**WEEK 2: DATA ANALYTICS BASICS**

**Task 2: Overview of Data Analytics**

Before we dive deep into the toolkit of a data analyst, let’s try to uncover what actually is data analytics.

Data analytics is the science of analyzing raw data in order to **make conclusions about that information**. Data analytics techniques can **reveal trends and metrics** that would otherwise be lost in the mass of information. This information can then be used to **optimize processes to increase the overall efficiency of a business or system.**

For example, manufacturing companies often record the runtime, downtime, and work queue for various machines and then analyze the data to better plan the workloads so the machines operate closer to peak capacity.

Gaming companies use data analytics to set reward schedules for players that keep the majority of players active in the game. Content companies use many of the same data analytics to keep you clicking, watching, or re-organizing content to get another view or another click.

Data analytics is important because it helps businesses **optimize their performances**. Implementing it into the business model means companies can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data. A company can also use data analytics to make **better business decisions** and help analyze customer trends and satisfaction, which can lead to new—and better—products and services

Let’s understand again, what is Data analytics using the tutorial below.

Video - <https://youtu.be/yZvFH7B6gKI>

# Task 3: Data Analytics Process

Data analytics process or framework is what drives the entire analysis cycle from asking the right business questions, collecting, cleaning, analyzing and visualizing insights powered by data. We’re going to look at phases of Data Analysis process includes (**SIX STEP PROCESS)**

You may take notes of this 6 step framework developed & used by Google. This would help you do analysis in structured way for upcoming case studies and examples.

1. **Ask** : Analysts in need to define what the project would look like and what would qualify as a successful result (or stakeholder’s expectations in the project). So, to determine these things, they asked effective questions and collaborated with leaders and managers who were interested in the outcome of their people analysis.
2. **Prepare**: The analysts identify and prepare what kind of data they need to achieve the successful result they identified in the previous step
3. **Process**: The data collected through surveys or any other way has cleaned up to make sure it is complete, correct, and relevant and free of errors and outliers
4. **Analyze** : Analyze data to find patterns, relationships, and trends. Discover the key indicator of the success of our project and document exactly what is found in the analysis, no matter what the results are.
5. **Share** Communicate the results with the right context and have productive team conversations about next steps.
6. **Act**: Work with a team within your company and decide how best to implement changes and take actions based on the findings



# Task 4: Analytical Thinking

Data analytics is not only about frameworks and tools. A great analyst has strong foundational thinking ability to crunch numbers and drive insights about the data. You may follow the below decision making framework and start abiding by the analytical thinking framework.

One of the most powerful ways you can put data to work is with **data-driven decision-making**. Data-driven decision-making is defined as using facts to guide business strategy. Analytical thinking involves identifying and defining a problem and then solving it by using data in an organized, step-by-step manner. The five key aspects to analytical thinking are:

1. **Visualization**: Visualization is the graphical representation of information. Some examples include graphs, maps, or other design elements.
2. **Strategy**: With so much data available, having a strategic mindset is key to staying focused and on track. Strategizing helps data analysts see what they want to achieve with the data and how they can get there. Strategy also helps improve the quality and usefulness of the data we collect.
3. **Problem-orientation**: In order to identify viable alternatives, a concrete formulation of the problem is required. It is the way that people approach problems, and how they set them into the context of their existing knowledge and ways of looking at the world.
4. **Correlation**: A correlation is like a relationship. You can find all kinds of correlations in data. Correlation does not equal causation. In other words, just because two pieces of data are both trending in the same direction, that doesn’t necessarily mean they are all related.
5. **Big-picture and Detail-oriented thinking**: Big picture thinking means being able to see the big picture as well as the details. A jigsaw puzzle is a great way to think about this. Big-picture thinking is like looking at a complete puzzle. You can enjoy the whole picture without getting stuck on every tiny piece that went into making it. If you only focus on individual pieces, you wouldn’t be able to see past that, which is why big-picture thinking is so important.

