RIGACODINGSCHOOL

DATU ANALĪZES (ar PYTHON pamatiem) apmācības

PAR MUMS

SKOLA
Akreditēta & moderna

FILIĀLES
Rūpniecības iela &
Republikas laukums

ABSOLVENTI
700+

KARJERAS CENTRS
Mēs mainām cilvēku
profesionālo karjeru

SADARBĪBA
70+ uzņēmumi

PIEREDZE
30+ pasniedzēji
5+ gadi apmācību
organizēšanā

PAR KURSIEM

Intensīvs &
koncentrēts kurss

Saņem **sertifikātu**

Lekcijas tiek
ierakstītas un
arhivētas

Atkārto kursu
bezmaksas 1
gada laikā

Apmeklē **bezmaksas**
vieslekcijas

20% teorija un **80%**
prakse

VALDIS

- Izglītība: Maģistra grāds datorzinātnēs
- Pieredze programmēšanā: 20+ gadi
- Specialitāte: grafu teorija sociālo tīklu analizēšanā
- Hobiji: prāta spēles, riteņbraukšana, šahs

Data Lake

Brief History of Data Analysis

- What is this ? :)

Brief History of Data Analysis

- ~ 18,000BC – Uganda, Ishango Bone
- ~ 2400BC – Babylon abacus, libraries
- 300BC – 48AD – Library of Alexandria

Brief History of Data Analysis

- How about this modern recreation of a 2000 years old device?

Brief History of Data Analysis

- **~ 100-200AD**
  Antikythera Mechanism

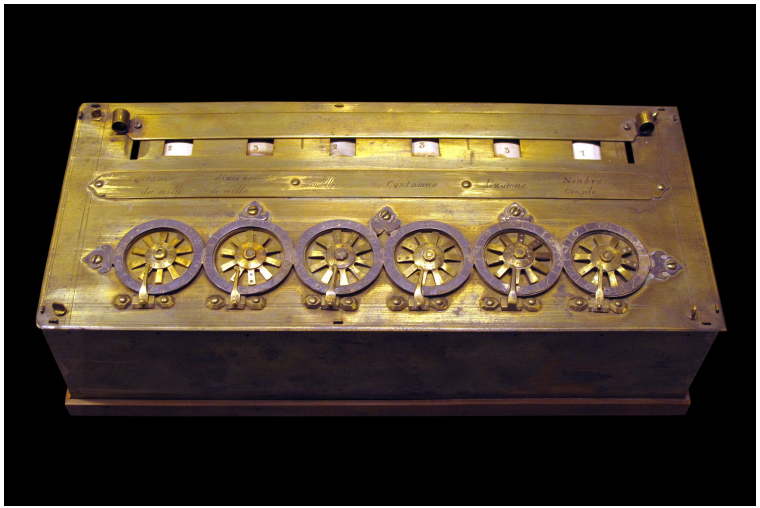
  Predicting:
  - Astrology
  - Astronomy
  - Olympics
  - Calendar

Brief History of Data Analysis

- **1642 Blaise Pascal's Pascaline**

Performs:

- Addition
- Subtraction
- Multiply/Divide using Add/Sub
- 1649 Royal Patent by Louis XIV

