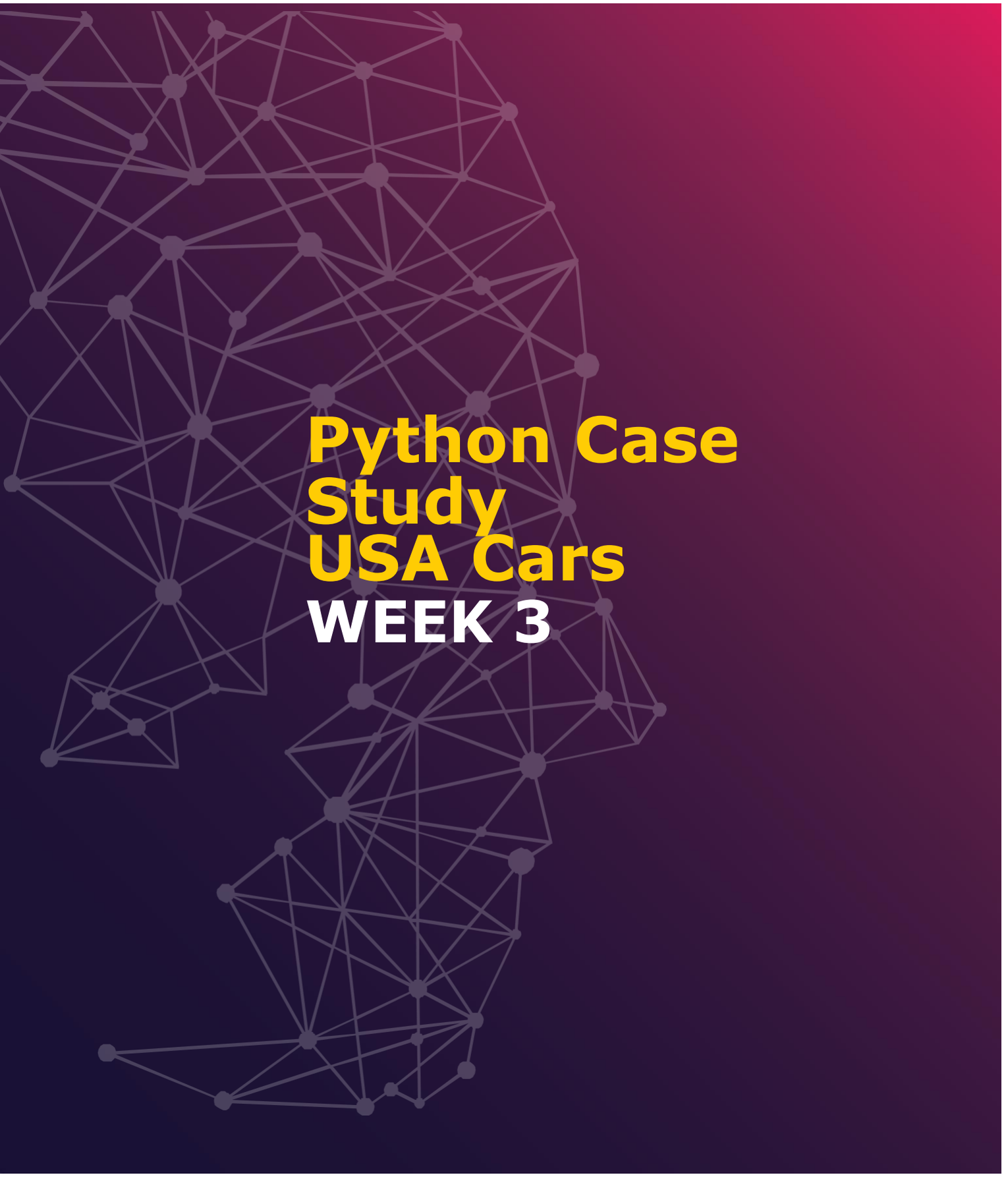

Python Case Study USA Cars

WEEK 3

Data Science Academy

Case Study

Disclaimer: This dataset contains new released car models between 1999 and 2008.

Variable dictionary is listed below:
Manufacturer: manufacturer name
Model: model name
Displ: engine displacement, in litres
Year: year of manufacture
Cyl: number of cylinders
Trans: type of transmission
Drv: the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd
Cty: city miles per gallon
Hwy: highway miles per gallon
Fl: fuel type
Class: "type" of car

Please, answer following questions:

1. Replace NaNs with mean for **highway miles per gallon** variable.
2. Delete observations that have NaN values.
3. Find correlation between variables. Which variables have highest correlation?
4. Show relationship between **highway** and **city** miles per gallon.
    a. Use scatter in matplotlib.
    b. Interpret how are variables correlated according to scatter.
5. Show distributions and scatters between all variables.While *hue* equals to **type of drive train**. Which variables have the lowest and the highest correlation? Find according to scatterplot.

Data Science Academy

6. Which **type of car** is most frequent in dataset? Show by using countplot.
7. Display number of **cylinders** for each **drive train** in bar chart. Which **drive train** is the most frequent?
8. Visualize **engine displacement** by each **class**, using boxplot. Do the same thing in violinplot.
9. Name your final case Script as "USA_Cars".
10. Write a lambda function which takes two arguments: a and b, the cathetus and return the length of Hypotenuse.
11. Write a lambda function which takes three arguments: a,b and c, the length of sides of triangle and return the area of triangle.

Data Science Academy