Homework 28.10.24 – Questions 1-7:

1. **Data Science** is a field that uses various techniques and methods to extract knowledge and insights, and to make predictions possible, which can enhance machine learning and AI applications. These insights can then be applied to personal, professional, and global domains, such as business decisions, weather predictions, economic strategies, and more. It is a foundational tool that makes AI and machine learning possible.
2. **Structured Data** is all data that is pre-structured in forms like tables and lists and is designed to be used in an organized database, usually a relational one (an example is SQL databases with interconnected tables). Structured data is easy to use due to its relational nature and pre-processed values and syntax, but its applications are limited by the code assigned to it, making it less flexible.

**Unstructured Data**, on the other hand, includes all types of data that are not pre-organized for relational use. This can include pictures, audio files, large text collections, etc. Unstructured data allows more flexibility in how it’s used across various fields due to its native format, which can also save time since there is no need for data organization into rigid formats. However, due to its unorganized nature, a high level of expertise is required to use it effectively, and it requires specialized tools for management and manipulation.

1. **Exploratory Data Analysis** (EDA) is a method used to examine and explore the characteristics of data prior to usage. This method often includes visual aids to identify data variables, determine data distribution, and recognize patterns and relationships between variables. The use of EDA helps prepare for the next steps in a project’s process.
2. **Companies use Data Science for two main purposes**:
   1. **Answer a question** to inform a decision: Business, production, management, and other issues can be approached using Data Science due to its predictive capabilities.
   2. **Solve a problem**: When facing an issue to be addressed, Data Science can be used to analyze the problem and generate a solution based on statistical analysis and specific needs.
3. **Gen AI Tools** are applications and platforms that use AI to generate content or outputs based on input data or prompts, typically handling unstructured data in formats like text, image, video, and audio, and creating related outputs.
4. **Three libraries that we will use in Python are**:
   1. **Pandas**: Enables high-level data manipulation and analysis.
   2. **NumPy**: Used for complex mathematical computations.
   3. **Matplotlib**: Supports static, animated, and interactive visualizations, such as graphs, which enhance data visualization.
5. **a. Jupyter Notebook** is a valuable Data Science tool because it allows users to experiment with one-line code and integrate it with visual aids or data visualization, making it easier to explore code, outputs, or algorithms.

**b. Markdown** is a Jupyter feature that lets users create visual enhancements for written code. Users can use Markdown to style text, add images or graphs, embed code, etc., making code, products, and projects more accessible and understandable.

**c. Typing "!" before a command in Jupyter** allows the user to run shell or terminal commands directly from Jupyter.