

Welcome to CDAC

- Data is everywhere
- Data is a collection of facts, these include numbers, pictures, videos, words, observations and more
- Data Analysis is collection, transformation and organization of data in order to draw conclusions, make predictions and drive informed decision-making.
- Data evolves over time, means this analysis can give us new information throughout data's lifecycle.
- Reading reviews before buying a product
- Keeping track of fitness statistics
- This is data analysis
- We create huge amounts of data every single day.
- 40,000 Searches per second \Rightarrow 3.5 B Every day
1.2 T Every year
2B YT users

Data Analysis is someone who performs 

High Demand

ASK
Prepare
Process
Analyze
Share
Act

Cleaning data doesn't require
Soap & water

Complete, correct & relevant

R1

Designed to make any type of learner a entry level junior data analyst / associate data analy

- > In future DA will be highest demanding career 10.9% vs 5.2%
- > All types of companies all over the world need business analysts to make best possible business decisions
- > Opportunity to create case study which you can highlight in your resume, to show what you learned

What to expect
Expect

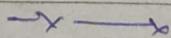
Reading: to introduce new idea & build from video

& Pay attention to video questions

16/12
R2:

"Learning Log" Everyday

Everyday
Reflection
Give & Response



Success =

Computer + Your Brain + Skills + Traits

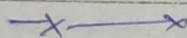
V2:
Finance
Healthcare
Communication
Government
Consumer product
Etc

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

What after the analysis is complete (Insights)

SHARE with others, decisions are made & businesses take action.

- > Help organization completely rethink what they do or guide them in new direction.



R: Human Resource Analytics / Workforce Analytics / People Analytics

: Practice of collecting & analyzing data on people who make up for a company's workforce

6 steps of DA Process: → Ask

Prepare
Process

Analyze

Share

Act

APPASA



3 Pillars of DS: Analytics, ML & statistics

THE UNKNOWN

↓
Don't
know

↓
Total
O

↓
Rev'D

Understanding the data Ecosystem

Data Ecosystems

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

Also includes hardware, software & people,
like me

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

Cloud, term used to address a virtual location
where the data is stored.

The job is to harness the power of data ecosystem
& find right information & provide team with
analyses that helps them make smart decision

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

*→ Data scientists create new questions using data
*→ Data Analyst answers existing questions by creating
insights from data.

Data Analysis & Data Analytics are Different

Page No.

Date

Data Analysis:

The collection, transformation & organization of data in order to draw conclusions, make predictions & drive informed decision-making

Data Analytics: → THE SCIENCE OF DATA

It's a broad concept that encompasses everything from job of managing and using data to tools and methods that data workers use everyday.



It's a umbrella term:

Data, Data analysis, data ecosystem all fit under it



v2

* Data alone will never be as powerful as data combined with human experience, observation & sometimes even intuition

Insight from people who are familiar with business problem are more valuable



These people are called "Subject Matter Experts" & they have ability to look ~~out~~ of DA & identify any inconsistencies, make sense of gray areas, and eventually validate choices being made

Don't focus on gut instinct if you can't back it up with data.

Page No.	
Date	

* Data & gut instinct:

- > The more you understand data related to a project the easier it will be to figure out what is required.
- & past experience may form connection that no one else would notice, so it's not just gut instinct.

* Data + Business knowledge = mystery solved.

Try asking yourself:

- How do I define success for this project?
- What kind of results are needed?
- Who will be informed?
- Am I answering the questions being asked?
- How quickly does a decision need to be made?

Ex. Rush projects depend on own knowledge & Exp more than usual

cause there's not enough time to analyze all available data.

Plenty time & resource: data-driven, blend data & knowledge in a lot of ways to make best possible choice

Practice → Perfect Blend.

Google DAI Certificate

	Dell EMC's	SAS's
1. Ask	Discovery	Ask
2. Prepare	Pre-processing data	Prepare
3. Process	Model planning	Explore
4. Analyze	Model building	Model
5. Share	Communicate results	Implement
6. Act	Operationalize	Act
7.	↓	Evaluate

Dell EMC

David Dietrich

cyclical

cyclical

return to ask again.



