**Summary: Defining Data Science**

**Data**

Data is everywhere from the textbooks we read to the phone numbers stored on the mobile.

**Data Science**

Data Science is a field that uses scientific methods to extract knowledge and insights from both structured and unstructured data. The main goal of data science is to understand data, find some hidden relationships and build a model.

These are some fields that are related to data science

* Databases: Structuring data for faster processing.
* Big Data: Storing and processing large data sets efficiently.
* Machine Learning: Building models to predict outcomes.
* Artificial Intelligence: Creating complex models that mimic human thought.
* Visualization: Making data comprehensible through visual means.

**Types of Data and their Sources**

|  |  |  |
| --- | --- | --- |
| **Structured** | **Semi-structured** | **Unstructured** |
| Data organized in rows and columns like tables. | Data that is not fully organized in a structure. | Data that is not organized in a structure. |
| Database Table and Excel spreadsheet. | Email and Json Files. | Images and videos |
| IoT sensors, surveys and user behavior analysis. | Social network graphs and photos for group dynamics. | Texts, images, videos and web server logs. |

**Data Science Process**

The data science process usually involves these steps

1. **Data Acquisition**: Collecting data.
2. **Data Storage**: Storing data in databases, NoSQL databases, or data lakes.
3. **Data Processing**: Converting data into a usable form.
4. Visualization/Human Insights: Using visualizations to understand data.
5. **Training Predictive Models**: Using machine learning to make predictions.

This is a short summary of what I have learned in Defining Data Science Task.