

Welcome to Data Science January 2018 batch!

---

Let's start the journey with

---

# Data Science Foundation: Data Analysis with Python



# Data Science

## The 'Why' of data







In the next 10 years, [REDACTED] and [REDACTED] will do more for medicine than all of the biological sciences together.

# About Me



Sai Charan

- Product Manager with 8+ years of experience
- MBA + Computer Science Background
- Manages a product that is powered by Machine Learning
- Implemented features like Voice-to-text, Virtual Assistant and Mentor-Matching
- Love Teaching. Ran my own GMAT training venture called 'GMAT Analytics' for an year. Taught Science and Maths to 6<sup>th</sup>/7<sup>th</sup> standard students in rural areas using Technology.
- Setup a Grid during my under-grad days to validate a Mersenne Prime number in under 15 mins
- **Interests :** Public Speaking, Brain-lifting, Football
- **Email :** saicharantej@upxacademy.com

# **Introduction to Data Science**

# Agenda for today's class

Introduction to  
Data Science

What is Data?

What is Data  
Science?

Industry  
applications

Problems  
solved by  
Data Science

Data Science  
Project  
Lifecycle

Data Types



Example to ponder

**amazon**

The Amazon logo consists of the word "amazon" in a bold, black, sans-serif font. Below the word is a curved orange arrow that starts under the letter 'a' and points towards the letter 'n'.

# Example to ponder upon

Guess type of customer !!

iPhone	Beer	Cornflakes			
iPhone	Diapers	Beer			
iPhone	Diapers	Biscuits			

# Example to ponder upon

Guess type of customer !!

iPhone	Beer	Cornflakes			Single M/F
iPhone	Diapers	Beer			Parent M/F
iPhone	Diapers	Biscuits			Parent M/F



# What do you recommend

???

iPhone	Diapers	Biscuits			
--------	---------	----------	--	--	--

Hint!!

iPhone	Beer	Cornflakes	97%	Biscuits	Single M/F
iPhone	Diapers	Beer	98.4%	Aftershave	Parent M/F
iPhone	Diapers	Biscuits	91%	Cosmetics	Parent M/F

# Example to ponder upon

Also buy

iPhone	Beer	Cornflakes	97%	Biscuits	Single M/F
iPhone	Diapers	Beer	98.4%	Aftershave	Parent M/F
iPhone	Diapers	Biscuits	91%	Cosmetics	Parent M/F
iPhone	Diapers	Biscuits	76%	Racquets	Single M

If I had further info

## Example to reflect on

35% of Amazon sales  
are generated  
through its  
**Recommendation  
Engine\***

amazon.com

**Recommended for You**

Customers who viewed this item also viewed these products



Dualit Food XL1500  
Processor

\$560

Add to cart



Kenwood kMix Manual  
Espresso Machine

★★★★☆

\$250

Select options



Weber One Touch Gold  
Premium Charcoal  
Grill-57cm

\$225

Add to cart



NoMU Salt Pepper and  
Spice Grinders

\$3

View options

\*Source: Venturebeat



**How far  
is too  
far?**



## Example to reflect on

---



- 60 member data mining team
- Conclude about shoppers
  - Specific
  - Private
  - Target Ads - Buying behaviour



If a women  
shopper is  
expecting, and  
their approx due  
date



Predict income  
levels based on  
purchase history



Attract customer  
attention through  
custom brand  
offerings

# What is Data?



# What is Data?

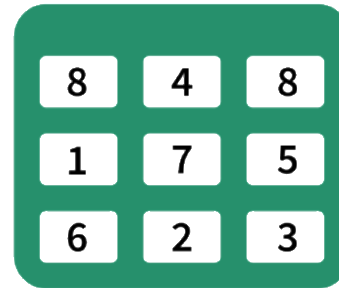
- Data are raw ingredients from which statistics are created.
- Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.“



