Why Data Science?

Course Induction

Areej Alasiry

About Me

- Areej Alasiry
- PhD in Computer Science and Information Systems
- Majored in Data Mining and NLP
- Assistant Professor at IS Department
- IBM Certified Data Science Instructor

- Vice Dean for Graduate Studies and Scientific Research
  College of Computer Science
  King Khalid University

Lecture Outline

- Data and Big Data
- Data Structures
- Business Intelligence
- Data Science
- Data Science Knowledge Areas
- Job Market
- Career Prospect
- Data Scientists Skills
- Course Specification
- Course Assessment
- Course References

Data

- What is data?

Example

Areej Alasiry

Big Data

- What is Big Data?
  - *How big is "Big"?*
  - Characteristics of Big Data

" *Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.*

Doug Laney, Analyst at Gartners Inc, 2011

Big Data Around Us

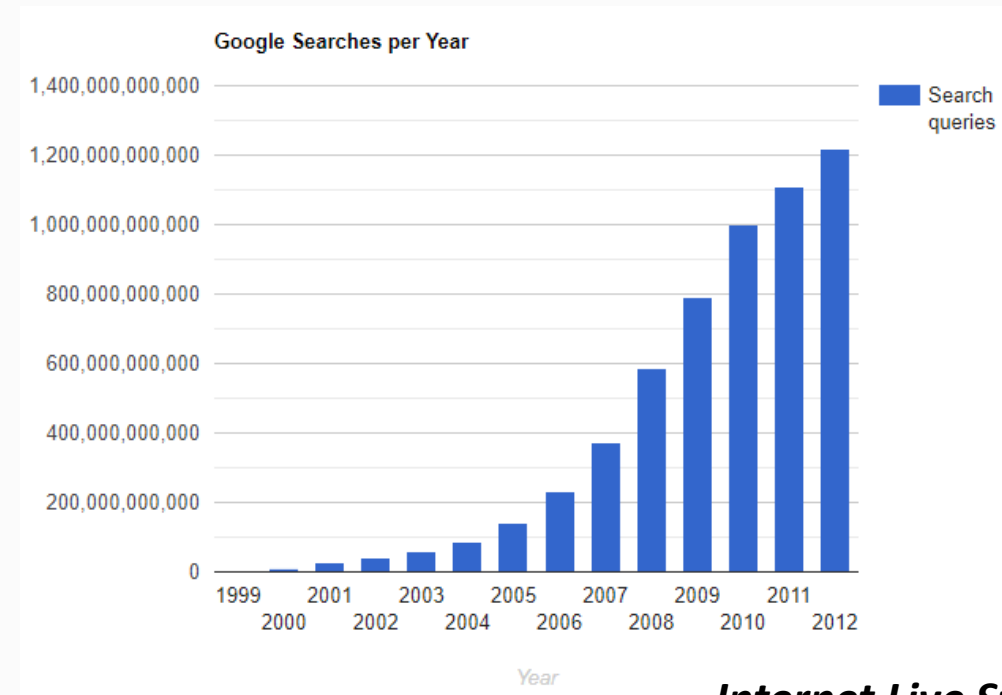Can you think of examples of Big Data **?**
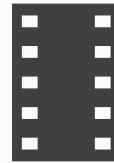
Google

" 1 second = 40000 search queries (on average)

