

## Exercise 2.7: Data Analysis and Visualization in Django

### Learning Goals

Work on elements of two-way communication like creating forms and buttons

Implement search and visualization (reports/charts) features

Use QuerySet API, DataFrames (with pandas), and plotting libraries (with matplotlib)

### Reflection Questions

- 1. Consider your favorite website/application (you can also take CareerFoundry). Think about the various data that your favorite website/application collects. Write down how analyzing the collected data could help the website/application.**

CareerFoundry would collect a lot of data, including how many people visit their website, what part of the course each student is on, how long a student stays on a task for and how many calls a student uses with their mentor. Analysing the above data could help CareerFoundry to market their company/business, assist students who may need extra support/time on their course and adjust estimated times for each task.

- 2. Read the Django official documentation on QuerySet API. Note down the different ways in which you can evaluate a QuerySet.**

Developers can use methods such as `filter()`, `get()`, `exclude()` and `annotate()` to evaluate the data in a QuerySet.

- 3. In the Exercise, you converted your QuerySet to DataFrame. Now do some research on the advantages and disadvantages of QuerySet and DataFrame, and explain the ways in which**

## **DataFrame is better for data processing.**

### QuerySet

#### Advantages

- integrates with Object-Relational Mapping system (easy integration with database)
- data is not fetched from database until requested (efficient)

#### Disadvantages

- doesn't provide extensive data analysis tools and functions like DataFrame have

### DataFrame

#### Advantages

- extensive range of functions (advantage when performing complex operations)
- easily read data from different sources (flexibility)

#### Disadvantages

- understanding all functions and methods available can take time, particularly for new developers
- can consume a lot of memory

DataFrame is better for data processing because there are many useful functions available for data processing, a large community for support and resources, data is versatile and can be read from many sources and lastly, complex operation can be performed through the wide range of functions.