INTRODUCTION TO DATA SCIENCE

VISHWA NATH JHA
CEO & CO FOUNDER
GAMUTDATA CONSULTING PVT LTD


WHO AM I?

ABOUT THE TRAINER

- *I'm a Technology Consultant working with different clients in personal capacity.*
- *As an Analytics Trainer, I train professionals on tools & technologies like **R, Python, SAS, Excel & VBA, Hadoop, MapReduce, Spark, HIVE, Pig, SQL, Statistics, Tableau, Machine Learning, etc.***
- *I'm busy building a couple of products dedicated to Data Science Industry in Ed-Tech & Human Resource space*

If you've any queries, feel free to reach out

eMail: Vishwanath.jha@gamutdata.in

LinkedIn: https://in.linkedin.com/in/vishwanathjha1

Phone: +91 9167240332 | Website: www.gamutdata.in

OUTLINE

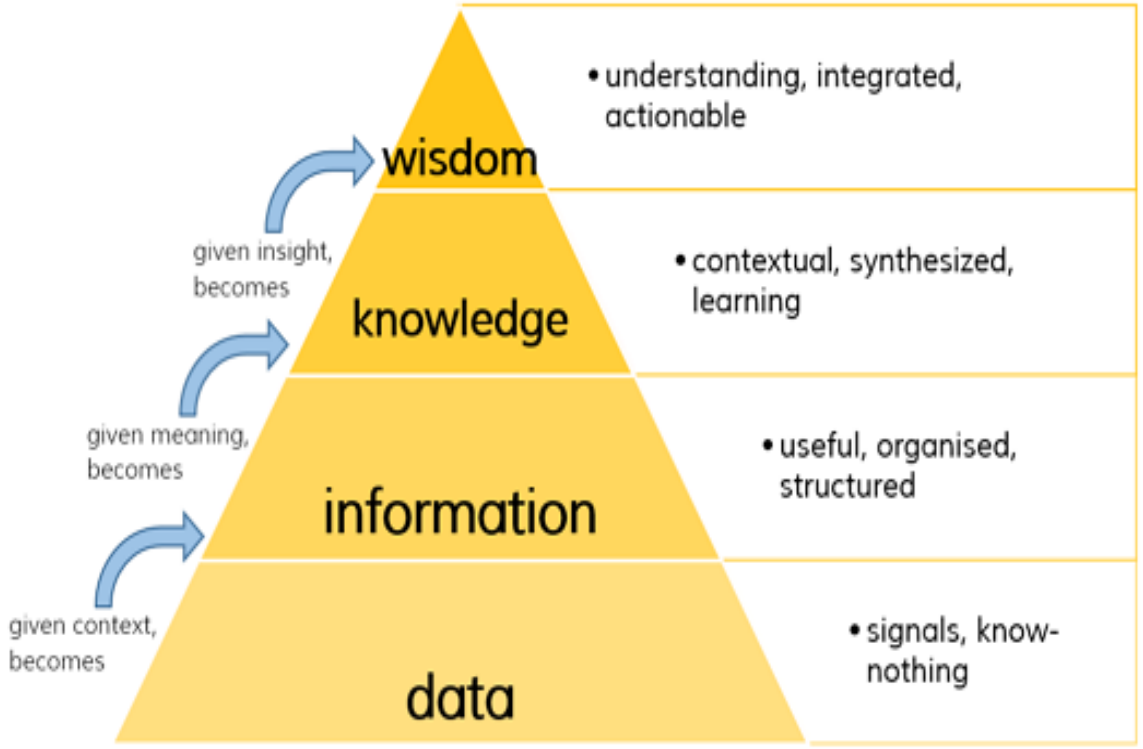
- ✓Data: An Introduction
- ✓The Growth Story of Data
- ✓Problems Associated with The Growth Story
- ✓The Rise of Data Scientist
- ✓Opportunity
- ✓Super Powers of *The Data Scientist*
- ✓About This Course
- ✓Q&A

DATA: AN INTRODUCTION

- Data are raw facts and figures that on their own have no meaning

- These can be any alphanumeric characters i.e. text, numbers, symbols

WHY DATA IS IMPORTANT?



