# 1 Marketing

### - Evolution of marketing concept
The evolution of marketing concepts is in 4 phases:

1. Production orientation: Demand is more then supply, there is a lack of goods. Strategy consist in rapid and efficient production.

2. Product orientation: Smaller gap between supply and demand, so companies create products or services that are better or different then competitors.

3. Sales Orientation: focus on sales, uninterested in costumers needs. Companies combine study of production and sales with the segmentation of market and target identification.

4. Marketing orientation: saturated demand, so companies create tailor-made solutions for consumers with aim of fainting trust and loyalty over time. Analysis of needs of costumers is fundamental, with control of production and distribution costs..

The marketing move form a product centricity, that had involved only supplier and reseller, to a customer centricity, that place accent on costumer needs, and aim in establish solid relationship capable of adding value in medium long term.

### - Corporate marketing: areas and strategic marketing
Corporate marketing is made by three interconnected areas:

- Strategic marketing: medium-long term corporate objectives. Involve systematic and continuous analysis of market needs.

- Operative marketing: short-term actions and plans to achieve the set objectives.

- Data-driven marketing: techniques and tools fore collection and analysis of relevant information to support in defining the strategy.

Strategic marketing means an analysis of the market and competitive context. Types of strategic analysis involve:

- Framework swot: identification of strength, weakness, opportunity and threat.

- 5 forces of porter: to understand company competitive position look for new entries, substitute product, power of consumers, power of suppliers and competitive rivalry.

- Customer study: study of current and potential customers aimed in identify reference target and segmenting in sub targets.

- Placement: position of the brand in relation of the needs, value and expectations of the customer

- KPI: Key Performance Indicators, business indicators to be monitored over time to implement corrective measures on a strategy.

### - Technology in evolution of the marketing concept
Evolution of marketing also brings with the evolution of technology.

1. Marketing 1.0: focus on production and sales, use one-way communication channels (form company to consumer), undifferentiated massage to large audience. Use traditional medias

2. Marketing 2.0: interest in knowing the final consumer, use two-way with one-to-one logic. Use digital tools, telephone survey and market analysis.

3. Marketing 3.0: focus also on ability of make a better world. Focus on three values ESG (environmental, social, governance).

4. Marketing 4.0: omnichannel marketing, use cloud computing and big data. Customer journey for 4A (aware, attitude, act, act) to 5A (aware, appeal, ask, act, advocate)

5. Marketing 5.0: merge physical and digital reality. Use technology for humanity to satisfy material needs of consumers, promoting immaterial ones. Use data-driven marketing, predictive marketing, contextual marketing and augmented marketing.

### - Technology as a marketing enabler for data analysis

Main technologies for management, acquisition and exploration of data have been introduced starting from simple systems MIS and DSS up to RDBMS. Growth of the amount of data lead development of new systems for standardize data coming in a fragmented and decentralized way form different vertical and centralized DB. An example are DWHs, that offer a single point of access on as single application.

MapReduce data processing framework, HDFS file systems and YARN resource manager were introduced to handle presence of unstructured data.

Data Lake (structured, semistructured and unstructured data) and DWHs (only structured data) are used for visualization and BI activities. Data Lakehouse provide benefits of a DWH using a low-cos storage that support data analysis and application of ML models, keeping performances and reliability of DWHm with use of visualization and BI tools in a single open-source environment.

### - Data, advanced analytics and AI landscape

The Data, advanced analytics and AI landscape include:

- Data sources and API: use external data sources and allow the democratization of data.

- Data resources: data services like kaggle allow public dataset, incubators&school like DataCamp provide online platforms with free resources and research like Google or Open AI allow for article publishing community.

- Open sources: the choice depend on amount and type of data.

- Infrastructure: can be on-cloud, on-promises or hybrid.

- Analytics: include a data catalog and discover tools to search in datasets, use a query engine and data analytics to building pipelines, and BI and data visualization to analyse the data.

- ML and AI

- Application - Enterprise: are useful form anging specific business area

- Application - Industry

### - Cloud architectures

Cloud computing is the possibility of offering computational services through the visualization of hardware resources, offer traversal service, transforming it into a utility. The four main providers (all data lakehouse) are Google Coloud Platform (services, processing, storage, ingestion), Amazon Web Services (services, processing, storage, ingestion), Microsoft Azure (services, processing, goverance, storage, ingestion), Databricks (services, processing, governace, storage, ingestion).

### - Microsecices in cloud computing

Micro services are a programming paradigm where application is obtained as a composition of smaller independent services that communicate through a message system (API or HTTP). The architecture is built as a suit of services, ensuring dynamism, flexibility and scalability. The strengths is free choice of development tool chains, which can be developed using different framework and languages.

