COLUMBIA BUSINESS SCHOOL

CS@CU

# Data Science and Technology Entrepreneurship

(Spring 2013)

## Business Model, Customer Validation, Linear Classifiers, Intellectual Property

Sameer Maskey

Week 2

# Announcement

- Looking for the 3rd TA
  - Computer Science
  - Experience in Machine Learning, MapReduce, Web programing
- If you have friends who may be interested let them know

# Topics for Today

- Guest Lecture
  - Intellectual Property, Patents, Columbia Technology Ventures
- Business Model
- Customer Validation
- Linear Classifiers

# Guest Lecture

- Orin Herskowitz
- Executive Director, Columbia Technology Ventures (CTV)

# Assignment 1 - Due Friday

- Assignment 1 was posted on the course website
- Due this Friday @6pm

- Submission Method
  - Each team create a google id
  - teamXYZ@gmail.com
  - put your doc in teams google drive
  - send us a shared link (email Sameer, Morgan and Jigar)
  - **Email Subject Line :**
    - **DSTE - Assignment 1 - TeamName**

# Extra Classes : Web Programming 101

- Starts in 1 or 2 weeks
- 2-3 extra lectures
- 1-2 hour lecture each
- Fridays @3:30 pm
- Web Programming Basics
  - If you know how to build web applications you don't have to come
  - Geared for MBA students who want to know bit more behind technology that runs new web startups
  - Geared for CS/Eng students who know programming but not necessarily web programming

# Teams

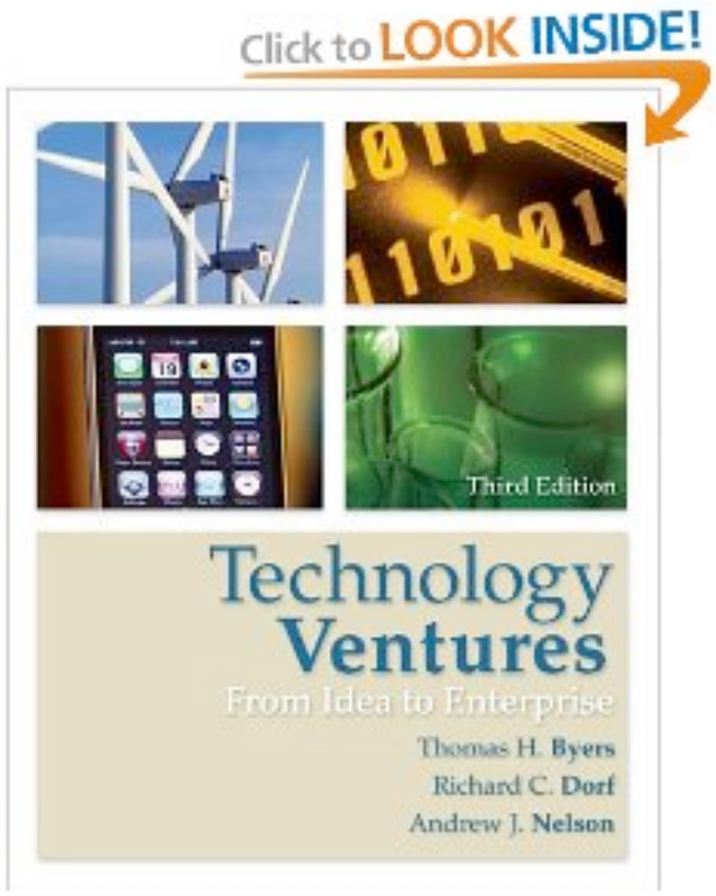
- Minimum (1 MBA + 1 CS)
- Maximum (2 MBA + 2 CS) - Not a strict rule
- Team of 4+?
  - Let me know

# Team Name

- As a part of your assignment you are supposed to come with a name for your team/start up
  - Test the name with your friends and family
    - Right name?
  - domain available?
  - social handles - facebook, twitter, handles available?

# Reminder : Textbook

- Technology Ventures: From Idea to Enterprise, 3rd edition
- Thomas Byers (Author), Richard Dorf (Author), Andrew Nelson (Author)



# Course Stages

**Stage 1** (3 weeks – Jan 30 – March Feb 20) Problem definition, Data collection, Customer development, Business Model Canvas, Data science methods for testing your hypothesis

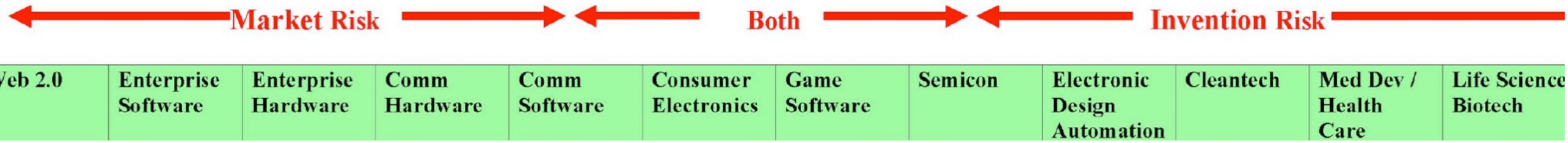
**Stage 2** (5 weeks – Feb 4 – March 10) Minimum Viable Product development, Quantifying customer feedback with classification and clustering techniques

**Stage 3** (2 weeks – March 11 – March 31) Agile development, Data analysis of feature surveys, Sequential prediction algorithms (costs, revenue, traction)

**Stage 4** (2 weeks – April 1 – April 29) Launching the product, Data driven marketing techniques, A/B testing

**Stage 5** (2 weeks – April 1 – May 5) Try to raise funds with VC network provided in the class

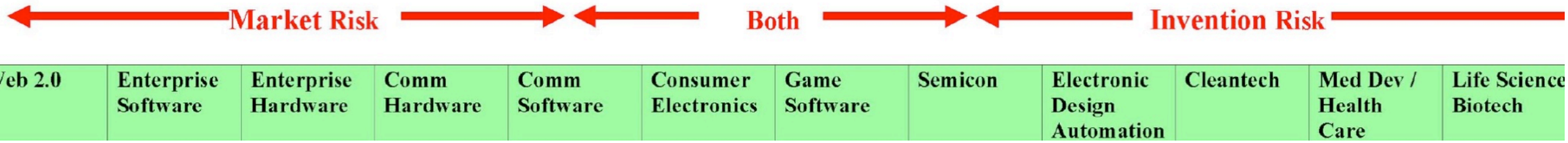
# Customer/Market Risk vs Invention Risk




Source : Steve Blank

Where does your startup fall?

# Customer/Market Risk vs. Invention Risk




Source : Steve Blank

More Invention Risk you have,
more important Intellectual Property become

# Guest Lecture

- Orin Herskowitz's slides are posted separately in the course website

# Business Model and Startups

- When you try to build a startup you can go through a business design exercise [Byers, et. al]
  - who are the customers?
  - how are needs of customers satisfied?
  - how are profits captured and profitability protected?
- What is a business model?

# Business Model and Startups

- When you try to build a startup you can go through a business design exercise [Byers, et. al]
  - who are the customers?
  - how are needs of customers satisfied?
  - how are profits captured and profitability protected?
- What is a business model?

> "A business model is a set of planned assumptions about how a firm will create value for all its stakeholders." [Magretta 2002]

# Business Model Hypotheses

- When you have a startup you have assumptions

# Business Model Hypotheses

- When you have a startup you have assumptions

My users will share funny pictures at least 20 times
My value to the client is the speed of our system
Our recommendation engine works well

.

.

.

Our revenue will hit 200K US$ in 4 months

# Business Model Hypotheses

- When you have a startup you have assumptions

My users will share funny pictures at least 20 times
My value to the client is the speed of our system
Our recommendation engine works well

.

.

.

