, predicted_y)
     print('test set')
     print(report_test)
```

```
logistic regression model - undersampled
training set
              precision    recall  f1-score   support

           0       0.69      0.62      0.65       429
           1       0.65      0.72      0.69       429

    accuracy                           0.67       858
   macro avg       0.67      0.67      0.67       858
weighted avg       0.67      0.67      0.67       858

test set
              precision    recall  f1-score   support

           0       0.96      0.60      0.74      1848
           1       0.18      0.76      0.30       218

    accuracy                           0.62      2066
   macro avg       0.57      0.68      0.52      2066
weighted avg       0.87      0.62      0.69      2066
```

```python
print(logreg_model_under.coef_, logreg_model_under.intercept_)
```

```
[[ 0.00368411 -0.03621944  0.00246363]] [-2.80649227]
```

```
[20] y_score_train = logreg_model_under.decision_function(X_undersampled)
     fpr_train, tpr_train, _ = roc_curve(y_undersampled,  y_score_train)
     auc_train = roc_auc_score(y_undersampled, y_score_train)
     plt.plot(fpr_train,tpr_train, color='red', label='undersampled - train , auc='+str(auc_train))

     y_score_test = logreg_model_under.decision_function(X_test)
     fpr_test, tpr_test, _ = roc_curve(y_test,  y_score_test)
     auc_test = roc_auc_score(y_test, y_score_test)
     plt.plot(fpr_test,tpr_test, color='Blue', label='undersampled - test , auc='+str(auc_test))

     plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
     plt.legend(loc=4)
     plt.show()
```



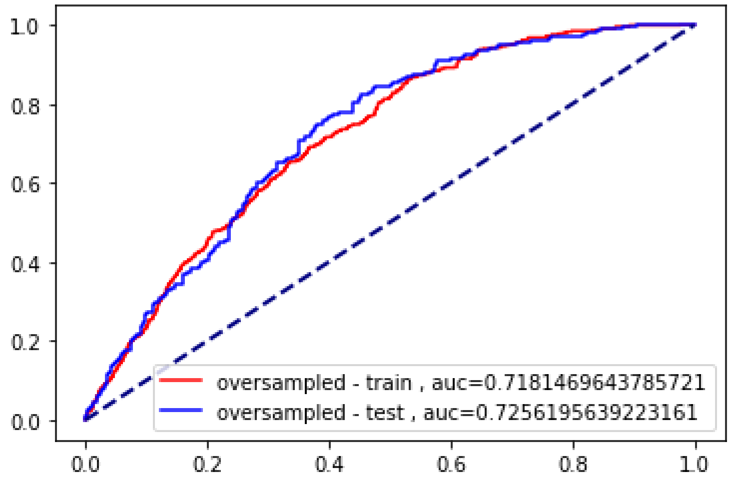
```
[21] print('logistic regression model - oversampled')
     logreg = LogisticRegression(solver='liblinear', class_weight='balanced')
     predicted_y = []
     expected_y = []

     logreg_model_over = logreg.fit(X_oversampled, y_oversampled)
     predictions = logreg_model_over.predict(X_oversampled)
     predicted_y.extend(predictions)
     expected_y.extend(y_oversampled)
     report_train = classification_report(expected_y, predicted_y)
     print('training set')
     print(report_train)

     predicted_y = []
     expected_y = []
     predictions = logreg_model_over.predict(X_test)
     predicted_y.extend(predictions)
     expected_y.extend(y_test)
     report_test = classification_report(expected_y, predicted_y)
     print('test set')
     print(report_test)
```

```
logistic regression model - oversampled
training set
              precision    recall  f1-score   support

           0       0.67      0.62      0.65      4389
           1       0.65      0.69      0.67      4389

    accuracy                           0.66      8778
   macro avg       0.66      0.66      0.66      8778
weighted avg       0.66      0.66      0.66      8778

test set
              precision    recall  f1-score   support

           0       0.95      0.63      0.76      1848
           1       0.19      0.72      0.30       218

    accuracy                           0.64      2066
   macro avg       0.57      0.68      0.53      2066
weighted avg       0.87      0.64      0.71      2066
```

```
print(logreg_model_over.coef_, logreg_model_under.intercept_)
```

```
[[0.00415889 0.02516016 0.00181298]] [-2.80649227]
```

```
[22] y_score_train = logreg_model_over.decision_function(X_oversampled)
     fpr_train, tpr_train, _ = roc_curve(y_oversampled,  y_score_train)
     auc_train = roc_auc_score(y_oversampled, y_score_train)
     plt.plot(fpr_train,tpr_train, color='red', label='oversampled - train , auc='+str(auc_train))

     y_score_test = logreg_model_over.decision_function(X_test)
     fpr_test, tpr_test, _ = roc_curve(y_test,  y_score_test)
     auc_test = roc_auc_score(y_test, y_score_test)
     plt.plot(fpr_test,tpr_test, color='Blue', label='oversampled - test , auc='+str(auc_test))

     plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
     plt.legend(loc=4)
     plt.show()
```



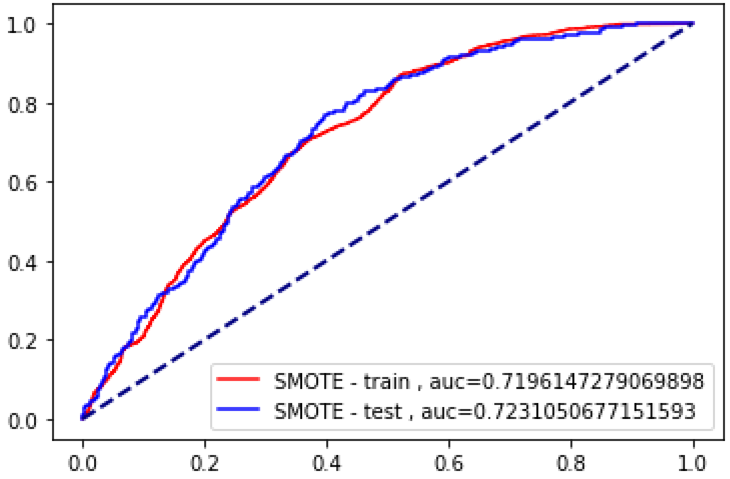
```
[23] print('logistic regression model - SMOTE')
     logreg = LogisticRegression(solver='liblinear', class_weight='balanced')
     predicted_y = []
     expected_y = []

     logreg_model_SMOTE = logreg.fit(X_SMOTE, y_SMOTE)
     predictions = logreg_model_SMOTE.predict(X_SMOTE)
     predicted_y.extend(predictions)
     expected_y.extend(y_SMOTE)
     report_train = classification_report(expected_y, predicted_y)
     print('training set')
     print(report_train)

     predicted_y = []
     expected_y = []
     predictions = logreg_model_SMOTE.predict(X_test)
     predicted_y.extend(predictions)
     expected_y.extend(y_test)
     report_test = classification_report(expected_y, predicted_y)
     print('test set')
     print(report_test)
```

```
logistic regression model - SMOTE
training set
              precision    recall  f1-score   support

           0       0.68      0.62      0.65      4389
           1       0.65      0.71      0.68      4389

    accuracy                           0.67      8778
   macro avg       0.67      0.67      0.67      8778
weighted avg       0.67      0.67      0.67      8778

test set
              precision    recall  f1-score   support

           0       0.95      0.63      0.75      1848
           1       0.18      0.71      0.29       218

    accuracy                           0.64      2066
   macro avg       0.57      0.67      0.52      2066
weighted avg       0.87      0.64      0.71      2066
```

```
print(logreg_model_SMOTE.coef_, logreg_model_under.intercept_)
```

```
[[0.00371859 0.00684075 0.00207781]] [-2.80649227]
```

```
[24] y_score_train = logreg_model_SMOTE.decision_function(X_SMOTE)
     fpr_train, tpr_train, _ = roc_curve(y_SMOTE,  y_score_train)
     auc_train = roc_auc_score(y_SMOTE, y_score_train)
     plt.plot(fpr_train,tpr_train, color='red', label='SMOTE - train , auc='+str(auc_train))

     y_score_test = logreg_model_SMOTE.decision_function(X_test)
     fpr_test, tpr_test, _ = roc_curve(y_test,  y_score_test)
     auc_test = roc_auc_score(y_test, y_score_test)
     plt.plot(fpr_test,tpr_test, color='Blue', label='SMOTE - test , auc='+str(auc_test))

     plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
     plt.legend(loc=4)
     plt.show()
```

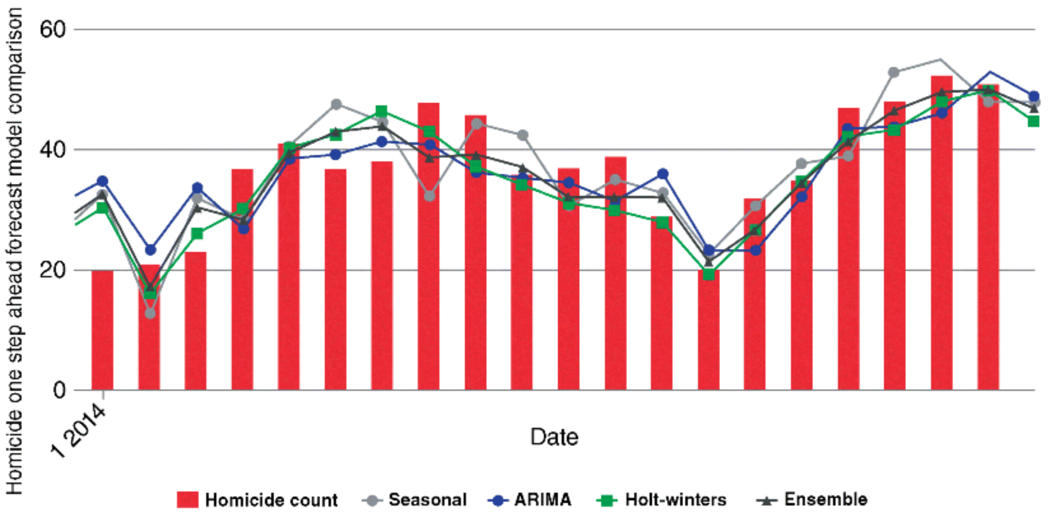


Ensemble Methods

- One of the most powerful techniques in machine learning is the ensemble, or a model federation that combines predictions from several (or many) machine learning algorithms. As a general rule, ensembles improve predictive performance at the cost of making the process for producing predictions far less interpretable.

Simple averaging

- The simplest approach to ensembles involves averaging the outputs of several different models to form a consensus prediction or classification.



https://towardsdatascience.com/predict-customer-churn-with-r-9e62357d47b4

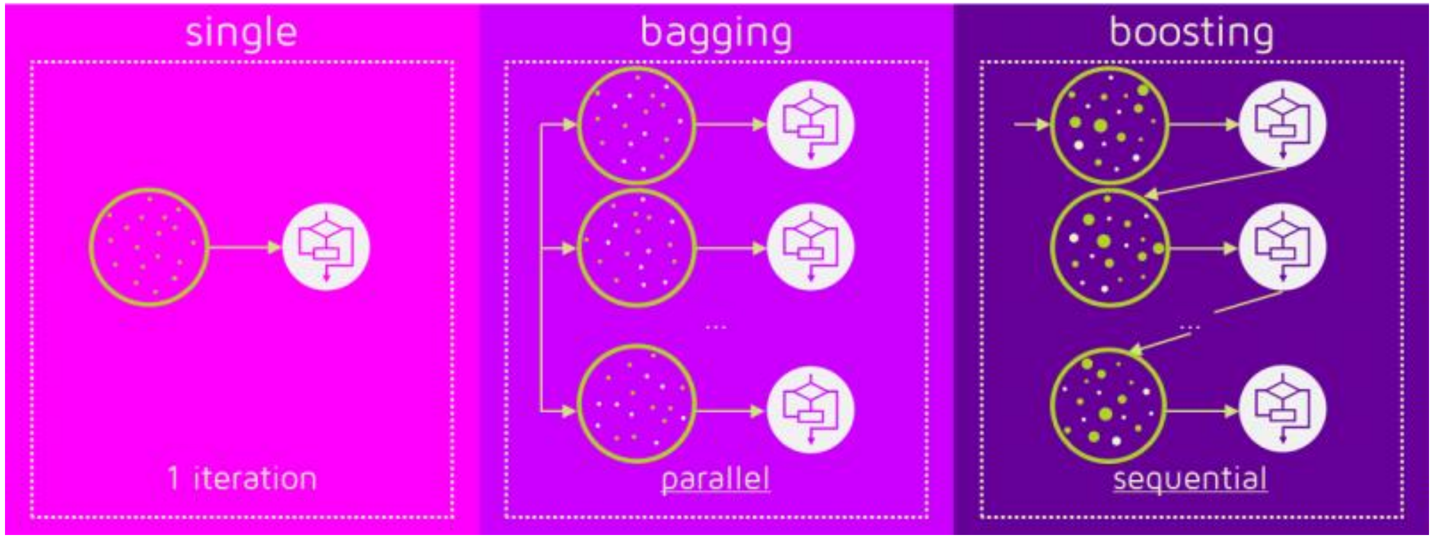
Ensemble Methods

Bagging

- Bootstrap aggregating, more frequently referred to as bagging, involves iteratively sampling from the training data set to create many training samples of the same size. A model is built from each sample, resulting in an ensemble of many prediction or classification models that have been fit with the same machine learning algorithm using slightly different training data sets. Example is Random Forest.

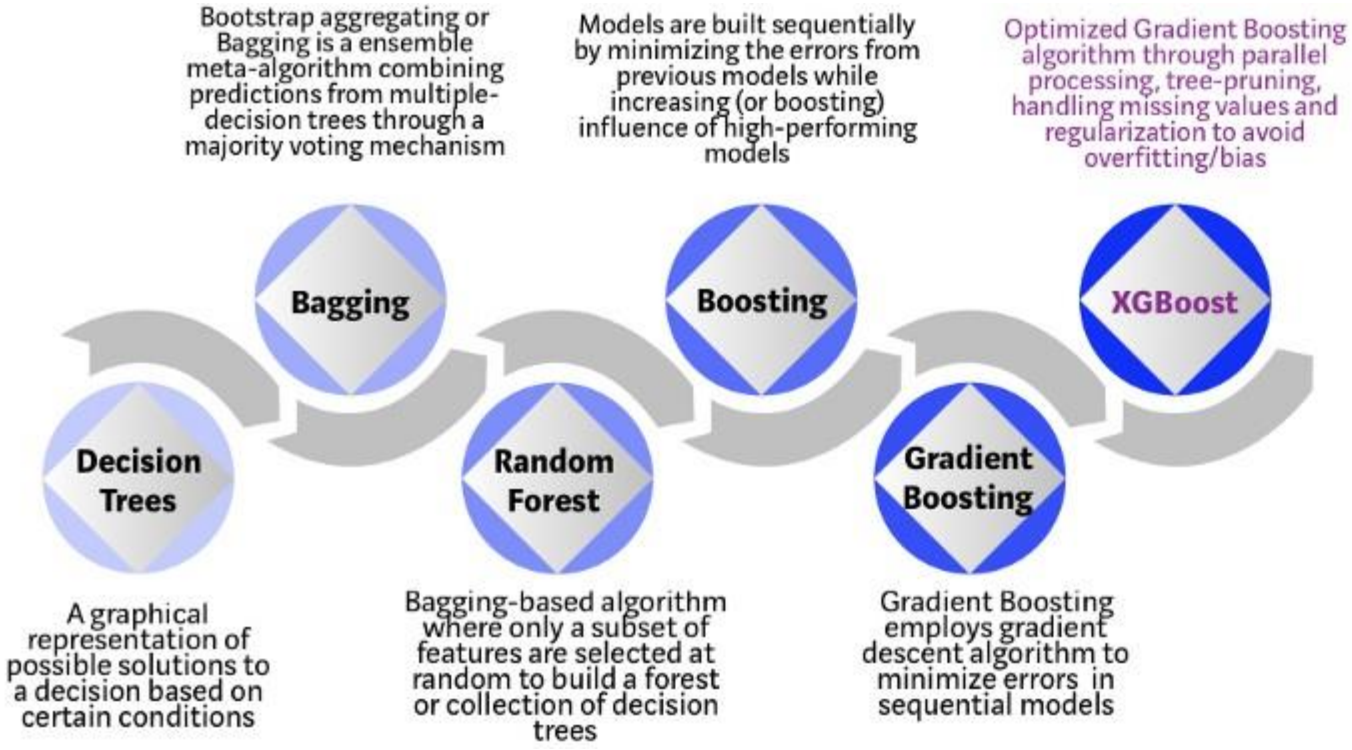
Boosting

- Boosting is designed to reduce model bias; it was originally intended for classification problems and then extended to regression problems. Boosting builds a sequence of models where each new model in the sequence is designed to improve performance on observations in the training data set that were misclassified in the previous model.



XGBoost

- XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework.

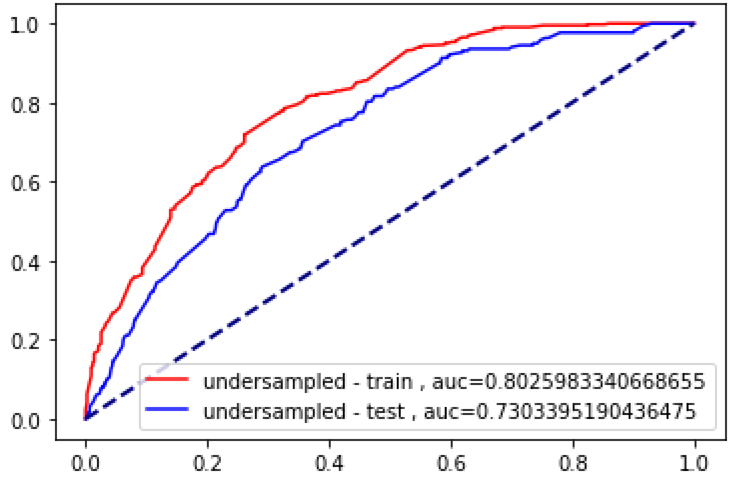


