MKTG6018
Customer Analytics &
Relationship Management
2022 S2 期末复习 Part1

Prepared by 糖糖

MKTG6018 – Customer Analytics & Relationship Management

Final Exam
Worth: 30%
Individual/Group: Individual
Due: TBA

The final exam will assess students' understanding of the concepts and tools covered in the course and their ability to apply these in a practical marketing context. The material covered will include all lectures, cases and prescribed readings discussed throughout the semester (weeks 1 - 13). The final exam will be closed book and consist of written extended response style questions. The focus will be not only on demonstrating your individual knowledge of the substantive material but also on your ability to apply that knowledge to a novel context. It is worth 30%.

Week 1: Introduction to Customer Analytics, CRM and Data Science

Learning Objectives:

- Understand what **customer analytics** and **CRM** are and **why** they are important and useful
- **How and why** customer data can be used for customer analytics and customer understanding
- Describe **differences** between BI and Data Science
- Have a deep understanding of the Data Science process
- Use effective visualisations to communicate insights from data (Workshop)

1. Marketing: From 4Ps to 4Cs

product, price, place, promotion →
consumer, costs, convenience, communication

Marketing Mix: 4Ps (Product, Price, Place and Promotion)

• **Product**

> o ***Definitions*:** A product refers to an item that satisfies the consumer's needs or wants.
> Products may be tangible (goods) or intangible (services, ideas or experiences).
>
> o ***Marketing Decisions*:**
>> ▪ Product design – features, quality
>> ▪ Product assortment – product range, product mix, product lines
>> ▪ Branding
>> ▪ Packaging and labelling
>> ▪ Services (complimentary service, after-sales service, service level)
>> ▪ Guarantees and warranties
>> ▪ Returns
>> ▪ Managing products through the life cycle

• **Price**

> o ***Definitions:***
>> ▪ Price refers to the amount a customer pays for a product.
>> ▪ Price may also refer to the sacrifice consumers are prepared to make to acquire a product (e.g. time or effort).
>> ▪ Price also includes considerations of customer perceived value.
>
> o ***Marketing Decisions:***
>> ▪ Price strategy
>> ▪ Price tactics
>> ▪ Price-setting
>> ▪ Allowances – e.g. rebates for distributors
>> ▪ Discounts – for customers ▪ Payment terms – credit, payment methods

- **Place**
    - o *Definitions:* Refers to providing customer access to products or services (location, distribution)
    - o *Marketing Decisions:*
        - ▪ Strategies such as intensive distribution, selective distribution, exclusive distribution
        - ▪ Franchising
        - ▪ Market coverage
        - ▪ Channel member selection and channel member relationships
        - ▪ Location decisions
        - ▪ Inventory
        - ▪ Transport, warehousing and logistics
- **Promotion**
    - o *Definitions:* Promotion refers to marketing communications. May comprise elements such as: advertising, PR, direct marketing and sales promotion.
    - o *Marketing Decisions:*
        - ▪ Promotional mix - appropriate balance of advertising, PR, direct marketing and sales promotion
        - ▪ Message strategy - what is to be communicated
        - ▪ Channel/ media strategy - how to reach the target audience
        - ▪ Message Frequency - how often to communicate

Marketing 4Cs: Consumer, Cost, Convenience and Communication
- (From Product to) **Customer Needs and Wants**
    - o *Definition and Marketing Decisions*: A company will only sell what the consumer specifically wants to buy. So, marketers should study consumer wants and needs in order to attract them one by one with something they want to purchase.
- (From Price to) **Cost to Satisfy**
    - o *Definition and Marketing Decisions*: Price is only a part of the total cost to satisfy a want or a need. The total cost will consider for example the cost of time in acquiring a good or a service, a cost of conscience by consuming that or even a cost of guilt "for not treating the kids". It reflects the total cost of ownership. Many factors affect cost, including but not limited to the customer's cost to change or implement the new product or service and the customer's cost for not selecting a competitor's product or service.
- (From Place to) **Convenience**
    - o *Definition and Marketing Decisions*: In the era of Internet, catalogues, credit cards and phones, consumers neither need to go anywhere to satisfy a want or a need nor are they limited to a few places to satisfy them. Marketers should know how the target market prefers to buy, how to be there and be ubiquitous, in order to guarantee convenience to buy. With the rise of Internet and hybrid models of purchasing, Place is becoming less relevant. Convenience takes into account the ease of buying the product, finding the product, finding information about the product, and several other factors.
- (From Promotion to) **Communication**
    - o *Definition and Marketing Decisions*: While promotion is "manipulative" and from the seller, communication is "cooperative" and from the buyer with the aim to create a dialogue with the potential customers based on their needs and lifestyles. It represents a broader focus. Communications can include advertising, public relations, personal selling, viral advertising, and any form of communication between the organization and the consumer

Customer Experience as Competitive Advantage

Competition moves: products → services → customer experiences.

2. Customer-related Data

	Traditional Approach	Dynamic Approach
	Traditional approach	High-value, dynamic approach - source of competitive differentiation
WHO?	Descriptive data •Attributes •Characteristics •Self-declared info •(Geo)demographics	Interaction data •E-Mail / chat transcripts •Call center notes •Web Click-streams •In person dialogues
WHAT?	Behavioral data •Orders •Transactions •Payment history •Usage history	Attitudinal data •Market Research •Social Media

3. Customer Analytics

Definition: Customer analytics refers to the processes and technologies that give organizations the customer insight necessary to deliver offers that are anticipated, relevant and timely at a profit.

Customer Behaviour Offerings:

Digital Analytics; Marketing Optimisation; Customer Understanding; Marketing Strategy

4. Customer Relationship Management (CRM)

Definition

- CRM is the core **business strategy** that integrates internal processes and functions, and external networks, **to create and deliver value to targeted customers at a profit**. It is grounded on high quality **customer-related data** and enabled by **information technology**.
- CRM is the business processes and supporting technologies that support the key activities of targeting, acquiring, retaining, understanding, and collaborating with customers.

The guiding objectives of CRM is Get-Keep-Grow.

Get → Keep → Grow

- Get: Employ awareness and acquisition tactics while reducing customer acquisition costs and getting right customers.
- Keep: Retain customers via loyalty programs, product updates and quality service.
- Grow: Increase customer lifetime value by up-selling and cross-selling.

CRM strategies can include all activities that can improve any of the stages in the customer lifecycle. So we can build and maintain better and stronger relationships with our customers or potential customers.

Why CRM matters?

Customers who have great relationship with us will buy again (customer retention), buy more (cross-selling and up-selling, "customer development"), and recommend us (customer advocacy).

Types of CRM

Note that we will focus more on strategic and analytical CRM. In contrast, operational CRM is enabled by CRM software. The use of CRM software is not our focus.

Type of CRM	Dominant characteristic
Strategic	Strategic CRM is a core **customer-centric business strategy** that aims at **winning and keeping profitable customers.**

Operational	Operational CRM focuses on the **automation of customer-facing processes** such as selling, marketing and customer service.
Analytical	Analytical CRM is the process through which organizations **transform customer-related data into actionable insights** for either strategic purposes.

Real Problems Examples

- Purchase: scanner data and consumer expenditure survey
- Consumer review: Amazon and Trip advisor data
- Consumer complaint database about financial products and services (CFPB Data)
- Flight on-time performance by airlines in the U.S and Australia
- Recycling and Donation Data - Social media data
- Cart-Item Analysis (recommender system)
- Covid-19
- AirBnB

Tools
Collection (Python, Power Query)
Analysis and Machine Learning (Excel, SPSS, STATA, LIWC, Python, RapidMiner)
Visualization (Tableau, Python)

5.Data Science
Definition: Data science is defined as a sophisticated new discipline requiring advanced skills and competencies in such areas as statistics, computer science, data mining, mathematics, and computer programming. As it has been said time and again, data scientists are the business 'rock stars' of the current century.
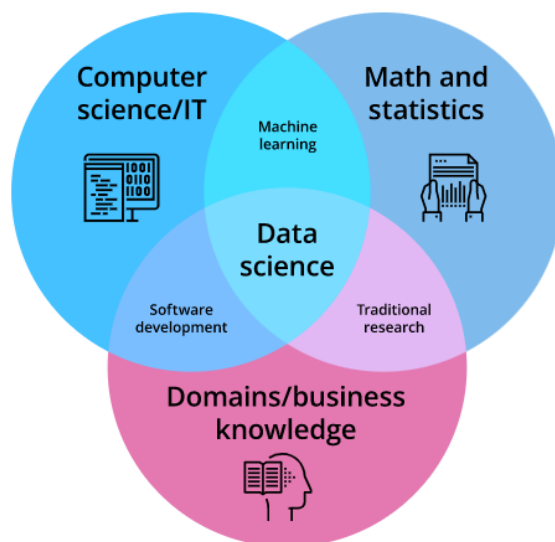
How is data science different to business intelligence
These two disciplines follow different objectives as they look for answers to different questions.

Business Intelligence	Data science

BI focuses on **descriptive analytics**, which pertains to **'What happened?'** types of questions. Some examples are as follows:

- How many products did I sell last month?
- What were last Christmas sales by postal code?
- How many products were returned last month?
- How much were revenues and profits of the company during the last quarter?
- How many employees did I recruit last year?

BI pertains to reporting on the current state of business, which is commonly known as Business Performance Management (BPM). BI provides retrospective reports helping business users monitor the current state of business and find answers to questions about the historical business performance. These reports and questions are central to business, sometimes needed for regulatory and compliance reasons.

BI may also apply some rudimentary analytics (time series analysis, previous period comparisons, indices, shares, and benchmarks) so that business users would be able to flag underperformance and overperformance areas. However, even these analytics are based on monitoring what occurs to business.

Data scientists look for variables and metrics that would **predict** business performance much better. Consequently, data scientists turn their attention to such types of questions as predictive analytics: **What is likely to happen?** and prescriptive analytics: **What should I do?**. Some examples are as follows:

- Predictive questions:
  o What is likely to happen?
  o How many products will I sell next month?
  o What were last Christmas sales by postal code?
  o How many products were returned last month?
  o How much were revenues and profits of the company during the last quarter?
  o How many employees did I recruit last year?

- Prescriptive questions:
  o What should I do?
  o Order [10,000] Component X to support widget sales for next quarter.
  o Hire [Z] new sales reps by these zip codes to deal with projected Christmas sales.
  o Put aside [$125K] in financial reserve to cover Product X returns.
  o Sell the following product mix to realise quarterly revenue and margin goals.
  o Increase hiring pipeline by 35% to realise hiring goals.

In order to find answers to answer these predictive and prescriptive questions, data scientists develop analytic models to quantify cause and effect.

Another difference between BI and data science lies in the attitudinal characteristics and work approach of the people who undertake these roles.

Area	BI Analyst	Data Scientist
Focus	Reports, KPIs, trends	Patters, correlations, models
Process	Static, comparative	Exploratory, experimentations, visual
Data sources	Pre-planned, added slowly	On the fly, as needed
Transform	Upfront, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	Good enough, probabilities
Data model	Schema on load	Schema on query
Analysis	Retrospective, descriptive	Predictive, prescriptive

Explore the different perspectives pertaining to 'data quality.' In view of a BI analyst dealing with historical data, the data must be 100% accurate. BI and data warehouse organisations have invested too much in data governance and master data management to obtain assurance that the data warehouse content is 100% accurate.

Meantime, the data scientist seeks to predict what is likely to happen in the future and based on that, they deal with probabilities, confidence levels, F-distributions, t-tests, and p-values. Data scientists acquire a notion of what is 'good enough' in predicting what is likely to happen and recommending what actions to take because the future is never 100% predictable.

Common terminology

Statistics; Database Query; Data warehousing; Regression Analysis; Machine learning; The data analytics process

6.Cross Industry Standard Process for Data Mining (CRISP)

Data scientists embrace failure as a learning tool. In their quest to discover new measures and variables that are better predictors of performance, data scientists learn to accept failure as part of their agile, fail-fast methodology. A common approach embraced by data scientists is modelled after Cross Industry Standard Process for Data Mining (CRISP).

CRISP

The CRISP process diagram sheds light on the fact that iteration is a rule rather than an exception. Generally speaking, it would not be a failure to go through the process once before the problem has been resolved. The whole process involves data exploration, and following the first iteration the data science team would get more knowledge. The following iteration may be better-informed. Let's now outline the steps.
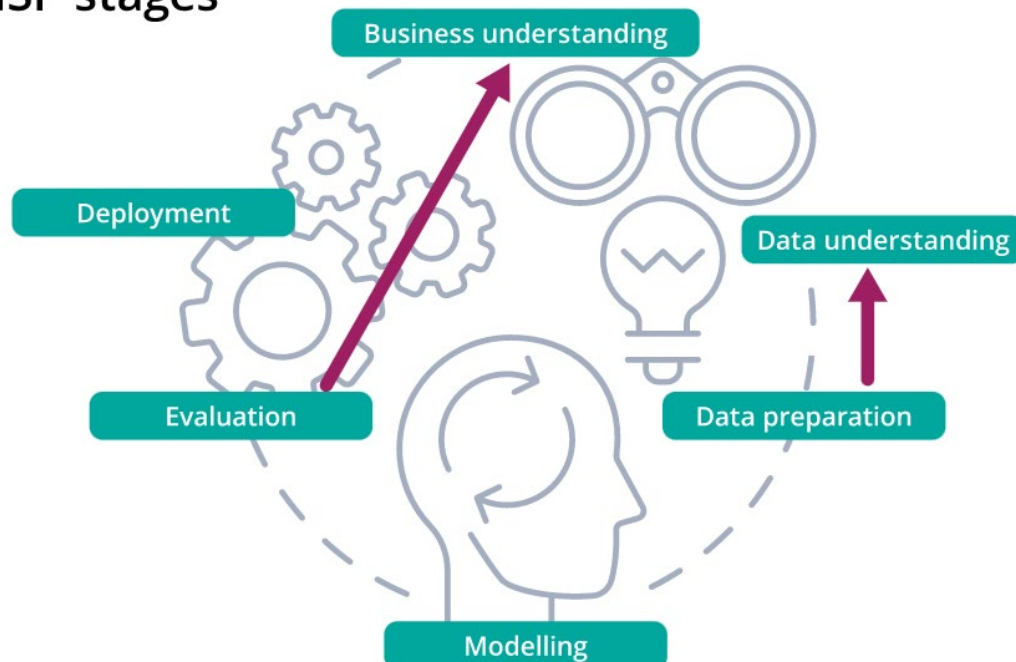
Business understanding

First and foremost, it is crucial to understand the problem before a solution may be found. That may sound easy, but seldom do business projects come pre-packaged as clear data mining problems without any ambiguity. Most of times, recasting the problem and developing a solution would offer an iterative process of discovery.

The diagram presented in this figure shows this as **cycles within a cycle**, rather than simply as a linear process. **Because the first formulation may not be comprehensive or ideal, numerous iterations may be required to arrive at a satisfactory solution formulation**. The business understanding stage is a portion of the craft in which the analysts' inventiveness is crucial. Although data science has a lot to say, the key to tremendous success is frequently a creative issue formulation by an analyst on how to turn the business challenge into one or more data science problems. Knowledge of the fundamentals at a high level can help creative business analysts experience new formulations.

In this first phase, the design team should carefully **consider the use scenario**.

- What exactly are we going to do?
- How exactly would we be able to do it?
- In this use scenario, what parts represent possible data mining models?

Further going through it, we will start with a simplified form of the use scenario, but as we go ahead we will loop back to find out that the use scenario must mostly be adjusted so as to reflect the actual business need more effectively.

Data understanding

If the objective is to solve the business problem, the data consists of available raw material of which the solution will be born out. **It is key to understand the strengths as well as constraints of the data as rarely is there an exact match with the problem.**

Most of times, historical data is gathered for **purposes** that have nothing to do with the current business problem or there may no explicit purpose at all. The information contained in a customer database differs from the information in a transaction database or a marketing response database. Therefore, they may cover different intersecting populations, while having varying degrees of reliability.