On the flip side, detail-oriented thinking is all about figuring out all of the aspects that will help you execute a plan. In other words, the pieces that make up your puzzle.

In data analysis, solutions are almost never right in front of you. You need to think critically to find out the right questions to ask. But you also need to think creatively to get new and unexpected answers. Some of the questions data analysts ask when they’re on the hunt for a solution are:

* **What is the root cause of a problem?**

A root cause is the reason why a problem occurs. If we can identify and get rid of a root cause, we can prevent that problem from happening again. A simple way to wrap your head around root causes is with the process called the Five Whys. In the Five Whys you ask “why” five times to reveal the root cause. The fifth and final answer should give you some useful and sometimes surprising insights.

* **Where are the gaps in our process?**

For this, many people will use something called gap analysis. Gap analysis lets you examine and evaluate how a process works currently in order to get where you want to be in the future. Businesses conduct gap analysis to do all kinds of things, such as improve a product or become more efficient. The general approach to gap analysis is understanding where you are now compared to where you want to be. Then you can identify the gaps that exist between the current and future state and determine how to bridge them.

* **What did we not consider before?**

This is a great way to think about what information or procedure might be missing from a process, so you can identify ways to make better decisions and strategies moving forward

# Task 5: Applications in Real World

Now that you know the traditional explanation of what Data analytics really is, the best way to actually understand it is to see **how it works in real life**. Let’s take a look at how some of the world’s largest brands made Data analytics a part of their core business.

One thing that companies in e-commerce, entertainment, healthcare, manufacturing, marketing, finance, tech, and hundreds of other industries all have in common is that they all use data.

**Data is basically a collection of facts or information**. This collection can include numbers, pictures, videos, words, measurements, observations, and more. It’s everywhere. Once you have data, analytics puts it to work through analysis. **Data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.** A data analyst is an explorer, a detective, and an artist all rolled into one.

Every minute of every hour of every day, more data is being created. Businesses need a way to control all that data so they can use it to improve processes, identify opportunities and trends, launch new products, serve customers, and make thoughtful decisions. **For businesses to be on top of the competition, they need to be on top of their data.** That’s why these companies hire data analysts to control the waves of data they collect every day, make sense of it, and then draw conclusions or make predictions.

Google processes more than 40,000 searches every second. That’s 3.5 billion searches a day and 1.2 trillion searches every year. YouTube has almost two billion users. With this kept in mind, let’s understand this better through a real life use case of data & it’s analysis in marketing & advertising.

**Facebook Data Use Case** in Advertising, Marketing & even Politics

Video - <https://youtu.be/JAO_3EvD3DY>

Video - <https://youtu.be/vyRsarL-dqk>

# Task 6: Amazon Case Study

Amazon is definitely the number 1 e-commerce shop at the moment and they have their database to thank for that. They are constantly using big data to improve their customer experience, so here are two examples that show just how well that works.

##### 1. Dynamic pricing

Everybody knows that airlines use this tactic when selling plane tickets – if you check out the same tickets over and over again, it probably means you really want them and are willing to pay more for them. That same logic is implemented on Amazon’s website. But what you probably didn’t know is that they change their prices up to 2,5 million times a day.

**What affects these price changes are factors like shopping patterns, competitor’s prices, and whether the product is a common one or not.**

##### 2. Product recommendations

It doesn’t matter if the person buys the products, puts it in the cart or even just takes a look at it – Amazon will use that data. That way they can learn what each customer wants and likes, and can recommend that same product or similar ones to them when they return to the shop.

**This is how the company earns 35% of their annual sales.**

Video - <https://youtu.be/O0DnYqkRWWQ>

# Task 7: Netflix Case Study

Netflix is unarguably the biggest online platform for streaming movies and TV shows, and it owes its success to Data analytics. Because they know their users, their retention rate is 93% which, compared to their main competitors, is a huge number. **So, how does Netflix use data analytics?**

By collecting data from their 151 million subscribers, and implementing data analytics models to discover customer behavior and buying patterns. Then, using that information to recommend movies and TV shows based on their subscribers’ preferences.