SOURCE OF DATA

ARCHIVES

Archives of scanned documents, statements, insurance forms, medical record and customer correspondence, paper archives, and print stream files that contain original systems of record between organizations and their customers

DOCS

XLS, PDF, CSV, email, Word, PPT, HTML, HTML 5, plain text, XML, JSON, etc.

MEDIA

Images, videos, audio, Flash, live streams, podcasts, etc.

DATA STORAGE

SQL, NoSQL, Hadoop, doc repository, file systems, etc.

BUSINESS APPS

Project management, marketing automation, productivity, CRM, ERP content management systems, HR, storage, talent management, procurement, expense management, Google Docs, intranets, portals, etc.

PUBLIC WEB

Government, weather, competitive, traffic, regulatory, compliance, health care services, economic, census, public finance, stock, OSINT, the World Bank, SEC/Edgar, Wikipedia, IMDb, and other Web services

SOCIAL MEDIA

Twitter, LinkedIn, Facebook, Tumblr, Blog, SlideShare, YouTube, Google+, Instagram, Flickr, Pinterest, Vimeo, Wordpress, IM, RSS, Review, Chatter, Jive, Yammer, etc.

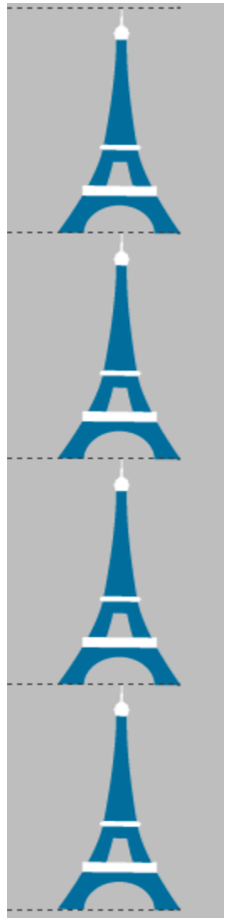
MACHINE LOG DATA

Event logs, server data, application logs, business process logs, audit logs, call detail records (CDRs), mobile location, mobile app usage, clickstream data, etc.

SENSOR DATA

Medical devices, smart electric meters, car sensors, road cameras, satellites, traffic recording devices, processors found within vehicles, video games, cable boxes or household appliances, assembly lines, office buildings, cell towers and jet engines, air conditioning units, refrigerators, trucks, farm machinery, etc.

DATA: THE GROWTH STORY



FEW FACTS ABOUT DATA EXPLOSION

- 90% of data generated is “Unstructured”
  - Includes tweets, photos, customer purchase history & even customer service call logs
- Global internet population grew 14.3% between 2011 & 2013
- 3 Billion: The number of people who have access to the internet today equals that of the world’s population in 1960
- Data generation by 2018 will be 50,000 GB per second
- Amount of data generated since the dawn of mankind has doubled in years from 2011 to 2013 and is expected to grow exponentially

