# SMA Module 4 - Data Analytics

https://docs.google.com/presentation/d/1M8W-IU8ZFIuSe5M2irPu81WKzcUxiXH6kymsDHudbJM/edit#slide=id.p

## Clickstream analysis

Click stream analysis **involves tracking the sequence of user interactions** (clicks) as they navigate through a website or online platform.

It helps in **understanding user behavior patterns**, preferences, and engagement levels.

By analyzing the clickstream data, businesses can gain insights into:

- **Most visited pages**: Identify popular pages that attract more user attention.
- **Conversion paths**: Understand the flow of users from landing pages to conversion points.
- **Drop-off points**: Identify where users leave the website without completing desired actions.
- **User journey**: Visualize the path users take to explore content and services.

### What is click stream data?

- Clickstream data and clickstream analytics are the **processes involved in collecting, analyzing and reporting aggregate data about which pages a website visitor visits** -- and in what order.
- **The path** the visitor takes through a website is called the clickstream.
- Clickstreams are categorized into **clickstream data and clickstream analytics**, which is also referred to as clickstream analysis.
- The **clickstream data** is the **information** collected about a user while they browse through a website or use a web browser.

Clickstream analytics is the **process of tracking, analyzing and reporting data** on the pages a user visits and user behavior while on a webpage.

**Clickstream data includes the following types of web analytics information:**

- whether the individual is a unique or repeat visitor to the website;
- the terms an individual plugs into a search engine;
- what page the individual lands on first;
- the amount of time a user spends on a page;
- the features on the page the user clicks on and engages with;
- when and where an item is added or removed from a cart;
- where the user goes next; and
- when the back button is used.

## Clickstream Analytics

There are two levels of clickstream analysis: traffic analytics and e-commerce analytics.

**Traffic analytics:**

This analysis **operates at the server level**. It collects and analyzes the following data sets:

- how many pages are served to a user;
- how long it takes each page to load;
- how often the user hits the browser's back button; and
- how much data is transmitted before the user moves to a different webpage.

**E-commerce analytics:**

This analysis uses clickstream data **to determine the effectiveness of a website in terms of conversions and transactions**. It is concerned with the following data points:

- what pages the shopper lingers on;
- what the shopper puts in or takes out of a shopping cart;
- what items the shopper purchases;
- whether the shopper belongs to a loyalty program;

- whether the shopper uses a coupon code; and
- the shopper's preferred method of payment.

## Benefits of clickstream data analysis:

- **User information.** The data collected can include search terms used, pages landed on, webpage features used and the addition or removal of items from a cart, all of which can lead to more actionable insights.
- **User routes.** Organizations can use data analysis to view the different routes their online visitors or customers take to reach a page or to make a purchase.
- **Customer trends and insights.** Collecting and analyzing the clickstreams of a large number of visitors lets an organization identify trends in the following areas:
  - how visitors get to the website;
  - what they do once there;
  - how long they stay on a page;
  - the number of page visits visitors make; and
  - the number of unique and repeat visitors.
- **UX.** If a majority of users quickly leave a page or website, it could be a sign that the page is poorly optimized or doesn't contain enough information of value. Clickstream data enables an organization to recognize UX shortcomings, enabling them to make necessary changes.
- **Digital marketing.** Clickstream data can be used to determine the amount of traffic coming from ad banners and campaigns. Such data provides insight as to which advertisements are most effective and lead to customer conversion rate optimization. Clickstream analysis can also derive what times of day, month or year a marketing strategy is most effective.

## A/B Testing

- A/B testing, also known as split testing, is a method **used in marketing and experimentation to compare two versions of a webpage, advertisement, or other digital content** to determine which one performs better in achieving a specific goal or outcome.
- It involves randomly **dividing a sample of users into two groups: Group A and Group B.**
- **Group A is exposed to the original or control version** (referred to as variant A) while **Group B is exposed to a modified version** (referred to as variant B).
- The two variants **differ in a specific aspect**, such as a different headline, call-to-action button, layout, or color scheme. **The performance of each variant is then measured**, and statistical analysis is applied to determine if there is a significant difference in their effectiveness.
- It helps answer questions like:

  1. Which version leads to higher click-through rates, conversions, or sales?

  2. Which variant improves user engagement or reduces bounce rates?

  3. Which design or content element generates more user interactions or time spent on the page?
- It helps identify areas of improvement, validate hypotheses, and optimize digital experiences to achieve desired business objectives.

