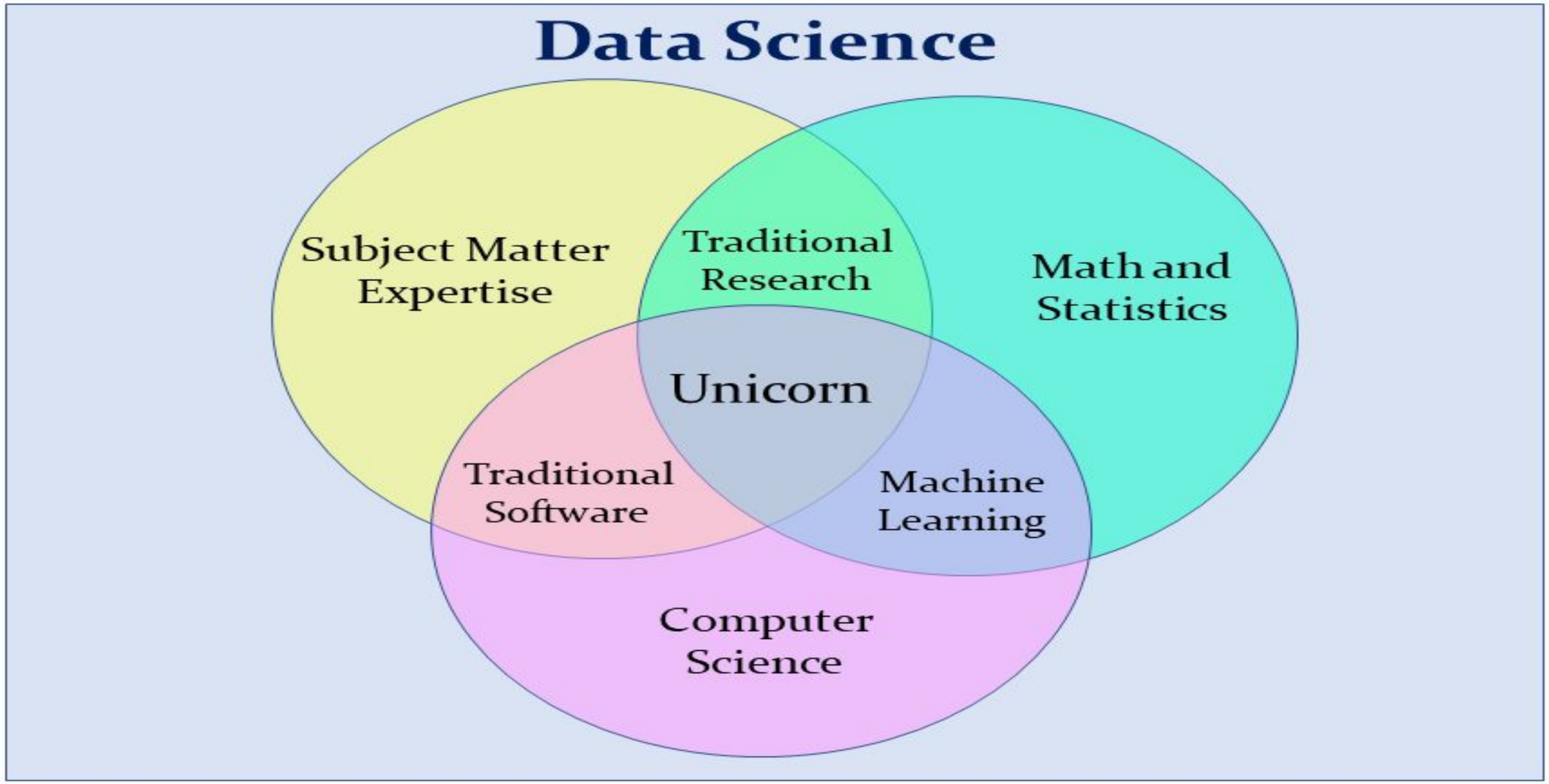
Data Visualization

Data Science

- **Data Science** combines several disciplines including:
  a. Math
  b. statistics
  c. computer science