1 day = 3.5 billion search queries

1 year = 1.2 trillion search queries

Internet Live Stats

Twitter

" 1 second = 6,000 tweets

1 minutes = 350,000 tweets

1 day = 500 million tweets
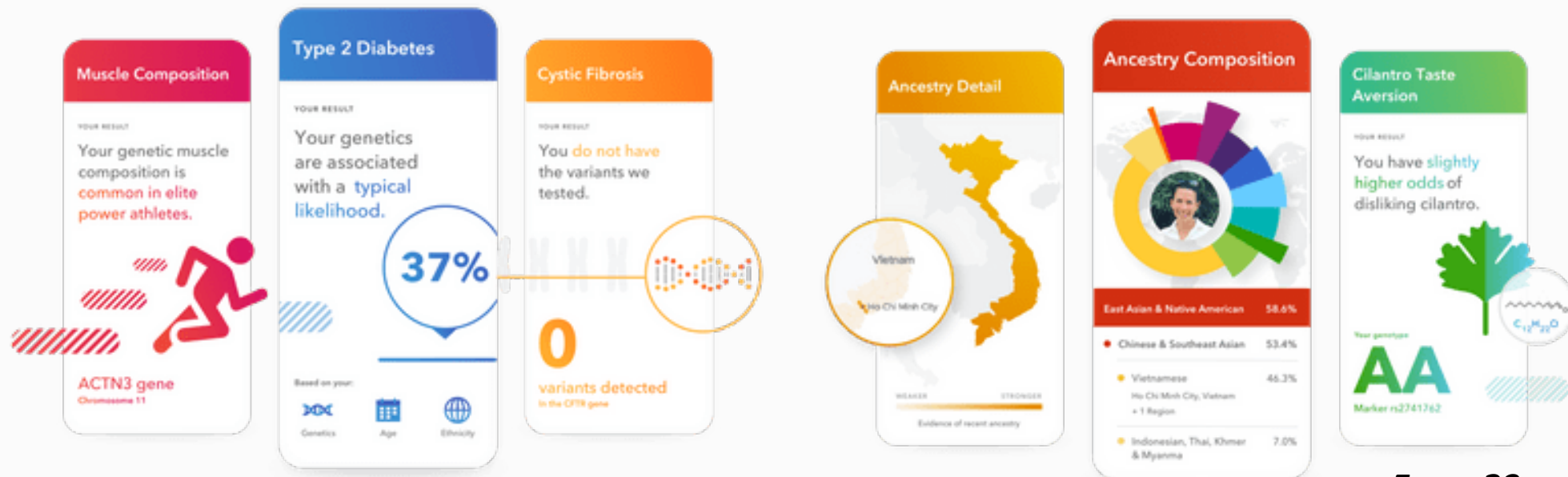
1 year = 200 billion tweets