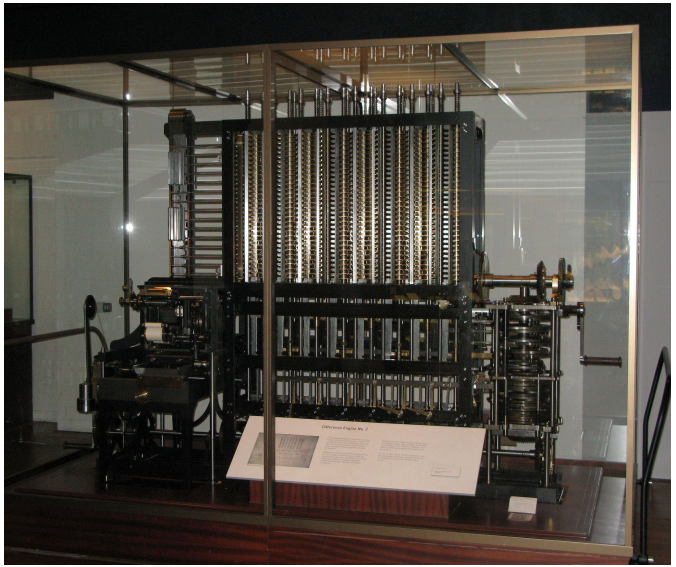


Brief History of Data Analysis

- **Charles Babbage**
  Difference Engine
- Ada Lovelace - first programmer?

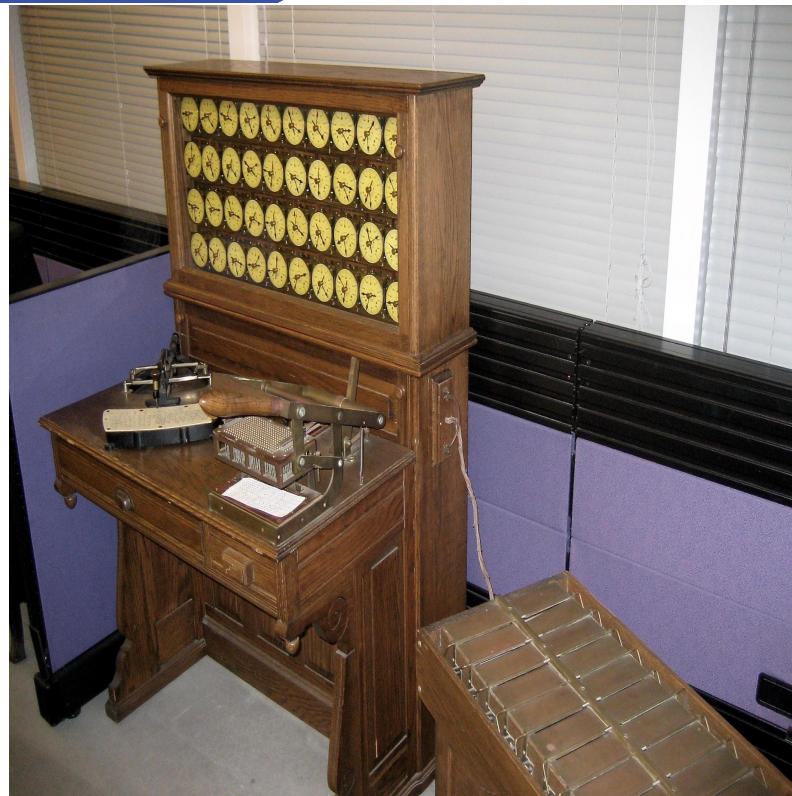
Performs:

- Arithmetic
- Derivation
- Power Series
- Curve Fitting



Brief History of Data Analysis II

- 1663 – London, J.Graunt mortality analysis
- 1865 – banker H. Furnese business intelligence
- 1880-90 US Census Hollerith Machine -> IBM
- 1928 – F. Pfleumer magnetic tape invention

