Course Glossary: Understanding Data Science

- **Data Science:** The practice of working with quantifyable data to draw conclusions. There are three main subareas of Data Science: Data Engineering, Data Analysis, and Machine Learning.
- **Descriptive Analytics:** The use of data to describe the current state of a system or process.
- **Diagnostic Analytics:** Determining why something happened by identifying the data's root causes, patterns, and relationships.
- **Predictive Analytics:** Using historical data and machine learning to identify possible outcomes and the probability that they will happen.
- **Data Analysis:** A set of methods for summarizing the current state of a company, product, or service by using data. Specialties include dashboarding, data visualization, and ad hoc analysis.
- **Machine Learning:** The science of predicting or classifying unknowns using a large set of data.
- **Data Engineering:** The practice of creating and supporting systems so that data scientists and data analysts can access data.
- **Data Pipeline:** A series of automated steps to collect, process, and store data for analysis.
- **Database:** A repository of data.
- **Internet of Things (IoT):** A network of connected devices that collect and share data through embedded sensors and software (e.g., smart thermostats, wearable devices).
- **Cloud service provider:** A company that gives access to large amounts of computing and storage power that is remote, secure, and reliable. Examples include AWS, Azure, and Google Cloud.
- **AI:** Any type of advanced prediction performed by a computer based on data and rules.
- **Deep learning:** A special type of predictive computing that uses a neural net to solve problems that classical machine learning cannot.
- **Supervised learning:** Machine learning where the input dataset is labeled, and the algorithm attempts to apply the same labels to new data according to rules learned by examining the input dataset.
- **Unsupervised learning:** Machine learning where the input dataset is unlabeled, and the algorithm attempts to group the input dataset and eventually assign those groupings to new data.
- **Regression:** Machine learning that predicts a number, such as the value of a stock at a future point in time.
- **Classification:** Machine learning that labels a piece of data, such as image recognition or customer churn prediction.
- **Time Series Analysis:** Techniques used to analyze trends, cycles, and seasonality in time-based data.