# Data-Driven Decision Making

Tushar B. Kute,
http://tusharkute.com

# Introduction

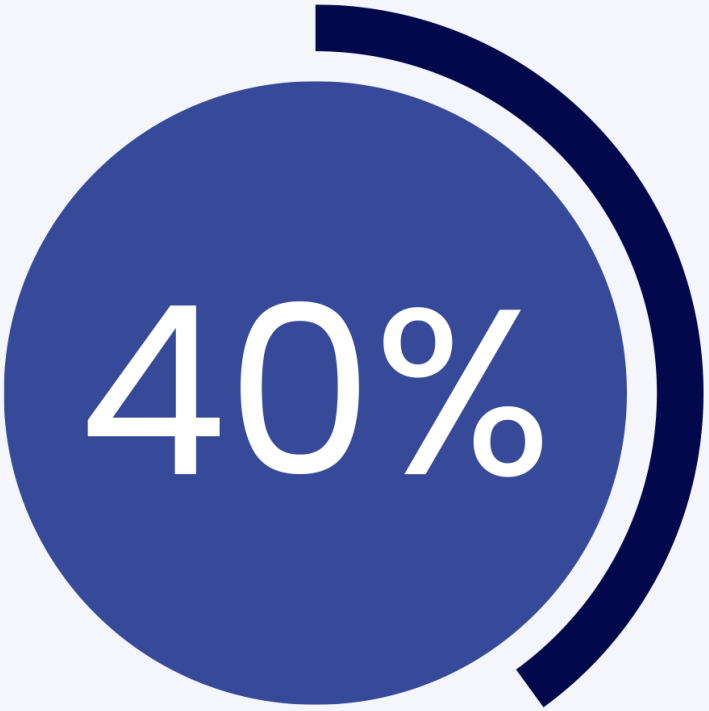
- As human beings, we often make split-second decisions based on feeling and intuition, without taking the outcome or consequence into account.
- We can’t always help it, we are emotional and empathetic creatures after all!
- But, when it comes to making business decisions, it is crucial to take all external factors into account before taking any decisive action.
- Through the process of data-driven decision-making (DDDM), organizations can ensure that their business goals and objectives are guided by cold, hard evidence.

# Data-Driven Decision Making (DDDM)

- Data-driven decision-making is defined as using facts, metrics, and insights to guide strategic business decisions that align with goals, strategies, and initiatives.
- It is a process that involves analyzing collected data through market research, and drawing insights, to benefit a business or organization.
- At its core, data-driven decision-making allows for a better understanding of business needs by leveraging real, verified data, instead of just making assumptions.

Reference: https://www.driveresearch.com

# Data-Driven Decision Making (DDDM)




***of developed products making it to market and of that 40%, only 60% generating any revenue at all***

Source: MarketingResearch.org

# Data-Driven Decision Making (DDDM)

- Though, the beauty of DDDM is that it has a positive impact on every area of your business.
  - Improves customer retention. Utilize customer surveys to identify areas of satisfaction, dissatisfaction, Net Promoter Score, likelihood to switch, and other critical KPIs to assure customers are happy with your organization.
  - Improves customer attrition. Rely on non-customer surveys to determine what drives prospects to use the products or services you sell. Understand things like sources of awareness, perception of your brand, and what competitors they are currently using.

