Course 1:

Foundations: Data, Data, Everywhere

Week 1: **Introducing Data Analytics**

15/12/2021

Introduction:

Data: a collection of facts

Data analysis: the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making

Data analyst: someone who collects, transforms, and organizes data in order to help make informed decisions

Transforming data into insights

Business need data to:

- Improve processes
- Identify opportunities and trends
- Launch new products
- Serve customers
- Make thoughtful decisions

Six steps of data analysis:

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

Case Study: New data perspectives

An organization was experiencing high turnover rates among new hires.

Question: **how can the organization improve the retention rate for new employees?**

ASK:

needed to define what the project would look like and what would qualify as a successful result.

- What do new employees need to learn to be successful in their first year of job?
- Is there a previous data for new employees? If yes, do we have access to it?
- Do managers with higher retention rates offer new employees something extra or unique?
- What is the leading cause for dissatisfaction among new employees?
- By what percentage would you like employee retention to increase in the next fiscal year?

PREPARE:

The group built a time line of three months and decided how they will progress in the time period. The analysts identified what data was needed to achieve the successful result identified in previous step – they chose to gather data from an online survey of new employees.

- They developed specific questionnaire for employee satisfaction with different business process.
- They established rules for access to the collected data.
- They finalized what information would be gathered and how to present it.
- Possible issues were brainstormed and solutions were proposed.

PROCESS:

The survey was sent out ensuring that employees provided data with consent and understood that their data would be collected, stored, managed and protected. **Data ethics is a major responsibility of data analysts**.

- They restricted access to data to their group
- They cleaned data to make sure it was complete, correct and relevant.
- They uploaded raw data to internal data warehouse for additional layer of security

ANALYZE:

From the completed surveys, the analysts discovered that employee experience with certain processes was a key indicator in overall job satisfaction.

- Employees who experienced a long complicated hiring process were most likely to leave the company
- Employees who experienced an efficient and transparent evaluation and feedback process were most likely to remain with the company.

The exact findings were document in results to truthful analysis.

SHARE:

- They shared the report with managers who met or exceeded minimum number of direct reports with submitted responses to the survey
- They presented the results to the managers to make sure they had the full picture
- They asked the managers to personally deliver the results to their teams

It needs to be ensured that the results are communicated in the right context.

ACT:

The last stage was to work with leaders and decide how to implement changes and take actions based on the findings.

- Standardize the hiring and evaluation process for employees based on the most efficient and transparent practices
- Conduct the same survey annually and compare results with those from the previous year.

20-12-2021

Dimensions of data analytics

A data analyst is an explorer, a detective and an artist all rolled into one.

Data science encompasses three discipline: machine learning, statistics, and analytics.

When considering data science and choosing which area to specialize in, go with your personality.

23-12-2021

What is the data ecosystem?

Data ecosystems are made up of various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data. These elements include hardware and software tools, as well as the people who use them.

Cloud: a place to keep data online, rather than a computer hard drive

Data science: Creating new ways of modeling and understanding the unknown by using raw data.

Data analysis: The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics: The science of data.

How data informs better decisions

Data-driven decision-making: using facts to guide business strategy.

Data and gut instinct

Data analysis depends on facts and clues to make decisions. If an analyst makes decisions based on their own experience then the data becomes biased and if the decision is made on gut instinct then it can cause mistakes.

The more we understand data related to a project, the easier it will be to figure out what is required. These efforts also help in identifying gaps and errors in data which helps in better communication of the findings. Past experience can play a role in making a connection between such gaps which would not be oblivious to someone with less knowledge of the subject.

Data + Business knowledge = mystery solved

Origins of the data analysis process

Data analysis is rooted in statistics, which dates back to ancient Egyptians and Indians.

Week 2: All about analytical thinking

25-12-2021

Analytical Skills: Qualities and characteristics associated with solving problems using facts.

1. Curiosity
2. Understanding context
3. Having a technical mindset
4. Data design
5. Data strategy

Context: the condition in which something exists or happens.

Technical mindset: the ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.

Data Design: how you organize information

Data Strategy: the management of the people, processes and tools used in data analysis

Analytical thinking: identifying and defining a problem and then solving it by using data in an organized step-by-step manner

Key skills for analytical thinking:

1. Visualization:
   a. the graphical representation of information
   b. it helps in understanding and explaining information more effectively
2. Strategy:
   a. helps to see what needs to be achieved with the data and how to get there
   b. it helps improve the quality and usefulness of the data we collect
3. Problem-orientation
   a. Helps to identify, describe and solve problems
4. Correlation
   a. Being able to identify relationship between two or more pieces of data.
   b. Correlation does not equal causation: 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
   a. Big picture is like looking at a complete puzzle

b. It helps to zoom out and see possibilities and opportunities
c. This leads to new ideas and innovations.
d. Detail oriented thinking is about figuring out all of the aspects that will help you execute a plan

Common questions asked by data analysts:

What is the root cause of the problem?

Root cause: the reason why a problem occurs

The five "why?" helps in solving the root cause. Try to ask five questions about why a problem is occurring and it will help to reveal the root cause for a problem.

Where are the gaps in our process?

Gap analysis: a method for examining and evaluating how a process works currently in order to get where you want to be in the future.

The general approach is to understand where we are and where we want to be.

What do we not consider before?

Helps to identify what information or procedure might be missing from the process. This helps in making better decisions and strategies moving forward.

Data-driven decision-making

Using facts to guide business strategy.

Quartiles: A quartiles divides data points into four equal parts or quarters.

Week 3: **The wonderful world of data**

28-12-2021

Data phases and tools

Data phases are:

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

Data analysis tools include spreadsheets, databases, query languages, and visualization software.

