ANALYTICS 101

PAM NAGARAJAPPA,
SEPTEMBER 26, 2024

SEPTEMBER 26, 2024, 9 AM TO 12 PM

DATA CONCEPTS (9 TO 9:20)

QUESTIONS (9:25 TO 9:30)

ANALYSIS CONCEPTS (9:30 TO 9:50)

QUESTIONS (9:50 TO 9:55)

BREAK (9:55 TO 10:05)

WORKSHOP (10:05 TO 11:30)

QUESTIONS AND WRAP-UP (11:30 TO NOON)

ANALYSIS AGENDA

WHY ANALYZE DATA?

TYPES OF DATA ANALYSIS

DATA CLEANING PROCESS

SUMMARY

WHY ANALYZE?

- Scenario:
  - **I am a store manager and I want to improve the sales through understanding customer behavior.**
  - To do this, I want to understand on average how much are my customers spending and the products they are buying.
  - The analyst is tasked off with this job.
  - **Data driven approach avoids subjective answers.**
- What does the analyst do?
  - They collect customer sales and product data.
  - They analyze the data to provide the necessary answers to the manager.
  - They provide a report or dashboard to the manager containing the necessary information.

TYPES OF ANALYSIS TO HELP ORGANIZATIONS MAKE DECISIONS

01

Descriptive analytics:

- helps answer what, when and where

02

Diagnostic analytics:

- Helps answer why

03

Predictive:

- Help forecast future

04

Prescriptive:

- Given the forecast, helps answer what if

Higher complexity

Diagnostic, Predictive and prescriptive are advanced analysis methods. They will not be discussed as part of this course.

ANALYSIS TYPES
TO HELP ORGANIZATIONS DRIVE DECISIONS

Descriptive

What is happening now or has happened in the past

Detailed account of past events, such as sales figures or customer demographics

Diagnostic

Identify underlying factors that contributed to a particular outcome or event

Help the retailer identify factors such as new marketing campaigns that contributed to the increase in sales.

Define Problem Statement

- What business problem is being solved?

Select Data

- Find relevant data that can answer the question

Clean Data

- Make data analysis ready

Analyze

- Discover patterns
- Interpret results
- Report

ANALYSIS EXAMPLE

- Scenario:
  - **I am a store manager and I want to improve the sales through understanding customer behavior.**
  - To do this, I want to understand on average how much are my customers spending and the products they are buying.
  - The analyst is tasked with this job.
  - **Data driven approach avoids subjective answers.**

WE WILL LOOK AT SOME DATA CLEANING TASKS

Data cleaning

- Identify and correct errors
- Ensure data is complete and accurate for reliable analysis

Common tasks

- Handle missing values
- De-duplicate data
- Data validation
- Data transformation
- Aggregation
- Remove invalid values
- Standardize/Normalize data
- Assess data quality

DATA CLEANING: HANDLE DUPLICATE VALUES

Original data

(age was incorrectly entered in row 3)

Customer ID	Name	Age
1	John	25
2	Jane	30
3	John	28

After de-duplication

Customer ID	Name	Age
1	John	25
2	Jane	30

Duplicate values are normally removed.

DATA CLEANING: DATA TRANSFORMATION

Tables before transformation

Address table

Customer Id	Name	Address	City	State
1	John Smith	123 Main St	New York	NY
2	Jane Doe	456 Elm St|	Chicago	IL
3	Bob Johnson	789 Park Ave	Los Angeles	CA

Phone table

Customer Id	Name	Phone
1	John Smith	555-1234
2	Jane Doe	444-5678
3	Bob Johnson	333-9012

Multiple tables must be joined for easier information retrieval and comparison.

DATA CLEANING: DATA TRANSFORMATION

Table after transformation

AddressPhone Table

Customer Id	Name	Address	City	State	Phone number
1	John Smith	123 Main St	New York	NY	555-1234
2	Jane Doe	456 Elm St	Chicago	IL	444-5678
3	Bob Johnson	789 Park Ave	Los Angeles	CA	333-9012

Address and phone numbers are now merged into one table

DATA CLEANING: AGGREGATION

Sales by product table

Customer ID	Product code	Sale price
1	C0005000	100
2	C0002011	200
1	C0005438	100
4	C0007832	50

Aggregate sales by customer table

Customer ID	Total sale price
1	200
2	200
4	50

Sum, average and median are some ways to aggregate the data.
Example: Aggregation provides information on high value customers.