According to Netflix, over 75% of viewer activity is based off **personalized recommendations**. Netflix collects several data points to **create a detailed profile** on its subscribers. The profile is far more detailed than the personas created through conventional marketing.

Most significantly, Netflix collects customer interaction and response data to a TV show. For example, Netflix knows the time and date a user watched a show, the device used, if the show was paused, does the viewer resume watching after pausing? Do people finish an entire TV show or not, how long does it take for a user to finish a show and so on.

Netflix even has screenshots of scenes people might have viewed repeatedly, the rating content is given, the number of searches and what is searched for. With this data, Netflix can create a detailed profile on its users. To collect all this data and harness it into meaningful information, **Netflix requires data analytics**. For example, Netflix uses what is known as the recommendation algorithm to suggest TV shows and movies based on user’s preferences.

Netflix’s ability to collect and use the data is the reason behind their success. According to Netflix, they earn over a billion in customer retention because the **recommendation system** accounts for over 80% of the content streamed on the platform. Netflix also uses its big data and analytics tools to decide if they want to greenlight original content. To an outsider, it might look like Netflix is throwing their cash at whatever they can get, but in reality, **they greenlight original content based on several touch points derived from their user base.**

For example, Netflix distributed ‘Orange is the New Black’ knowing it would be a big hit on their platform. How? Because ‘Weeds’, Jenji Kohan’s previous hit performed well on Netflix in terms of viewership and engagement.

Netflix even uses big data and analytics to conduct **custom marketing**, for example, to promote ‘House of Cards’ Netflix cut over **ten different versions of a trailer** to promote the show. **If you watched lots of TV shows centered on women, you get a trailer focused on the female characters**. However, if you watched a lot of content directed by David Finch, you would have gotten a trailer that focused the trailer on him. Netflix did not have to spend too much time and resources on marketing the show because they already knew how many people would be interested in it and what would incentivize them to tune in.

Video - <https://youtu.be/sAQmj8a_aI8>

Detailed case study of Netflix - <https://neilpatel.com/blog/how-netflix-uses-analytics/>

# Task 8: Data Lifecycle Management

Till now, we covered the overview of data analytics, **processes & frameworks and some business cases.** We would soon be starting with actual **tools of a data analysis** and solving case studies in upcoming weeks. But before that, we are left with the final step. That is to understand the **Data Lifecycle and the most important tools** used by a data analyst on a day to day basis.

The data lifecycle represents all of the stages of data throughout its life from its **creation for a study to its distribution** and reuse. The data lifecycle begins with a researcher(s) developing a concept for a study; once a study concept is developed, data is then collected for that study.

After data is collected, it is processed for distribution so that it can be archived and used by other researchers at a later date. Once data reaches the distribution **stage of the lifecycle**, it is stored in a location (i.e. repository, registry) where it can then be discovered by other researchers. Data discovery leads to the **repurposing of data, which creates a continual loop back** to the data processing stage where the repurposed data is archived and distributed for discovery.

**Data Life Cycle:** The life cycle of data is **plan, capture, manage, analyze, archive and destroy**.

1.**Plan:** This actually happens well before starting an analysis project. During planning, a business decides what kind of data it needs, how it will be managed throughout its life cycle, who will be responsible for it, and the optimal outcomes.

2. **Capture:** This is where data is collected from a variety of different sources and brought into the organization. With so much data being created everyday, the ways to collect it are truly endless.



3. **Manage:** This involves how we care for our data, how and where it’s stored, the tools used to keep it safe and secure, and the actions taken to make sure that it’s maintained properly.

4. **Analyze:** In this phase, the data is used to solve problems, make great decisions, and support business goals.

5. **Archive:** Archiving means storing data in a place where it’s still available, but may not be used again. During analysis, analysts had a huge amount of data. So it makes sense to archive the data which is no longer relevant to us.

