

Foundations of Data Analytics

Saturday, January 28, 2023 22:20

Get started

Tuesday, September 6, 2022 20:45

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.

Processes for Data Analysis :

Ask - Prepare - Process - Analyze - Share - Act

Entry for the learning log :

Create a list of at least 5 questions that you might use data to answer. (situations in which you may use data to make decisions)

1. Which roads should I take to get to work quickly?
2. Should I buy coffee every day?
3. What is the best time to work on a difficult task?
4. What is the best seating arrangement at my next dinner party?
5. Which jobs should I apply to to have the best chances at starting a new position?

Choose one question

- What are some considerations or preferences you want to keep in mind when making a decision?

Which jobs would I actually like doing if I got an offer?
Which jobs am I qualified for? Do I know anyone in my who works at this company? Will there be a lot of other applicants?

- What kind of information or data do you have access to that will influence your decision?
 - job description, public records about the company, LinkedIn

- job description, public records about the company, LinkedIn features, my own qualifications
- Are there other things you might want to track associated with this decision?
 - Applicant pool, # of applicants, interview process, compensation

Transforming data into insights

Wednesday, September 7, 2022 20:56

Data Analytics in everyday life :

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

Analysis: Turning data into insights.

↳ might help a company to develop a new product or a new form of customer experience

Reading "New data perspectives"

Case Study Q: how can the organization improve the retention rate for new employees?

① Ask effective questions

- eg.: what do new employees need to learn to be successful in their first year on the job?
- What do you suspect is the leading cause of dissatisfaction among new employees?
- etc.

② Prepare

Build a timeline, establish communication channels, identify what data is needed to succeed, how to best present the data visually, what problems might come up and how to avoid them, establish rules on who will have access to the data etc.

③ Process

for their Survey: obtain consent from participants (employees), explain to participants how their data would be collected, stored, managed, and protected.

- restricted access to the data
- aggregated data to not reveal individual responses
- use internal data warehouse to store raw data

④ Analyze

- discovered that certain employee experiences are indicators of job satisfaction
- document each finding, no matter what the result

⑤ Share

- communicate the results with the right context
- share findings with a relevant group of people

⑥ Act

- implement changes and take actions based on the findings
e.g.: standardize hiring procedure, conduct the same survey annually to compare results



Learning Log what you've learned from the case study

- Did the details of the case study help to change the way

- Did the details of the case study help to change the way you think about data analysis? why or why not?

Yes, the case study helped me realize that a data scientist's job is not just to analyze data, but also to store & share it responsibly

- Did you find anything surprising about the way the data analysts approached their task?

I thought it was interesting that they primarily relied on survey data rather than some other internal company data source

- What else would you like to learn about data analysis?

How to actually analyze the data, run regressions and find correlations, how to create a good survey and maximize survey responders

Understanding the data ecosystem

Wednesday, September 7, 2022 21:53

Data ecosystems :

The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data.

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

vs

Data Analytics

- Find answers to existing questions by creating insights from data sources

Data Analysis ≠ Data Analytics

"the science of data"

Data-driven decision making:

Using facts to guide business strategy

① Figure out the business need

eg. how to source parts from a more sustainable supplier

② Data analyst finds data, analyzes it and uses it to uncover trends, patterns, and relationships.

To get the most out of data-driven decision making, it's important to include insights from people who are familiar with the business problem.

Embrace your data analyst skills

Saturday, September 10, 2022 16:57

Analytical skills:

Qualities and characteristics associated with solving problems using facts.

Key data analyst skills:

- ① Curiosity
- ② Understanding Context
- ③ Having a technical mindset
- ④ Data design - how you organize information
- ⑤ Data strategy - the management of the people, processes and tools used in data analysis

Learning Log: Explore data from your daily life

Pick one area of your everyday life and collect some data on it

- Distance I've moved in the past week

	Walking	Biking
9/3	0.4 mi + 1.9 mi + 0.7 mi 9 min + 58 min + 19 min	
9/4	0.9 mi 39 min	
9/5		1.1 mi + 1 mi 7 min + 7 min
9/6	2 mi + 1.7 mi 29 min + 70 min	3.1 mi + 3.3 mi 38 min + 36 min
9/7	1.6 mi + 1.9 mi 60 min + 60 min	3 mi + 3.3 mi 24 min + 26 min
9/8	2.3 mi + ??? 68 min + 76 min	3.1 mi + 3.2 24 min + 26 min

