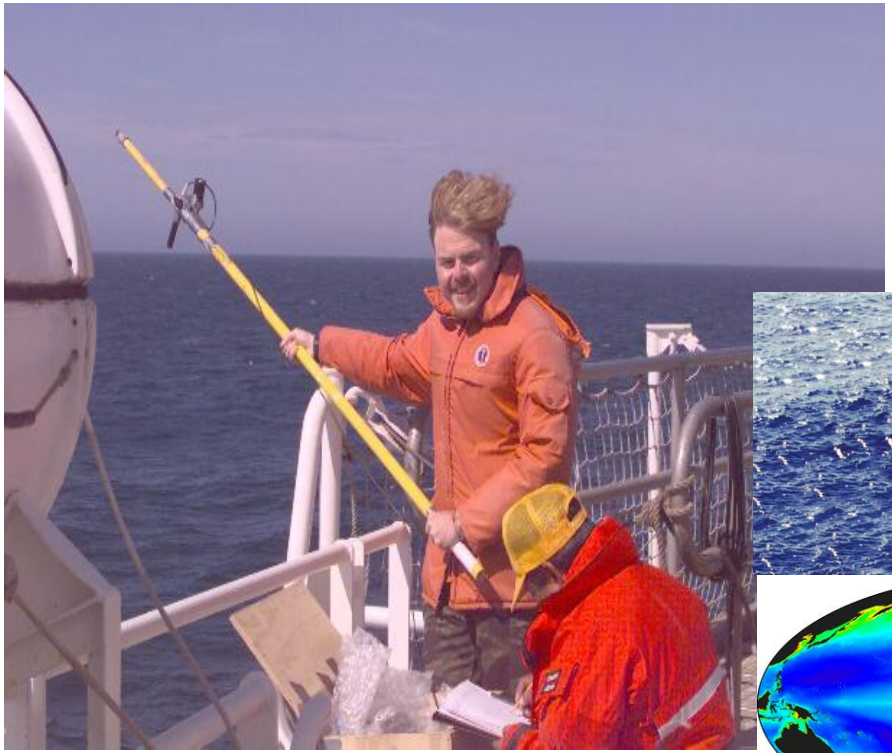
Veribilimine Giriş

Mustafa Akgül
Özgür Yazılım Kış Kampı
2020

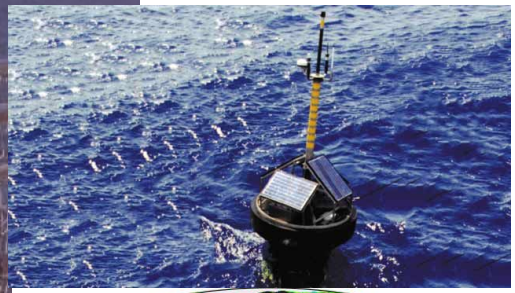
Servet Ahmet Çizmeli
ahmet@pranageo.com

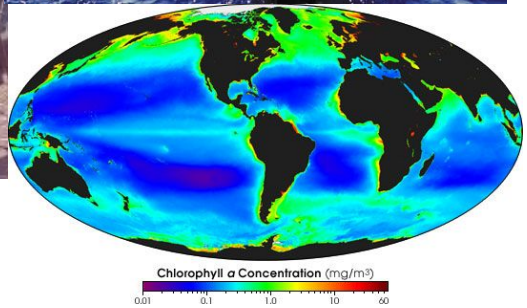
PRANAGE

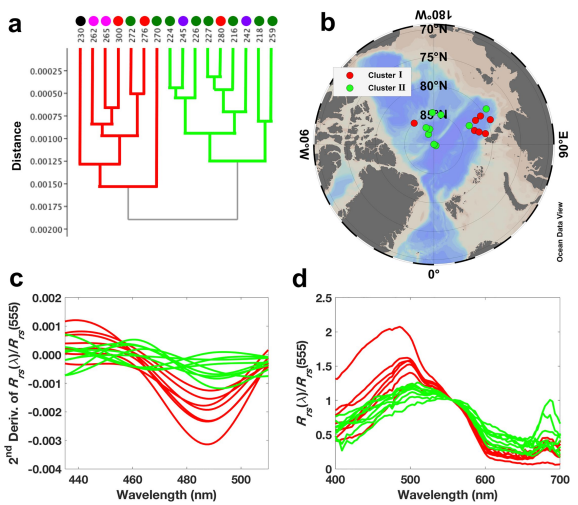
About me



My name is Servet Ahmet Çizmeli. I am researcher in **environmental sciences**. Now I decided to go in the direction of **tech entrepreneurship**.