Project-based data analytics life cycle

1. Identifying the problem
2. Designing data segment
3. Pre-processing data
4. Performing data analysis
5. Visualizing data

developed by Vignesh Prabhati

(doesn't include act phase)

Brg data analytics life cycle
Created by Thomas Grl, Paul Buhler

1. Business case evaluation
2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation & representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

prepare &

process are broken into detailed steps hence more than 6 steps

Analytical Skills

- > You already possess analytical skill, you might just not know yet
- > People often overlook their existing skills while learning something new.
- > But it's important to ack them, you're probably more prepared than you think.

Defn: Qualities and characteristics associated with solving problems using facts

5 Essential points: →

- ① Curiosity ✓
- ② Understanding context ✓
- ③ Having a technical mindset ✓
- ④ Data Design ✓
- ⑤ Data Strategy ✓

Technical Mindset:

Ability break things down into smaller steps or pieces that work with them in an orderly and logical way

paying
bills

contacts
arrangement

mowing
lawn

Data Design: How you organize information

Data Strategy:

Management of people, processes and tools used for data analysis.

(you, people, processes, tools, key, gas)

If you only focus on individual pieces, you wouldn't be able to see past that!

Zoom-out & zoom-in

Page No.	
Date	

- > People don't think about thinking, thinking is second nature to us
- > It just happens automatically!
- > There are many different ways to think
 - > Creative Thinking
 - > Critical Thinking
 - > Abstract Thinking
 - > Analytical Thinking

Analytical thinking :-

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

* 5 key aspect to analytical thinking

1. Visualization : Graphical representation of information

2. Strategy

3. Problem-Orientation : Identify, describe & solve problem

4. Correlation : Btw 2 or more pieces of data

5. Big-picture and detail-oriented thinking

* Data analysts ask a lot of questions, this helps improve communication and saves time while working on a solution

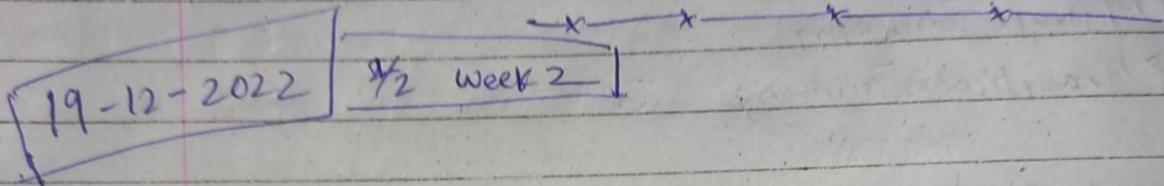
Ex- Surveying customers about their experiences using a product and building insights from those question to improve that product.

* Correlation does not equal causation, two pieces of data trending in same direction doesn't mean they're all related.

necessarily

Q A Data visualization includes/involves using graphs, maps & chart for representing present data.

Q A To execute a plan using detail oriented thinking a DA considers the specifics.



* Exploring Core Analytical Skills :-

- > Some type of thinking might not come to you easily
- > But you can absolutely grow and develop them.
- > This means you can become a versatile thinker which is an import aspect of data analysis.
- > The more ways you can think the easier it is to think outside the box & come up with fresh ideas.
- > Why its important to think in different ways? cause the solutions are almost never in front of you.
- > You need to critically to find out right questions to
- > And creatively to get new & unexpected answers.

Questions that DA ask often?

- What is the root cause of the problem?

Root cause: The reason why a problem occurs.

The 5 why's: Ask 5 "why?" to reveal root cause

Q. Want to make blueberry pie but couldn't find blueberry to make any pie!

- ① Why can't I make any pie? No blueberry at store
- ② Why there are no blueberries at the store?
⇒ Blueberry don't have enough fruit this season
- ③ Why was there not enough fruit?
⇒ Birds were eating all the berries
- ④ Why were birds eating blueberry?
⇒ Although birds don't like blueberry, they like mulberries but mulberries didn't produce any fruit.
- ⑤ Why mulberries yield this season?
⇒ The late frost damaged the mulberry bushes.