Our revenue will hit 200K US$ in 4 months

These assumptions need to be validated
In fact, in the beginning, many assumptions tend to be off the mark or outright wrong

# Business Models - Variation in View

- Different interpretations on what constitutes a business model
- What assumptions are relevant?
- We are going to focus on ones that are particularly relevant for startups
  - Elements of Business Model [Byers, et. al]
  - Business Model Canvas [Osterwalder, A]
  - Lean Canvas [Maurya, A]

# Elements of Business Model [Byers, et. al. ]

| | |
|---|---|
| ■ **Customer selection:** | Who is the customer?<br>Is our offering relevant to this customer? |
| ■ **Value proposition:** | What are the unique benefits? |
| ■ **Differentiation and control:** | How do we protect our cash flow and relationships?<br>Do we have a sustainable competitive advantage? |
| ■ **Scope of product and activities:** | What is the scope of our product activities?<br>What activities do we do, and what do we outsource? |
| ■ **Organizational design:** | What is the organizational architecture of the firm? |
| ■ **Value capture for profit:** | How does the firm capture some of the total value for profit?<br>How does the firm protect this profitability? |
| ■ **Value for talent:** | Why will good people choose to work here?<br>How will we leverage their talent? |

# Business Model Canvas [Osterwalder, A]

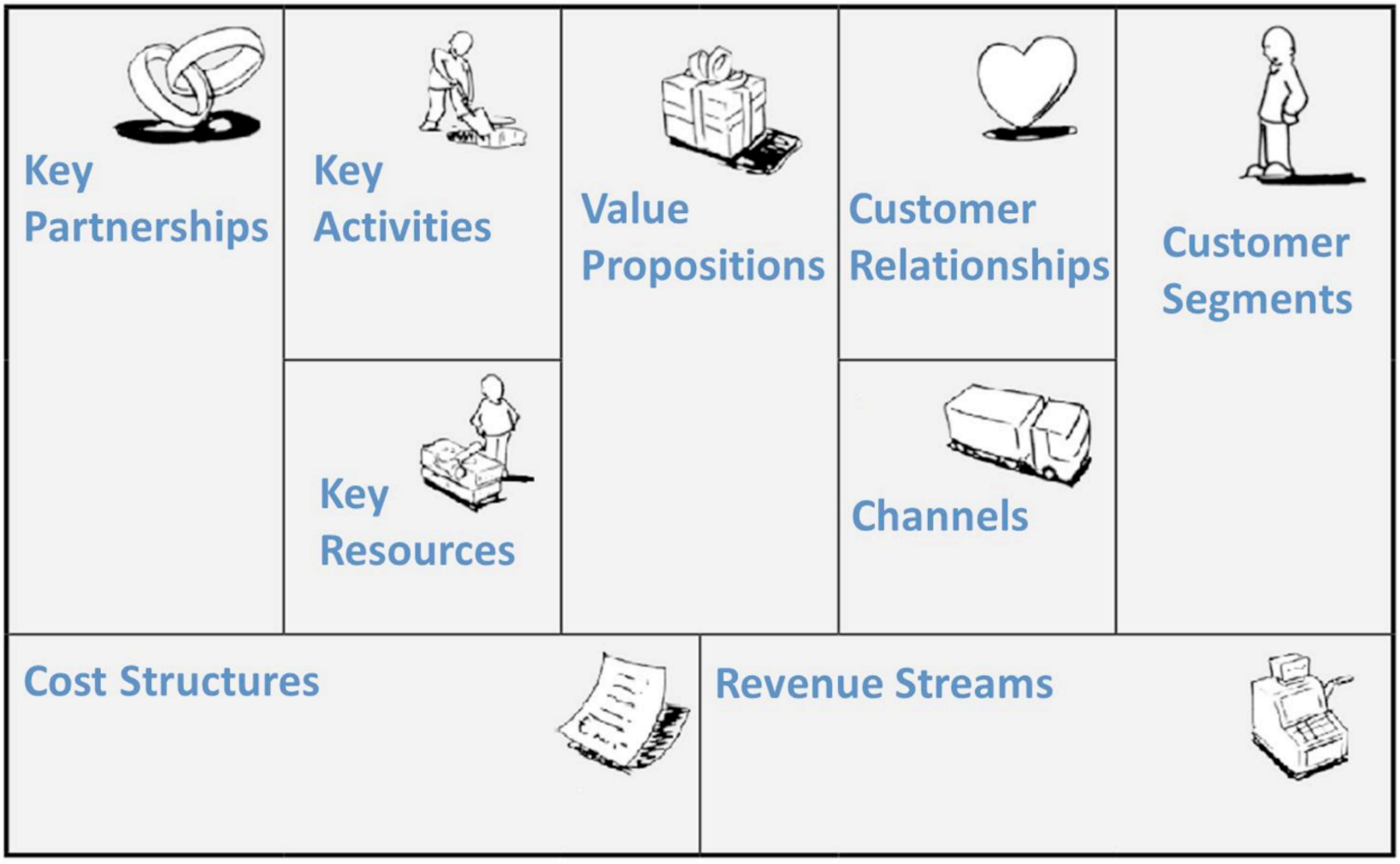
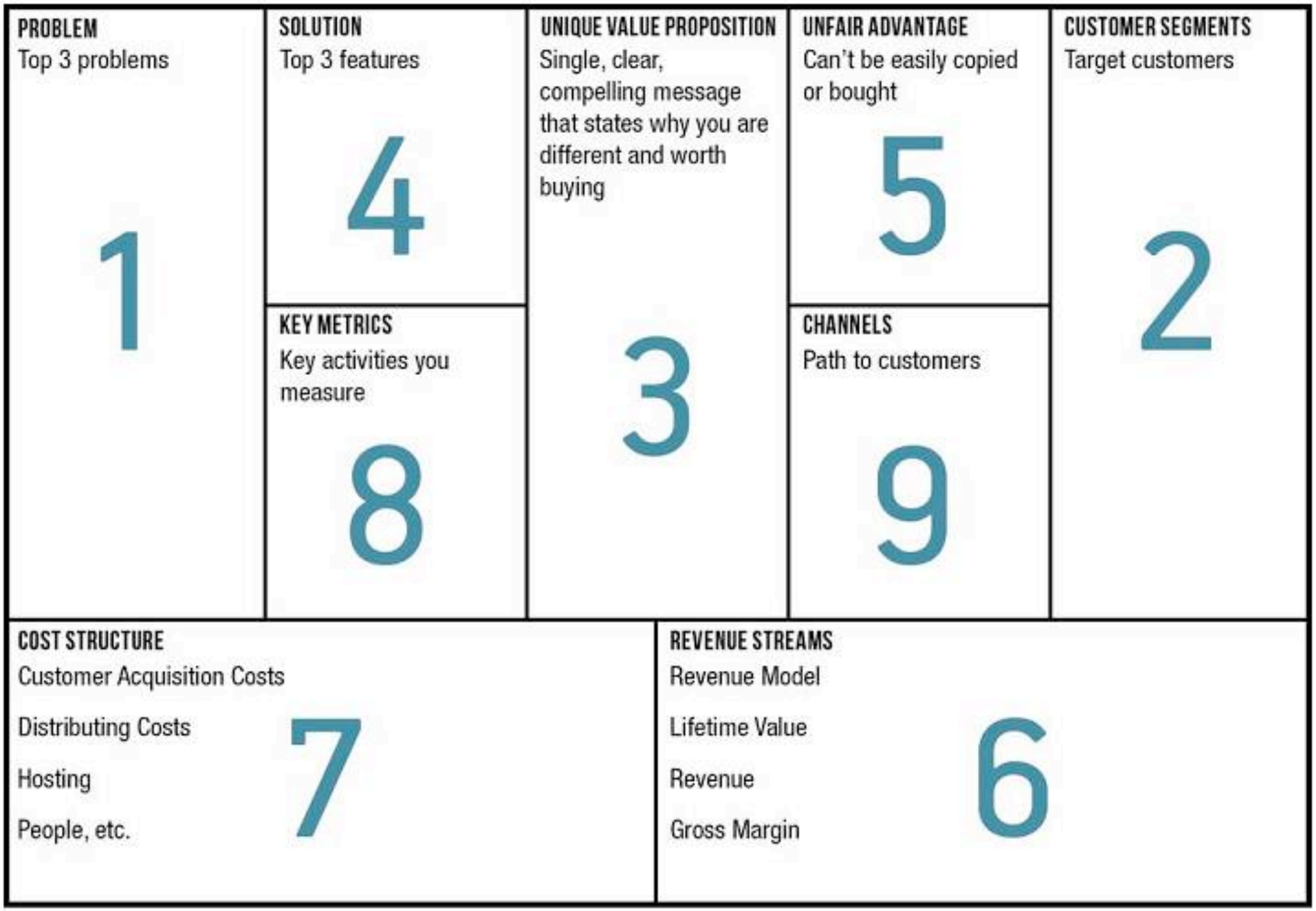
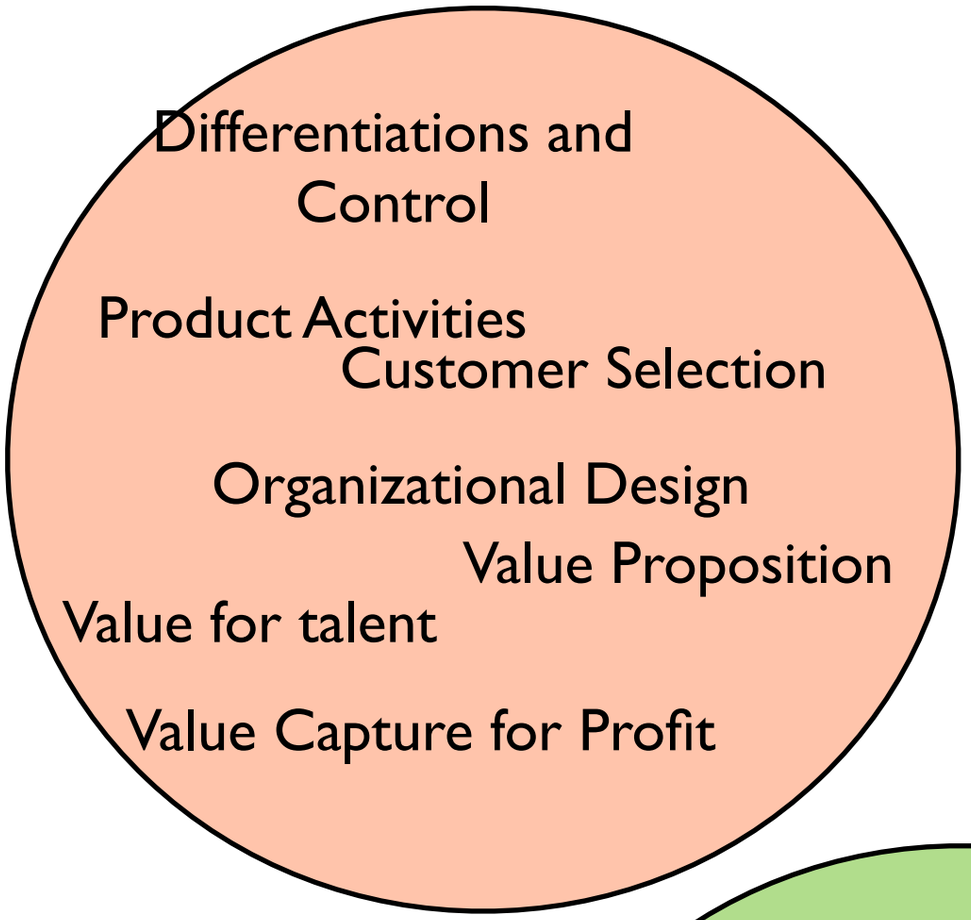



