Data Science Career Track

Introduction to Data Science

Developer Student Clubs
Al-Azhar University

Some Basic Rules
Before we Start!

AGENDA

01 **What is Data Science**

What's Data Science is All about?

02 **Data Science Ecosystem**

Data science World and commercial landscape

03 **Data Science Building Blocks and Team Roles and skills**

04 **Data Science Methodology**

Steps of a Data science Project

HELLO!

Meet Mohamed

I am here because I love to Start a Career in Data Science!

Join me throughout my journey ☺!

1

What is Data Science

What's Data Science is All about?

Data

They care about volume and velocity and whatever other buzzwords describe data that is too big for you to analyze in Excel.

Science

Data science is only useful when the data are used to answer a question.

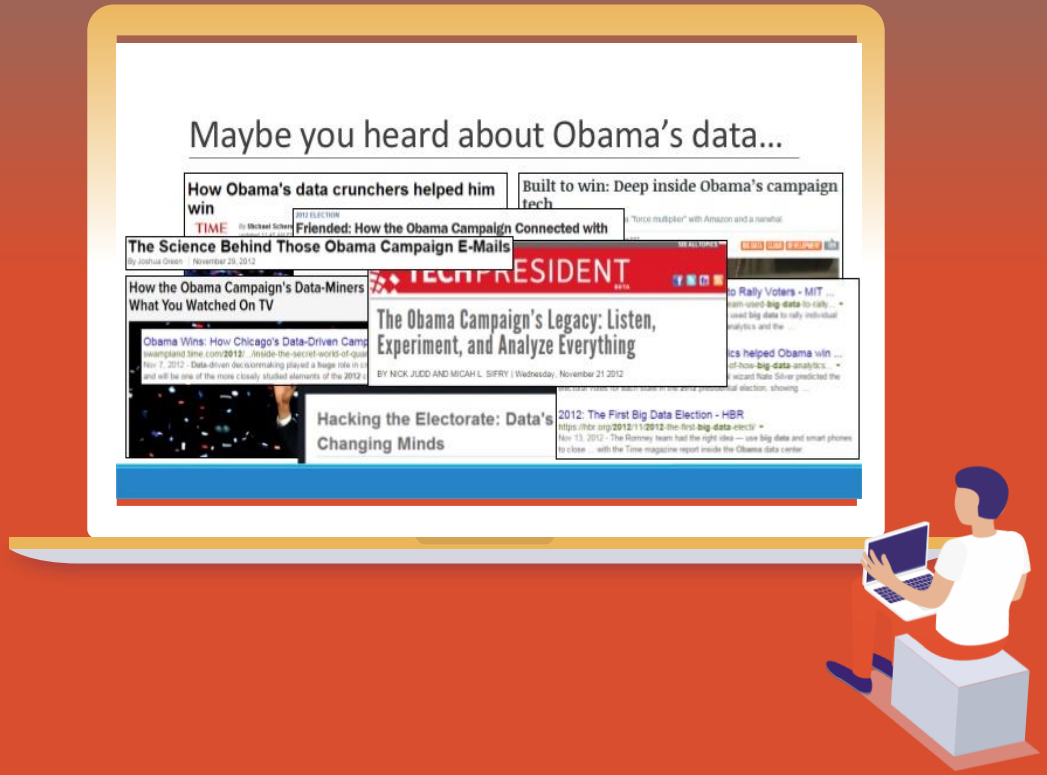
"I have this really hard question, can I answer it with my data?"

Moneyball

can we build a winning baseball team if we have a really limited budget?.

Voter Turnout

“How do we find the people who vote for Barack Obama and make sure that those people end up at the polls on polling day?”



“The goal is to turn data into information, and information into insight.”

–Carly Fiorina

“Numbers have an important story to tell. They rely on you to give them a voice.”