CHALLENGES

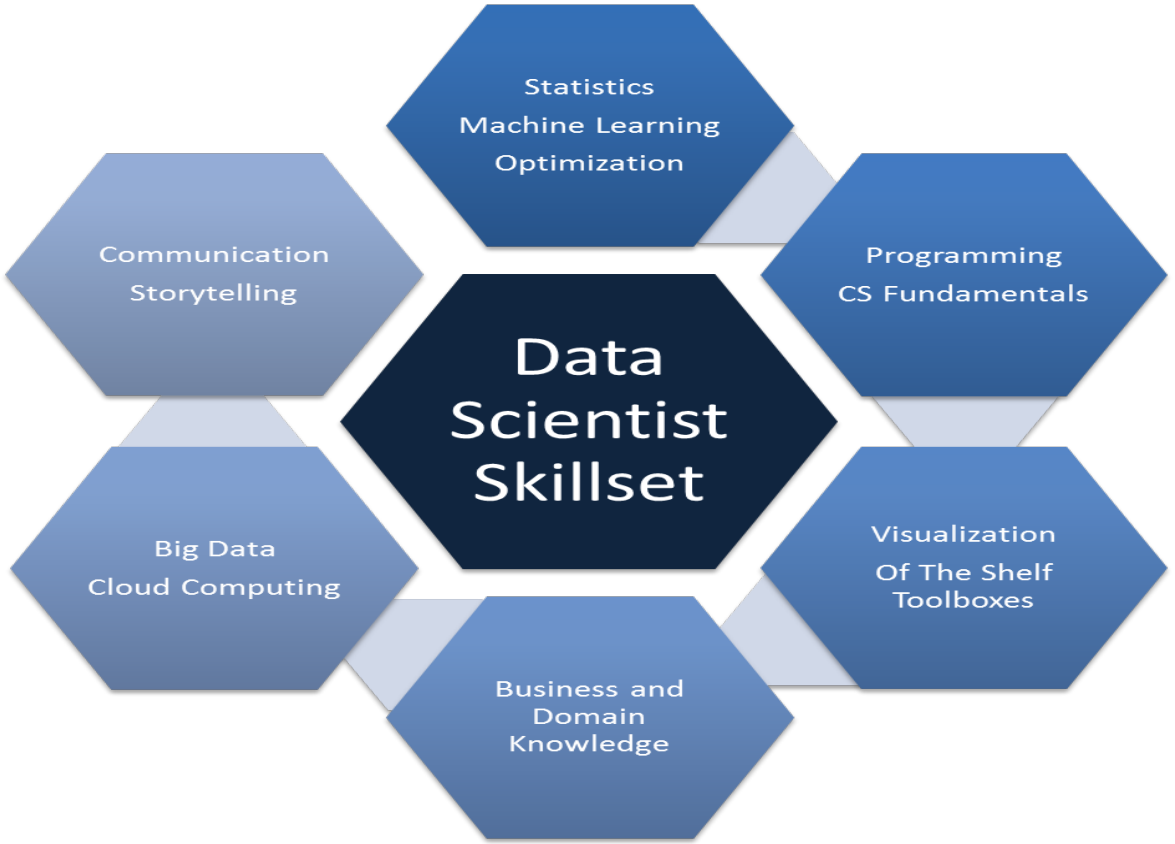
- Multiple Data Sources
  - Data is spread out like universe
  - Finding Relevant Data
- Data Crunching On Scale
  - Less than 0.5% of all data we create is ever analyzed and used.
  - Conventional Infrastructure Fails to Tackle Diversity & Volume
- Customers Insist on Interacting
  - Communication of Insights in form of Interactive Dashboards
- Knowledge of the future
  - State-of-the-art Algorithms
  - Computing Powers