Image Source : Web [FiveWhys]

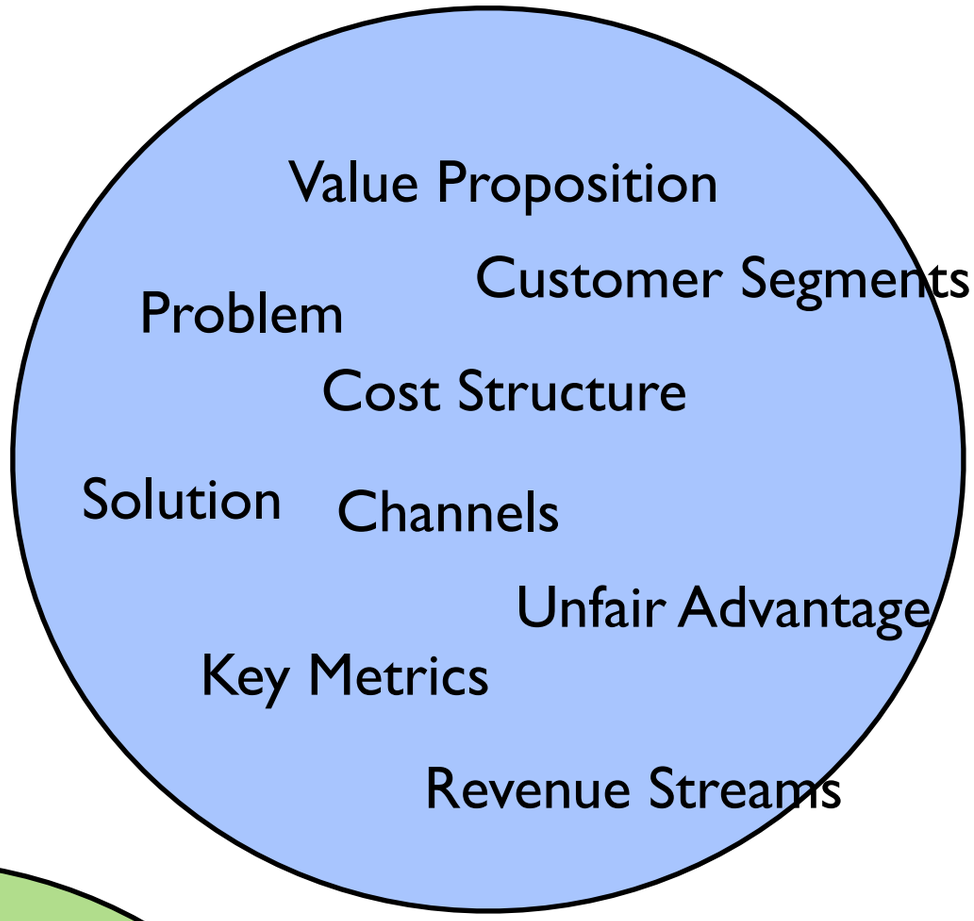
# Lean Canvas [Maurya, A]



