BIG DATA
Modern usage of the term “Big Data”

- predictive analytics,
- user behaviour analytics,
- certain other advanced data analytics methods that extract value from data
- by the year 2020, 1.7 megabytes of data will be generated every second for every person on the planet

Big Data

- Big data is a term used to refer to the study and applications of data sets that are too complex for traditional data-processing application software to adequately deal with

Big Data Challenges

- capturing data,
- data storage,
- data analysis, search, sharing, transfer,
- visualization, querying, updating,
- information privacy and data source

Case Study – Insurance Company

1. Risk Assessment - Hancock’s Vitality program
2. Fraud Detection
3. Customer Insights
4. Marketing
5. Customer Experience

The Jobs Landscape in 2022

LinkedIn

These will be the 10 most in-demand skills in 2017 according to LinkedIn
• Big data is a term used to refer to the study and applications of data sets that are too complex for traditional data-processing application software to adequately deal with
Big Data Challenges

- capturing data,
- data storage,
- data analysis, search,
- sharing, transfer,
- visualization, querying,
- updating,
- information privacy and
- data source
Modern usage of the term “Big Data"

- predictive analytics,
- user behaviour analytics,
- certain other advanced data analytics methods that extract value from data
- by the year 2020, 1.7 megabytes of data will be generated every second for every person on the planet
A Modern Data Platform is born

New data sources

Social → Mobile → Data Services → Sensors → Location

“New” relational databases
Open Source SQL
- MariaDB
- PostgreSQL
- EDB EnterpriseDB
- SQL compatible
- MySQL compatible

New data models - NoSQL
Document, Column, Key Value, Graph
- MongoDB
- Cassandra
- Redis
- Kinetica
- Neo4j
- TITAN
- Hortonworks

New data models
Hadoop/Spark
Unstructured Data
- Hadoop
- Spark

Conventional data platform

Transactional
Case Study – Insurance Company

1. Risk Assessment - Hancock's Vitality program
2. Fraud Detection
3. Customer Insights
4. Marketing
5. Customer Experience

<table>
<thead>
<tr>
<th>Where Big Data is Expected to Help</th>
<th>Now*</th>
<th>Two years from now</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing, underwriting, risk selection</td>
<td>42%</td>
<td>77%</td>
</tr>
<tr>
<td>Better management decisions</td>
<td>19%</td>
<td>60%</td>
</tr>
<tr>
<td>Loss control and claim management</td>
<td>17%</td>
<td>58%</td>
</tr>
<tr>
<td>Understanding customer needs</td>
<td>17%</td>
<td>50%</td>
</tr>
<tr>
<td>Product development</td>
<td>19%</td>
<td>48%</td>
</tr>
<tr>
<td>Marketing/Distribution/Sales</td>
<td>15%</td>
<td>48%</td>
</tr>
</tbody>
</table>

*Survey fielded September 9 – November 2, 2015
The Jobs Landscape in 2022

Top 10 Emerging
1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. General and Operations Managers
4. Software and Applications Developers and Analysts
5. Sales and Marketing Professionals
6. Big Data Specialists
7. Digital Transformation Specialists
8. New Technology Specialists
9. Organisational Development Specialists
10. Information Technology Services

Top 10 Declining
1. Data Entry Clerks
2. Accounting, Bookkeeping and Payroll Clerks
3. Administrative and Executive Secretaries
4. Assembly and Factory Workers
5. Client Information and Customer Service Workers
6. Business Services and Administration Managers
7. Accountants and Auditors
8. Material-Recording and Stock-Keeping Clerks
9. General and Operations Managers
10. Postal Service Clerks

These will be the 10 most in-demand skills in 2017 according to LinkedIn
Learn Python?

**Features**

- Simple - feels almost like reading English
- Easy to learn - extraordinarily simple syntax
- Free and Open Source
- High-level Language
- Interpreted - does not need compilation to binary. You just run the program directly from the source code
- Object Oriented
- Packages: Numpy, Matplotlib (visualisation data) & Scikit-learn (machine learning models)

**TREK PYTHON FOR PROGRAMMER**

- Introduction to Python for Data Science
- Intermediate Python for Data Science
- Python Data Science Toolbox (Part I)
- Python Data Science Toolbox (Part 2)
- Importing Data in Python (Part 1)
- Importing Data in Python (Part 2)
- Cleaning Data in Python
- Pandas Foundation
- Visualising Distributions with Matplotlib
- Introduction to Databases in Python
world's fastest growing and most popular programming language used by software engineers, analysts, data scientists, and machine learning engineers alike.
Features

- Simple - feels almost like reading English
- Easy to learn - extraordinarily simple syntax
- Free and Open Source
- High-level Language
- Interpreted - does not need compilation to binary. You just *run* the program directly from the source code
- Object Oriented
- Package - Numpy, Matplotlib (pemvisualan data) & Scikit-learn (pembelajaran mesin).
Learn Python?

SYNTAX
Programming rules

ALGORITHM
Solving Problem
Learn Python?

**indexing**

```python
>>> print data [2]
```

**slicing**

```python
Extracting data
>>> print data [2:4]
```
Program Pembangunan Bakat Analitik Data Raya (Pioneering Big Data Analytics Talent Development Program)

Melalui penubuhan empat belas (14) unit Makmal Data Raya di kampus-kampus di UiTM dijangka akan membantu negara dalam meningkatkan jumlah data profesional dalam bidang analitik data raya di Malaysia.

Program ini dijangka akan menghasilkan seramai 2,340 bakat analitik data raya setahun yang mana sekurang-kurangnya seramai 180 orang perlu dihasilkan di setiap Fakulti/Kampus UiTM setahun.
TREK PYTHON FOR PROGRAMMER

Trek ini diajar kaedah menggunakan kemahiran pengaturcaraan untuk mengumpulkan data (data wrangling) dan membina alatan (tool development) untuk analisa data

1. Intro to Python for Data Science
2. Intermediate Python for Data Science
3. Python Data Science Toolbox (Part 1)
4. Python Data Science Toolbox (Part 2)
5. Importing Data in Python (Part 1)
6. Importing Data in Python (Part 2)
7. Cleaning Data in Python
8. Pandas Foundation
9. Manipulating DataFrames with Pandas
10. Introduction to Databases in Python
PENSIJILAN

STATEMENT OF ACCOMPLISHMENT

HAS BEEN AWARDED TO

Stephane Gilabert

FOR SUCCESSFULLY COMPLETING

Intro to Python for Data Science Course

DataCamp
Certification LINKEDIN

Accomplishments

1 Certification

Intro to Python for Data Science
License ce409e020c62c42c637a4377477db41a191cafd2

DataCamp

See certificate
THE SMARTEST WAY TO

Learn Data Science Online

The skills people and businesses need to succeed are changing. No matter where you are in your career or what field you work in, you will need to understand the language of data. With DataCamp, you learn data science today and apply it tomorrow.

Start Learning For Free

Create Your Free Account

Email address

Password

Create Free Account

By continuing you accept the Terms of Use and Privacy Policy, that your data will be stored outside of the EU and that you are 16 years or older.