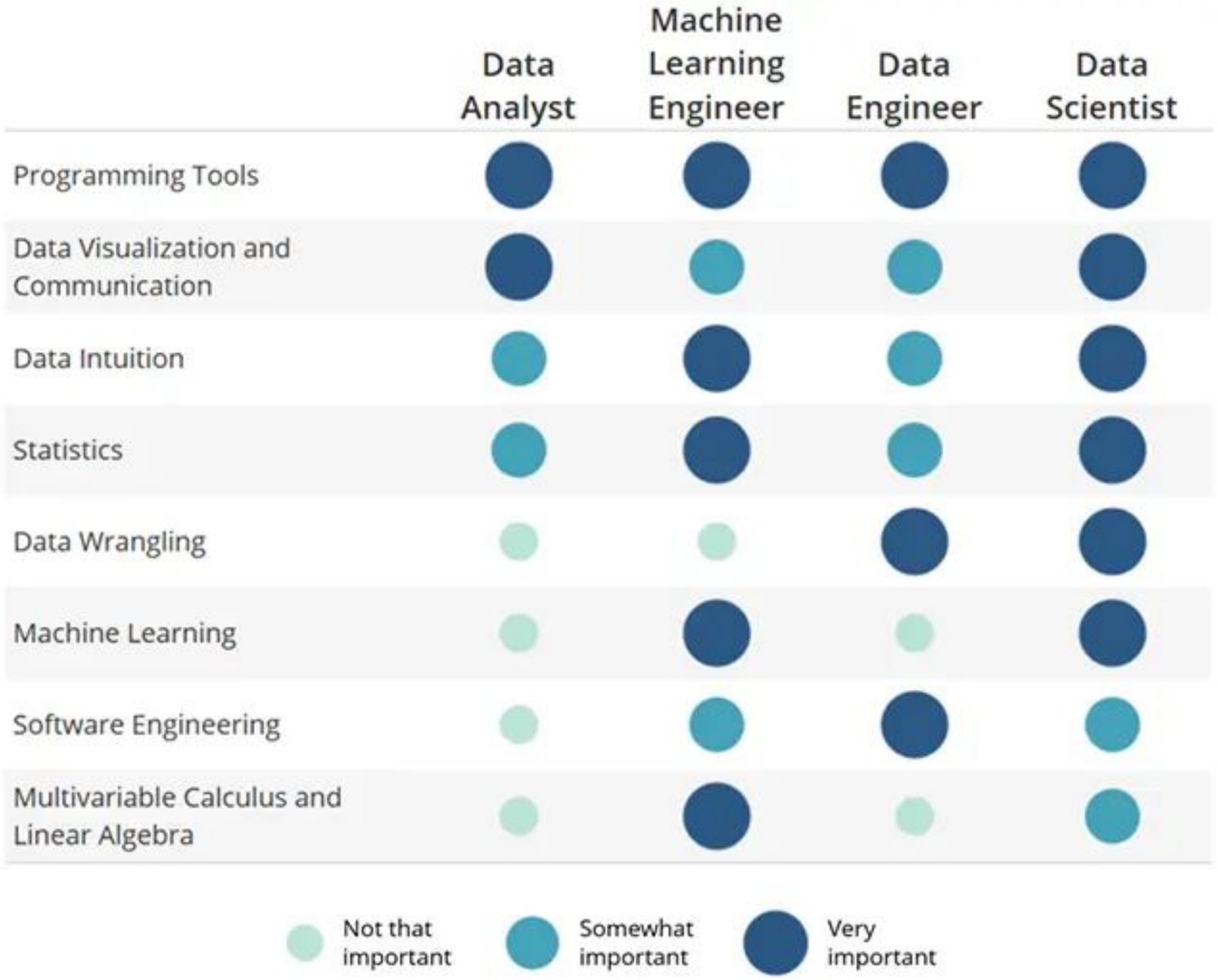
- How to become a **Unicorn** in **Data Science**?

Data Science Venn Diagram 2.0



Data Science Job Types

Long-lasting Career



Data Science Job Types 1) Data Analyst

- Job might consist of tasks like:
  - pulling data out of **SQL** databases
  - becoming an **Excel** or **Tableau** master
  - producing basic **Data Visualizations** and **Reporting Dashboards**
  - On occasion:
    - analyze the results of an **A/B Test**
    - take the lead on company's **Google Analytics account**
- Some companies: **Data Scientist** is synonymous to **Data Analyst**

Data Science Job Types 2) ML Engineer

- Some companies: **data** or **data analysis platform** is the product
  - so, **Data Analysis** or **ML** can be pretty intense
  - so, there is a need for someone who:
    - has a formal mathematics or statistics background
    - is hoping to continue down a more academic path
- **ML Engineers**
  - often focus more on producing great **data-driven products**
  - less focus on answering operational questions for a company

