**Task 1**

**Understanding Data Science**

**Data Science and Related Fields:**

**>Data in Everyday Life:**

Data is ubiquitous, from text to contacts, influencing decisions. It spans basic counting to complex communication.

**>Understanding Data Science:**

The primary purpose data science serves is to extract insights from data using scientific methods including ML and AI for predictive modeling.

**>Related Fields:**

Related fields that Data science overlaps with are databases, big data, ML, AI, and visualization. This Facilitates scalable data handling and decision-making.

**>Types of Data:**

Data could be structured, semi-structured, and unstructured data with examples including spreadsheets, text files, and HTML or JSON files respectively. These vary in organization and analysis complexity.

**>Data Handling Steps:**

Data Handling steps include Acquisition, storage, pre-processing, visualization, model training, and further evaluation. Following all these steps is essential for deriving insights and predictive models.

**Data Ethics and Applied AI:**

**>Ethics Concepts:**

Data Ethics studies moral issues in data and AI and defines shared values and ethical challenges.

**>Ethics Principles:**

It encompasses of foundational values like accountability, transparency, fairness, and privacy. it further guides ethical AI frameworks across organizations.

**>Ethics Challenges:**

Ethical Challenges data scientists face include data ownership, informed consent, privacy, bias, and fairness. It usually raises questions on user rights and ethical data use.

**>Case Studies:**

Examples like the Netflix data breach and algorithmic biases fit the description and illustrate real-world impacts of ethical lapses in data use.

**>Applied Ethics:**

It suggests the use of professional codes, ethics checklists, and regulations. and to Promote ethical AI practices and governance.

**Defining Data:**

**>Definition:**

Data encompasses facts, information, observations, and measurements used for discoveries and decisions. A data point is a single unit within a dataset, which can vary in format and structure based on its source.

**>Data Characteristics:**

* Raw data is initial, unanalyzed data that requires organization for human and technological understanding.
* Quantitative data numerical observations enabling mathematical analysis (e.g., population, earnings).
* Qualitative data is subjective information not measured objectively (e.g., product reviews, colors).

**>Data Structures:**

* Structured datais organized into rows and columns with defined rules (e.g., spreadsheets, databases).
* Unstructured data lacks strict organization (e.g., text files, video files), allowing flexibility but posing challenges in analysis.
* Semi-structured datacombines structured elements with flexibility (e.g., JSON, HTML), often using metadata for organization.

**>Data Sources:**

* Primary vs. Secondary data originates from user-generated sources vs. collected for general use.
* Common sources in the data world are Databases, files (audio, image, video, spreadsheets), internet-hosted data, APIs, and web scraping.