# What is Data: Examples

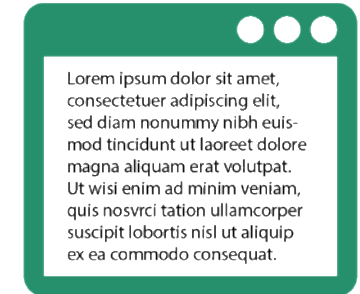


Sequence of bits



8	4	8
1	7	5
6	2	3

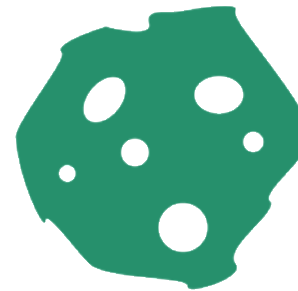
Table of Numbers



Characters on a page



Recording of sounds

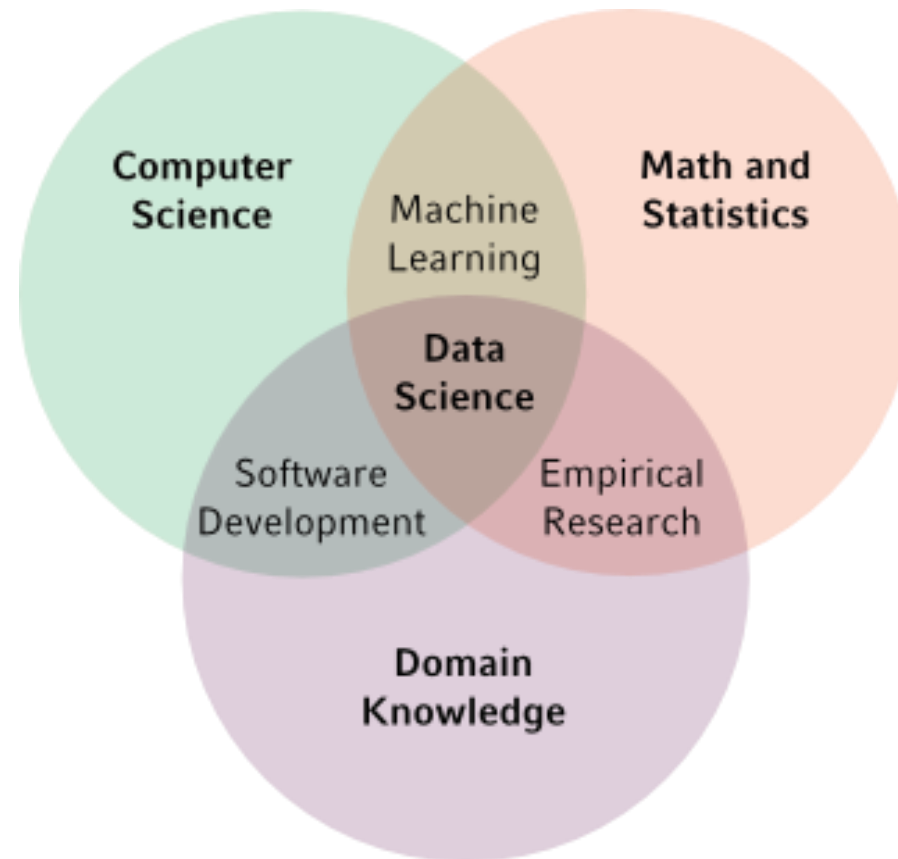


Moon rock specimen

# What is Data Science?

# What is Data Science?

---





# What is Data Science?

## Analytics/BI



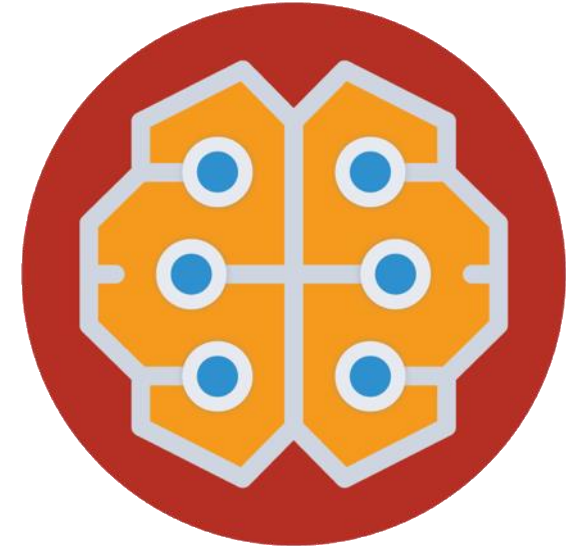
Shows **WHAT** has happened

## Data Science(EDA)



Shows **WHY** it happened

## Machine Learning



Shows **HOW** we can predict & automate decision making systems

# Who are Data Scientists?

# Who is a Data Scientist?



**Josh Wills**

@josh\_wills

Follow

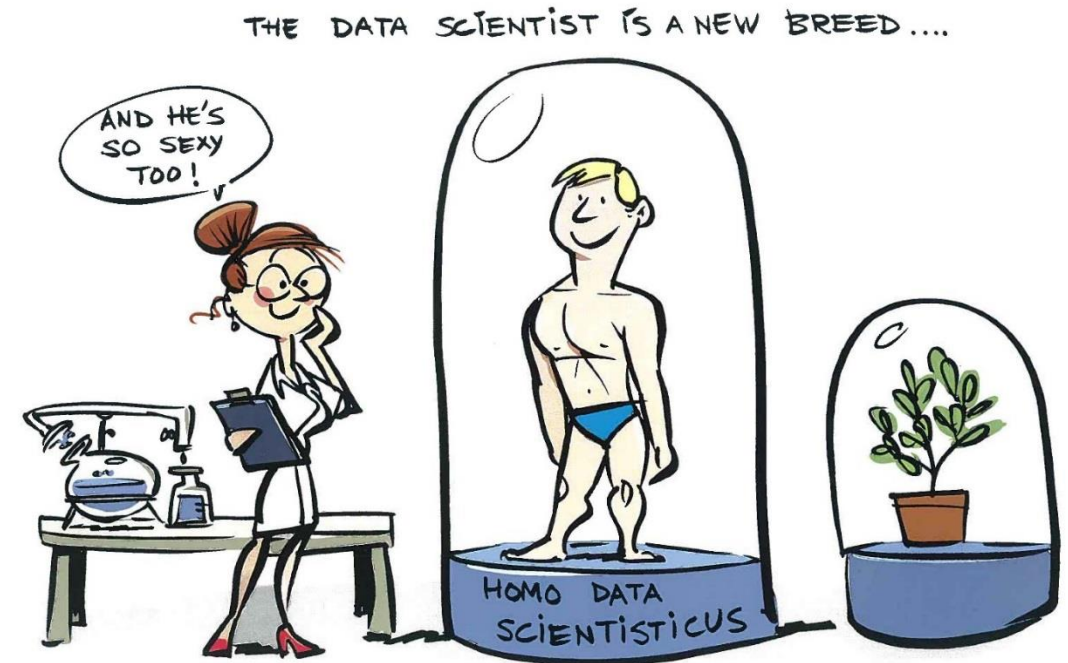
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

9:55 AM - 3 May 2012

1,673 Retweets 1,365 Likes



51 1.7K 1.4K



©Marion van de Wiel 2014

DSC/e 2014

# Data Science: Hype or Reality?

**Data scientists... are in such high demand & low supply in the new era of big data that they are earning almost three times Australia's average salary.**

news.com.au Aug 2015

- In the press, we have heard a great deal of hype about the “data scientist” role.
- These are rare superstars who are great at math and statistics, also great programmers, and have great business understanding.



Source:  
*Harvard Business Review*, Oct. 2012

# **Diving Deep into Data Science**



# Data Science: Science or Art?

- It is more **Art** and less Science
- Ways in which **creativity can be expressed in data science include:**
  - applying techniques in unconventional settings.
  - combining techniques in unconventional ways.
  - developing new and unusual hypothesis.