```python
[25] print('XGBoost model - undersampled')
     xgb_model = xgb.XGBClassifier(objective='binary:logistic', eval_metric="auc")
     predicted_y = []
     expected_y = []

     xgb_model_under = xgb_model.fit(X_undersampled, y_undersampled, early_stopping_rounds=5, eval_set=[(X_test.to_numpy(), y_test)])
     predictions =  xgb_model_under.predict(X_undersampled)
     predicted_y.extend(predictions)
     expected_y.extend(y_undersampled)
     report_train = classification_report(expected_y, predicted_y)
     print('training set')
     print(report_train)

     predicted_y = []
     expected_y = []
     predictions = xgb_model_under.predict(X_test.to_numpy())
     predicted_y.extend(predictions)
     expected_y.extend(y_test)
     report_test = classification_report(expected_y, predicted_y)
     print('test set')
     print(report_test)
```

```
XGBoost model - undersampled
[0]	validation_0-auc:0.705229
Will train until validation_0-auc hasn't improved in 5 rounds.
[1]	validation_0-auc:0.713788
[2]	validation_0-auc:0.716027
[3]	validation_0-auc:0.716674
[4]	validation_0-auc:0.715722
[5]	validation_0-auc:0.716981
[6]	validation_0-auc:0.715989
[7]	validation_0-auc:0.717465
[8]	validation_0-auc:0.716698
[9]	validation_0-auc:0.718954
[10]	validation_0-auc:0.718443
[11]	validation_0-auc:0.720803
[12]	validation_0-auc:0.721902
[13]	validation_0-auc:0.721018
[14]	validation_0-auc:0.722452
[15]	validation_0-auc:0.723688
[16]	validation_0-auc:0.723585
[17]	validation_0-auc:0.726791
[18]	validation_0-auc:0.727038
[19]	validation_0-auc:0.727042
[20]	validation_0-auc:0.727542
[21]	validation_0-auc:0.727999
[22]	validation_0-auc:0.728137
[23]	validation_0-auc:0.729541
[24]	validation_0-auc:0.729021
[25]	validation_0-auc:0.728669
[26]	validation_0-auc:0.728642
[27]	validation_0-auc:0.73034
[28]	validation_0-auc:0.730049
[29]	validation_0-auc:0.730086
[30]	validation_0-auc:0.728374
[31]	validation_0-auc:0.7291
[32]	validation_0-auc:0.727795
Stopping. Best iteration:
[27]	validation_0-auc:0.73034
```

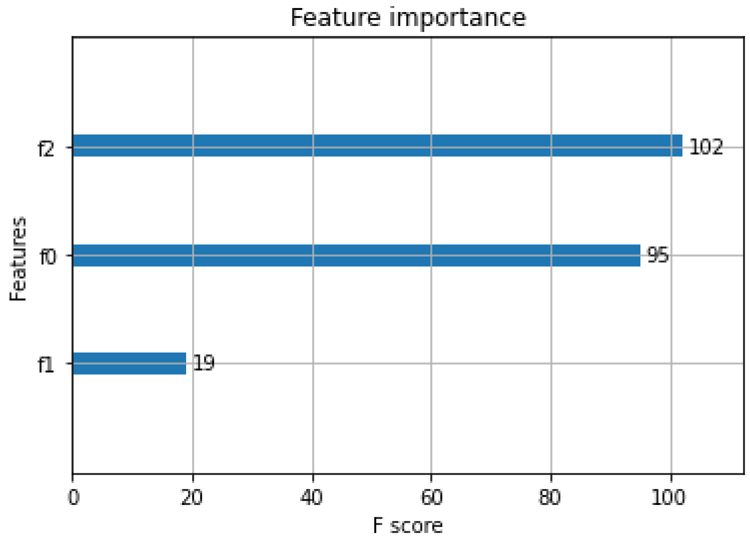
```
training set
              precision    recall  f1-score   support

           0       0.77      0.61      0.68       429
           1       0.68      0.82      0.74       429

    accuracy                           0.72       858
   macro avg       0.73      0.72      0.71       858
weighted avg       0.73      0.72      0.71       858

test set
              precision    recall  f1-score   support

           0       0.96      0.53      0.69      1848
           1       0.17      0.80      0.28       218

    accuracy                           0.56      2066
   macro avg       0.56      0.67      0.48      2066
weighted avg       0.88      0.56      0.64      2066
```

```python
[26] y_score_train = xgb_model_under.predict_proba(X_undersampled)
     fpr_train, tpr_train, _ = roc_curve(y_undersampled,  y_score_train[:,1])
     auc_train = roc_auc_score(y_undersampled, y_score_train[:,1])
     plt.plot(fpr_train,tpr_train, color='red', label='undersampled - train , auc='+str(auc_train))

     y_score_test = xgb_model_under.predict_proba(X_test.to_numpy())
     fpr_test, tpr_test, _ = roc_curve(y_test,  y_score_test[:,1])
     auc_test = roc_auc_score(y_test, y_score_test[:,1])
     plt.plot(fpr_test,tpr_test, color='Blue', label='undersampled - test , auc='+str(auc_test))

     plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
     plt.legend(loc=4)
     plt.show()
```



```
[27] xgb.plot_importance(xgb_model_under)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f738c6b3208>
```



```
print('XGBoost model - SMOTE - parameter tuning')

xgb_model = xgb.XGBClassifier(objective='binary:logistic', eval_metric="auc",
 base_score=0.5,
 learning_rate =0.01,
 n_estimators=5000,
 max_depth=3,
 min_child_weight=2,
 gamma=0.1,
 subsample=0.4,
 colsample_bytree=0.4,
 nthread=4)
predicted_y = []
expected_y = []

xgb_model_SMOTE_tuned = xgb_model.fit(X_SMOTE, y_SMOTE, early_stopping_rounds=5, eval_set=[(X_test.to_numpy(), y_test)])
predictions =  xgb_model_SMOTE_tuned.predict(X_SMOTE)
predicted_y.extend(predictions)
expected_y.extend(y_SMOTE)
report_train = classification_report(expected_y, predicted_y)
print('training set')
print(report_train)

predicted_y = []
expected_y = []
predictions = xgb_model_SMOTE_tuned.predict(X_test.to_numpy())
predicted_y.extend(predictions)
expected_y.extend(y_test)
report_test = classification_report(expected_y, predicted_y)
print('test set')
print(report_test)
```

```
XGBoost model - SMOTE - parameter tuning
[0]	validation_0-auc:0.611373
Will train until validation_0-auc hasn't improved in 5 rounds.
[1]	validation_0-auc:0.733395
[2]	validation_0-auc:0.728699
[3]	validation_0-auc:0.738274
[4]	validation_0-auc:0.738525
[5]	validation_0-auc:0.732941
[6]	validation_0-auc:0.740621
[7]	validation_0-auc:0.741142
[8]	validation_0-auc:0.740015
[9]	validation_0-auc:0.739549
[10]	validation_0-auc:0.741946
[11]	validation_0-auc:0.737816
[12]	validation_0-auc:0.738247
[13]	validation_0-auc:0.737253
[14]	validation_0-auc:0.736756
[15]	validation_0-auc:0.736153
Stopping. Best iteration:
[10]	validation_0-auc:0.741946

training set
              precision    recall  f1-score   support

           0       0.81      0.63      0.71      4389
           1       0.70      0.86      0.77      4389

    accuracy                           0.74      8778
   macro avg       0.76      0.74      0.74      8778
weighted avg       0.76      0.74      0.74      8778

test set
              precision    recall  f1-score   support

           0       0.95      0.63      0.76      1848
           1       0.19      0.74      0.31       218

