# Scoring - Marketing applications

Arthur Mello (amello@avisia.fr)
Maud Laval (mlaval@avisia.fr)

# Scoring in marketing

# Scoring in marketing

1. Marketing
2. Statistics applications in marketing
3. Scoring applications in marketing
4. How to set up a data project

# Marketing

- Definition:

> *“Marketing is the set of tools made available to a company to increase its revenue through the sale of products/services to its customers.”*

- It includes all the steps involved in the sale of a product/service: development, pricing, communication, distribution of the product.

# Marketing

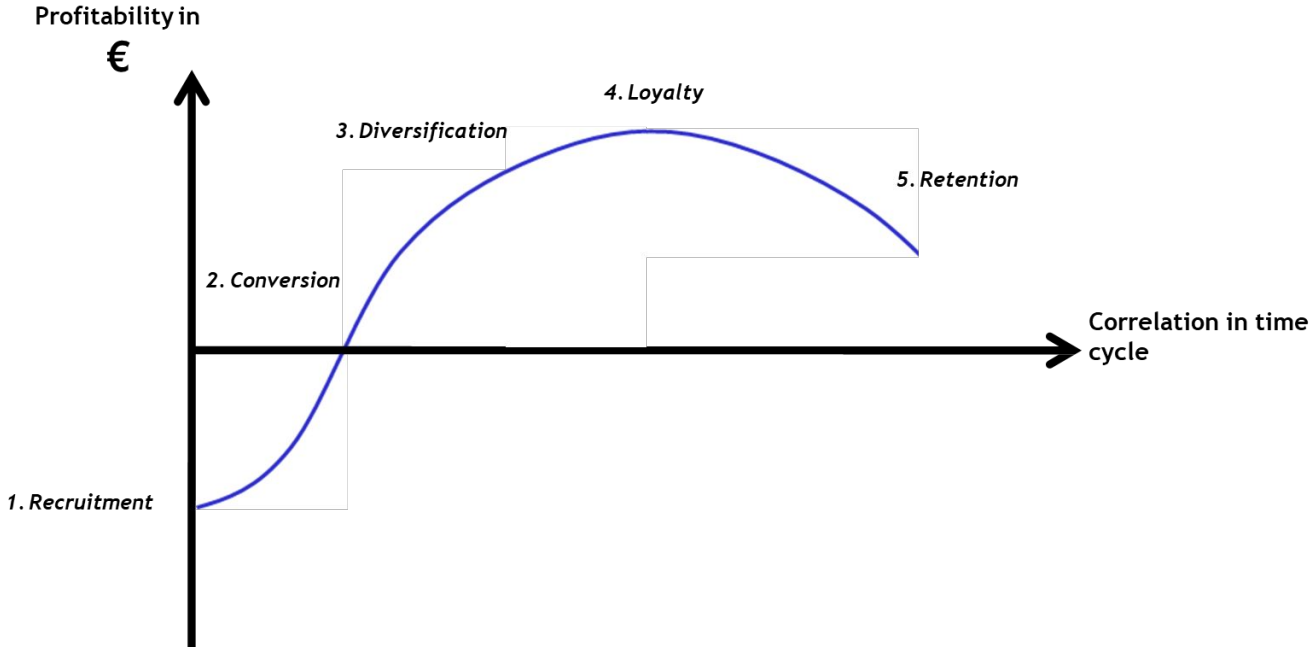
- The four P’s of marketing:
  - Product: characteristics, specificities, quality…
  - Place: which distribution network
  - Price: tariff positioning compared to other products
  - Promotion: advertising, communication, discounts…

# Marketing

- Customer-centric marketing:
  - “Know your customers well to better serve them”
- CRM (Customer Relationship Management)
  - All marketing measures/operations aimed at optimizing the quality of customer relations, building loyalty and maximizing revenue or margin per customer
- CRM database
  - All information related to customers (personal info., interactions, transaction history, source of the leads, website visits…)