The **costs** of data may also vary. Some data will be available virtually freely but some other will necessitate effort to obtain. There is also data which may be purchased. Yet other data simply won't exist, thereby requiring the entire ancillary projects to arrange their own collection. Estimating the costs and advantages of each data source, as well as assessing if future investment is warranted, is an important aspect of the data comprehension process. Even after the acquisition of all datasets, their collation may require extra effort.

For instance, customer records and product identifiers are extremely variable and noisy. Cleaning and matching customer records to verify solely one record per customer is a complicated analytics problem by itself. With the progress of data understanding, solution paths are likely to change direction in reaction, and team efforts may even go off the right track. Fraud detection would help illustrate this.

When it comes to data understanding, we should **dig beneath** the surface to explore the business problem structure and the available data. Then comes the phase of matching them to one or more data mining tasks for which we may apply substantial science and technology.

It is not uncommon for a business problem to necessitate the fulfilment of several data mining tasks, mostly of diverse types, and combining their solutions will be a must.

Data preparation

The analytical technologies we can use are powerful, but they have specific requirements on the data they use. They may require data to have a **form** different from the natural provision of the data, and in some cases conversion may be needed.

A data preparation phase thus proceeds mostly along with the understanding the data, where the data are manipulated and converted into forms to produce better results. Some typical examples of data preparation include data conversion to tabular format, removal or inference of missing values, and conversion of data to diverse types. Certain techniques of data mining are developed for symbolic and categorical data, with others handling merely numeric values.

Furthermore, numerical values should be normalised or scaled to become comparable. Standard techniques and general rules may be used in doing such conversions.

More generally speaking, data scientists may spend a big amount of time throughout the process on defining the variables used later on in the process.

This is one of the major areas where human creativity, common sense, and business knowledge are proving instrumental.

Modelling

Modelling is the topic of the upcoming weeks and we will not focus on it here, just to say that **the output of modelling is a kind of model or pattern capturing data regularities**. The stage of modelling is the primary place where data mining techniques apply to the data. Having some understanding of the fundamental ideas of data mining, including the sorts of techniques and algorithms that exist, is important since it is the part of the skills required in most branches of science and technology.

Evaluation

The purpose of the evaluation is **to rigorously assess the results of data mining and to gain confidence about their validity and reliability before taking any further steps ahead**. By looking hard enough at any dataset we can find patterns; however, they are unlikely to survive careful scrutiny.

We want to be assured that **the models and patterns extracted from the data are true regularities and that they are not just idiosyncrasies or sample anomalies**.

We may deploy results shortly after data mining, but it is not recommended; it would be much easier, cheaper, faster, and safer to test a model in a controlled laboratory setting first. In the meantime, the evaluation stage also helps ensure that the model satisfies the original business goals.

Let's keep in mind that the primary objective of data science for business is to support decision-making. We should also remember that we initiated the process by concentrating on the business problem we intend to solve. A data mining solution may only be a piece of the larger solution, which would need to be evaluated as such.

Deployment

In deployment, data mining results—and more importantly the data mining techniques themselves—are **put into real use with a view to realising some return on investment**. The clearest cases of deployment pertain to implementation of a predictive model in a information system or throughout a business process.

Churn analysis involves a model for predicting the likelihood of churn, which could be integrated with the business process for churn management, for example, through special offers to customers who are likely to be particularly at risk.

Two major reasons justifying deployment of the data mining system itself rather than the models developed thereby are as follows:

1. the pace of changes in the world may be faster than the speed of adaptation of data science team, as is the case with fraud and intrusion detection
2. a business offers too many modelling tasks for its data science team to manually create each model individually.

In these cases, the best option would be to **deploy the data mining phase into production**. For that purpose, it is vital to design the process to alert the data science team on any seeming anomalies and to provide fail-safe operation.

Furthermore deployment may be much less 'technical'. In a known case, data mining led to the discovery of a set of rules that help diagnose a common error in industrial printing quickly in order to fix it. Data scientists should generally remain involved in the project until final deployment, serving as advisors or developers depending on their expertise. Regardless of the success or failure of the deployment, the process often returns to the phase of business understanding. The mining data process produces too much insight into the business problem as well as difficulties of its solution. A second iteration may yield an improved solution. Merely the experience of reflecting on the business, data, and the performance objectives often results in the development of new ideas for business performance improvement, and even new lines of business or new ventures.

A very similar approach is pursued by **data science**:

1. develop a business hypothesis or question
2. discover different data and analytics combinations to build, test, and refine the analytic model and
3. wash, rinse, and repeat until the model proves its ability to provide the required 'analytic lift' while giving a satisfactory goodness of fit.

At the end of the day, the analytics are deployed or operationalised, including possibly rewriting the analytics in a different language to speed up the model execution (i.e., in-database analytics) and integration of the analytic models. It finally results in the organisation's operational and management systems.

7. What is data visualization and why do we do it?
Data visualisation
Data visualisation is how we communicate numerical or quantitative information in a visual manner. To do this, we use things like charts, graphs, maps, and sometimes even pictures and iconography. The medium can be anything you want: print an interactive web page, a static infographic, crayons on paper, or even three-dimensional objects like modelling clay, among many, many other things.

What is data visualisation and why do we do it?
According to Krishna in his article, ***data visualisation is a process that consumes data as input and transforms it into insights and stories as output***. There are no stipulations on how you do it. So, the sky's the limit. **Why** did we do data visualisation? Have you ever tried to scan a list of numbers such as the one shown here to find the highest value, not too terrible? But what about the second highest or the 10th and the lowest? ***It's very difficult for us humans to read, comprehend, and store*** all these numbers in our memory, and then organize them on the fly in a way that makes sense. This is why we rely on visualisation to help us with this problem.
Check out the list now, it's the same data, but we've **converted the numbers into bars and sorted them in descending order**. Since presumably, the biggest number is the most important, we want to have it front and centre for us to internalize it easily and quickly. This is only one small example. You'll learn about why we need data visualisation, especially in our world today where big data is so prevalent.

Imagine you own a company with hundreds of products and millions of customers. You can't be expected to sift through every individual order to gain insights about what your most popular product is or what product is causing you to lose the most money. Again, this is where data visualisation comes in handy. You can learn these things and so much more by building a simple viz that does the heavy lifting for you.
Lots of people do data visualisation, probably more than you might first imagine. Basically, anyone who must use some written or visual medium to communicate numbers to someone else is in the business of data visualisation.
So have a look at some of the various roles where data viz skills might be applicable. What kind of jobs can you think of where data visualisation skills would be useful? It is so important when you're entering a new field to get a feel for the history of it. To do so helps you build a good foundation and better appreciate the advances that we have and sometimes haven't made in modernity.
- Data and Analytics Manager
- Business analyst
- Database administrator
- Statistician
- Data engineer, analyst or architect
- Data Scientist

History
Though it doesn't quite predate human language the way art does, data visualisation, surprisingly has about as lengthier history as astronomy. In fact, detailed astronomical information about the locations of stars can be found in the Lascaux cave, dating all the way back to the Pleistocene era in Southern France.
We can also see the visualisation of quantitative information in the Quipu or talking knots used by the Inca people around 2600 BCE. These were used to record and keep records about everything from census data to taxes.
The earliest documented visualisations undoubtedly are maps such as the Turin papyrus map dating back to 1160 BCE in Egypt, which accurately depicts a location of stone needed

to build statues of King Ramesses the fourth. Now for the next couple of thousand years, the most notable advances in data visualisation were made in the world of cartography. But around the 10th or 11th century, we start once again seeing evidence of people recording astronomical information about planetary movement across the sky. By the 1500s, people were regularly using instruments and tools to precisely measure geographical and astronomical locations as well as other physical quantities.

By about the 1600s, we start getting into the realm of iconic infographics according to RJ Andrew's interactive timeline of the most iconic infographics, I highly encourage you to take a look. See if you can find Florence Nightingale's contribution in the timeline.

One of my favourites missing from RJ's list though is Napoleon's march during his Russian campaign of 1812. This map expertly depicts six dimensions of data using only two dimensions on the page. It shows the number of troops, the distance they travelled, the temperatures they endured, the latitude and longitude of their travels, the direction of their travel, and time. As you grow in your data viz skills, you will increasingly appreciate what a feat of visualisation this is. I'll leave it to explore RJ's interactive timeline on your own. I think you'll really enjoy not only the information within, but the adorably illustrated way he depicts that information.

Week 2: Customer Analytics and Big Data Concepts

Learning Objectives:
- Understand the **fundamental shift in marketing** from the transaction (product) to a relationship (customer)
- Describe and identify the steps in the **marketing planning process**
- Explain how **data-driven decision-making** impacts marketing
- Identify **types of analytics** used in marketing
- Apply lessons from how Airbnb used analytics to improve its marketing strategy to other contexts.
- Describe the **Big Data landscape**, including examples of real-world big data problems, including the three key sources of Big Data: people, organizations, and sensors.

1. Shifting Paradigm in Marketing

There was a shifting paradigm from 4Ps to 4Cs. Our marketing decisions are now centring around our customers.

The emphasis is on communicating the right stuff to the right customer in a convenient and cost-efficient way.

- **The consumer wants and needs (vs. Products)**
    - **You can't develop products and then try to sell them to a mass market**. You have to study consumer wants and needs and then attract consumers one by one with something each one wants. In most cases, you have to find out what people want and then "build" it for them, their way.
- **Cost to satisfy (vs. Price)**
    - **You have to realize that price - measured in dollars - is one part of the cost to satisfy**. For example, if you sell hamburgers, you have to consider the cost of driving to your restaurant, the cost of conscience of eating meat, etc. One of the most difficult places to be in the business world is the retailer selling at the lowest price. If you rely strictly on price to compete, you are vulnerable to competition - in the long term.
- **Convenience to buy (vs. Place)**
    - **You must think of convenience to buy instead of place**. You have to know how each market subset prefers to buy - on the Internet, from a catalogue, on the phone, using credit cards, etc. Lands End

clothing, Amazon Books and Dell Computers are just a few
businesses that do very well over the Internet.

- **Communication (vs. Promotion)**
    - ○ **You have to consider communication instead of a promotion**.
      Promotion is manipulative (ouch!) - it's from the seller.
      Communication requires a give and take between the buyer and
      seller (that's nicer). Be creative, and you can make any advertising
      "interactive". Use phone numbers, your website address, etc., to
      help here. And listen to your customers when they are "with" you.

2.Rethinking Marketing
In the past, companies were mainly concerned about product profitability, current sales,
brand equity, and market share.
These metrics are still usable, but there are better metrics for a new model.

New Metrics for a New Model

OLD APPROACH	NEW APPROACH
PRODUCT PROFITABILITY	CUSTOMER PROFITABILITY
CURRENT SALES	CUSTOMER LIFETIME VALUE
BRAND EQUITY	CUSTOMER EQUITY
MARKET SHARE	CUSTOMER EQUITY SHARE

customer profitability: how profitable your customers are
CLV: Customer Lifetime Value.. It is the total sum of profit that a given customer will
generate over her or his lifetime. It is a forward-looking metrics.
customer equity: the total sum of the customer lifetime values of your company
customer equity share: the value of a company's customer base divided by the total value of
the customers in the market

3.Customer Analytics
Definition: Descriptive Analysis (the 'what') & Predictive and Prescriptive Analytics (the "so
what")
Marketing Processes

How to Use Customer Analytics to Turn Customer Data in Value: 5 Steps

- **Define a customer data and analytics strategy (Step 1)**
  Customer-centric decision support
  Step one involves taking a detailed look at consumer data and analytics capabilities. It's imperative to ensure strategic alignment of data, analytics, and technology capability development priorities to measure brand strategy, consumer ambition, and business objectives.

- **Build a customer data management, analytics and technology ecosystem (Step 2)**
  'Voice of the customer' analytics with blended data
  Step two focuses on creating a data management environment encompassing multi-sourced, internal, external and linked data in various environments.

- **Develop data models and analytics engines (Step 3)**
  Actionable insights
  For step three, create data models by blending, enriching, analysing and sharing the data via visualisation tools or reports for various analytics requirements such as social media insights, 'voice of the customer', 'voice of the employee', behavioural segmentation, marketing campaign effectiveness and sales and service optimisation.

- **Generate and interpret insights (Step 4)**
  Business dashboards
  Turn insights into actions by integrating the information with business and customer processes to improve customer engagement, service delivery and products. Build systems and processes of insights for marketing, sales, customer experience, product, service and channel teams.

- **Establish management controls (Step 5)**
  Governance and Performance Framework
  A critical final step is to set up the governance of your data. Define organisation structures, define roles and responsibilities, target operating models and performance and reporting mechanisms. Identify the data sources needed to calculate metrics, and establish measurements for a continuous improvement process.

4. Predictive Analysis - Six Steps in Predictive Analytics

Step 1: Identify data from a variety of sources. Potentially valuable data often exists in multiple hard-to-access locations, both internally (data silos in enterprise applications) and externally (social media, government data, and other public or licensed data sources). Advanced data visualization tools can help to explore the data from various sources to determine what might be relevant for a predictive analytics project.

Step 2: Wrangle the data. Data preparation for predictive analytics is a key challenge. Many users of predictive analytics spend more than three quarters of their time preparing the data: calculating aggregate fields, stripping extraneous characters, filling in missing data, or merging multiple data sources.

Step 3: Build a predictive model. Predictive model building: can choose to run the best predictive model. The best algorithm(s) to choose depend on the type and completeness of the data and the type of prediction desired. Analysts run the analysis on a subset of the data called "training data" and set aside "test data" that they will use to evaluate the model.

Step 4: Evaluate the model's effectiveness and accuracy. Predictive analytics is not about absolutes; it is about probabilities. To evaluate the predictive power of the model, data analysts use the model to predict the "test data" set. If the predictive model can predict the test data set, it is a candidate for deployment.

Step 5: Use the model to deliver actionable prescriptions to your business peers. There is little value in a prediction if it doesn't enable the seizing of a predictive opportunity or avoiding a negative event. Business peers need to learn to trust in the predictions of models and those creating the models need to learn from their partners in the business what the most actionable insights may be.

Step 6: Monitor and improve the effectiveness of the model. Predictive models are only as accurate as the data fed into them, and over time they may degrade or increase their effectiveness. To monitor models for ongoing effectiveness and value, newly accumulated data is rerun through the algorithms. If and when the model becomes less accurate, AD&D pros (application development and delivery professionals) will have to adjust the model (e.g., by adjusting parameters in the algorithms) and/or seek additional data.

4.Big Data

Big Data: a growing torrent

Cloud computing: Computing anywhere and anytime

Computing anywhere and anytime + Data Torrent

→dynamic and scalable data analysis

→ Big data Era

Applications: What makes big data valuable

Big Data→ Better Models→ Higher Precision

Personalized Marketing

Recommendation Engines

Sentiment Analysis

Natural language processing

Mobile Advertising (Customer profile/Recent purchases + Geolocation data)

Consumer Growth to Guide Product Growth(Collective Consumer Behavior)

Where Does Big Data Come From?

Three major sources of Big Data

Machines: Data generated by real-time sensors in industrial machines or cars that track users' online behaviour, environmental sensors, personal health trackers, and a variety of other sense data resources are referred to as machine-generated data.

People: when we talk about human-generated data, we're referring to the massive amounts of social media data, status updates, tweets, images, and other material.

Organisations: organisational generated data includes more typical data kinds such database transaction information and structured data kept in data warehouses.

Machine-Generated Data: It's Everywhere and There's a Lot!

Machine data is the largest and most complex source of big data when compared to all other sources.

What makes a smart device smart?

> *Connect to other devices / networks*
>
> *Collect and analyze data autonomously*
>
> *Provide environmental context*

E.g.: Activity Tracker: Body temperature; Stairs climbed; Heartbeat; Distance walked

The Internet of Things (IoT) was coined as a result of the increased availability of smart gadgets and their interconnectedness.