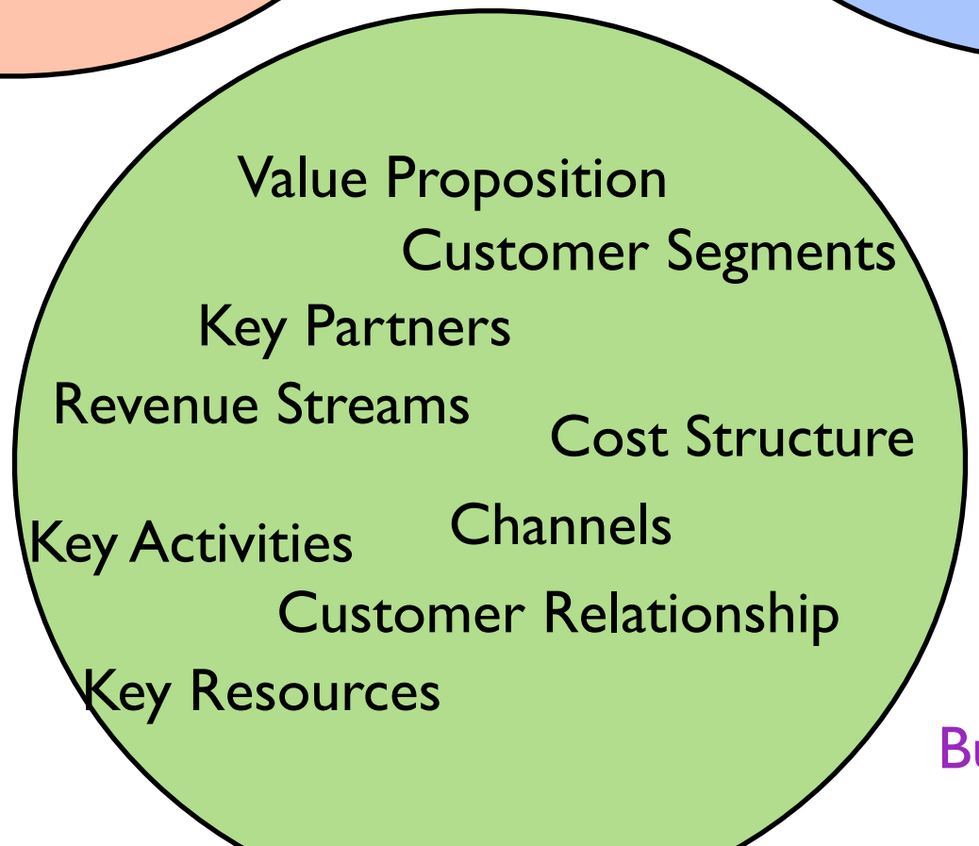
# Common Concept Across Frameworks




Elements of Business Model

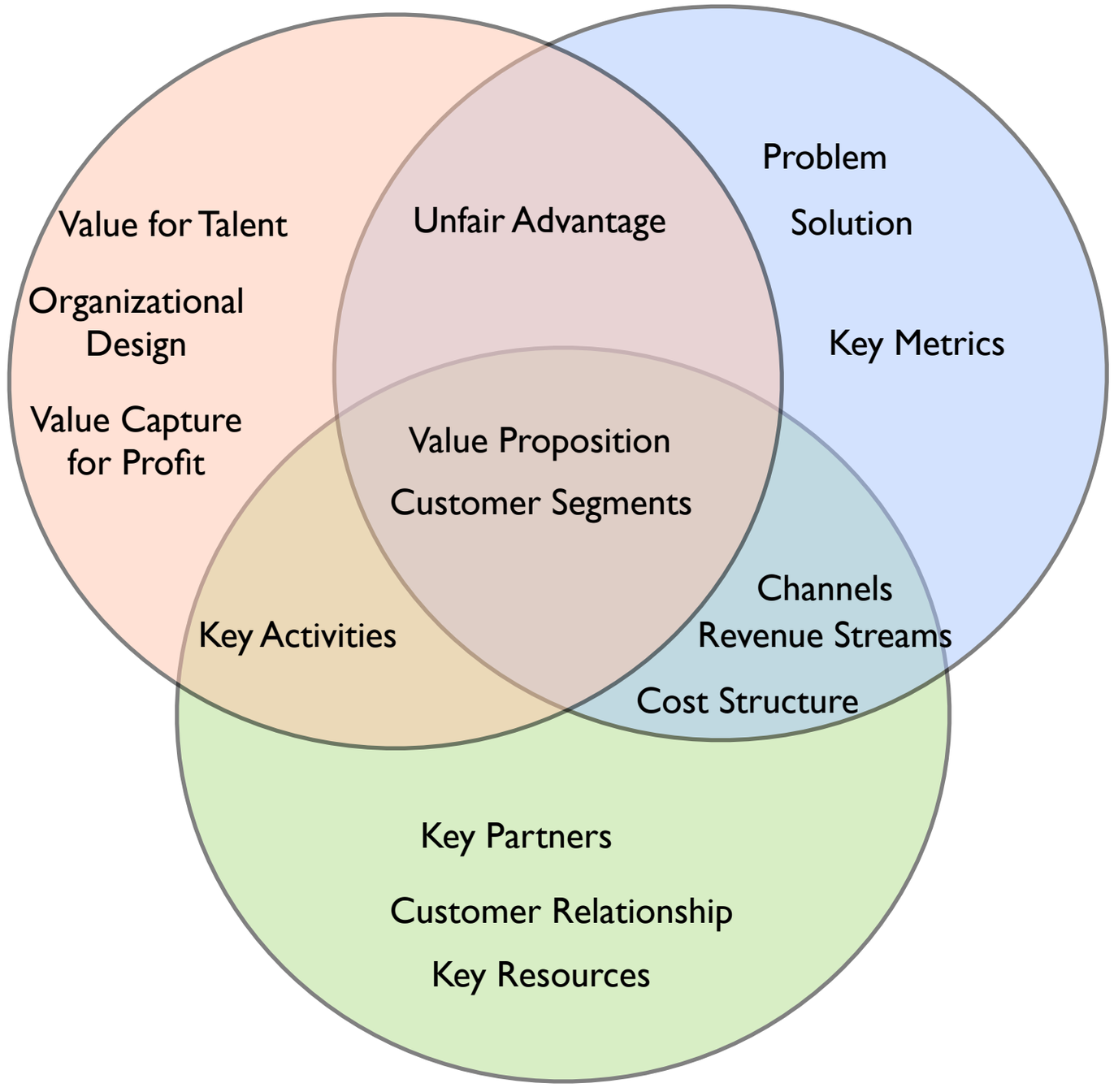



Lean Canvas

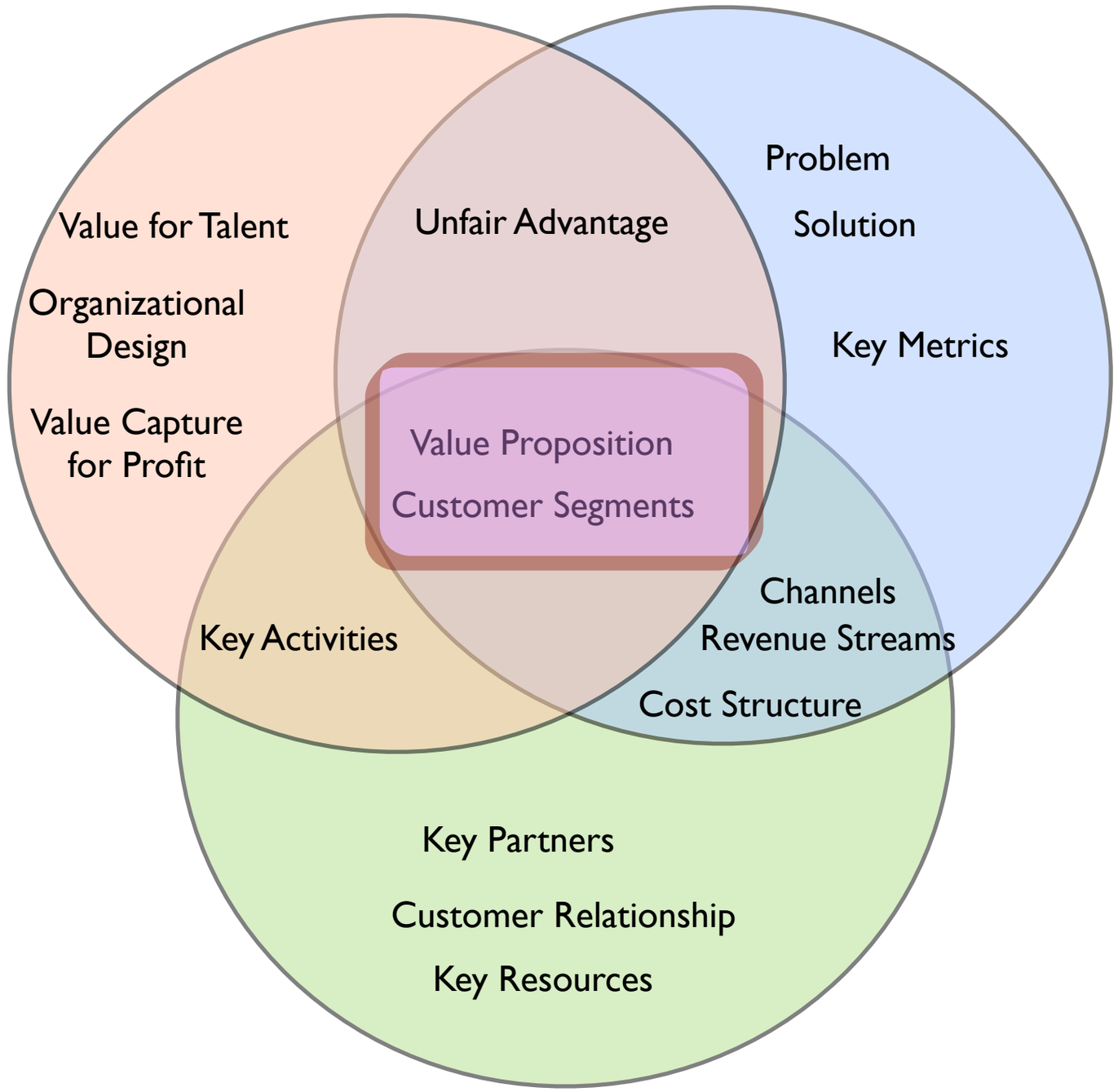



Business Model Canvas

# Common Concept Across Frameworks



# Common Concept Across Frameworks



# Value Proposition & Customer Segments

- This is common theme across all business model frameworks

- 2 important questions to ask yourself
  - Who are your customers?
  - What is the value proposition for them?

# Customers

- Customer Segments
  - Who are your customers?
  - Mass market? Niche?
  - People, Small Businesses (SMEs), Large Corporations?
    - Business travelers?
    - SMEs in retail industry?
      - Food?
        - Canned food?
          - Beans?
            - Organic?
              - ...
  - Customer segments can be dissected or categorized in many ways
- How to really find out your assumptions about customer segment are correct?

# Value Proposition

- Next relevant question is :Value Proposition for your customer?
- What features of your product provides unique value to your customer?
- Why should they use or buy your product?
  - Lowest price?
  - Make things easier?
  - Fastest?
  - Newest technology?
  - Customization?

# Value Proposition

- Founder Ken Croc: “McDonalds stands for friendliness, cleanliness, consistency, and convenience”
- Coca-Cola - unique taste
- Ryanair - extremely cheap
- Walmart - price
- Your company’s value proposition?

# Five Key Values

| | | |
|---|---|---|
| 1 | Product | Performance, Quality, Features, Brand, Selection, Search, Safe, Ease |
| 2 | Price | Fair, Visible, Consistent, Reasonable |
| 3 | Access | Convenient, Location, Nearby, At-hand, Easy to find, in a reasonable time |
| 4 | Service | Ordering, Delivery, Return, Check-out |
| 5 | Experience | Emotional, Respect, Ambiance, Fun, Intimacy, Relationships, Community |

Source : Tech Ventures book

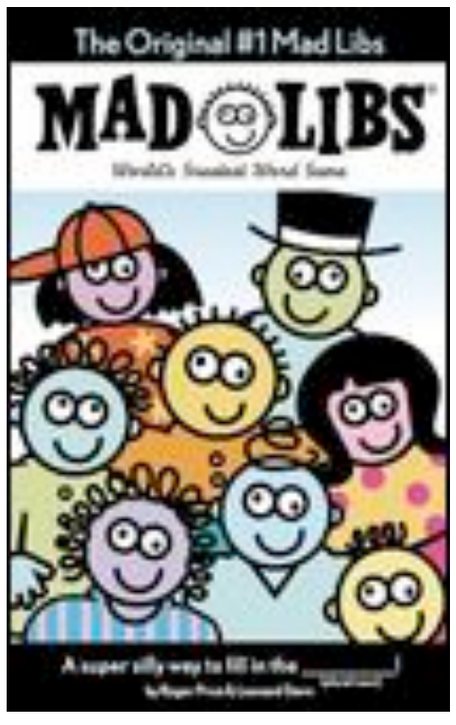
# Primary and Secondary Value

| | | Primary Value | | | | |
|---|---|---|---|---|---|---|
| | | **Product** | **Price** | **Access** | **Service** | **Experience** |
| **Secondary Value** | **Product** | —— | Wal-Mart | Amazon.com | Honda | Harley-Davidson<br>Disney World |
| | **Price** | Target | —— | Holiday Inn | Wal-Mart | Olive Garden |
| | **Access** | Google<br>Barnes & Noble | Priceline<br>Visa | —— | Dell Computer | Starbucks |
| | **Service** | Toyota<br>Home Depot<br>Intel | Southwest Airlines | McDonald's | —— | Carnival Cruise Line |
| | **Experience** | Mercedes | Virgin Atlantic<br>Best Buy | AT&T | Nordstrom | —— |