9/8	2.3 mi + ???	3.1 mi + 3.2
	68 min + 76 min	24 min + 26 min
9/9	2.6 mi + 0.2 mi + ???	4 mi + 3.4 mi
	76 min + 14 min = 53 min	20 min + 25 min
9/10	1.6 mi + 2.4 mi	
	47 min + 57 min	

- Are there any trends you noticed in your behavior?
 - biking on weekdays because I'm going to work
 - highly dependent on my dog
 - extra walks when I get lunch at work bc. lunch place is far away
- Are there factors that influence your decision making?
 - work, dog, reason to leave the house (ie going shopping)
- Is there anything you identified that might influence your future behavior?
 - Set myself tasks I normally do on weekends so I use the bike
 - buy small items at workplace canteen to get extra steps in on workdays

Thinking about analytical thinking

Saturday, September 10, 2022 20:00

Analytical thinking :

Identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

The 5 key aspects to analytical thinking :

- ① Visualization
- ② Strategy
- ③ Problem orientation
- ④ Correlation
- ⑤ Big-picture and detail-oriented thinking

Visualization - the graphical representation of information

- graphs, maps, other design elements help DA's understand and explain information more effectively

Strategy

- stay focused and on track, what do we want to achieve and how can we get there, improves quality and usefulness of the data we collect

problem-oriented - identify, describe, and solve problems, keep problem top of mind, ask a lot of questions

Correlation

- finding relationships between variables
≠ causation

big picture

- being able to look at the whole, finished puzzle without focusing too much on the details, zoom out and see opportunities and possibilities

details, zoom out and see opportunities and possibilities

detail-oriented - figuring out all the aspects that make the plan succeed (puzzle pieces)

Think **CRITICALLY**

Think **CREATIVELY**

Think **ANALYTICALLY**

A question that often comes up on the hunt for a solution:

what is the root cause of a problem?

Root Cause: The reason why a problem occurs

To reveal the root cause, **ask "why?" five times.** (The five whys)

Eliminating the root cause eliminates the problem.

Another question: **Where are the gaps in our process?**

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

> understand where you are now compared to where you want to be, identify the gaps, determine how to bridge them

What did we not consider before?

> what information or procedure might be missing from a process?

Learning Log

<u>Analytical Skill</u>	<u>Level</u>	<u>Comments</u>
Curiosity	Developing	need to be more persistent in asking questions and not accepting the status quo
Context	Developing	need to make more connections

		accepting the status quo
Context	Developing	need to make more connections between subjects and think about more stakeholders
Technical mindset	Strength	
Data Design	Strength	
Data Strategy	Developing	still don't know much about data structures and storage

Thinking about outcomes

Sunday, September 11, 2022 10:53

Data-driven decision-making:

using facts to guide business strategy

- gain valuable insights, verify theories / assumptions, better understand opportunities and challenges, support an objective, help make a plan etc.

Real-life example

Google's HR department wanted to know if there is value in having managers. Google's people analytics team looked at past performance reports and employee surveys. Data was plotted on a graph.

> Googlers had positive feelings about managers

> needed to dig deeper

Divided data into quartiles (four equal parts)

> Revealed BIG difference between top and bottom quartiles

> teams with best managers were significantly more productive & happier

» managers were valued and make a big difference

»» what makes a great manager? (need actionable insight)

They launched an awards program to identify best managers, the submission had to specify what made the manager great

Then they interviewed managers from the top and bottom quartiles

> identified best manager behaviors and most common reasons for needing improvement

Shared these insights and put in place manager evaluations with these qualities in mind

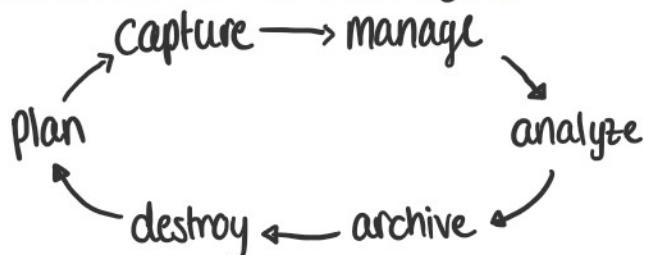
Follow the data lifecycle

Sunday, September 11, 2022 11:33

Phases of data analysis :