75%
Businesses Market on Twitter
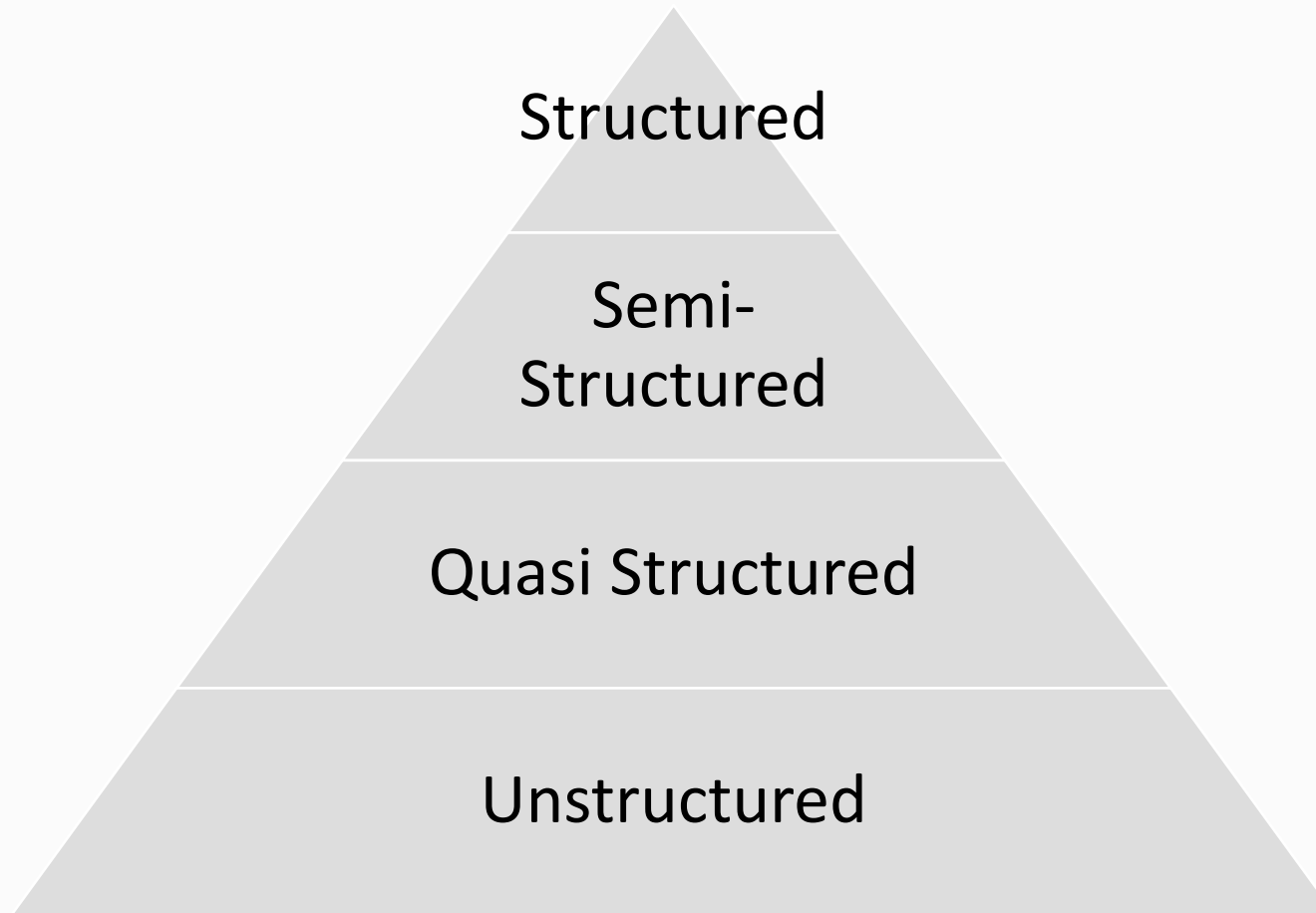
143,199
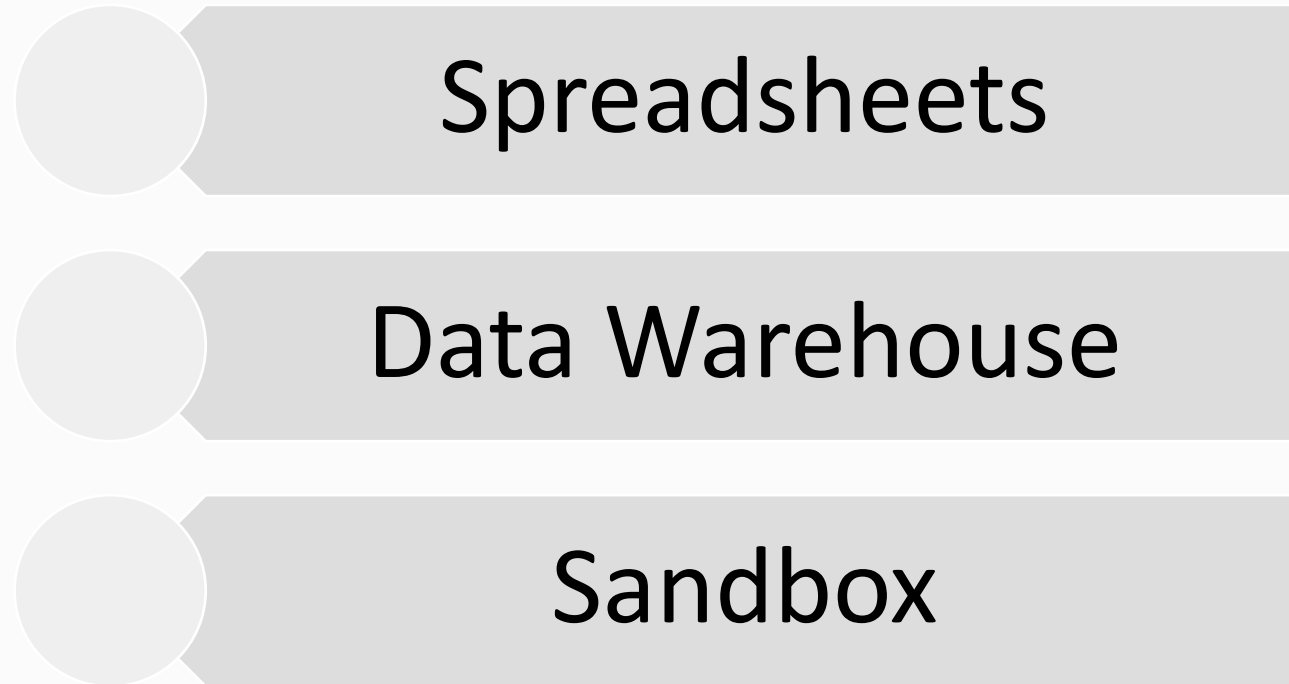Tweets/sec. Japan

