

Introduction to Data Science



In your groups discuss

- What are excited to learn? 😄
- What are your nervous about? 😅
- What can Jess do to help you be successful? ★
- What can you do to be successful? 💪

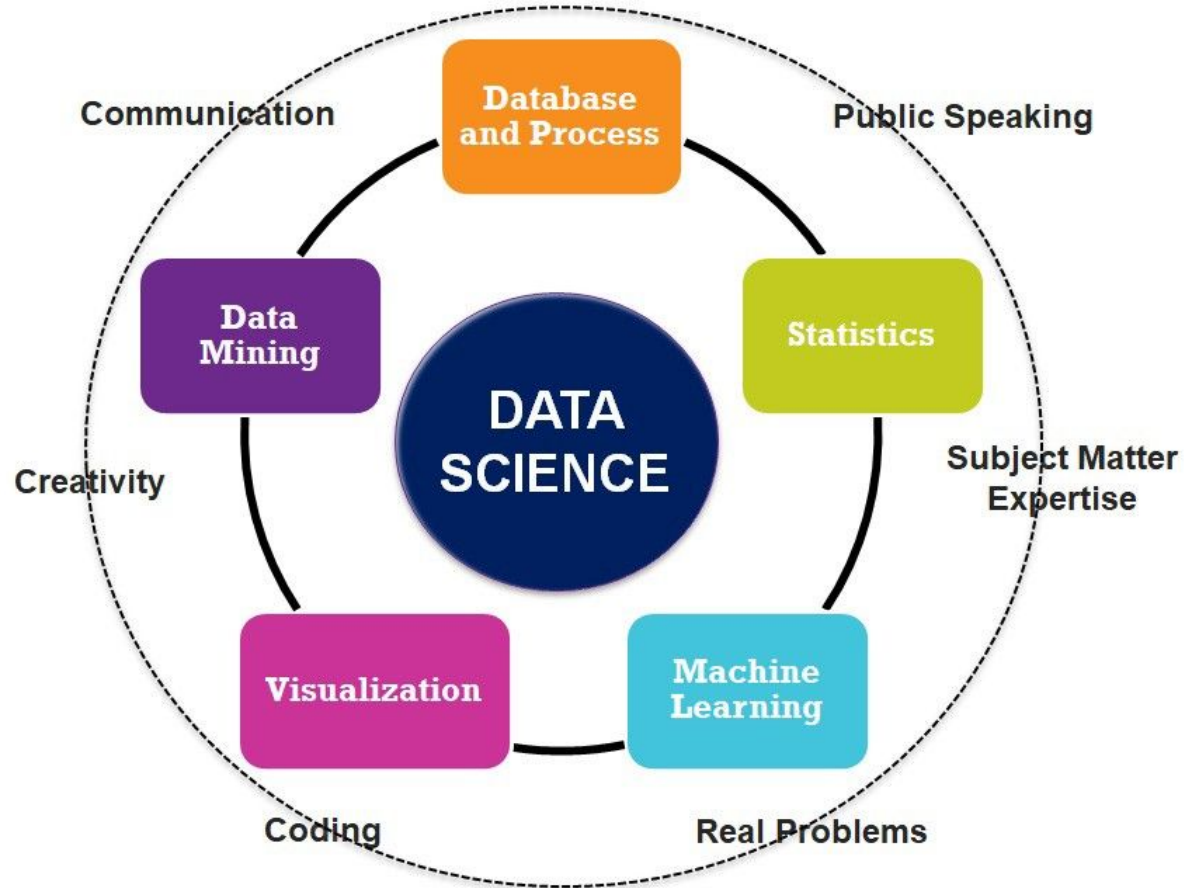


Students, write your response!

In this class we will explore

- Manipulating/Exploring data
- Visualizing data
- Analyzing data in meaningful ways
- Common tools and vocab
- Creating basic machine learning models and performing analysis

What is Data Science?



Why are you interested in Data Science?



Students, write your response!



Data Science

Advantages

A It's in Demand

Abundance of Positions **B**

C Highly Paid Career

Highly Prestigious **D**

E Versatile

Disadvantages

A It is a Blurry Term

Mastering Data Science is near to impossible **B**

C Large amount of domain knowledge required

Arbitrary Data May Yield Unexpected Results **D**

E Problem of Data Privacy

What are the different data science roles?