- 1 Ask: Define the problem and confirm stakeholder expectations
- 2 Prepare: Collect and store data for analysis
- 3 Process: Clean and transform data to ensure integrity
- 4 Analyze: Use data analysis tools to draw conclusions
- 5 Share: Interpret and communicate to others to make data-driven decisions
- 6 Act: Put your insights to work to solve the original problem

Stages of the data life cycle :



Planning: what data is needed?

how will it be managed throughout its life cycle?

who will be responsible for it?

what are the optimal outcomes?

Capture: data is collected from a variety of different sources and brought into the company
from outside sources (publicly available datasets)
from a company's own documents and files from a database

Manage: how we care for our data
how and where it is stored
the tools used to keep it safe and secure
actions taken to maintain it properly

Analyze: data is used to solve problems, make great decisions, and support business goals

- decisions, and support business goals
- archive:** storing data in a place where it is still available but may not be used again
- destroy:** use a secure data erasure software, shred paper files
important for protecting a company's private information

Outlining the data analysis process

Sunday, September 11, 2022 12:11

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

① Ask

Defining the problem: look at the current state and identify how it's different from the ideal state

Understanding stakeholder expectations: determine who the stakeholders are and what they expect from the project
> making sure you stay on track and engaged
> developing strong communication skills is important
five why's can be helpful here

② Prepare

Collect and store data: which kinds of data to collect to get an objective and unbiased result

③ Process

Find and eliminate errors and inaccuracies: cleaning data, transforming data, combining datasets, complete information, remove outliers, fix typos, missing or inaccurate data, verify data cleansing

④ Analyze

Use tools to transform and organize the data to draw useful conclusions
make predictions, and drive informed decision-making
spreadsheets and structured query language (SQL)

⑤ Share

interpret results and share them with others: visualization is essential, need presentation skills

⑥ Act

Business takes the data analyst's insights and puts them to work to solve the original problem

The data analysis toolbox

Sunday, September 11, 2022 13:06

Spreadsheets • Microsoft Excel Google Sheets

A **spreadsheet** is a digital worksheet.

It stores, organizes, and sorts data.

Spreadsheets have **formulas** and **functions**.

Formula: a set of instructions that performs a specific calculation using the data in a spreadsheet.
add, subtract, multiply, divide, find averages,
lookup values, return the sum of a set of values, etc.

Function: A preset command that automatically performs a specific process or task using the data in a spreadsheet.

Query language: A computer programming language that allows you to retrieve and manipulate data from a database.

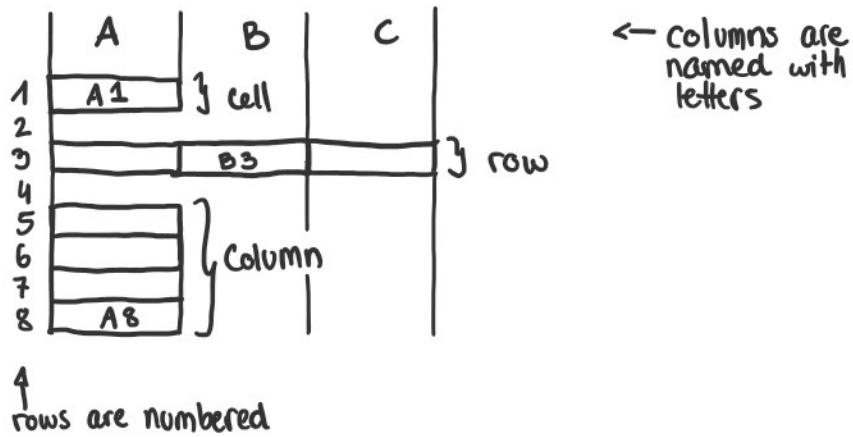
Structured Query Language (**SQL**) lets analysts communicate with a **database**. Insert, delete, update data.