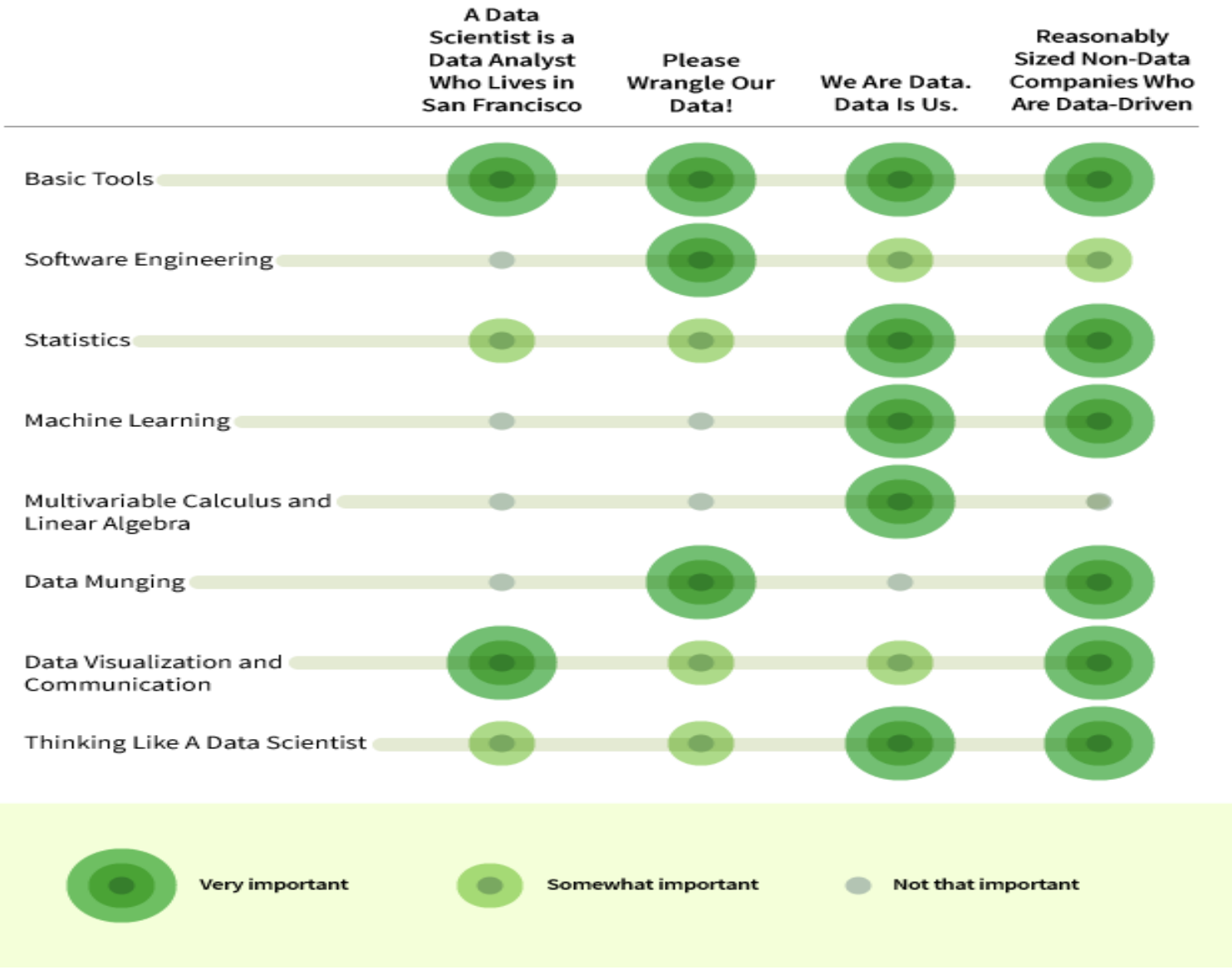
THE RISE OF DATA SCIENTIST

WHAT THE WORLD HAS TO SAY?