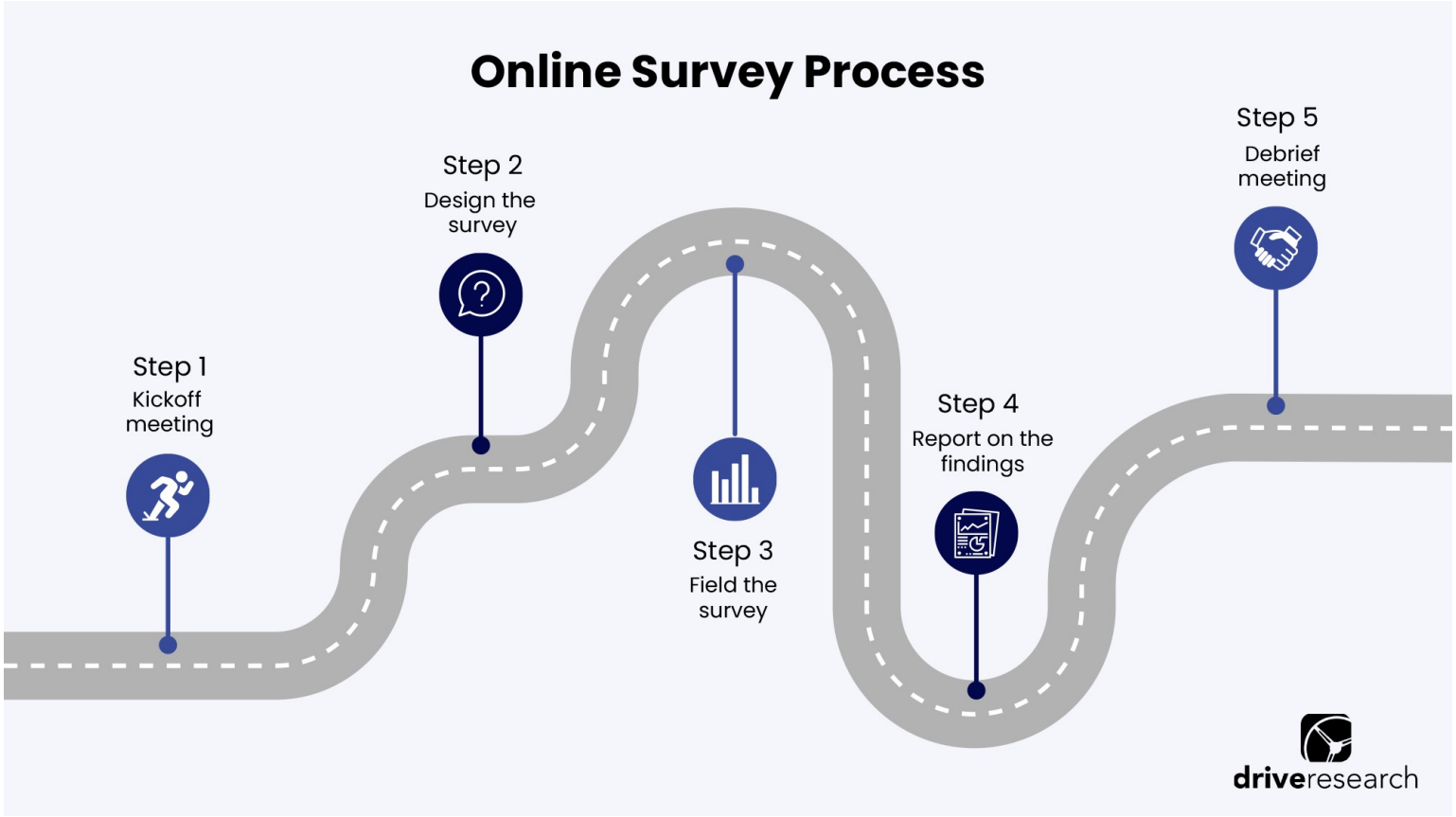
# Data-Driven Decision Making (DDDM)

- Improves employee satisfaction.
  - With the help of an employee survey company, learn what areas your team wants to be improved.
  - The insight will lead to cultural changes that have a direct impact on employee engagement and retention.

# Data-Driven Decision Making (DDDM)

- In order to get the best quality data, the following steps must be taken:
  - Determine your objectives
  - Write survey questions
  - Collecting survey data
  - Analyzing the results
  - Act on the data

# Data-Driven Decision Making (DDDM)



# Data-Driven Decision Making (DDDM)

- Determine your objectives
  - For any market research study, it is important to have clear goals and objectives in mind. And when working with a market research company, these will be discussed in a kickoff meeting.
  - This introductory meeting is also a great way for both the team and the client to get acquainted with each other.

# Data-Driven Decision Making (DDDM)

- Kickoff meetings cover a wide variety of topics.
- No matter what's discussed, there are certain factors that will always be covered in a kickoff meeting:
  - Project objectives
  - Key audiences to target
  - Timeline review
  - Reporting needs
  - Additional questions

# Data-Driven Decision Making (DDDM)

- Survey design
  - Now that the project goals are clear, the market research team will begin to design and write the survey.
  - There's more to writing a survey than ensuring questions are well-written.