Stages of data lifecycle

- Planning
    - Happens well before starting a project
    - 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.
- Capture
    - Data is collected from a variety of different sources and brought into the organization
- Manage
    - 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
    - Important phase for data cleansing
- Analyze
    - Data is used to solve problems, make great decisions, and support business goals.
- Archive
    - Archiving means storing data in a place where it's still available, but may not be used again
- Destroy
    - Important for protecting a company's private information, as well as private data about it's customers.

Database: a collection of data stored in a computer system.

Phases of data analysis

Ask:

- We define the problem to be solved and make sure that we fully understand stakeholder expectations.
- Defining a problem means looking at the current state and identifying the difference from the ideal state.
- Communicate with the stakeholders to make sure what their expectations are and to stay engaged and on track throughout the project.
- *We want to ask all the right questions at the beginning of the engagement so that we better understand what our leaders and stakeholders need from this analysis.*

Stakeholder: people who have invested time and resources into a project and are interested in the outcome.

Prepare:

- data and results should be unbiased, i.e., they should be focused on facts and be fair and impartial.
- *We need to be thinking about the type of data we need in order to answer the questions that we've set out to answer based on what we learned when we asked the right questions.*

Process:

- cleaning data, transforming it into a more useful format, combining two or more datasets to make information more complete and removing outliers, which are any data points that could skew the information. This phase is all about getting the details right.
- *This is where you get a chance to understand its structure, its quirks, its nuances, and you really get a chance to understand deeply what type of data you're going to be working with and understanding what potential that data has to answer all of your questions.*

Analyze:

- using tools to transform and organize that information so that you can draw useful conclusions, make predictions, and drive information decision-making. Some tools are spreadsheets and structured query language (SQL).

- *This is the point where we have to take a step back and let the data speak for itself.*

Share:

- visualization is essential to getting others to understand what your data is telling.
- *All of this work form asking the right questions to collecting your data, to analysing and sharing doesn't mean much of anything, if we aren't taking action on what we've just learned.*

Act:

- business takes the insights and puts them into work in order to solve the original business problem.

Tools for Data analysts

- Spreadsheets
  - Microsoft Excel
  - Google sheets
  - Spreadsheet is a digital worksheet
  - It stores, organizes, and sorts data
  - It helps to see patterns, group information and easily find the information
  - They have useful features called formulas and functions to perform calculation using the data and perform preset commands on data.
- Query languages for databases
  - A programming language that allows to retrieve and manipulate data from a database.
    - SQL (Structured Query Language)
      - Easy to use and works well with all kinds of databases.
  - Allows analysts to isolate specific information from a database
  - Make it easier for you to learn and understand the requests made to database
  - Allow analysts to select, create, add, or download data from a database for analysis
- Visualization tools
  - The insights from data are visualized so that complex data can be understood easily and shared.
    - Tableau
    - Looker

Spreadsheets	Databases
Software applications	Data stores – accessed using a query language
Structure data in a row and column format	Structure data using rules and relationships
Organize information in cells	Organize information in complex collections
Provide access to a limited amount of data	Provide access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a data base management system

Week 4: **Set up your toolbox**

29-12-2021

Spreadsheets

		This	
This is a cell		Is	
This	is	A	row
		Colum	

Attribute: a characteristic or quality of data used to label a column in a table.

Observation: all of the attributes for something contained in a row of a data table.

Formula: a set of instructions that performs a specific action using the data in a spreadsheet

https://support.google.com/a/users/answer/9282959?visit_id=637361702049227170-1815413770&rd=1

https://support.google.com/a/users/answer/9300022

https://support.microsoft.com/en-us/office/excel-for-windows-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb

30-12-2021

SQL

SQL enables data analysts to talk to their databases. It is one of the most useful data analyst tools, when working with large datasets. It can help investigate huge databases, track down text and numbers, and filter for the exact kind of data you need – much faster than a spreadsheet can.

- Store
- Organize
- Analyze

Query: a request for data or information from a database

Syntax: is the predetermined structure of a language that includes all required words, symbols, and punctuation, as well as their proper placement.

Basic structure of a SQL query:

```
SELECT
      [choose the column(s) you want]  #2
FROM
      [from the appropriate table]     #1
WHERE
      [a certain condition is met]     #3
```

This is the suggested order in which you write your SQL queries. Start big (data table) and go small (specific conditions).

https://www.w3schools.com/sql/default.asp

https://towardsdatascience.com/sql-cheat-sheet-776f8e3189fa

Data Visualization

Data visualization is the graphical representation of information. Data visualizations are pictures. They are a wonderful way to take very basic ideas around data and data points and make them come alive.

https://public.tableau.com/en-us/s/resources

https://rstudio.com/

https://rstudio.com/resources/cheatsheets/

Week 5: **Endless Career Possibilities**

30-12-2021

Issue: a topic or subject to investigate

Question: designed to discover information

Problem: an obstacle or complication that needs to be worked out

Business Task: the question or problem data analysis answers for a business

Fairness: ensuring that your analysis doesn't create or reinforce bias

Decoding the job description

	Data Analysts	Data Scientists	Data Specialists
Problem solving	Use existing tools and methods to solve problems with existing types of data	Invent new tools and models, ask open-ended questions, and collect new types of data	Use in-depth knowledge of databases as a tool to solve problems and manage data
Analysis	Analyze collected data to help stakeholders make better decisions	Analyze and interpret complex data to make business predictions	Organize large volumes of data for use in data analytics or business operations
Other relevant skills	• Database queries • Data visualization • Dashboards • Reports • Spreadsheets	• Advanced statistics • Machine learning • Deep learning • Data optimization • Programming	• Data manipulation • Information security • Data models • Scalability of data • Disaster recovery