Increasing number of machines that sense **X** Data collected by each device **=**

Machines → Biggest Source

Big Data Generated By People, The Unstructured Challenge

Facebook, YouTube, Twitter → Text-heavy Unstructured

The majority of this data is unstructured and text-heavy and hence does not conform to a well-defined data model. However, it's possible that the data generated contains content with some sort of description. This level of activity results in a massive increase in data.

How Is It Being Used?

Hadoop; APACHE STORM; Spark

Traditional Data Warehouse:

> Multiple data sources → Extract Transform Load→Data warehouse

NoSQL Data Storage in the Cloud:

> Beyond relational databases,
>
> Organize data to suit the problem and objectives

Sentiment Analysis

> Predict Customer Behaviour
>
> Personalized Recommendations → Happy Customer → More Profit

Unstructured People Generated Data → Emerging Big Data Tools and Approaches ⬚

→ Better decisions, increased productivity, profits and societal impact

Organization-Generated Data: Structured but often siloed

How organizations produce data: Commercial Transactions; Banking/Stock Records; Credit Cards; Government Open Data; E-Commerce; Medical Records

Sale transaction data:

Detect correlated products;
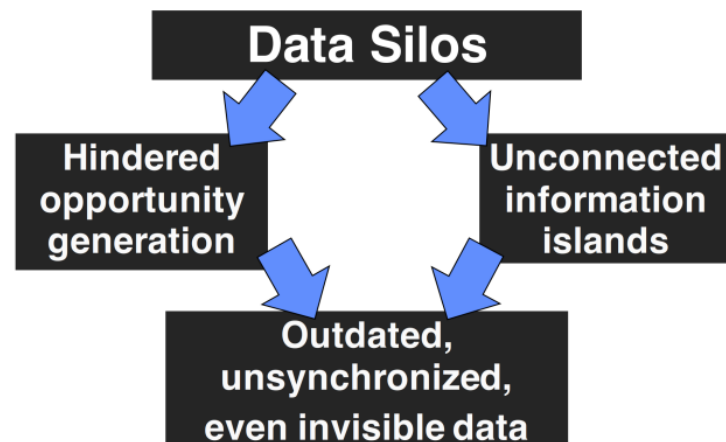
Estimate demand;

Capture fraudulent activity

Sale transaction data + open data + Analytics = Better predictions

Highly structured data: e.g. Sales Transaction Records

Challenge: Data silos within an organization!

Data Silos: → Hindered opportunity generation

→ Unconnected information islands

Organization-Generated Data: Benefits Come From Combining With Other Data Types

Real-World Examples

The UPS success

reducing each driver's route by just 1 mile save 50 million dollars

Large operational data + Optimization algorithms → Route Optimization → Savings

The Walmart success

Twitter data, local events, local weather, in-store purchases, online clicks →

Launch new products, Improve predictive analytics, Customize recommendations

Future Trends for Organizations

THE COMPANIES THAT USE ANALYTICS BEST ARE...

2 times more likely to have top-quartile financial performance

5 times more likely to make decisions much faster than competition

3 times more likely to execute decisions as intended

2 times more likely to use data very frequently when making decisions

Major benefits:

Efficient Operations; Higher Sales; Improved Safety; Customer Satisfaction; Better Profit Margins; Improved Product Placement

Data integration

Data integration, often known as knowledge, is the process of combining data from several sources to create more coherent and valuable information. The basic goal and objective of this project is to tame (or, to put it another way, manage) data and turn it into something that can be used programmatically.

A data integration process has several **stages**:

- data discovery
- access
- monitoring
- data modelling
- transformation from various sources.

Characteristics of Big Data and Dimensions of Scalability

The FOUR V's of Big Data

Volume; Variety; Velocity; Veracity

Characteristics of Big Data

Volume; Velocity; Variety; Veracity; Valence; Value

Big data features: Volume

The big data dimension of volume refers to the sheer size of large data. This volume can come from the sharing of massive datasets or the collection of many little data bits and occurrences over time. 204 million emails are sent per minute, 200,000 photos are uploaded, and 1.8 million Facebook likes are made. 1.3 million videos have been seen on YouTube.

Scalability: Variety

Now we'll discuss variety, which is a type of scalability. In this situation, scale refers to greater variety rather than the size of the data.

Big data features: Velocity

Velocity refers to the growing rate at which large amounts of data are generated and must be stored and analysed. Big data analytics has a specific purpose of processing data in real-time to match its production rate as it is generated. For example, based on your recent search, viewing, and purchase history, this feature could allow for the customisation of advertisements on the web pages you've visited.

If a company is unable to make use of data as it is generated or at the speed at which analysis is required, it risks missing out on opportunities.

(Real-time processing
Being able to keep up with the speed of big data and analyse it as it is generated can have a significant impact on human life quality.)

Big data features: Veracity

The veracity of Big Data is sometimes referred to as validity or volatility, and it refers to the data's lifetime. It highlights the relevance of data quality. Veracity is critical for making large data operational because it can be noisy, ambiguous, and full of biases. Data is worthless if it isn't correct, and big data analysis outcomes are only as good as the data being analysed. In analytics, this is referred to as 'garbage in, garbage out' (GIGO). While big data offers many options for data-driven decision-making, the evidence offered by data is only valuable if the data is of sufficient quality.

Valence

Valence is a measure of data's connectedness; the more connected data is, the greater its valence. The term valence was originally used in the field of chemistry. The core electrons and valence electrons of an atom are discussed in chemistry. Valence electrons are located in the outermost shell, have the highest energy level, and are responsible for atom bonding. Greater bonding or connectivity is indicated by a higher valence.

In the context of big data, this idea is carried over into our definition of the term valence. Data objects are frequently linked to one another, just like a city is linked to the country to which it belongs.

Value

We have discussed the five dimensions of big data. Each V represented a difficult aspect of big data, such as scale, complexity, speed, quality, and connectivity.

Although there are more dimensions that can be listed depending on the situation, we choose to identify these five as the fundamental dimensions that big data can assist you with.

However, translating all of the other aspects into a truly usable commercial value is at the heart of the big data challenge. The goal of digesting all of this enormous data in the first place is to add value to the problem.

5.Customer Persona

Customer Persona is a detailed picture of the (hypothetical) customer who is the most representative of the target segment, including demographics, personal motivations, what they value in a brand, what kind of communication they prefer, etc.

This description should be based upon your understanding of the customers via market research such as interviews, surveys, observations, internal customer data, social listening and monitoring, etc.

1. Since a customer persona is a detailed portrait of a person, demographics are a good place to start. Demographics form the skeleton of the persona such as age, gender, income, occupation, household size, etc.
2. Then, we can enhance it with psychographic indicators such as personality, self-image, lifestyle, interests, motivations, frustrations (pain points), etc.
3. The final touch is often based upon behaviours such as preferred marketing/communication channels, commonly-used platforms, purchase patterns, and preferred brands.

Week 3: CLV and RFM Update

Learning Objectives:
- Recency, Frequency, and Monetary Value (RFM) Analysis
- Customer Lifetime Value (CLV)
  - Customer Centricity
  - CLV: What and Why
  - Simple CLV Calculation

1. Customer Centricity

10 habits of customer-centric organizations
1. Continuously Listening to Customers
2. Consistently Following Up With Customers on Their Feedback
3. Acting Proactively to Anticipate Needs
4. Building Customer Empathy Into Processes and Policies
5. Respecting Customer Privacy
6. Sharing Knowledge Internally and With Customers
7. Motivating Employees to Stay Engaged
8. Acting Systematically to Improve the Customer Experience Improvements
9. Creating Accountability for Customer Experience Improvements
10. Adapting to Customer Demands and Circumstances in Real-Time

2. Customer Lifetime Value

Definition: It computes the dollar value of an individual customer relationship
　　　　It is both backward looking and forward looking
　　　　　- Compute value of past customers
　　　　　- Using that information to project forward

Expected Customer Lifetime in Months	20
Average Gross Margin per Month per Customer	50$
Average Marketing Costs per Month per Customer	0$
Average Net Margie per Month 50 x 20=1000	
Customer Lifetime Value	1000$

What is CLV use for?

Just like we use Net Present Value (NPV) to evaluate investments, we use CLV to evaluate customer relationship
CLV is the expected NPV of the cash flows from a customer relationship
CLV is defined as the discounted sum of all future customer revenue streams minus product and servicing costs and remarketing costs

› Net Margin per Netflix Customer = M – R = 50 $
› Retention Rate = r = 80 %
› Number of Customers joined Netflix in August 2021 = 100

Month	Number of Customers	Total Net Profit	Total Net Profit	Present Value of Total Net Profit
August	100	100*[M-R]	50*100=5000	
September	r*100=80	r*100*[M-R]	80*50=4000	
October	r*(r*100)=0.8*0.8*100	r^2*100*[M-R]	0.64*100*50=3200	
November	r*(r*(r*100))= 0.64*0.8*100	r^3*100*[M-R]	51.2*50=2560	

$M	Contribution per period from active customers. Contribution = Sales Price – Variable Costs
$R	Retention spending per period per active customer
r	Retention rate (Fraction of current customers retained each period)
d	Discount rate per period
Present Value of net present value is extended up to infinity	

$$CLV = (\$M - \$R) \times \left(\frac{1+d}{1+d-r}\right)$$

$$CLV = \boxed{(\$M - \$R)} \times \boxed{\left(\frac{1+d}{1+d-r}\right)}$$

Short-Term Margin

Long-Term Multiplier

Why does CLV Matters?
1. Revenues grow over time as customers buy more.
2. Cost-to-serve is lower for existing customers because both supplier and customer understand the other.
3. Higher prices are paid by existing customers than new customers.
4. Value-generating referrals are made by existing, satisfied customers through their unpaid advocacy.

Simple CLV Calculation:
Simple CLV = (Profit × Lifetime) − (Acquisition Cost) = (Revenue − Cost) × [1/(1 − Retention Rate)] − (Acquisition Cost)

The Base CLV Model

› Netflix charges 19.95$ per month.
› Variable costs are about 1.5$ per account per month.
› With marketing spending of 6$ per year, their attrition is only 0.5% per month.
› At a monthly discount rate of 1%, what is the CLV of a customer?

$M	19.95-1.5=18.45$
$R	6/12 = 0.5$
r	1 - 0.005 = 0.995
d	0.01

$$CLV = (\$M - \$R) \times \left(\frac{1+d}{1+d-r}\right)$$

CLV = [$M-$R] x [(1+d) / (1+d-r)]
CLV = [18.45 − 0.5$] x [(1+0.01) / (1+0.01-0.995)]
CLV= [17.95] x [67.333]
CLV= 1,209 $

› Netflix cuts retention spending from 6$ to 3$ per year

› They expect attrition will go up 1% per month

› Should they do it?

› To decide, we need to recalculate CLV under these new assumptions

› If the new CLV is higher, we should do it

› Otherwise, we should not

› **Netflix charges 19.95$ per month.**

› **Variable costs are about 1.5$ per account per month.**

› **With marketing spending of 3$ per year**

› **Their attrition will be 1% per month.**

› **At a monthly discount rate of 1%, what is the CLV of a customer?**

› **What is the CLV of a customer?**

$M	19.95-1.5=18.45$
$R	3/12 = 0.25$
r	1 - 0.01 = 0.99
d	0.01

CLV = [$M-$R]x[(1+d)/(1+d-r)]

CLV = [18.45 – 0.25$]x[(1+0.01)/(1+0.01-0.99)]

CLV= [18.20] x [50.5]

CLV= 919 $

What is the time horizon?
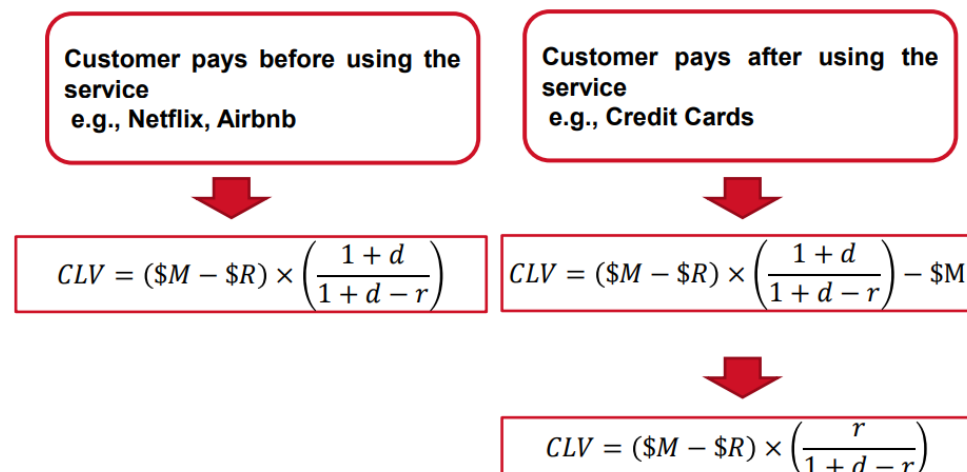What this concept of cohort and incubators and why it is important

Percent of CLV Accruing in First Five Years						
Discount Rate	Retention Rate					
	40%	50%	60%	70%	80%	90%
2%	99	97	93	85	70	47
4%	99	97	94	86	73	51
6%	99	98	94	87	76	56
8%	99	98	95	89	78	60
…						
…						
20%	100	99	97	93	87	76

Source: Farris, Paul et al (2010), Marketing Metrics, 2nd Edition, FT Press.

波动性大的没办法预测未来，所以前五年的价值就是预估的全部价值

波动性小的可以预测未来也会有稳定价值，所以前五年的价值占比较小

CLV-Initial Margin

Customer pays before using the service e.g., Netflix, Airbnb	Customer pays after using the service e.g., Credit Cards

$$CLV = (\$M - \$R) \times \left(\frac{1+d}{1+d-r}\right)$$

$$CLV = (\$M - \$R) \times \left(\frac{1+d}{1+d-r}\right) - \$M$$

$$CLV = (\$M - \$R) \times \left(\frac{r}{1+d-r}\right)$$
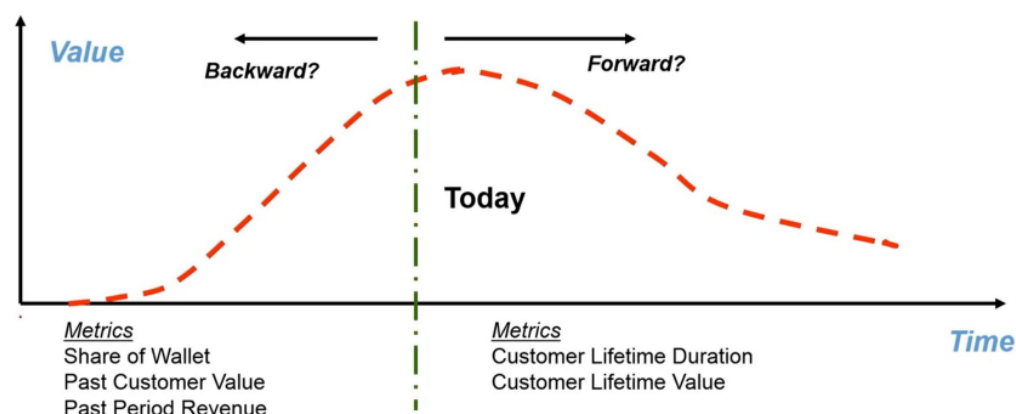
Cohort and Incubate
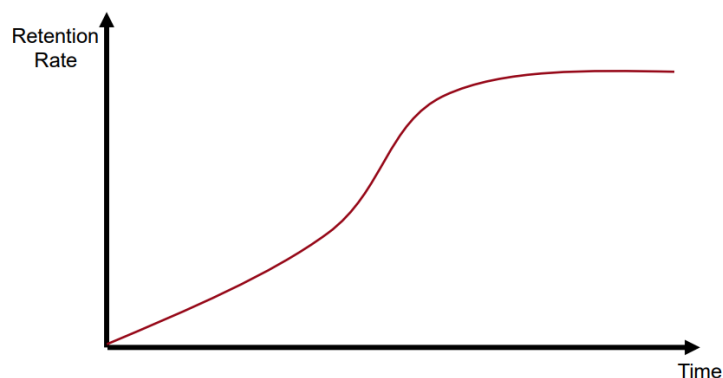
Retention Rate Depends on Time since customer acquisition

Cohort = Customers acquired at the same time period (month, quarter or year)
Since retention changes with time since acquisition, CLV calculations are better if they are done separately for each cohort
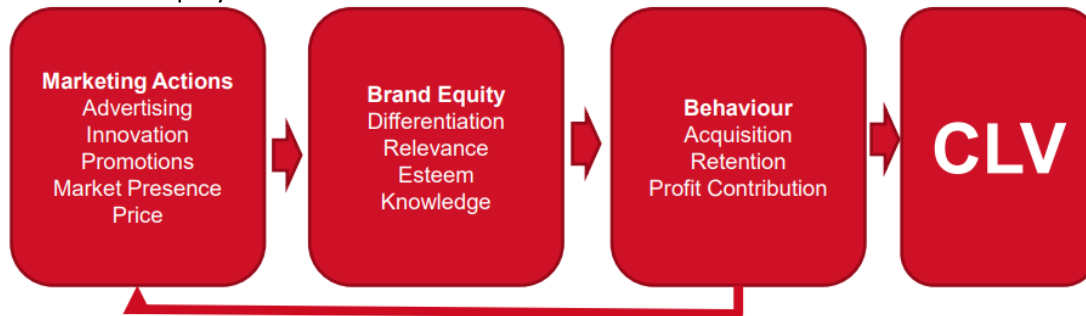
Customer Lifecycle: Where should we be looking?