# Show me the money!

## Why do Data Scientists get paid a lot?



Severe  
shortage of  
talent



Directly impact  
business value



Organizations  
face enormous  
data challenges



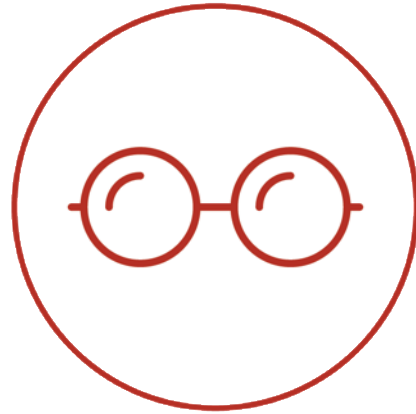
Need not restricted  
to Tech Giants  
anymore

# Good vs Great Data Scientists?

---



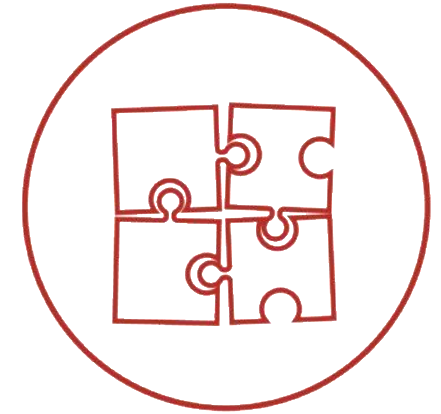
Effective  
Communication



Intellectual  
Curiosity



Domain / Industry  
Knowledge



Analytic Problem-  
Solving

# **Problems that Data Science solves**

# Problems that Data Scientists solve

---

- Is this A or B?
- Is this weird?
- How much or how many?
- How is this organized?
- What should I do now?

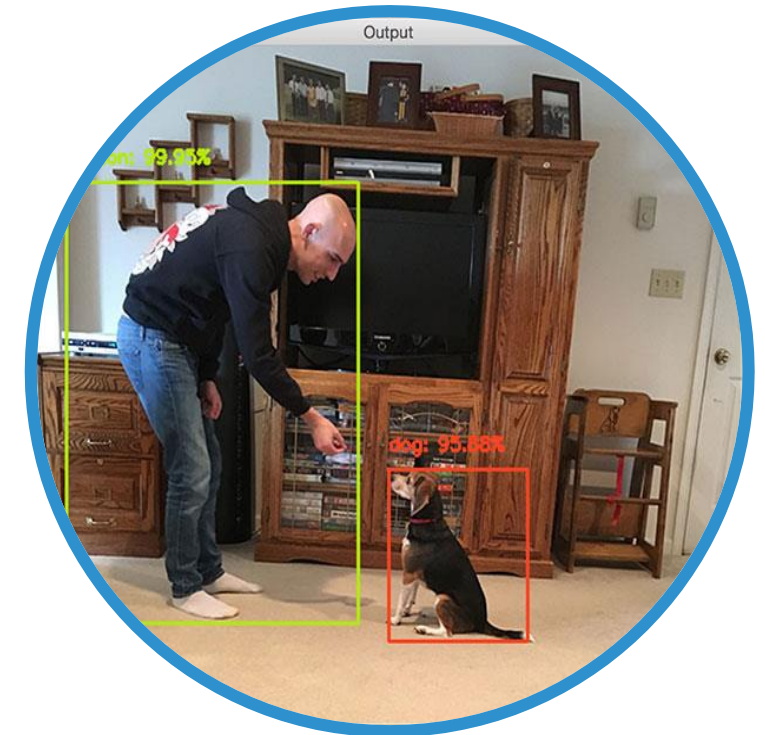
# Problems that Data Scientists solve

Is this  
A or B?

Classification



Will this applicant be able  
to repay the loan?



Identify a human & a  
dog?



# Problems that Data Scientists solve

**Is this  
Weird?**

Anomaly



Is this mail in your inbox  
spam?



Is someone riding bike on  
a walkway?

