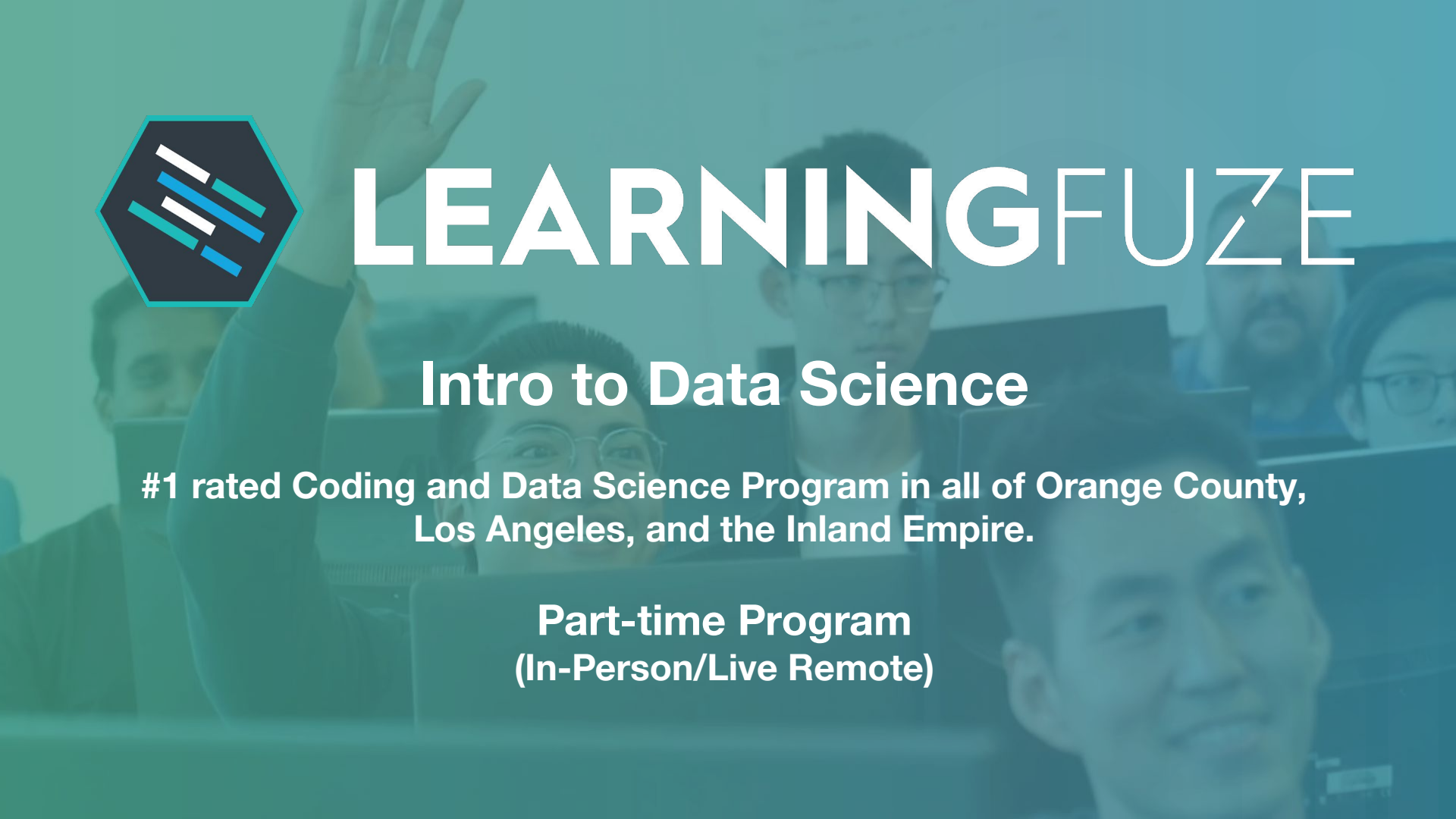
