



**OBÓZ NAUKOWY
POLITECHNIKI WARSZAWSKIEJ**

Sztuczna Inteligencja, dzień 2, część 1

Data Science – kto, po co i dlaczego?



Kto?

Data scientist – termin określający zawód osoby zajmującej się analizą danych nieuporządkowanych, w szczególności big data.

Musi on:

- wykazać wszechstronne umiejętności poruszania się w niespójnych, różnorodnych zbiorach danych, związane z technologiami, które obecnie bardzo dynamicznie ewoluują
- odnosi się do jego cech osobowości: jest zaangażowany w zrozumienie biznesu - potrafi dostrzec szeroki kontekst swoich analiz, wyjść poza schematyczne rozwiązania, przenieść rozwiązanie z innej gałęzi biznesowej, myśli innowacyjnie i strategicznie, jest ciekawy świata i patrzy na niego przez pryzmat danych z różnorodnych źródeł, rozwiązuje problemy, ma zacięcie hackerskie
- odnosi się do jego miękkich umiejętności: zdolność do komunikacji - także z wyższą kadrami zarządzającą, zdolność do komunikacji poprzez wysokiej jakości wizualizacje, umiejętność wyjaśnienia wykonanych złożonych analiz
- często posiada kompetencje analityka i / lub administratora baz danych służących do przetwarzania dużych zbiorów danych (ang.



Po co?

- Oszczędność
 - Czasu
 - Zasobów
 - Pieniędzy
- Zarabianie
- Bezpieczeństwo
- Rozwój innych gałęzi nauki
 - Fizyki
 - Chemii
 - Biologii
- Dostarczanie wygodniejszych usług



THE LANDSCAPE OF BIG DATA

Big data is projected to grow into a **\$53.4 BILLION** market by 2017, up from **\$10.2 BILLION** in 2013

All of the world's digital data equals about **900 exabytes**, 70% of which is created by individuals

China will account for more than **1/5** of the world's data by 2020

90% of the data in the world today has been created in the last two years alone

1 terabyte = 1000 gigabytes
1 petabyte = 1000 terabytes
1 exabyte = 1000 petabytes
1 zettabyte = 1000 exabytes

1 EB = 1 billion gigabytes or 250 billion DVDs
1 EB = nearly 2 times as large as the web archive at the US Library of Congress

WWW
MORE THAN 570 NEW WEBSITES are created every minute of the day

Average online cart sizes increase 10%-15% as a result of personalized advertising using customer information obtained through big data

MOBILE
Global mobile data traffic grew 81% IN 2013
OVER 500 MILLION mobile devices and connections were added in 2013
Mobile traffic in 2013 was nearly 18 TIMES the size of the entire global Internet in 2000

IMPACT OF BIG DATA
Poor data across businesses and the government costs the U.S. economy \$3.1 TRILLION/YEAR
Accessing 10% more data equates to an additional income of \$65.7 MILLION for the average Fortune 1000 company

VIDEO
Percentage of all online traffic that was mobile video: 53% IN 2013, 69% IN 2018
100 HOURS of video is uploaded to YouTube every minute

WHICH IS WHY
34% of companies implemented big data initiatives in 2013, up from 27% in 2012
68% of companies are running two or more big data projects as part of their big data initiatives
15% of companies with Big Data initiatives spent at least \$100 million each on those initiatives in 2012: 7% invested at least \$500 million and 24% invested less than \$2.5 million a piece

BY 2015 4.4 MILLION IT JOBS globally will be created to support big data, generating 1.9 MILLION IT jobs in the United States

