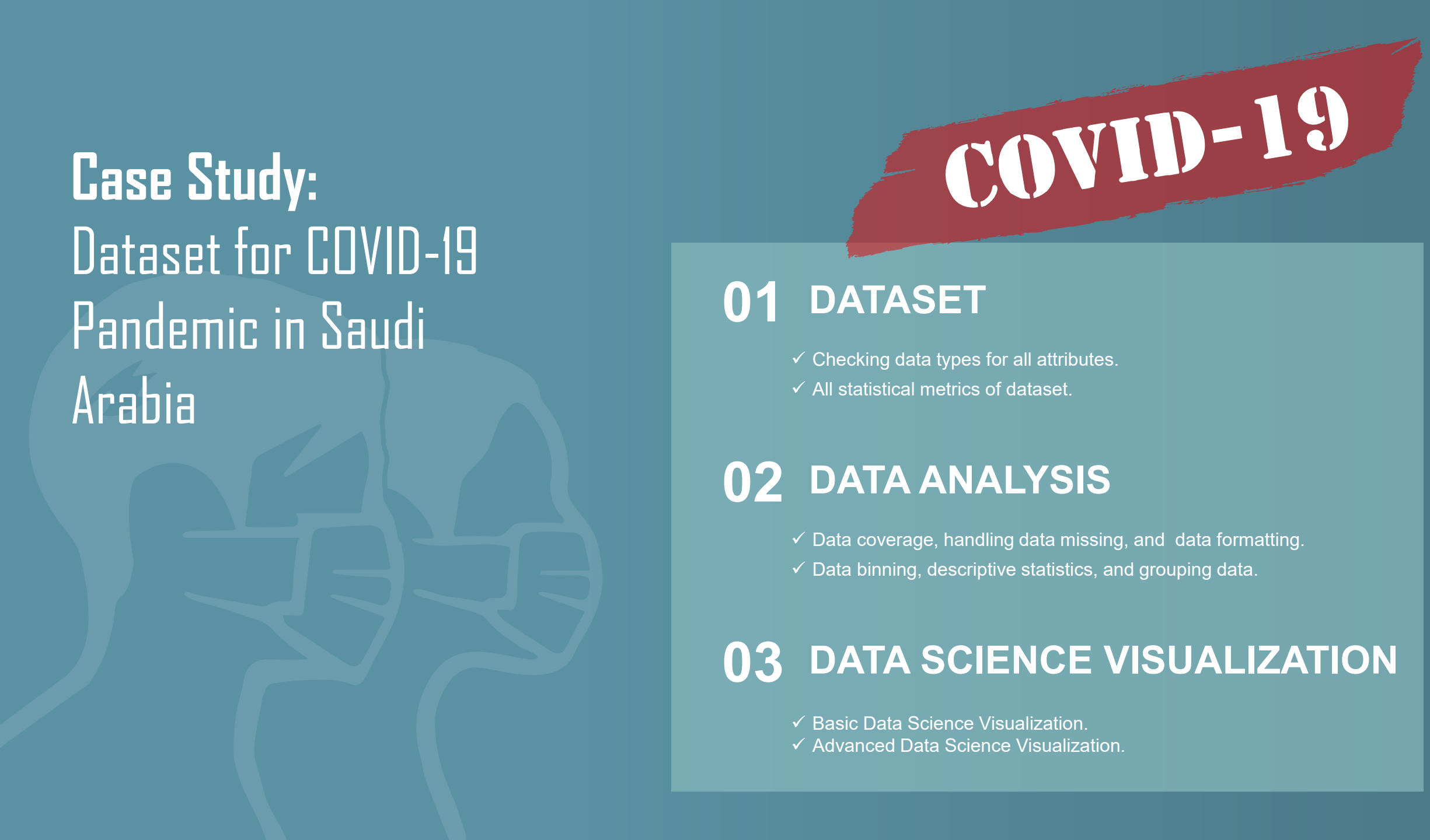
Exercise:
COVID-19 DATASET

Data Science & Big Data Bootcamp

https://github.com/galshammri/Applied-Data-Science

Ghalib Alshammri, Ph.D.

Case Study:
Dataset for COVID-19 Pandemic in Saudi Arabia



01 DATASET

- ✓ Checking data types for all attributes.
- ✓ All statistical metrics of dataset.

02 DATA ANALYSIS

- ✓ Data coverage, handling data missing, and data formatting.
- ✓ Data binning, descriptive statistics, and grouping data.