Source : Byers et. al. book

# Articulating Value Proposition

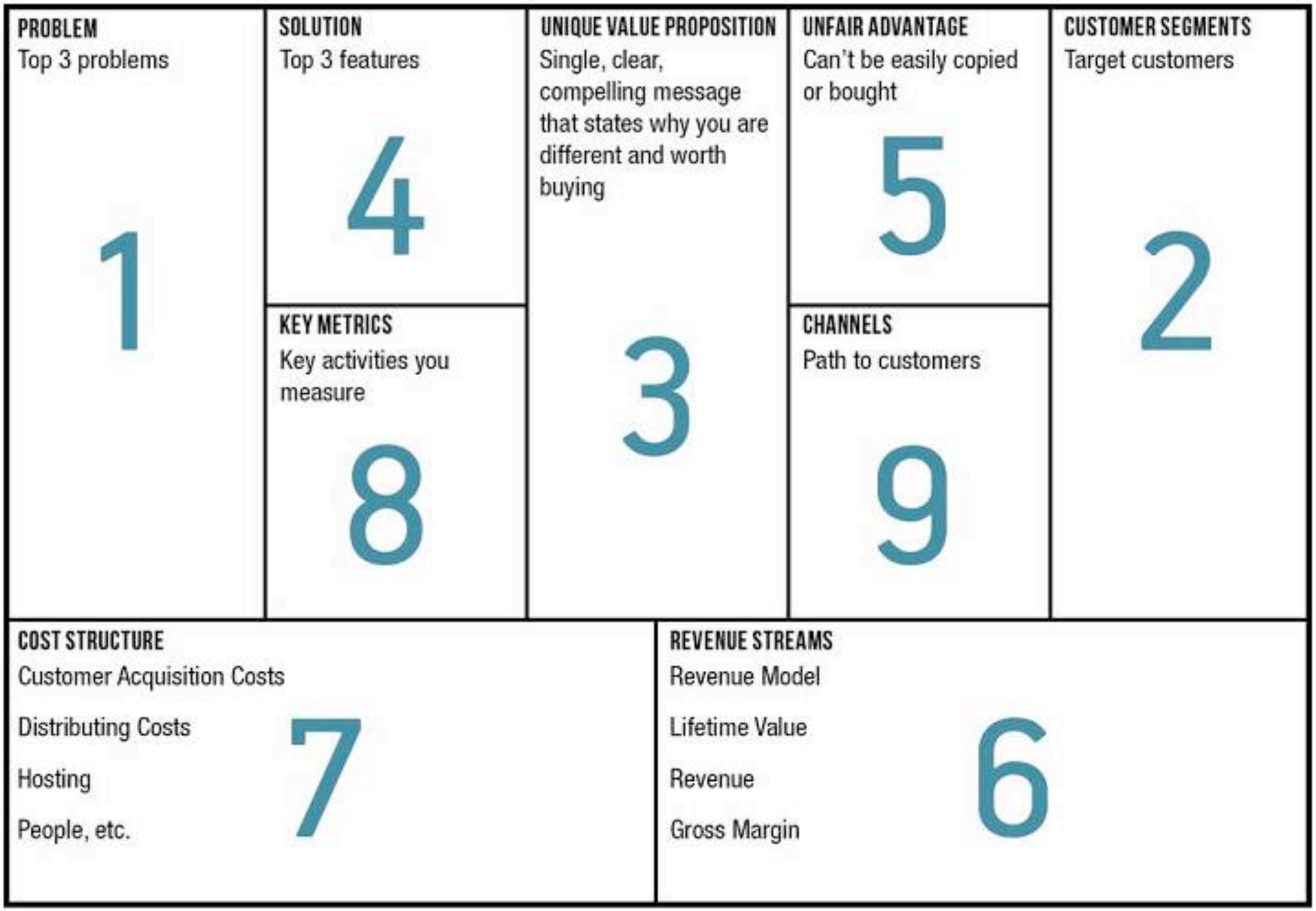
Our company, (company name),
is developing (a defined offering)
to help (a target audience)
(solve a problem)
(with secret sauce)




Adeo Ressi's Mad Lib Template

# Lean Canvas [Maurya, A]

https://docs.google.com/drawings/d/1RCczinVGbE1FJ0geyOwpGWWm5FYkvmLSXnRenf9dY_o/edit



# Are Your Assumptions Valid?

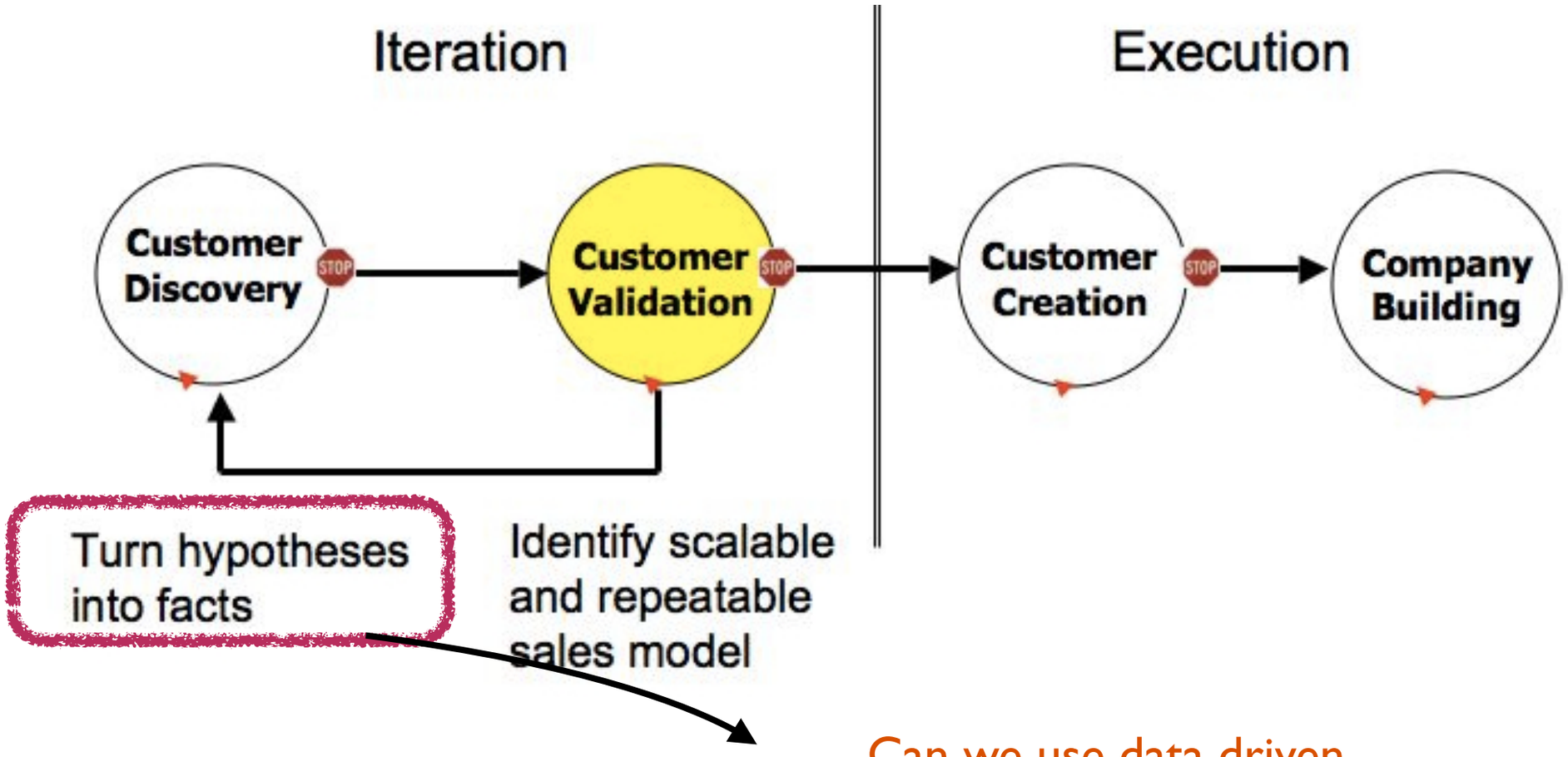
- Assume you have come up with a great value proposition
- How can you validate that value proposition is really valid?
- Validate by testing your hypotheses with customers!