Data Science Courses

| Course | Course Title | Prerequisites |
|--------|----------------------------------|---------------|
| DS 1.1 | Data Analysis & Visualization | CS 1.2, QUANT |
| DS 2.1 | Introduction to Machine Learning | DS 1.1 |
| DS 2.2 | Neural Networks & Deep Learning | DS 2.1 |
| DS 2.3 | Production Data Science Tools | DS 2.1 |
| DS 2.4 | Advanced Topics in Data Science | DS 2.2 |

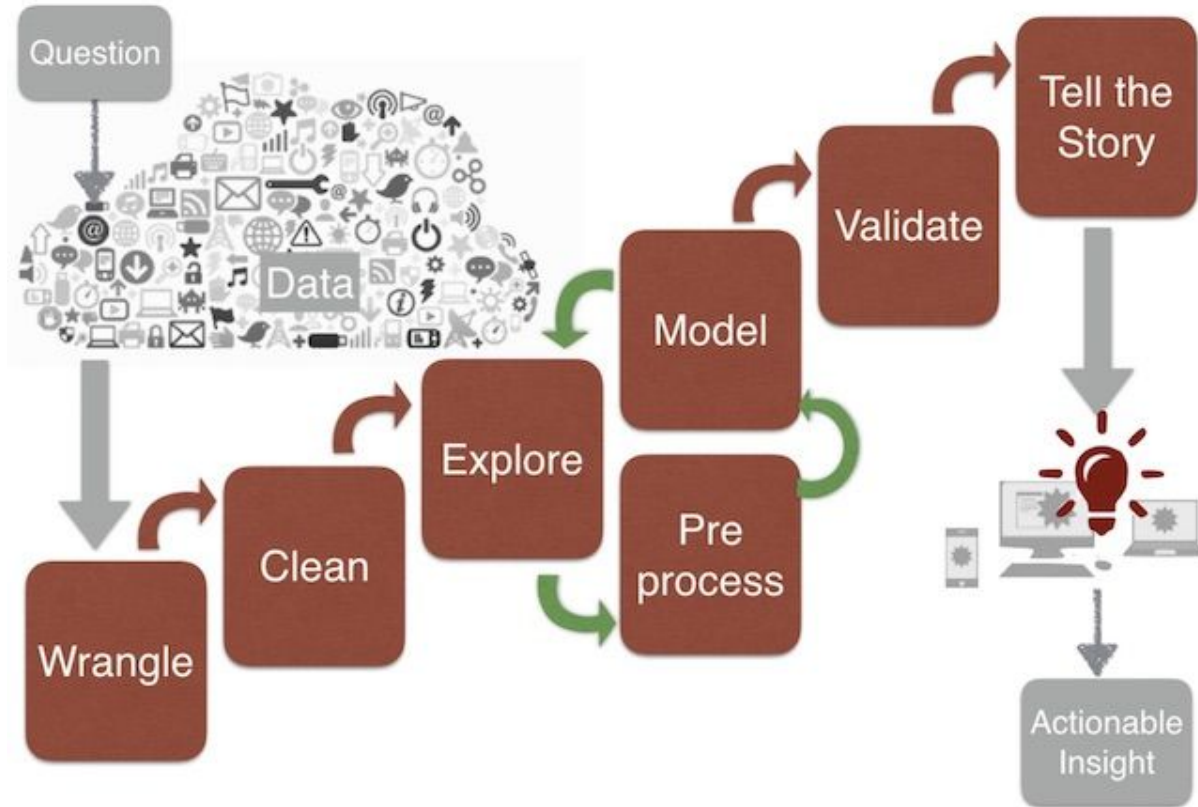
Core Technical Courses

| Course | Course Title | Prerequisites |
|---------|--|---------------|
| BEW 1.1 | RESTful & Resourceful MVC Architecture | - |
| BEW 1.2 | Authentication, Authorization & Advanced | BEW 1.1 |
| CS 1.1 | Programming Fundamentals | - |
| CS 1.2 | How Data Structures Work | CS 1.1 |
| CS 1.3 | Core Data Structures & Algorithms | CS 1.2 |

+ SPD Every Term

Data Science Track Roadmap

Exploratory Data Analysis



Toolsets

KEY LANGUAGES



DATA ANALYSIS

Pandas



VISUALIZATION TOOLS

seaborn

matplotlib

ESSENTIAL RESOURCES



kaggle







Python Pandas **Features**

Multiple file
formats supported

Python
support

Input and
output tools

Optimized
performance

Great handling
of data

Handling
missing data

Grouping

Lot of time
series

Unique data

Cleaning
up data

Alignment
and indexing

Merging and
joining of datasets

Visualize

Mask data

kaggle™

A platform for predictive modeling competitions.



“We're making data science into a sport.”

Video Game Sales Exploration