Finally: So we can't make blueberry pie cause of late frost months ago.

Another question commonly asked by DA:
"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 future.

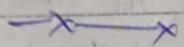
Basically: where you are now vs where you want to be?

3rd most common question:

"What did we not consider before?"

Great way to think ↓ information or procedure might be missing from process to make better decisions and strategies moving forward.

- * The way DA asks & answers impact business decisions.



* Data driven decision-making is nothing but using facts to guide business strategy.

Real world Examples →

Google > manager, is it worth keeping them?
→ most of staff had fine feeling ☺

In sufficient data

Wanted to learn more
Split data into quartiles

Quartiles divides data into four equal parts

There was a big difference in observations of top & bottom quartiles.

- ① Team with best managers, were more happy productive & wanted to keep working at google
- Turned out its worth keeping mangers.

Now, what is a great manager?

→ They took 2 additional steps →

- ① Launched award program, where employees had to provide data such as "what makes that manager great?"
- ② Interviewing managers grouped on top 2 bottom quartiles.

Finally sharing insights & putting a procedure in place for evaluating managers with these qualities in mind.

— x — x

1/3 week 3

20/12/22

Follow Data Life Cycle; →

Bring data to life:

Includes spreadsheets, databases, query languages and visualization softwares.

Six DA phases → APP ASA

Analyze Share Act

Stages of Data Life Cycle →

Planning → Capture → Manage → Analyze

↓

Achieve

↓

Destroy

Variations of DLC

Google DAC

- ① plan : Decide type of data reqd, how to manage, who uses it
- ② Capture : collect from diff sources
- ③ manage : maintain data, store & tools used
- ④ Analyse : use data to solve problems, make deci, support Bgrnd
- ⑤ Archiver : keep relevant data for future reference
- ⑥ Destroy : Remove data from storage & any shared copies of

→ X → 2/3 weeks

21/12/22 SIX phases of Data Analysis ↗

ASK

Stakeholders :

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

Defining a problem ↗

Look at the current state and identify how it's different from the ideal state.

Understanding stakeholder expectations

- ① Determine who the stakeholders are
- ② Communication is the key

Prepare

- Collect & store data
- we'll further learn different type of data & which types of data are most useful.
- Data & results should be **UNBIASED**, any decisions you make must be based on facts, be fair & impartial.

Process

- > Eliminate errors & inaccuracies (typos, missing, inaccurate, etc)
- > Clean, transform into useful format
- > Combining two or more datasets to make info more complete, removing outliers (they skew the info).

Analyze

- > Using tools to transform & organise the information
- > To draw useful conclusions, make predictions & drive informed decision making.
- > we'll see spreadsheets and SQL

Share

- > Why visualization is so important
- > With right visuals, facts & figures it becomes much easier to understand.
- > we'll explore different kinds of visuals and tools.
- > we'll also practice presentation skills by creating compelling slideshows, & be fully prepared to answer questions. > we'll see [R]

Action

- > Business takes the insights provided by DA & puts them to use.
- > prepare for job search
- > Complete case study project. (Helps stand out).

Y

— X — X —

3/3 week 3

Molly: Example of the data process →

First of all ask the eight questions →

- ① what is the problem that we're trying to solve?
- ② What is the purpose of this analysis?
- ③ what are we hoping to learn from it?

Now the preparation →

- ④ what type of data we need to answer these key questions
- ⑤ Qualitative or quantitative
- > Cross sectional or points in time VS
longitudinal over a period of time
- > Do we need to collect data?
- > How do we collect?

Process: The most fun part according to miss molly:

- ④ Cleaning : Handshake to your data
- we understand its structure, quirks, nuances
- and we get the chance to deeply understand what potential that data has to answer your questions.
- Is the data correct or it has inconsistencies, missing, outliers.

Analyze In the most objective & unbiased way possible.

- As we get better and better at our analysis job we begin to intuit and expect from data.
- This is where we need to take a step back and let the data speak for itself.

Share:

- High level planning with executive team
- Melting, presentation
- Communication & interpersonal skills

Act All of this doesn't mean anything if we aren't taking action on the insights.