    accuracy                           0.65      2066
   macro avg       0.57      0.69      0.53      2066
weighted avg       0.87      0.65      0.71      2066
```
Co

Exercise

https://www.kaggle.com/rodsaldanha/arketing-campaign

AcceptedCmp1 - 1 if customer accepted the offer in the 1st campaign, 0 otherwise
AcceptedCmp2 - 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
AcceptedCmp3 - 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
AcceptedCmp4 - 1 if customer accepted the offer in the 4th campaign, 0 otherwise
AcceptedCmp5 - 1 if customer accepted the offer in the 5th campaign, 0 otherwise
Response (target) - 1 if customer accepted the offer in the last campaign, 0 otherwise
Complain - 1 if customer complained in the last 2 years
DtCustomer - date of customer's enrolment with the company
Education - customer's level of education
Marital - customer's marital status
Kidhome - number of small children in customer's household
Teenhome - number of teenagers in customer's household
Income - customer's yearly household income
MntFishProducts - amount spent on fish products in the last 2 years
MntMeatProducts - amount spent on meat products in the last 2 years
MntFruits - amount spent on fruits products in the last 2 years
MntSweetProducts - amount spent on sweet products in the last 2 years
MntWines - amount spent on wine products in the last 2 years
MntGoldProds - amount spent on gold products in the last 2 years
NumDealsPurchases - number of purchases made with discount
NumCatalogPurchases - number of purchases made using catalogue
NumStorePurchases - number of purchases made directly in stores
NumWebPurchases - number of purchases made through company's web site
NumWebVisitsMonth - number of visits to company's web site in the last month
Recency - number of days since the last purchase

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

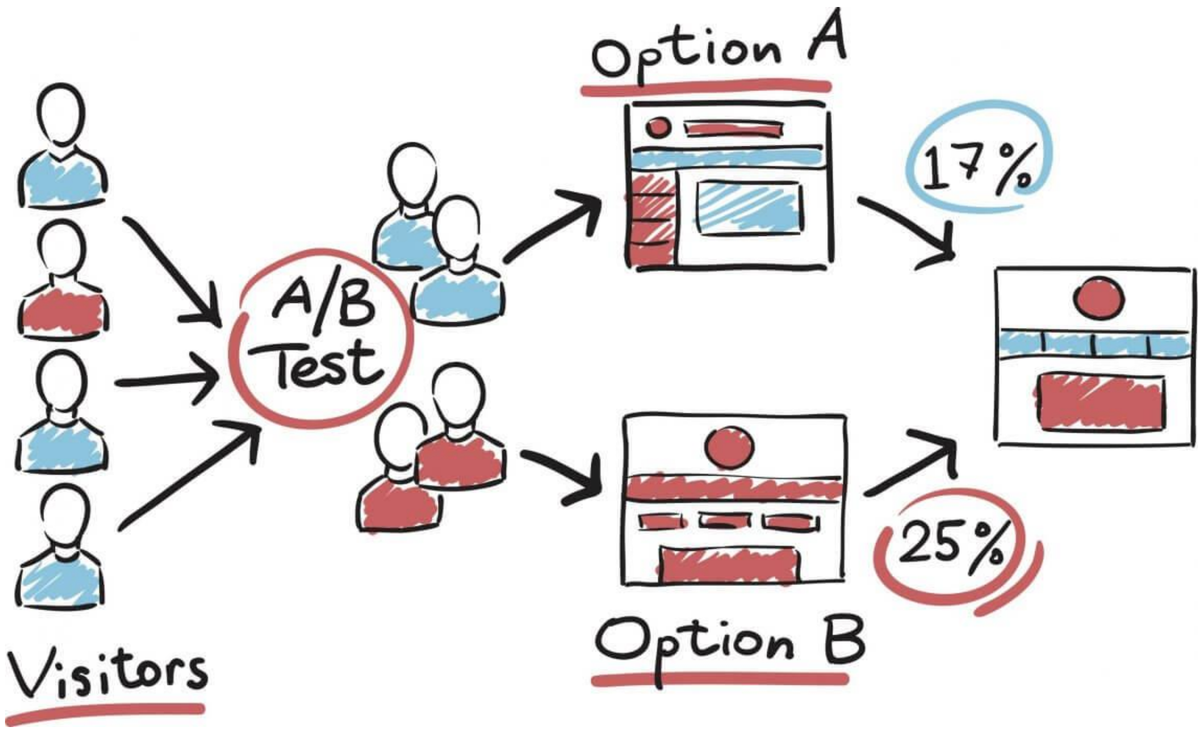
Topic 11
A/B Testing


BUSINESS
Learn from the best to ensure success
Reasons we will be successful
Issue 764
Monday, Jun 14, 2016
Citydailynews
25 great jobs for people who love to travel
Business Company

A/B Testing

A/B Testing

A/B testing (sometimes called split testing) is comparing two versions of a web page to see which one performs better. You compare two web pages by showing the two variants (let's call them A and B) to similar visitors at the same time. The one that gives a better conversion rate, wins!



"Always Be Testing"

https://vwo.com/ab-testing/

What Can You Test?

- Headlines
- Sub headlines
- Paragraph Text
- Testimonials
- Call to Action text
- Call to Action Button
- Links
- Images
- Content near the fold
- Social proof
- Media mentions
- Awards and badges

Which test won?

มุ้ขข ขข ไหนได้ใจกว่ากัน

Between – Subjects

ไข่มุก อยู่ด้านบน หรือ ด้านล่างแก้ว แบบไหน
กระตุ้นต่อมความอยากกินมากกว่ากัน

A

B

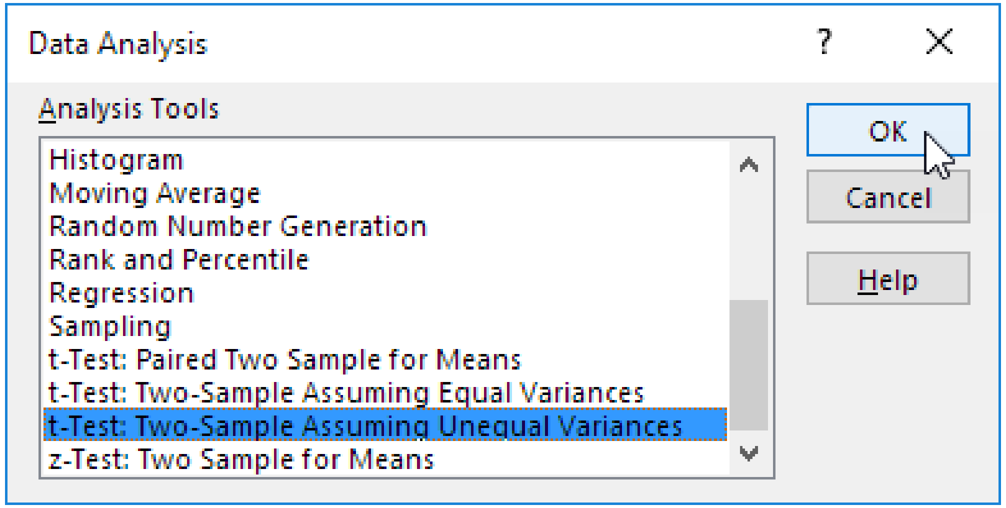
Mean Comparison

t-test

	A	B	C
1	Female	Male	
2	26	23	
3	25	30	
4	43	18	
5	34	25	
6	18	28	
7	52		
8			

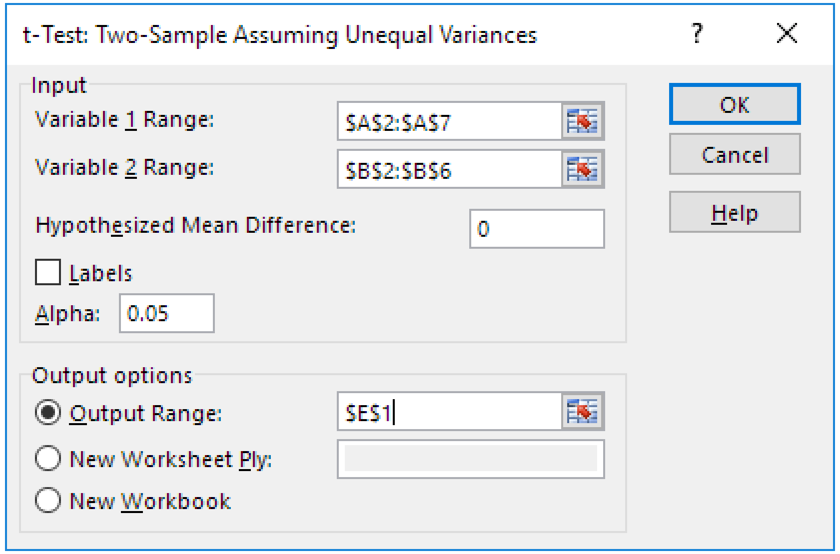
- This example teaches you how to perform a t-Test in Excel. The t-Test is used to test the null hypothesis that the means of two populations are equal.
- Below you can find the study hours of 6 female students and 5 male students.
- H_0: $\mu_1 - \mu_2 = 0$
  H_1: $\mu_1 - \mu_2 \neq 0$

t-test



- On the Data tab, in the Analysis group, click Data Analysis.
- Select t-Test: Two-Sample Assuming Unequal Variances and click OK.

t-test



- Click in the Variable 1 Range box and select the range A2:A7.
- Click in the Variable 2 Range box and select the range B2:B6.
- Click in the Hypothesized Mean Difference box and type 0 (H_0: $\mu_1 - \mu_2 = 0$).
- Click in the Output Range box and select cell E1.
- Click OK.

t-test

E	F	G
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	33	24.8
Variance	160	21.7
Observations	6	5
Hypothesized Mean Difference	0	
df	7	
t Stat	1.47260514	
P(T<=t) one-tail	0.092170202	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.184340405	
t Critical two-tail	2.364624252	

- Conclusion: We do a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. This is not the case, -2.365 < 1.473 < 2.365.
- Therefore, we do not reject the null hypothesis. The observed difference between the sample means (33 - 24.8) is not convincing enough to say that the average number of study hours between female and male students differ significantly.

BUSINESS
Issue 764
Monday, Jun 14, 2016
Citydailynews
25 great jobs for people who love to travel
Business Company

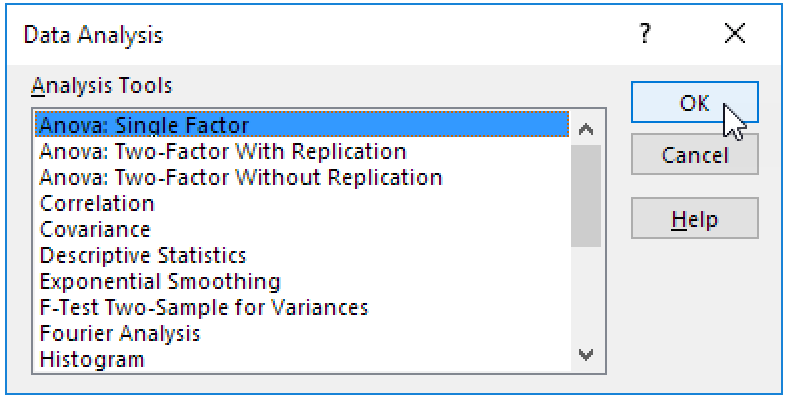
ANOVA

ANOVA

	A	B	C	D
1	economics	medicine	history	
2	42	69	35	
3	53	54	40	
4	49	58	53	
5	53	64	42	
6	43	64	50	
7	44	55	39	
8	45	56	55	
9	52		39	
10	54		40	
11				

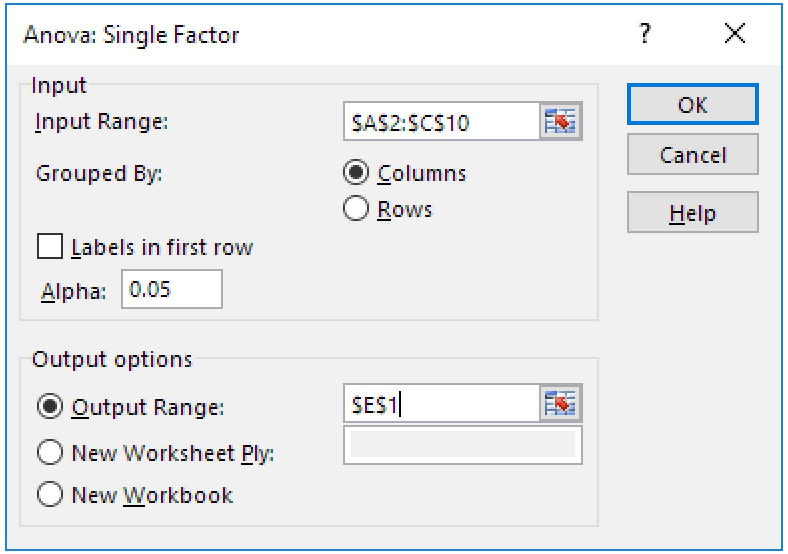
- This example teaches you how to perform a single factor ANOVA (analysis of variance) in Excel. A single factor or one-way ANOVA is used to test the null hypothesis that the means of several populations are all equal.
- Below you can find the salaries of people who have a degree in economics, medicine or history.
- H_0: $\mu_1 = \mu_2 = \mu_3$
  H_1: at least one of the means is different.

ANOVA



- On the Data tab, in the Analysis group, click Data Analysis.
- Select ANOVA: Single Factor and click OK.

ANOVA



- Click in the Input Range box and select the range A2:C10.
- Click in the Output Range box and select cell E1.
- Click OK.

ANOVA

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	9	435	48.33333	23.5		
Column 2	7	420	60	32.33333		
Column 3	9	393	43.66667	50.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

- Conclusion: if F > F crit, we reject the null hypothesis. This is the case, 15.196 > 3.443. Therefore, we reject the null hypothesis. The means of the three populations are not all equal. At least one of the means is different.
- However, the ANOVA does not tell you where the difference lies. You need a t-Test to test each pair of means.


Learn from the best to ensure success
Reasons we will be successful
BUSINESS
Issue 764
Monday, Jun 14, 2016
Citydailynews
25 great jobs for people who love to travel
Business Company

CHI-SQUARE

Learning Objectives

- Understand the principle of chi-square test
- Conduct hypothesis testing using chi-square test

Contingency Tables

Contingency Tables

- Useful in situations involving multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called a cross-classification table.

One-Way Contingency Table

Shows number of observations in *k* independent groups (outcomes or variable levels)

Outcomes (k = 3)

Candidate

Tom	Bill	Mary	Total
35	20	45	100

Number of responses

A Test of a Hypothesis : One-Way Table

H_0: $p_1 = p_{1,0}, p_2 = p_{2,0}, \ldots, p_k = p_{k,0}$

where $p_{1,0}, p_{2,0}, \ldots, p_{k,0}$ represent the hypothesized values of the multinomial probabilities.

H_a: At least one of the multinomial probabilities does not equal its hypothesized value.

The Chi Square Test

- The general formula is

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

- where
  - O = observed data in each category
  - E = observed data in each category based on the experimenter's hypothesis

Chi-Square: Goodness of Fit Test with Excel

- As an HR director, you want to test the perception of fairness of three methods of performance evaluation. Of 180 employees,

  63 rated Method 1 as fair,

  45 rated Method 2 as fair,
  72 rated Method 3 as fair.

- At the .05 level of significance, is there a difference in perceptions?

Chi-Square: Goodness of Fit Test with Excel

=CHISQ.TEST(D5:D7,E5:E7)

B	C	D	E	F	G	H	I

	Observed	Expected
Method 1	63	60
Method 2	45	60
Method 3	72	60

Chi-Square 0.042852

Testing Hypothesized Relationship

Relationship between two categorical variables

- Suppose we are interested in the relationship between gender and attending college.

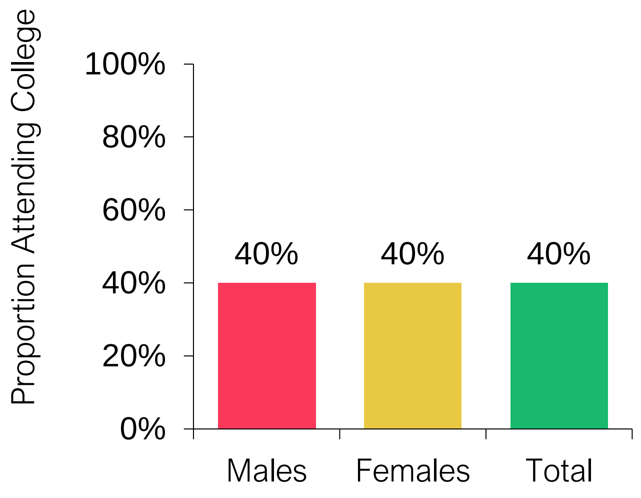
- If there is no relationship between gender and attending college and 40% of our total sample attend college, we would expect 40% of the males in our sample to attend college and 40% of the females to attend college.

- If there is a relationship between gender and attending college, we would expect a higher proportion of one group to attend college than the other group, e.g. 60% to 20%.

Displaying Independent and Dependent Relationships

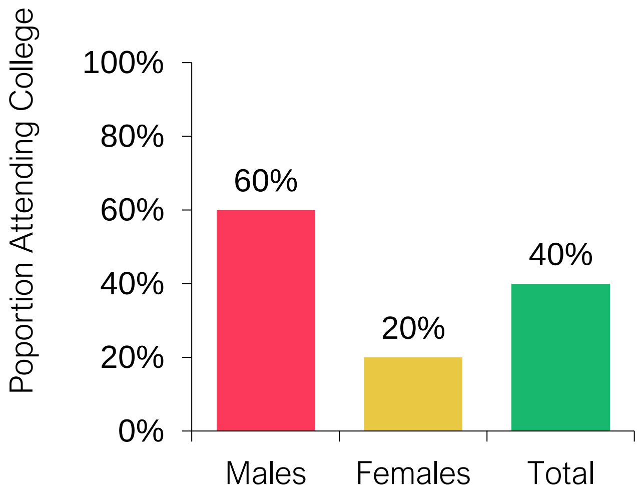
When the variables are independent, the proportion in both groups is close to the same size as the proportion for the total sample.

Independent Relationship between Gender and College



When group membership makes a difference, the dependent relationship is indicated by one group having a higher proportion than the proportion for the total sample.

Dependent Relationship between Gender and College



Chi-square: Test of independence

- Shows if a relationship exists between two qualitative variables
  - One sample is drawn
  - Does not show causality
- Uses two-way contingency table

Two-way Contingency Table

Levels of variable 2

House Style	House Location: Urban	Rural	Total
Split-Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Levels of variable 1

Finding Expected Cell Counts for a Two-Way Contingency Table

- The estimate of the expected number of observations falling into the cell in row *i* and column *j* is given by

$$E_{ij} = \frac{R_i C_j}{n}$$

where R_i = total for row i, C_j = total for column j,
and n = sample size.

$\chi 2$ -Test for Independence

H_0: The two classifications are independent.
H_a: The two classifications are dependent.

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

Rejection region: $\chi^2 > \chi^2_\alpha$, where χ^2_α has $(r - 1)(c - 1)$ df.

*** The sample size n must be large. This is satisfied if for every cell, the expected count E_i will be equal to 5 or more.

Expected Count Example

Marginal probability = $\frac{112}{160}$

	Location		
	Urban	Rural	
House Style	Obs.	Obs.	Total
Split–Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Expected Count Example

Marginal probability = $\frac{112}{160}$

House Style	Location: Urban Obs.	Location: Rural Obs.	Total
Split–Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Marginal probability = $\frac{78}{160}$

Expected Count Example

Joint probability = $\frac{112}{160}\frac{78}{160}$

Marginal probability = $\frac{112}{160}$

	Location		
House Style	Urban Obs.	Rural Obs.	Total
Split–Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Marginal probability = $\frac{78}{160}$

Expected count $= \cancel{160} \cdot \frac{112}{\cancel{160}}\frac{78}{160}$

$= 54.6$

Expected Count Example

$$E_{ij} = \frac{R_i C_j}{n}$$

	House Location				
	Urban		Rural		
House Style	Obs.	Exp.	Obs.	Exp.	Total
Split-Level	63	54.6	49	57.4	112
Ranch	15	23.4	33	24.6	48
Total	78	78	82	82	160

$\frac{112 \cdot 78}{160}$ $\frac{112 \cdot 82}{160}$ $\frac{48 \cdot 78}{160}$ $\frac{48 \cdot 82}{160}$

Chi-Square: Test of Independence

- As a realtor you want to determine if house style and house location are related. At the .05 level of significance, is there evidence of a relationship?

	House Location		
House Style	Urban	Rural	Total
Split-Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Chi-Square: Test of Independence

=CHISQ.TEST(D5:E6,D10:E11)

B	C	D	E	F

Chi-Square 0.112298302

Observed	Urban	Rural	Total
Split-Level	63	49	112
Ranch	15	33	48
Total	78	82	160

Expected	Urban	Rural	Total
Split-Level	54.6	57.4	112
Ranch	15	33	48
Total	78	82	160

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

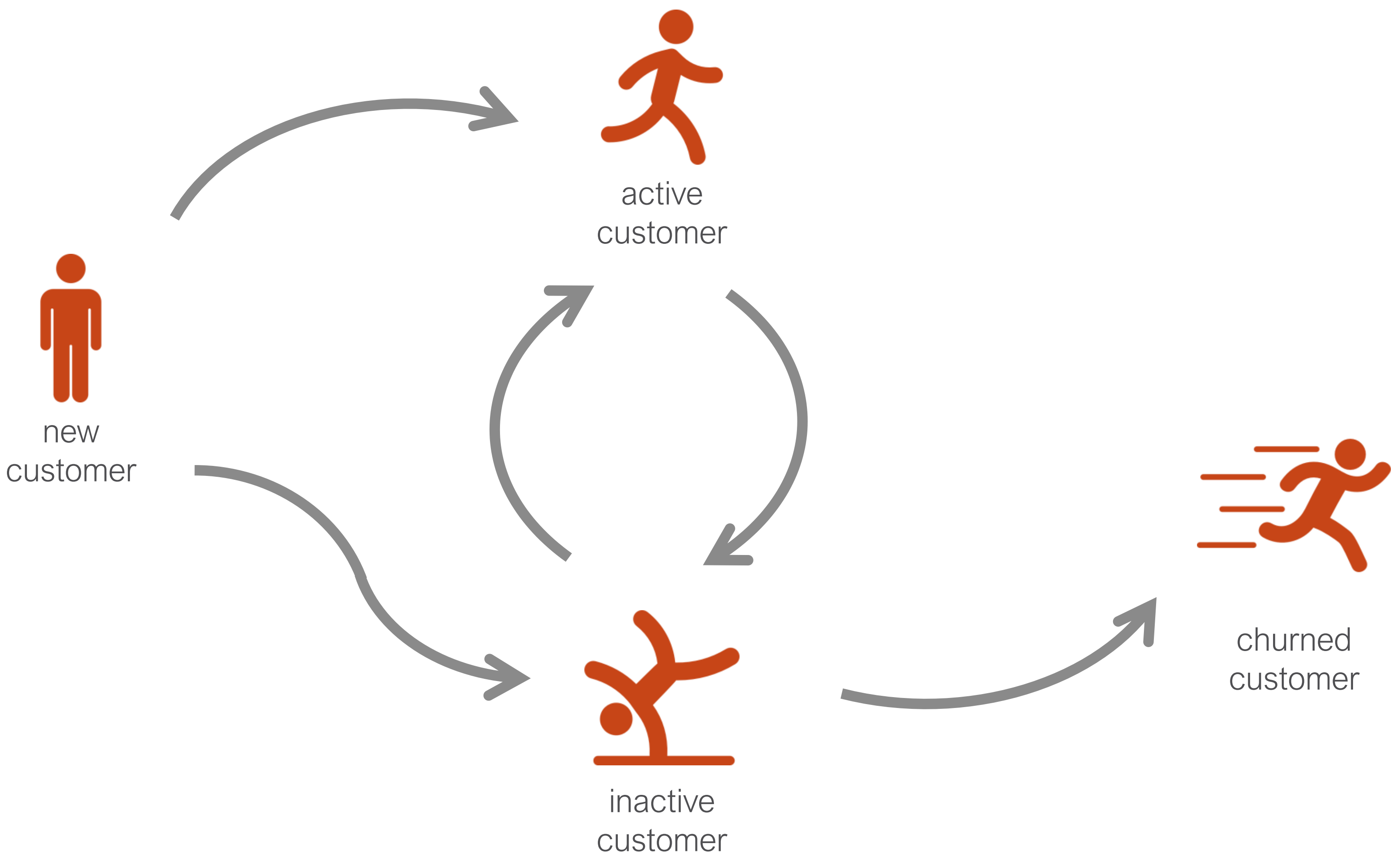
Topic 12
Churn Prediction Model

Churn


Contractual
Non-Contractual
Clear Ending
Unclear Ending
churned

Churn Prediction – Contractual

Churn Prediction – Non-Contractual



Customer Churn vs. Customer Retention

$$\text{CUSTOMER CHURN RATE} = \frac{\text{(CUSTOMERS BEGINNING OF MONTH - CUSTOMERS END OF MONTH)}}{\text{CUSTOMERS BEGINNING OF MONTH}}$$

$$\frac{(500-450)}{500} = 10\%$$

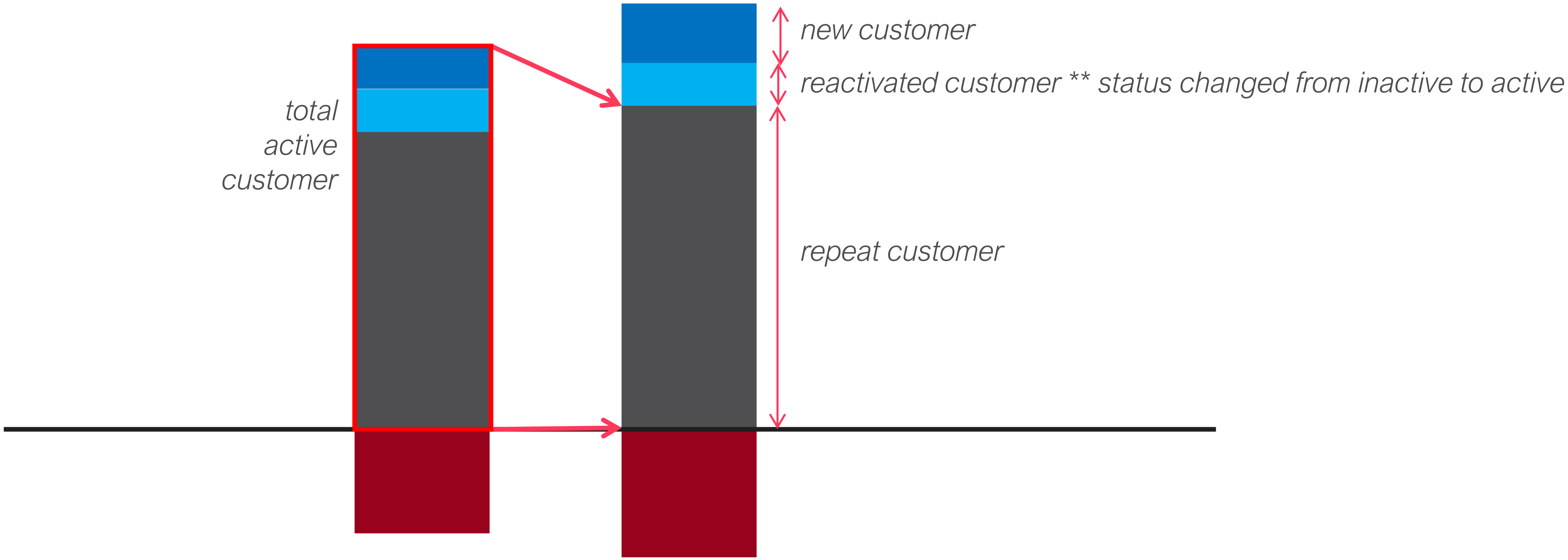
The Formula for Calculating Your Customer Retention Rate (CRR)

$$CRR = \left[\frac{E - N}{S}\right] X\ 100$$

E = The number of customers at the end of the period
N = The number of new customers acquired during that period
S = The number of customers at the start of the period

https://www.salesforce.com/

Customer Movement Analysis



Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 13
Voice of Customer Analytics

Installing libraries

Load libraries and data