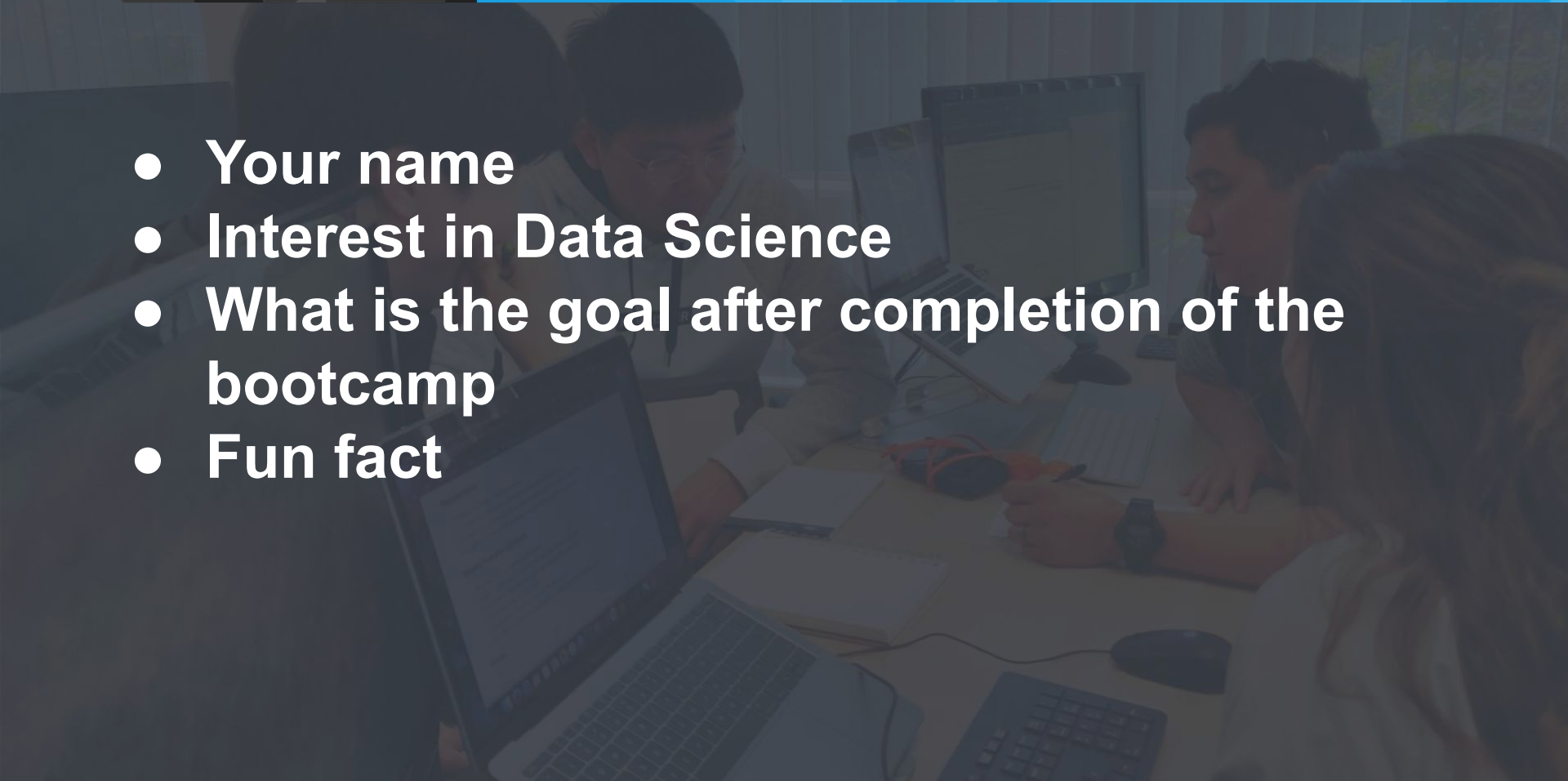