# Data-Driven Decision Making (DDDM)

- A few recommendations from our online survey agency include:
  - Keep it concise. If you're going with an online survey, you won't want to make it longer than 20 minutes. In fact, 10 to 15 minutes are ideal here. Respondents can lose interest fast, especially anything over 20 minutes.
  - Change up the question style. Variety always plays a role in good market research. When asking questions, it's important to ensure they're not all formatted in the same way. This helps keep the respondent engaged.

# Data-Driven Decision Making (DDDM)

- Collecting survey data
  - After the survey is fully written and complete, it will be taken from its original document and programmed into an online survey platform.
  - Then, it's onto fieldwork. This is the period of time when responses are collected. Typically, this phase lasts a couple of weeks depending on your timeline.

# Data-Driven Decision Making (DDDM)

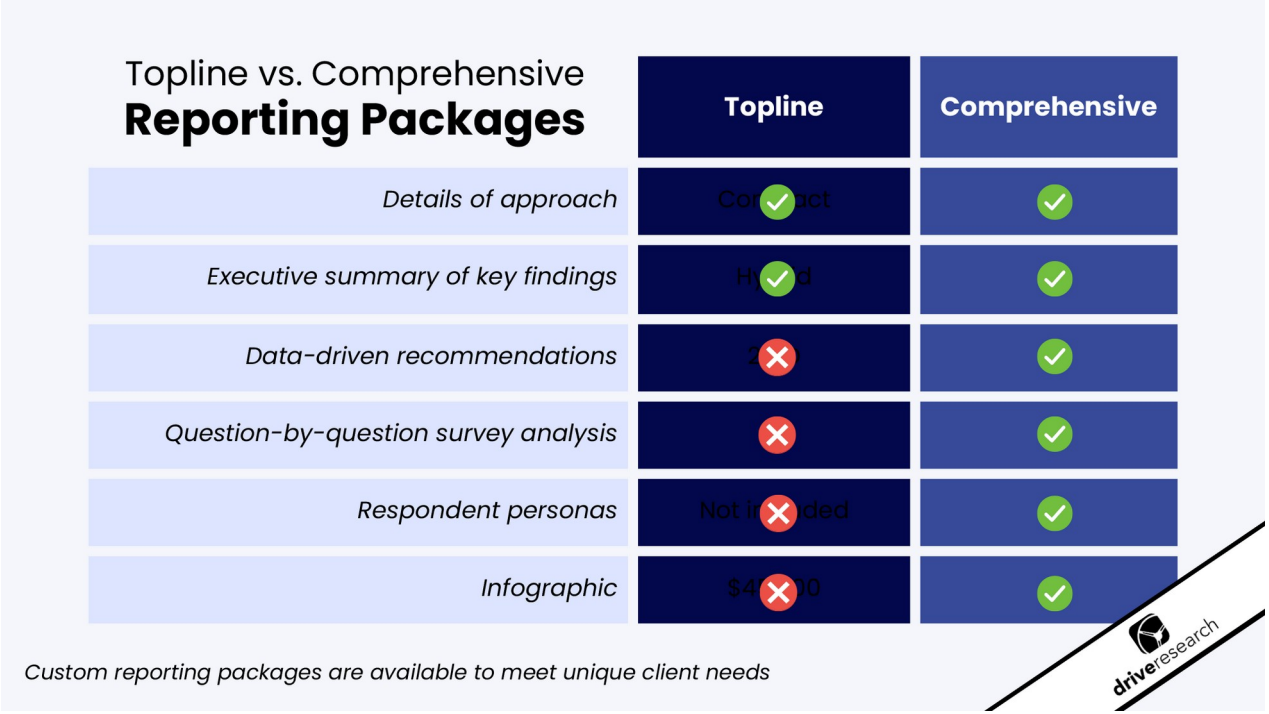
- Analysis and customized report
  - The final two steps to any solid market research process include a full analysis and report of the research data.
  - Prior to the analysis, the survey data will have been cleaned. This process simply ensures only the highest-quality data made it through.
  - And then we're onto the report.