By David Sayce

Genetic Sequencing

From 23andme.com

Data Structures

Structured

Semi-Structured

Quasi Structured

Unstructured

Data Repository

Spreadsheets

Data Warehouse

Sandbox
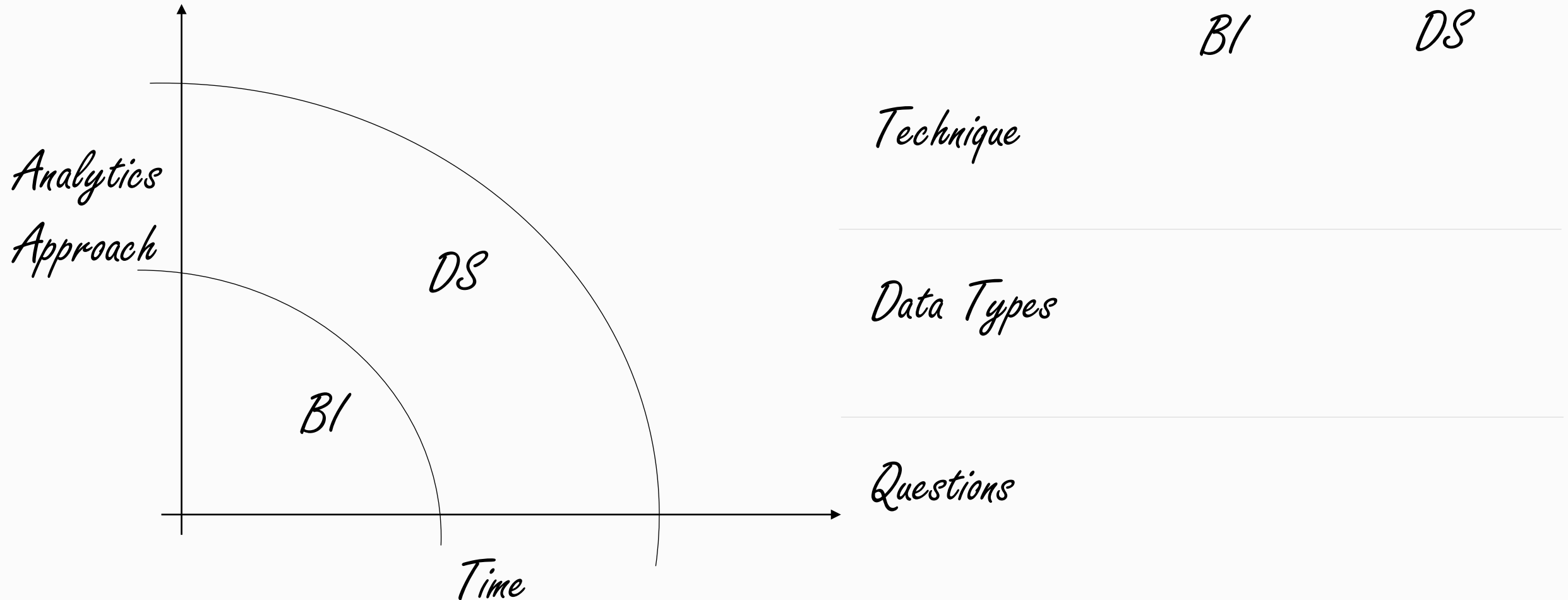
Why Businesses Analyse Data?

- Optimise operations
- Identify risks
- Predict Opportunities
- Comply with laws requirements

Business Intelligence and Data Science

Analytics
Approach

DS

BI

Time

Technique

Data Types

Questions

BI   DS

Data Science

- Datafication!

> **"** *taking all aspects of life and turning them into data.*

The Rise of Big Data, 2013
Kenneth Neil Cukier
Viktor Mayer-Schoenberger

What is Data Science

" *Data science is the civil engineering of data. Its acolytes possess a practical knowledge of tools and materials, coupled with a theoretical understanding of what's possible.*