LEARNINGFUZE
Intro to Data Science
#1 rated Coding and Data Science Program in all of Orange County,
Los Angeles, and the Inland Empire.
Part-time Program
(In-Person/Live Remote)

LEARNINGFUZE

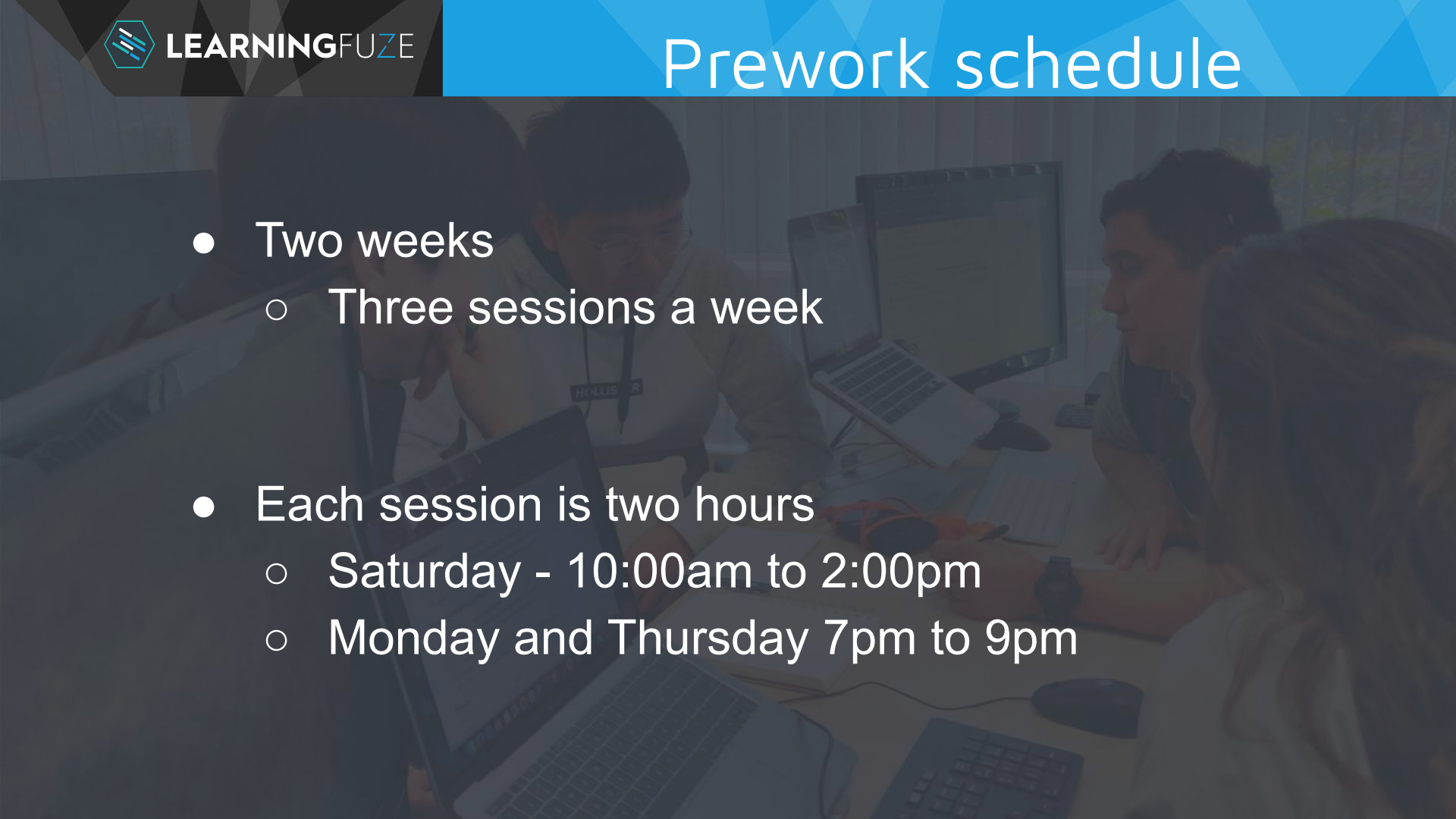
# Introductions



- Your name
- Interest in Data Science
- What is the goal after completion of the bootcamp
- Fun fact