# Data-Driven Decision Making (DDDM)

- A good market research report will cover:
  - Review of the methodology used
  - Coverage of main themes
  - Additional context
  - In-depth client suggestions

# Data-Driven Decision Making (DDDM)

- Though, based on your budget, there are ways to lower or increase the cost of the final deliverable.



# Data-Driven Decision Making (DDDM)

- Making confident decisions
  - One of the greatest advantages to come with data-driven decision-making is the ability to make decisions more confidently than ever before.
  - Save money and increase ROI
- Secondly, there's a good chance you can save on costs if you rely on data from the get-go.
  - Instead of taking an approach where you guess and see what happens, with data-driven decision making you're already up to date with all of the mainstream trends.

# Data-Driven Decision Making (DDDM)

- Become a proactive decision-maker
  - Last but certainly not least, data-driven decision-making turns you and your company into proactive decision-makers.
  - Too often, we find ourselves reacting to events in our lives that could have easily been avoided if only we had been willing to identify them beforehand.

# Data-Driven Decision Making (DDDM)

- Unite your team
  - Data-driven decision-making can also be used to promote your internal team.
  - When all members of your team are aligned with the proper data, everyone is on the same page. As a result, employees will work together instead of separately to meet company goals.

# Data-Driven Decision Making (DDDM)

- Create personal connections with buyers
  - One of the greatest benefits of data-driven decision-making is the connection made with clients and customers.
  - By consistently measuring market data, businesses will always be on top of the trends within their consumer base.

# Examples : Lufthansa

- With over 500 subsidiary companies, the Lufthansa Group is the second-largest airline company in Europe in terms of passengers carried.
- It brings in billions in revenue, but at one point, there was no uniformity in terms of data analytics across the many subsidiaries of this massive company.
- However, after deciding to use one analytics platform company-wide, efficiency skyrocketed by 30% across the company.

# Examples : Lufthansa

- Decision-makers across the company's subsidiaries were able to make better-informed decisions after careful data collection and analysis, which allowed business objectives to be streamlined and more efficient.
- By creating a data culture, Lufthansa empowered its employees to make better, more informed decisions.

# DDDM: Summary

- Data-driven decision-making or DDDM for short is the process of gathering insights and other figures to make informed business decisions.
- DDDM benefits all aspects of a company. From client relationships to consumer outreach to employee satisfaction, it should be at the core of all strategic decisions.

# Enterprise Data Management

- Enterprise data management (EDM) is the process of inventorying and governing your business's data and getting your organization onboard with the process.
- In other words, EDM is as much about managing people as it is about managing data.
- Data management means making sure your people have the accurate and timely data they need, and that they follow your standards for storing quality data in a standardized, secure, and governed place.

Reference: https://tableau.com

# Enterprise Data Management

- Enterprise data managers are most often database administrators, IT administrators, or IT project managers.
- They are in charge of the process of managing your business's entire data life cycle. They document and direct the flow of data from ingestion, and they control the process of removing data the business doesn't need.
- This life cycle is also referred to as a data lineage. By managing your data lineage, your data is less vulnerable to breaches, incorrect analysis, and legal complications.

# Enterprise Data Management

- By making data management a priority, you are ensuring that your data is in a secure place and available when your business users need it. This benefits your teams by enabling the following:
  - Accessing high-quality data for accurate analysis
  - Ensuring your data is secure and compliant under regulations
  - Consolidating data across multiple sources for increased efficiency
  - Having a consistent data architecture that scales with your enterprise

# Enterprise Data Management

- Keep these simple best practices in mind when starting on your own data management program:
  - Data managers need executive leadership, like the chief technical officer (CTO) or chief data officer (CDO), to buy in
  - Educate teams on the importance of data management and following your guidelines.
  - Prioritize data security and governance
  - Catalog your data
  - Improve data access to appropriate teams
  - Leverage modern data cataloging technologies to scale

# Data Preparation

- Data preparation is the process of preparing raw data so that it is suitable for further processing and analysis.
- Key steps include collecting, cleaning, and labeling raw data into a form suitable for machine learning (ML) algorithms and then exploring and visualizing the data.
- Data preparation can take up to 80% of the time spent on an ML project.
- Using specialized data preparation tools is important to optimize this process.