The problem of putting a single service into production and isolating it is obviated by the use of virtualization techniques:

- VM: computing resources that uses a software layer to run applications. Share the same physical server (host), Providing guest systems with a virtual environment.

- Container: collection of one or more process that maintain a separate execution context form the rest of the system, and share the same host operating system.

- Serverless calculation: supports runtimes such as Phyton, Java. The application can be executed on-demand in response to events that trigger its execution.

- Kubernets: open-source container orchestration platform. It automate software development and management. Use a master-slave architecture by controlling compute and storage resources as objects while its components are managed through an API. Control Plane is the master.

- Apache Mesos: master-slave architecture that isolate resources in cluster management, consist of frameworks that deal with task execution and workload distribution. It takes care of the orchestration of the Mesos and Docker containers.

- Apache service mix: opens ouce project for implementing a distributed ESB.

- Openshift: softwere containerization products, PaaS platform and includes Docker images together with the Kubernetes platform and all its features.

- Kaas: Cloud solutions that manage the orchestration through Kubernetes and take care of managing development, maintenance, hosting, updating, monitoring, automation support, load balancing and scalability.

### - Python

Python is an open source high-level object-oriented programming language. Provides a program that can run other programs directly from their source code, without prior compilation. Python is a general purpose language, which can be used as a scripting language, intended as a tool for coding system functionality as an alternative to shell scripts. The Python language is highly portable, in the sense that most programs can be run on different platforms without making any changes to the code.

The Python interpreter is accessed via the interactive prompt or command line. The Python prompt is a REPL (read-evaluate-print-loop). Ipython provides a REPL that backends Jupyter notebooks accessible via web browsers. Python is a case sensitive language. Python is not a typed language.

Virtual environments are a tool that allows you to create spaces independent from the rest of the system where you can install different versions of Python and its modules.

### - Data-driven marketing approach

Data-driven marketing approach is strategic corporate decision where companies depend their knowledge of their customer base through the analysis, study and synthesis of data collected and processed, this generate insight to support the strategic decision with the ultimate goal of defining a personalized proposal for different customer segments, to maximize the profitability of relationship along the entire customer life cycle. Data-driven approach improves business performance by 20%.

The main pillars are:

- Analytics: analysis of data and grouping into segments

- Advanced analytics: automatic mode of analysing data to discover relationship and correlation, to develop forecast and recommendations

- AI: useing computers to understan human intelligence, in two sub fields: ML and DL.

### - Model development stages

The process of model develpment involve four players: data scientist, business owner, data engineer and data architect.

The process has 4 phases:

- Definition of business objective: identification of analysis objective, tools, type of output and business objectives.

- Data understanding and data collection: mapping of internal and external data sources to understand data and needs they satisfy.

- Model development: definition of model components, account for complexity and specificity of each industry.

- Model deployment: summary documentation with main insight and technical details, that describe decision made during development of the model. Code testing using unit-test and UAT, and preparation of logging and monitoring systems.

The four pillar for development of advanced analytics or AI algorithms are:

- Maturity (entry level - data focused - data driven)

- Analytics type (advanced analytics - ML -DL)

- Focus (customer - product - selling point - marketing optimization)

- Lifecycle (attract - engage - retain)

Main customer models, regardless of the company's maturity level towards a data-driven approach and the type of Analytics, useful for generating insights along the customer lifecycle are:

- Repurchase model (attract)

- CLTV (engage)

- Churn model (retain)

- Cluster analysis (engage - retain)

- Customer satisfaction (attract - engage - retain)

- Neuromarketing (attract - engage - retain)

- RFM (attract - engage - retain)

Where attract means to acquire new customers, engage means to get customer engaged, retain means to keep customers.

### - Deterministic segmentation - RFM

RFM alorithm is based on three components:

- Recency: length of time since last purchase

- Frequency: number of purchases in a period

- Monetary: amount spent

Dividing each variable into categories, is given the probability of response to the commercial offer.
RFM's methodological approach is

1. definition of reference period and analysis time interval,

2. verify correctness and distribution of the variables to define the number of optimal categories, to divide in the three dimensions,

3. Cartesian product of categories to identify RFM segments, ordered triplet (R,F,M),

4. verification of segments to exclude of a reduction of number of categories or unification of segments,

5. summary indicator (score).

The benefits are:

- a priori choice of dimensions,

- easy interpretation of segments,

- model useful for entry-level companies for evolutionary roadmap in data-driven field,

- strategic insight on customer base to increase loyalty and retention and develop up-selling and cross-selling actions,

- tailor-made marketing actions,

- intervention on customer migration to a lower segment.