LEARNINGFUZE

# Prework schedule

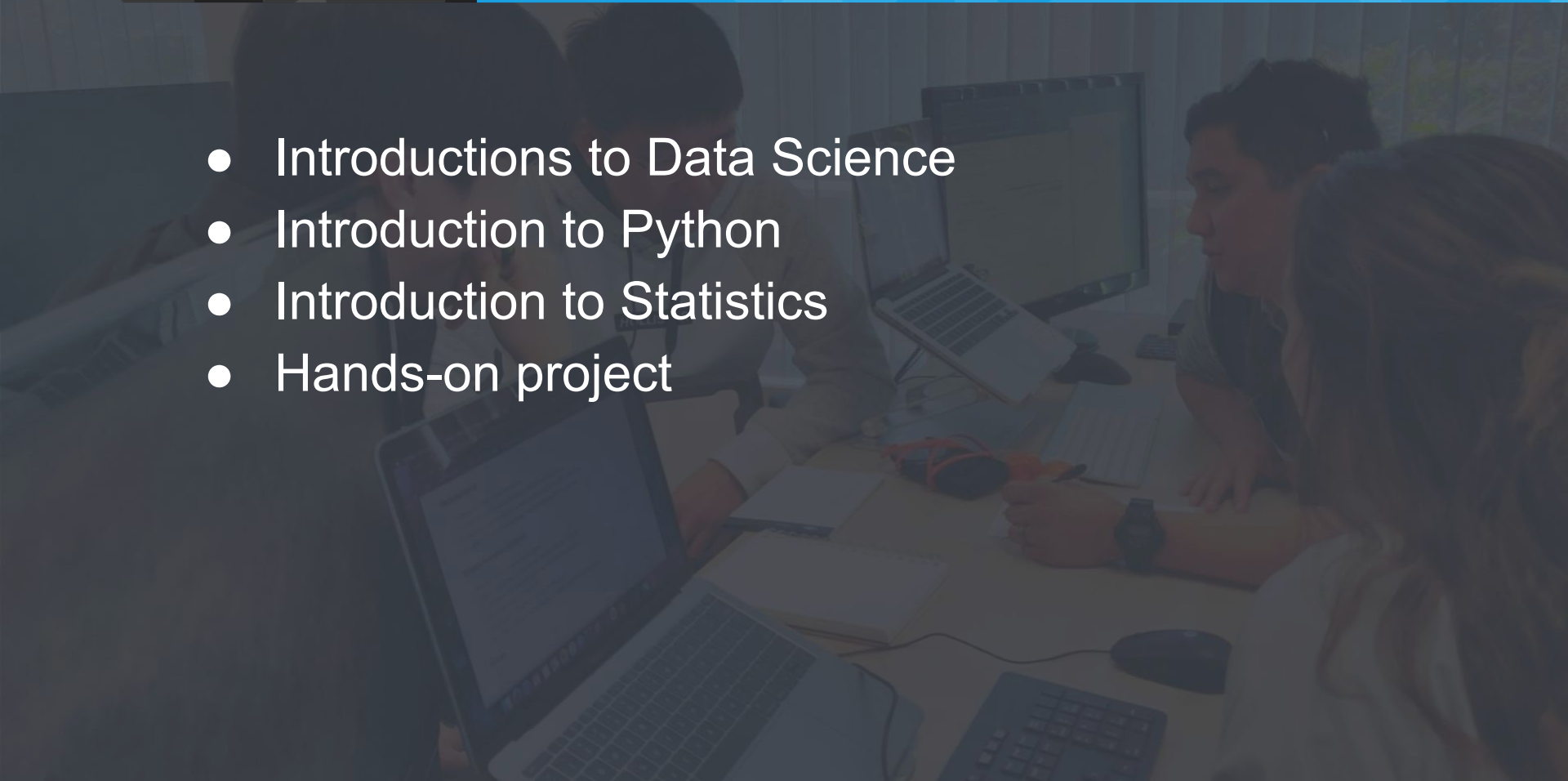


- Two weeks
  - Three sessions a week

- Each session is two hours
  - Saturday - 10:00am to 2:00pm
  - Monday and Thursday 7pm to 9pm

LEARNINGFUZE

# Intro to Data Science



- Introductions to Data Science
- Introduction to Python
- Introduction to Statistics
- Hands-on project

LEARNINGFUZE

# Intro to Data Science

## Session 1

- What is Data Science?
- Data Science methodologies
- Industry terminology (Data Engineering, Artificial Intelligence, Data Science, Big Data, Machine Learning, Deep Learning)
- Set-up Visual Studio Code
- Introduction to Git