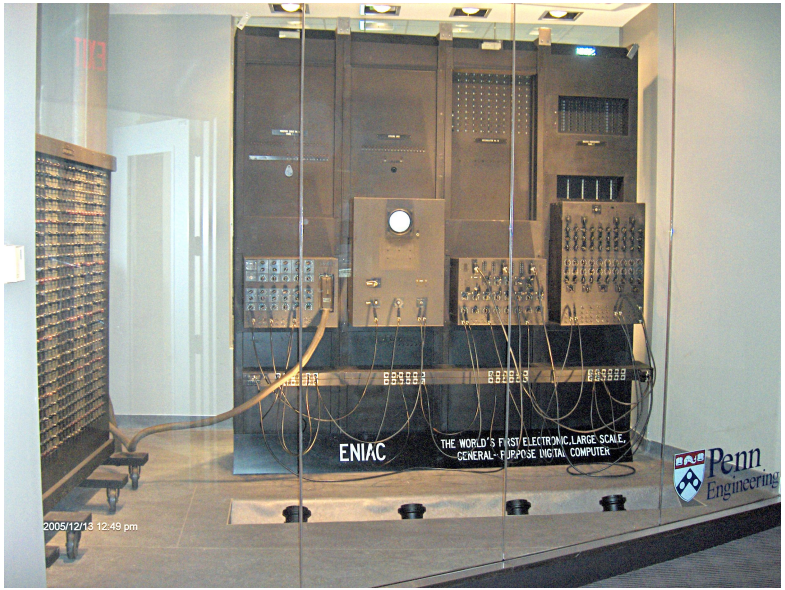


Brief History of Data Analysis III

- 1940s - First General Purpose Electronic Computer ENIAC (Zuse mechanical)

- Turing complete
- von Neumann architecture

- most computers work the same today



Brief History of Data Analysis III

- 1950s - Flat Files
- 1958 – IBM's Luhn defines Business Intelligence
- 1960s - CODASYL
- 1970s – Codd's relational DBs -> SQL
- 1980s – Data Werehouses / Marts
- 2000s – Big Data / noSQL DBs

BIG DATA LANDSCAPE 2017

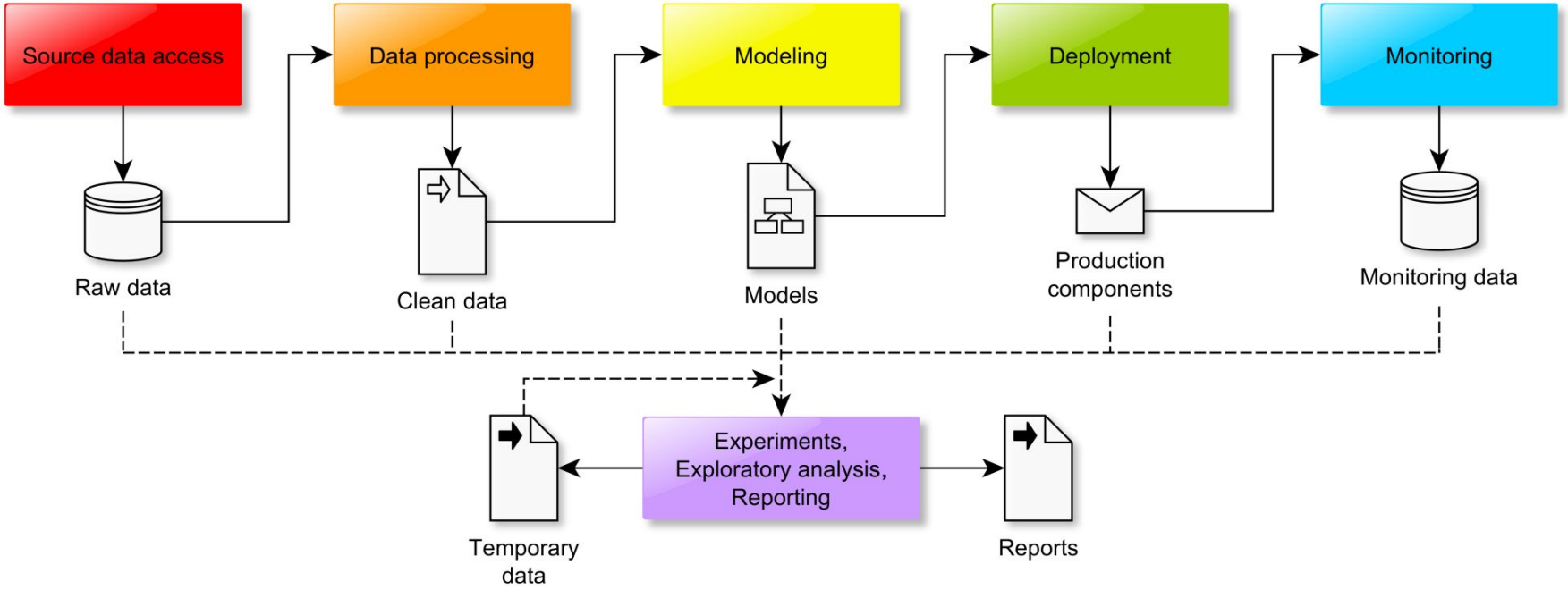
Buzzword bingo

- Big Data
- Data Mining (datizrace)
- Machine Learning – subset of AI
- Data Science – statistics