Data Science Job Types 3) Data Engineer

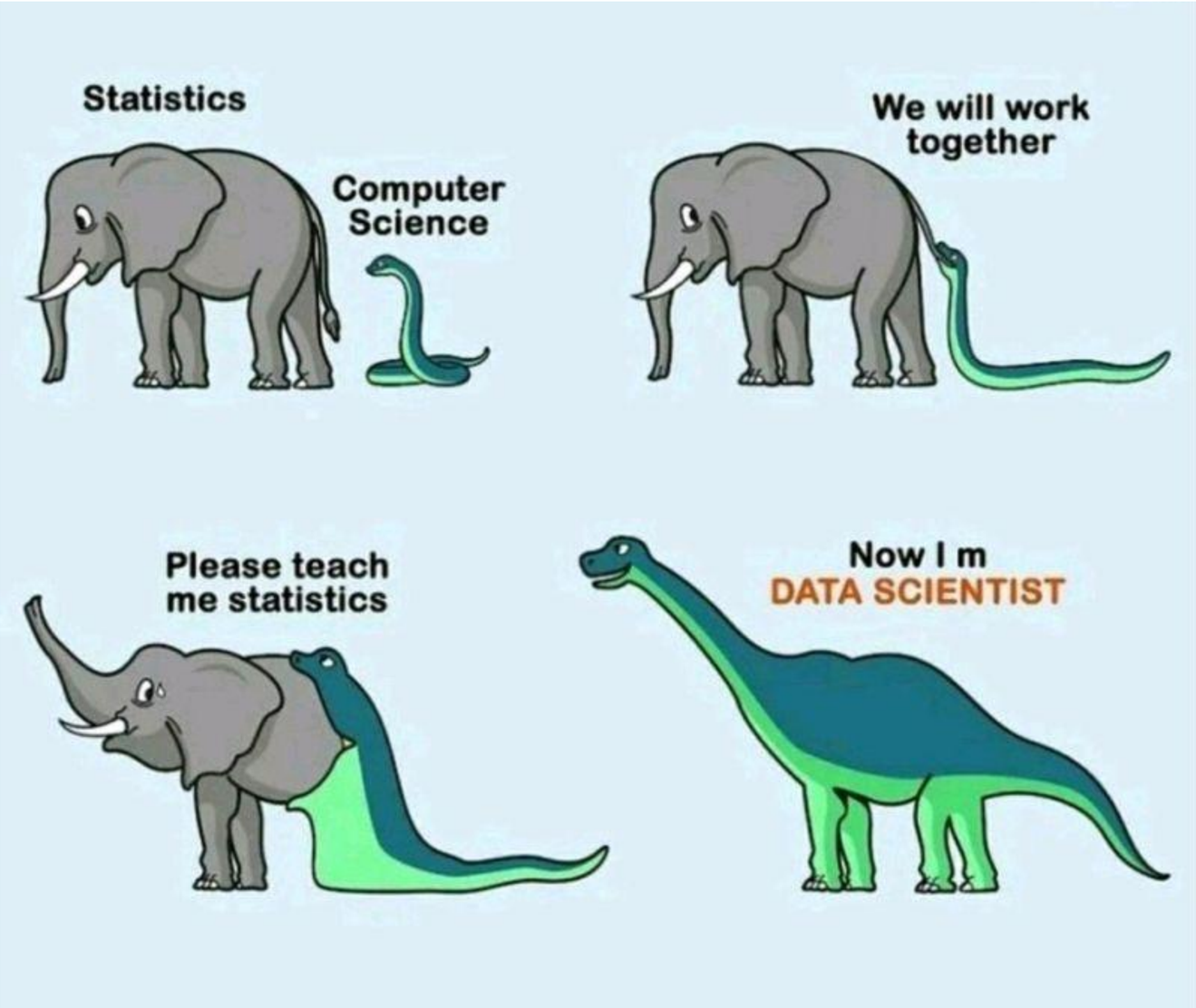
- Some companies: have **lot of traffic** and **large amount of data**
  - so, there is a need for someone who:
    - set up lot of **data infrastructure** that the company will need moving forward
    - can provide analysis
- Job postings listed under **Data Scientist** and **Data Engineer**
  - strong software engineering skills are more important
  - heavy statistics and ML expertise are less important