```
[1] !pip install --upgrade pythainlp
    !pip install pyLDAvis
```

```
Requirement already up-to-date: pythainlp in /usr/local/lib/python3.6/dist-packages (2.2.4)
Requirement already satisfied, skipping upgrade: requests>=2.22.0 in /usr/local/lib/python3.6/dist-packages (from pythainlp) (2.23.0)
Requirement already satisfied, skipping upgrade: python-crfsuite>=0.9.6 in /usr/local/lib/python3.6/dist-packages (from pythainlp) (0.9.7)
Requirement already satisfied, skipping upgrade: tinydb>=3.0 in /usr/local/lib/python3.6/dist-packages (from pythainlp) (4.2.0)
Requirement already satisfied, skipping upgrade: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests>=2.22.0->pythainlp) (3.0.4)
Requirement already satisfied, skipping upgrade: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests>=2.22.0->pythainlp) (1.24.3)
Requirement already satisfied, skipping upgrade: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests>=2.22.0->pythainlp) (2.10)
Requirement already satisfied, skipping upgrade: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests>=2.22.0->pythainlp) (2020.6.20)
Requirement already satisfied: pyLDAvis in /usr/local/lib/python3.6/dist-packages (2.1.2)
Requirement already satisfied: wheel>=0.23.0 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (0.35.1)
Requirement already satisfied: pytest in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (3.6.4)
Requirement already satisfied: numpy>=1.9.2 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (1.18.5)
Requirement already satisfied: scipy>=0.18.0 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (1.4.1)
Requirement already satisfied: jinja2>=2.7.2 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (2.11.2)
Requirement already satisfied: joblib>=0.8.4 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (0.17.0)
Requirement already satisfied: pandas>=0.17.0 in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (1.1.4)
Requirement already satisfied: numexpr in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (2.7.1)
Requirement already satisfied: funcy in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (1.15)
Requirement already satisfied: future in /usr/local/lib/python3.6/dist-packages (from pyLDAvis) (0.16.0)
Requirement already satisfied: pluggy<0.8,>=0.5 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (0.7.1)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (1.15.0)
Requirement already satisfied: more-itertools>=4.0.0 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (8.6.0)
Requirement already satisfied: atomicwrites>=1.0 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (1.4.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (20.2.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (50.3.2)
Requirement already satisfied: py>=1.5.0 in /usr/local/lib/python3.6/dist-packages (from pytest->pyLDAvis) (1.9.0)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.6/dist-packages (from jinja2>=2.7.2->pyLDAvis) (1.1.1)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.6/dist-packages (from pandas>=0.17.0->pyLDAvis) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages (from pandas>=0.17.0->pyLDAvis) (2018.9)
```

Importing libraries and data

```
[2]  import pandas as pd
     import pythainlp
     import gensim
     import pyLDAvis.gensim
     pyLDAvis.enable_notebook()
     import warnings
     warnings.filterwarnings("ignore", category=DeprecationWarning)
```

```
[3]  df = pd.read_csv('Wongnai Reviews - Small.csv')
```

```
[4]  df.tail()
```

	Review ID	Review
295	296	ค่านี้คุณเพื่อนอยากส้มตำ หมูเฮาเลยพากันลงมากิน...
296	297	ร้านสะอาดดี ตกแต่งสวยงาม มีที่จอดรถ ราคาเมนูต่...
297	298	เช้าๆ รีบๆ วิ่งมาเข่าห้องเรียนแทบไม่ทันแต่ต้อง...
298	299	ร้านนี้เป็นร้านกาแฟเล็กๆ ข้างๆ ร้านๆ Happy Man...
299	300	ทรูคอฟฟี่สาขาซีคอนอยู่ในศูนย์บริการของทรู ชั้น...

Tokenize Words with pythainlp

Tokenize Words

```
[5] stopwords = list(pythainlp.corpus.thai_stopwords())
    removed_words = [' ', '  ', '\n', 'ร้าน', '(', ')']
    screening_words = stopwords + removed_words

    def tokenize_with_space(sentence):
      merged = ''
      words = pythainlp.word_tokenize(str(sentence), engine='newmm')
      for word in words:
        if word not in screening_words:
          merged = merged + ',' + word
      return merged[1:]
```

```
[6] df['Review_tokenized'] = df['Review'].apply(lambda x: tokenize_with_space(x))
```

```
[7] df.tail()
```

	Review ID	Review	Review_tokenized
295	296	ค่านี้คุณเพื่อนอยากส้มตำ หมูเฮาเลยพากันลงมากิน...	ค่ำ,เพื่อน,ส้มตำ,หมู,เฮา,ลงมา,กิน,ส้มตำ,ออฟฟิศ...
296	297	ร้านสะอาดดี ตกแต่งสวยงาม มีที่จอดรถ ราคาเมนูต่...	สะอาด,ดี,ตกแต่ง,สวยงาม,ที่จอดรถ,ราคา,เมนู,เทีย...
297	298	เช้าๆ รีบๆ วิ่งมาเข่าห้องเรียนแทบไม่ทันแต่ต้อง...	เช้า,รีบ,วิ่ง,เข่า,ห้องเรียน,แทบ,ต้องหา,ของกิน...
298	299	ร้านนี้เป็นร้านกาแฟเล็กๆ ข้างๆ ร้านๆ Happy Man...	ร้านกาแฟ,Happy,Mango,อาทิตย์,นัด,เพื่อน,นั่ง,ค...
299	300	ทรูคอฟฟี่สาขาซีคอนอยู่ในศูนย์บริการของทรู ชั้น...	ทรู,คอ,ฟ,ฟี่,สาขา,ซีคอน,ศูนย์,บริการ,ทรู,ชั้น,...

Creating Dictionary

Create Dictionary

```
[8] documents = df['Review_tokenized'].to_list()
    texts = [[text for text in doc.split(',')] for doc in documents]
    dictionary = gensim.corpora.Dictionary(texts)
```

```
[9] print(dictionary.token2id.keys())
```

```
dict_keys(['20', 'Macchiato', 'กาแฟ', 'กาแฟร้อน', 'กิน', 'คน', 'ครึ่ง', 'ความคิด', 'ชอบ', 'ดื่ม', 'ตอน', 'ทาน', 'นึง', 'บาท', 'ปริมาณ',
```