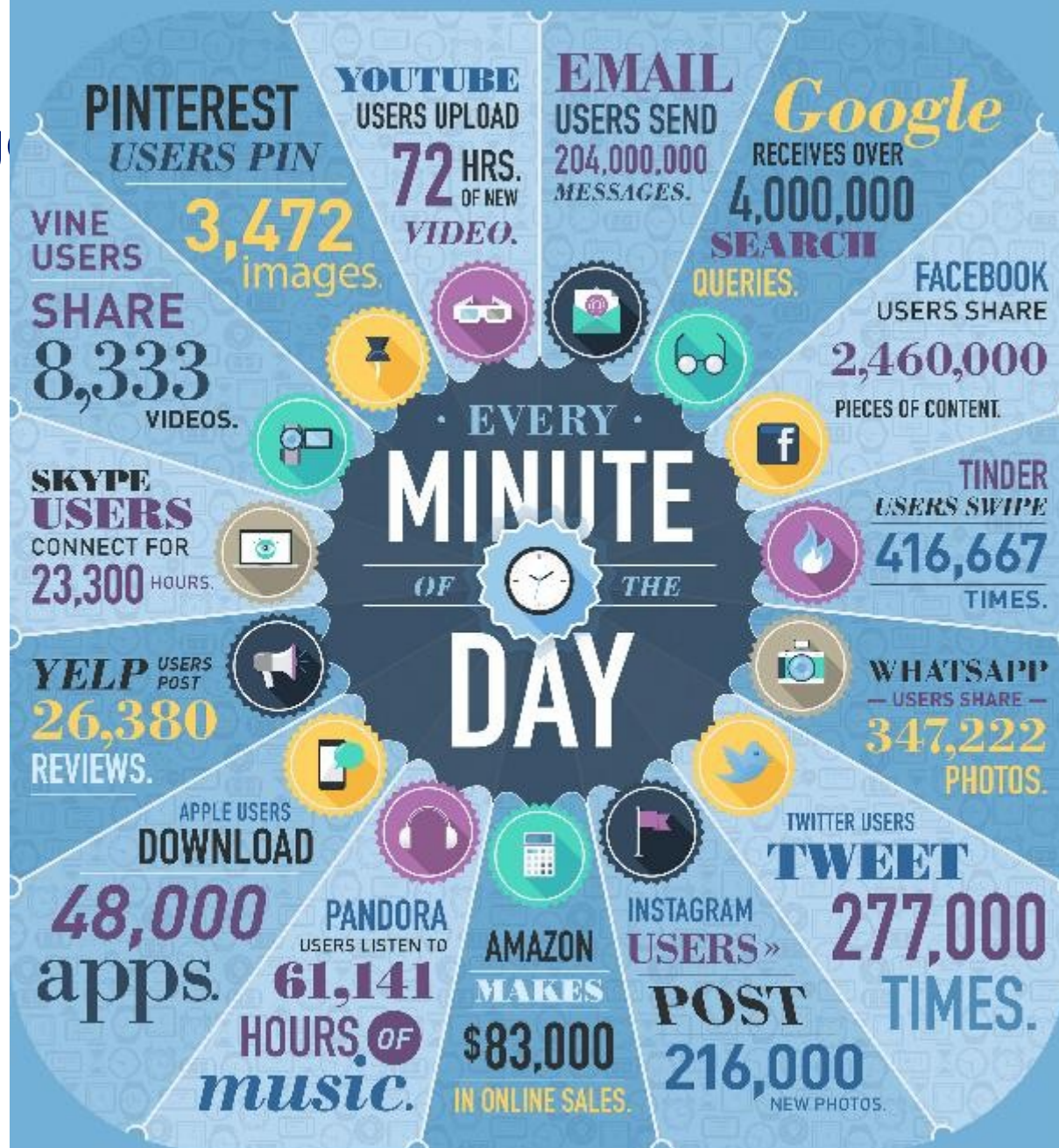
BY 2018 The United States alone could face a shortage of 140,000 to 190,000 PEOPLE with deep analytical skills as well as 1.5 MILLION managers and analysts with the know how to use the analysis of big data to make effective decisions.

The White House administration is investing \$200 MILLION into big data research projects