03 DATA SCIENCE VISUALIZATION

- ✓ Basic Data Science Visualization.
- ✓ Advanced Data Science Visualization.

COVID-19 Dataset

Basic Insights of Dataset

Data Type:

Check All Data Types

Statistical Metrics:

Check All Statistical Matric of Dataset

Understanding Dataset

Data Coverage:

Only Cover Daily Cases
From **Col: 2** To **Col: 1275**

Drop Record:

e.g. drop unnecessary record (“Total”)

Saving:

Save a New Data Frame after Changing

Data Pre-Processing

Handle Missing Value:

Check All Missing Values in Dataset

Data Formatting:

Change Wrong Format in Dataset

Header Dataset:

Drop Attributes:

- Daily / Cumulative
- Event

Advannced Processing

Data Binning:

Create Data Binning for
Total Attribute → BTotal

Descriptive Statistics:

Para.: **Indicator vs. Cases**

Grouping Data:

Para.: **Indicator vs. Region vs. Cases**

Dataset Sample

	A	B	C	D	E	F	G
1	**Daily / Cumulative**	**Indicator**	**Date**	**Event**	**region**	**City**	**Cases**
2	Daily	Cases	3/2/2020	First case of COVID-19	Eastern Region	Qatif	1
3	Daily	Cases	3/4/2020	Umrah suspension	Eastern Region	Qatif	1
4	Daily	Cases	3/7/2020		Eastern Region	Qatif	2
5	Daily	Cases	3/8/2020		Total	Total	4
6	Daily	Cases	3/10/2020		Eastern Region	Dammam	4
7	Daily	Cases	3/12/2020		Mecca	Mecca	21
8	Daily	Cases	3/13/2020		Mecca	Mecca	26
9	Daily	Cases	3/16/2020	Gov / private suspension	Riyadh	Riyadh	4
10	Daily	Cases	3/16/2020	Gov / private suspension	Mecca	Mecca	2
11	Daily	Cases	3/16/2020	Gov / private suspension	Mecca	Jeddah	5
12	Daily	Cases	3/17/2020		Eastern Region	Qatif	10
13	Daily	Cases	3/18/2020		Eastern Region	Dammam	23
	Da	Cases	3/19/2020		Eastern Region	Qatif	4
15	Da	Cases	3/19/2020		Eastern Region	Dammam	3
16	Dail	Cases	3/20/2020		Eastern Region	Qatif	?
17	Daily	Cases	3/20/2020		Al Bahah	Al Bahah	?
18	Da	Cases	3/20/2020		Asir	Bisha	1
19	Daily	Cases	3/20/2020		Mecca	Jeddah	11
20	Daily	Cases	3/23/2020	Curfew st ed for 2 days (6am -7 pm)	Mecca	Taif	6
21	Daily	Cases	3/23/2020	Curfew rted for 21 ys (6am -7 pm)	Riyadh	Riyadh	18
22	Daily	Cases	3/24/2020		Eastern Region	Dammam	6
23	Daily	Cases	3/24/2020		Total	Total	205
24	Daily	Cases	3/24/2020		Mecca	Jeddah	82
25	Daily	Cases	3/26/2020	Riyadh, Makkah and Madinah lockdown -	Eastern Region	Hufof	2
26	Daily	Cases	3/26/2020	Riyadh, Makkah and Madinah lockdown -	Mecca	Mecca	26
27	Daily	Cases	3/26/2020	Riyadh, Makkah and Madinah lockdown -	Mecca	Jeddah	13
28	Daily	Cases	3/27/2020		Qassim	Buraydah	2
29	Daily	Cases	3/27/2020		Eastern Region	Dammam	4

https://datasource.kapsarc.org/explore/dataset/saudi-arabia-coronavirus-disease-covid-19-situation/export/?sort=-daily_accumulative&dataChart=%3D&disjunctive.daily_accumulative=&flg=ar&disjunctive.city&disjunctive.daily_cumulative&disjunctive.indicator&disjunctive.event&disjunctive.region

Data Science Visualization

Data Science Visualization

04 Visualizing Geospatial Data

Use Folium Library to draw total cases in regions.

03 Pie Chart

Use Pie chart to draw total cases in regions.

02 Area Plot

Use area plot to draw a Top 5: Total Cases - Cities

01 Line Plot

Use line plot to draw a line for every week and only “Cases”
Parameters: Every Week + Indicator: Cases

THANK YOU