–Stephen Few

The outputs of a data science Projects

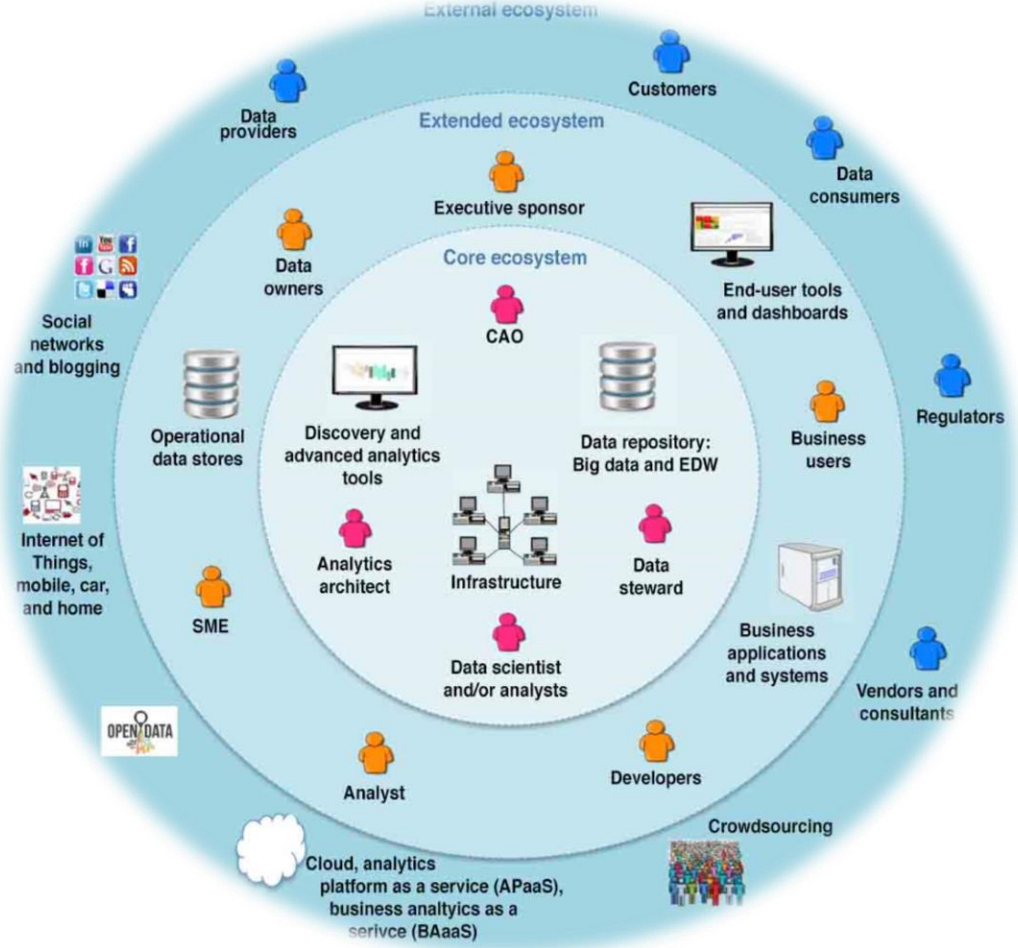
2

Data Science Ecosystem

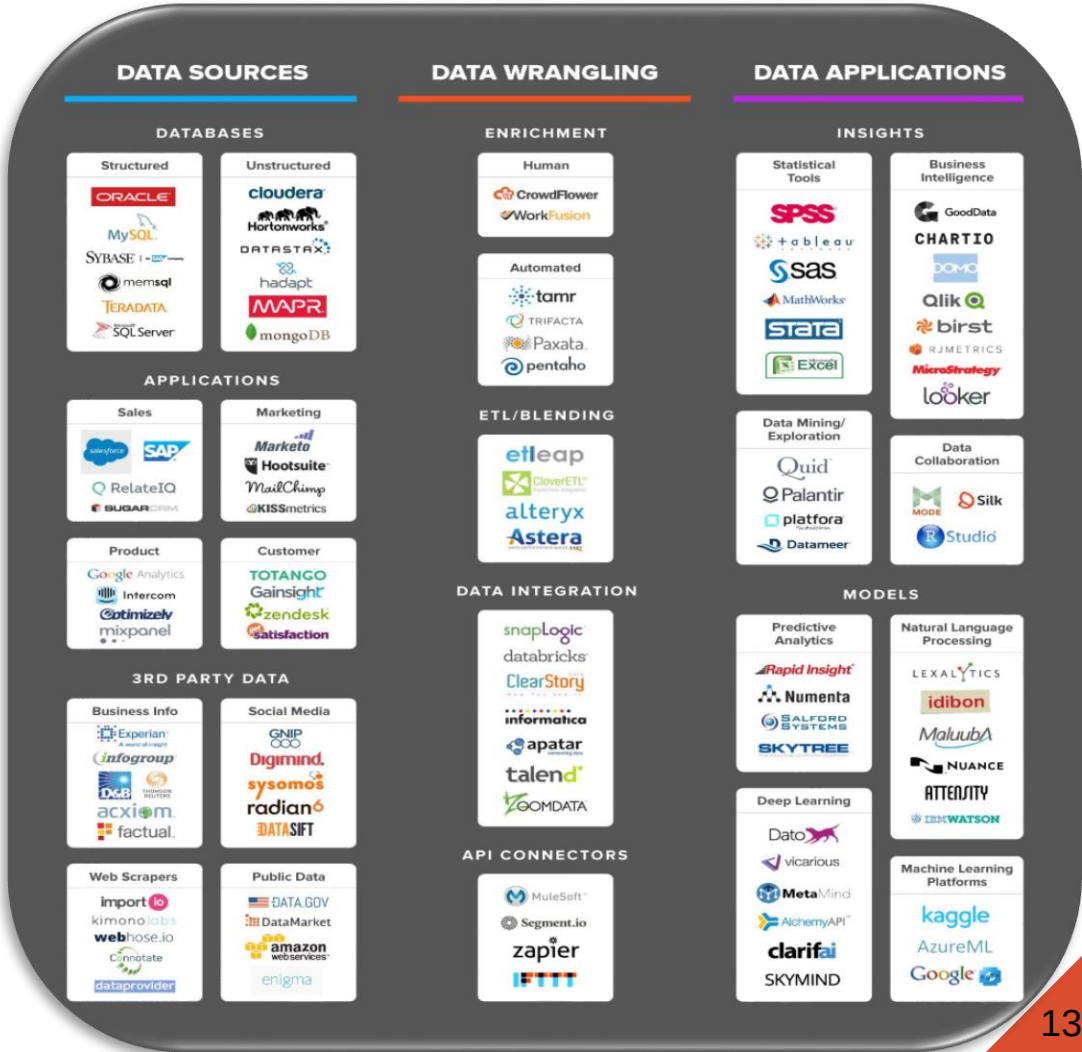
Data Ecosystem

An ecosystem: is the network of organizations—including suppliers, distributors, customers, competitors, government agencies, and so on—involved in the delivery of a specific product or service through both competition and cooperation.