```
[10] gensim_corpus = [dictionary.doc2bow(text, allow_update=True) for text in texts]
     word_frequencies = [[(dictionary[id], frequence) for id, frequence in couple] for couple in gensim_corpus]
```

Topic Modeling using LDA

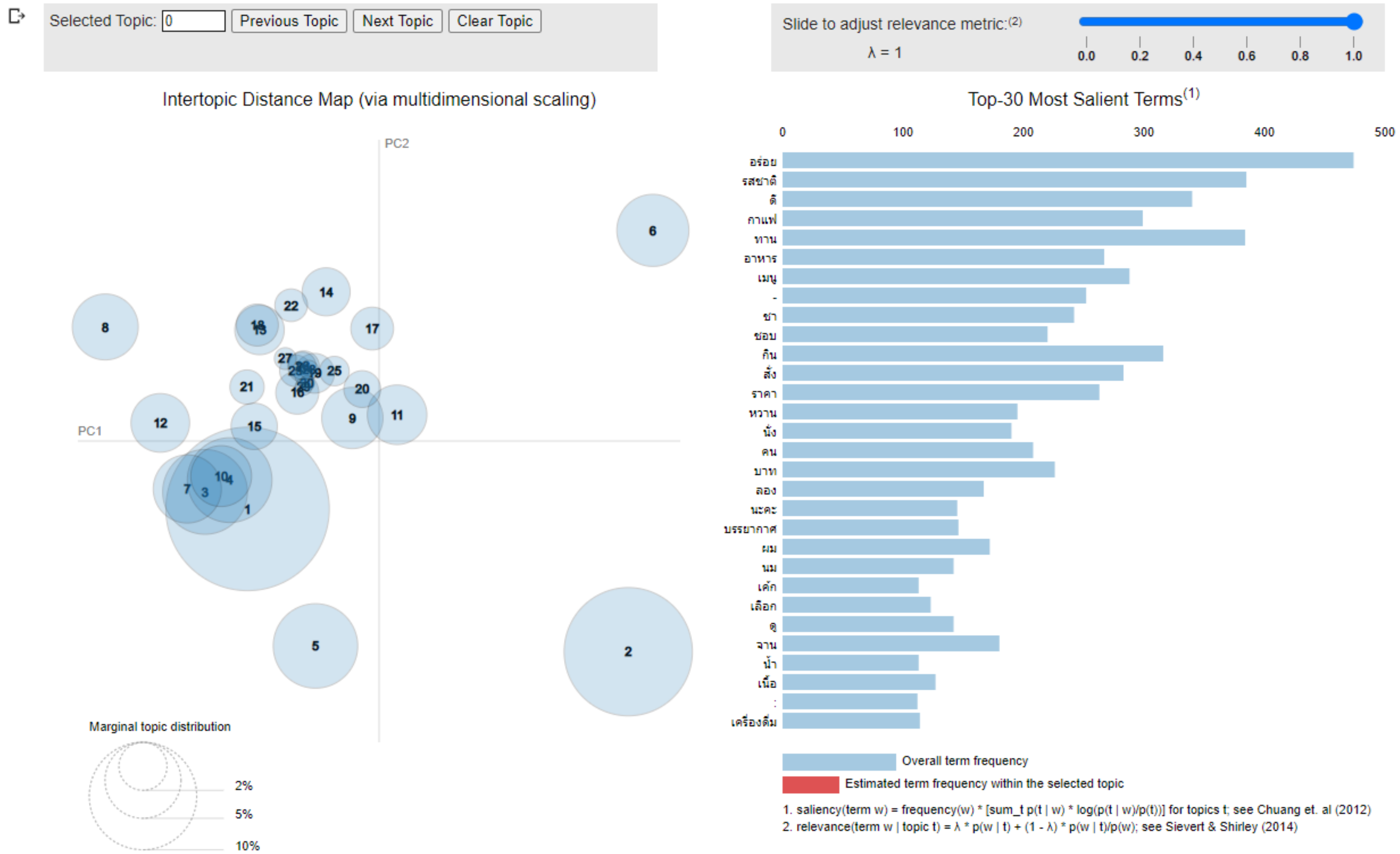
Topic Modeling

```
[11] num_topics = 30
     chunksize = 4000 # size of the doc looked at every pass
     passes = 20 # number of passes through documents
     iterations = 50
     eval_every = 1  # Don't evaluate model perplexity, takes too much time.

     # Make a index to word dictionary.
     temp = dictionary[0]  # This is only to "load" the dictionary.
     id2word = dictionary.id2token

     %time model = gensim.models.LdaModel(corpus=gensim_corpus, id2word=id2word, chunksize=chunksize, \
                            alpha='auto', eta='auto', \
                            iterations=iterations, num_topics=num_topics, \
                            passes=passes, eval_every=eval_every)
```

```
CPU times: user 16.4 s, sys: 11.4 s, total: 27.7 s
Wall time: 14.2 s
```

Visualizing Results

```
pyLDAvis.gensim.prepare(model, gensim_corpus, dictionary)
```



Predict Topics

```
[13] model.show_topic(1)
```

```
[('เค้ก', 0.014751856),
 ('อร่อย', 0.01075815),
 ('รสชาติ', 0.009799483),
 ('ทาน', 0.0094283),
 ('ดี', 0.0076978332),
 ('ชา', 0.00694221),
 ('ราคา', 0.0064805807),
 ('หอม', 0.006462373),
 ('หวาน', 0.006441255),
 ('กาแฟ', 0.0057419683)]
```

```
[14] df['topics'] = df['Review_tokenized'].apply(lambda x: model.get_document_topics(dictionary.doc2bow(x.split(',')))[0][0])
     df['score'] = df['Review_tokenized'].apply(lambda x: model.get_document_topics(dictionary.doc2bow(x.split(',')))[0][1])
```