The limitations are:

- limited number of variables penalizes the results,

- R, F, M are often correlated, generating an overlapping effect,

- not allow to obtain a single summary indicator.

### *- The behavioral segmentation - Cluster Analysis*

Cluster Analysis are a set of algorithms that divide the observation in subgroups named clusters, so an element is most similar withe other in the same cluster and most heterogenus to ones in other clusters.

It is used to divide the customer base into clusters to understand the characteristic and needs of each group, for tailor-made marketing activities.

Given $n$ observation the total number of subdivision in $k$ cluster of dimensionality $x$ is $\frac{1}{k!} \sum_{x=1}^{k} \left(\frac{k}{x}\right) x^n$.

Main cluster methods are:

- Partition Methods: subdivide the set of observations into a predetermined number of non-empty clusters.

- Hierarchical Methods: define a hierarchy within the cluster using a tree structure and incising similarity threshold.

- Other Methods

Cluster methods use similarity measures that depend on the type of variables. For numerical variables the distances used are: Euclidean, Manhattan and Minkowski. While for categorical variables, it depend if the variable is ordinal or nominal. Ordinal variables are reduced to numerical values and then a measure of distance is used, while nominal variables are represented as binary attributes of possible values that the variable can assume, then Jaccard distance is calculated.

PCA is used to dimensionality reduction for using a minimal subset of relevant variables, it create a linear transformation that map the variables in a smaller latent space, whit minimum loss of information.

Partition methods apply iterative procedure that reallocates the observations within the cluster in order to improve their representativeness. The best known is the k-means algorithm:

1. Initialization: random choice of initial centroids

2. each obs is assigned to cluster with centroid most similar

3. new centroid is calculated for each cluster as average of the cluster observation

The algorithm converge when there is no reallocation. The down side is that it is strongly influenced by outliers, so data cleaning is necessary.

The hierarchical methods overcome the problem of partition methods which require to specify the number of cluster a priori and it is hard to make one that meet the business needs. This method is based on distance between two cluster that must be defined in advance. Hierarchical methods can be:

- Bottom up

    1. initialization: many cluster as the observations

    2. distance between cluster merging the two most similar

    3. procedure end when all observation are in one cluster

- Top down

    1. initialization: all observation in a single cluster

    2. subdivide the cluster to maximise distance between subdivision

    3. procedure end when there is a cluster for each observation

The effective cluster subdivision is chosen cutting the dendrogram.

The evaluation of an unsupervised algorithm cannot take place on the basis of a target, but must refer to the main objective of the algorithm itself.

- Silhouette's coefficient: measures how much an observation is similar to the other in the same cluster, is between -1 and 1, where 1 means well clustered and -1 means wrong assignment.

- Elbow method: number of clusters on the x axis and SSE on the y axis, to show how the quality of the segmentation improves as the number of clusters increases, the elbow represent trade-off between increase in complexity and benefit obtained.

- Pseudo-F statistic: compare the ratio between inter-cluster variance and intra-cluster variance, a large value indicate separate cluster, peaks in graph identify the ideal number of clusters.

The benefits are:

- strategic insights on the customer base that increase loyalty, acquisition and develop up/cross-selling actions

- insights to develop tailor-made marketing actions

- insight to marketing automation

- find best acquisition and communication channel for each group

- compare distribution of segments from geographical point of view

The limitations are:

- difficult to find optimal set of variables

- the need of small number of variables and standardized, make difficult to interpret clusters

- the use of a specific offer to the customers of each group without applying further filters afterwards, in the selection of the target

### - Churn and repurchase models
Churn and repurchase model study respectively the abandonment and the propensity to repurchase of new customers.

Churn predict which customers are most likely to abandon and allow for an intervention, it focus on repeater customers and analyzes the main characteristics of those how left. In relation to the type of abandonment (total or partial) is possible to divide the churn into scenarios (contractual/non contractual) and types (voluntary/involuntary).

The repurchase focus on one-shooter customers, with aim of estimating probability of repurchasing.

Repurchase model is at the star of customer lifecycle in the attract part while the churn is at the end in the retain part.

The main factor that influence customers in purchase and return are: reference sector of the industry, customer experience, omnichannel and tuchpoints, personalized onboarding process, natural repurchase cycles.

The methodological approach is in 4 parts:

- Target definition: in both churn and repurchase model the first step is the target definition.