Quora, 2010
Mike Driscolls - Metamarket CEO
Answer to What is Data Science
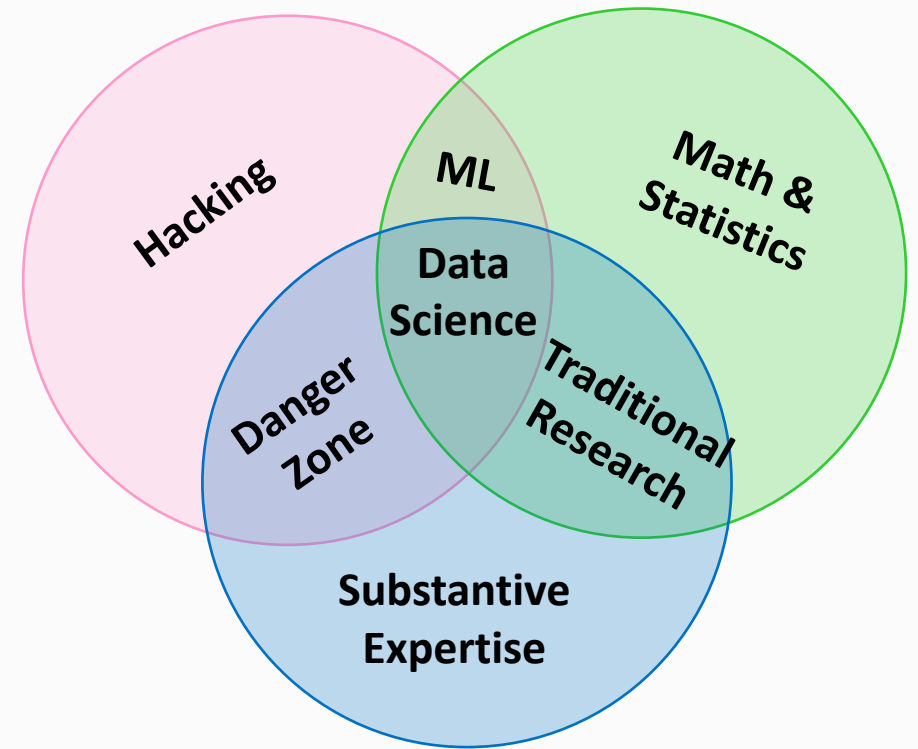
Drew Conway's Venn diagram

Data Science Knowledge Areas

- ***Education for Data Intensive Science to Open New Science Frontiers (EDISON project provides a Data Science framework), 2017***

- Data Science Engineering

- Data Science Analytics

- Data Management

- Research Methods and Project Management
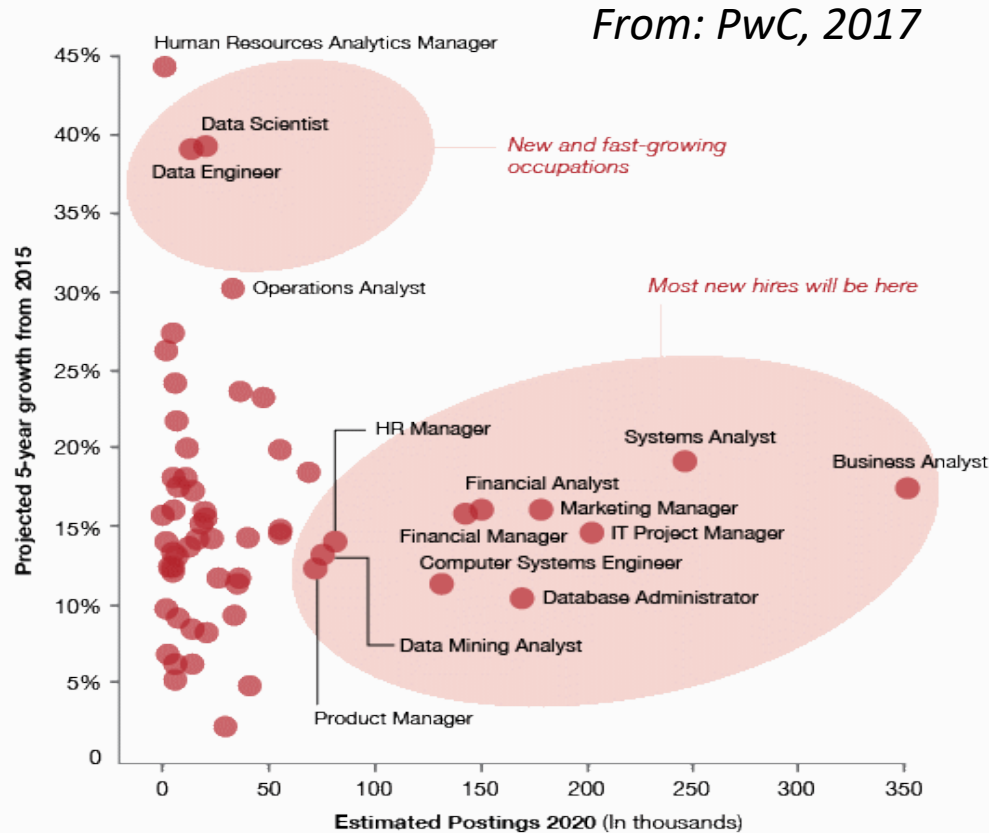
- Business Analytics

Data Scientist

- DJ Patil at LinkedIn and Jeff Hammerbacher at Facebook coined the term 'Data Scientist' as a job title in 2008.