Metrics
Share of Wallet
Past Customer Value
Past Period Revenue

Metrics
Customer Lifetime Duration
Customer Lifetime Value

Marketing

Link Brand Equity and CLV

Complete CLV

In the real business practice, there are more to consider when calculating CLV. For example, customer acquisition cost would be a function of the costs associated with all of the following activities:

- **Promotional elements:** All forms of advertising. Sales and trade promotion – such as in-store displays and point-of-purchase promotions. Direct marketing expenditure – such as direct mail and email campaigns. Most other promotional methods – such as events, sponsorships, online activities, and so on.
- **Online expenses:** Some social media costs – where the goal is to attract new customers. Some website expenditure should also be allocated to customer acquisition.
- **Pricing incentives:** Upfront discounts used to provide an incentive to purchase (this is also considered a sales promotion as well as a pricing tactic). Sales and service staff costs.
- **Retailing costs:** Potentially a retailer in a high traffic location could allocate a proportion of its rental costs to customer acquisition. This is because its retail location and service scape design contributes to attracting new customers.
- **Product fulfilment Supply and installation of equipment** will also be customer acquisition cost, if not immediately offset by a fee or charge. An example here could be a cable TV provider, or an office bottled water delivery firm – both of which provide the consumer with the necessary equipment.
- **Data and analysis**: Some companies will consider market research and data analysis costs to be a customer acquisition cost – particularly if it is used to identify and target potential customers (sometimes called prospects).

3. RFM Analysis
R = time elapsed since last purchase (Recency)
F = number of purchases in a given time period (Frequency)
M = monetary value of purchases in a given time period (Monetary Value)
How to Implement RFM Analysis Used in Customer Segmentation

Customer ID	Recency (Day)	Frequency (Number)	Monetary (Total)
1	4	6	540
2	6	11	940
3	46	1	35
4	23	3	65
5	15	4	179
6	32	2	56
7	7	3	140
8	50	1	950
9	34	15	2630
10	10	5	191
11	3	8	845
12	1	10	1510
13	27	3	54
14	18	2	40
15	15	1	25

Customer ID	Recency (Day)	Rank	R Score
12	1	1	5
11	3	2	5
1	4	3	5
2	6	4	4
7	7	5	4
10	10	6	4
5	15	7	3
15	15	8	3
14	18	9	3
4	23	10	2
13	27	11	2
6	32	12	2
9	34	13	1
3	46	14	1
8	50	15	1

Customer ID	Frequency (Number)	Rank	F Score
9	15	1	5
2	11	2	5
12	10	3	5
11	8	4	4
1	6	5	4
10	5	6	4
5	4	7	3
4	3	8	3
7	3	9	3
13	3	10	2
6	2	11	2
14	2	12	2
3	1	13	1
8	1	14	1
15	1	15	1

Customer ID	Monetary (Total)	Rank	M Score
9	2630	1	5
12	1510	2	5
8	950	3	5
2	940	4	4
11	845	5	4
1	540	6	4
10	191	7	3
5	179	8	3
7	140	9	3
4	65	10	2
6	56	11	2
13	54	12	2
14	40	13	1
3	35	14	1
15	25	15	1

Customer ID	R Score	F Score	M Score	RFM Score
1	5	4	4	544
2	4	5	4	454
3	1	1	1	111
4	2	3	2	232
5	3	3	3	333
6	2	2	2	222
7	4	3	3	433
8	1	1	5	115
9	1	5	5	155
10	4	4	3	443
11	5	4	4	544
12	5	5	5	555
13	2	2	2	222
14	3	2	1	321
15	3	1	1	311

Analysing RFM Segmentation
Champions
Potential Loyalists
New Customers
At Risk Customers
Can't Lose Them

RFM analysis is based on a simple technique.
RFM (Recency, Frequency, Monetary) analysis is a proven marketing model for behaviour based customer segmentation. It groups customers based on their transaction history – how recently, how often, and how much they bought. RFM helps divide customers into various categories or clusters to identify customers who are more likely to respond to promotions and future personalization services.

Using RFM for Customer Segmentation and Strategies
Finally, you can segment your customer database into different groups based on this Recency – Frequency – Monetary score.
Think about what percentage of your existing customers would be in each of these segments. And evaluate how effective the recommended marketing action can be for your business.

RFM segmentation readily answers these questions for your business…
Who are my best customers?
Which customers are on the verge of churning?
Who has the potential to be converted into more profitable customers?
Who are lost customers that you don't need to pay much attention to?
Which customers must you retain?
Who are your loyal customers?
Which group of customers is most likely to respond to your current campaign?

Some advantages of RFM
RFM is useful for different types of businesses – online, retail, direct marketing, subscriptions, non-profits…
You get to know different customer segments and can identify your best customers.
RFM helps craft highly targeted marketing campaigns
It aids customer relationship marketing and customer loyalty
Combine it with other tools to get detailed customer analytics and customer insights
RFM reduces marketing costs due to optimizing targeting
It decreases negative reactions from customers due to controlled targeting

Some limitations of RFM
It may not be useful when most customers are just one-time purchasers
When you sell just one product and that too only once, RFM may not be suitable

Without a software/tool, calculating RFM scores and segments can be complex
Sending too many campaigns to one particular segment can upset customers

Week 4: CEM, NPS and Customer Journey Mapping

Learning Objectives:
- Customer Experience Management (CEM)
  - What CEM is and Why CEM is important
  - How to measure success in CEM
- Customer Journey Mapping (CJM)
  - Defining Customer Journey Mapping
  - Key considerations when mapping the customer journey
  - Step-by-step procedure for Customer Journey Mapping
- Net Promoter Score (NPS)
  - What and Why NPS
  - How to calculate NPS
  - Explain how NPS works as a system in an organization
- Business Model Maturity Index

- o Use the Big Data Business Model Maturity Index as a framework for organizations to measure how effective they are at leveraging data and analytics to power their business models
- o Discuss the objectives and characteristics of each of the five phases of the Big Data Business Model Maturity Index: Business Monitoring, Business Insights, Business Optimization, Data Monetization, and Business Metamorphosis
- o Discuss how the economics of big data and the four big data value drivers can enable organizations to cross the analytics chasm and advance past the Business Monitoring phase into the Business Insights and Business Optimization phases
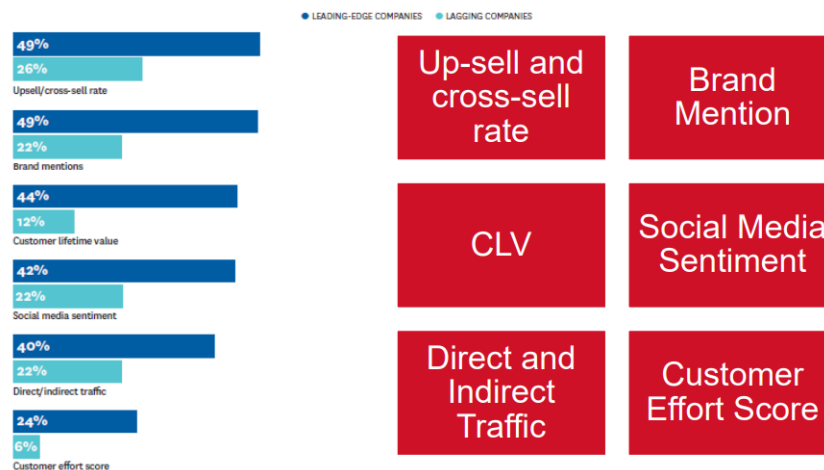
1.Customer Experience Management (CEM)

Definition

• Practice of designing and reacting to customer interactions to meet or exceed customer expectations and, thus, increase customer satisfaction, loyalty and advocacy during the duration of their relationship.

• It is measured by the individual's experience during all points of contact (touchpoints) against the individual's expectations.

Measuring Success in CEM

Leading-Edge Companies Measure Success Differently
Please indicate whether or not your organization uses the following metrics to measure the success of customer experience efforts.

● LEADING-EDGE COMPANIES   ● LAGGING COMPANIES

49% — Upsell/cross-sell rate	
26%	
49% — Brand mentions	
22%	
44% — Customer lifetime value	
12%	
42% — Social media sentiment	
22%	
40% — Direct/indirect traffic	
22%	
24% — Customer effort score	
6%	

Up-sell and cross-sell rate	Brand Mention
CLV	Social Media Sentiment
Direct and Indirect Traffic	Customer Effort Score

Customer Effort Score (CES)

* The company made it easy for me to handle my issue.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree

* How much effort did you personally have to put forth to handle your request?

Very low effort	Low effort	Neutral	High effort	Very high effort

CEM: Satisfying Left and Right Brain

Left Brain (rational):

- Deliver as promised

- My needs are met

- I get value for money

- Interactions are easy

- Consistently exceed my expectations

Right Brain (emotional)

- I trust them

- They treat me with respect

- They care about myself

- I feel connected with them

- I am proud to be a customer

The Six Pillars of Experience Excellence (from KPMG)

Pillar 1 - Integrity: Acting with integrity and engendering trust.
- o Do the right thing for me personally.
- o Be seen to be doing the right things for customers in general
- o Stand for something more than profit
- o Have not been associated with negative stories on the TV, internet or in the newspapers
- o Are recommended by my friends or family
- o Take corporate social responsibility seriously
- o Contribute to my local community

Pillar 2 - Resolution: Turning a poor experience into a great one.
- o Own the resolution and fix it with urgency
- o Keep me fully informed of issue resolution progress
- o Offer a warm and sincere apology
- o Go the extra mile if required
- o Provide a temporary solution while trying to resolve the problem
- o Assume my innocence

Pillar 3 - Expectations: Managing, meeting and exceeding customer expectations.
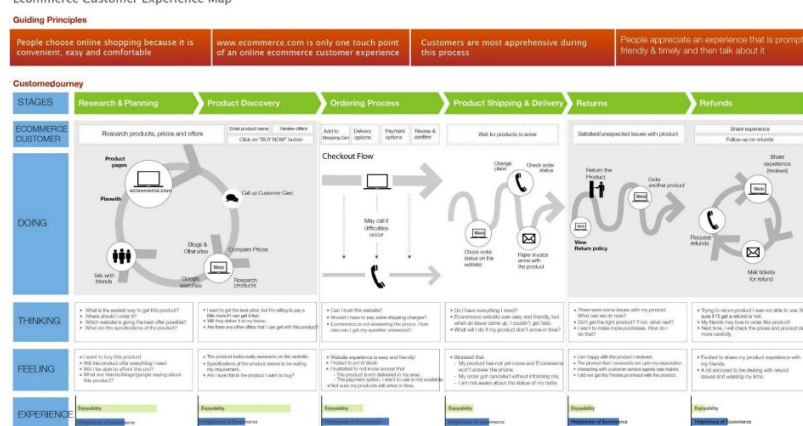- o Do what they said they would
- o Provide a consistent service every time I use them
- o Use plain English – no jargon
- o Keep me informed as to what is happening when I have a query or place an order.
- o Check whether I'm happy with their product/services
- o Set my expectations accurately

Pillar 4 - Empathy: Achieving an understanding of the customer's circumstances to drive deep rapport
- o Understand my particular situation
- o Explain things in a way I can easily understand
- o Demonstrate they care
- o Acknowledge how I'm feeling and act appropriately
- o Invest time to understand me and my needs
- o Provide the right emotional responses for my situation
- o Are you willing to bend the rules to help me out

Pillar 5 - Personalization: Using individualised attention to drive emotional connection.
- o Offer products or services relevant to me
- o Understand my specific personal needs
- o Make me feel valued
- o Treat me as an individual

o Know who I am from my customer history when I'm dealing with them
o Make me feel in control
o Show an interest in me

Pillar 6 - Time and Effort: Minimising customer effort and creating frictionless processes.
o Provide information/answers I need when I contact them
o Minimise waiting times
o Manage queues/call waiting effectively
o Explain exactly what I need to do next at all points in my task or purchase
o Make my time investment feel worth it.
o Offer alternative ways to complete my task or make my purchase.

2.Customer Journey Mapping

Customer journey mapping is the process of creating a customer journey map, a visual story of your customers' interactions with your brand. This exercise helps businesses step into their customer's shoes and see their business from the customer's perspective.

Customer touchpoints are where customers interact with your brand, product, service, etc. Developing an understanding of each touchpoint means that you can design better customer experiences. This understanding can also be used to enhance user and customer journey mapping exercises.

❑ Who is the customer to map (customer persona)

❑ When and Where Do Important Interactions take place (touchpoints)?

    ➢ Critical Moments

    ➢ Communication Channels

❑ What need(s) would customers have at each touchpoint?

❑ What emotions (both positive and negative) could customers experience at each touch point?

❑ How can we better meet customer needs and improve customer emotions?

❑ Do we deliver a customer experience that our competitors can't match?

Customer Journey Mapping: E-Commerce

Steps

1. Select a specific customer to map

2. Map a customer's step-by-step experiences (major stages)

3. Map touchpoints and channel

4. Add customer's needs and emotions

5. Identify problems & opportunities and set a priority

3.Net Promoter Score (NPS)
• NPS: a popular way of measuring customer loyalty through understanding word-of-mouth marketing
• "How likely are you to recommend [product or service] to a friend or colleague?"

Net Promoter or Net Promoter Score is a management tool that can be used to gauge the loyalty of a firm's customer relationships. The Net Promoter Score (NPS) is a single, easy-to-understand metric that predicts overall company growth and customer lifetime value. But NPS can work as a system to improve the overall performance of a company.
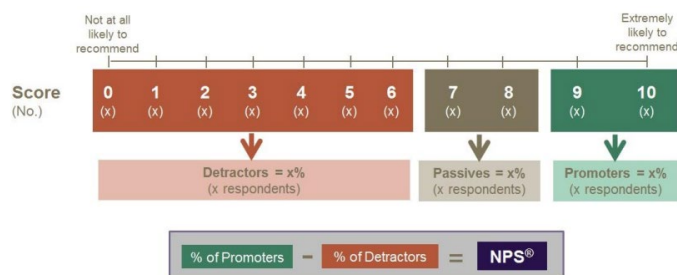
NPS as Primary Customer Outcomes
NPS: Vary by Industry and Company

How NPS as a System Works?
Step 1 – Sort Customers
Step 2 – Close the Loop
Step 3 – Make Loyalty a Top Priority

Some More Thoughts on NPS
1. The value of NPS is its simplicity
2. Meeting expectations for brand experience is linked to a higher NPS
3. NPS also lacks diagnostic value – a low NPS score would require further investigation to find out where the problem lies (root causes).
4. NPS underplays negative attitudes and under-represents noncustomers.
5. Cannot have NPS as the only measure.
6. Tying NPS to executive bonuses drives focus on customer experience.