## Session 2

- Introduction to Python
- Flow control

## Session 3

- Data structures
- Functions

## Session 4

- Introduction to Statistics
- Descriptive statistics

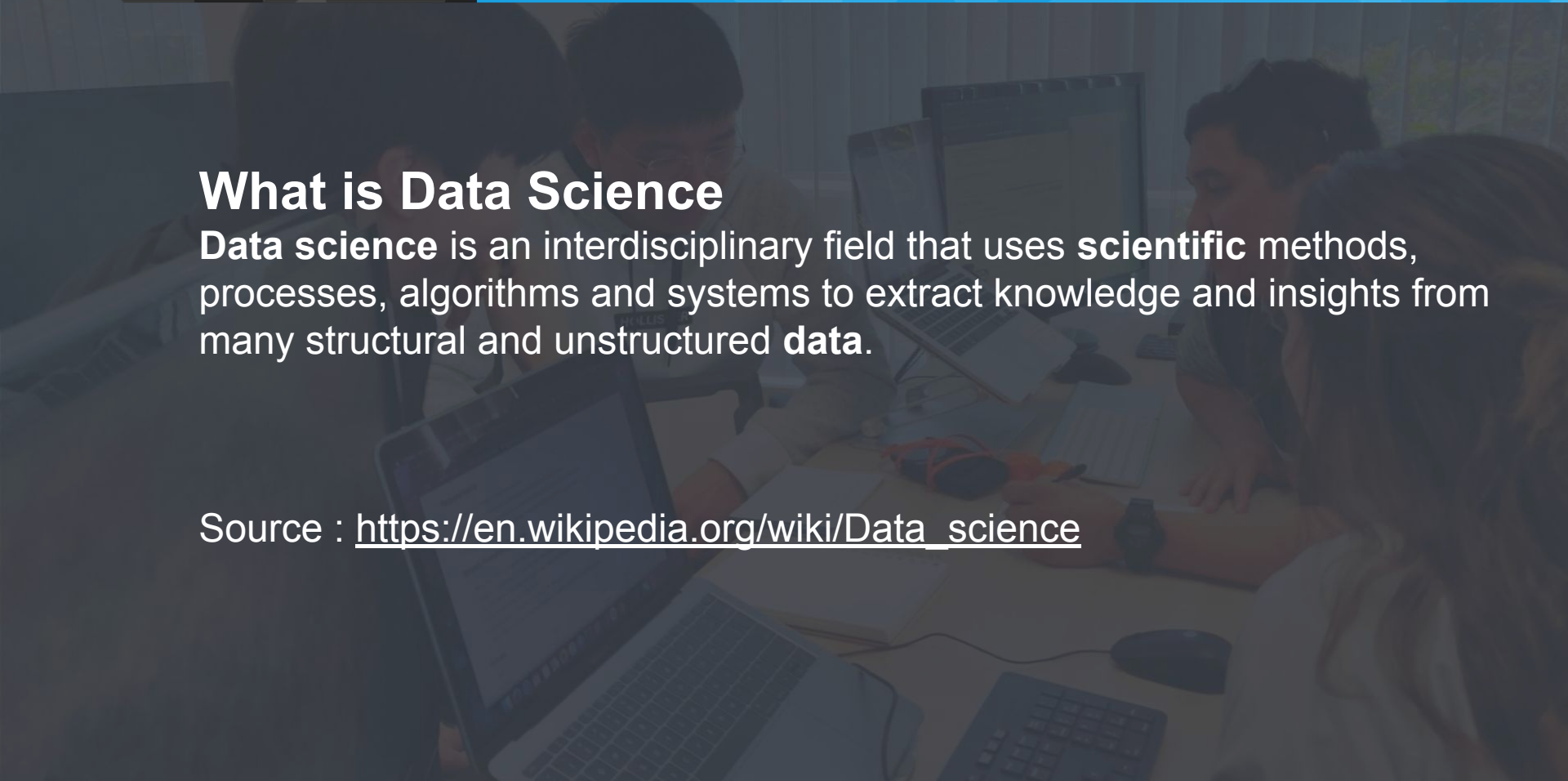
## Session 5

- Descriptive statistics
- Standard deviation and variance

## Session 6

- Hands-on Data Science Project from Kaggle
- Predict who is at risk of having a heart stroke