  - Churn
    * evaluate the type and scenario of abandonment to deal with
    * identify outliers and exclude them
    * make exploratory analysis to verify the indicators to use
    * define the analysis interval, as: observation period (analysis time window where to calculate variables), latency period (latency window where to calulate variables, update scores, and carry on marketing actions), churn period (time window where verify the customer's behavior to define the target variable)

  - Repurchase
    * identify new costumer for the target
    * calculate inter-purchase distance
    * analyze the repurchase curve and define the time frame

- Explanatory variables: definition of the explanatory variables divided into macro categories, use correlation analysis and tests of significance

- Balancing: verification of balance of the dataset using under-sampling, over-sampling or synthetic-data generation.

- Train-test cross validation: split in train and test or use cross-validation for training the ML model to predict abandonment or repurchase

The main performance measures for churn and repurchase models are:

- AUC: sensitivity and 1-specificity curve, to visually compare models

- Accuracy: number of correct predictions over all the predictions

- percentage of customer correctly recognised as positive over all the costumes

It is also important to monitor the KPIs related to the customer group with a high probability of abandonment. The insights found and the characteristics of the groups allow for the construction of a tailor-made promotional activity plan.

The benefits are:

- knowing the reasons that drive a customer to abandonment and implement data-driven strategies

- insights generated can be integrated to customer care activities

- model score used to optimize marketing contact plan

The limitations are:

- churn: difficulty of changing the outcome of the model through marketing actions, the model is developed on churn not on ones willing to leave

- repurchase: scarcity of available data, only the first act of purchase is considered.

The evolution are:

- churn: the "uplift" models divide the customer base into four categories along two dimensions, conditioned by the marketing actions carried out

- repurchase: enrichment through the use of third-party data

### - CLTV model

CLTV (customer lifetime value) model aims to predict the economic value of a customer throughout the life cycle (form acquisition to abandonment), maximizing the relationship over time.

Main applications are: evaluation of acquisition cost (segmenting by acquisition channels), service models according to current and prospective value of customers, and identification of segments with highest future value.

This strategy result in benefit of representing the prospective value of each customer. The limit is the difficulty in finding historical data to allow correct application in long term, and difficulty in obtaining immediate indications on the marketing actions to implement.

The methodological approach can be:

- Historical approach: algorithm based on variables like

  - CLTV amount, frequency, duration of relationship
  - CLTV amount, frequency, duration of relationship and profit
  - CLTV weighted by interest rate and churn

- Predictive approach - ML: ML models

  - define time horizion of customers
  - create of model features (socio-demographic and behavioral variables)
  - Data quality, exploratory analysis and statistical tests
  - model testing
  - model evaluation

- Predictive approach - economic model: econometric models, the distribution curves of some parameters are analyzed for probability distribution for estimating customer CLTV, most used are BTYD models (based on purchase frequency estimation model and a churn estimation model)

The evaluation of CLTV model consist in backtesting (applying the model to customer tables of historical data) and calculate prediction errors like MSE, RMSE, MAE, MAPE.

### - Measurement of customer satisfaction

Customer satisfaction is the customer's opinion that emerges form the difference (negative/positive) between perception of the services used or product purchased and its expectations. It is subjective and depend on multiple factor, usually a trade-off between the perceived benefits and the cost. To have better economic results it is essential to satisfy customers.

Customer satisfaction is measured using surveys, dimensions are: tangible aspects, reliability, ability to respond, ability to reassure, empathy.

The virtuous circle of CS is that a satisfied costumer is more inclined to repurchase, is more likely to recommend the brand, so the brand get more visibility and customers are more satisfied.

The main brand recommendation indicator is BRS (brand recommendation score) and it measure the derives from a single question asked to the customer.

The benefits is the monitoring of the interactions between the company and customers, using customer feedback, allow to isolate and problem and intervene.

The limitations is that customers are more inclined to complain and remember negative experience.

### - Neuromarketing and AI

Rational investigation measure what people say and discover the rational, conscious and verbal level; while emotional investigation complete the profile measuring what people fell in the non-verbal, subconscious and emotional aspects. Companies capable of identifying emotions have and unbeatable competitive advantage.

Neuromarketing is based on:

- AIDA model: composed of four stages

    – Awareness: aware of the product

    – Interest: advantages deriving form the purchase

    – Desire: preference for the brand or product

    – Action: make the purchase

- fMRI: map of functions of brain

- ECG: cardiac activity to study emotional impact

- EEG: brain wave activity

- x-ray and neuroimaging

The four main areas of applications are: advertising area, digital area, customer experience, and brand reputation.
The benefit of neuromarketing are:

- understand subconscious behavior of customers,

- use reliable neurometric indicators through a innovative approach,

- capturing customers' instinctive responses,

- learn form clients' subconscious,

- improve marketing strategies with orientation to the innovation

- reduce time between data collection and insight generation compared to traditional survey or focus group

- find out which marketing stimuli are most effectve in attracting, engaging and motivating a customer to stay loyal to the brand.