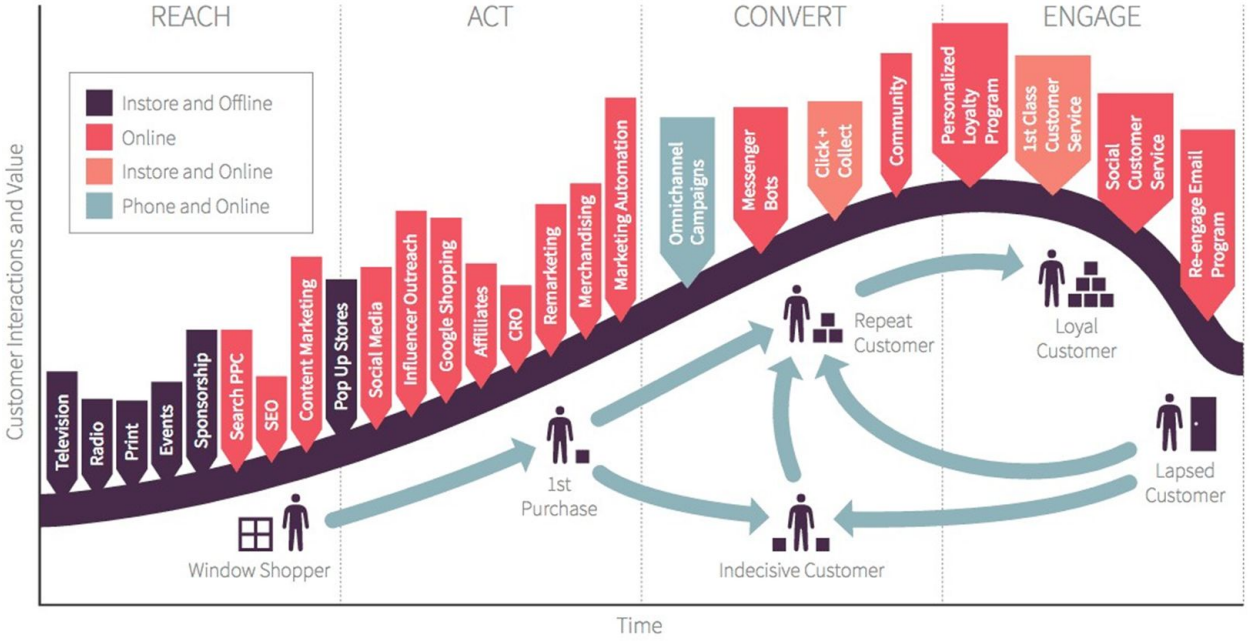
# Marketing

## Customer life cycle



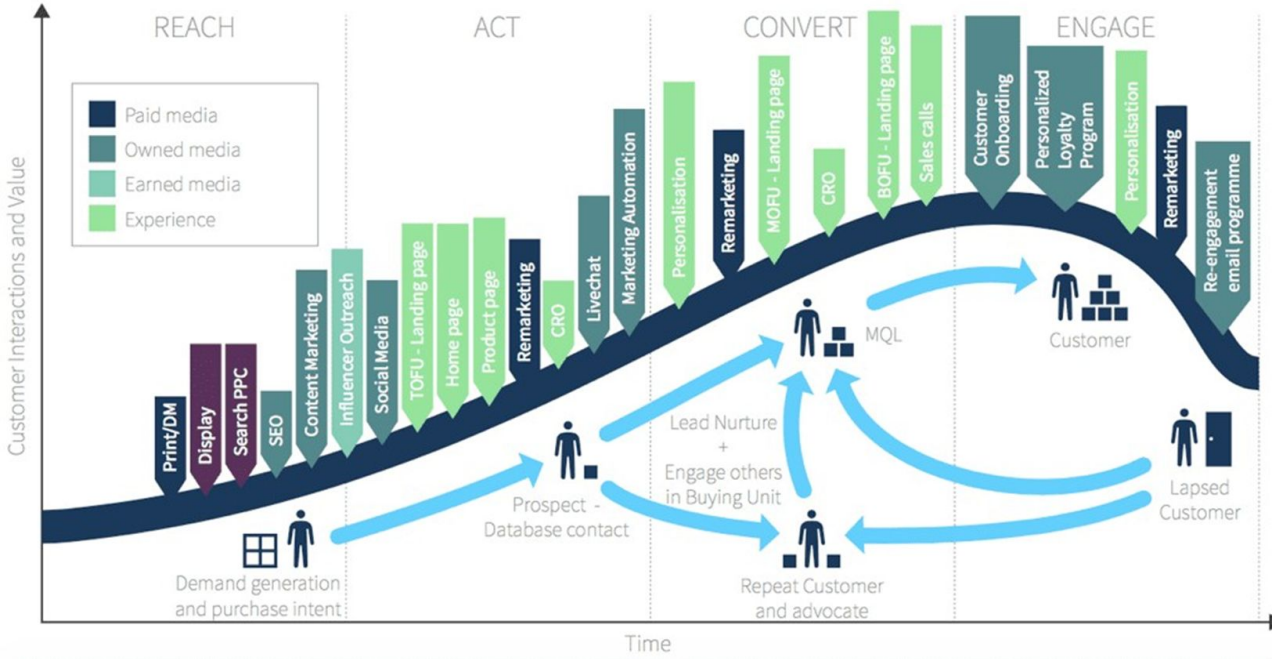
# Marketing

Customer life