```
[15] df.tail()
```

	Review ID	Review	Review_tokenized	topics	score
295	296	ค่ำนี้คุณเพื่อนอยากส้มตำ หมูเฮาเลยพากันลงมากิน...	ค่ำ,เพื่อน,ส้มตำ,หมู,เฮา,ลงมา,กิน,ส้มตำ,ออฟฟิศ...	13	0.996950
296	297	ร้านสะอาดดี ตกแต่งสวยงาม มีที่จอดรถ ราคาเมนูต่...	สะอาด,ดี,ตกแต่ง,สวยงาม,ที่จอดรถ,ราคา,เมนู,เทีย...	10	0.992349
297	298	เช้าๆ รีบๆ วิ่งมาเข่าห้องเรียนแทบไม่ทันแต่ต้อง...	เช้า,รีบ,วิ่ง,เข่า,ห้องเรียน,แทบ,ต้องหา,ของกิน...	23	0.247305
298	299	ร้านนี้เป็นร้านกาแฟเล็กๆ ข้างๆ ร้านๆ Happy Man...	ร้านกาแฟ,Happy,Mango,อาทิตย์,นัด,เพื่อน,นั่ง,ค...	23	0.996443
299	300	ทรูคอฟฟี่สาขาซีคอนอยู่ในศูนย์บริการของทรู ชั้น...	ทรู,คอ,ฟ,ฟี่,สาขา,ซีคอน,ศูนย์,บริการ,ทรู,ชั้น,...	24	0.994567

Customer Relationship Management Analytics and Intelligence

Thanachart Ritbumroong, Ph.D.

Topic 14
AI and CRM

Creative Morphology

Without Creativity,
Mankind Would Not Progress

Creativity vs. Innovation

- Creativity is the ability to develop new ideas an to discover new ways of looking at problems and opportunities.
- Innovation is the ability to apply creative solutions to problems and opportunities to enhance or to enrich people's lives. (Creative destruction)
- In a nutshell, creativity is thinking new things, innovation is doing new things.
- Innovation = Creativity + Implementation

Creativity

- Results from the interaction of a system consisting of
  Domain: contains symbolic rules
  Person: brings novelty into the symbolic domain
  Field of Experts: recognize and validate the innovation

- Creativity in the process by which a symbolic domain in the culture is changed.

Domain

- consists of a set of symbolic rules and procedures.
- Three dimensions of domains related to creativity

  The clarity of structure

  The centrality within the culture

  The accessibility

- Ex. Mathematics; the internal logic is strict; there is a high degree of clarity. Thus, it is easy for a young person to assimilate the rules quickly and jump to the cutting edge of the domain in a few years.

Field

- A field is necessary to determine the worth of a new idea.
- Fields can affect the rate of creativity in three ways;

  Being either reactive or proactive

  Screening new ideas (conservative/liberal)

  Connected to the rest of the social system to channel support into their own domain

Person

- Creativity is facilitated by a genetic predisposition for a given domain.

  A person whose nervous system is more sensitive to color and light will have an advantage in painting.
- Important ingredient of creativity

  Interest in the domain

  Access to a domain

  Access to a field

Creative Person

- Complex personality; express full range of traits
- A great deal of physical energy
- Smart, yet also naïve
- Playfulness and discipline
- Able to use well two opposite ways of thinking; the convergent and the divergent
  - Convergent thinking

    Solving well-defined, rational problems that have one correct answer
  - Divergent thinking

    Generating a great quantity of ideas, to switch from one perspective to another

Phases of Creative Process

- Preparation

  Learn about the problem

  Examine problem from various perspectives

  Similar to structuring the problem

  Understanding the structure of the problem and how elements relate to one another is preparation for the creative process

Phases of Creative Process

- Preparation
- Incubation

  Explore new paths and alternatives

  Many include unconscious processing of information

    Find solutions to problems in a dream

  Position of pieces of information yields a creative solution

  How many have wanted to "think about it for a while?"

Phases of Creative Process

- Preparation
- Incubation
- Insight
  When all the pieces come together
- Evaluation
  Does the solution have merit?
  Return to the hard logic of the problem
  Are all constraints being satisfied?
  How well does it perform with respect to the fundamental objectives?

Phases of Creative Process

- Preparation
- Incubation
- Insight
- Evaluation
- Elaboration

  Slow, routine work

Blocks to Creativity

- A “block to creativity” interferes with creativity
- Why should we be concerned about them?

  They hinder our decision analytical process

  If we understand what they are, and why they hinder the process, we can hopefully avoid them
- Framing and Perceptual Blocks

  Arise in the ways we tend to perceive, define, and examine the problem

Perceptual Blocks

- Stereotyping – fit into some standard category
- Tacit assumptions – impose artificial constraints
- Saturation

  Focus too quickly on “obvious” problem

  Focusing to much on details

  Getting overwhelmed with data
- Inability to see problem from other viewpoints

  Multiple objectives will be at play

  Must understand other’s values and objectives

Emotional or Value-based Blocks

- Fear of taking a risk
  Risk aversion is a key decision analysis concept
  May be counterproductive to not offer “wild” ideas
- Status quo bias
  Various levels of bias to current state of affairs
  Change can be hard to accept
- Reality versus Fantasy