Unlike the results achieved with traditional marketing research where the answers may contain rational bias, neuromarketing makes it possible to scientifically decipher the customer's unconscious thoughts.

Three type of neuromarketing analysis are:

- Brain imaging: monitor brain response of customers following exposure to physical and neurological stimuli. EEG is used to detect changes in level of attention while viewing different versions of an advertisement or website. fMRI is used to distinguish decisions made following the activation for analytical thinking from those activation on emotional state.

- Physiological analysis: analysis of physiological behaviours in response to a stimulus. ECG is used to record variations of emotions, combined with other techniques. GSR is used to monitor the emotional response of the testers, it measure the sweat.

- Behavior-mental analysis: study behavior of customers during purchase phase. Use eye-tracker to records pupil movements in response to a stimulus to get which elements attract attention. The IRT is a question-answer association test used for benchmark analysis between brands.

### - Methodological approaches for analyses in product key
Together with the customer needs is important to take into consideration the relevance of the product or service offered.
The main models are:

- Algorithms for defining associative rules: recommendation algorithms that aim to understand dynamics and predict purchasing by using association rules that emerge form customer's purchases. It is based on probability. Main example is MBA.

- Price sensitivity models: evaluate the elasticity of demand for product as function of a price change. Main examples are: regression model, price ladder method, Van Wastendorp model, Gabor-Granger method.

- Propensity to purchase models: ML supervised models to identify customer's probability of purchases.

### - Algorithms for defining associative rules
Algorithms for defining associative rules between products or services make possible to discover hidden association between sets of jointly purchased products. It identify recurring behaviours among customers. Also identify recurring behaviors among customers.
Use probabilistic associative rules: if X is true then Y is also true, hat have a probability of occurrence.
MBA analyze the transactions made by customers to understand their purchase pattern by finding associations between the different product in their cart.
The methodological approach is:

1. Build a dataset with $m$ binary variables indicating the presence or absence of an item

2. Definition of an associative rules (X defined antecedent to the rule and Y consequent to the rule)

3. Support (percentage of transitions that contain both X and Y) and confidence (conditional probability of X given Y) of the associative rules

4. Define business conditions that must have not have too high a level of triviality and must be able to identify significant patterns

5. Search for "storng association rules", which are rules that exceed minimum support value and a minimum confidence value.

   - generation of frequents item-sets (groups of products)
   - Generation of strong rules starting form frequents item-sets

Another approach is "sequence analysis", that is identification of associations, taking into consideration also the temporal order of the purchase.
The main evaluation methods are used for the construction of the rules:

- Lift: confidence of the rule compared to the support of its consequent, if it is great then 1 then the rule is effective, if close to 1 the products are independent.

- Leverage of a rule: difference between the probability of co-occurrence of antecedent and consequent versus probability of independence, 0 indicate Independence, positive values indicates correlation.

- Conviction: product between probability of antecedent and probability of lack of consequent, compared to probability of occurrence of the antecedent and consequent.

The benefit are:

- Study of correlations that lead the purchase of goods together

- Campaigns over time to increase both cross-selling and up-selling

- A priori evaluate the introduction of a new product

- Design the best possible arrangements of the product on she shelf

- Analyze the consequence of removing obsolete products

### - Price sensitivity models

Price sensitivity is the analysis of how much changes in the price of a product affect the customers' willingness to purchase them. It is measured as the price elasticity of demand.

The factors that influence customer's price sensitivity are:

- type of product (essential goods has least sensitivity)

- reference price (price offered by other companies)

- uniqueness (forced to suffer price increase when there are no alternatives)

- ease of exchange (difficulty in evaluating the cost)

- disposable income

- customer attitude

Methodological approaches possible are:

- Mondanic proce testing: interview in which the customer is given various purchase price hypotheses for a good, indention to purchase is recorded. Allow to quickly get initial feedback.

- Conjoint analysis: customers evaluate a set of alternative products and asked to choose their favourite.

- Linear regression model: allow to understand the percentage change in quantities sold of a product with respect to price change. Identify elastic and inelastic products. Not conider all potiential factors.

- Gabor-Granger: identify for each customer maximum price they are willing to pay for a product, start form purchase response form a randomly chosen price and depending on the response increase or decease the price until obtain the highest possible. It allow to reach the oprimum point between demand curve and revenue curve. This method is suitable for products already on the market. The limitation is that it ignore substitute products.