The whole process is rigorous & lengthy!
 The import aspect for analysis is they get excited about diving deep into data to find out its capabilities & potential.

Spreadsheets, SQL, normalization

↓ to fetch large amounts of data in seconds.



Spreadsheets:

The sooner you understand it, the easier time you'll

have!

	A	B	C	D
F			This	
2	This is a cell		is	
3	This	is	a	row
4			Column	

Damn

columns → vertically, ordered by letter
rows → horizontally, ordered by number

locate a cell → col-letter + rowNumber

example "d" is present at cell C3

Length & height of cell [format]

Column labels are called attributes

Tony Francis & The cool instructor

Page No.	
Date	

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

Attributes are also referred to as by column name, column label, headers or header row.

A row is also called as observation

To sort > select Required
> Data tab > sort

Formula \Rightarrow A set of instructions that perform a specific action using the data in a spreadsheet

formulas are always preceded with "=" sign

$$= \text{C2} + \text{C3} + \text{C4} = \underline{10} \quad \text{cool}$$

SQL

why?

If you have a very large dataset then

No matter what database you use SQL basically works the same for each.

SQL queries are universal.

+

A request for data/information from a database

Basic structure of SQL query →

SELECT column
FROM Table
WHERE Condition

To select all of data →

SELECT *

FROM movie_data.movies ;

Adding where is

SELECT *

FROM movie_data.movie

WHERE genre_id = 'Action';

we get only movies with 'Action' genre.

④ WHERE is used to filter the data

We can have single as well as multiple columns returned by a query.

```

SELECT col1, col2, col3...
FROM T-name
WHERE condition1
    AND condition2
    AND condition3 ... ;
  
```

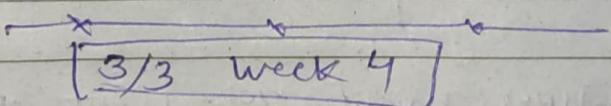
Also there can be more than one condition separated by ~~commas~~ operator.

- - - - -

~~Angie~~ Angie :

- "Everytime I'm learning something feels like I'm learning to speak all over again"

- know that "you can ask for help"

Visualization :  [3/3 week 4]

Graphical representation of information.

For most Data analyst this is the most exciting part of the job, as they get to see their hard work pay off with something interesting.

Florence Nightingale → Philosophy of Modern Nursing.

→
was a Data Analyst

During 1850

Crimian War

1000s of
soldiers
died each day

Created a visualization on paper to demonstrate
soldiers dying vs Existing.

Turned out most soldiers died ~~existing~~ due to
preventable causes.

The visualization had data of several months
blue section showed preventable deaths.

we can explain spreadsheets to someone,
but Visualization charts, graphs, maps would
econ as a better approach.

we must keep visualizations easy to understand
and interesting to look at

stakeholders might not want to see the data, so
our job is to make them time worthwhile

→ Bar chart

Just a few steps away from next course

Keep
it
opt

Lilah :

Data visualizations are pictures. They are wonderful way to take very basic ideas around data and data points and make them come alive.

Stakeholder : Should we open a new branch in mumbai?

Data Analyst : Here's why we should backed by visualizations & data points.

13 weeks

24/12/22

Role of DA

Business tasks to DA

Fairness in analysis

Opportunities for you

And your future success

Coca-cola : Advertisements based on consumer feedback

② vending machine (make your flavour)

: Built in AI & DA tools in machine

* data goes to Coca-cola about which flavour consumer like most

Using Data Analytics gives edge over things like:

Understanding consumer buying pattern, effective SMM Messaging

Visitors based on climate (Zoo/Aquarium)

Tony

finance program
manager at google

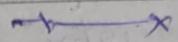
Page No.

Date

zoo; rainy days, predict based on past data

Hospital; long wait times

Data analysis for how many doctors needed
when etc.



Job:

"In analytics, I feel like the key to success is being able to blend the personal side with the technical side."

You need to grow them equally.



first 2 years → Be a generalist

Next 2-5 years → A specialist

(Tony):

"There's so many different paths that you can take from the starting point that you really can't predict your end"