What is Data Science?

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

Roles and Responsibilities of Data Scientists

Help Organizations to:

Understand their market environment

Gather insights to support a hypothesis(an assumption that a company has on how the business is happening)

Reveal any hidden trends and opportunities

Data Analysis is a huge part of data science

It help by adding on the knowledge of an organization.

Helps to understand how to use the data for the betterment of the business

Process of Data Analysis:

1. Data Scientists can concentrate on specific problem and answer specific questions of an organization. Data Scientist needs to understand the problem at hand. He/She can do that by asking questions about the problem.
2. Investigating which type of data we need and where to get it.
   1. Data can Structured and Unstructured.
   2. There can be multiple sources
   3. Data Scientist can choose a particular model depending on the type of the data.
3. Applying models reveals patterns, Architype, outliers.
4. Sometimes we need more than 1 model to get the insights.
5. The result might be something that is already know . Sometimes the result might be completely new that might help the business to move towards a completely new approach.
6. Data Scientist needs to become a story teller often using graphs and charts to explain the trends and the finding to the stakeholder.

Many Paths to Data Science:

Data Science is new

Stats and Math

Business

All Engineers work with Data.

Modelling Traffic

Green Pipe emissions

Contemporary data scientists come from different backgrounds such as engineering, mathematics, and even psychology. The secret skill is passion for continuous learning of new tools and patience to clean and analyze data.

# Advice for New Data Scientists

Be Curious, Extremely Argumentative and Judgemental.

Without Curiosity We don’t know what to do with data

Judgemental - Have a pre conceived notion on what the data provides

Argue --> Learn new things from the data that might prove you wrong--> incorporate the learning.

Flexibity with data science tools.

Ability to convey findings

Which industry?

Understand your Compatity Advantage

## Data Science: The Sexiest Job in the 21st Century

https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTAxRU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZSUyMDEvRGF0YV9TY2llbmNlX1RoZV9TZXhpZXN0X0pvYi5tZCIsInRvb2xfdHlwZSI6Imluc3RydWN0aW9uYWwtbGFiIiwiYWRtaW4iOmZhbHNlLCJpYXQiOjE2Njg2MDk1NzV9.E\_5CUNsuJtxnXNRoqcKxZlzghHV3b2olVxPD55vdhZ8

In this lesson, you have learned:

* Data science is the study of large quantities of data, which can reveal insights that help organizations make strategic choices.
* There are  many paths to a career in data science; most, but not all, involve a little math, a little science, and a lot of curiosity about data.
* New data scientists need to be curious, judgemental and argumentative.
* Why data science is considered the sexiest job in the 21st century, paying high salaries for skilled workers.

Part 2:

A Day in the life of a Data Scientist:

Projects that the people did in the video.

Recommendation System

Predict Algae Blooms (Artificial Neural Network)

Complaints with Transit.(tabular data and Unstructured data like emails and phone call recordings ) -- find pattern or may not find a pattern. Data was related to weather.(rainy days , heavy win)

Old, New Problems and Data Science Solutions:

Lot of data is available. Organizations can take advantage of it in a lot of ways

However, lot of them use Data Science to find optimum solutions for existing problems

Examples

Uber collect realtime data to collect how many drivers are available and check if they need more.And figure out if they need to apply a surge charge, Uber analyses data to put the drivers in optimal place and add surge to maximizre the pay.

Toronto Transport Corporation solved the problem of managing transport between cities .They gathered data for street carts and figured out if they need interventions, Analysed customer complaints data, used probe data to understand the trafic in the main routes and created a team to better capitalize on big data for planning and operation of the traffic .By focusing on peak hour clearances and most congested routes ,peoples time lost dropped from 4.7 hours to 3 hours from 2010 to 2014.

Environment Application:

Fresh water is necessary for a ecological needs( drinking water, producing food).

But lakes are being polluted by Cyanobacterial blooms.

Lot of teams work on the project to reduce this.

Robotic boats, Buoys and Camera-equipped drones are being used to gather physical, chemical and biological data to analyze the effects of Cyanobacterial blooms.

Also new Algorithms are being bult to better predict when and where Cyanobacterial blooms tak.

It’s a multi deciplinary problem.

It take lot data gathering, cleaning and analizing the data to get better insights on the data and to apply this onto the field.

How to get a Better Solution which is efficient

1. Understand the problem clearly.
2. Gather data for analysis.
3. Identify the tools needed.
4. Develop a strategies.
5. Case studies are also helpful
6. Once the clean data is available, develop a machine learning model to predict.
7. Refine the best practices to make the process efficient.

Data Scienc topics and Algorithms,

Regression,

Visualization

Artificial Neutral Networks.

K-Nearest Neighbor.

Sometimes simple algorithm like K-Nearest Neighbor gives you better result than a very complex algorithms.

Structured Data – tables

Unstructured data – not tabular. Most from internet.text, video, audio.

Mostly transform unstructured data to get some kind of structure.

Then we analyse,

Regression:

In a cab there is a fixed amount first.

Every n meters, m amount increases.

There is a relationship between distance and the amount we pay + constant value at the starting.

Regression is similar. It tell what is the base fare and what is the relationship between the time and the distance you travelled and amount you pay.

Knowing the relationship relationship between the time and the distance you travelled and amount you pay, regression allows you to find the base pri

Cloud for Data Science:

Cloud is very important for Data Science.

It allows us to store data on a centralized system and eleminates the limiation of having exensive pysical servers.

It allows us to deploys advanced algorithms aswell.

Allows collaboration.

Allows to get access to open source tech like apache spark which allows to avoid installation and configuring them locally.

Also gives most up todate tools and libraries.No maintainance.

Can access in different devices.

Tools like jupyter notebook and spark are readily available.

IBM – Skills Network lab

What Makes Someone a Data Scientist?

https://author-ide.skills.network/render?token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJtZF9pbnN0cnVjdGlvbnNfdXJsIjoiaHR0cHM6Ly9jZi1jb3Vyc2VzLWRhdGEuczMudXMuY2xvdWQtb2JqZWN0LXN0b3JhZ2UuYXBwZG9tYWluLmNsb3VkL0lCTURldmVsb3BlclNraWxsc05ldHdvcmstRFMwMTAxRU4tU2tpbGxzTmV0d29yay9sYWJzL01vZHVsZSUyMDEvUmVhZGluZ19XaGF0X01ha2VzX1NvbWVvbmVfRGF0YVNjaWVudGlzdC5tZCIsInRvb2xfdHlwZSI6Imluc3RydWN0aW9uYWwtbGFiIiwiYWRtaW4iOmZhbHNlLCJpYXQiOjE2Njg2MTAzMDl9.ruZLaBvWfl-pKvpKfQz3T9ydJPLa6P-GypnsTgUJr0E

In this lesson, you have learned:

* The typical work day for a Data Scientist varies depending on what type of project they are working on.
* Many algorithms are used to bring out insights from data.
* Accessing algorithms, tools, and data through the Cloud enables Data Scientists to stay up-to-date and collaborate easily.