Introduction to Data Science with Python
Lecture 1: General Introduction

Vladimir Osin

Data Scientist/Engineer

Signify Research (formerly known as Philips Lighting)
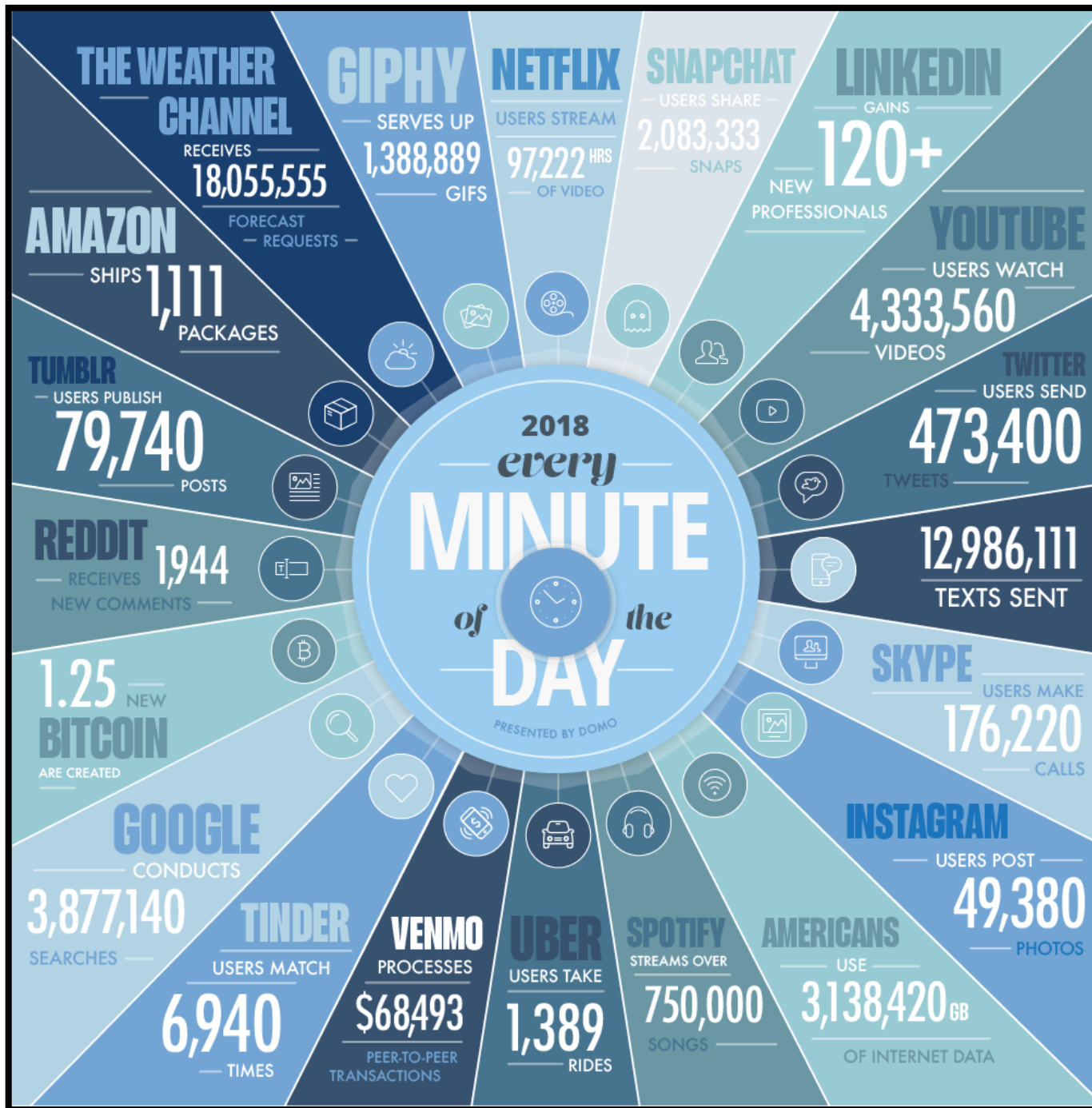
Outline

- Data is the new oil? Spoiler Alert: YES
  - Artificial Intelligence/Machine Learning - unhyped
  - Some amazing results