Reference: https://aws.amazon.com

# Data Preparation and ML

- Data flows through organizations like never before, arriving from everything from smartphones to smart cities as both structured data and unstructured data (images, documents, geospatial data, and more).
- Unstructured data makes up 80% of data today. ML can analyze not just structured data, but also discover patterns in unstructured data.
- ML is the process where a computer learns to interpret data and make decisions and recommendations based on that data.
- During the learning process—and later when used to make predictions—incorrect, biased, or incomplete data can result in inaccurate predictions.

# How do you prepare your data?

- Collect data
  - Collecting data is the process of assembling all the data you need for ML.
  - Data collection can be tedious because data resides in many data sources, including on laptops, in data warehouses, in the cloud, inside applications, and on devices.

# How do you prepare your data?

- Clean data
  - Cleaning data corrects errors and fills in missing data as a step to ensure data quality. After you have clean data, you will need to transform it into a consistent, readable format.
  - This process can include changing field formats like dates and currency, modifying naming conventions, and correcting values and units of measure so they are consistent.

# How do you prepare your data?

- Label data
  - Data labeling is the process of identifying raw data (images, text files, videos, and so on) and adding one or more meaningful and informative labels to provide context so an ML model can learn from it.
  - For example, labels might indicate if a photo contains a bird or car, which words were mentioned in an audio recording, or if an X-ray discovered an irregularity.
  - Data labeling is required for various use cases, including computer vision, natural language processing, and speech recognition.

# How do you prepare your data?

- Validate and visualize
  - After data is cleaned and labeled, ML teams often explore the data to make sure it is correct and ready for ML.
  - Visualizations like histograms, scatter plots, box and whisker plots, line plots, and bar charts are all useful tools to confirm data is correct.
  - Additionally, visualizations also help data science teams complete exploratory data analysis. This process uses visualizations to discover patterns, spot anomalies, test a hypothesis, or check assumptions.
  - Exploratory data analysis does not require formal modeling; instead, data science teams can use visualizations to decipher the data.

# Data Cleaning

- Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
- When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.
- If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct.
- There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset.



Reference: https://www.tableau.com

# Data Cleaning vs. Transformation

- Data cleaning is the process that removes data that does not belong in your dataset.
- Data transformation is the process of converting data from one format or structure into another.
- Transformation processes can also be referred to as data wrangling, or data munging, transforming and mapping data from one "raw" data form into another format for warehousing and analyzing.

# Steps


IMPORTING DATA
MERGING DATA SETS
REBUILDING MISSING DATA
STANDARDIZATION
NORMALIZATION
DE-DUPLICATION
VERIFICATION & ENRICHMENT
EXPORTING DATA
DATA CLEANING CYCLE

# Steps

- While the techniques used for data cleaning may vary according to the types of data your company stores, you can follow these basic steps to map out a framework for your organization.
    - Step 1: Remove irrelevant data
    - Step 2: Deduplicate your data
    - Step 3: Fix structural errors
    - Step 4: Deal with missing data
    - Step 5: Filter out data outliers
    - Step 6: Validate your data

# Remove irrelevant data

- First, you need to figure out what analyses you'll be running and what are your downstream needs.
- What questions do you want to answer or problems do you want to solve?
- Take a good look at your data and get an idea of what is relevant and what you may not need. Filter out data or observations that aren't relevant to your downstream needs.

# Remove irrelevant data

- If you’re doing an analysis of SUV owners, for example, but your data set contains data on Sedan owners, this information is irrelevant to your needs and would only skew your results.
- You should also consider removing things like hashtags, URLs, emojis, HTML tags, etc., unless they are necessarily a part of your analysis.