4. ACURA Strategy
ACURA is an acronym for Acquisition, Cross-selling, Up-selling, Retention and Advocacy.
Acquisition Strategies

- Acquisition strategies are necessary to feed the company's pipeline with prospects, since companies can lose 20-40 percent or more of their customers each year.
- Mass media is still widely used to acquire new customers.
- When developing acquisition strategies, it is important to reduce adverse selection: i.e., targeting individuals who will have no interest in your offering or, if they do, will not qualify. Adverse selection wastes time and money and leads to negative word of mouth.
- Companies should develop their acquisition programs based on solid qualitative and quantitative marketing research. Key attributes in successful acquisition programs include proper timing and elimination of switching costs.
- Prospects tend to be enticed by offers for which they enjoy a relative advantage. This is referred to as the idiosyncratic-fit heuristic. Acquisition programs that appear difficult to qualify for (but for which recipients have the requisite characteristics) will result in more acceptances than programs that everyone would appear to qualify for.

- Lead Generation: the generation of consumer interest or inquiry into products or services of a business.
    - Mass media (e.g., advertising)
    - Direct marketing (e.g., telemarketing, personal contact)
    - Search engine optimization (SEO) and referrals
- Lead Conversion Rate
    - Remove switching cost if necessary
    - Strong intro Incentives with reduced costs
- Prospect Profiling and Message/Product Customization
- Sales Training

Retention Strategies
Companies should retain only customers whose retention will be profitable.
Retention strategies in general can be based on rewards, bonds, or service structure.
Bonds can be programmatic, such as rewards programs, or humanistic—preferential treatment given to customers by highly trained personnel.
Retention strategies are then itemized with examples of how each can be employed:

- Providing and attaining intimacy through conversational marketing, dialogue marketing, and event-based marketing. Conversational marketing technology provides a real-time view of the customer, a catalog of marketing offers, and a recommendation engine that delivers personalized offers based on an individual's profile and behavior. Dialogue marketing generates a running dialogue based on every customer interaction with the business or special events, such as a change in the customer's life stage.
- Moments of truth occur when a company can assist in a customer emergency or customer complaint.
- Preferential treatment is the customer's perception of how much better they are treated than the company's other customers.
- Rewarding occurs when customers are offered tangible benefits in return for loyalty.
- Personalization is a consumer's perception about how personally they are treated; for example, Tesco delivers promotions to individuals based on their lifestyles or profiles. Personal shopping assistants are another example. Personalization can also be delivered over a company's website when it configures itself based on a customer's personal preferences.

- Customization occurs when a company can adjust its 4 P's to suit a particular individual customer or can modify its website to suit the needs of a customer.
- Managing migration (the change in customer value over time) can be done when a company recognizes a customer is becoming less profitable and generates promotions to reverse the downward trend.
- Conversion of transactional buyers to relational buyers occurs through programs such as frequent shopper or rewards programs.

- Moment of Truth
  - a customer needs medical attention and contacts American Express. The company finds the nearest hospital, sends a map with its location to the customer's smart phone, and reminds the customer that he can charge the bill on his AmEx card.
- Preferential Treatment
  - How much better they are treated than the company's other customers.
- Reward and Loyalty Program

- Data-driven customer retention tactics
  - Focus on initial customer reactions
  - Offer custom products and services
  - Create engaging educational content
  - Interact with customers on their preferred channels
  - Offer personalized incentives to keep customers coming back
  - Hire quality service representatives
  - Implement customer feedback
  - Protect your customers' data

Cross-selling Strategies
- Cross-selling occurs when additional products or services are sold. For example, Amazon recommends books based on a customer's previous book purchases.
- Cross-selling is most effective when you stick to extra items the customer would have a genuine interest in anyway - items that are well-connected to their original choice.
- Social influence: 'Other people who bought this also bought.' In 2006, Amazon reported that a whopping 35% of the year's sales came from cross-sells. Their introduction of 'Customers who bought this item also bought' and 'Frequently bought together' is nothing short of cross-selling genius.
- Consider the basket size and the relative value of the add-on item (cross-selling item). The value of an add-on sale should not increase the overall order by more than 25 per cent. For example, if the original order is $100, you should be cautious in your attempts to exceed that order by $25. In addition, the value of the initial order is big (e.g., $500) it might be easier to cross-sell the second item at the value of $50. But doing so is quite difficult when the value of the initial item is $100.
- Understand the customer habit. Customers could differ by the intrinsic inclination to accept the cross-sell offer.
- "Pushy" cross-selling attempts can cause the customers to discard the original sale.
- Reasonable discounts with bundling or other benefits (e.g., reduced shipping costs, discounted second buy) helps.
- Build strategic partnership with relevant other businesses (e.g., credit card, airlines, banking, travel agency, car rentals, insurance, etc). Don't need to cross-

sell internally. External cross-selling could be more profitable while giving more choices for your customers.

- Cross-selling occurs when additional products or services are sold.
  - Amazon recommends products based on a customer's previous purchases.
  - Well connected to the original choice
- Social influence: 'Other people who bought this also bought.'
  - 35% of Amazon's sales from cross-sells
  - User reviews and recommendations are effective
- Bundling and discount offers
- Strategic partnership with other businesses

Up-selling Strategies
- Upselling occurs when a company convinces a customer of the value of purchasing a better quality offering; a suite instead of a room, for example.
- Value for money still matters: Bundling can use for both up-selling and cross-selling. Justify the enhanced value for money from up-selling (offer discounts: e.g., value meals at fast food restaurants, car + home insurance).
- Selling more expensive products or services is not the only way: Quantity-based up-selling can product more profits. You increase the discount by how many products they order. If they order 3 it's a 10% discount, if they order 5 it's a 15% discount.
- Remove risk for customers: Offer incentives such as a money-back guarantee or a free trial to prove that the product you are up-selling is worth the additional value. You can even offer a bonus such as a free download or free shipping when they purchase an additional item or spend over a certain amount.
- Anticipate customer needs and know your products: Your products or services should always serve as a solution to a problem that your customers have. Up-sell products that can further improve the results the customer is trying to achieve. Go beyond a simple transactional relationship with your customers. Whether it's through social media, email marketing or catalogs – track their purchases, map out their interests and learn what makes them tick.

- Upselling occurs when a company convinces a customer of the value of purchasing a better-quality offering
  - a suite instead of a standard room
- Value for money still matters
  - Enhanced value for money from up-selling (offer discounts: e.g., value meals at fast food restaurants, car + home insurance).
- Selling more expensive products or services is not the only way
  - Quantity-based up-selling can produce more profits
- Remove risk for customers
  - Offer incentives such as a money-back guarantee or a free trial to prove that the product you are up-selling is worth the additional value

Advocacy Strategies
- Short-Term Buzz: If your objective is to get a whole lot of people talking about your product launch or latest initiative, then look to 'celebrity' endorsements or influencer coverage that would engineer online conversations to get your message across.
- Long-Term Love: Creating a sustainable network of advocates – customers who support your brand and will talk about it to their friends – will take more effort to build, but will guarantee long-term commitment from your fans. These program can empower super-fans or even your employees or partners to talk on your brands' behalf as ambassadors.

- Create an open feedback loop: Companies often ask for feedback – but actually acting on it is another thing. Brands that ensure that comments, complaints and suggestions becomes part of a loop that actually feeds back into the company's products and services make customers feel more involved and appreciated – and will drive better product development.
- Organic vs. Paid Advocacy: Paid advocacy (through power bloggers, expert reviewers, magazine and other forms of media) is less sustainable. If your goal is to create long-term buzz, don't pay. But pay for the platform or events in which organic lovers can advocate for your brands.
- Getting customers to talk about your company and products. Meaning happy shoppers spread the word about your brand on social media or direct word-of-mouth, leading to people in their network potentially becoming customers too.

- Short-Term Buzz
  - to get a whole lot of people talking about your product launch or latest initiative, then look to 'celebrity' endorsements or influencer coverage that would engineer online conversations to get your message across
- Long-Term Love
  - Creating a sustainable network of advocates -customers who support your brand and will talk about it to their friends - will take more effort to build, but will guarantee long-term commitment from your fans

Win-Back Strategies

Win-back is the process of revitalizing relationships with customers who have defected or are about to defect. Winning back such customers is in fact easier than successfully closing the sale on a brand new customer. Win-back strategies should be developed after analyzing defections; i.e., researching the reasons why customers leave in the first place. In the context of online marketing, you might use the "exit survey" in order to identity the root cause of the defection.

ACURA Strategies Examples

Dialogue marketing—a bank's computer reviewing data in its data bank notices a customer's child is entering college and generates a conversation regarding college loans. This could be done over the phone, via e-mail, etc.

Moments of truth—a customer needs medical attention and contacts American Express. The company finds the nearest hospital, sends a map with its location to the customer's smart phone, and reminds the customer that he can charge the bill on his AmEx card.

Preferential treatment—Burberry invites its best customers to be its guests at a runway fashion show.

Rewarding—a chemical company's best distributors are invited to fish above the Arctic Circle for one week as a reward for buying the requisite amount of chemicals for that year.

Personalization—a tailor keeps specific measurements for its best customers and recommends fabric based on their location.

Customization—a company adopts payment plans to suit the cash flow of a particular customer

Cross-selling—special training seminars for companies buying a particular mix of computer products and services

Upselling—selling an expensive Audi 8 to a customer who initially came into the showroom wanting to buy a less expensive Audi 6.

Managing migration—American Express notices a customer's charges have decreased from $10,000 per month to $1,800 per month. It offers triple miles for each dollar spent over the next six months instead of single miles to generate more habitual usage.

Brand-building tools—Disney staff provides such a warm experience for their kids, a couple decides to stay at a Disney Hotel the next time in Orlando.

Moments of Truth

Moments of truth (MOT) in marketing, is the moment when a customer/user interacts with a brand, product or service to form or change an impression about that particular brand, product or service. Marketers strive to use moments of truth to create positive, customer-centric outcomes. MOT is particularly important when consumers are unable to evaluate the actual quality of products/services at the time of the purchase and the interaction with products/services is with high impact and less frequency (e.g., health insurance, car insurance, vacation package, etc.). See the following video about how a car insurance company uses a mobile app to improve the customer experience at MOT.

Word-of-Mouth (WOM) or advocacy marketing strategies have received growing attentions from marketers and academics.

Big data Business Model Maturity Index

Business users and business management, and in particular organisations that wish to survive beyond their immediate term, can no longer relinquish their control of data and analytics to IT departments. The discussions regarding big data are currently required to focus on methods that can contribute to helping organisations connect new sources of customers, products and their relative operational data to advanced analytics (data science). Organisations must come to understand that they do not require a big data strategy as much as a business strategy that incorporates big data.

The big data Business Model Maturity Index is an important basic concept supporting the big data implementation. It's necessary to lay a strong foundation in how organisations can use the big data Business Model Maturity Index in order to answer the fundamental big data business question of 'How effective is my organisation at use of data analytics in our business model?'

Phase 1: Business monitoring. In the phase of Business monitoring, organisations apply technique and tools of data warehousing and Business intelligence a with a view to monitoring their business performance (also known as Business Performance Management).

Phase 2: Business insights. In this phase, organisations largely broaden their data assets by gathering all of their detailed transactional and operational data and coupling it with new sources of both internal data and external data. Then, they apply predictive analytics to discover the hidden customer, product, and operational insights through these data sources.

Phase 3: Business optimisation. In the phase of Business optimisation phase, organisations apply prescriptive analytics to make key business processes optimal and build on the insights uncovered throughout the phase of Business insights. Here, organisations relay the analytic results (like recommendations, scores, rules) to frontline employees and business managers to help them optimize the targeted business processes with the help of better decision-making.

Phase 4: Data monetisation. Organisations aim to generate new revenue streams during the Data monetisation phase. This includes selling data or insights into new markets, integrating analytical insights into products and services, and/or repackaging customer, product, and operational insights to create entirely new products and services that could help them enter new markets and target new customers or audiences.

Phase 5: Business metamorphosis. The Big Data Business Model Maturity Index's holy grail is when a company uses data, analytics, and insights to transform its business. This necessitates a significant change in the organisation's core business model, which is driven by the Big Data Business Model Maturity Index's findings.

Organisations that go from selling things to offering business-as-a-service are one example.

Phase 1: Business monitoring

The business monitoring phase is when companies use Business Intelligence (BI) and data warehousing technologies to track and monitor their ongoing business performance.

In this phase, also known as business performance management, organisations create reports and dashboards for monitoring the current state of the business, flagging under- and/or over-performance areas, and alerting key business stakeholders with pertinent information when special 'out of bound' performance situations occur. For most big data journeys, the business monitoring phase is a fantastic place to start.

Companies have spent significant time, money, and effort identifying and documenting their critical business processes as part of their BI and data warehousing activities, those that contribute to making their organisations unique and successful.

They have painstakingly constructed a supporting data model and data architecture, as well as many reports, dashboards, and alerts of the important actions and indicators that support the business process.

Many excellent assets have already been generated, and these assets serve as a springboard for our big data adventure. Moving beyond the business monitoring phase is, unfortunately, a big difficulty for many firms.

Some disadvantages

Decades of BI and data warehouse efforts have created resistance to the 'think differently' approach that is required to effectively harness big data for its economic value.

Not only that, but the large financial payback isn't fully recognised and understood until the organisation moves from the business insights phase to the business optimisation phase.

Phase 2: Business insights

The business insights phase blends predictive analytics with the organisation's growing amount of internal and external, structured and unstructured data to uncover hidden customer, product, and operational insights.

This entails finding uncommon (or out-of-the-ordinary) behaviours, trends, and patterns in the data that need further research. In this phase, enterprises must take advantage of big data's economics, which indicates that big data technologies are 20 to 50 times cheaper to store, manage, and analyse data than traditional data warehouses.

The economics of big data empowers organisations to rethink how they collect, integrate, manage, analyse, and act on data, and they lay the groundwork for moving beyond the business monitoring phase and into the analytics gap.

Big data economics enables four new capabilities that will help the organisation overcome the analytics chasm and go from the business monitoring phase to the Business insights phase.

Big data value drivers
Drive 1 – Access to all the organisation's transactional and operational data
In the age of big data, we must go beyond the data warehouse's summarised and aggregated data and be prepared to store and analyse the organisation's whole history of detailed transactional and operational data.

This could involve taking into account 25 years of *specific point-of-sale (POS)* transactional data, rather than just the 25 months of *aggregated POS data* in the data warehouse. Consider the business potential of being able to analyse each POS transaction at the individual customer level for the past 15 to 25 years (thanks to loyalty programs).

Grocers would be able to detect when individual customers are experiencing financial difficulties by looking at their likely changes in purchase behaviour and product preferences (i.e., buying lower-quality products, replacing branded products with private label products, and increasing the use of discounts and coupons), something that would otherwise go unnoticed if only looking at aggregated data stored in the data warehouse. Organisations can collect, analyse, and act on the entire history of every purchase occasion made by each customer, including the combinations, prices of their purchases, coupons they may have used, what and when they purchased on sale, which stores they frequented at what time of day and day of week, as well as the weather and or local economic conditions of their purchases, using big data. You can uncover insights about each customer and their product behaviour, tendencies, propensities, preferences, and usage patterns by analysing customers' transactional and operational data, and organisations can take action based on these insights.

It's quite difficult to get actionable data at the aggregated level of store, zip code, or consumer behavioural categories.

Driver 2 – Access to unstructured internal and external data

Unstructured data is disliked by data warehouses, which prefer structured data. Because data warehouses are based on relational database management systems (RDMBS), they require data to be organised in rows and columns.