  Some people only want realistic solutions
  Such people are comfortable “in their box”

Emotional or Value-based Blocks

- Judgment and Criticism
  Do not apply your values too soon in creative process
  Need to let ideas flow freely
- Inability to Incubate
  Not well understood
  Accepted as a phase
  Are we always given time to incubate an idea?

Cultural Blocks

- Taboos
  Views of culturally accept behavior may block ideas
- Humor
  Good ideas can be obtained in an informal setting
  Often want to let the joking free-wheel for a time
- Reason and Logic prevails
  Overly analytical thinking (even though it is important)
- Tradition and change
  Often a strong resistance to changes
  The status quo got the decision maker where they are

Environmental Blocks

- Non-supportive environment
- Environment that dissuades humor and playfulness
- Organization is overly structured and routine
- Strictly hierarchical structure
- Autocratic bosses

  Bosses that have all the answers
- Over focus on awards, competition and oversight
- Strict timelines

  Often a tight suspense can lead to good results

Brainstorming

- Introduced in 1930s by Osborn
- Based on idea of eliminating perceptual blocking filters
- Two Principles:
  - Defer judgment
  - Quantity breeds quality
- Four rules
  - Rule out criticism
  - Welcome freewheeling
  - Seek large quantities of ideas
  - Encourage combination and improvement of ideas

Brainstorming

- Works due to its synergistic effect
  - Among participants
  - Combining of ideas is not just additive
  - Combine pairs, triples, etc of ideas to get new ideas
- Generally regarded as a group technique based on a specific objective
  Specificity focuses the efforts
- Useful in situations calling for idea generation rather than judgment

Synectics

- Gordon in 50s found novel ideas expressed as analogies
  Research suggested use of analogies a key insight
- Reduce problem to barest essentials and search for a natural analogy
- Two distinguishing characteristics
  - Attack of the underlying concept of the problem
  - Examination of problem from many angles
- Three types of analogy (metaphorical thinking)
  - Fantasy – idealistic versus realistic
  - Direct – find personal parallel experiences
  - Personal – place yourself in role of problem

Checklists

- Very simple means of generating ideas
- Ask and list answers to series of questions. For instance

  Are there other uses?

  Can something be adapted?

  Can something be modified?

  Can components be re-arranged?

  Can components be combined?

  Can some substitution be made?
- Osborn (1963) offered a series of idea spurring questions.

Obsorn’s Questions

- Put to other uses
  - New ways to use as is
  - Other uses if modified
- Adapt
  - What else is like this?
  - What other idea does this suggest?
  - Does the past offer a parallel?
  - What could I copy?
  - Whom could I emulate?

Obsorn’s Questions

- Modify
  - New twist?
  - Change meaning, color, motion, sound, odor, form shape?
  - Other changes?
- Magnify
  - What to add?
  - More time? Greater frequency? Stronger? Higher?
  - Longer? Thicker? Extra value? Plus ingredient?
  - Duplicate? Multiply? Exaggerate?

Obsorn's Questions

- Minify?
  What to subtract? Smaller? Condensed? Minature?
  Lower? Shorter? Lighter? Omit? Streamline?
  Split up? Understate?
- Substitute?
  Who else instead? What else instead? Other ingredient? Other material? Other process?
  Other power? Other place? Other approach? Other tone of voice?

Obsorn's Questions

- Rearrange

  Interchange components? Other pattern? Other layout? Other sequence?
  Transpose cause and effect? Change pace? Change schedule?

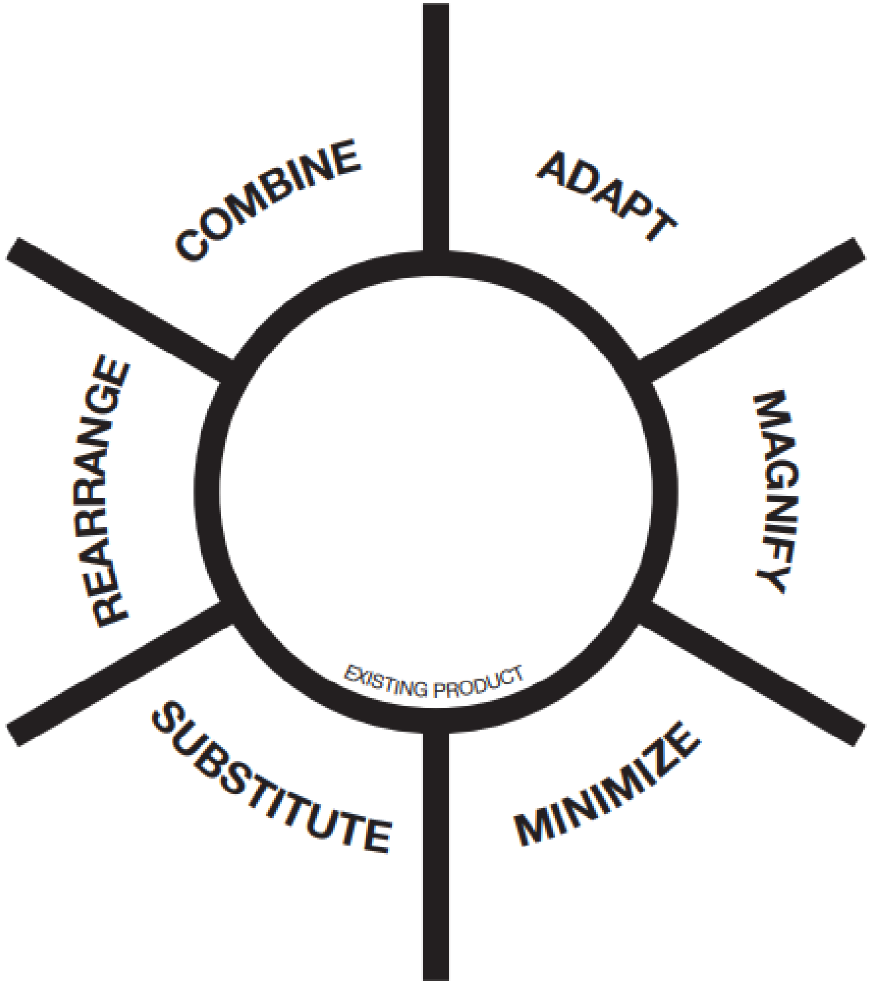
- Reverse?

  Transpose positive and negative? How about opposites?
  Turn it backward? Turn it upside down? Reverse roles?
  Change shoes? Turn tables? Turn other cheek?

Obsorn’s Questions

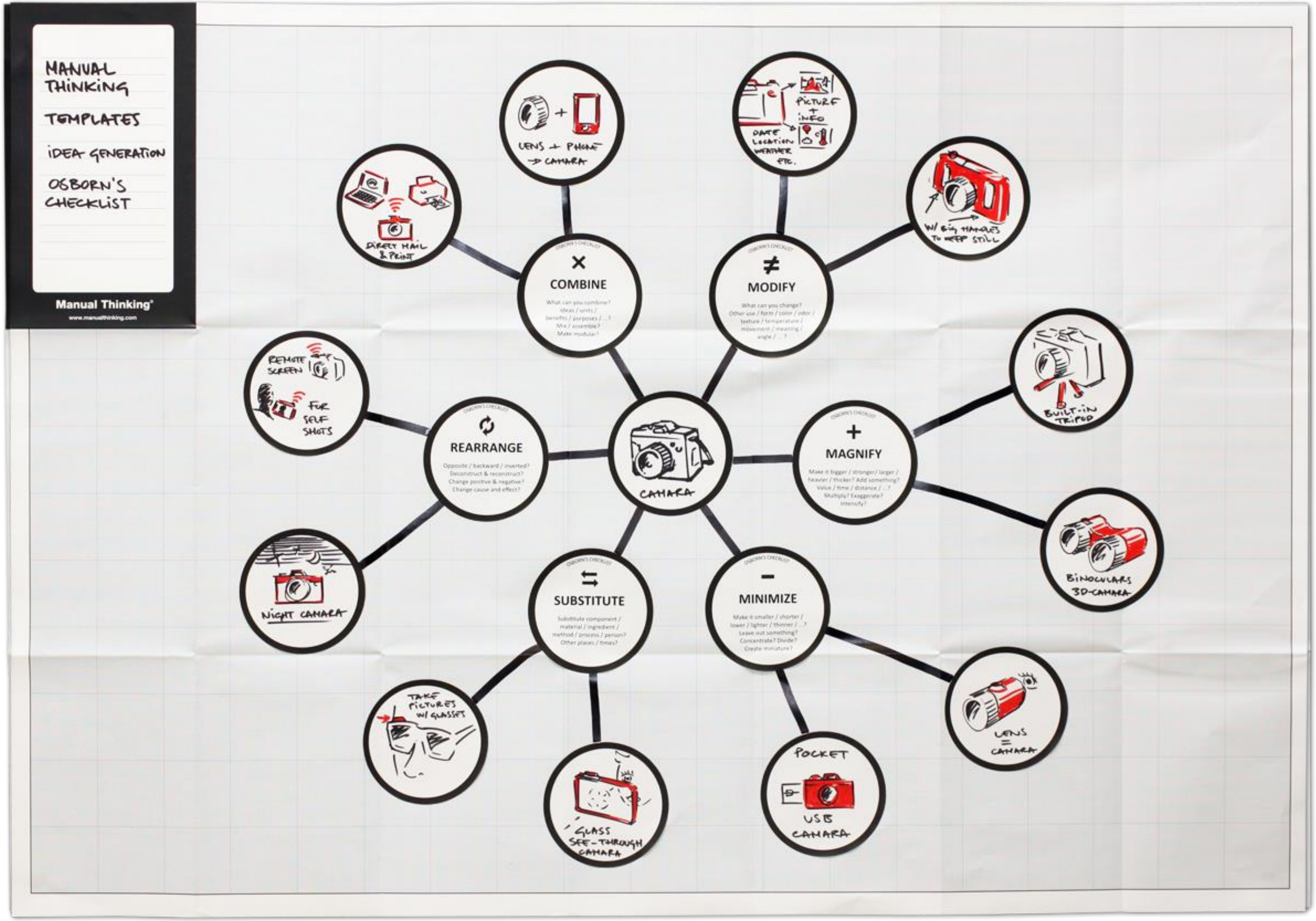
- Combine?
  How about a blend, an alloy, as assortment, an ensemble?
  Combine units?
  Combine purposes?
  Combine appeals?
  Combine ideas?

Osborn's Checklist



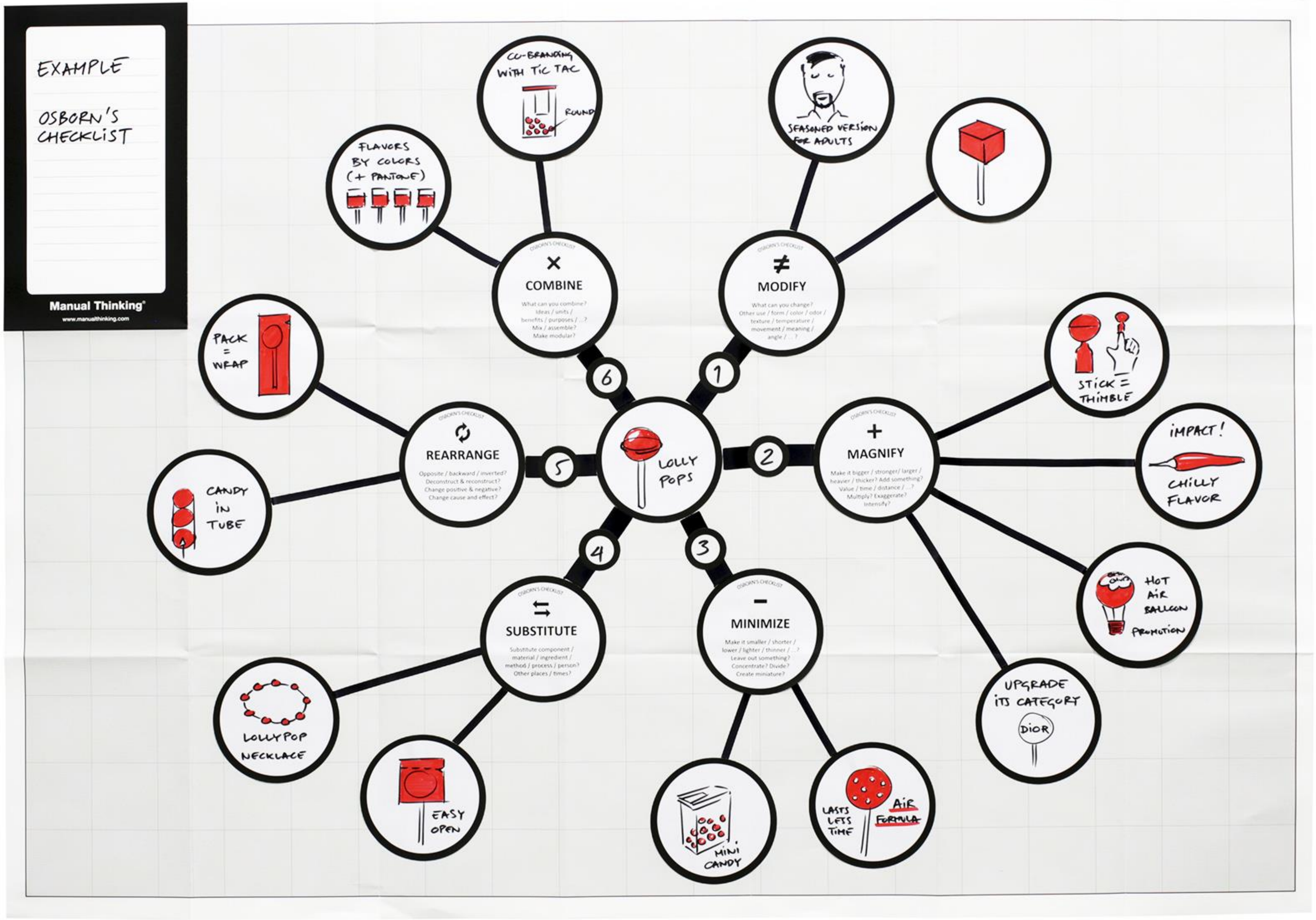
https://manualthinking.com/wp-

Example



MANUAL THINKING
TEMPLATES
IDEA GENERATION
OSBORN'S CHECKLIST
Manual Thinking
www.manualthinking.com
CAMARA
COMBINE
What can you combine?
Ideas / units /
benefits / purposes / ...?
Mix / assemble?
Make modular?
MODIFY
What can you change?
Other use / form / color / odor /
texture / temperature /
movement / meaning /
angle / ...?
MAGNIFY
Make it bigger / stronger/ larger /
heavier / thicker? Add something?
Value / time / distance / ...?
Multiply? Exaggerate?
Intensify?
MINIMIZE
Make it smaller / shorter /
lower / lighter / thinner / ...?
Leave out something?
Concentrate? Divide?
Create miniature?
SUBSTITUTE
Substitute component /
material / ingredient /
method / process / person?
Other places / times?
REARRANGE
Opposite / backward / inverted?
Deconstruct & reconstruct?
Change positive & negative?
Change cause and effect?
LENS + PHONE → CAMARA