# Let's get started

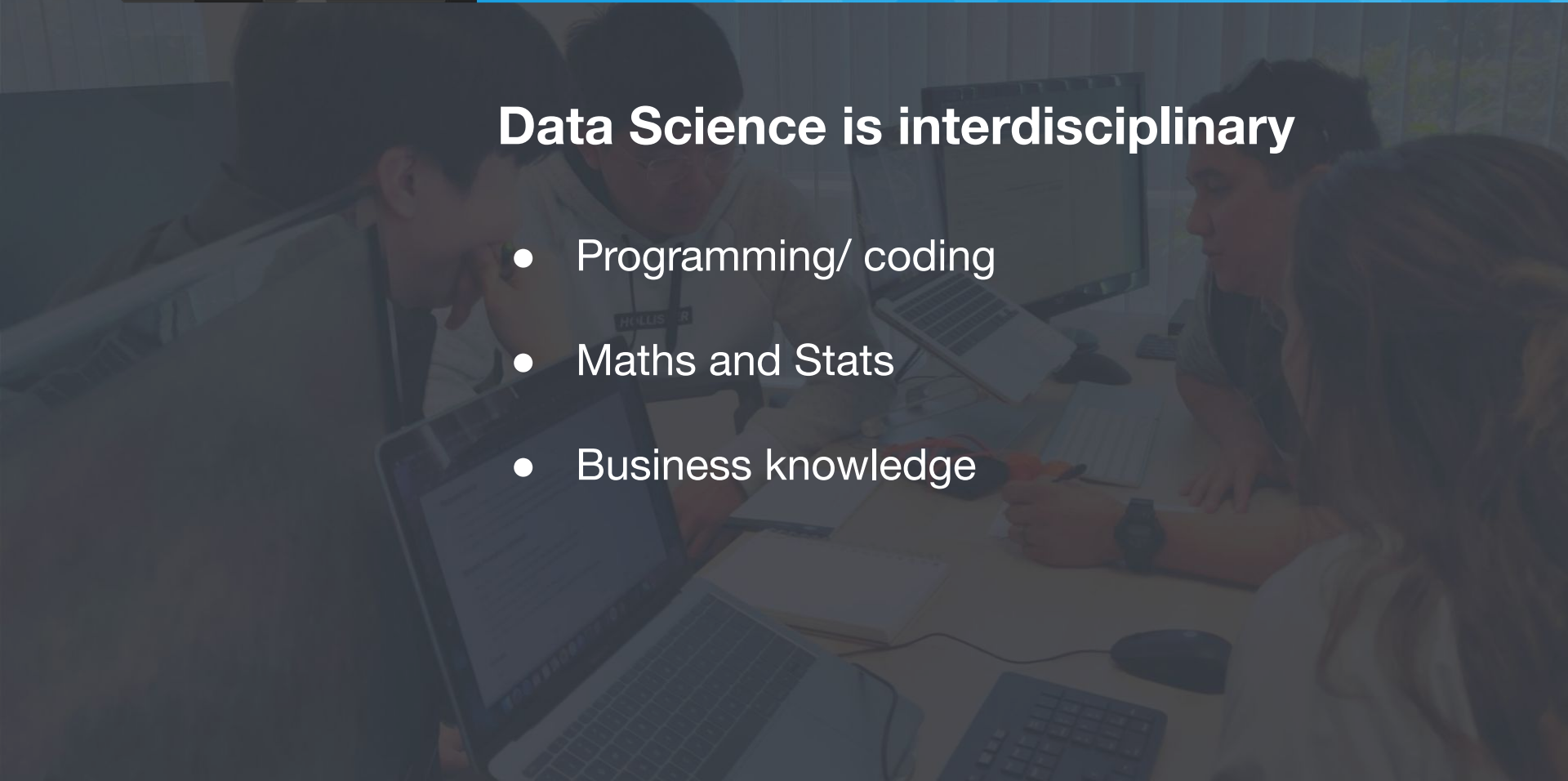


## What is Data Science

**Data science** is an interdisciplinary field that uses **scientific** methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured **data**.

Source : https://en.wikipedia.org/wiki/Data_science

# Interdisciplinary field



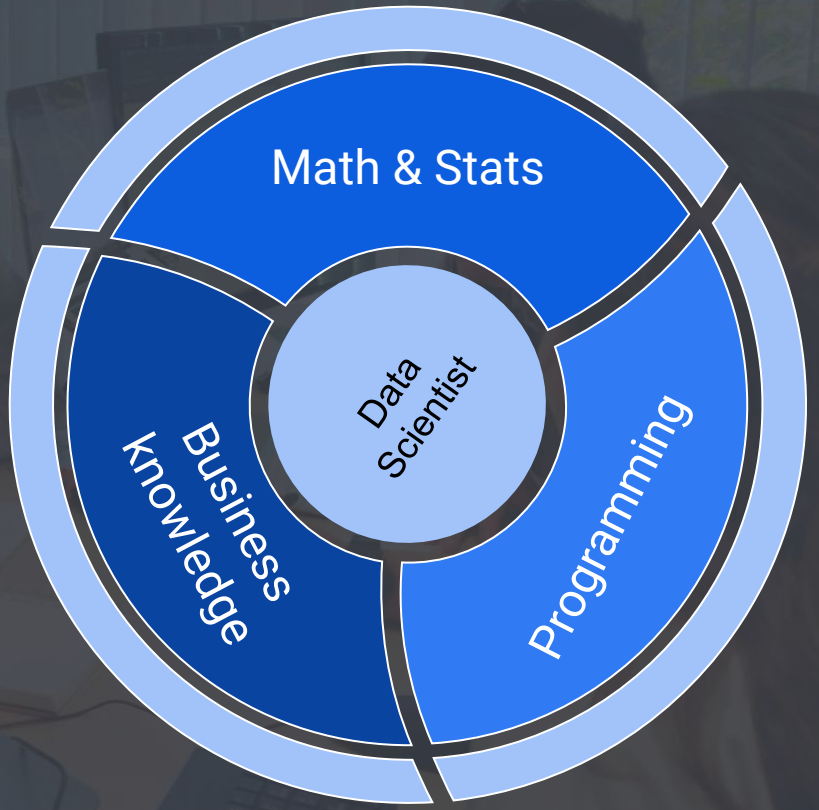
## Data Science is interdisciplinary

- Programming/ coding
- Maths and Stats
- Business knowledge

LEARNINGFUZE

# Who is the Modern Data Scientist?

- Passionate about building, creating, and assembling
- Curious and consider yourself a lifelong learner
- Enjoys challenges and problem solving
- Meticulous and detail-oriented
- Logical thinker
- Enjoy working with data