mushroom NETWORKS
www.mushroomnetworks.com

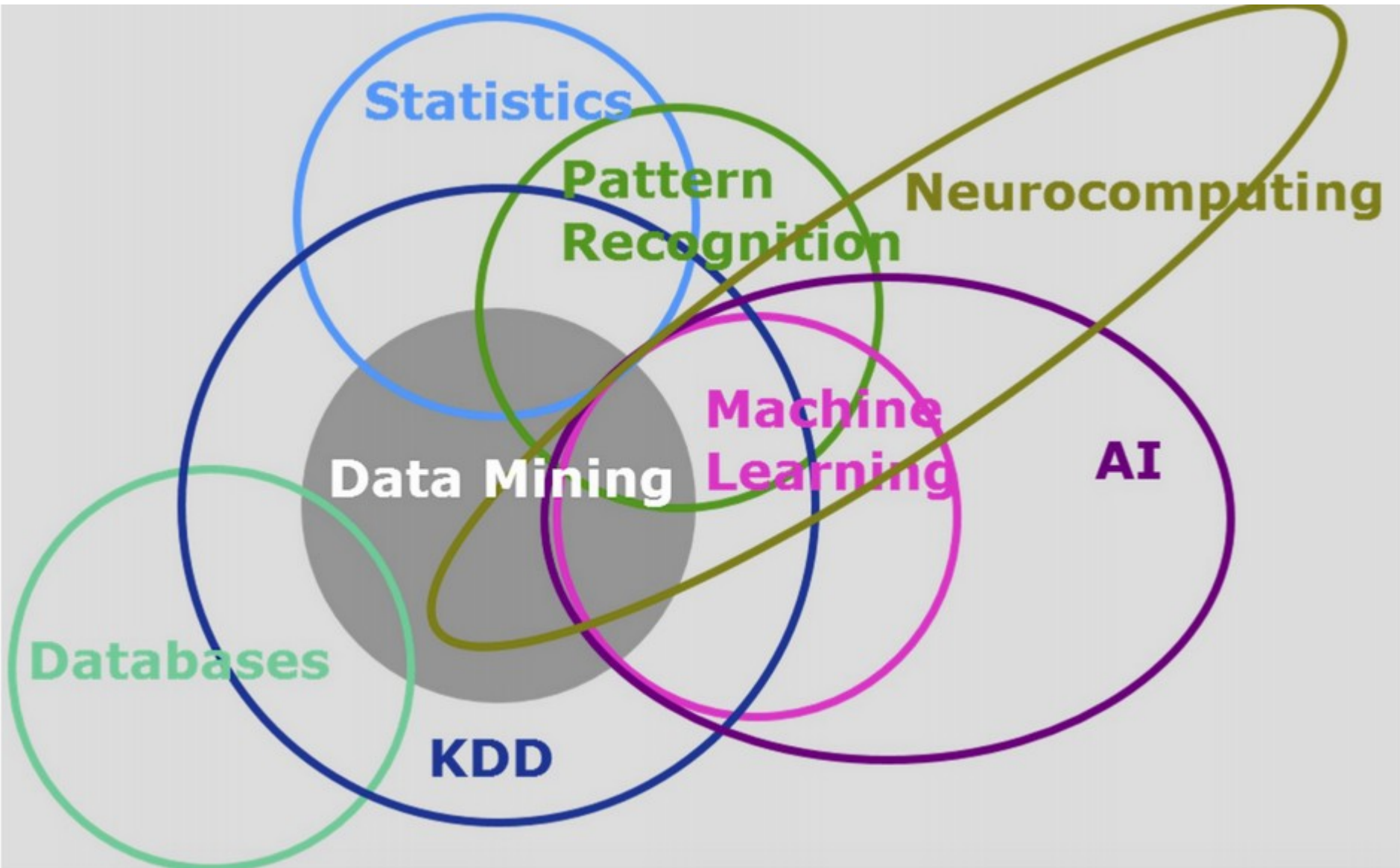
SOURCES:
http://www.cisco.com/c/en/us/solutions/collateral/service_provider/visn/it-environment/index.html
http://www.youtube.com/yt/press/vid/traffic.html
http://www.wired.com/wired/archive/19.03/big_data.html
http://www.boston.com/business/analyses/big_data/abstracts/reporting

Dlaczego



THE GLOBAL INTERNET POPULATION GREW **14.3%** FROM 2011 - 2013 AND NOW REPRESENTS

2.4 BILLION PEOPLE.



DATA PRODUCTS

“A PRODUCT THAT FACILITATES AN END GOAL
THROUGH THE USE OF DATA”

— DJ PATIL, CHIEF DATA SCIENTIST OF THE U.S.





data oil

is the new

we need to find it,
extract it, refine it,
distribute it and
monetize it.

David Buckingham



OBÓZ NAUKOWY
POLITECHNIKI WARSZAWSKIEJ

Wykorzystywane narzędzia

- **Python**
 - Numpy
 - Pandas
 - Scikit-learn kit
- R
- Julia
- SAP