6. **Destroy:** This is important for protecting a company’s private information, as well as private data about its customers.

Video - <https://youtu.be/5I2bYqeFQy4>

# Task 9: Toolkit of a Data Analyst

Let’s look at the tools you need to be **familiar with to become a successful data analyst.** In coming modules, we will go in-depth with each tool along with practical assignment & case studies.

* **Spreadsheets**: Data analysts rely on spreadsheets to collect and organize data. Two popular spreadsheet applications you will probably use a lot in your future role as a data analyst are Microsoft Exceland Google Sheets. Digital worksheets structure data in a meaningful way by letting you

1. Collect, store, organize, and sort information
2. Identify patterns and piece the data together in a way that works for each specific data project
3. Create excellent data visualizations, like graphs and charts.

* **Databases and query languages**: A database is a collection of structured data stored in a computer system. Some popular Structured Query Language (SQL) programs include MySQL, Microsoft SQL Server, and Big Query. Query languages;

1. Allow analysts to isolate specific information from a database(s)
2. Make it easier for you to learn and understand the requests made to databases
3. Allow analysts to select, create, add, or download data from a database for analysis

* **Visualization Tools**: Data analysts use a number of visualization tools, like graphs, maps, tables, charts, and more. Two popular visualization tools are Tableau and Looker. Tableau’s simple drag-and-drop feature lets users create interactive graphs in dashboards and worksheets while Looker communicates directly with a database, allowing you to connect your data right to the visual tool you choose These tools

1. Turn complex numbers into a story that people can understand
2. Help stakeholders come up with conclusions that lead to informed decisions and effective business strategies
3. Have multiple features

* **Statistical Analysis and Visualization**: Involves using programming languages, like R andPython

Gear up, in the next module we are going to start with these tools. Let’s start our Data Analytics journey!

# Task 10: Summary

* Data analytics is the science of analyzing raw data in order to**make conclusions about that information.**
* Data analytics techniques can**reveal trends and metrics** that would otherwise be lost in the mass of information.
* Data analytics is important because it helps businesses **optimize their performances.**
* Data analytics process or framework is what drives the entire analysis cycle from asking the right business questions, collecting, cleaning, analyzing and visualizing insights powered by data.
* Phases of Data Analysis process includes (**SIX STEP PROCESS):**

1. Ask
2. Prepare
3. Process
4. Analyze
5. Share
6. Act

* **Data-driven decision-making** is defined as using facts to guide business strategy.
* Analytical thinking involves **identifying** and **defining** a problem and then solving it by using data in an organized, step-by-step manner.
* The five key aspects to analytical thinking are:

1. Visualization
2. Strategy
3. Problem – Orientation
4. Correlation
5. Big – picture and Detailed – oriented thinking

* Some of the questions data analysts ask when they’re on the hunt for a solution are:

1. What is the root cause of a problem
2. Where are the gaps in our process
3. What did we not consider before

* **Data is basically a collection of facts or information.**
* **Data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.**
* Companies hire data analysts to control the waves of data they collect every day, make sense of it, and then draw conclusions or make predictions.
* The data lifecycle represents all of the stages of data throughout its life from its **creation for a study to its distribution** and reuse.
* Data discovery leads to the **repurposing of data, which creates a continual loop back** to the data processing stage where the repurposed data is archived and distributed for discovery.
* The life cycle of data is –

1. Plan
2. Capture
3. Manage
4. Analyze
5. Archive
6. Destroy

* The **tools** that we need for data analyst are –

1. Spreadsheets – Microsoft Excel and Google Sheet
2. Databases and query languages – MySQL , Microsoft SQL server and Big Query
3. Visualization tools – Tableau and Looker
4. Statistical Analysis and Visualization – R and Python

# Task 11: Intro to Data Analytics : Virtual Session

Now , you have an overview of what is data analytics and its time for you to know what are the different kinds of data analytics and what are the objectives we are trying to achieve with these types of analysis.

Video - <https://vimeo.com/665072901>