- Van-Westandrop method: series of questions following a description of the product, and discover hypothetical price of customers. The questions are:

    – What price would be so low as to question the quality of the service?
    – At what price does the product begin to be considered a bargain?
    – At what price does the product start to seem expensive?
    – At what price is the product overpriced?

- Neuromarketing pricing: use fMRI because if a price is not appropriate lead to brain response of physical pain, while preference for a product give pleasure sensation. This method also take in consideration the context in which the purchase is made.

### - Propensity to purchase a product of a category of products

The Propensity to purchase model define which customers are most likely to purchase and so on which activate a specific marketing campaigns, that aim to increase redemption rate and cross-selling/up-selling rates.

Predictive models estimate the probability of a purchase based on the patterns.

The approach used to identify possible buyers are:

- Pilot campaign

- Product uptake

- Profiling of already owners

The meteorological approach deepen on the situation of the company. It starts from the definition of the target variable, then continue with the selection of customers divided into training and validation. It is important to dwell on how to use insight coming form the model for marketing campaigns. The definition of purchase probability classes are: up to first tertile (low probability), between first and second tertiles (average probability), beyond the second tertile (high probability). It associates the propensity score to each costumer, and based on that a campaign will be different for each class (aims at increase the probability). Usually percentage between 5 and 10 of customers with same characteristics as the target are used as stratified control group for main variables of interest.

Optimization is needed when customers are eligible for more than one product, so results form descriptive analysis and propensity models are used (Next Best Offer increases CLTV). This require the development of specific cross-selling and up-selling models, purchase probability estimates are combined with the expected Net Present Value (NPV) for each product to estimate the propensity value of each offer (VPO) The best offer for each customer maximise the estimated propensity value.

### - Optimization of marketing channels and communications
Computer vision is a the center of digital transformation. It use NN to recognise objects within images. Using CNN recognise shapes and colors and follow moving objects. It can be applied in quality control or retail form a predictive prospective.

Another approach is social analytics. It involve a digital monitoring activities to analyze the metric of all social media. It consists of three phases:

1. identification of data flows and data collect

2. descriptive analysis and AI model development

3. interpretation of results and insight (conversion rate, engagement rate, brand reputation)

Another data source come form neuromarketing research. Their main objectives are:

- determine the most attention catching graphic elements

- analyze the emotional, attention and cognitive involvement during interaction with the brand

- identify elements of a video/post that generate most emotional engagement

- compare effectiveness of two posts/videos

Another strategy is NLP, which involve automatically processing information written or spoken, to obtain concise, useful and structured information. It is applied in "sentiment analysis" which involve using NLP on different platforms to determining the tone of opinions. The applications are:

- Understand what people think

- Track mentions online

- Track competitor mentions for benchmarking

- SWOT analysis

- understand brand reputation

### - Computer vision
Computer vision consist of both classic and AI based techniques for image processing and information extraction form them. The main source of digital imaging process is electromagnetic spectrum (also ultrasonic acoustic waves and electron beams). A digital image is a matrix of values in which each pixel takes on a discrete value representing its intensity, can be two-dimensional (gray scale) or three channel (RGB). The main file format in which images are saved, that also define the compression, are jpeg, tiff and png. DICOM is a way to asave aslo metadata.

Images are represented:

- Spatial domain: discretization of a matrix in which pixels represent intensity, use convolutions.

- Frequency domain: use wavelet transformation technique, where the signal is broken down on the basis of small waves of different frequencies and time duration, can process the same image at different resolution.

CNNs make image processing at different scales conditioned by filters defined in a data-driven way. The filters that make up the network have kernel values initialized randomly and modified during training.

- Autoencoder is composed by encoder (compression) and decoder (reconstruction).

- GAN is composed by a generator and a discriminator that train together, the generator try to fool the discriminator generating images which are as real as possible, while the discriminator aim to discover which images are real.

- Transformers use attention mechanism to weigh each part of the image.