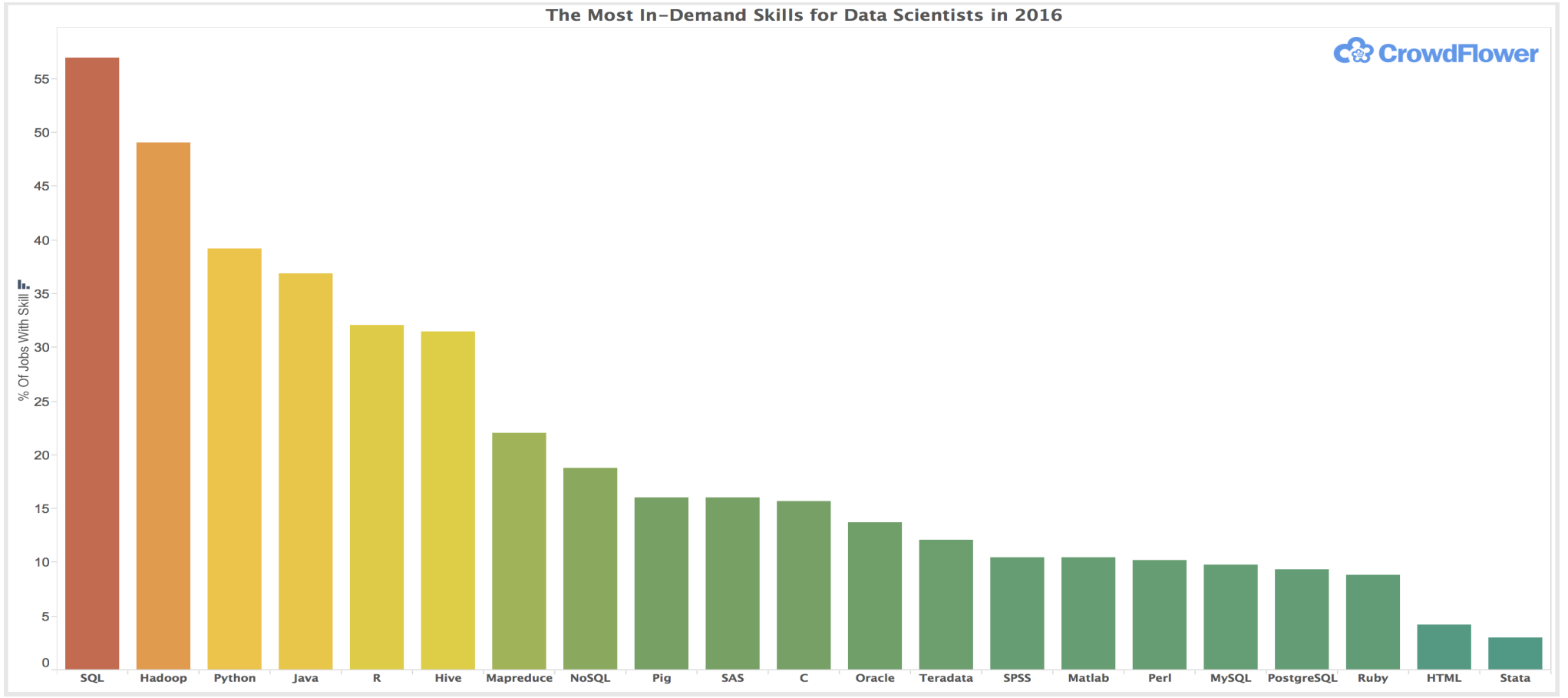
"The nerdy-cool job that companies are scrambling to fill¨	**Fortunes, 2016**
"Data Scientists: The Definition of Sexy"	**Forbes, 2012**
"The Sexiest Job of 21st Century"	**Harvard Business Review, 2012**
"The Best Job of Year 2016"	**Glassdoor**
"Why "Data Scientist" Is The Best Job To Pursue In 2016"	**Forbes**