Data Science Job Types 4) Data Generalist

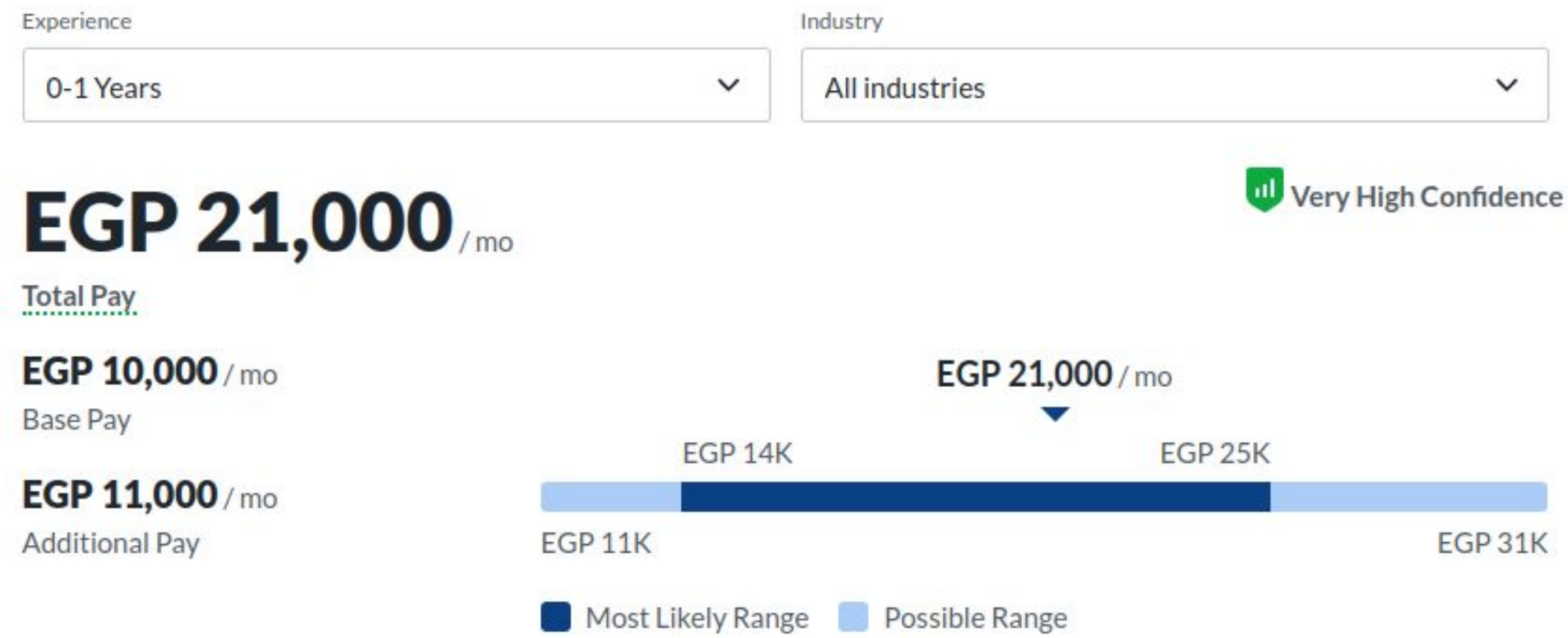
- Some companies: look 4 **Generalist** to join data scientists team
  - these companies cares about data but probably isn't a data company
- **Data Generalist** most important skills:
  - familiarity with tools designed for **Big Data**
  - experience with messy, real-life datasets
  - …
- **Data Generalist** can:
  - perform analysis
  - touch production code
  - **Visualize Data**