Source: Investopedia.com



Commercial Landscape



3

Data Science Building Blocks

Data Science Building Blocks

Math And Statistics

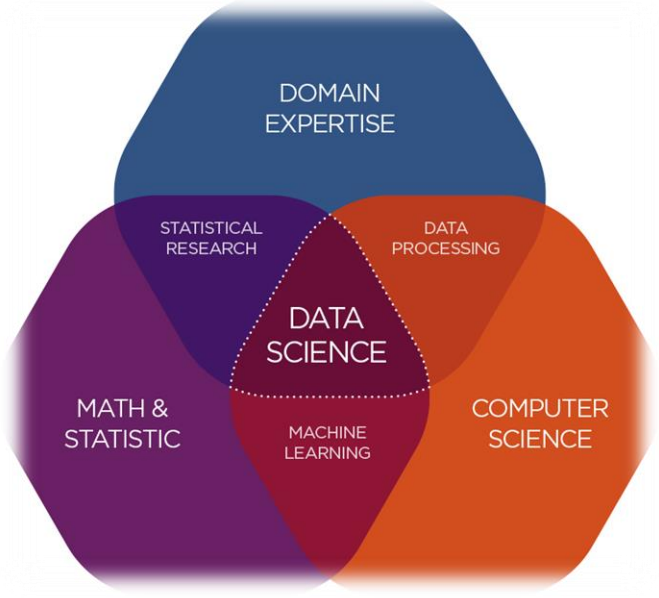
the foundation of knowledge and techniques that are used in Data analysis and inference.

Computer Science

knowledge of programing languages, data structures, data bases and algorithms .

Human Expertise

the knowledge and experience of the business domain problems where data science will try to solve.



Programming Language

- Python
- R

Data Visualization

- Tableau
- Power BI
- Matplotlib
- Seaborn
- ggplot
- D3.js
- Gephi

Data Analysis

- Feature Engineering
- Data wrangling
- EDA

DATA SCIENCE

IDE

- Jupyter
- Colaboratory
- Spyder
- PyCharm

Machine Learning

- Classification
- Regression
- Reinforcement learning
- Deep learning
- Clustering

Math

- Statistics
- Linear Algebra
- Differential calculus

Deploy

- AWS
- AZURE

Web Scraping

- Scrapy
- URLLIB
- Beautiful Soup

3

Data Science Team

Data science is a team sport

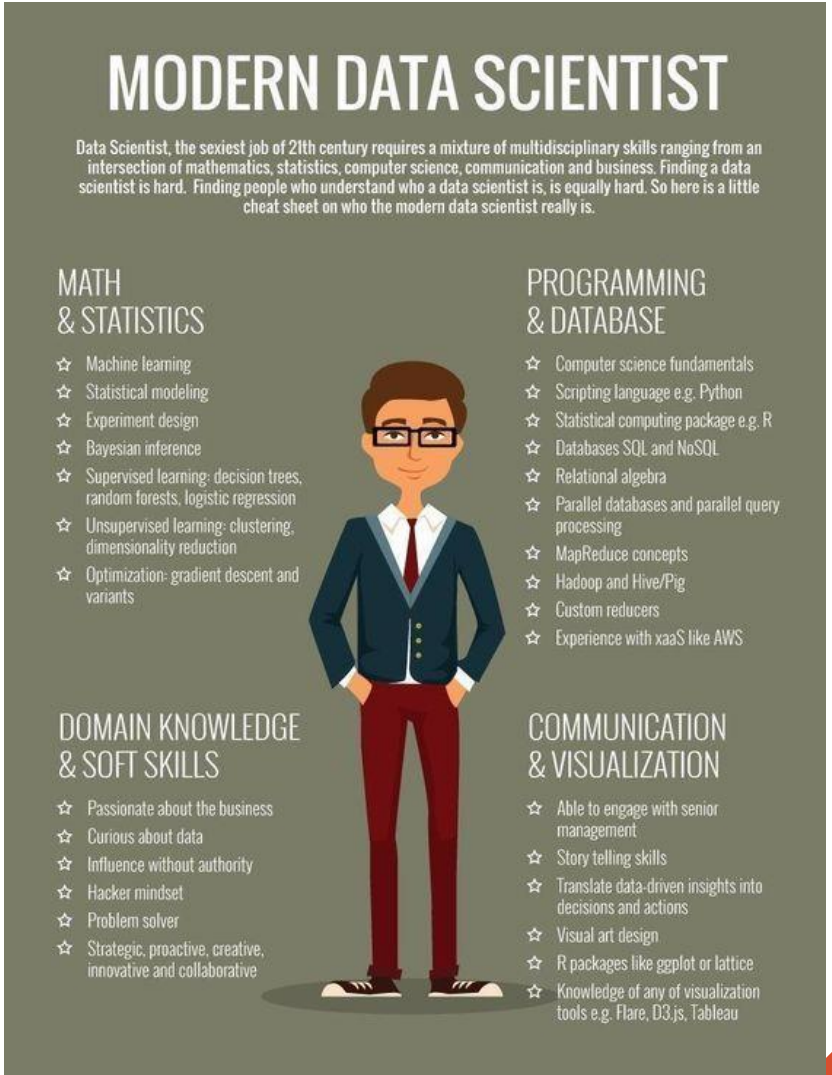
Data Sciene "Unicorn"

The **data science unicorn** is a somewhat mythical person who is a leader in **data science**, technology, and business. ...

Data Scientist

What does a data scientist do?

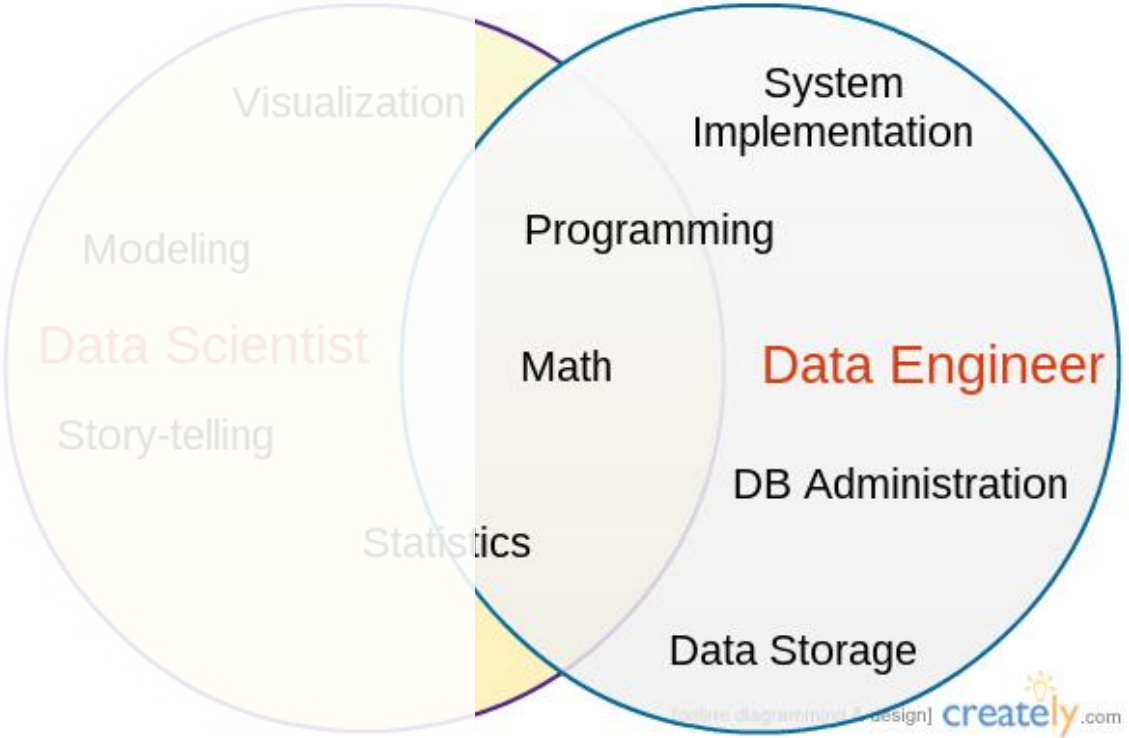
Design experiments
Pull and clean data
Analyze data
Communicate results



Data Engineer

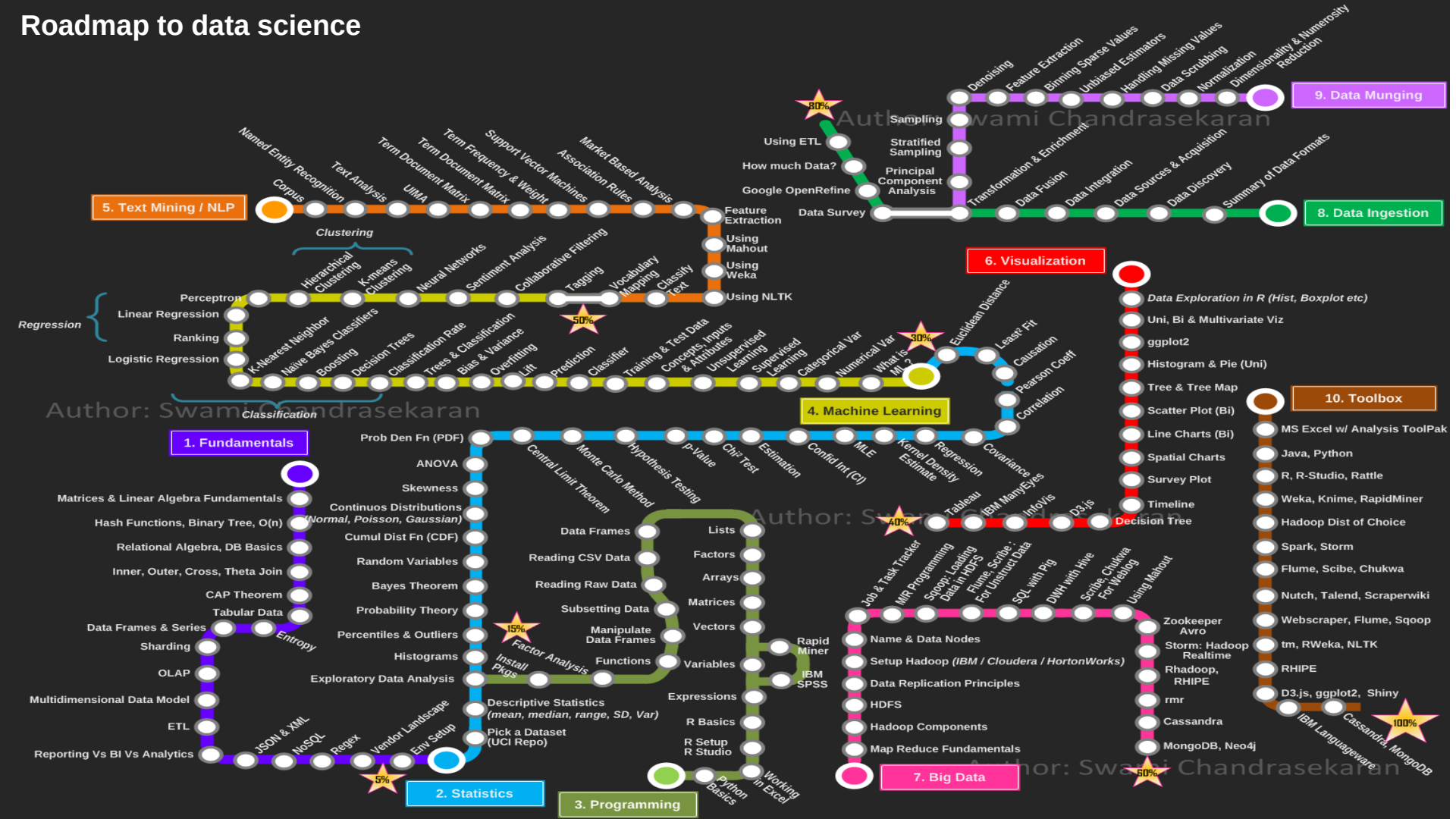
What does a data engineer do?