- In 2010, the first page on Wikipedia of data science appeared.
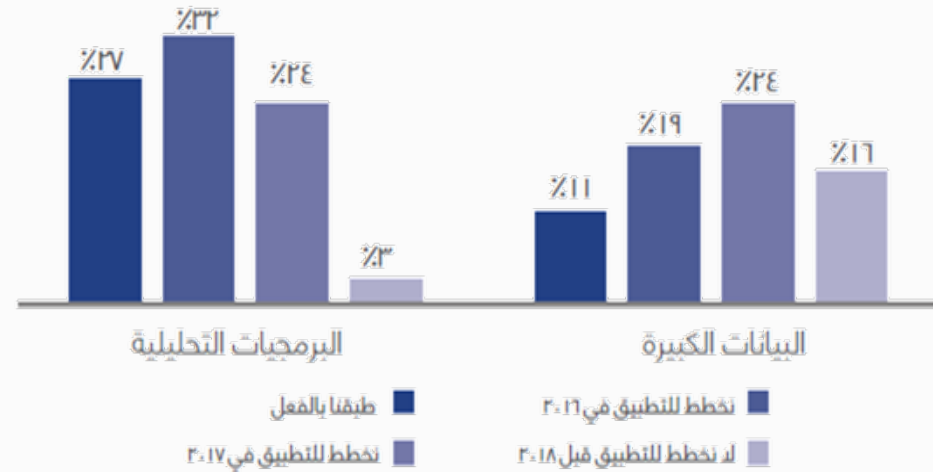
" *Sexiest Job of 21st Century*

Harvard Business Review 2012

Job Market

From: PwC, 2017

From: Vision of the Digital Government in Kingdom of Saudi Arabia, 2018

Job Market

2019

" *Data Scientist is the best job in America for the 4th year in a row*

Data Scientist Leads 50 Best Jobs In America For 2019 According To Glassdoor Lious Columbus