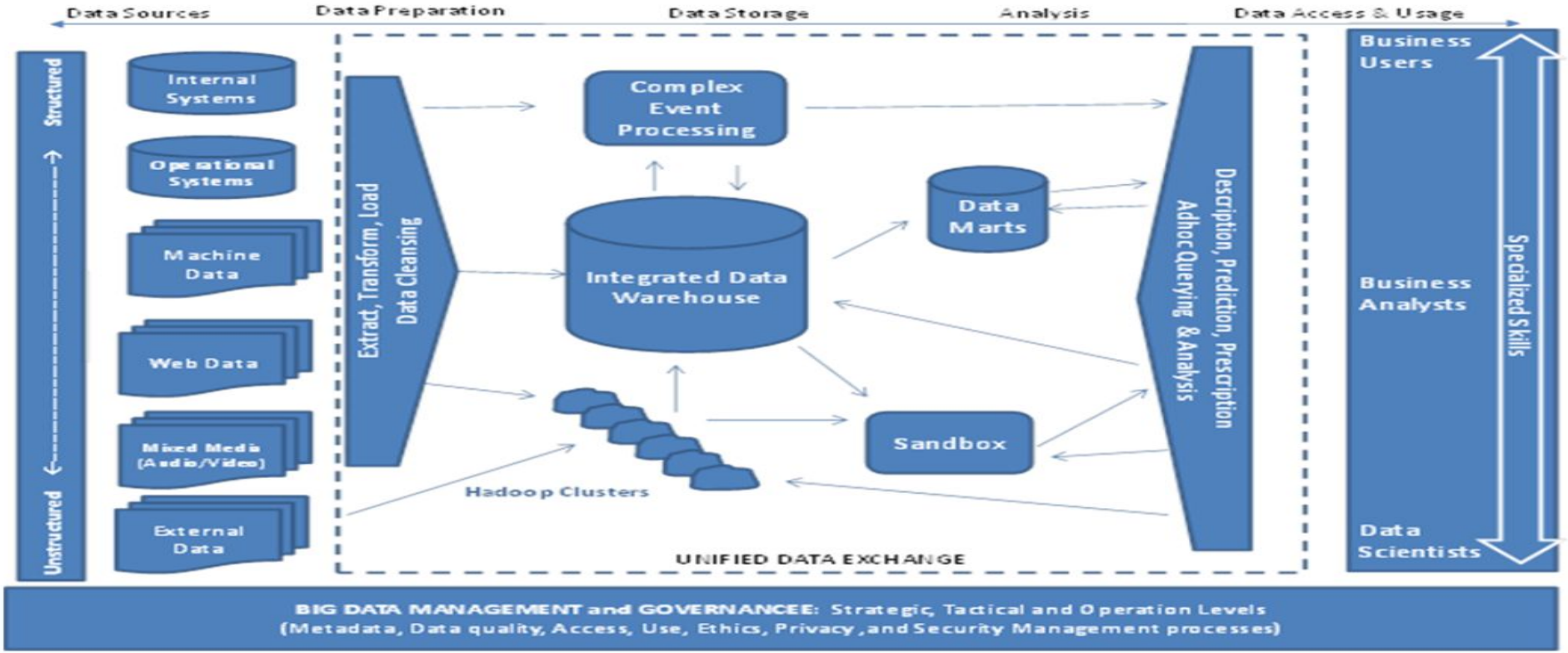
- Big Data or Pokemon
- https://pixelastic.github.io/pokemonorbigdata/

