Exercise 2.7: Data Analysis and Visualization in Django

Reflection Questions :

- *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.*

  - **A:** Benefits of analyzing collected data on www.imdb.com :
    - Content popularity – Analyzing user interactions could help which movies, TV shows, and personalities in show business are popular overall and/or at particular points in time, beyond/differently what viewership of said-shows could represent.
    - User preferences – Complex data could be derived and measured based on what varying users (users of various movie/show/people interests) interact with while browsing show business information.
    - User ratings – Users are encouraged to rate movies and shows, which generates priceless marketing and popularity data for producers, agents, and show runners.
    - Advertising – All of the above can be analyzed to guide advertisers to better place and craft their advertising resources.

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

  - **A:** Evaluating a QuerySet is to actually fetch the data from the database in a deliberate or calculated manner.
    - Iteration – Looping through the data elements to be returned in chunks.
    - Slicing – Returns a specified subset of the data.
    - Pickling (Caching) – Data sets, already evaluated, to be stored for access at another point in time.

- *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.*

  - **A: QuerySet Advantages:**
    - <u>Integration</u> – Easy to work with due to being tightly integrated with Django models.
    - <u>ORM</u> – Model relationships, database migrations, and models methods are a few examples of ORM features leveraged by Django.
    - <u>Agnosticism</u> – QuerySet can work seamlessly with different Django databases by allowing easy switching between them, without requiring very much change to the code.

  - **A: DataFrame Advantages:**
    - <u>Functionality</u> – The wider range of functions for data manipulation make it more versatile for general-purpose data processing.
    - <u>In-Memory Processing</u> – Performing data operations in-memory, as opposed to database-driven operations, makes it faster for certain types of operations.
    - <u>Importing</u> – Easily imports data from various sources such as SQL, Excel, CSV, etc.
    - <u>Indexing</u> – Advanced indexing and hierarchical labeling allow for more sophisticated data structures.
    - <u>Parallel Processing</u> – scalable for large datasets due to supporting parallel processing and distributed computing.