DIRECT MAIL & PRINT
PICTURE + INFO
DATE LOCATION WEATHER ETC.
W/ BIG HANDLES TO KEEP STILL
BUILT-IN TRIPOD
BINOCULARS 3D-CAMARA
LENS = CAMARA
POCKET USB CAMARA
GLASS SEE-THROUGH CAMARA
TAKE PICTURES W/ GLASSES
NIGHT CAMARA
REMOTE SCREEN
FOR SELF SHOTS

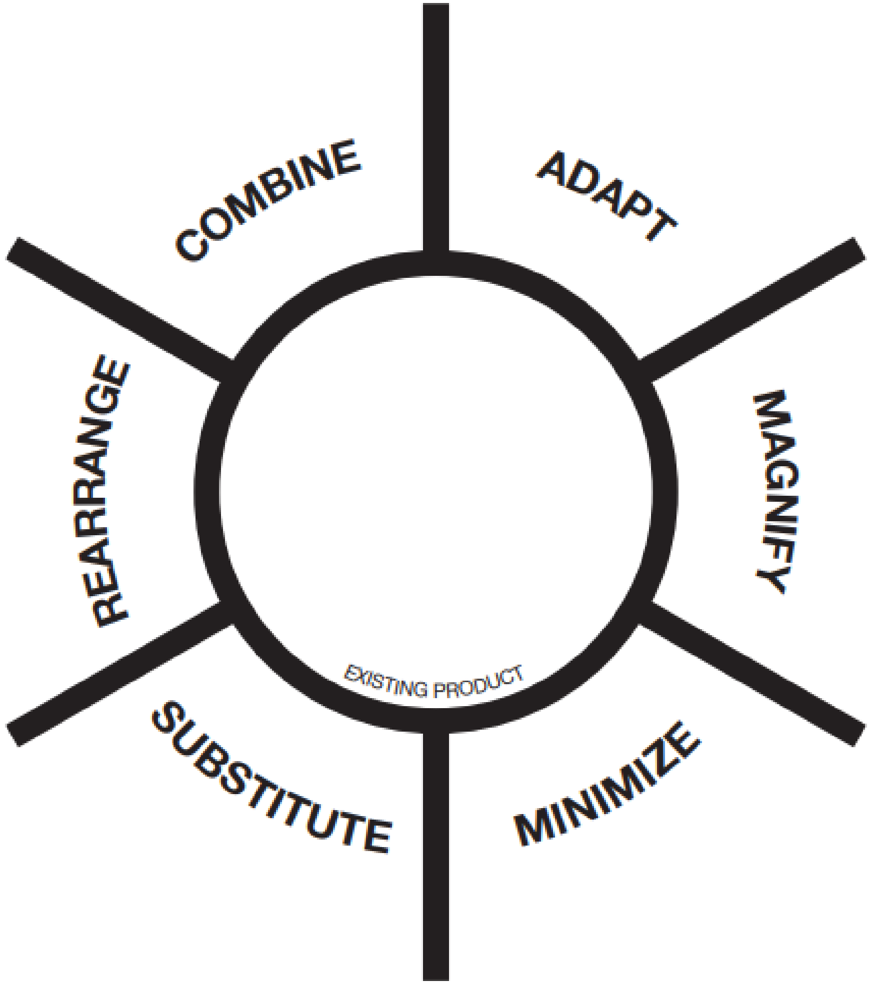
Example


EXAMPLE
OSBORN'S CHECKLIST
Manual Thinking
www.manualthinking.com
LOLLY POPS
1
MODIFY
2
MAGNIFY
3
MINIMIZE
4
SUBSTITUTE
5
REARRANGE
6
COMBINE
CO-BRANDING WITH TIC TAC
ROUND
FLAVORS BY COLORS (+ PANTONE)
SEASONED VERSION FOR ADULTS
STICK = THIMBLE
IMPACT!
CHILLY FLAVOR
HOT AIR BALLOON PROMOTION
UPGRADE ITS CATEGORY
DIOR
LASTS LESS TIME
AIR FORMULA
MINI CANDY
EASY OPEN
LOLLYPOP NECKLACE
CANDY IN TUBE
PACK = WRAP

Example

MANUAL THINKING USED TO CREATE AN OSBORN CHECKLIST
YOGURT DRINK WINTER VERSION
26.01.2012
Manual Thinking
REVERSE
COMBINE
ADAPT TO SIMILAR THINGS
PUT TO OTHER USES
MODIFY
MAGNIFY
MINIMIZE
SUBSTITUTE
REARRANGE
IN A TUBE FOR DRESSINGS & TOPPINGS
YOGURT DELUXE GOURMET DESSERT
PROMO PACK WINTER + A FREE HAT
YOGURT INSIDE CHOCOLATE EGG FOR KIDS!
YOGURT REFORCE YOUR DEFENSE
YOGURT CUPS TO DRAW ON
WINTERS YOGURT = FROZEN
SUMMER VERSION!
+ 40°C
YOGURT FOR MICRO WAVE
IN CASE YOUR HAVING A COLD
FAMILY PACK
DISPENSER
YOGURT PILLS
SALE IN PHARMACY
+ ADDED VALUE
FRESH STRAWBERRIES + YOGURT SAUCE
Actimel

Exercise



https://manualthinking.com/wp-

Forced Relationships

- Generate ideas by relating seemingly unrelated ideas
  - Use combining concept from brainstorming
  - Ideas may be related yielding more mundane solutions
- Use ideas related to the problem and possible to each other if more practical ideas are required
  - Less effort validating the ideas
- Start with the more general ideas and increase the specifics used later
- Somewhat related is “Attribute Listing”

Morphological Analysis

- Develop a grid of attributes along several dimensions
- Examine combinations of attributes
- Try to determine a solution/alternative to each combination
- Really provides a framework within which to screen all combinations and determine the most appropriate combinations
- Strategy-generation table is closely related

Creative Morphology

- MORPH
  - Shape/Form
  - Appearance
  - Composition
  - Layout
  - Structure
- OLOGY
  - Knowledge About

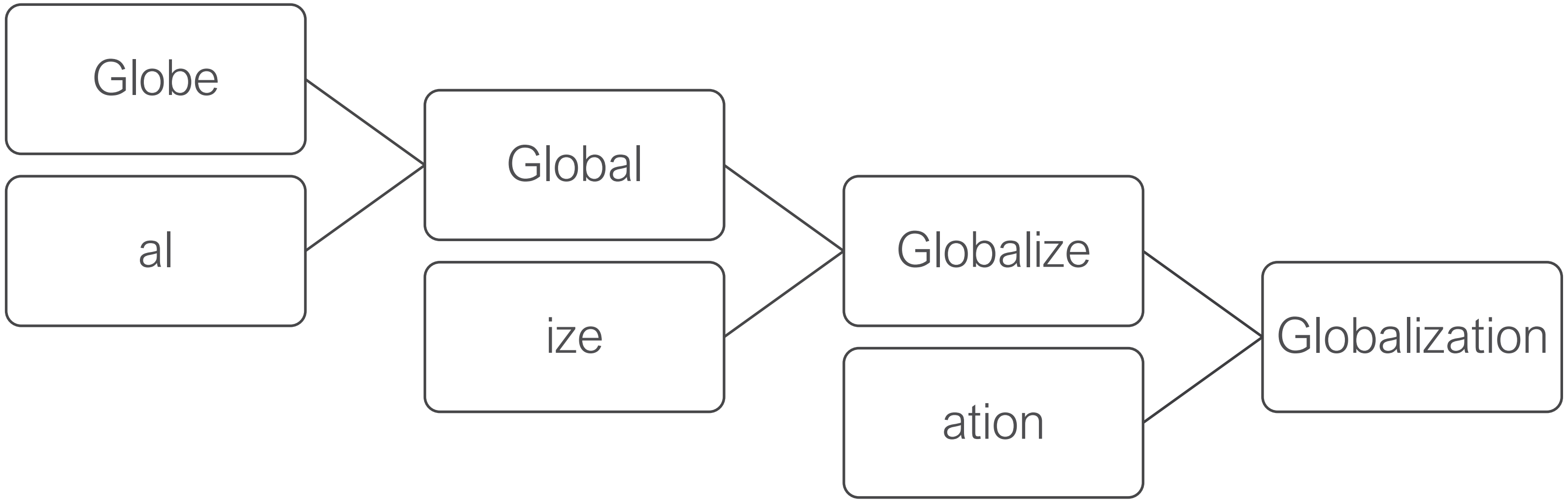
- Creative Morphology
  - The study of the form(s) of things
  - Break and put together in a new way
- Morphology, the study of pattern and form, is crucial to design because it constitutes an essential part of its corpus of coherent knowledge.

Creative Morphology

ㅇ	+	ㅘ	+	ㄴ	=	완
ㅅ	+	ㅓ	+	ㅇ	=	성、
ㅈ	+	ㅗ			=	조
ㅎ	+	ㅏ	+	ㅂ	=	합
ㅎ	+	ㅏ	+	ㄴ	=	한
ㄱ	+	ㅡ	+	ㄹ	=	글
						。

완성, 조합 한글.

Creative Morphology



Creative Morphology

Creative Morphology

Design Thinking

DESIGN THINKING

DESIGN THINKING

ỗດວŸບỷ ỗສດỢ́ປ̣ະỢ́ờssỗsŸŸ

- 1969 : The Sciences of the Artificial (Thinking)
- 1973 : Experience in Visual Thinking (Engineer)
- 1980 : How Designer Think (Designer)
- 1987 : Design Thinking (Creative)
- 2004 : d. School Stanford University

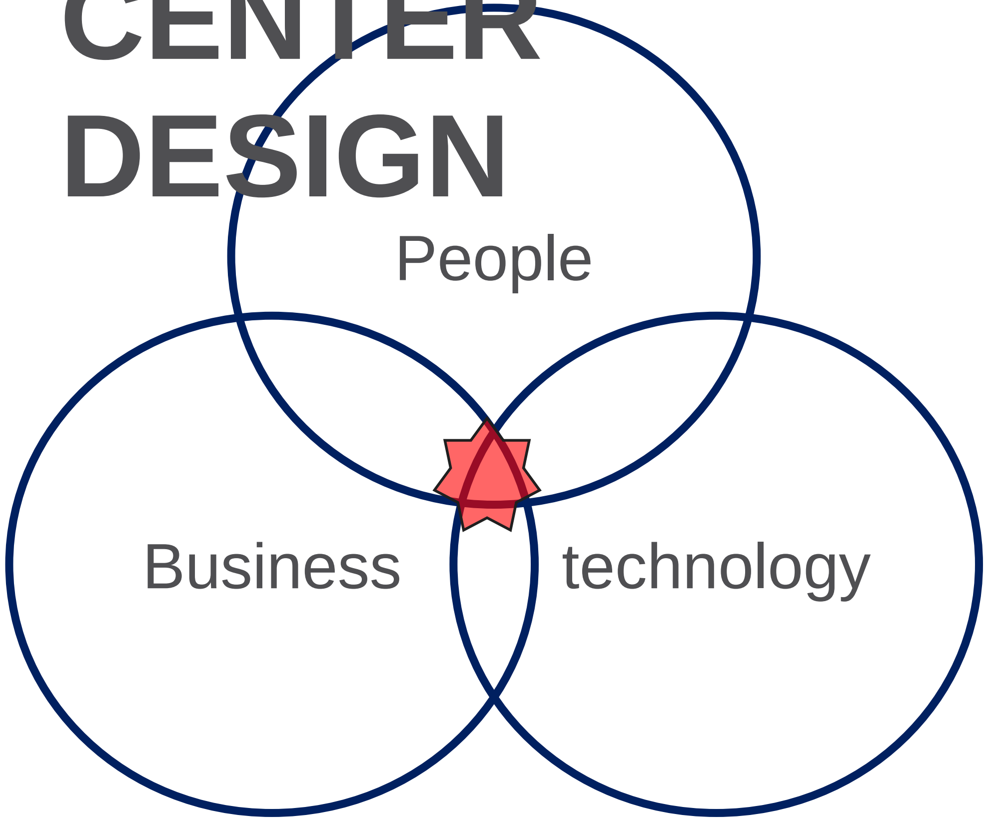

NOTHING IS A MISTAKE.
THERE'S NO WIN
AND NO FAIL.
THERE'S ONLY
MAKE


524A

"**Design thinking** is a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success."

Tim Brown, President and CEO, IDEO

HUMAN-CENTER DESIGN



Oral-B®

DESIGN
THINKING

DESIGN
THINKING

Oral-B®

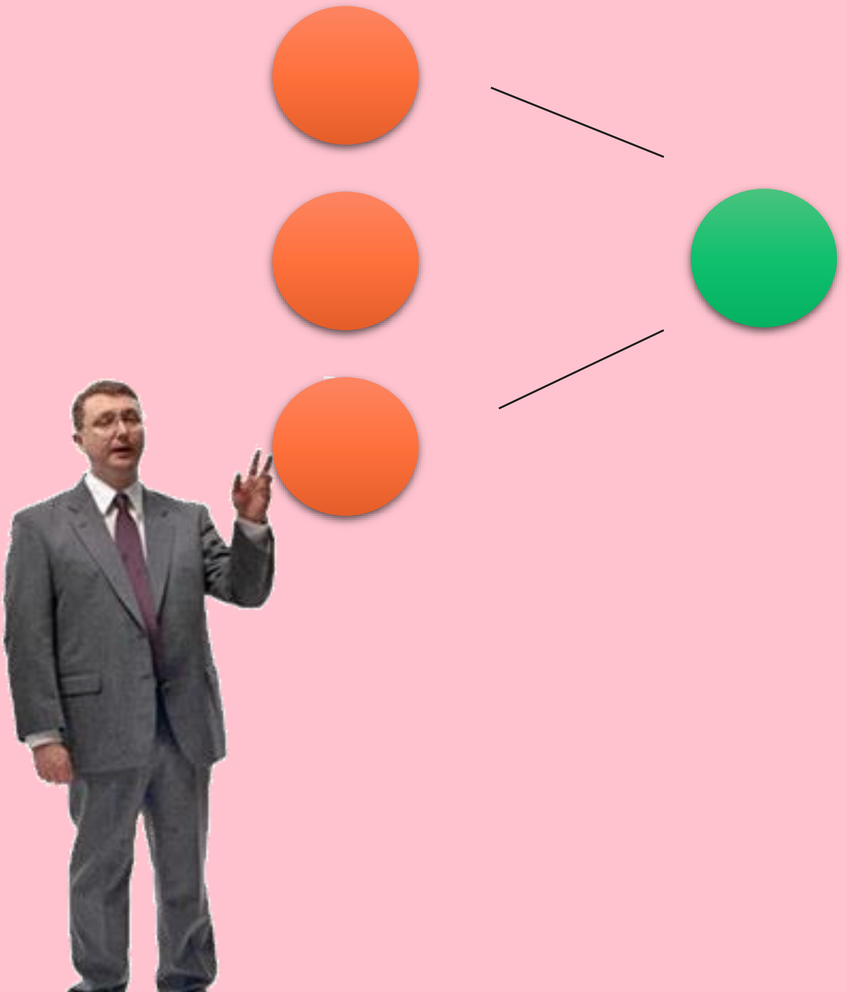
Oral-B

DESIGN
THINKING

NEED & INSIGHT

Business Thinking

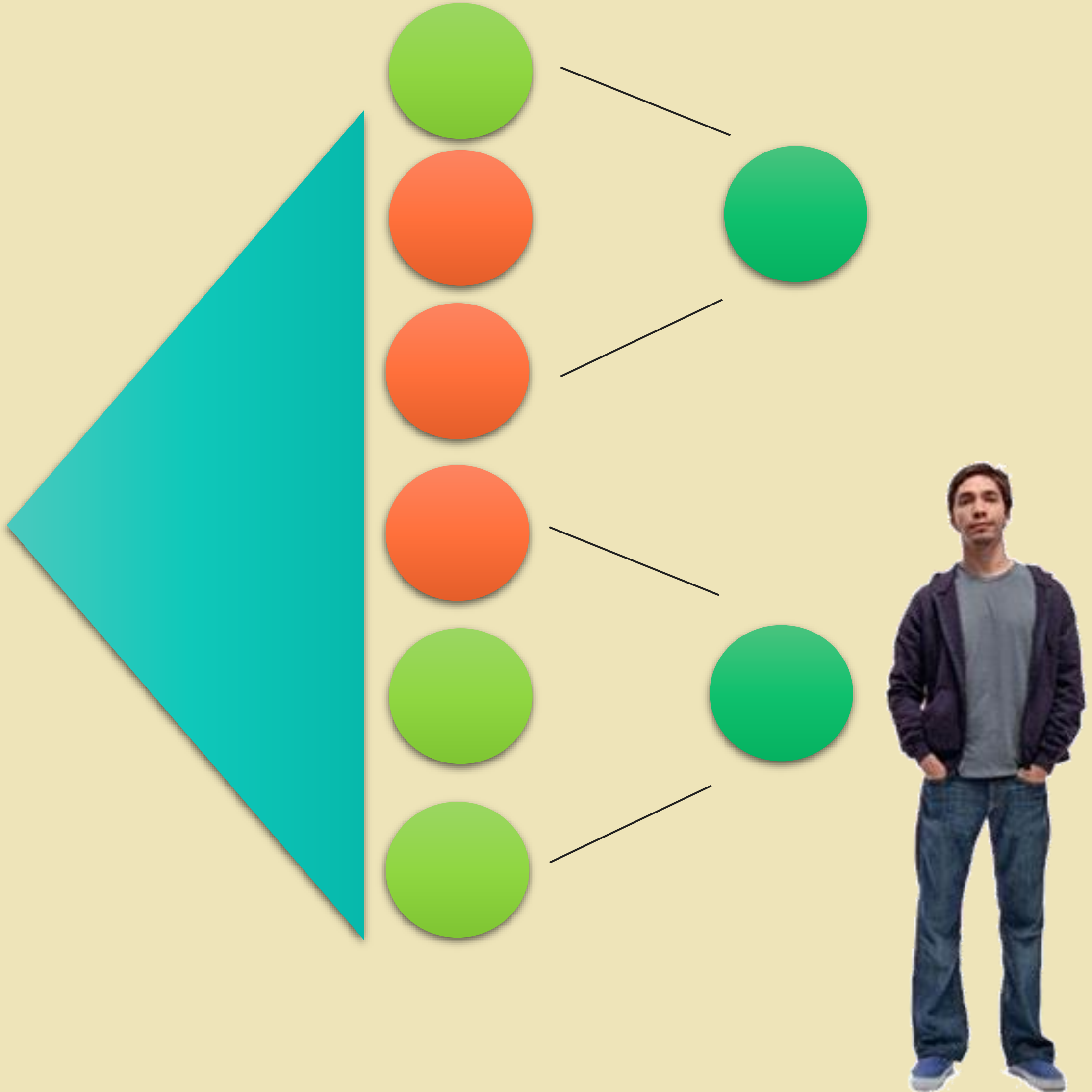
Problem Solution



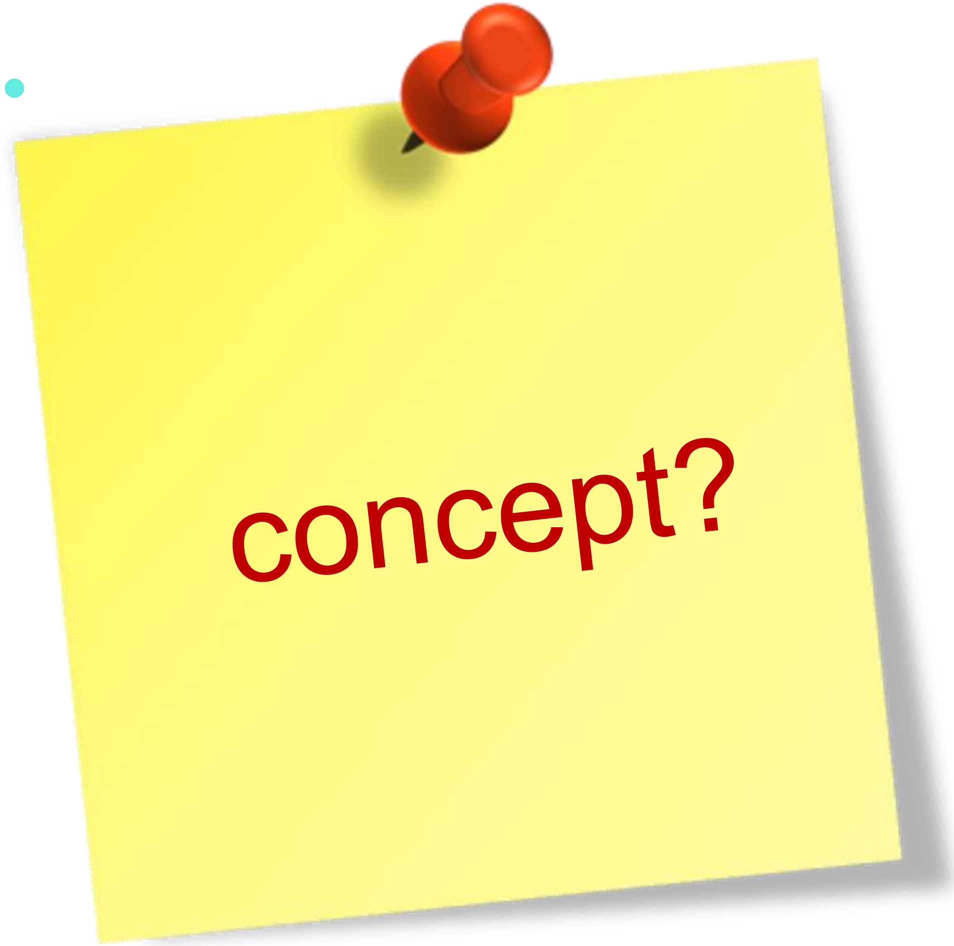
Design Thinking

Understand Problem Solution

People

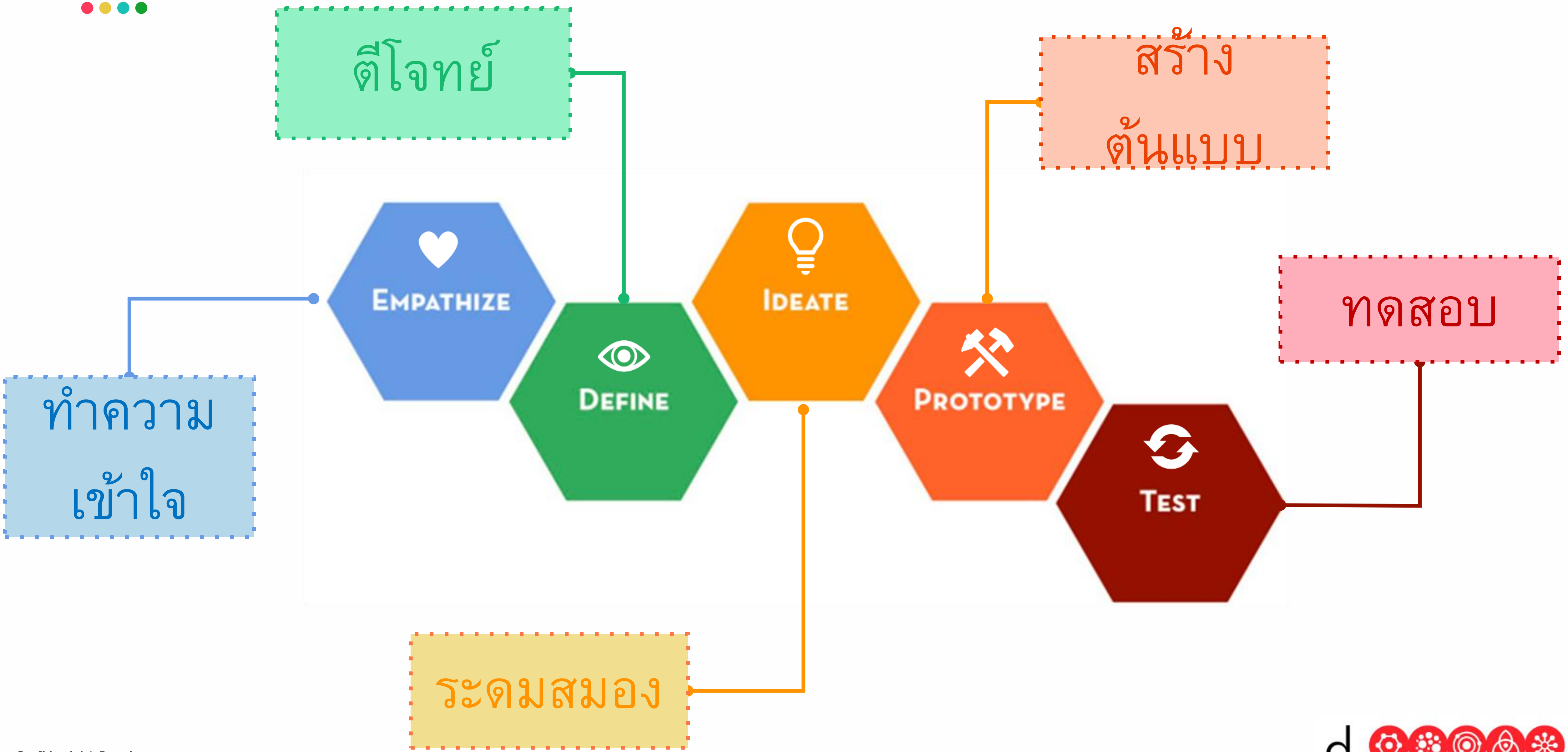


DESIGN THINKIN G



DESIGN THINKING

Design Thinking Mode



Activity

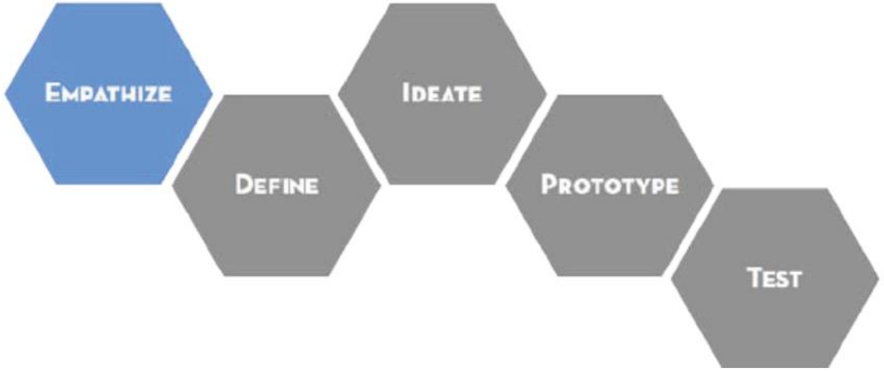
Find your partner

1 Empathize

ทำความ

เข้าใจ

Empathize



What ?

การเอาใจเขามาใส่ใจเรา
หรือ
การเข้าใจกลุ่มเป้าหมาย

Why ?

ปัญหาที่เราพยายามแก้ไข
ส่วนใหญ่ ไม่ใช่ปัญหาของตัว
เราเอง แต่เป็นปัญหาของ
ผู้อื่น

Interview

Empathize

Noted from your interview

Name:

Profile:

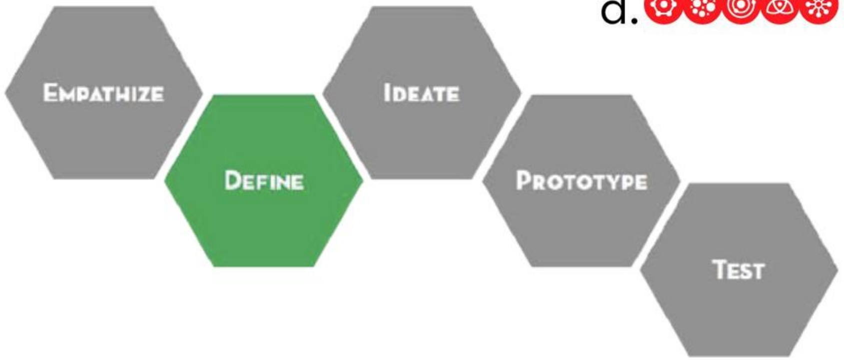
Life Style:

2

Define

ตีโจทย์

Define



Why ?

กำหนดกรอบปัญหาเพื่อนำไปใช้กระตุ้นไอเดียในการหาทางออกให้ปัญหาต่อไป

How?

- Storytelling
- Capturing key points
- Establish a point-of-view

กำหนดกรอบของปัญหา

DEFINE

User Description

Need a way to (user's need)

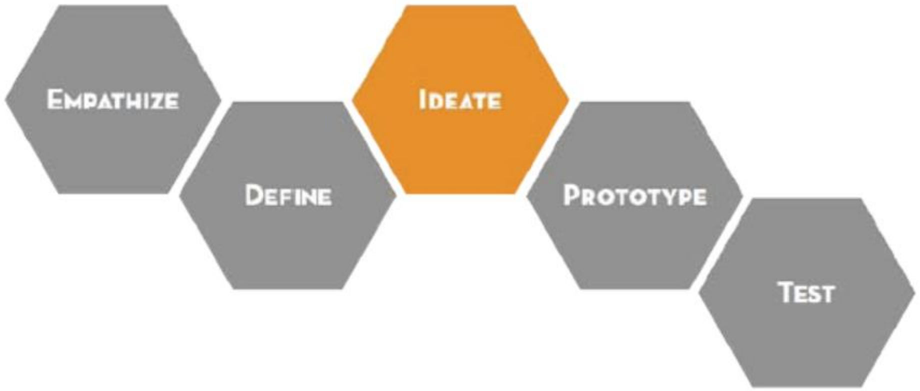
Surprisingly/ Because/But (user's insight)

3

Ideate

ระดมความคิด

Ideate



What ?

- Generating Ideas
- Go for volume
- Not Evaluating Ideas
- Write down your idea

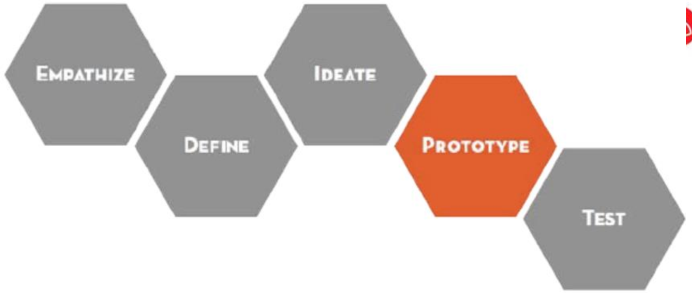
IDEATE

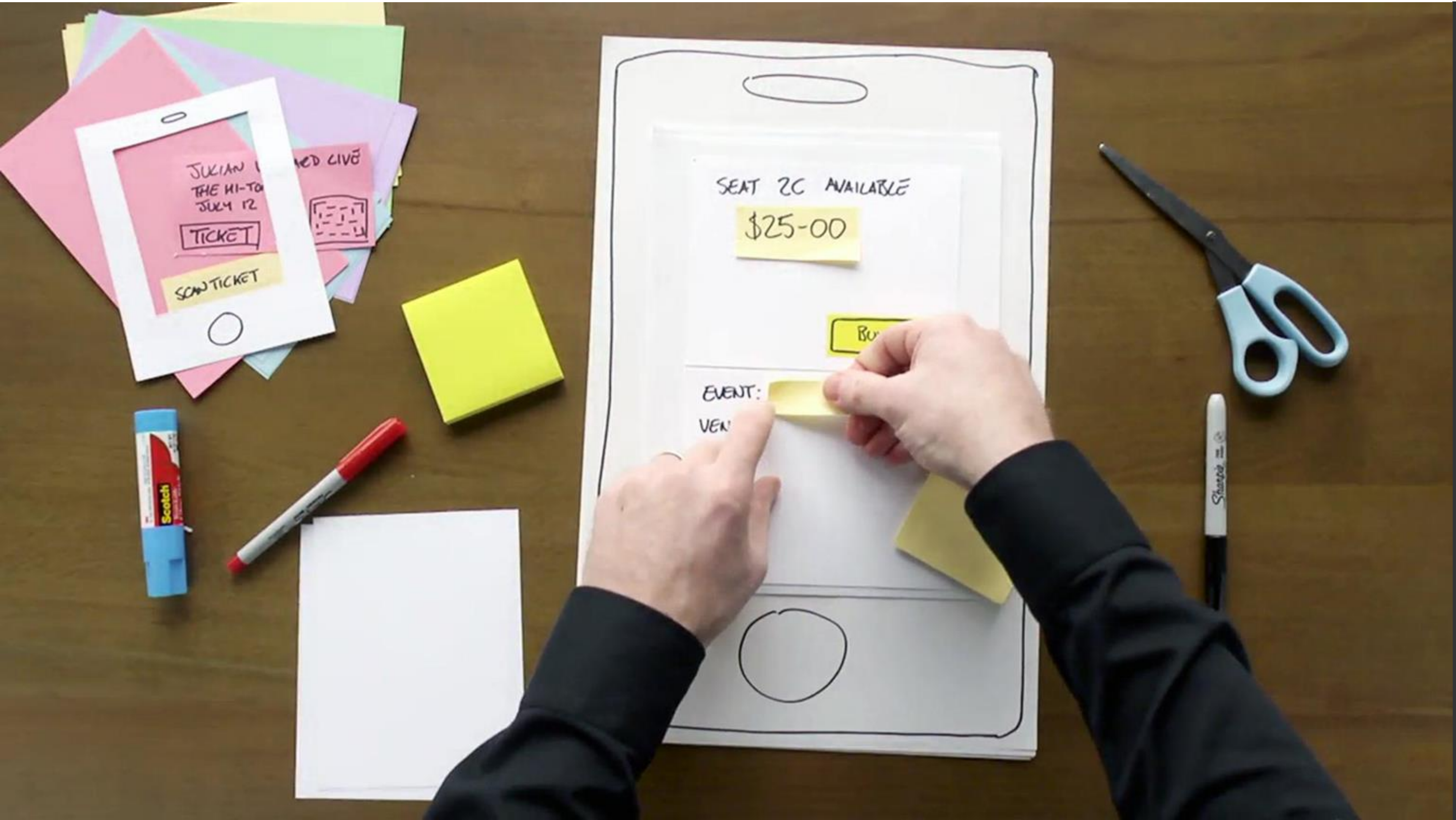
Sketch solution to meet your user’s needs

4 | prototype

สร้างต้นแบบ

prototype





Make Ideas Tangible

Draw your prototype

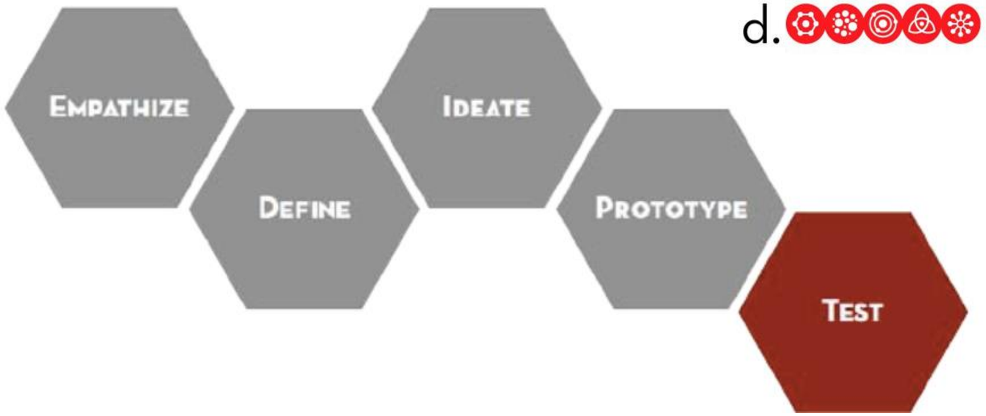
PROTOTYPE

5

test

ทดสอบ

test



Why ?

- To refine your prototypes and solutions.
- To learn more about your user.
- To test and refine your POV.

Share your solution and get feedbac

test

Like

Dislike

Question

Idea