Random Forests	Neural Network	Reinforcement Learning	Supervised Learning	Cognitive Computing
Caffe	Support Vector Machine	Artificial Intelligence	Python	Cloud
Unstructured Data	Bot	DATA SCIENCE BUZZWORD BINGO (free square)	K-means	GPU
Spark	Data Wrangling	Deep Learning	Ensemble	Machine Learning
Keras	Tensorflow	Big Data	Algorithm	Feature Engineering

Full Analysis Framework



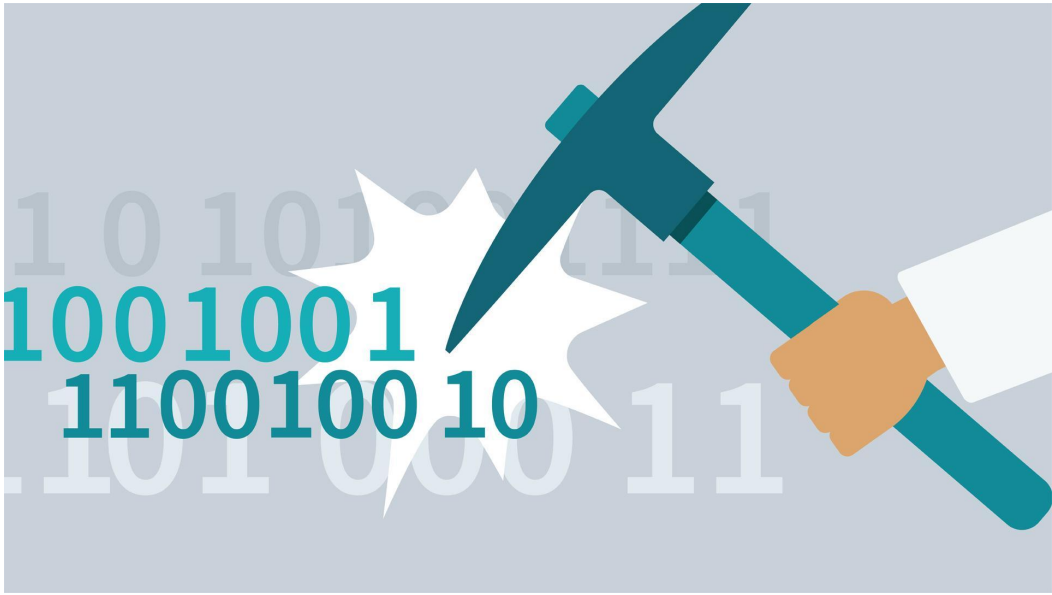
Full Analysis Framework



Data Analysis Sandbox

Data Mining

- Anomalies
- Classification
- Clusters
- Dimension Reduction
- Regression
- Relationship finding
- Summarization / Visualization