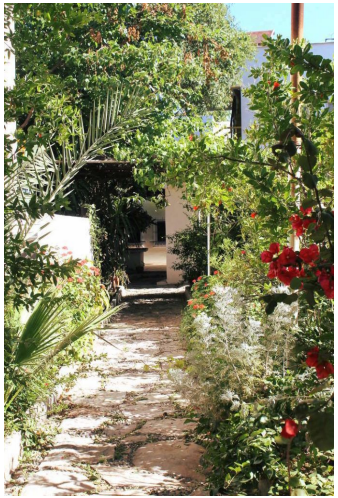






About us

PranaGEO.com is a young **Research** and **Development** company in Bodrum, Turkey. We provide data analysis & software solutions for **Data Science** projects.



We would like to create software to make the dream of **Reproducible Research** come true.

Reproducible Research

Problem

"an article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result."

(Donoho, 2010)

we need a RR-aware platform. If possible, **ONE** platform that binds all the required services/stages

PRANAGEO

Reproducible Research

Why?

Computational research today needs to be 100% **open and reproducible**. Why?

- Research advances when others construct **on top of** what we build, and vice versa.
- Scientific results are subjective and multi-dimensional. Full access to **datasets**/**figures**/**code** is needed to prove/**refute**/**compare**/**improve** existing research
- On the contrary to common, old fashioned belief, **we get richer when we share!**

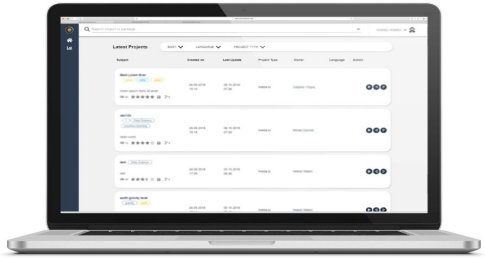
Our product : melda.io

Modern, Elegant Data Analysis

melda.io is a new, innovative, web-based cloud-native data science platform. You can create data analysis projects with R, python, publish your work, interact and co-create with others.

Start using melda.io for free

it was designed to promote reproducibility, online **collaboration**, to make it **easy** to conduct environmental/data scientific studies, hence speed up **innovation**



save time

PRANAGEO

Data science

Innovation + business+ data + coding

Innovation

Definition of Innovation

innovation

/ɪnəˈveɪʃ(ə)n/

noun

the action or process of innovating.

"innovation is crucial to the continuing success of any organization"

Similar: change | alteration | revolution | upheaval | transformation

- a new method, idea, product, etc.

Example

Definition of Healthcare Innovation

Medical Innovation leads to the discovery and development of products and therapies that **save**, **improve** and **extend** lives — the primary goal of any healthcare program.

And by doing that, innovation helps keep healthcare **costs down** and drives **economic growth**.

PRANAGEO

Data science

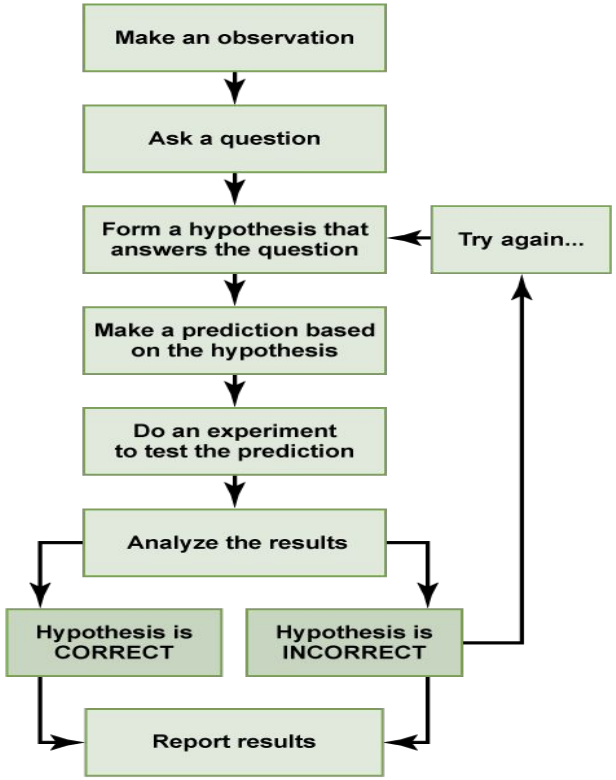
It's in the business world!