As a result, corporations and their business users have been educated that unstructured data is essential. Big data, on the other hand, challenges this viewpoint by providing all companies with a cost-effective approach to acquire, store, manage, and analyse a wide range of *unstructured data*.

The combination of this data with enterprises' extensive structured data allows for the discovery of new consumer, product, and operational insights and perspectives.
Despite the fact that the majority of the buzz around unstructured data appears to be focused on the potential of external unstructured data (e.g., social, blogs, newsfeeds, annual reports, mobile, third-party, publicly available), many organisations' true value lies in their internal unstructured data. Consider a project for improving the predictive maintenance of wind turbines. It was found out that while technicians are at the top of turbine scaling a wind turbine to replace a ball bearing, they may also make other observations. Various unstructured comments like this, specially with the operational sensor readings, error codes, and vibrations coming off that particular wind turbine, may provide invaluable insights into the wind turbine's predictive maintenance.

Driver 3 – Exploiting realtime analytics
Newly developed big data technologies offer the technical potential to flag and act on special or rate circumstances in *real-time* for organisations. Data warehouses were traditionally batch environments and they struggled to discover and support the real-time opportunities hidden in data. A relevant example is the 'trickle feeding' of data into the data warehouse. which has long been a data warehouse challenge, because as soon as new data enters the data warehouse, all the supporting

measures, aggregate tables, and materialized views need to be updated with it, which makes real-time analysis possible.

Whereas the majority of organisations do not have numerous use cases requiring a real-time analytics environment, there are many of tem for 'right-time' analytics, in which the opportunity time is likely to vary between minutes, hours, and even days. For instance, nurses and admissions personnel in a hospital have probably four to fivr minutes to reach the likelihood of a patient being contaminated with a hospital-acquired infection (staph infection) during the admission process. Other examples include location-based services targeting shoppers who meet a certain demographic and/or behavioural category of characteristics while visiting a store.

The best solution to reveal these right-time analytic opportunities is to break down the targeted key business initiative into data events constituting the business initiative. *Discovering the data events whose prediction (within minutes, hours, maybe even a day sooner) is the next step that may provide a monetisation opportunity.*

Driver 4 – Integrating Predictive Analytics

In the end, we may apply predictive analytics to mine the value of the structured and unstructured data in *discovering areas of "unsualness" in the data*. To put it otherwise, using predictive analytics could be instrumental in revealing occurrences within the data not being considered normal behaviours or engagement patterns. These techniques may be applied by organisations to identify operational insights, products and customers in the massive volumes of structured and unstructured data.

Examples

For instance, organisations may apply basis statistics, data mining, and predictive analytics to their growing structured and unstructured data to uncover such insights as,

- marketing campaigns faring *two to three times bette*r than the average campaign performance in some markets on special days of the week
- customers showing *two to three standard abnormal deviations* in their purchase patterns for certain product categories under specific weather conditions
- suppliers whose components are operating *not within the upper or lower limits of a control chart* under extreme cold weather conditions.

Some disadvantages

The phase of Business insights is the *most difficult* of the big data Business Model Maturity Index because of requiring organisations to *'think differently'* about the way they should approach data and analytics. Application of rules, techniques, and approaches already tended to work in Business Intelligence and data warehouses to the world of big data would not be possible. It is known as the true moment of 'crossing the analytics chasm'.

For instance, business intelligence analysts have learnt to 'slice and dice' the data to discover insights hidden in the data. This approach had a satisfactory performance in dealing with gigabytes of data, five to nine dimensions, and 15 to 25 metrics. However, this technique is not functional enough when dealing with petabytes of data, 40 to 60 dimensions, and hundreds of metrics.

Furthermore, before the organisation reaches the phase of Business optimisation, we may not have learnt much about the return on investment (ROI) or big data financial payback. It is thus essential to focus your big data journey on a key business initiative; a goal the business is striving to achieve over 9 to 12 months.

Concentration on a business initiative may create the necessary financial and organisational motivation to go through the business insights phase and realise the financial return and payback created in the phase of business optimisation.

Phase 3: Business optimisation
The phase of business optimisation phase is the stage of the big data Business Model Maturity Index where organisations develop the ***predictive analytics*** (forecasting what may happen) and ***prescriptive analytics*** (recommending preventive actions) required for the targeted key business process optimisation.

This phase relies on the analytic insights discovered during the phase of business insights. It develops predictive and prescriptive analytic models through insights passing S.A.M. criteria. One client has identified it as the phase of ***'tell me what I need to do'***.

While this part of the maturity index may be considered by many to be where organisations turn the process of optimisation over to the machines, it is the phase of business optimisation that delivers the actionable insights (e.g. recommendations, scores, rules) to frontline employees and managers in addition to helping them decide about supporting the targeted business process in a better manner.

Examples
Examples include delivering recommendations on:
- resource scheduling to store managers based on such factors as purchase history, buying behaviours, seasonality, and local weather and events.
- distribution and inventory to logistic managers based on current and predicted buying patterns, combined with local traffic, demographic, weather, and events data.
- product pricing to product managers based on the competitive prices, current buying patterns, inventory levels, and product interest insights gleaned from social media data.
- financial investment to financial planners and agents based on a customer's current financial asset mix, financial goals, market and economic conditions, risk tolerance and savings objectives.
- maintenance, scheduling, and inventory to wind turbine technicians based on error codes, sensor readings, vibration readings, and recent comments by the technician in previous maintenance.

Business optimisation of customer behaviour
The phase of business optimisation phase also intends to ***affect customer purchase and engagement behaviours through analysing the past purchase patterns, behaviours, and tendencies of the customer with a view to delivering relevant and actionable recommendations***.

Some examples are Amazon's 'Customers who bought this item also bought', Netflix movies, and music recommendations by Deezer. The efficacy of these recommendations is associated with the capture and analysis of the purchase, usage, and engagement activities of an individual customer in building analytic profiles which codify the customer's behaviours, preferences, propensities, tendencies, trends, patterns, affiliations, interests, passions, and associations.

In the end, the phase of business optimisation is required to integrate the product, operational prescriptive analytics and customer or recommendations back into the operational systems (e.g. call centre, sales force automation, direct marketing, procurement, logistics, inventory) as well as systems of management applications (reports, dashboards). For instance, consider an 'intelligent' store manager's dashboard, that instead of just presenting tables and charts of data, really delivers recommendations to the store manager to improve store operations.

Phase 4: Data monetisation
In the phase of data monetisation, organisations leverage the insights gathered from the phases of business insights and business optimisation phases and ***develop new revenue opportunities***.

Opportunities
Some of these new monetisation opportunities are as follows:

- Packaging data (with analytic insights) to be sold to other organisations. Consider a smartphone vendor who may capture insights about customer behaviours, product performance, and market trends to sell to advertisers, marketers, and manufacturers. Another case may be MapMyRun (purchased by Under Armour for $150M) in packaging the customer usage data from its smartphone application to generate audience and product insights to be sold to various companies, like sports apparel manufacturers, sporting goods retailers, insurance companies, and health care providers.
- Direct analytic insights integration into the products and services of an organisation for creating such 'intelligent' products or services as cars that learn the driving patterns and behaviours of a customer and adjust seats, driver controls, brake pedals, mirrors, etc. to match their driving style.
- Televisions and DVRs learning shows and movies a customer may like and search across various channels to find and record similar shows pleasing that specific customer.
- Ovens learning how a customer may like certain foods prepared and cook them in that manner automatically as well as recommendations for other foods and recipes that 'others like you' enjoy.
- Jet engines ingesting weather, elevation, wind speed, and other environmental data for adjustment to blade angles, tilt, yaw, and rotation speeds with a view to minimising fuel consumption during flight.
- Reshaping insights to develop totally new products and services to help organisations win foothold in new markets and win new customers or audiences.

To that effect, organisations may capture, analyse, and package customer, product, and operational insights across the entire market for helping channel partners to more effective markets and sell to their customers like those listed here:

- Online digital marketplaces (Yahoo, Google, eBay, Facebook) may leverage trends of the general market and other data associated with merchant performance to offer recommendations to small merchants about inventory, ordering, merchandising, marketing, and pricing.
- Financial services organisations may create a financial advisor dashboard for their agents and brokers, which would capture the customers' investment goals, current income levels, and current financial portfolio besides creating investment, risk, and asset allocation recommendations to help the brokers and agents more effectively in serving their clients.
- Retail organisations can mine customer loyalty transactions and engagements to discover customer and product insights empowering the organisation to embrace new product categories or new geographies.

Although the phase of data monetisation is clearly the big data Business Model Maturity Index phase that catches everyone's attention, an organisation is required to go through phases of business insights and business optimisation to capture the customer, product, operational, and market insights that would lay the foundation for these new opportunities of monetisation.

Business metamorphosis
The phase of *business metamorphosis* in the big data Business Model Maturity Index is required to be the ultimate goal of organisations. This maturity index phase is the point organisations seek to leverage data, analytics, and analytic insights with a view to metamorphosing or transforming the business model of the organisation. It is at this phase that organisations integrate the insights that they captured about the usage patterns of their customers, product performance behaviours, and general market trends to transform their business models. *Organisations use this business model metamorphosis to provide new services and capabilities to their customers.*

That is much easier for customers to consume and would also facilitate the organisation's engagement in higher-value and more strategic services.

Examples

An example in this regard is the data, analytics, and analytic insights that Boeing would require for shifting from selling aeroplanes to selling air miles. The data, analytics, and insights Boeing is required to discover about passengers, airlines, airports, routes, holidays, economic conditions, etc. to improve its business models, processes and people, and successfully implement this business change. Consider the business requirements needed to encourage third-party developers to build and market value-add services and products on the new business model of Boeing. Some other examples of the business metamorphosis phase are as follows:

- Energy companies shifting to the 'home energy optimisation' business by recommending when to replace appliances (based on predictive maintenance) and even which brands and models to buy based on your usage patterns, local water quality, local weather and local environmental conditions such as local water conservation efforts and energy costs.
- Retailers shifting to 'shopping optimisation' business by customising their recommendations for specific products based on the clients' current buying patterns, and recommendations for products that they may not even sell (think 'Miracle on 43rd Street').
- Airlines shifting to 'travel delight' business through which they offer discounts on air travel based on customers' travel behaviours and preferences and proactively recommend deals on hotels, sporting or musical events, rental cars, limos, shows, restaurants, and shopping in the areas in light of the areas of interest and preferences of their customers.

In spite of the big difficulties which organisations have to deal with in reaching the phase of business metamorphosis, this objective may both encourage and empower an organisational catalyst to move more aggressively along the maturity index.

Week 5: Big Data Strategy, BMS, Data Exploration

Learning Objectives:

- **Big Data Strategy**
  - demonstrate theoretical knowledge and in-depth understanding of big data strategy and business initiative
  - independently develop the big data strategy document for an organisation
  - critically evaluate and determine the business value and implementation feasibility of the data sources
  - independently develop the prioritisation matrix to prioritise the use cases based on business value and implementation feasibility.
- **Business Model Canvas**
  - Why BMC?
  - Nine Building Blocks
  - Build Business Model Canvas for Raiz Investment
- **Data Exploration for Customer Analytics**
  - demonstrate advanced theoretical knowledge and in-depth understanding of data exploration process and methods
  - independently prepare data for informed analysis

1.Big Data Strategy

Common Business Terminology

Corporate Mission. Defines what an organization is, its purpose, and, more broadly, "why it exists." The Walt Disney Company, for example, has as its corporate aim to be one of the world's foremost creators and providers of entertainment and information.

Business Strategy. Over the next two to three years, how the organization plans to achieve its purpose.

Strategic Business Initiatives. What the organization intends to do over the following 9 to 12 months to fulfill its business strategy; this usually includes business objectives, financial targets, metrics, and time frames.

Business Entities. Customers, patients, students, doctors, wind turbines, and trucks are examples of physical objects or entities that the business effort will aim to understand, forecast, and affect behavior and performance (also known as strategic nouns of the business).

Business Stakeholders. Those business functions that influence or are influenced by the strategic business effort (sales, marketing, finance, store operations, logistics, and so on).

Business Decisions. The decisions that the stakeholders in the firm must make in order to support the strategic business goal.

Big Data Use Cases. The analytic use cases that support the strategic business effort (decisions and actions).

Data. The structured and unstructured data sources that will be identified throughout the big data strategy document process, both within and external to the organization.

Big Data Strategy Document

The big data strategy document establishes a framework for connecting a company's business strategy and supporting business objectives to its big data endeavours. It can help the company break down its business strategy and initiatives into possible big data business user cases, as well as the data and analytic requirements that go with them.

why the big data strategy paper is effective:

- It is succinct.
- It can be rapidly examined by anyone in the organization who wants to verify they are working on the highest priority items because it fits on a single page.
- It's obvious.
- It is business relevant.

The following sections make up the big data strategy document:

- Business strategy
- Critical business efforts
- Critical business entities
- Critical decisions
- Financial motivators (use cases)

Key Business Initiatives

The process of creating a big data strategy document starts with determining the organization's business initiatives for the following 9 to 12 months. These initiatives outline what the company hopes to accomplish in the following nine to twelve months. This time range is crucial because it:

1- concentrates the organization's big data efforts on something that is of immediate benefit and significance to the company.

2- Instills in the organization a sense of urgency to act fast and diligently.

3- Provides a more realistic opportunity for the big data project to generate a good Return on Investment (ROI) and a financial payback in 12 months or less business initiative qualities

A business initiative is a type of project that supports a company's overall strategy and contains the following characteristics:

- Critical to short-term company and/or financial success (usually 9- to 12-month time frame)
- Cross-functional • Communicated (internally or publically) (involves more than one business function)
- Has a measurable financial goal
- Has a well-defined delivery time frame
- Delivers compelling financial or competitive advantage
- Is owned or championed by a top business executive

Consider a wireless provider with a key business initiative to reduce attrition among its most profitable customers by 20% in the next year, or a public utility with a key business initiative to improve customer satisfaction by a certain number of basis points while reducing water consumption by 20% in the next year.

Where can you find out about a company's most important business initiatives?

The primary business initiatives of a company can be found in a variety of places. If the company is publicly traded, the financial statements are a good place to start. However, a number of publicly available sources can be utilized to identify an organization's core commercial efforts for both private and non-profit organizations, including:

- Annual reports
- 10-K (filed annually)
- 10-Q (filed quarterly)
- Quarterly analyst calls
- Executive presentations and conferences
- Executive blogs
- News releases
- Social media sites
- com
- Web searches using Google, Yahoo, and Bing

Key Business Entities and Key Decisions

We attempt to identify the key business entities relevant to the targeted business effort ("raise same store sales") after defining our targeted business initiative. Strategic nouns must be the focus of the intended business endeavor. A single business endeavor is unlikely to include more than three to five business entities, or strategic nouns.

Around these business entities, we'll want to capture the behaviors, tendencies, patterns, trends, preferences, and so on that are presented at the individual entity level. To better detect fraud and optimize merchant marketing offerings, a credit card firm, for example, would want to gather a customer's individual travel and buying behaviors and preferences.

Big Data Strategy Document Process

This example will aid in the development of the template that will be used to support the big data strategy document process. Our business initiative to "raise same-store sales" has already been defined.

The following are the three business entities:

- Stores
- Local events (sporting, entertainment, social)
- Local competitors

We then discuss the analytic insights that we might want to collect for each of the business units separately. These insights specify the details we need to know about each particular business entity in order to support the strategic business endeavor.

Make a list of some analytic insights you'd like to know about each of the important company entities listed below:

- Stores
- Local events
- Local competitors

We'd like to learn more about the following for each store: In-store traffic patterns

- Nearby customer demographics
- Most popular market basket combinations
- Customer product preferences
- Weather conditions,
- Outside traffic conditions
- Local economic situation
- Local home values
- Nearby schools and colleges
- Yelp rating
- Social media sentiment, and so forth.