Customer Discovery Process

For more read : Startup Owner's Manual

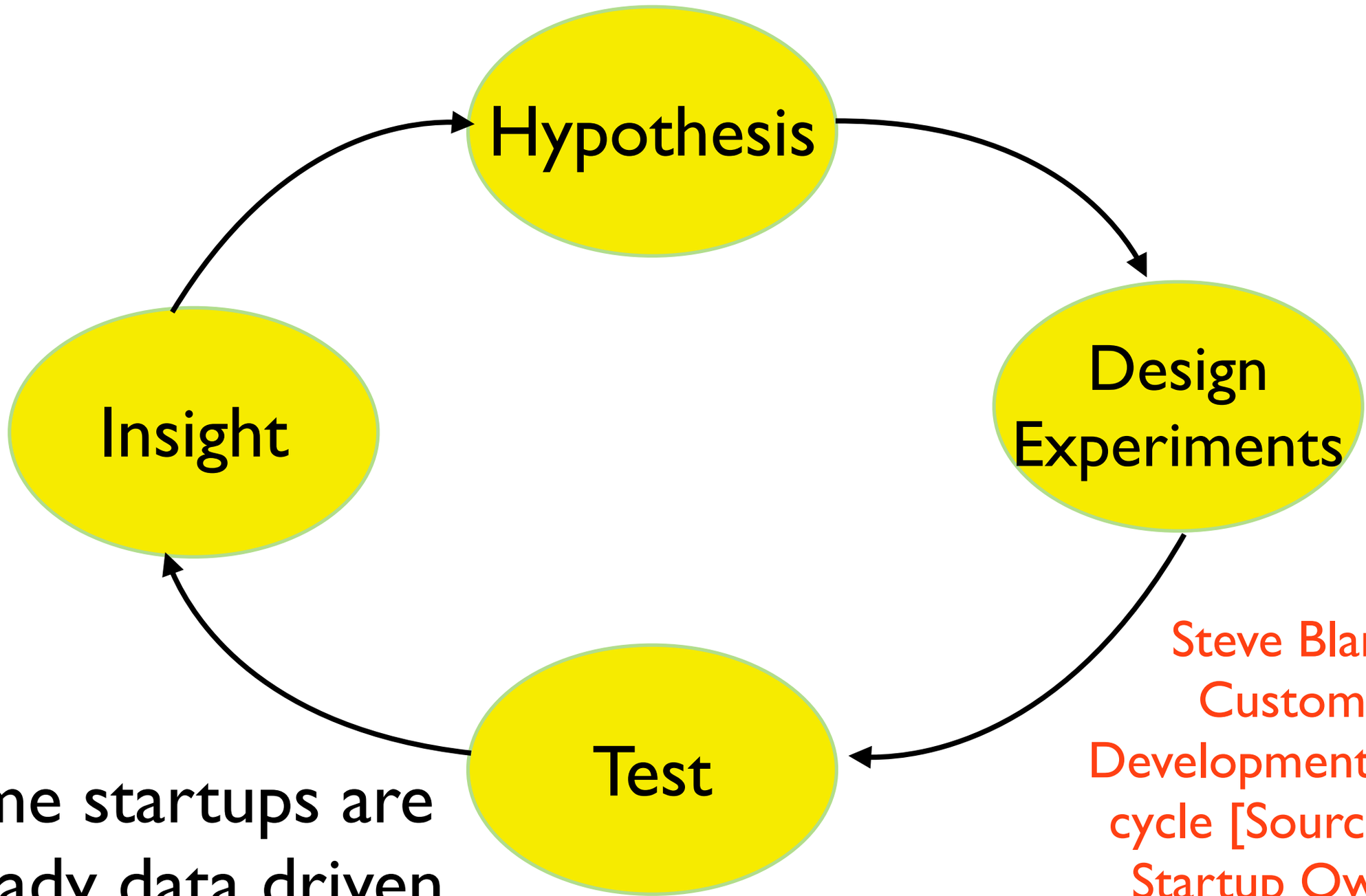
# Customer Discovery Process

- Source : Startup Owner's Manual - Steve Blank and Bob Dorf




Can we use data driven methods ?
What kind of methods?

# Data Driven Decision and Startups




Some startups are already data driven

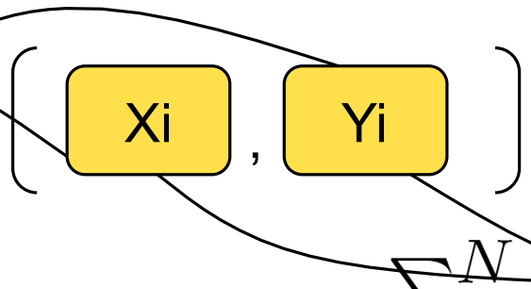
Steve Blank's Customer Development Insight cycle [Source : The Startup Owner's Manual]

# Recap: Data to Scores

- We looked at how we can use a linear regression model to predict if a product is likely to clicked
- Converted raw data (product information) into insights (likelihood of clicks)
- The method of scoring raw data can also be used
  - to test if your customer acquisition rate prediction is valid
  - to test if your pricing model is valid

# Data to Scores

- Raw Data => Processed Data => Model => Prediction Score
  - Given our training data
    - (1, 4)
    - (0, 2)
    - .
    - .
    - .
    - (1, 9)




**Training Our Regression Model:**
Just need to implement for loop that computes numerators and denominators in equations here. And we get optimal thetas

$$\theta_1 = \frac{\sum_{i=1}^{N} x_i y_i - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i^2 - \frac{1}{N}\sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i}$$

**For Prediction/Testing:**
Given optimal thetas, plug in the x value in our equation to get y

$$\theta_0 = \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\theta_1 \sum_{i=1}^{N} x_i$$

# User Data History

Product Clicks

(  , 4 )

Stauer Compendium Hybrid Watch
$29.00
+ No Shipping Info
STAUER

(  , 5 )

Michael Kors Watches Bradshaw Gold - Michael
$250.00
+ Free Shipping
eBags.

(  , 2 )

Men's Pro Diver Chronograph Black
$199.99
+ Free Shipping
World of Watches

# Process Raw Data

Product Clicks

(  , 4 )

**Stauer Compendium Hybrid Watch**
$29.00
+ No Shipping Info
**STAUER**

(  , 5 )

**Michael Kors Watches Bradshaw Gold - Michael**
$250.00
+ Free Shipping
**eBags.**

(  , 2 )

**Men's Pro Diver Chronograph Black**
$199.99
+ Free Shipping
**World of Watches**

Raw Data

Process

Count number of word "Watch" in Product Description

$(x_i, y_i)$

Product Clicks

( 1 , 4 )

( 1 , 5 )

( 0 , 2 )

Processed Data

# Data to Predicted Scores

Process

( Casio Men's PRW2500-1 Pathfinder Triple Sensor Tough Solar Digital Multi-Funtion Pathfinder … , ? )

Count number of word "Watch" in Product Description

( 1 , ? )

Raw Data

Processed Data

$$f(x)=\theta_0+\theta_1 x$$

$$\theta_0=0;\ \theta_1=2$$

Our Model

2

This watch will be potentially clicked twice

# Business Model Assumptions

- The company Zoolaster sells Zoola watches online
- Zoolaster buys watches from wholesaler for cheaper price and sells them online
- Zoolaster assumes certain types of Zoola watches sell well
- Zoolaster executives want to quantify which Zoola watches may sell well so that they just buy those kind from the wholesaler

# Sales Data

Product Buy?

(  , 1 )

**Stauer Compendium Hybrid Watch**
$29.00
+ No Shipping Info
**STAUER**

(  , 1 )

**Michael Kors Watches Bradshaw Gold - Michael**
$250.00
+ Free Shipping
**eBags.**

(  , 0 )

**Men's Pro Diver Chronograph Black**
$199.99
+ Free Shipping
**World of Watches**

1 = Bought
0 = Didn't Buy

# Sales Prediction Model

Product Buy?

(  , 1 )

**Stauer Compendium Hybrid Watch**
$29.00
+ No Shipping Info
**STAUER**

(  , 1 )

**Michael Kors Watches Bradshaw Gold - Michael**
$250.00
+ Free Shipping
**eBags.**

(  , 0 )

**Men's Pro Diver Chronograph Black**
$199.99
+ Free Shipping
**World of Watches**

Customer Will Buy?

Zoolaster can potentially buy more watches from wholesaler that have higher potential of selling online

# Machine Learning and Business Model Testing

- We will look at four main ways to analyze data that are all useful in decision making for businesses in general
- Data to Scores
- Data to Classes
  - Discriminative Methods
  - Generative Methods
- Data to Clusters

# Data to Classification

- Given a set of features

$$X=(x_1, x_2, x_3, \ldots, x_n)$$

- we want to predict Y

$$Y=\{0,1\}$$

How about x?