cycle - B2C (Business to Customer)



# Marketing

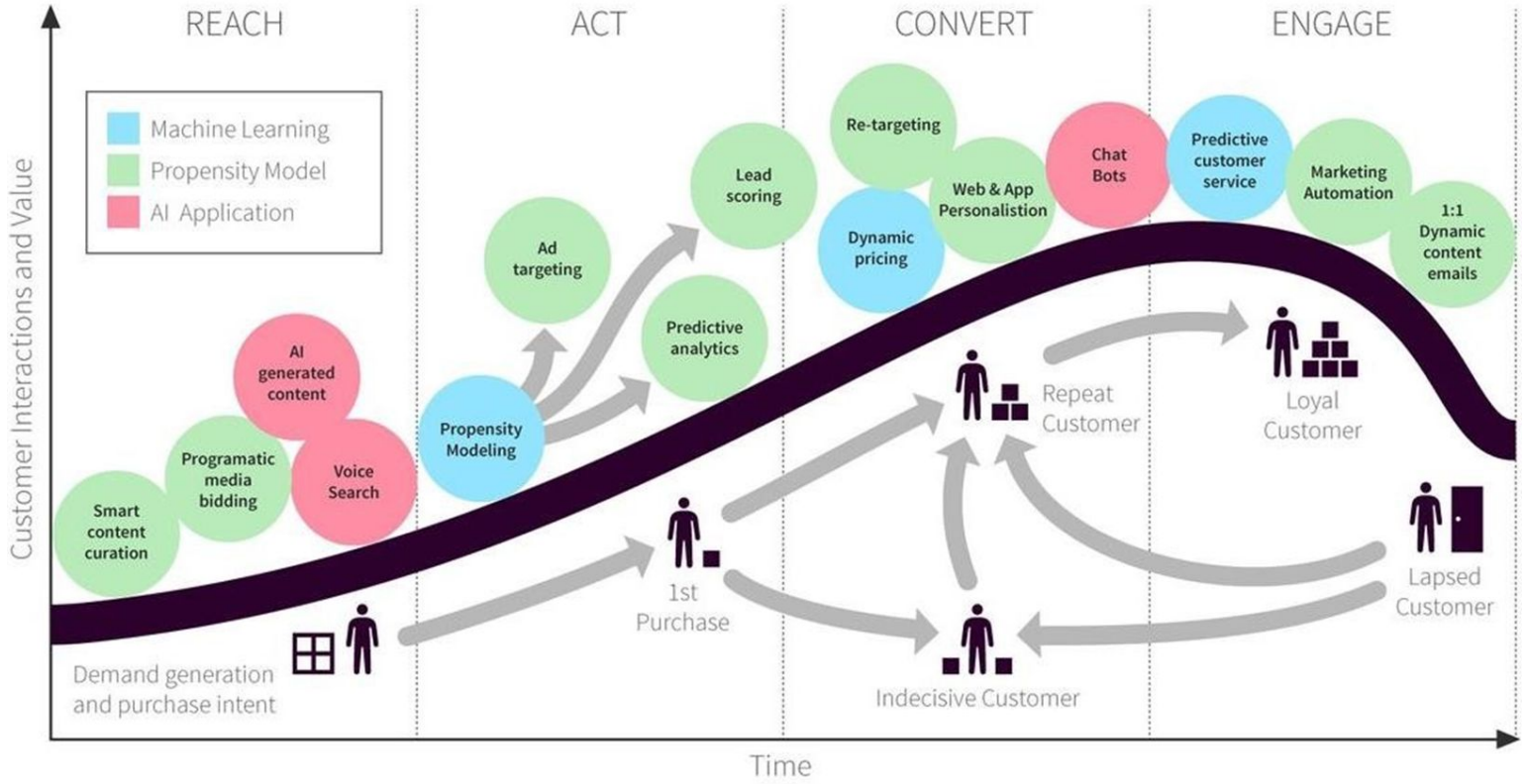
Customer life cycle - B2B (Business to Business)