## Online Surveys

- Online surveys are a popular method of collecting data and gathering insights from a targeted audience.
- They involve the distribution of a set of questions through digital platforms, such as websites, email, social media, or survey platforms, to collect respondents' opinions, feedback, or demographic information
- Key Aspects:

  1. **Design**: Online surveys can be designed using various **question formats**, including multiple-choice, open-ended, Likert scale, or rating scales. The design should be user-friendly, visually appealing, and easy to navigate to encourage higher response rates.

2. **Target Audience**: Online surveys allow researchers to **reach a wide range of participants**, including specific demographics or individuals with particular interests. Targeting the right audience ensures that the collected data is relevant to the research objective.

3. **Convenience and Accessibility**: Respondents can **complete online surveys at their convenience**, from any location with internet access. This convenience factor leads to higher response rates compared to other data collection methods, such as phone or in-person interviews.

4. **Data Collection**: Online surveys automatically **collect and store responses in a centralized database**, eliminating the need for manual data entry. This streamlines the analysis process and reduces the chances of data entry errors.

5. **Data Analysis**: Online survey platforms often **provide tools for data analysis**, including generating summary statistics, creating visualizations, and exporting data for further analysis. Researchers can gain insights by analyzing response patterns, identifying trends, and drawing conclusions from the collected data.

6. **Cost and Time Efficiency**: Online surveys are generally **more cost-effective than traditional methods**, such as paper-based surveys or phone interviews. They eliminate printing and mailing costs and can be conducted within a shorter time frame, allowing for quicker data collection and analysis.

7. **Flexibility**: Online surveys **offer flexibility in terms of survey length**, question branching, and customization options. Researchers can tailor surveys to specific research objectives and easily modify them as needed.

## Google Analytics

- Google Analytics is a web analytics service that provides statistics and basic analytical tools for search engine optimization (SEO) and marketing purposes.
- The service is part of the Google Marketing Platform and is available for free to anyone with a Google account.
- Google Analytics is used to track website performance and collect visitor insights.
- It can help organizations determine top sources of user traffic, gauge the success of their marketing activities and campaigns, track goal completions (such as

purchases, adding products to carts), discover patterns and trends in user engagement and obtain other visitor information such as demographics.

- Small and medium-sized retail websites often use Google Analytics to obtain and analyze various customer behavior analytics, which can be used to improve marketing campaigns, drive website traffic and better retain visitors.

## Web Crawling

- Web crawling, also known as web scraping or spidering, refers to the automated process of **systematically browsing the World Wide Web** to discover and gather information from web pages.
- It involves the **use of web crawlers** or bots, which are computer programs that **navigate through web pages by following hyperlinks.**
- **The web crawler starts with a list of seed URLS** and proceeds to visit each URL, **extracting the content and following any outgoing links on the page**.
- **This process continues recursively, enabling the crawler to access and collect data from a large number of web pages.** The collected data can include text, images, metadata, and other relevant information.
- Web crawling is vital for search engines as it **allows them to continuously discover and update their index of web pages.**
- It enables search engines **to provide up-to-date and comprehensive search** results by ensuring that new content is indexed and existing content is reindexed.

## Indexing

- Once web pages are crawled, **the next step is indexing**. Indexing is the process of organizing and storing the collected web page data in a **structured manner** to facilitate quick and efficient search retrieval.
- The data is typically stored in an index database that allows for fast searching and retrieval of relevant information. During indexing, various techniques are employed to analyze and extract key information from web pages.
- This can include extracting text content, identifying important keywords, generating metadata, and establishing relationships between different web pages.

- Web crawling and indexing are continuous processes as new web pages are constantly added to the web and existing pages are updated. Search engines employ sophisticated algorithms to ensure efficient crawling, effective indexing, and accurate retrieval of search results.

## Natural Language Processing Techniques for Micro-Text Analysis:

- Micro-text analysis refers to the **process of analyzing short and concise text data**, such as social media posts, tweets, product reviews, and chat messages.
- Natural Language Processing (NLP) techniques play a crucial role in **extracting meaningful insights from micro-text data.**

**Actions that generate click stream data :**

- User login
- Account registration
- Newsletter sign-up
- Search performed
- Product added to cart
- Product purchased
- Feedback provided

**What click stream data is used to analyse:**

- Ad campaigns
- Marketing campaigns
- Campaign tools
- Product performance
- A/B testing trials