We'd like to know more about:

- The type and frequency of the event
- When it takes place (time of day, day of week, time of year)
- When it starts and ends
- The number of attendees
- The demographics of participants and attendees (age, gender)
- The event administrator/coordinator
- Event sponsors and so on for each individual local event.

We want to know the following about each local competitor:

- The competition's category and type
- Whether it's a chain or a mom-and-pop competitor
- Its size and proximity to us
- The sort of service provided, the cuisine served, and the price ranges Yelp ratings
- Social media sentiment
- Length of time in service
- Customer demographics, etc.

key business decisions

The next phase is to identify the essential business decisions that must be taken in support of the targeted business effort about the main business entities. In our case, this would be the business entity decisions that Chipotle and corporate management must make to support the "raise same-store sales" business strategy.

This step allows us to brainstorm decisions with our important business stakeholders, as well as those that influence or are impacted by the key business endeavor.

Here are some examples of important business decisions taken by a company:

Business entity: Stores

- Given the upcoming weekend's events, how much staffing, inventory, and ingredients do I require?
- Given the impending holidays and seasonal activities, how much manpower, inventory, and ingredients do I require?
- For Friday local company catering, how much manpower, inventory, and ingredients do I need?
- How much staffing, inventory, and ingredients do I need to meet local high school demand (in-session vs. out-of-session)?
- For the approaching high school football season, what are the best hours of operation?

Business entity: Local events

- How much more personnel will I require for which local events?
- For which local events will I require additional inventory and ingredients?
- Which local activities would I like to sponsor, and how much would it cost?
- What kind of promotions do I want to run to support local events?

Business entity: Local competitors

- What are the most efficient offers or promotions to counter competitors' attempts to steal my customers?
- What are the most effective promotions used by my competitors?
- In light of significant rival activity, what pricing and production changes do I need to make?
- Do I need to reply to any local competitors' promotions?
- Given competitors' promotional actions, what is the most effective response or promotion?

Some of the decisions will be relatively similar, which is fine because it will allow the organization to approach them from different angles.

Identifying Financial Drivers (use cases)

Then we group the decisions into "common use cases" or "common themes," which means we look for and group decisions that have comparable commercial and financial goals. The financial drivers or opportunities that "create more money" for our targeted business venture are the use cases that result.

Financial drivers Example

The following financial drivers or use cases are more likely for Mad Mex's "raise same store sales" business initiative:

- Increase foot traffic in the store (acquire new customers, increase frequency of repeat customers)
- Boost sales and profit margins from shopping bags (cross-sell complementary products, up-sell)
- Increase the amount of business events (catering, repeat catering events)
- Improve promotional efficacy (Halloween Boo-ritto, Christmas gift cards, graduation, holiday, and special event gift cards)
- Improve new product introduction effectiveness (Halloween Boo-ritto, Christmas gift cards, graduation, holiday, and special event gift cards) (seasonal, holiday)

The entire process of creating a big data strategy document was created with the goal of revealing these use cases and identifying the financial drivers that support our targeted business activity. They are the section of the big data strategy

document where we draw attention to the big data opportunities that can "Make Us More Money."

Use Cases	Potential Analytic Models
Increase Store Traffic	Store Marketing Effectiveness
	Store Layout Flow Analysis
	Store Remodelling Lift Analysis
	Store Customer Targeting
Increase Shopping Bag Revenue and Margin	In-store Merchandising Effectiveness
	Pricing Optimization
	Up-sell/Cross-sell Effectiveness
	Market Basket Analysis
Increase Number of Corporate Events	Campaign Effectiveness
	Pipeline and Sales Effectiveness
	Pricing Optimization
	Customer Lifetime Value Score
	Likelihood to Recommend Score
Improve Promotional Effectiveness	Promotional Effectiveness
	Pricing Optimization
	Market Basket Analysis
	Up-sell/Cross-sell Effectiveness
Improve New Product Introductions	Pricing Optimization
	New Product Introductions Effectiveness
	Up-sell/Cross-sell Effectiveness

Data Sources Identification and Prioritization

We begin the data and metrics envisioning process when we've defined the use cases and financial drivers. During this phase, we brainstorm data sources (whether or not you already have access to them) that may be useful in giving fresh insights for the targeted business project. We want to unleash the creative thinking of the business and IT teams by giving them the opportunity to discuss data sources that could generate new customer, product, store, campaign, and operational insights that could increase the success of the many use cases.

The following Mad Mex data sources were discovered as part of the conceptualizing process:

Point of Sales Transactions   Market Baskets   Product Master
Store Demographics   Competitive Stores Sales   Store Manager Notes
Employee Demographics   Store Manager Demographics
Consumer Comments   Weather   Traffic Patterns   Yelp
Zillow/Realtor.com   Twitter/Facebook/Instagram
Twellow/Twellowhood   Zip Code Demographics   EventBrite
MaxPreps   Mobile App

Potential data sources' business value

As you can expect, not all data sources have the same business value or implementation feasibility, and they must be evaluated based on the following criteria: • Business value that they can provide in support of each use case

The ease with which they can be acquired, cleaned, aligned, normalized, enriched, and analyzed

Assessment of implementation feasibility: (1=worst, 4=best) We're now trying to figure out how feasible it is to implement each of the various data sources. IT groups are typically in charge of this component because they have a better

understanding of the implementation issues and hazards associated with each data source. The simplicity of data collecting, data cleanliness, data accuracy, data granularity, data acquisition cost, organizational skill sets, and tool proficiency are some of the obstacles and hazards.

Data Source	Ease of acquiring	Cleanliness	Accuracy	Granularity	Cost
Point of sales transactions	4	4	4	4	4
Market baskets	4	4	4	4	4

Prioritization Matrix

The final phase in the big data strategy paper process would be to go through a use case prioritization process with the business and IT stakeholders.

We identified the use cases that support the organization's major business initiative using the big data strategy document, brainstormed additional data sources, and assessed the applicability of those data sources using a business value and implementation feasibility assessment. Currently, we are prioritizing the use cases based on their proportional business value and practicality of implementation over the next 9 to 12 months.

We must remember to base this process on the next 9-to 12-month time frame, as this will ensure that the big data project delivers immediate-term commercial value and business relevance. This would also keep the big data effort from devolving into a "boil the ocean" mission, which will certainly fail.
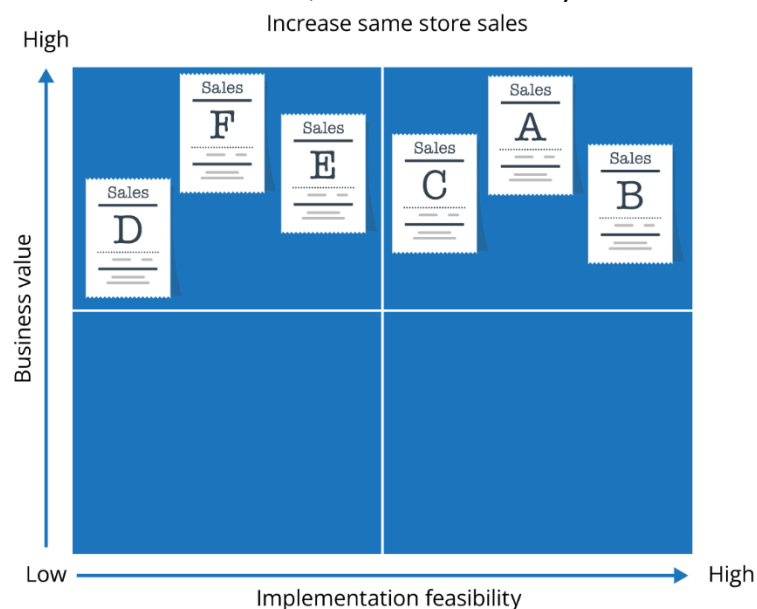
2. Business Model Canvas (BMC)

Business Model Canvas is a strategic management and lean startup template for developing new or documenting existing business models. It is a visual chart with elements describing a firm's or product's value proposition, infrastructure, customers, and finances. It assists firms in aligning their activities by illustrating potential trade-offs.

The Business Model Canvas
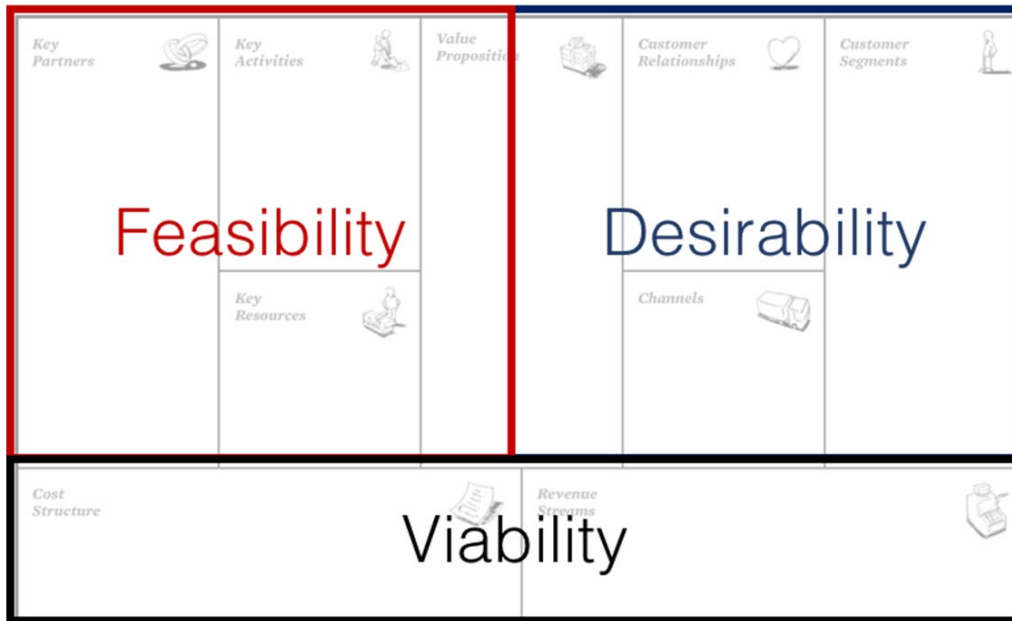
Description

Formal descriptions of the business become the nine building blocks for its activities.

- ***Customer Segments (CS, Block 1):*** To build an effective business model, a company must identify which customers it tries to serve. Various sets of customers can be segmented based on their different needs and attributes to ensure appropriate implementation of corporate strategy to meet the characteristics of selected groups of clients. The different types of customer segments include:
  - Mass Market: There is no specific segmentation for a company that follows the Mass Market element as the organization displays a wide view of potential clients. e.g. Car
  - Niche Market: Customer segmentation based on specialized needs and characteristics of its clients. e.g. Rolex
  - Segmented: A company applies additional segmentation within existing customer segment. In the segmented situation, the business may further distinguish its clients based on gender, age, and/or income.
  - Diversify: A business serves multiple customer segments with different needs and characteristics.
  - Multi-Sided Platform / Market: For a smooth day-to-day business operation, some companies will serve mutually dependent customer segments. A credit card company will provide services to credit card holders while simultaneously assisting merchants who accept those credit cards.
- ***Value Propositions (VP, Block 2):*** The collection of products and services a business offers to meet the needs of its customers. According to Osterwalder (2004), a company's value proposition is what distinguishes it from its competitors. The value proposition provides value through various elements such as newness, performance, customization, "getting

the job done", design, brand/status, price, cost reduction, risk reduction, accessibility, and convenience/usability. The value propositions may be:
  - o Quantitative – price and efficiency
  - o Qualitative – overall customer experience and outcome
- ***Channels (CH, Block 3, Connecting VP and CS):*** A company can deliver its value proposition to its targeted customers through different channels. Effective channels will distribute a company's value proposition in ways that are fast, efficient and cost-effective. An organization can reach its clients through its own channels (store front), partner channels (major distributors), or a combination of both.
- ***Customer Relationships (CR, Block 4, connecting VP and CS):*** To ensure the survival and success of any businesses, companies must identify the type of relationship they want to create with their customer segments. Various forms of customer relationships include:
  - o Personal Assistance: Assistance in a form of employee-customer interaction. Such assistance is performed during sales and/or after sales.
  - o Dedicated Personal Assistance: The most intimate and hands-on personal assistance in which a sales representative is assigned to handle all the needs and questions of a special set of clients.
  - o Self Service: The type of relationship that translates from the indirect interaction between the company and the clients. Here, an organization provides the tools needed for the customers to serve themselves easily and effectively.
  - o Automated Services: A system similar to self-service but more personalized as it has the ability to identify individual customers and their preferences. An example of this would be Amazon.com making book suggestions based on the characteristics of previous book purchases.
  - o Communities: Creating a community allows for direct interactions among different clients and the company. The community platform produces a scenario where knowledge can be shared and problems are solved between different clients.
  - o Co-creation: A personal relationship is created through the customer's direct input to the final outcome of the company's products/services.
- ***Revenue Stream (R$, Block 5):*** The way a company makes income from each customer segment. Several ways to generate a revenue stream:
  - o Asset Sale – (the most common type) Selling ownership rights to a physical good. e.g. retail corporations
  - o Usage Fee – Money generated from the use of a particular service. e.g. UPS
  - o Subscription Fees – Revenue generated by selling access to a continuous service. e.g. Netflix
  - o Lending/Leasing/Renting – Giving exclusive right to an asset for a particular period of time. e.g. Leasing a Car
  - o Licensing – Revenue generated from charging for the use of a protected intellectual property.
  - o Brokerage Fees – Revenue generated from an intermediate service between 2 parties. e.g. Broker selling a house for commission
  - o Advertising – Revenue generated from charging fees for product advertising.

- ***Key Resources (KR, Block 6):*** The resources that are necessary to create value for the customer. They are considered assets to a company that are needed to sustain and support the business. These resources could be human, financial, physical and intellectual. What are the most important assets required to make a business model work?
- ***Key Activities (KA, Block 7):*** The most important activities in executing a company's value proposition. An example for Bic, the pen manufacturer, would be creating an efficient supply chain to drive down costs. What are the most important activities a company must do to make a business model work?
    - Production
    - Problem solving: solution to individual customer problems
    - Platform/Network: platform management, service provisioning, and platform promotion
- ***Key Partnership (KP, Block 8):*** In order to optimize operations and reduce risks of a business model, organizations usually cultivate buyer-supplier relationships so they can focus on their core activity. Complementary business alliances also can be considered through joint ventures or strategic alliances between competitors or non-competitors. There are four different type of partnerships: strategic alliance between non-competitors, cooperation such as strategic alliance between competitors, joint ventures to develop new businesses, buyer-seller relationships to assure reliable supplies and platform operation. The goal of building partnership is multifaceted.
    - Optimization and economy of scale
    - Reduction of risk and uncertainty
    - Acquisition of particular resources and activities
- ***Cost Structure (C$, Block 9):*** The Cost Structure describes all costs incurred to operate a business model. This describes the most important monetary consequences while operating under different business models. What are the costs associated with key resources and key activities (blocks 6 and 7). Some companies could be more cost-driven whereas others could be more value-driven:
    - Cost-Driven – This business model focuses on minimizing all costs and having no frills. e.g. Low-cost airlines
    - Value-Driven – Less concerned with cost, this business model focuses on creating value for products and services. e.g. Louis Vuitton, Rolex

Characteristics of Cost Structures:
- Fixed Costs – Costs are unchanged across different applications. e.g. salary, rent
- Variable Costs – Costs vary depending on the amount of production of goods or services. e.g. music festivals
- Economies of Scale – Costs go down as the amount of goods are ordered or produced.
- Economies of Scope – Costs go down due to incorporating other businesses which have a direct relation to the original product.

Taken together, these nine blocks can be categorized into desirability, feasibility and viability clusters. ***Desirability*** is about understanding your customer, what motivates them, how they engage with you and what makes or breaks a purchase decision. ***Feasibility*** is about how you make everything happen behind the scenes. This includes hiring the right people, using the right tools, working with the right partners, and focusing on the right set of core activities. ***Viability*** is about the dollars: How many we earn and how many we spend. No matter your legal structure, you'll need a surplus to survive – meaning there's money left over at the end of the day

3. Data Exploration

Data terminology

Sample: A sample is a representation or illustration of an entity in your data. This is usually one of your dataset's rows. This graph depicts Bitcoin price.