# Statistics applications in marketing

- Demand analysis and forecasting
  - Predict the demand across time (seasonality, price, competitors, economical context…)
- Pricing optimization
  - Measure the price elasticity to fix the right price
- Optimization of the product distribution
  - Analysis of the performance of different types of POS, (supply chain + operational research)
- Customer knowledge and personal communication
  - Customer segmentation and customers' behaviors
  - Customer lifetime value or potential
  - Probability to churn
  - Satisfaction vs dissatisfaction
  - Propensity to purchase
  - Preferred communication channel
  - Communication frequency
  - Impact of marketing campaigns
  - Product recommendation

# Statistics applications in marketing



# Statistics applications in marketing

- Credit
  - There are twice as many loans for red cars worth more than €15,000 in the Paris region than the national average
  - People with @hotmail.fr email address are riskier than people with @orange.fr email address on average, mainly because it is a signal for the age of the customer
- Retail
  - Customers who received a significant discount on their first purchase and whose purchases show little diversification in the first half of the year are five times more likely not to renew their loyalty card
  - Luxury: Chinese customers tend to purchase higher amount but with a lower frequency than US customers

# Scoring applications in marketing

- Scoring aims to order observations according to their probability to be in the target.
- Often about predicting a future behavior -> need to think about train / test periods (seasonality, data history depth, target period calculation…)
- Main use cases in marketing:
  - Churn
  - Conversion probability
  - Risk

# Scoring applications in marketing

- Churn
- Lead conversion probability
- Customer conversion probability
- Risk

# Churn

- Goal
  - Identify customers that are likely to leave and target them in campaigns
- Widely used in subscription services (phone, internet, streaming, etc.)
- Methodology
  - Define a time frame (“will the customer churn within 6 months?”)
  - Use samples of churned and non-churned customers, given the time frame
  - Include demographic and transactional data
  - Run classification algorithms on the data
  - Use the score as threshold for campaigns

# Lead conversion probability

- Goal
  - Identify leads that are likely to become customers to prioritize them
- Often used with CLTV (customer lifetime value) predictions
- Methodology
  - Define a time frame (“will the lead become a customer within 6 months?”)
  - Use samples of converted and unconverted leads, given the time frame
  - Include demographic and contact data
  - Run classification algorithms on the data
  - Use the score to prioritize sales efforts

# Customer conversion probability

- Goal
  - Identify customers that are likely to buy a specific product/service
- Methodology
  - Define a time frame (“will the customer use the service within 6 months?”)
  - Use samples of converted and unconverted customers, given the time frame
  - Include demographic and transactional data
  - Run classification algorithms on the data
  - Use the score to prioritize marketing campaigns

# Risk

- Goal
  - Identify customers that are likely to default
- Methodology
  - Use samples of defaulted and non-defaulted customers, given the time frame
  - Include demographic and transactional data
  - Run classification algorithms on the data
  - Use the score to screen customers and/or to apply optimal interest rates

# How to set up a data project

1. Understanding business concern
2. Data understanding
3. Data preparation
4. Establishment of a model
5. Assessment – Test
6. Deployment

# How to set up a data project

1. Understanding business concern
    a. Context description
    b. Defining objectives and criteria of success
    c. Inventory of available resources
2. Data understanding
    a. Collecting data
    b. Description / Exploration
    c. Quality control
3. Data preparation
    a. Selection of necessary data for the study
    b. Construction
    c. Consolidation
4. Establishment of a model
    a. Choosing the technique(s) to solve the problem
    b. Choosing the best model
5. Assessment – Test
    a. Applying the model on the test sample
    b. Checking the classification
6. Deployment
    a. Definition of deployment rules
    b. Analysis of the results / recommendations for actions