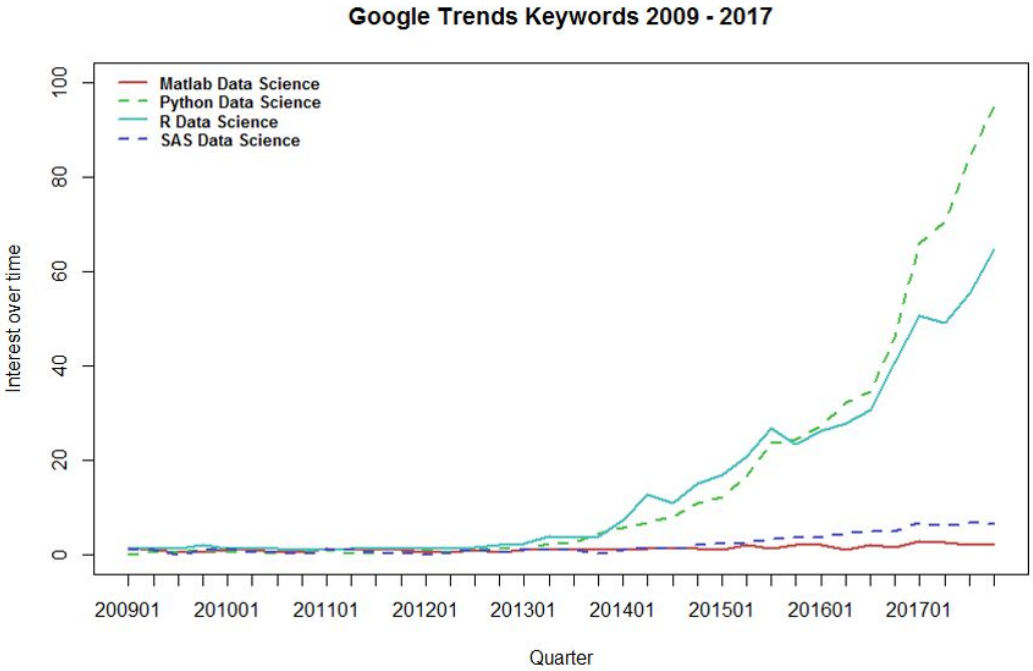


Building a Pipeline

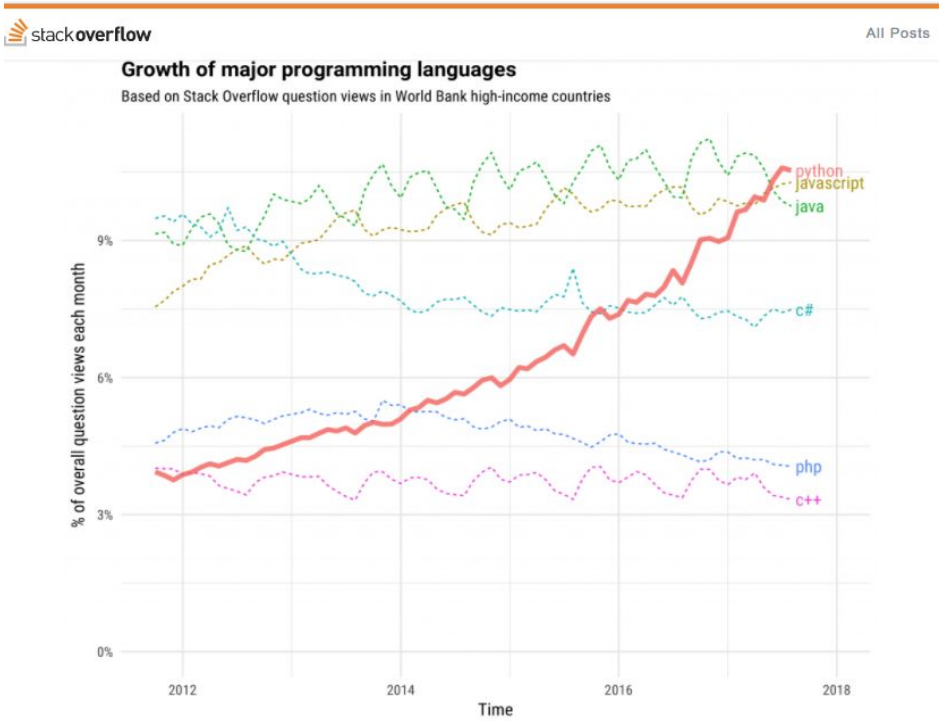
- Data Security
- Cleanup
- Organising Database
- Analysis
- Visualization – Dashboard