The main technological application of CV are:

- classification

- semantic segmentation (pixel-level prediction)

- object detection (classification and localization of multiple objects)

- image restoration (high quality from damaged input)

- facial detection

- object tracking (ID for each initial detection and following it through the frames of a video)

- image retrieval systems (find similar images)

### - Natural language processing
The main terms in NLP are the following:

- Corpus: collection of textual documents

- Terms: keywords that convey the information

- Index: list of terms that appear only once

- Reverse index: list of documents in which each terms appears

- Tokenization: division of texts into tokens, removing stop wards

- Stop words: meaningless words

- Stemming: root in compound words of conjugated verbs

In the "space-vector model" key tokens are translated into numerical language to apply an algorithm. Using word embeddings, terms of the index are mapped in a vector space. Terms are pairwise orthogonal. The mathematical model of a text is a term-document matrix, in which a document is a linear combination with positive coefficients of the vectors corresponding to terms in it. Each document is associated with a vector of length equal to the index. Different approaches are:

- BOW: presence of a term is 1 (not take into account frequency)

- Count vectorizer: weight indicating the number of occurrence and the document of occurrence

- TF-IDF: use concept of reverse index

An alternative to term-document matrix is the "symmetric co-occurrence matrix", which is the numerical representation of the corpus using a matrix $m \times m$, where $m$ is the number of terms. The matrix is symmetric and the entrances $a_{ij}$ indicate the time of appearance of term $i$ near term $j$. The similarity between two texts is the cosine similarity (or Jaccard, or Manhattan).
Feature selection techniques involve SVD and NNF (favors the interoperability of the model).
Text representation through projections on vector spaces do not reproduce a correct semantic and contextual understanding, so embedding model based on NN try to learn the representation of a text one step at a time and can be used to predict the probability of a word in a given context.

- Word2Vec: two templates used to produce word embeddings. CBOW predict the target word given the context, while Skip-Gram predict the context given a target word.

- Glove: weighted least square model trained on global co-occurrence between terms. Predict the probability that a word is in the context of another.

- RNN: multiple hidden NN that can use non linear operation to predict a word given the sequential context (most common LSTM)

- Transformer: use transformers to get a bi-directional context. GPT and BERT are pre-trained and satisfy any NLP benchmark.

Chat-bots are examples of applications of NLP in digital marketing. They can be distinguished bye type of intelligence:

- Command-based: limited set of commands, structured prompts, not use natural language

- AI-based: linguistic and dialogue management capability, based on semantic engines or third-party AI platforms

They are also distinguished based on type of response:

- Retrieval-based: repository of predefined responses

- Generative: use DL to translate form input to output

Also on domain extension:

- Open domain: ask anything on any topic

- Closed domain: only specific question related to a certain topic

Main functions are: customer care, entertainment, increase reach, offer something more, sell, increase productivity, get to know your customers better, recruiting, sentiment analysis.

### - Main approaches to the measurement of marketing activities
Companies rely on correct metrics to measure their success, without assessing whether these metrics accurately reflect the economic value generated by specific marketing activities.
Understanding the effect of each marketing activity on brand perception, incremental sales and customer base growth is essential for proper corporate budget allocation, management and planning.
Three main methodological approaches form measuring marketing activities are:

- Main metrics: evaluation on KPIs

- Attribute models: allow the results of marketing activities to be attributed to specific tuchpins with the customer, to identify the most influence ones.

- Marketing Mix Models

### - Methods for measuring marketing activities
Objective of each marketing activity is to carry out a specific action in order to impact their purchasing behavior (change to achieve business objectives).
Objective of a company is to obtain an increase in the margin compared to baseline through increase in revenues, so increase in conversion rate or decrease in costs.
Main indicators that evaluate an incremental margin increase are ROI and time to break-even (payback time).
To have an a priori assessment of the impact of an initiative and ex-ante business case is drawn up:

- The customer or prospect target is indicated

- The engagement mechanic is identified

- The unique selling preposition is defined

The definition of these drivers is base on the anlysis for the customer base, and allow to estimate costs to be incurred and expected revenue. Ex-post it is essential to periodically monitor the progress of the initiative.
To monitor the changes in its turnover considering:

- ROI (Return on investment, must be positive)

- Cost per sale (cost of transformation form prospect to customer

- Cost per lead (conversion cost form user to prospect)

- Conversion rate for inbound channels (ratio between who fill out the contact form and who enters the company platforms)

- ADS

- SEO (site indexing strategy)

- Brand awareness growth (using questionnaires)

- Construction of control group (benchmark group)

- A/B testing (to evaluate the effectiveness of two version of campaigns)

Control group is a set of customers similar to each other whose behavior is observed in the same conditions as the target without receiving the marketing action. The groups are homogeneous (through representative sampling stratification, they have the same proportions of the reference population). The study reveals insights into the effectiveness of marketing activity. Is possible to use historical control group, but is less accurate because not take into account the context and time. Sampling techniques involve random sampling and stratified sampling.

A/B testing involve different variants of marketing actions proposed to the target divide in two homogenise groups, and evaluate main KPIs. It is used mainly to creativity, graphics, informative texts, promotional offers and calls-to-action. The MBA allow to dynamically allocates traffic between different versions.