# Gathering data

# Gathering data

- Data sources in companies
  - Databases (SQL)
  - .xlsx/.csv files
- Data sources for personal/academic projects
  - Kaggle
  - UCI Machine Learning Repository
  - US Government
  - French Government
- Data to enrich original data (public or paid APIs)
  - Weather
  - Economic
  - Demographic

# Gathering data

[Food delivery customers dataset]

| Feature | Description |
|---|---|
| AcceptedCmp1 | 1 if costumer accepted the offer in the 1st campaign, 0 otherwise |
| AcceptedCmp2 | 1 if costumer accepted the offer in the 2nd campaign, 0 otherwise |
| AcceptedCmp3 | 1 if costumer accepted the offer in the 3rd campaign, 0 otherwise |
| AcceptedCmp4 | 1 if costumer accepted the offer in the 4th campaign, 0 otherwise |
| AcceptedCmp5 | 1 if costumer accepted the offer in the 5th campaign, 0 otherwise |
| Response (target) | 1 if costumer accepted the offer in the last campaign, 0 otherwise |
| Complain | 1 if costumer complained in the last 2 years |
| DtCustomer | date of customer's enrollment 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 in the last 2 years |
| MntSweetProducts | amount spent on sweet products in the last 2 years |
| MntWines | amount spent on wines 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 |

# Gathering data

## Bank marketing dataset

Bank client data:

- Age (numeric)
- Job : type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
- Marital : marital status (categorical: 'divorced', 'married', 'single', 'unknown' ; note: 'divorced' means divorced or widowed)
- Education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
- Default: has credit in default? (categorical: 'no', 'yes', 'unknown')
- Housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
- Loan: has personal loan? (categorical: 'no', 'yes', 'unknown')

Related with the last contact of the current campaign:

- Contact: contact communication type (categorical: 'cellular','telephone')
- Month: last contact month of year (categorical: 'jan', 'feb', 'mar', …, 'nov', 'dec')
- Day*of*week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
- Duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Other attributes:

- Campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- Pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- Previous: number of contacts performed before this campaign and for this client (numeric)
- Poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

# Gathering data

## Telco churn dataset

### Context

"Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs." [IBM Sample Data Sets]

### Content

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

# Gathering data

## Lead scoring dataset

Variables Description

- Prospect ID - A unique ID with which the customer is identified.
- Lead Number - A lead number assigned to each lead procured.
- Lead Origin - The origin identifier with which the customer was identified to be a lead. Includes API, Landing Page Submission, etc.
- Lead Source - The source of the lead. Includes Google, Organic Search, Olark Chat, etc.
- Do Not Email -An indicator variable selected by the customer wherein they select whether of not they want to be emailed about the course or not.
- Do Not Call - An indicator variable selected by the customer wherein they select whether of not they want to be called about the course or not.
- Converted - The target variable. Indicates whether a lead has been successfully converted or not.
- TotalVisits - The total number of visits made by the customer on the website.
- Total Time Spent on Website - The total time spent by the customer on the website.
- Page Views Per Visit - Average number of pages on the website viewed during the visits.
- Last Activity - Last activity performed by the customer. Includes Email Opened, Olark Chat Conversation, etc.
- Country - The country of the customer.
- Specialization - The industry domain in which the customer worked before. Includes the level 'Select Specialization' which means the customer had not selected this option while filling the form.
- How did you hear about X Education - The source from which the customer heard about X Education.
- What is your current occupation - Indicates whether the customer is a student, umemployed or employed.
- What matters most to you in choosing this course An option selected by the customer - indicating what is their main motto behind doing this course.
- Search - Indicating whether the customer had seen the ad in any of the listed items.
- Magazine

# Project brief

# How to set up a data project

1. Find a data source and define the problem
2. Validate step 1 with us
3. Analyse the data
4. Transform the variables
5. Apply scoring models
6. Iterate
7. Create visualizations
8. Present your project