- Emphasis on Analysis less on Infrastructure

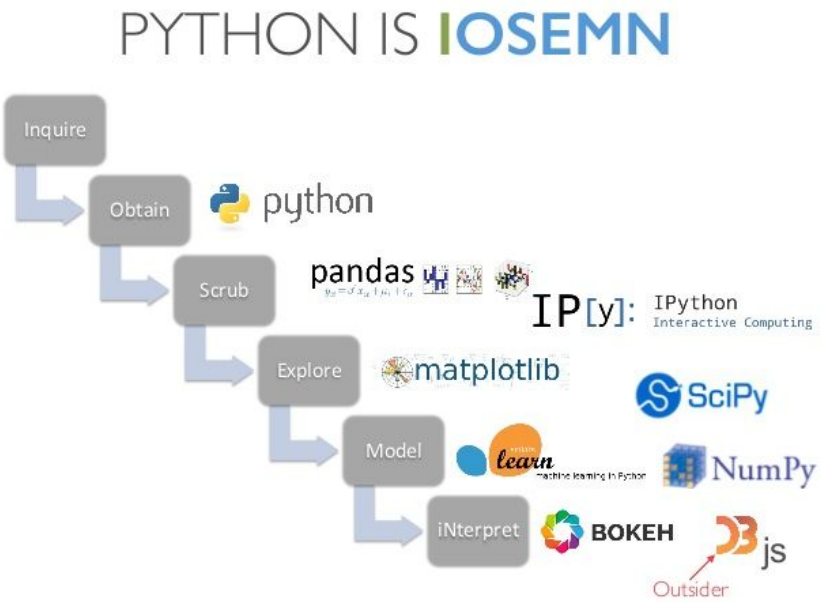
Why Python?


Google Trends Keywords 2009 - 2017
Matlab Data Science
Python Data Science
R Data Science
SAS Data Science
Interest over time
0
20
40
60
80
100
200901
201001
201101
201201
201301
201401
201501
201601
201701
Quarter