SUPER POWERS OF DATA SCIENTISTS

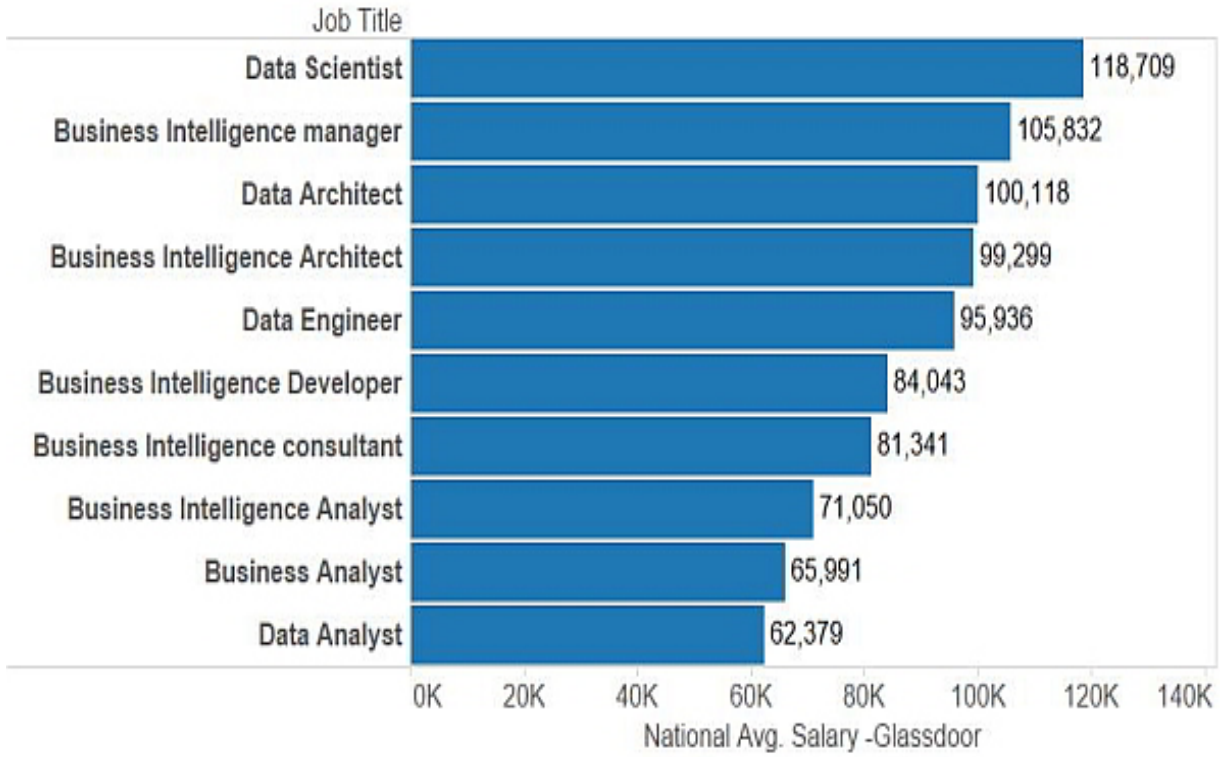



A Data Scientist is a Data Analyst Who Lives in San Francisco
Please Wrangle Our Data!
We Are Data. Data Is Us.
Reasonably Sized Non-Data Companies Who Are Data-Driven
Basic Tools
Software Engineering
Statistics
Machine Learning
Multivariable Calculus and Linear Algebra
Data Munging
Data Visualization and Communication
Thinking Like A Data Scientist
Very important
Somewhat important
Not that important

MOST SOUGHT AFTER SKILLS IN DATA SCIENTISTS



OPPORTUNITY



OPPORTUNITY

- Four in ten (43%) companies report their lack of appropriate analytical skills as a key challenge
- The median salary of a junior level data scientist is $91,000 but those managing a team of ten or more data scientists earn base salaries of well over $250,000, according to *Burtch Works*.
- When changing jobs, data scientists see a 16 percent increase in their median base salary
- International Data Corporation (IDC) predicts a need for 181,000 people with deep analytical skills in the US by 2018 and a requirement for five times that number of positions with data management and interpretation capabilities.

ABOUT THIS COURSE

- Fundamentals of R
- Statistics with R
- Data Manipulation with R
- Visualization with R
- Machine Learning With R

- Fundamentals of Python
- Statistics with Python
- Data Manipulation with Python
- Visualization with Python
- Machine Learning with Python

- Tableau- Guest Lecture
- Conclusion


Q&A

thank you
danke
рахмат
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
mersi
kia ora
barka
welalin
tack
spas
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
gracias
vinaka
спасибі
blagodaram
kiitos
dankie
dhanyavad
nanni
nandri
mauruuru
köszönöm
hvala
gràçie
bayarlalaa
bedankt
enkosi
dziękuję
akun
dankon
ačiū
хвала
asante
manana
obrigada
tenki
djiere dieuf
tau
mochchakkeram
дякую
mamnun
murakoze
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
sobodi
děkuji
obrigado
sagolun
sukriya
kop khun krap
arigatô
takk
dakujem
trugarez
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
grazie
kam sah hamnida
rahmat
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
mercé
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci