The Data Science Process

8/26/2019

Today's Lesson

Learning Objectives

- Describe the phases of the CRISP-DM data science process
- Use the CRISP-DM framework to analyze and plan out data science projects

Activities

- Reflecting on Your Data Science Experience
- Data Science Process Overview
- Revisiting Your Data Science Experience
- Reviewing Your Projects
- Exit Ticket

Your Data Science Experience

Think about a time when you have personally experienced a data science application in real life.

Task: Take turns with a neighbor, answering the questions below:

- What was the experience?
- What about it made you think it was “data science”?
- What, if anything, do you know about the model that was used?
- How did the experience affect you or others?

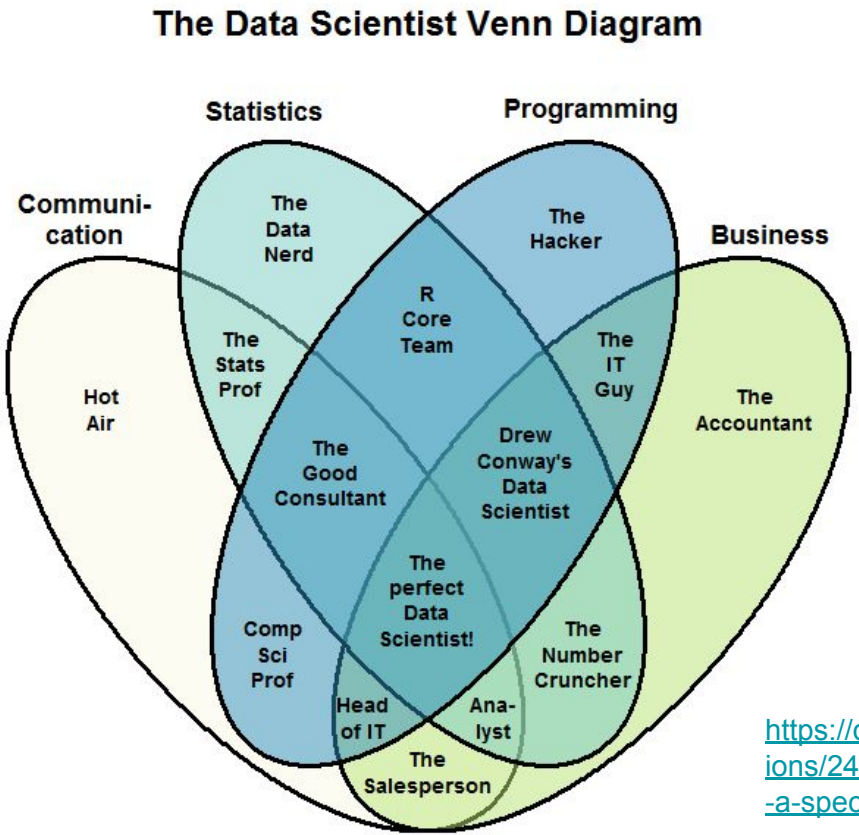
With your partner, pick one experience to share with the class.

Why bother defining a “data science process”?

- Common language for data scientists and business colleagues
- Project planning
- Evaluate your and others’ work
- Ensure you remember the big picture

//

Defining “Data Scientist”



https://datascience.stackexchange.com/questions/2403/data-science-without-knowledge-of-a-specific-topic-is-it-worth-pursuing-as-a-ca

Defining “Data Science”

Data Science **OSEMN** Model

Adapted from: KDNuggets

Obtain

- from other location
- Query from database or API
- Extract from another file
- Generate data (e.g. Sensors)

Scrub

- Filtering lines
- Extracting columns or words
- Replacing values
- Handling missing values
- Converting formats