How do we get them?

# Data to Classification

- Given a set of features

$$X=(x_1, x_2, x_3, \ldots, x_n)$$

- we want to predict Y

How about x?

How do we get them?

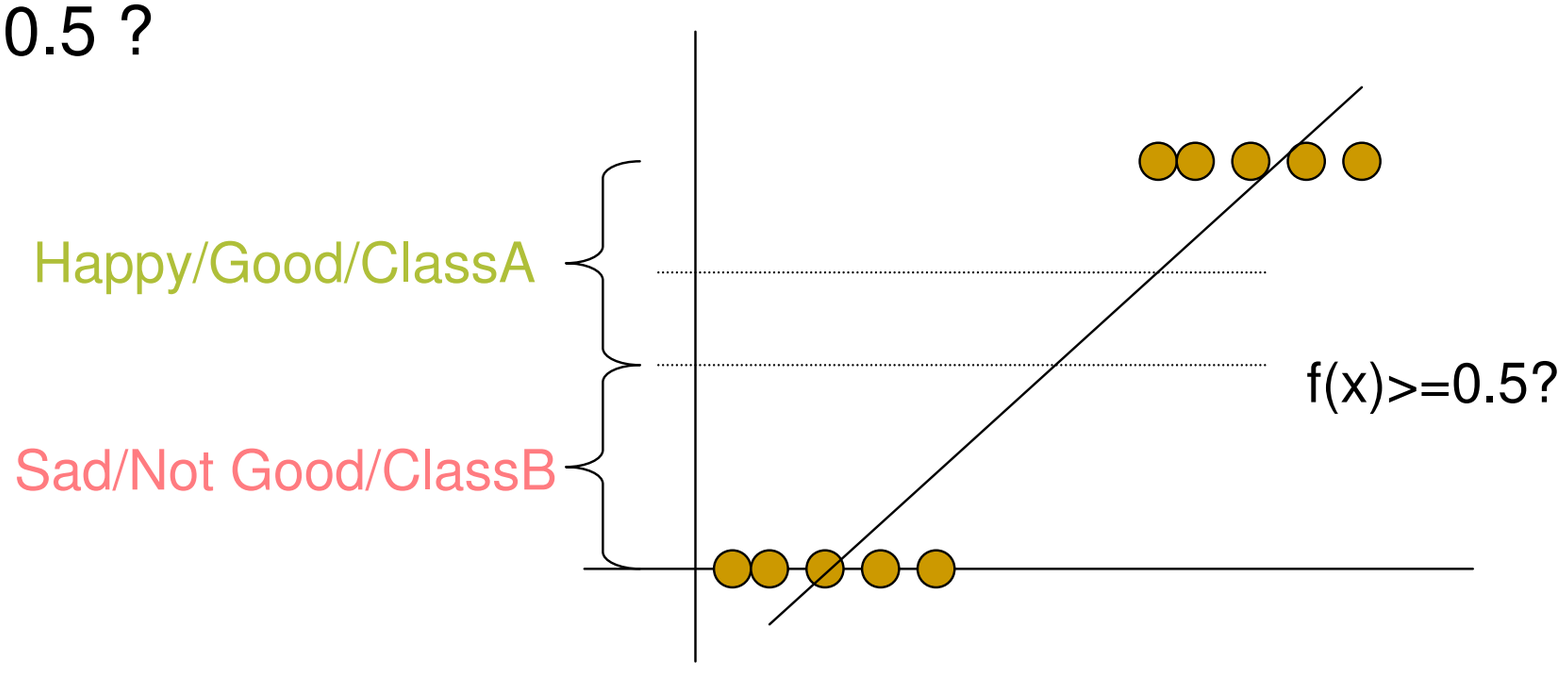
$$Y=\{0,1\}$$

( 
**Stauer Compendium Hybrid Watch**
$29.00
+ No Shipping Info
**STAUER**

, { will not buy OR will buy } )

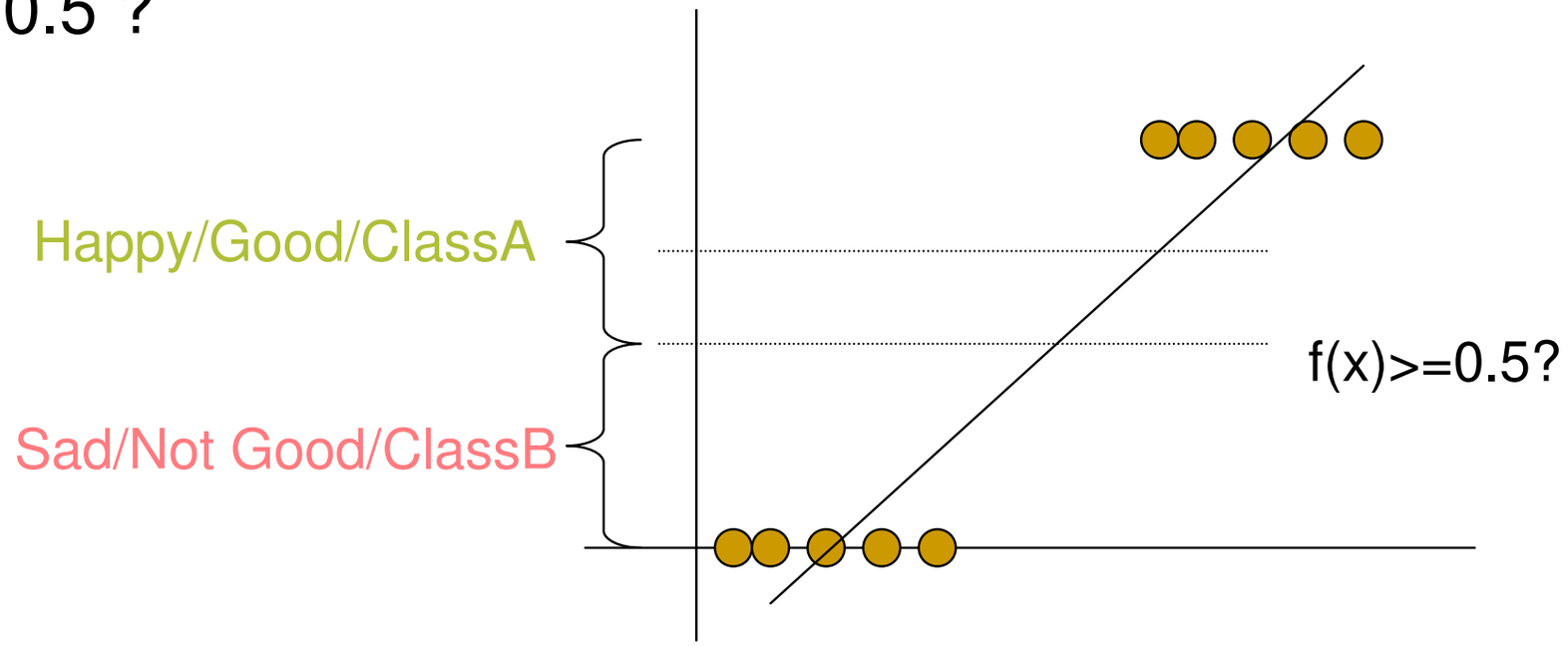
# Data to Classification

- Can we build a regression model to model such binary classes?
- Train Regression and threshold the output
  - If f(x) >= 0.7 CLASS1
  - If f(x) < 0.7 CLASS2
  - f(x) >= 0.5 ?



# Data to Classification

- Can we build a regression model to model such binary classes?
- Train Regression and threshold the output
  - If f(x) >= 0.7 CLASS1
  - If f(x) < 0.7 CLASS2
  - f(x) >= 0.5 ?