4.7 Job score

4.3 Job Satisfaction
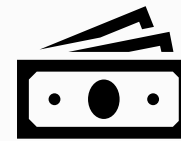
$108,000 Median Salary

6,510 open position

Data Scientist Skills

Quantitative

Technical

Curious

Sceptical

Communicative
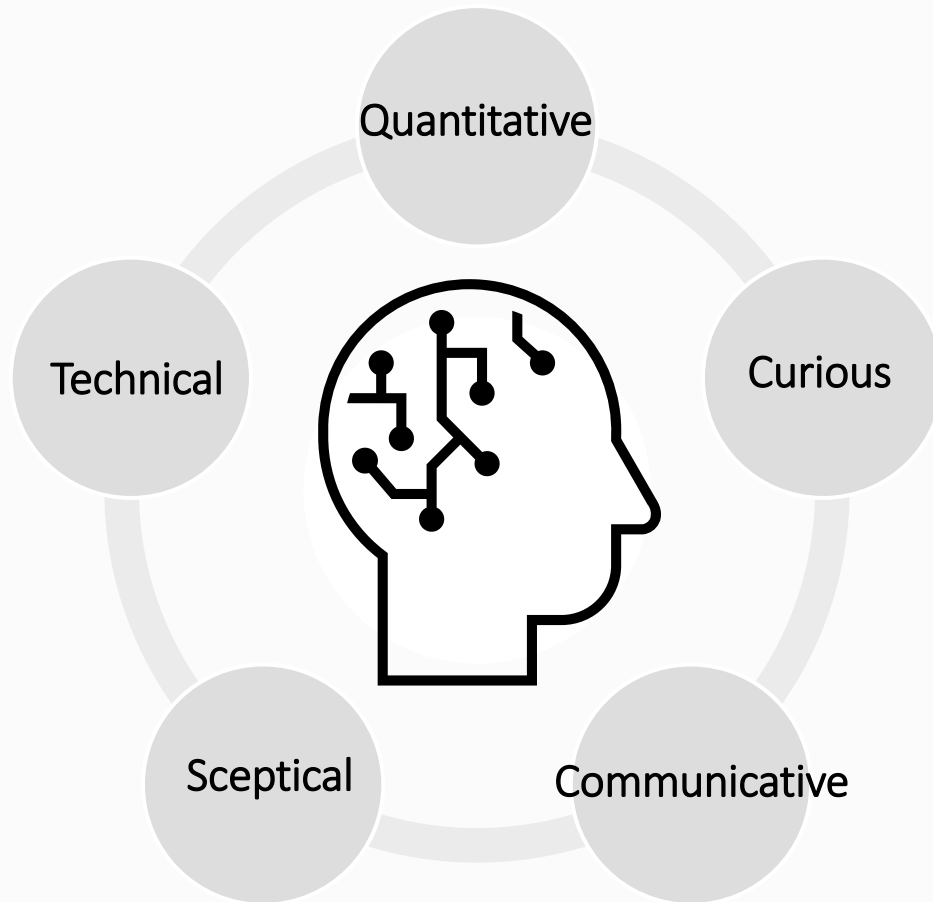
65% need to know Python

53% need to have R programming skills

Machine learning development skills

50% need to know SQL.

15 Data Scientist Jobs That Pay $100K Or More, 2016 Lious Columbus

Career Prospect

- Data Scientist
- Machine Learning Engineer
- Machine Learning Scientist
- Applications Architect
- Enterprise Architect
- Data Architect

- Infrastructure Architect
- Data Engineer
- BI Developer
- Statistician
- Data Analyst

11 Data Science Careers
Shaping the Future
Kelsey Miller, 2020

CIS 6211 Foundations Data Science

Course Overview

Topic	No. of Weeks
Introduction: What is Data Science?	1
Data Science Process	1
Introduction to Machine Learning	2
Data Munging	1
Mathematical Model	2
Data Science Application (Linear Regression, Logistic Regression)	1
Overfitting and Underfitting	1
Visualizing Model Performance	2
Decision Analytic Thinking	2
Other Data Science Tasks and Techniques	2

Course Overview

Lab Topics	
A brief background in python	1

Jupyter and Numpy:
  - Jupyter notebooks
  - learn how to use numpy for data analysis.
  - Simple Python Example
	2

Pandas: - Pandas, built on top of numpy, adds data frames which offer critical data analysis functionality and features.	2
Visualization: - How to visualize large data to gain a better understanding of it.	4
Machine Learning: - Introduce the basics of machine learning	2

Working with Text and Databases:
  - Provide the skills required to access text data.
  - How to analyze text data
	2

Final Project: - Pick the dataset, do the data munging, visualize the data, draw conclusions, and present the results.	2

Course Overview

	Assessment task (i.e., essay, test, quizzes, group project, examination, speech, oral presentation, etc.)	Week Due
1	Assignments	5, 10
2	First Midterm Exam	6
3	Second Midterm Exam	10
4	Lab Continuous Evaluation	1-14
5	Practical Project	14
6	Final Exam	15
	Assessment task (i.e., essay, test, quizzes, group project, examination, speech, oral presentation, etc.)	Week Due

Course Learning Outcomes

1	Knowledge and Understanding
1.1	Describe what Data Science is and the skill sets needed to be a data scientist.
1.2	Describe the Data Science Process and how its components interact.
1.3	Identify the appropriate set of algorithms for a given problem statement
1.4	Identify the appropriate evaluation metric based on a given problem goal
1.5	Recognize the principles and techniques for optimizing the performance of Python numeric applications
2	Skills :
2.1	Create effective visualization of given data
2.2	Develop a solution strategy to data science problems, including problem formulation, exploratory analysis, modeling, evaluation, implementation, and feedback
2.3	Develop Python applications that utilize big data services
2.4	Operate basic exploratory analysis to identify abnormalities in data
2.5	Write relatively advanced, well structured, computer programs in Python
3	Values:
3.1	Work effectively in teams on data science projects.
3.2	Justify model selection using cross-validation

References

- Data Science: Concepts and Practice. 2nd Edition by Vijay Kotu and Bala Deshpande, ISBN: 978-0-12-814761-0©2019.

- THE Data Science Design MANUAL. Steven S. Skiena, ISBN: 978-3-319-55444-0 ©2017.

- Data Science for Business. Foster Provost &Tom Fawcett, ISBN-13: 978-1449361327, O'Reilly ©2013.

Areej Alasiry

1/29/2024

Thanks