Explore

- Understanding data
- Deriving statistics
- Creating visualization

Model

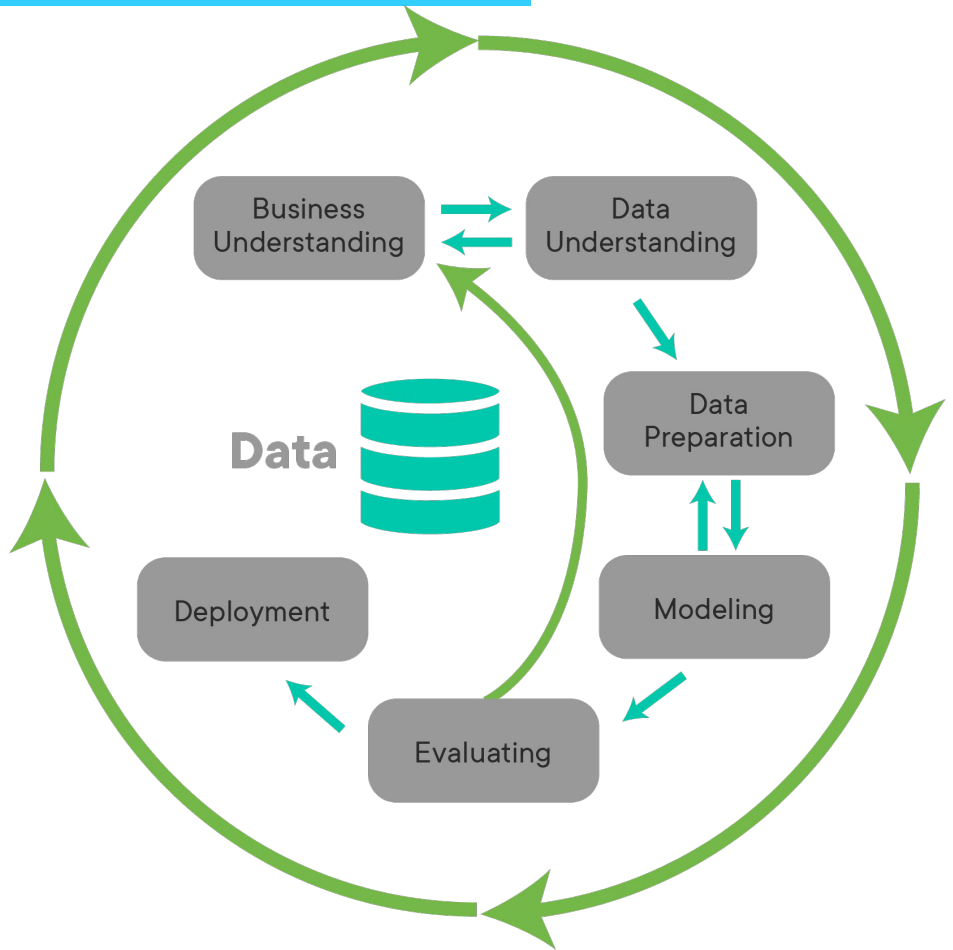
- Clustering
- Classification
- Regression
- Dimensionality reduction

INterpret

- Drawing conclusion from data
- Evaluating meaning of results
- Communicating result

Defining “Data Science”

Cross-Industry Standard Process for Data Mining (CRISP-DM)



https://www.sv-europe.com/crisp-dm-methodology/

http://cdn.intechopen.com/pdfs/5937/InTech-A_data_mining_amp_knowledge_discovery_process_model.pdf

Defining “Data Science”

Business understanding	Data understanding	Data preparation	Modeling	Evaluation	Deployment
Determine business objectives	Collect initial data	Select data	Select modeling techniques	Evaluate results	Plan deployment
Assess situation	Describe data	Clean data	Generate test design	Review process	Plan monitoring & maintenance
Determine DM objectives	Explore data	Construct data	Build model	Determine next steps	Produce final report
Produce project plan	Verify data quality	Integrate data	Assess model		Review project
		Format data			

Your Data Science Experience #2

Task: Work with your previous partner to revisit one experience you discussed with them.

Step into the shoes of the data scientist responsible for creating the experience you had. What do you imagine that person did or said during the Business Understanding or Evaluation phases?

What goal were they trying to achieve?
Do you think that they think they were successful?

Your Data Science Project

Task: With a new partner, take turns showing each other your regression project and discuss answers to the following questions:

- Which of the phases were strong for your project, and which were limited or didn't occur?
- Evaluate your model from a business perspective, rather than a statistical one. What real-world need might your project have served, and how well would it have done so?
- Imagine that you were going to take what you've learned and redo your project for a real client. What specific activities might you engage in to improve each phase of the process?

Summary + Exit Ticket

Presented by David Braslow

FLATIRON SCHOOL