# Deduplicate your data

- If you’re collecting data from multiple sources or multiple departments, use scraped data for analysis, or have received multiple survey or client responses, you will often end up with data duplicates.
- Duplicate records slow down analysis and require more storage.
- Even more importantly, however, if you train a machine learning model on a dataset with duplicate results, the model will likely give more weight to the duplicates, depending on how many times they’ve been duplicated. So they need to be removed for well-balanced results.

# Fix structural errors

- Structural errors include things like misspellings, incongruent naming conventions, improper capitalization, incorrect word use, etc.
- These can affect analysis because, while they may be obvious to humans, most machine learning applications wouldn’t recognize the mistakes and your analyses would be skewed.

# Fix structural errors

- For example, if you're running an analysis on different data sets – one with a 'women' column and another with a 'female' column, you would have to standardize the title.
- Similarly things like dates, addresses, phone numbers, etc. need to be standardized, so that computers can understand them.

# Deal with missing data

- Scan your data or run it through a cleaning program to locate missing cells, blank spaces in text, unanswered survey responses, etc.
- This could be due to incomplete data or human error. You'll need to determine whether everything connected to this missing data – an entire column or row, a whole survey, etc. – should be completely discarded, individual cells entered manually, or left as is.

# Deal with missing data

- The best course of action to deal with missing data will depend on the analysis you want to do and how you plan to preprocess your data.
- Sometimes you can even restructure your data, so the missing values won't affect your analysis.

# Filter out data outliers

- Outliers are data points that fall far outside of the norm and may skew your analysis too far in a certain direction.
- For example, if you're averaging a class's test scores and one student refuses to answer any of the questions, his/her 0% would have a big impact on the overall average.
- In this case, you should consider deleting this data point, altogether. This may give results that are "actually" much closer to the average.

# Filter out data outliers

- However, just because a number is much smaller or larger than the other numbers you're analyzing, doesn't mean that the ultimate analysis will be inaccurate.
- Just because an outlier exists, doesn't mean that it shouldn't be considered.
- You'll have to consider what kind of analysis you're running and what effect removing or keeping an outlier will have on your results.

# Validate your data

- Data validation is the final data cleaning technique used to authenticate your data and confirm that it’s high quality, consistent, and properly formatted for downstream processes.
  - Do you have enough data for your needs?
  - Is it uniformly formatted in a design or language that your analysis tools can work with?
  - Does your clean data immediately prove or disprove your theory before analysis?

# Validate your data

- Validate that your data is regularly structured and sufficiently clean for your needs. Cross check corresponding data points and make sure nothing is missing or inaccurate.
- Machine learning and AI tools can be used to verify that your data is valid and ready to be put to use.
- And once you've gone through the proper data cleaning steps, you can use data wrangling techniques and tools to help automate the process.

# Data Cleaning Tips

- Create the right process and use it consistently
  - Set up a data cleaning process that's right for your data, your needs, and the tools you'll use for analysis.
  - This is an iterative process, so once you have your specific steps and techniques in place, you'll need to follow them religiously for all subsequent data and analyses.

# Data Cleaning Tips

- Use tools
  - There are a number of helpful data cleaning tools you can put to use to help the process – from free and basic, to advanced and machine learning augmented.
  - Do some research and find out what data cleaning tools are best for you.
  - If you know how to code, you can build models for your specific needs, but there are great tools even for non-coders.

# Data Cleaning Tips

- Pay attention to errors and track where dirty data comes from
  - Track and annotate common errors and trends in your data, so you'll know what kinds of cleaning techniques you need to use on data from different sources.
  - This will save huge amounts of time and make your data even cleaner – especially when integrating with analysis tools you use regularly.

# Summary

- It’s clear that data cleaning is a necessary, if slightly annoying, process when running any kind of data analysis.
- Follow the steps above and you’re well on your way to having data that’s fully prepped and ready for downstream processes.
- Remember to keep your processes consistent and don’t cut corners on data cleaning, so you’ll end up with accurate, real-world, immediately actionable results.

# Thank you



@mitu_skillologies

@mITuSkillologies

@mitu_group

@mitu-skillologies

@MITUSkillologies

kaggle

@mituskillologies

**Web Resources**
https://mitu.co.in
http://tusharkute.com

@mituskillologies

**contact@mitu.co.in**

**tushar@tusharkute.com**