Data science is the field of exploring, manipulating, and analyzing data, and using data to answer questions or make recommendations also,

Data science is the study of large quantities of data, which can reveal insights that help organizations make strategic choices.

Data Scientist finds a solution to problems small/ big using appropriate tools and then communicate his/ her findings to the stakeholders. the amount of data one is dealing with doesn't determine if the person is a data scientist or not and a mere analytic tool doesn't restrict someone to be a data scientist.

a data scientist must be able to deal with unstructured situations.

**the steps that a data scientist follows -**

**1.** classify the questions that the organization needs to be answered

**2.** clarify the business needs

**3.** what data we need to solve the problem and where will it come from

**4.** data can be structured and unstructured from many sources and depending on the nature of the problem, they can choose to analyze the data in different ways.

**5.** use multiple models to explore the data, sometimes pre-existing models will give what organization demands, but sometimes a new approach needs to be followed.

**6.** after analyzing the data, a data scientist needs to communicate his findings to the organization, powerful data visualization tools are used to convey the results to the organization

**Examples of some of the organizations using data science -**

**Uber:** uber uses real-time customer data to decide price surges and ensure the availability of drivers in the region.

**Toronto Transportation commission:** uses complaint data to make better traffic flows to reduce average transit time.

**Cyanobacterial blooms:** in the east coast, us there is a problem of cyanobacterial blooms in freshwater lakes so using robot-driven boat data is collected and analyzed to predict when will the blooms occur.

**how to approach a solution to a problem -**

gathering a lot of data, cleaning and preparing it.

analyzing it to gain the insight needed to develop better solutions.

Identify the problem and establish a clear understanding of it.

Identify the right tools to use.

Develop a data strategy.

**A casual take on regression -**

let's suppose you take a cab ride you will be paying some base fees and then you will pay for the distance you travel and the time you were stuck,

regression allows you to device those constants when only the total distance and total fees are given.

**why the cloud is considered a godsent for a data scientist?**

- cloud doesn't just allow you to store large amounts of data to the cloud it also allows you to apply your algorithms to the data by using clouds computing power.

- also, one advantage of the cloud is that multiple people can work on the same data simultaneously.

- no need to install open source tech such as apache spark to your machine

- the cloud provides the most updated tools and libraries without worrying about updating and maintaining them