Data Science Job Types 5) Data Scientist

Often used as a blanket title to describe jobs that are drastically different

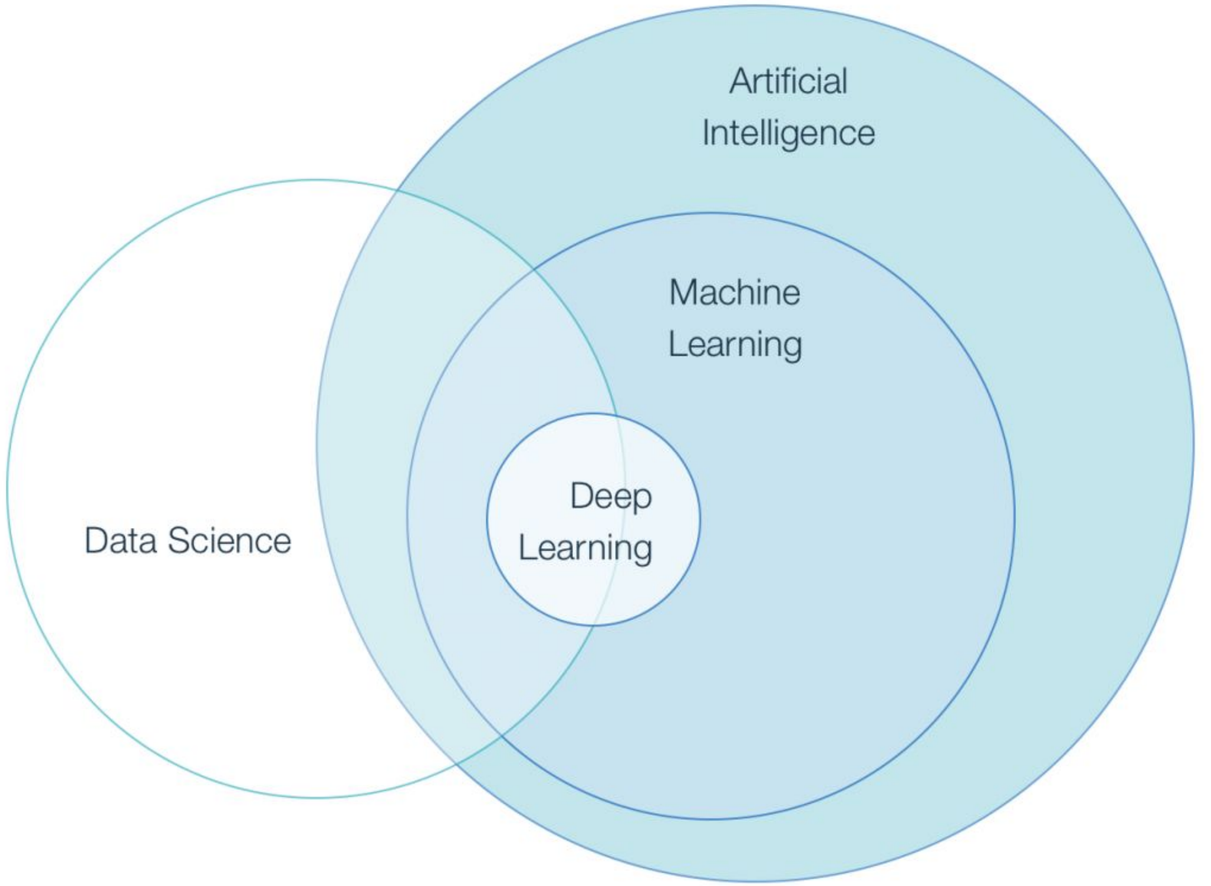


Data Scientist Average Salaries: Egypt 2022



https://www.glassdoor.com/Salaries/cairo-data-scientist-salary-SRCH_IL.0,5_IM1175_KO6,20.htm