SCIENCE | BUSINESS

Scientific Method

Defined

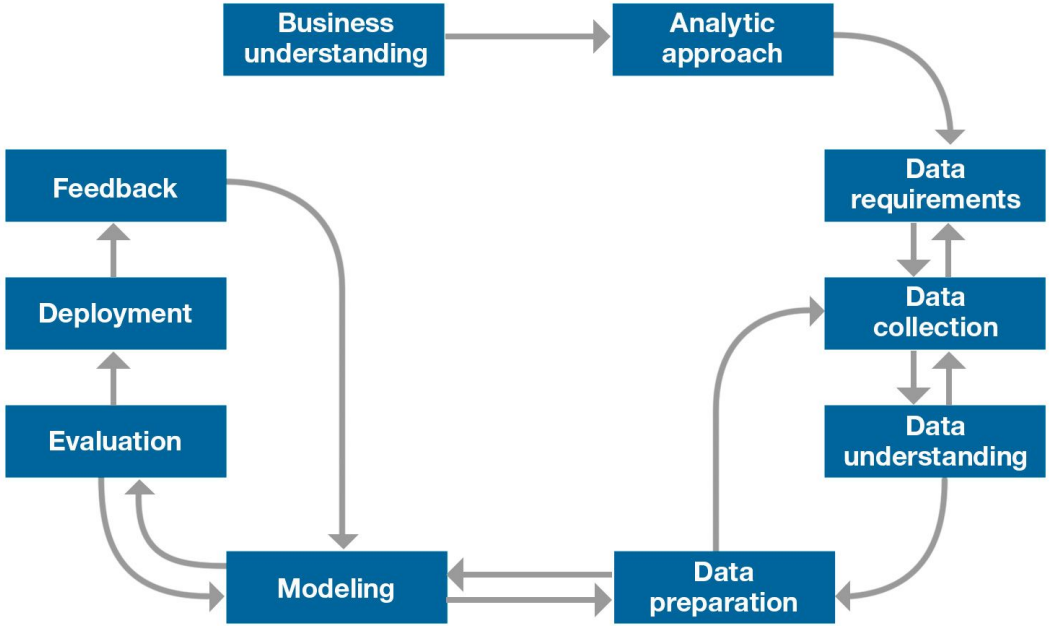
https://biologydictionary.net/scientific-method



https://courses.lumenlearning.com/boundless-psychology/chapter/the-scientific-method/

Scientific Method

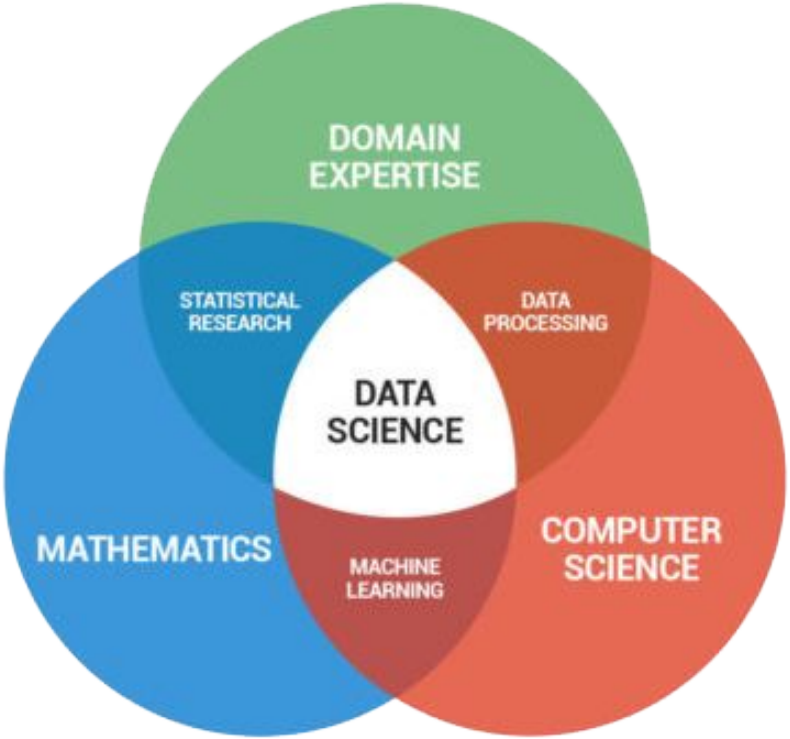
Implemented in data science



https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science

Data science

components



Data science

Who is a data scientist?