- Build data infrastructure
- Manage data storage and use
- Implement production tools



Data Science is a Marathon

Roadmap to data science

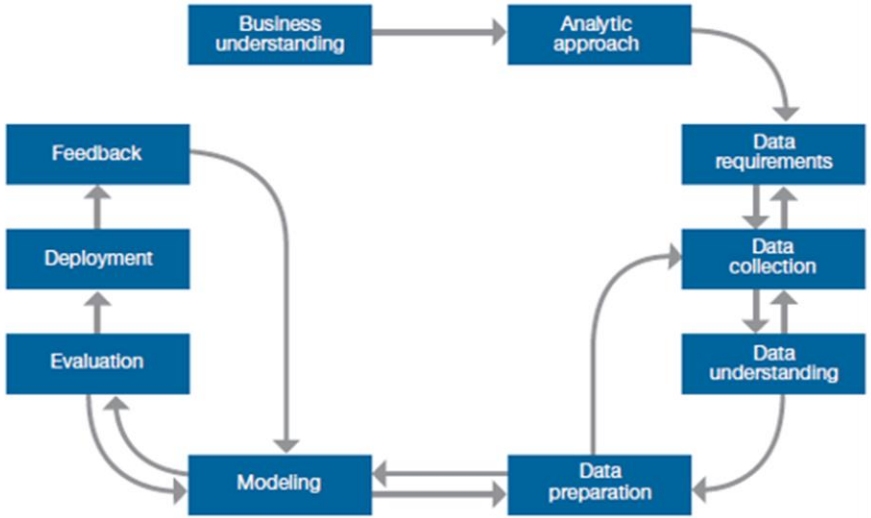


4

Data Science Methodology

CRISP-DM

Cross Industry Standard Process for Data Mining



IBM Data Science Methodology

From Problem to Approach

Every project begins **with business understanding.**

- Project objective?
- Business sponsors play the most critical role
- What are we trying to do – what is the goal?
- How do you define “success” and how can you measure it?

Example : Traffic

Problem: Traffic congestion wastes time and money

Clear question: How can we optimize traffic light duration using data on traffic patterns, weather, and pedestrian traffic?

Measurable outcomes:

- % decrease in commute time
- % decrease in length/duration of traffic jams

From Problem to Approach

Analytic Approach

Expand machine learning techniques

Regression:

"Predicting press problem in context of statistical revenue in the next quarter?"

Classification:

"Does this patient have cancer A, cancer B, or are they healthy?"

Clustering:

"Are there groups of users that seem to behave similarly to each other?"

Recommendation/Personalization:

"How can I target discounts to specific customers?"

Outlier Detection

From Requirements to Collection

The chosen analytic approach determines the **data requirements.**

- Content,
- formats,
- representations

Initial **data collection** is performed.

- Available Data?
- Obtain data?
- Revise data requirements or collect more data?

From Understanding to Preparation

Then **data understanding** is gained.

- Initial insights about data
- Descriptive statistics and visualization
- Additional data collection to fill gaps, if needed

Data preparation encompasses all activities t
construct and clean the data set.

Data cleaning

- Missing or invalid values
- Eliminating duplicate rows
- Formatting properly

Combining multiple data sources

Transforming data

Feature engineering

Text analysis

Accelerate data preparation by automating commor
steps

From Modeling to Evaluation

Modeling:

- Developing predictive or descriptive models
- May try using multiple algorithms
- Highly iterative process

Model **evaluation** is performed during model development and before model deployment

- .Understand the model's quality
- Ensure that it properly addresses the business problem

Diagnostic measures

- Suitable to the modeling technique used
- Training/Testing set
- Refine model as needed

From Deployment to Feedback

Once finalized, the model is **deployed** into a production environment.

- • May start in a limited / test environment

Involves other roles:

- Solution owner
- Marketing
- Application developers
- IT administration

Getting **Feedback** :

- How well did the model perform?
- Iterative process for model refinement and redeployment
- A/B testing

Let's start our Data science Hands on Journey!

Stay Tuned☺

THANKS!

Any questions?