# Generative vs. Discriminative Classifier

# Generative vs. Discriminative Classifier

- Generative Classifier

# Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability p(x,y) where x are inputs and y are labels

# Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability p(x,y) where x are inputs and y are labels
  - Make prediction using Bayes rule to compute p(y|x)

# Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability p(x,y) where x are inputs and y are labels
  - Make prediction using Bayes rule to compute p(y|x)
- Discriminative Classifier

# Generative vs. Discriminative Classifier

- Generative Classifier
  - Model joint probability p(x,y) where x are inputs and y are labels
  - Make prediction using Bayes rule to compute p(y|x)
- Discriminative Classifier
  - Try to predict output directly

# Generative vs. Discriminative Classifier

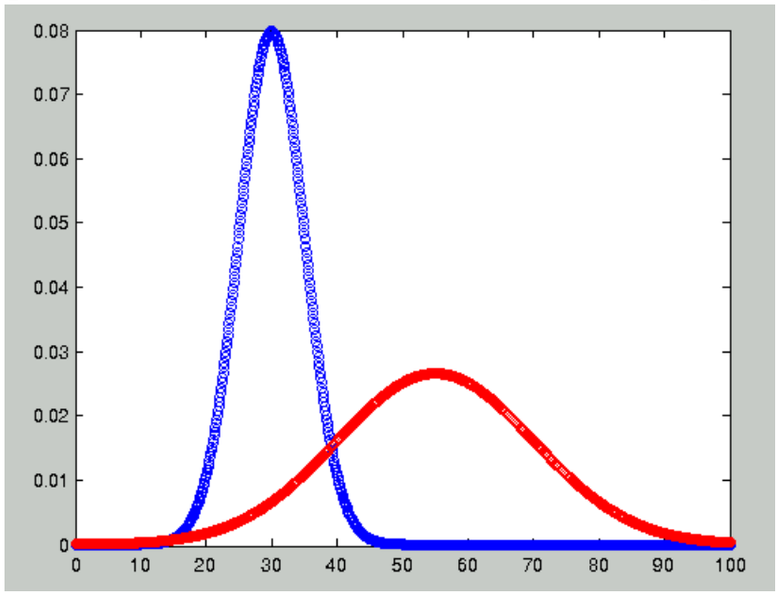
- Generative Classifier
  - Model joint probability p(x,y) where x are inputs and y are labels
  - Make prediction using Bayes rule to compute p(y|x)
- Discriminative Classifier
  - Try to predict output directly
  - Model p(y|x) directly

# Regression to Classification

- Thresholding on regression function does not always work
- Gaussian assumption on noise
- When the output is binary class, we may want to dry a different technique of modeling than regression
- Many modeling techniques that will better produce class category values we want for Y
- Linear Classifiers is one such method

# Generative Classifier

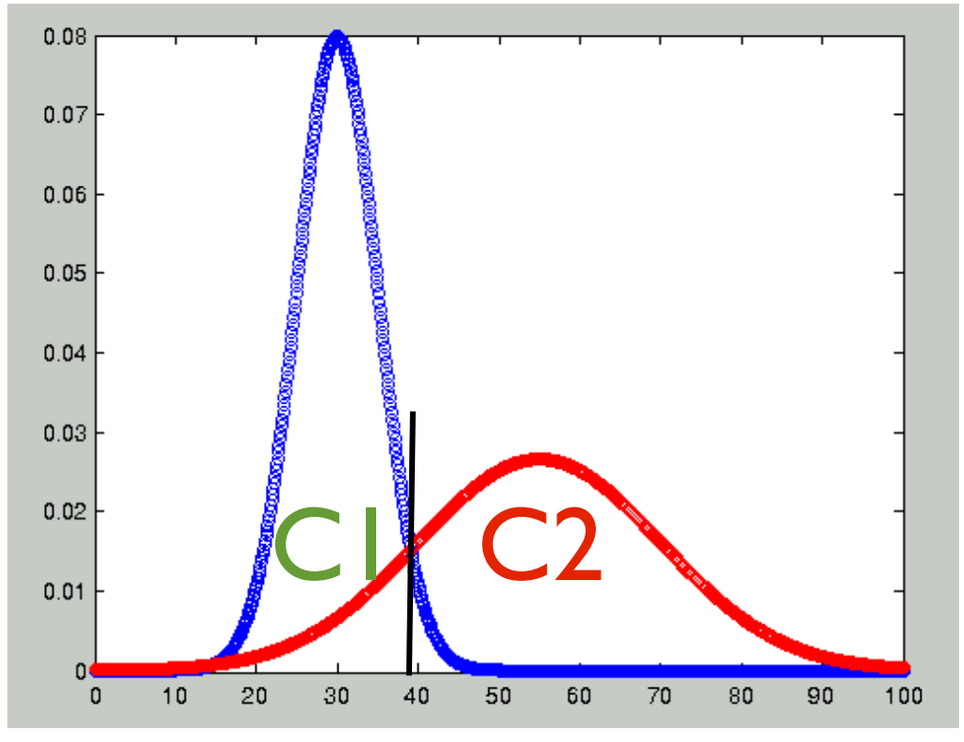
- We can model class conditional densities using Gaussian distributions
- If we know class conditional densities
  - p(x| y=C1)
  - p(x|y=C2)
- We can find a decision to classify the unseen example



# Bayes Rule

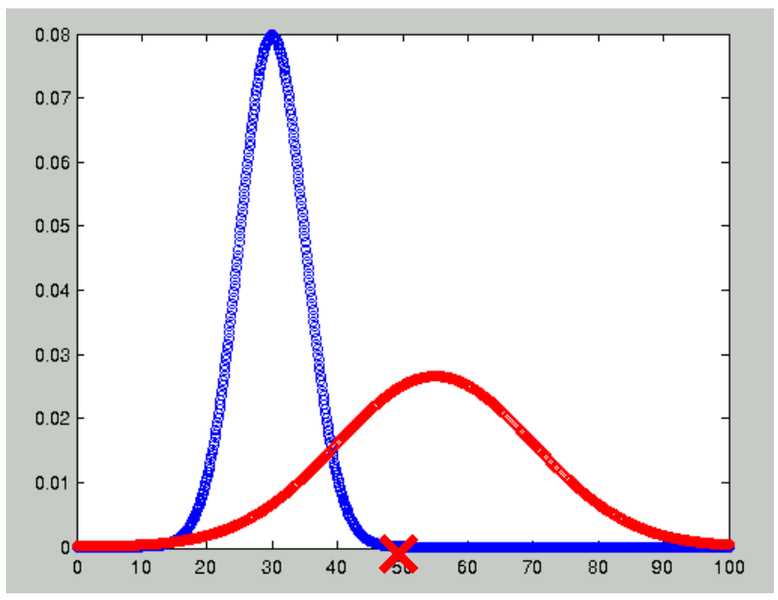
$$P(Y|X) = \frac{P(X|Y)\ P(Y)}{P(X)}$$

C1 = Buys
C2 = Doesn't Buy



# Generative Classifier

- Given a new data point find out posterior probability from each class and take a log ratio
- If higher posterior probability for C1, it means new x better explained by the Gaussian distribution of C1




$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

$$p(y = 1|x) \propto p(x|\mu_1, \sum_1)p(y = 1)$$

# Naive Bayes Classifier

- Naïve Bayes Classifier a type of Generative classifier
- Compute class-conditional distribution but with conditional independence assumption
- Shown to be very useful for many classification tasks

# Naive Bayes Classifier

- Conditional Independence Assumption

$$P(X_1, X_2, ..., X_N|Y) = \Pi_{i=1}^{N} P(X_i|Y)$$

# Naive Bayes Classifier

$$P(Y_k, X_1, X_2, ..., X_N) = P(Y_k)\Pi_i P(X_i|Y_k)$$

Prior Probability of the Class

Conditional Probability of feature given the Class

# Naive Bayes Classifier

$$P(Y = y_k | X_1, X_2, ..., X_N) = \frac{P(Y{=}y_k) P(X_1, X_2, .., X_N | Y{=}y_k)}{\sum_j P(Y{=}y_j) P(X_1, X_2, .., X_N | Y{=}y_j)}$$

$$= \frac{P(Y{=}y_k) \Pi_i P(X_i | Y{=}y_k)}{\sum_j P(Y{=}y_j) \Pi_i P(X_i | Y{=}y_j)}$$

$$Y \leftarrow argmax_{y_k} P(Y = y_k) \Pi_i P(X_i | Y = y_k)$$

# Assignment 1

- Form a team
- Name the team
- Write a short summary about your business concept (5-10 sentences)
  - Problem addressed, your customers, your proposed idea/ solution, value proposition, team
- Write 5 bullets on possible data sets you can collect to test your value proposition
- Due coming Friday @ 6pm

# Assignment II

- Fill up Lean Canvas
  - https://drive.google.com/previewtemplate?id=16uOd158UzJM9oqGWgJOtbppzGNPmZ4fWMSV6_xBz3Z8&mode=public#
- Field Assignment
  - Prepare a set of 8+ (minimum of 8) questions that you want your potential customers to answer that can help in validating your assumptions/hypotheses that you have created while filling up Lean Canvas. Questions can have qualitative or quantitative answers.
    - Example questions with quantifiable answers
      - e.g. Will you buy our product? Yes No
      - e.g. Do you think it will improve your customer service? 1 2 3 4 5
      - e.g. How important is the problem that our product solves for you? 1 2 3 4 5 6 7 8 9 10
    - Example questions with qualitative answers
      - Please tell us more about your problem?
      - What do you like about our product?
  - Interview minimum of 5 potential customers
    - minimum of 3 potential customers have to be strangers (not on campus)
    - Ask the questions you have prepared
    - Interview the potential customers with any other questions that you think are relevant
    - Collect data based on questions
    - Write a few sentences about your experience with each customer
- After you are done with all interviews, write one line elevator pitch on your value proposition by filling this Mad Lib Template of Adeo Ressi
  - Our company, __(insert name of company)__, is developing __(a defined offering)__ to help __(a defined audience)__ __(solve a problem)__ with __(secret sauce)__.
- Due Feb 18th Monday @ 6pm