### - Digital marketing attribution models
Attribution models allow to measure the value of campaigns and understand which tuchpoints have the greatest influence on the conversion journey of buyers. The key concepts are:

- Conversion (desired goal)

- Conversion Value (total cost of actions)

- Tuchpoints (in the puchasing process)

- Contributions (different moments in pyrchasing decsion)

The goals are the maximization of volumes and a budget efficent spending.

Whit the algorithm choice is also important to understand influence of key characteristic of the path such as:

- length of user path (number of tuchpoints)

- information recency (location of tuchpoins in user journey)

- tuchpoint granularity

Attribution models can be based on:

- Rules: simple approach that assign values to each tuchpoint based on a predefined formula.

  - single-touch attribution: 100% credit to only one tuchpoint (the first)
  - linear attribution: equal credit to each tuchpoint
  - custom attribution: arbitrary credit to each tuchpoint

- Data-driven: use historical data to decide the value of tuchpoint, it is based on estimated probabilities of conversion during the various tuchpoints. Algorithms are Shapley's value, Markov models and other regression models.

  - Markov model: probabilistic model that represent customer journey as a graph (nodes = tuchpoints, edges = observed transitions). Get the probability of success of a given customer journey given the history of all customer journeys, get the importance of each campaign with the take-down effect. Calculate attribution value.
  - Shapley's value: allocate each tuchpoint contribution to the conversion using the following rules
    * Marginal contribution

- * interchangeable players have the same value
- * factious players have zero value
- * when there are multiple parts, the cost can be broken down

Channel coalition does not take into account the sequential effect of the channels. The results are built using existing, non-simulated customer journeys.

- Regression: modelled as a classification problem where each customer journey makes an observation with the binary conversion indicator as dependent variable and use logistic regression.

### - Marketing Mix Models

MMM derive form 4P marketing concept (product, price, place, promotion). It consists of statistical models to measure the effectiveness of communication spending using historical data in aggregate form, with the aim of analyzing sales trends, according to advertising spending variables and in various marketing channels. The digital channel allowed the digital attribution.

Data collection is necessary in statistical methodology, Data is grouped by variables:

- target: concerns sales, divided according to sales channels (others can be: purchase conversions, number of new customers in a period of time, volume of traffic on the site and subscriptions in the stores)

- independent: average metric/marketing/control

Statistical representation is linear model.

Response pattern to advertising are:

- Current effect: variation in sales by advertising that occur in same period

- Curryover effect: occur in time following the launch

- Shape effect: change in sales compared to the increase in investment in advertising

- Competitive effect: reactivity in adv compared to competitors

- Dynamic effect: variations of effects of adv over time

- Content effect: change in sales in response to change in adv content

- Media effect: changes in sales in response to promotional stagiest on different channels

  - Halo effect: customer approaching a product thanks to the positive experiences of purchasing products of the same brand

  - Cannibalization effect: negative impact on a product due to negative performance of other products of the same brand

The implementation of the econometric model can be:

- Linear regression model: consider sales as the dependent variable and all marketing activities and other external control factors among the independent variables. The limitation is the possibility of only estimate the current effect. Adeguate for a company with a stationary product, operating in a mature market, with moderate seasonality and made up of several competitors.

- Multiplicative regression model: multiplicative relationship between the independent variables of the MMM, describes the effect of $m$ possible variables on sales. Logarithmic transformation of the target variable linearizes the mode (additive). The main benefits are:

  - the dependent variable undergoes the effect of the interaction of all the marketing variables considered

  - include a variety of forms of the response variable, making the model flexible

  - the coefficient both estimates the effects of the independent variable and represents a sales elasticity value with respect to advertising

The limitation is the possibility to estimate current effect and shape effect.

- Non-linear and hierarchical model: allow to estimate of two other effects:

  - Carryover effect: by transforming the time series of the media variables through the adstock function. Assumes that the highest values occur in the same exposure period as the ad. To obtain an effect of the ADV require more time and therefore higher values may not coincide with the exposure period, so delayed adstock function is introduced (control on the delay effect).

  - Saturation of the adv (S-shaped effect): transforming the time series of the media variables through the Hill function. Allow for different effects for different media, a multiplicative coefficient is applied to the Hill function.

  The combination of the two separate effects is possible, where the adstock function is applied first, then passing to the application of the Hill function or inverting the order of the transformations.

To estimate parameter so the MMM it is possible to use: maximum likelihood estimation or Bayes theorem.
The main issues in MMM are:

- dataset size

- limited range of data selection bias:

  - wrong target

  - seasonality

  - funnel effect due to a distortion caused by the interference of the impact of an adv channel with others inserted for the same marketing campaign