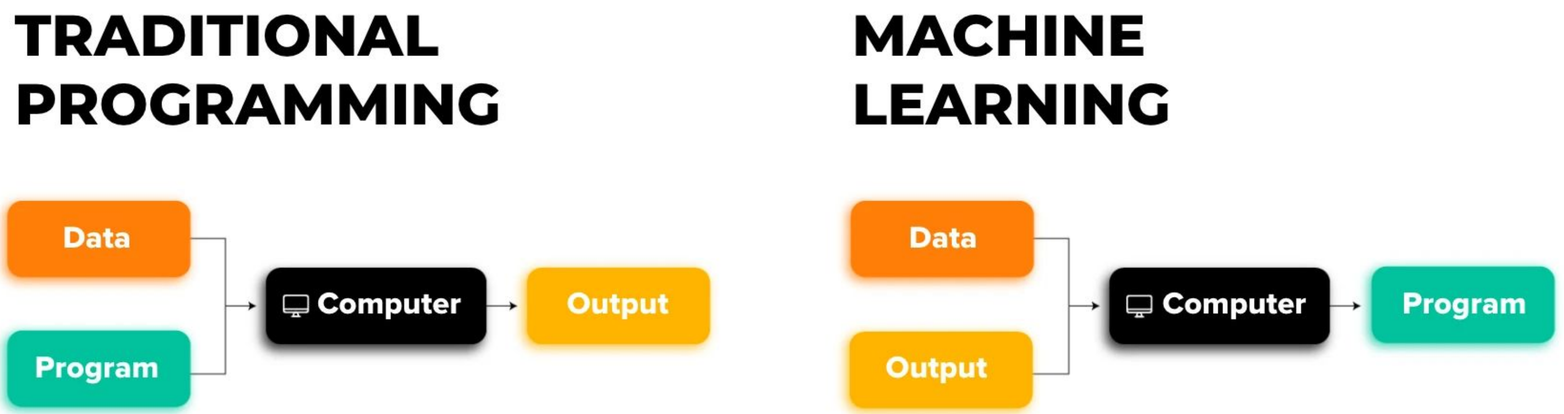
Data Science & AI

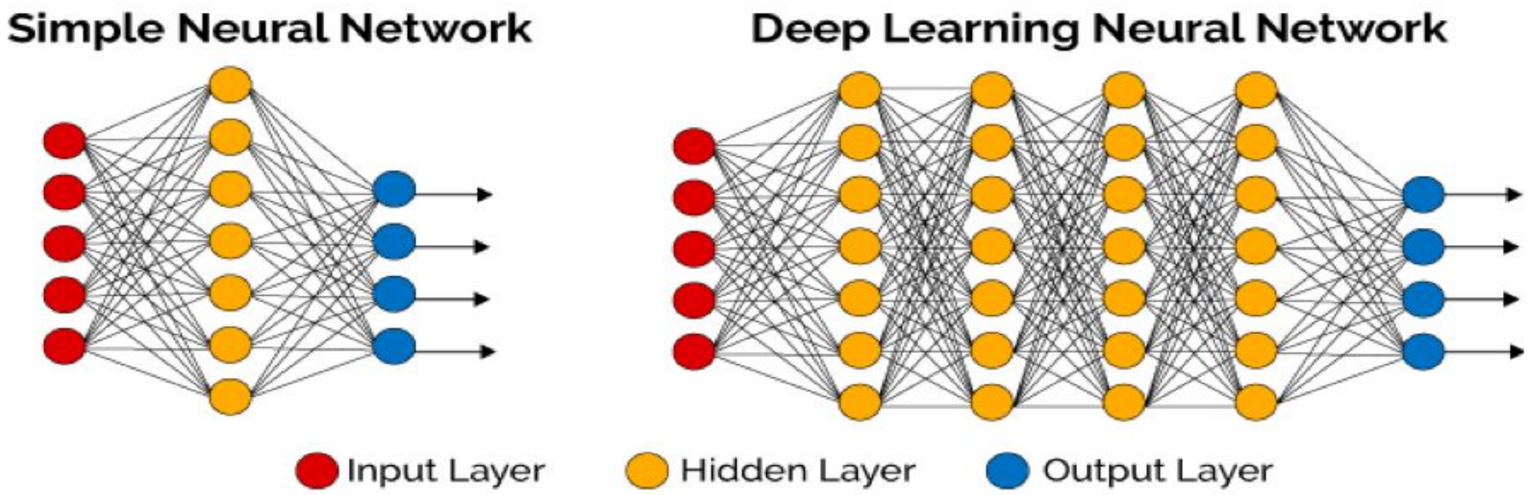


AI

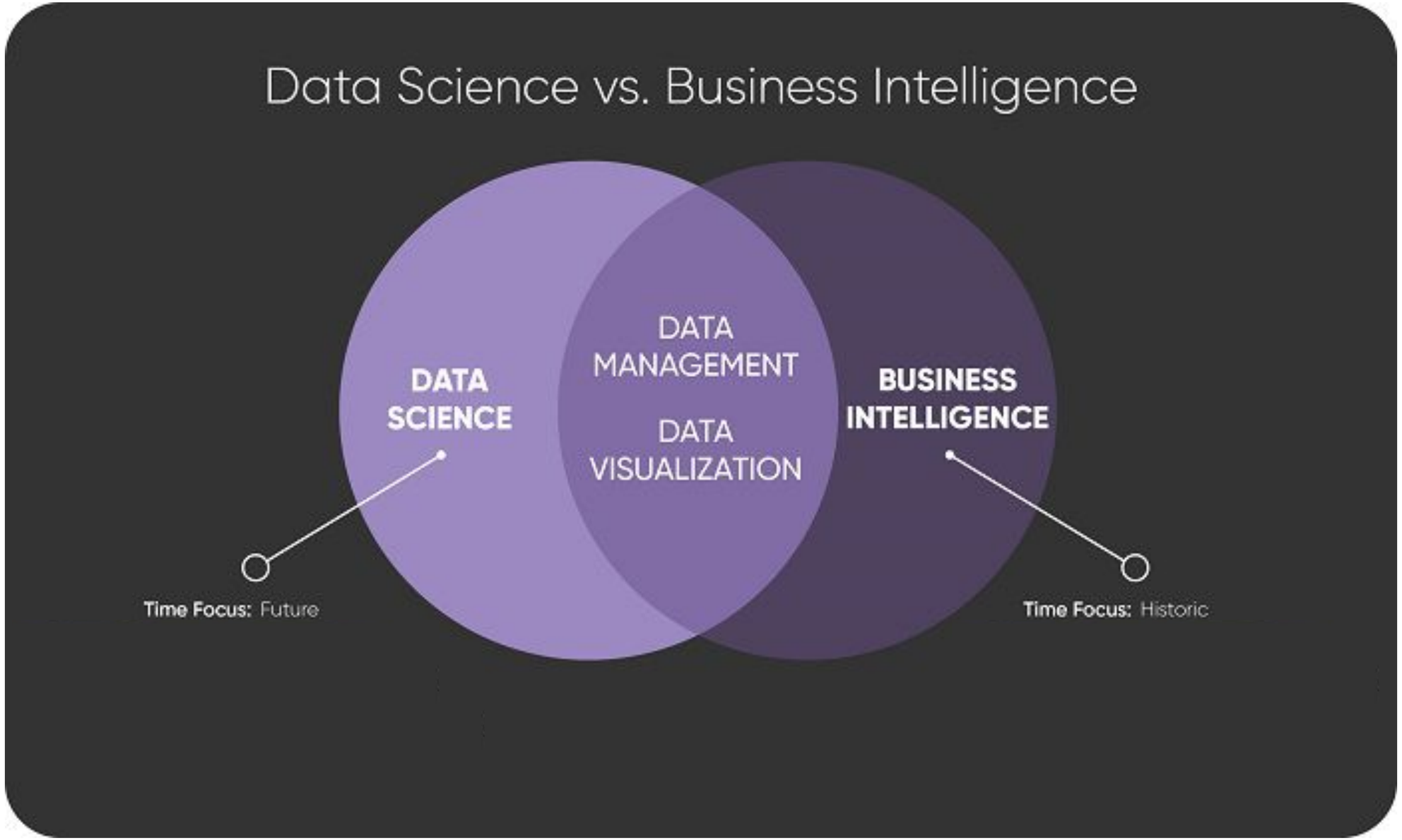
- According to the Turing test,
- is the machine's ability to exhibit intelligence behavior indistinguishable from that of a human

Data Science & AI/ML/DL





Data Science & BI



Data Analysis Steps

1. Question
2. Wrangle
3. Explore
4. Draw Conclusions
5. Communicate

Data Analysis Steps

Step 1: Ask Questions

- Given data then ask questions, or
- Ask questions then **gather** data

Step 2: Wrangle Data

a. **Gather** data to answer question

b. **Assess** data to identify any problems in your data's quality or structure

c. **Clean** data by modifying, replacing, or removing data

Data Analysis Steps

Step 3: Perform Exploratory Data Analysis (EDA)

- **Explore then augment** data to maximize the potential of analyses, **visualizations**, and models.
- **Exploring** involves:
  - finding **patterns** in data
  - **visualizing** relationships in data
  - building **intuition** about what you're working with
- **Remove Outliers:** (**optional**)
- **Feature Engineering:** create better features from data (**optional**)

Data Analysis Steps

Step 4: Draw Conclusions (or even make predictions)

- typically approached with **ML** or **inferential statistics**

Step 5: Communicate Results

- often need to **justify** and **convey** meaning in the insights
- if your end goal is to build a system, you usually need to:
  - **share** what you've built
  - **explain** how you reached design decisions
  - **report** how well it performs
- communicate results by: report | slides | presentation | post | email | conversation
- **Data Visualization** will always be very valuable

Questions

Links

https://github.com/fcai-b/dv

References

1. https://www.udacity.com/blog/2018/01/4-types-data-science-jobs.html
2. https://www.udacity.com/course/data-analyst-nanodegree--nd002