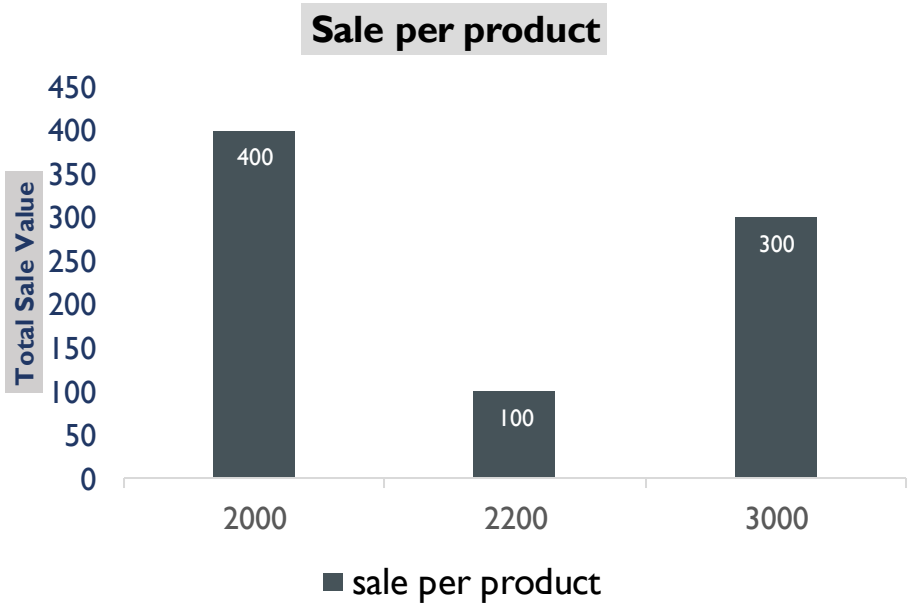
DATA CLEANING SUMMARY

- The purpose of data cleaning is to present data that is suitable for analysis
  - Cleaning can include data aggregation
  - We want to make the data "analysis ready"
- The analysis must not be influenced by invalid, missing, duplicate values
- Larger the data, better is the analysis.
- With the current compute power, companies can process millions of rows of data.

Data size is smaller than what is normally used. This data is shown for illustration only.

ANALYST DISCOVERS CUSTOMER BEHAVIOR

Id	Sale price	Product code
1	200	2000
2	300	3000
3	200	2000
1	100	2200



WHY ANALYZE?

- Scenario:
  - **I am a store manager and I want to improve the sales through understanding customer behavior.**
  - To do this, I want to understand on average how much are my customers spending and the products they are buying.
  - The analyst is tasked off with this job.
  - **Data driven approach avoids subjective answers.**
- What does the analyst do?
  - They collected customer sales and product data.
  - Analyst report: states the problem statement, data collection, cleaning and analysis steps. Captures insights in the report.
  - Analyst report summary: On average customers are spending $X and Product code 2000 seems to be generating the most value and bought the most.

SUMMARY

- DATA ANALYSIS IS USED TO ENABLE DATA-DRIVEN DECISION MAKING
- WE LOOKED AT DESCRIPTIVE ANALYTICS TODAY
  - ANSWERS THE WHAT, WHERE AND WHEN
- ANALYTICS SKILLS INVOLVE SOME KNOWLEDGE OF MATH AND DOMAIN UNDERSTANDING

5 MINUTES FOR QUESTIONS

9:50 TO 9:55 AM

QUIZ

- What are the four steps in a data driven problem solving approach?
- What are the different types of data analysis?
- What is descriptive analytics?
- How much data do you need for analysis?
- What are some of the data cleaning methods?

BREAK TIME (10 MINS FROM 9:55 TO 10:05 AM)

PLEASE ANSWER THE QUIZ QUESTIONS AND SHARE VIA TEAMS

- TODAY, WE WILL WALK YOU THROUGH A DATA WORKSHOP FOLLOWED BY A PYTHON BASED ANALYSIS WORKFLOW.
- THE PROBLEM SOLVING CONCEPTS ARE IMPORTANT RATHER THAN KNOWING THE CODE.
- IF YOU ARE A PYTHONISTA , PLEASE GO AHEAD AND WORK THROUGH THE CODE.