Male
70% of Data Scientists in our research were male

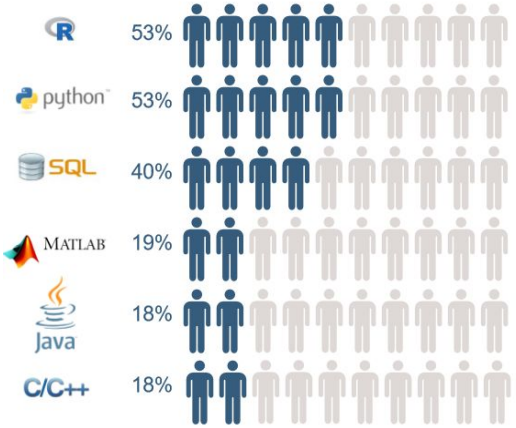
2 Languages
Data scientists speak at least 1 foreign language on average

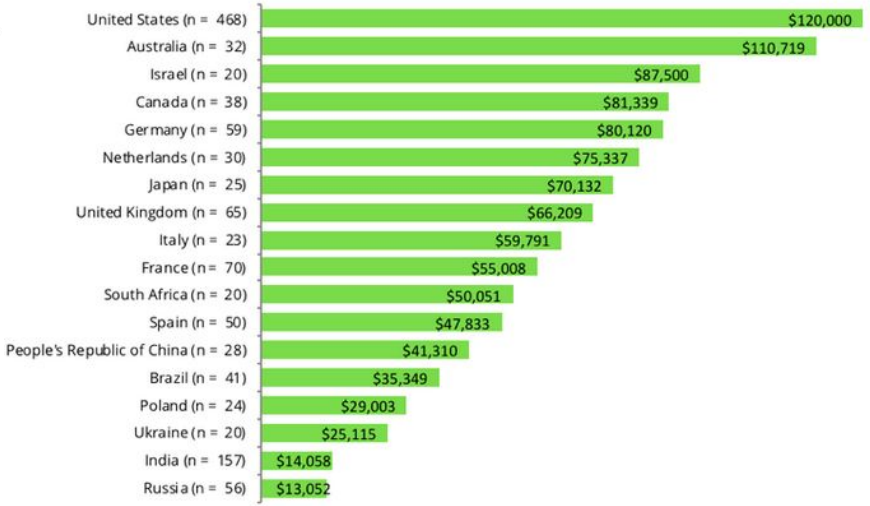
2 years
This is a new profession. The median experience as data scientists of professionals in our research was 2 years

4.5 years
People who work as data scientists currently have a median work experience of 4.5 years (including previous positions)

R and/or Python
More than 50% of the data scientists in our research work in R and/or Python

Master or PhD
75% of data scientists have a PhD (27%) or a Master (48%) degree





Data Science

Preliminary questions to answer

Do you already have data?

If not, do you know where to get it?

What kind of data analysis software/platform do you currently use?

What kind of analysis do you plan on using?

Does that involve use of hardware?

PRANAGEO

Data Science

Things to know

PRANAGEO

Data Science

Open source tools we prefer

Exercise 7.1

Find which data science models could be applied to the business problems given in column *Example Tasks* :

Example tasks	Machine learning terminology	Typical algorithms
Identifying spam email Sorting products in a product catalog Identifying loans that are about to default Assigning customers to customer clusters	Classification: assigning known labels to objects	Decision trees Naive Bayes Logistic regression (with a threshold) Support vector machines
Predicting the value of AdWords Estimating the probability that a loan will default Predicting how much a marketing campaign will increase traffic or sales	Regression: predicting or forecasting numerical values	Linear regression Logistic regression
Finding products that are purchased together Identifying web pages that are often visited in the same session Identifying successful (much-clicked) combinations of web pages and AdWords	Association rules: finding objects that tend to appear in the data together	Apriori
Identifying groups of customers with the same buying patterns Identifying groups of products that are popular in the same regions or with the same customer clusters Identifying news items that are all discussing similar events	Clustering: finding groups of objects that are more similar to each other than to objects in other groups	K-means
Making product recommendations for a customer based on the purchases of other similar customers Predicting the final price of an auction item based on the final prices of similar products that have been auctioned in the past	Nearest neighbor: predicting a property of a datum based on the datum or data that are most similar to it	Nearest neighbor

https://melda.io

Contact us:

melda.io

ahmet@pranageo.com

PRANAGE