Database: A collection of data stored in a computer system

Data visualization • Tableau
• Looker

Mastering spreadsheet basics

Monday, September 12, 2022 13:51



Text wrapping feature → set to overflow, wrap, or clip
to adjust how much text is visible

Labelling columns makes it easier to reference and find
data later on

» **Attribute**: A characteristic or quality of data used
to label a column in a table. = column names
column labels, headers, the header row

Row = **Observation**: All of the attributes for something
contained in a row of a data table.

to type formula in spreadsheet:

= C2+C3+C4 to sum cells C2-C4

Structured Query Language (SQL)

Monday, September 12, 2022 15:19

- SQL - store, organize, analyze data
- SQL needs a database that will understand its language
- Databases that use SQL:
 - Oracle, MySQL, PostgreSQL, Microsoft SQL Server
- in SQL, queries are universal

QUERY: A request for data or information from a database

Basic structure of a SQL query:

SELECT

[choose the columns you want]

#2

FROM

[from the appropriate table]

#1

WHERE

[a certain condition is met]

#3

suggested order

multiple columns and conditions in a query:

SELECT

column A
column B
column C

FROM

dataset-name.table-name

WHERE

Condition 1
AND condition2
AND condition3

Some SQL statements will be terminated with a **semicolon**. The semicolon is part of the **ANSI SQL-92 Standard**, which recommends a common syntax for adoption by all SQL databases, but not every database is enforcing it. So if a statement works without the semicolon that's fine.

WHERE conditions

WHERE field1 = 'Chavez' looks for an exact match.

WHERE field1 LIKE 'Ch%' looks for all values starting with 'Ch'. % is used as a wildcard. In some databases a * is used instead of %.

Use * (as in SELECT * to select all columns) sparingly.

Comments

Comments are text placed between /* comment */ or after -- (two dashes)

Aliases : assigning a new name (or alias) to column or table name to make them easier to work with using an SQL AS clause. Aliases are good for the duration of the query only and don't change the actual name of a column or table in a database.

field1 AS last_name

table AS customers

<> means "does not equal"

Data Visualization

Monday, September 12, 2022 20:42

Steps to plan a data visualization

- 1 Explore the data for patterns
- 2 Plan your visuals
 - want to explain findings quickly and effectively
- 3 Create your visuals
 - Line charts can track sales over time
 - Maps can connect sales to locations
 - Donut charts can show customer segments
 - Bar charts can compare total visitors and visitors that make a purchase

You can use the visualization tools in your spreadsheet for simple visuals, more advanced tools such as Tableau, or visualization tools in R Studio, depending on the size of your data and the process used for analyzing.

- Spreadsheets - built-in charts, simple, bar graphs, pie charts, maps, waterfall & funnel diagrams
- Visualization software (Tableau) - allows you to interact with the data, built-in visual best practices make analyzing and sharing data fast, easy, and useful
- Programming language (R with R Studio) - dashboard-style data visualizations, useful, if data analysis was done in R

Let's get down to business

Monday, September 12, 2022 21:27

The job of a data analyst

- Technology
- Marketing
- Finance
- Healthcare

Coca Cola uses data from consumer feedback to create advertising that speaks directly to different audiences with different interests. Coca Cola vending machines have built-in AI and data analysis tools that help Coca Cola see how customers are combining flavors which can give CC inspiration for a new product.

Small businesses can use data analytics to better understand their customers' buying habits, create more effective social media messaging, or predict the number of daily visitors based on the weather. In the healthcare industry, data analysts look at clinic attendance to help hospitals and doctor's offices predict when rush hours will hit so they can be ready for it.

Learning Log

- which parts of the data analysis process did you enjoy the most?
what did you enjoy about it?
 - process - I like formatting spreadsheets, making them look "pretty"
 - visualization - I am artistically inclined and like creating visuals

The importance of fair business decisions

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

Sometimes, conclusions based on data can be true and unfair
→ concluding from data that shows that only men are successful in a certain company that more men should be hired because they're the most successful

instead, one could conclude that something about the company culture prevents non-men from succeeding

Data Ethics: what is the good and right way of using data?
not just about minimizing harm but about how do we improve the lives of people by using data. Are the people represented in this data going to be benefited by this? At the end of the day, data often comes from people, and data analysts have a responsibility towards these people. How to keep their data private and give them control over it?