- Data Scientist, huh?
  - Path to success
  - CRISP-DM

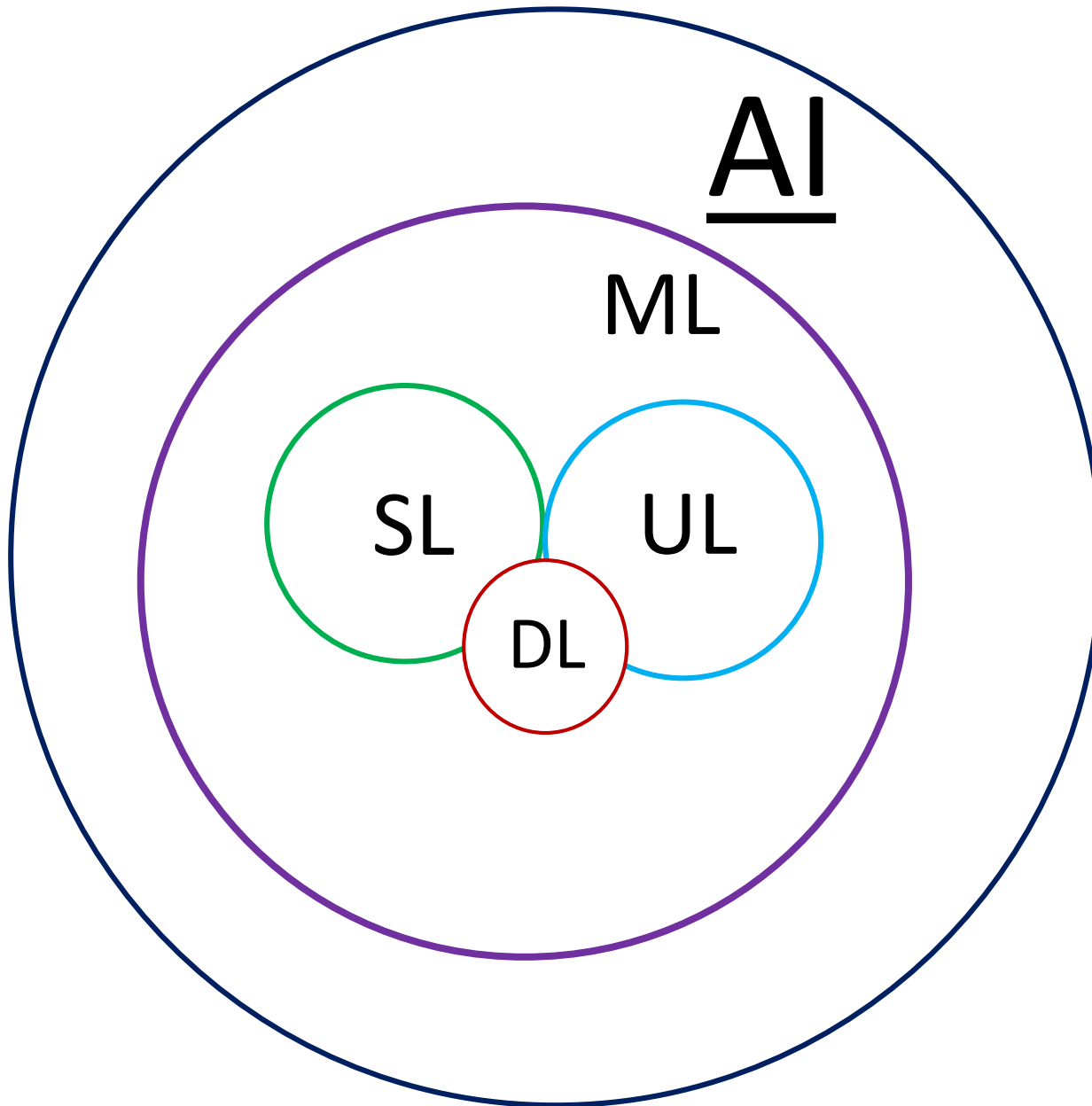- Exploratory Data Analysis
  - Tools/Tricks/Tips

"Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."

Clive Humby

Every **Minute** Of The Day

- **AI research** is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals. (Smart Toaster?)

- **Machine learning** is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with data, without being explicitly programmed.
  - Supervised learning
  - Unsupervised learning

- **Deep Learning** is part of a broader family of machine learning **methods** based on learning data representations, as opposed to task-specific algorithms.