Why Python? Part 2



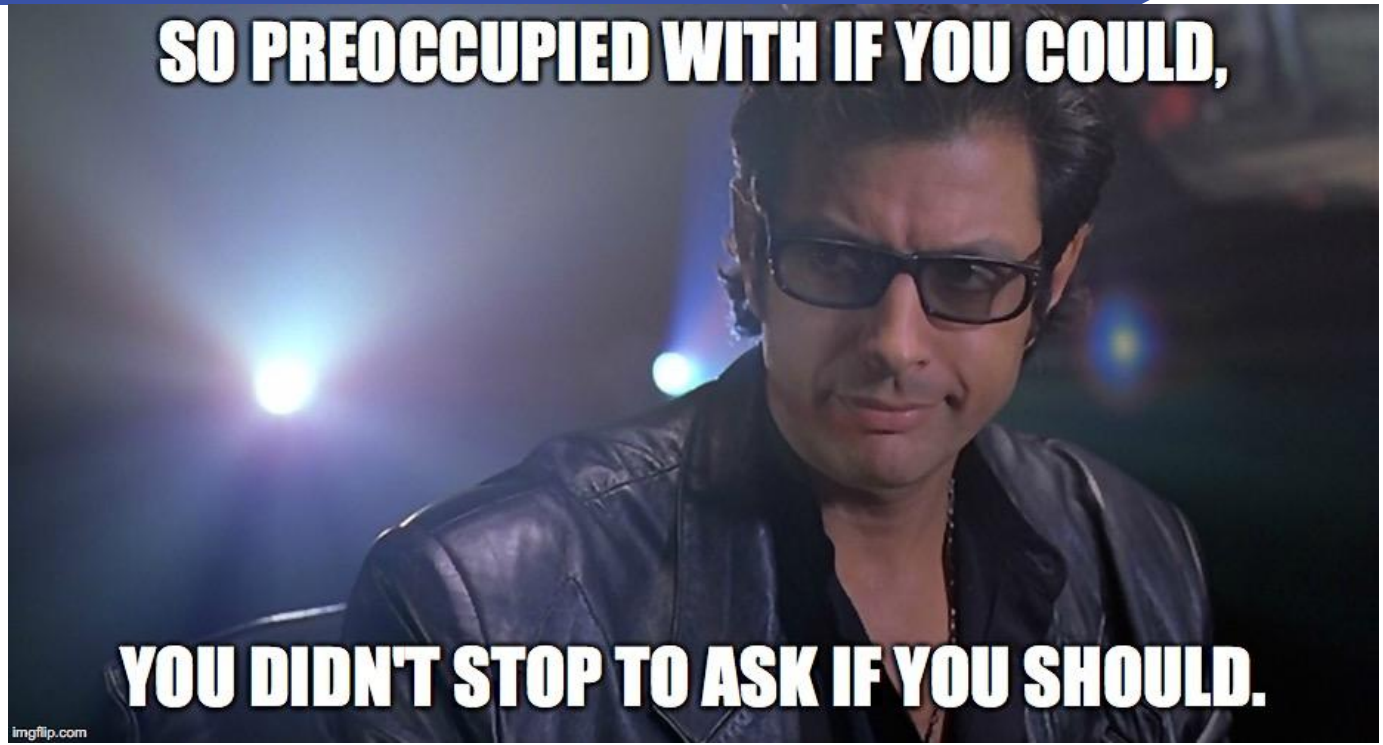
Python Ecosystem



Avoiding Pitfalls

- Overfitting
- Data Dredging / p-hacking
- https://xkcd.com/882/

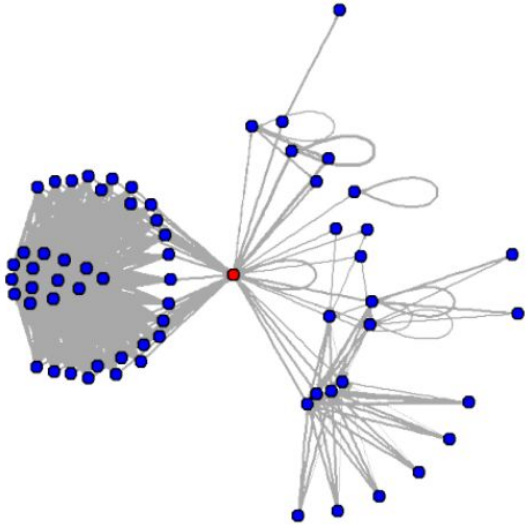
GDPR, ethics


SO PREOCCUPIED WITH IF YOU COULD,
YOU DIDN'T STOP TO ASK IF YOU SHOULD.
imgflip.com

Planned Projects

- Recommendation System / Churn Prediction
- Web Comments Sentiment Analysis
- Network Analysis (possibly some blockchain)

- Visualizations with PowerBI(or Tableau)
- Dashboards with Dash/plotly



Requirements

- Analytical / Logical Mind

- Helpful but NOT required knowledge:
- Python (will use Anaconda / Scipy)
- Comfortable in command line
- SQL
- Statistics

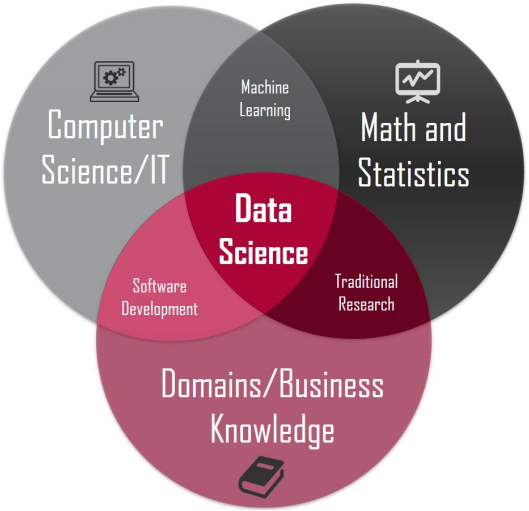
python™

ANACONDA®

- Helpful: a computer with a minimum of 8GB RAM
- https://www.anaconda.com/download/ (3.6+)

Goals

- Access structured/unstructured data
- Clean data
- Apply correct methods for analysis
- Visualize Results



Kādēļ Tu esi šeit?

- **Kādēļ Tu esi šeit?**
  - Vārds
  - Kādēļ pieteicies tieši Rīgas Programmēšanas skolas kursiem
  - Iepriekšējā pieredze/saistība ar tematiku
- **Ko Tu sagaidi no kursa?**
- **Tavi jautājumi?**

PALDIES!