



Data Analytics Training

PROFESSIONAL
DEVELOPMENT SERIES

Session 1 Recap



What is data analytics?

Data analytics is the process of **combining and examining datasets** to identify patterns and **develop actionable insights**.



What is data?

Data is information that can be considered and used **to help decision-making**.



Data Types

Structured and unstructured, categorical and numerical, nominal and ordinal, discrete and continuous



Google Sheets Tutorial

Basic cell operations, filtering, and structure of formulas.

Session 2 Agenda

Exploratory Data Analysis

01

Data Storage

02

Data Analysis Phases

03

What is Exploratory Data Analysis?

04

Key Components

05

Hands-on Activity

06

Project Update

How is data stored?

```
"UK", "Smith", 16753, "Qtr 3"  
"USA", "Johnson", 14808, "Qtr 4"  
"UK", "Williams", 10644, "Qtr 2"  
"USA", "Jones", 1390, "Qtr 3"  
"USA", "Brown", 4865, "Qtr 4"  
"UK", "Williams", 12438, "Qtr 1"  
"UK", "Johnson", 9339, "Qtr 2"  
"USA", "Smith", 18919, "Qtr 3"  
"USA", "Jones", 9213, "Qtr 4"  
"UK", "Jones", 7433, "Qtr 1"  
"USA", "Brown", 3255, "Qtr 2"  
"USA", "Williams", 14867, "Qtr 3"  
"UK", "Williams", 19302, "Qtr 4"  
"USA", "Smith", 9698, "Qtr 1"  
"USA", "Jones", 18978, "Qtr 2"
```

Text Files / Flat Files (.txt, .csv)

Is a computer file that only contains text and has no special formatting such as bold text, italic text, images, etc.

	A	B	C	D	E
	ID	Last Name	First Name	City	State
1					
2	1	DOE01	JANE01	Los Angeles	California
3	2	DOE02	JANE02	Sedona	Arizona
4	3	DOE01	JOE01	Elmira	New York
5	4	DOE02	JOE02	Lackawana	New York
6	5	DOE03	JOE03	Defiance	Ohio
7	6	DOE04	JOE04	Tel Aviv	Israel
8	7	DOE05	JOE05	Cimax	North Carolina
9	8	DOE03	JANE03	Liberal	Kansas
10	9	DOE04	JANE04	Montreal	Canada
11	10	DOE05	JANE05	New York	New York
12	11	DOE06	JOE06	Hot Coffe	Mississippi
13	12	DOE06	JANE06	Java	Virginia
14	13	DOE07	JOE07	Yama	Russia

Spreadsheets (.xlsx)

an file in which data is arranged in the rows and columns of a grid and can be manipulated and used in calculations. Data can be numeric, text, formulas, references and functions.

Table Beaches			
id	name	region	country
1	Costa da Caparica	Lisbon	Portugal
2	Balos beach	Crete	Greece
3	Aquitaine	Gironde	France
4	Umag	Istria	Croatia
5	Sun	Pag	Croatia

Table Places

id	name	region
1	Costa da Caparica	Lisbon
2	Balos beach	Crete
3	Aquitaine	Gironde
4	Umag	Istria
5	Sun	Pag

Table Areas

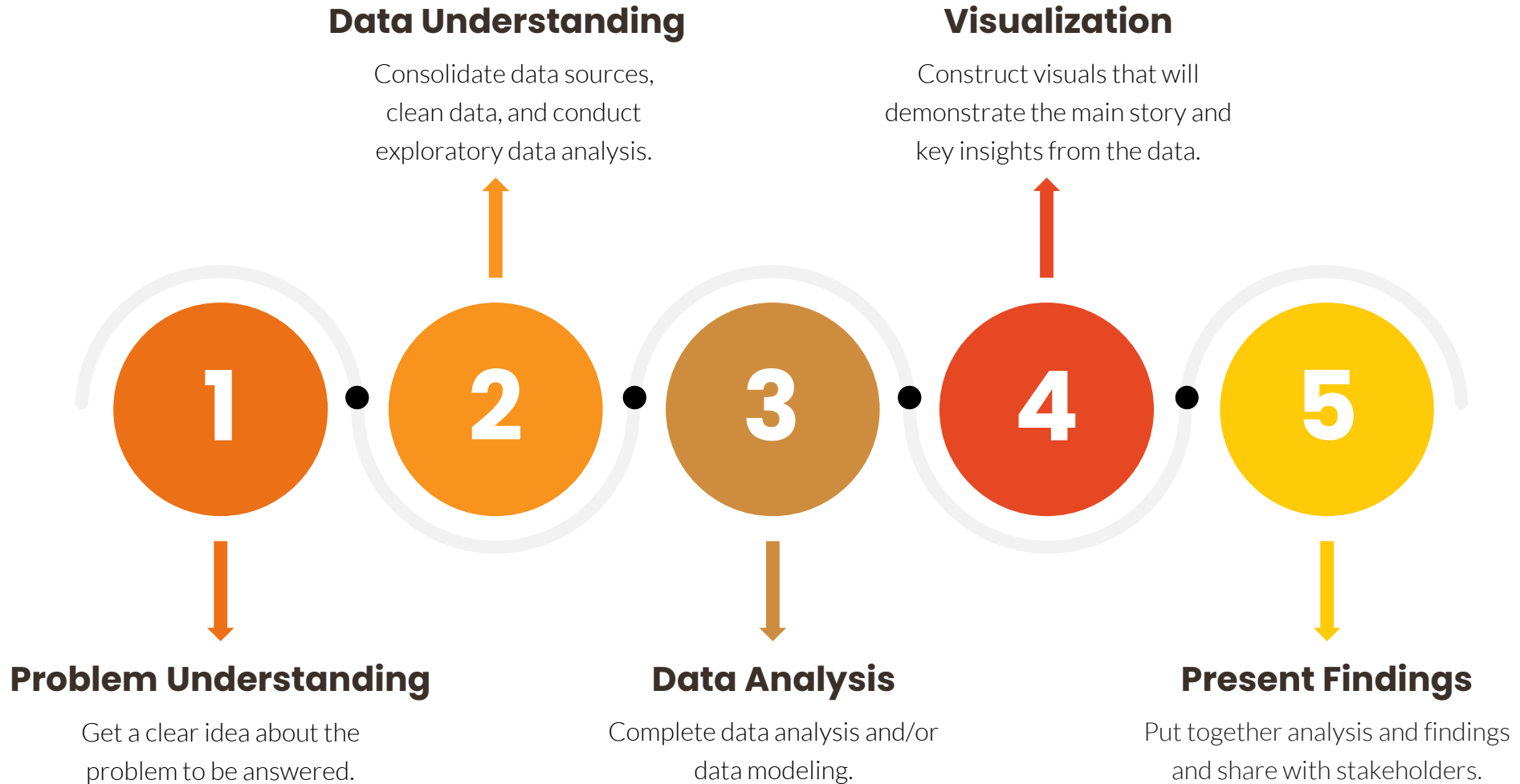
region	country
Lisbon	Portugal
Crete	Greece
Gironde	France
Istria	Croatia
Pag	Croatia

Databases

is an organized collection of structured information.

Databases, hold larger collections of organized information and allow multiple users at the same time.

Data Analysis Project Lifecycle



What is Exploratory Data Analysis (EDA)?



EXPLORATORY DATA ANALYSIS

“the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.”



Variables

Understand your variables.



Cleaning

Clean the dataset.



Relationships

Analyze relationships between variables.

Google Sheets Tutorial | EDA