U.S. fatalities associated with human choice or error (2014)

9,262
Speeding

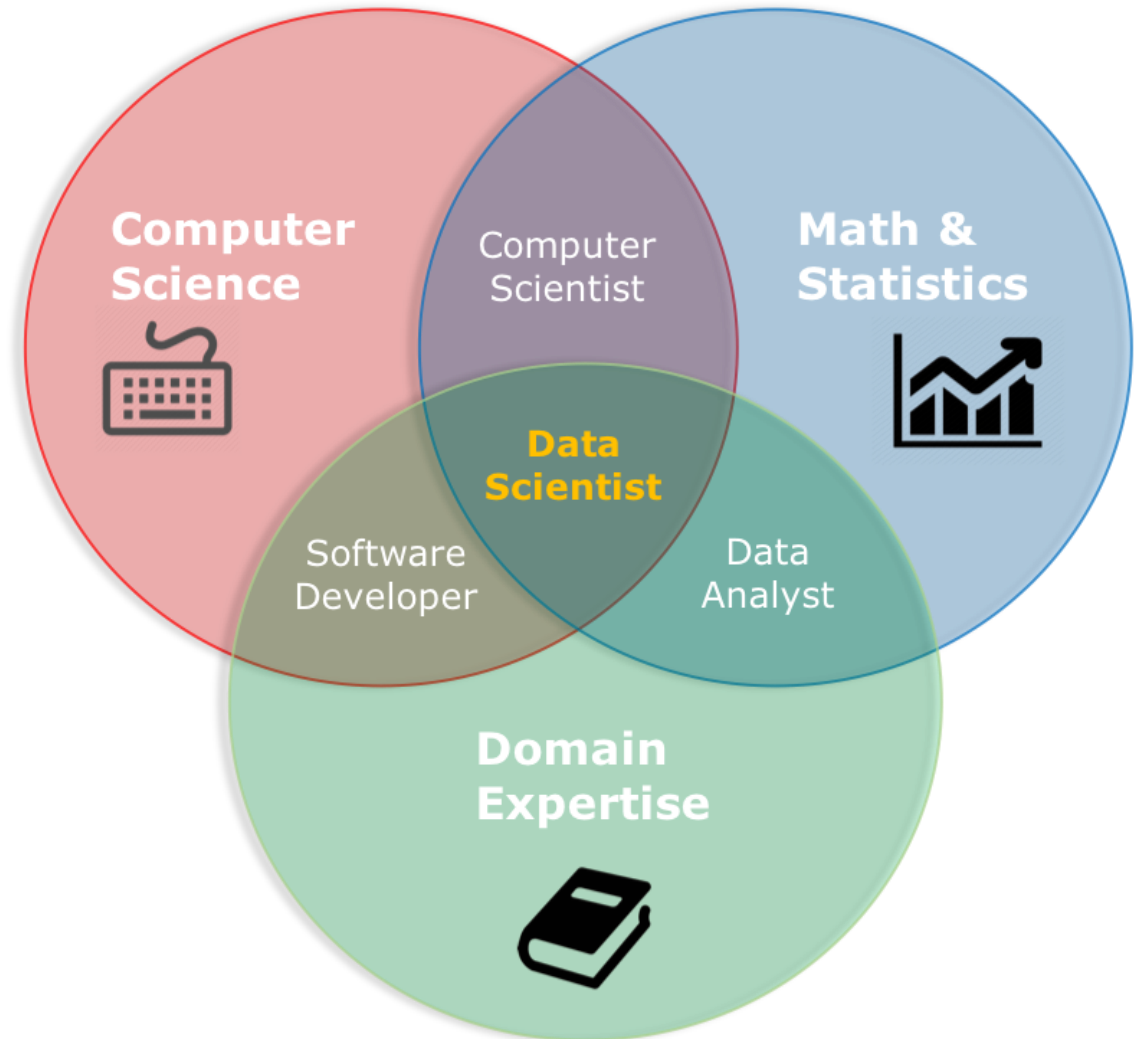**9,967**
Alcohol

3,179
Distraction

846
Drowsiness

''A **data scientist** is someone who is better at **statistics** than any software engineer and better at **software engineering** than any statistician.''
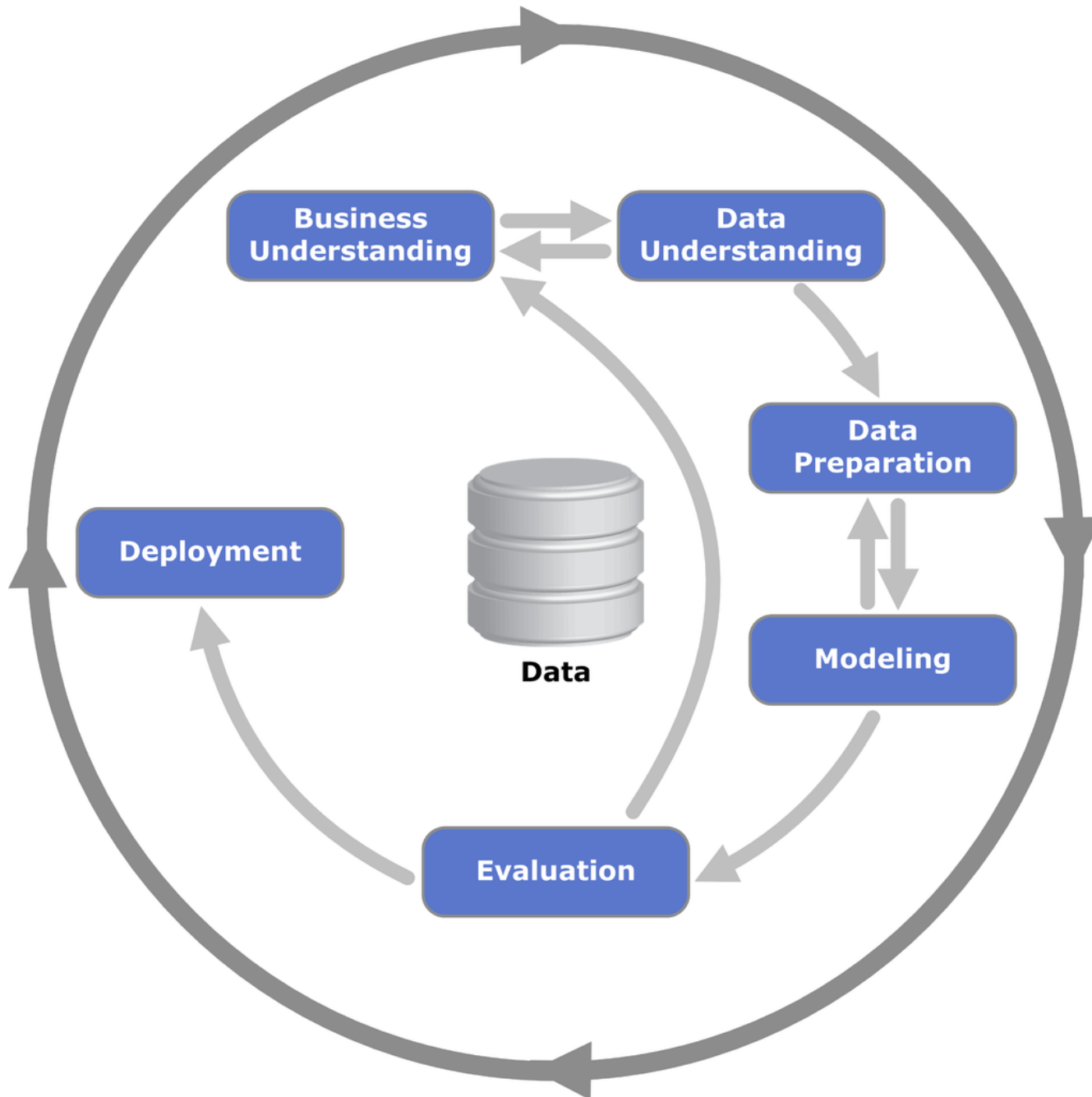
Josh Wills

Data Scientist

- Path to success:
  - **Courses**:
    - Programming: **Python** or R
    - Linear algebra/Calculus
    - Statistics
    - Big Data Tools: **Spark**
  - **Kaggle**
  - Stay hungry and curious
  - The best time is now
  - Keep going, never give up

CRISP-DM
Cross-industry standard process for data mining

Exploratory Data Analysis

- Approach to analyzing data sets to summarize their main characteristics, often with visual methods.