LET'S MOVE TO THE WORKSHOP NOW

WE WILL PICK A PROBLEM AND SOLVE

10:05 to 11:30 am

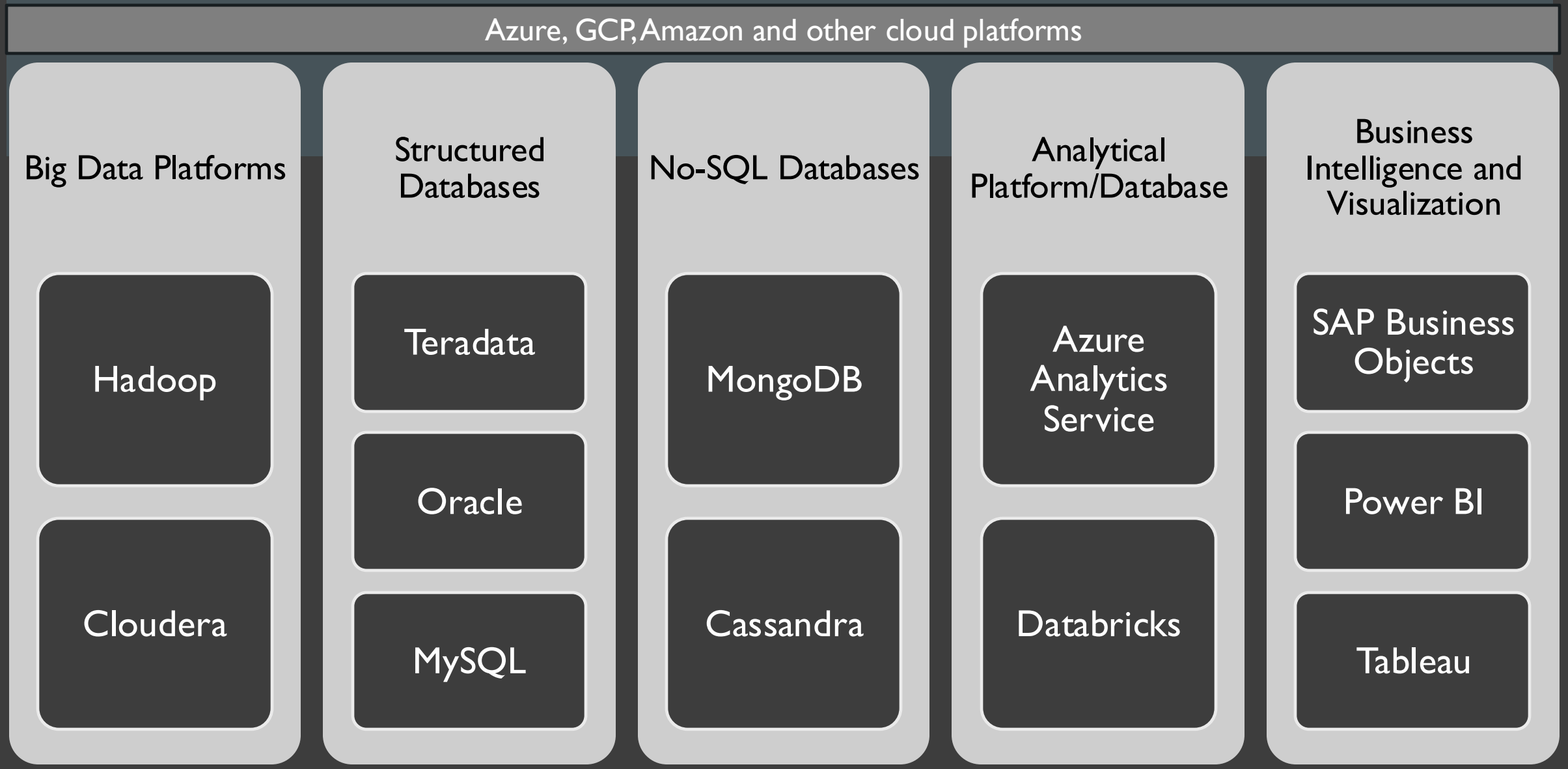
ASSIGNMENT

STEPS:

- EVALUATE YOUR RESULTS FROM THE ANALYSIS YOU DID IN THE WORKSHOP. STATE THE PROBLEM, BRIEFLY EXPLAIN THE DATA AND ANALYSIS YOU PERFORMED TO HELP ANSWER THE QUESTION, INTERPRET THE RESULT THAT THE DATA IS INFORMING AND STATE YOUR RECOMMENDATION. ALSO, THINK ABOUT HOW YOU CAN EXTEND THE ANALYSIS.
- WRITE A REPORT
  - If you like an evaluation, write a report in the tool of your choice, save it as a pdf and email it to aifeatures2000@gmail.com
- **NOTE: PLEASE ENSURE YOUR NAME IS IN THE REPORT'S FILENAME.**

NEXT STEPS AFTER THE WORKSHOP

- LEARN
- PRACTICE
- GROW



TECHNOLOGY, TOOLS, PLATFORMS

PRACTICE AND GROW

- Python or Excel
- Can use ChatGPT to help with Python or Excel coding
- There are a ton of data around. Try to apply what you learnt in this workshop in your daily life.
  - Solve a problem at work
  - Solve a problem as a hobby
  - Upload your shareable analysis to Github
- Platforms such as Azure, GCP or Amazon provide free tools for students
  - Sign up and learn their tools
- Repeat

GOOD LUCK AND THANK YOU

KEEP PRACTICING

PAM NAGARAJAPPA
AIFEATURES2000@GMAIL.COM