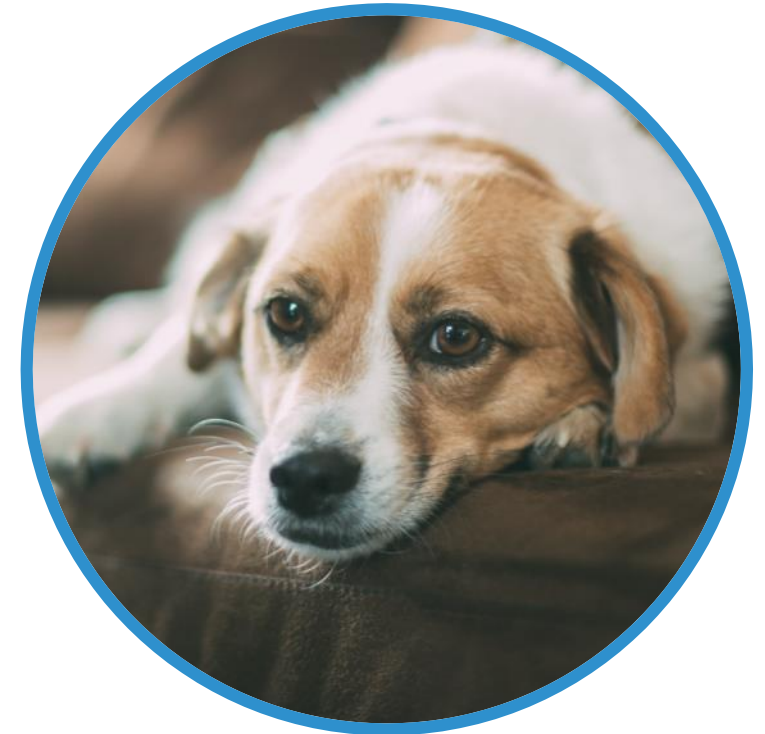
# Problems that Data Scientists solve

## How is this organized?

Clustering



Which viewers like the same kind of movies?



Which pet owner groups have the same purchase behaviour for pet foods?



In the next 10 years, data science and software will do more for medicine than all of the biological sciences together.

Vinod Khosla

# **Data Science Foundation Course Structure**

## Objectives of the Foundation Course

---

- Systematic approach towards answering the ‘Why’ questions about data
- Use programming and statistics together to uncover the ‘Why’
- Hands-on experience with real-world data sets
- Learning things one step at a time
- Most important of them all – learning by sharing!

# Data Analysis with Python Course Structure

---

- Class 1:** Welcome to the World of Data Science
- Class 2:** Let's learn Python for Data Science
- Class 3:** How to use Statistics (Descriptive) for understanding data
- Class 4:** Play around with data using Numpy
- Class 5:** How to manipulate data with Pandas
- Class 6:** Intro to Inferential Stats: The science behind prediction



## Data Analysis with Python Course Structure

---

**Class 7:** Inferential statistics contd.

**Class 8:** Visualize to understand more

**Class 9:** Data analyst at work - Day 1: Understanding spread of TB worldwide

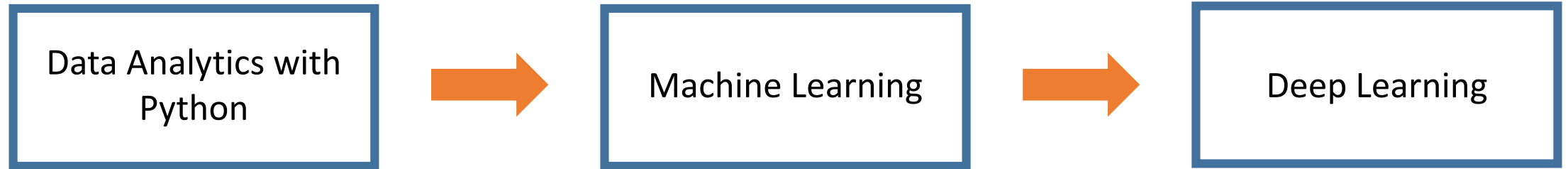
**Class 10:** Data analyst at work - Day 2: Use movie sets data to identify gender-movie preference

**Class 11:** Data analyst at work - Day 3: Identifying patterns in stock prices

**Class 12:** Let's Wrap up

# 6 Month Certificate Program in Data Science

## Core Modules



## Electives