Sample **Row** Record Instance Observation Example
Field Attribute Variable Feature **Column** Dimension

data type: Numeric; categorical; string; date

Numeric variables, as the name implies, are variables that take on numerical values. It is possible to measure numerical variables and categorise their values in some way. It's worth noting that a numeric variable can have either integer or continuous values. It can also have only positive numbers, only negative numbers, or both at the same time.

Let's look at a few examples of numeric variables. The height of a person is a positive, continuous quantity. An test score is a positive number that ranges from zero to one hundred percent. The number of transactions per hour is a positive statistic, although a stock price change can be positive or negative.

Categorical variables are variables that have labels, names, or categories for values instead of numbers. A variable that defines the colour of an item, such as the colour of a car, for example, can have values like red, silver, blue, white, and black. These are non-numerical values that describe an entity's quality or feature. These values can be thought of as labels or names that can be categorised. As a result, category variables are sometimes known as qualitative or nominal variables. Gender, marital status, and client type, such as teenager, adult, or senior, are examples of categorical variables. Electronics, kitchen, bathroom, and colour of an item are examples of product categories.

Each entity's distinctive characteristic is captured through a variable. As a result, a sample has a lot of variables to define it. Data from real-world applications is frequently multidimensional, which means that each sample has multiple dimensions or variables.

Data Exploration

The objective is to obtain a better knowledge of the data you'll be dealing with. If you understand the qualities of your data, you'll be able to make the most of it in any further processing and analysis. You should note that data exploration is also referred to exploratory data analysis, or simply EDA

Ways to explore data
Summary Statistics + Visualisation
Summary Statistics: mean; median; standard deviation
Something to look for
       Correlation: Provide information about relationship between variables
   Trends: Indicate general characteristics of data
       Outliers: Indicate potential problems with data
Visualisation: Histograms, line plots, scatter plots, etc.

Exploring Data through Summary Statistics

Measures of location

The central or typical value in your dataset is described by measures of location, which are summary statistics. These statistics indicate where the dataset's middle or centre is located. The terms ***mean, median, and mode*** are examples of these terms. A dataset's mean is simply the average of its values. If you ranked the values in your dataset, the median is the value in the middle. Half of the numbers in a sorted list will be less than the median, while the other half will be more. The median of two values is the mean of the two middle values when the number of data values is even. The mode is a value that appears more frequently than any other.
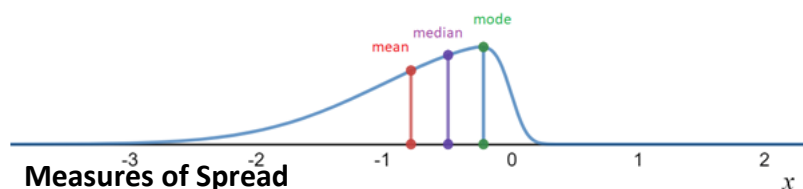
Measures of Spread

Spread measures describe how diverse or distributed your dataset is. ***Minimum, maximum, range, standard deviation, and variance*** are all common spread measures. The smallest and greatest values in your dataset are, of course, the minimum and maximum. The difference between the maximum and smallest values is range, and it indicates how evenly distributed your data is. The amount of variance in your dataset is described by the standard deviation.

A low standard deviation value indicates that your dataset's samples are near to the mean. Furthermore, a high standard deviation number indicates that the data samples are dispersed. The standard deviation and variance are inextricably linked. In fact, the variance equals the standard deviation squared. As a result, it also reveals how much the data samples deviate from the mean.
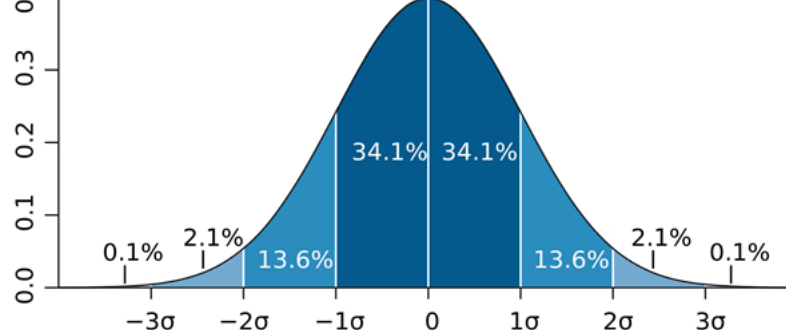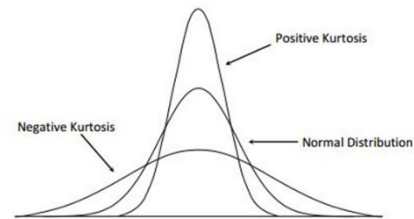

Shape measures

Shape measures describe the shape of a set of values' distributions. **Skewness and kurtosis** are two common shapes. Skewness measures how asymmetrically distributed the data values are. A skewness value of around zero suggests that the data distribution is approximately normal. A negative skewness number means the distribution is skewed to the left, as shown in the top diagram's left figure. A positive skewness score, on the other hand, suggests a right-skewed data distribution. Kurtosis is a measurement of the data distribution's tailedness, or how heavy or fat the tails are. A high kurtosis value indicates the presence of outliers by describing a distribution. A low kurtosis value, represents a distribution with shorter and lighter tails and a lower and larger centre peak, implying that outliers are absent.
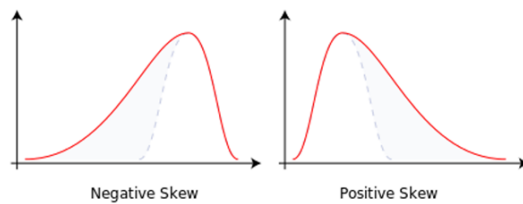
Measure of dependence

Measures of correlation determine whether there is a relationship between variables. A popular metric of reliance is a pairwise correlation

Statistics on Categorical variables

Check dimensions

Besides typical summary statistics for numerical variables and **category counts** for categorical variables, you should look at the following.

We also want to look at some extra statistics for machine learning challenges to validate the data quickly. **The number of rows and columns** in your dataset should be one of the first things you look at.

Is the number of rows equal to the number of samples expected?

Is the number of columns equal to the number of variables expected?

Check values

Another simple data validation check is to **examine the values in your dataset's initial and final few samples to see if they are appropriate**.

Do the Bitcoin price data, for example, appear to be in the correct units of measurement?

Are your variables' data types correct, such as the date field being captured as dates or timestamps?

Is it a string or a number value that you want to capture?

This will have an impact on how these fields are processed.

Check missing values

Checking for missing values is another crucial step. You must **figure out how many samples have missing values**. You should also check to see **if any variables have a significant percentage of *missing* data**. Missing value handling is a critical stage in data preparation that we shall discuss later. This information will **be extremely useful in evaluating how missing values in**

data preparation should be handled. We looked at various summary statistics that can be used for data exploration and machine learning. If you want to have a deeper grasp of your data, you should properly review the statistics.

Exploring Data through Plots

Types of Plots

To visualise your data, you can utilise numerous different types of graphs. The terms **histogram, line plot, scatter plot, bar plot, and box plot** will be discussed. These are the most popular plots, however there are plenty others

Histogram

A histogram is a visual representation of a variable's distribution. The variable's range of values is divided into a number of bins, with the number of values falling into each bin being counted. The height of each bin is determined by this.

A histogram can show a lot about a variable in your data. For example, you can typically figure out what the variable's central tendency is, which is where the majority of the values are.

You can also see the most common value of that variable's values. A histogram also reveals if the values for that variable are skewed, and if so, whether the skewness is to the left toward smaller values or to the right toward larger values.

Outliers can also be selected from the histogram, as seen in the bottom plot
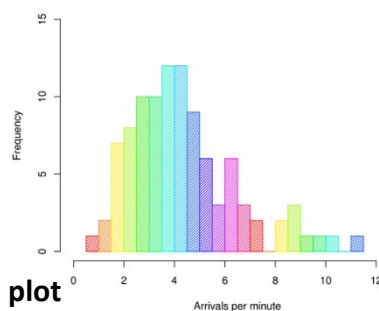

Line plot

A line plot depicts the evolution of data values across time. The Y axis depicts the values of a variable or variables, while the X axis depicts the passage of time. The data values are displayed over time on the resulting line.

A line plot can help you see trends in your data. For example, in this plot, a cyclical pattern can be seen where the numbers start high, then drop, and then rise again. Trends may also be seen in the upper-right plot, where the numbers change over time but show a general rising tendency.

On a single line plot, such as the one in the centre bottom, it's also simple to compare how different variables change over time.
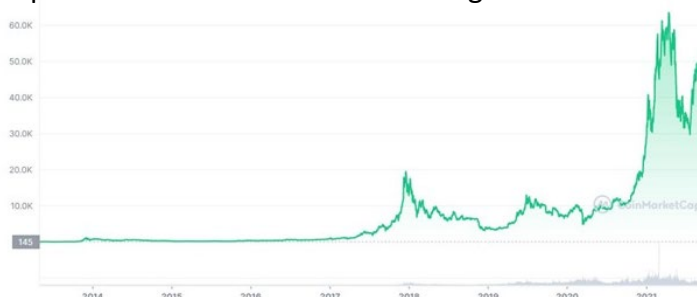
Scatter plot

The relationship between two features can be visualised using a scatter plot. On the x axis, there is only one variable. On the y axis is the other variable. The values of the

two variables aspects and the Y coordinates are used to create each sample. The plot that results depicts how one variable varies as the other is altered.
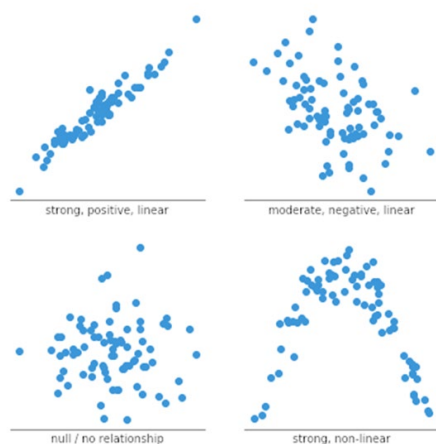
The correlation between two variables can be displayed using a scatter plot.

As illustrated in this graphic, two variables such as the day's high temperature and the day's low temperature might have a positive association.

A positive correlation shows that as the value of one variable rises, the value of the other variable rises in lockstep.

A negative correlation between two variables can be seen in the upper right scatter plot. This means that as one variable's value rises, the value of the other variable falls. Two variables can also have a non-linear relationship. So a change in one variable may or may not always correlate to a change in the other.

A lack of correlation between two variables is also possible. In this situation, you'll see something akin to randomly placed dots in the plot, showing that there's no relationship between how the two variables move in relation to one another.

Bar Plot

The distribution of categorical variables is depicted using a bar plot. Remember that a histogram is also used to examine the distribution of the variable's values. A histogram is typically used for numeric values, whereas a bar plot is typically used for categorical variables. In a bar chart, the different categories of a categorical variable is shown along the x-axis, and the count of instances for each category is displayed on the y-axis. This is an effective way to compare the different categories. For example, the most frequent category can be easily determined.
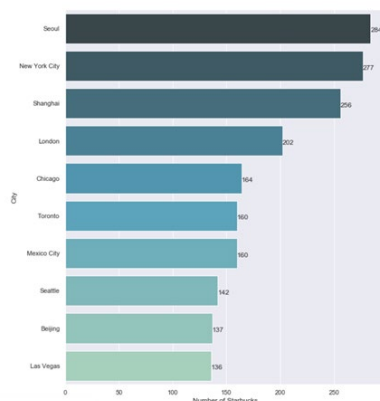
Comparing two categorical variables with a bar plot is also a good idea. A Grouped Bar Chart is the name for this form of Bar Plot. In addition, the various product factors are compared side by side. A Stacked Bar chart, can be used to make a

different kind of comparison. For each category, the accounts for the two variables are piled on top of one another.

Box Plot

A box plot is another figure that depicts the distribution of a numeric variable, but it does it in a different way than a histogram.

The grey section of the graphic is the box part of a box plot, which shows the distribution of values for a variable. The 25th and 75th percentiles are represented by the box's lower and upper edges, respectively. This indicates that the box represents the middle 50% of the data, and the median is the 50th percentile, which means that 50% of the data is larger than this figure and 50% of the data is less. The Whiskers at the top and bottom represent the 10th and 90th percentiles, respectively. As a result, 80% of the data is in the range indicated by the upper and lower extremes. Outliers are data value⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚ s a single point on the box plot.

It's worth noting that the box plot come⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚ whiskers reflecting various types of extr

Fig. 2 – Comparing Two Distributions
Base Salary Comparison
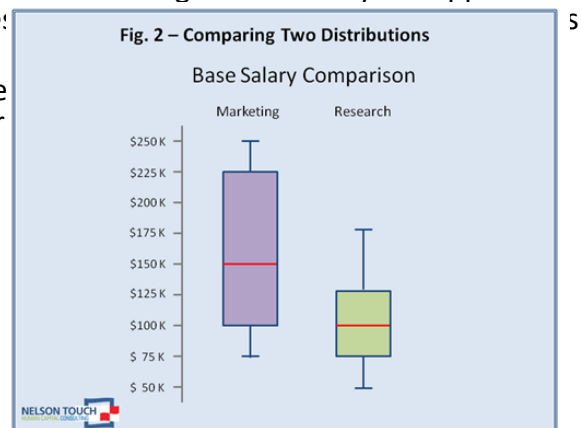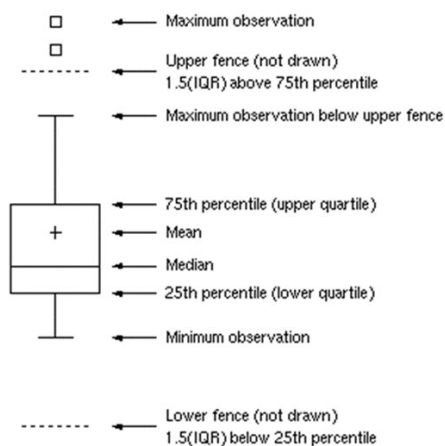
Box plots are commonly used to compare variables because they give a compact approach to demonstrate how variables are distributed.

For example, the above box plot compares the basic salaries for two different roles.

The median value, range, and spread of the two different variables may all be rapidly determined using this graphic.

We can tell right away that the median compensation for a marketing position is higher than for a research position.

Due to the bigger area of the purple box, we can also observe that the variation or spread of the values for marketing is greater than for research.

Distribution Shape and The Boxplot

Image source: http://www.slideshare.net/mido02/chap-3gbu
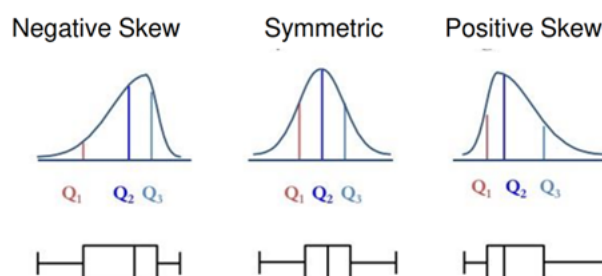
A box plot can also reveal whether the data value distribution is symmetrical, positively skewed, or negatively skewed. A box plot can also be presented on its side, as shown above.

If the line in the box that determines the median is in the middle of the box, it indicates a symmetric distribution.

When the median lies to the right of the box's centre, the skew is negative.

This indicates that there are more values below the median than above the median.

A positive skew is also suggested when the median is to the left of the box's centre.

To recap, data visualisation allows you to explore your data in a quick and intuitive manner. To explore data, data visualisation should be used in conjunction with summary statistics. The many types of plots discussed in this session will also come in handy when reporting your results throughout your machine learning project.