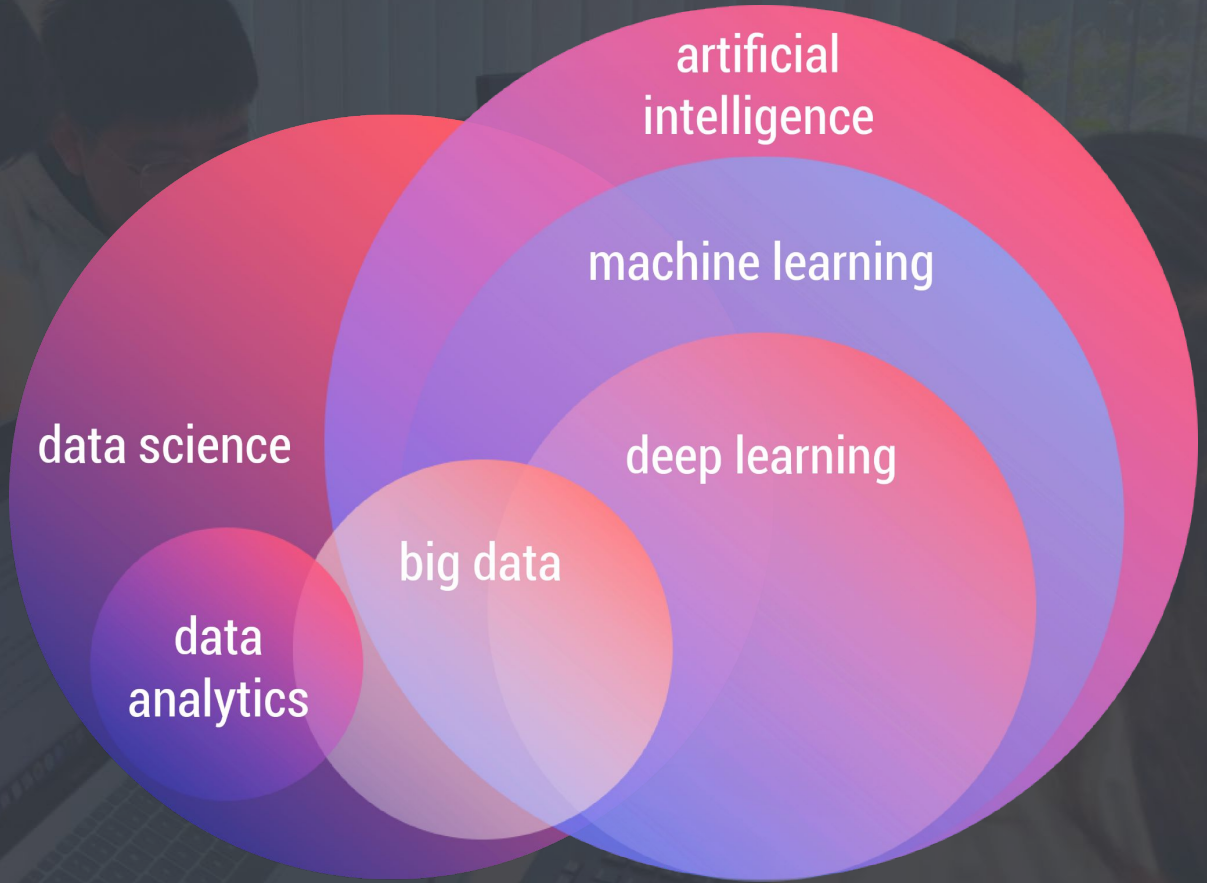


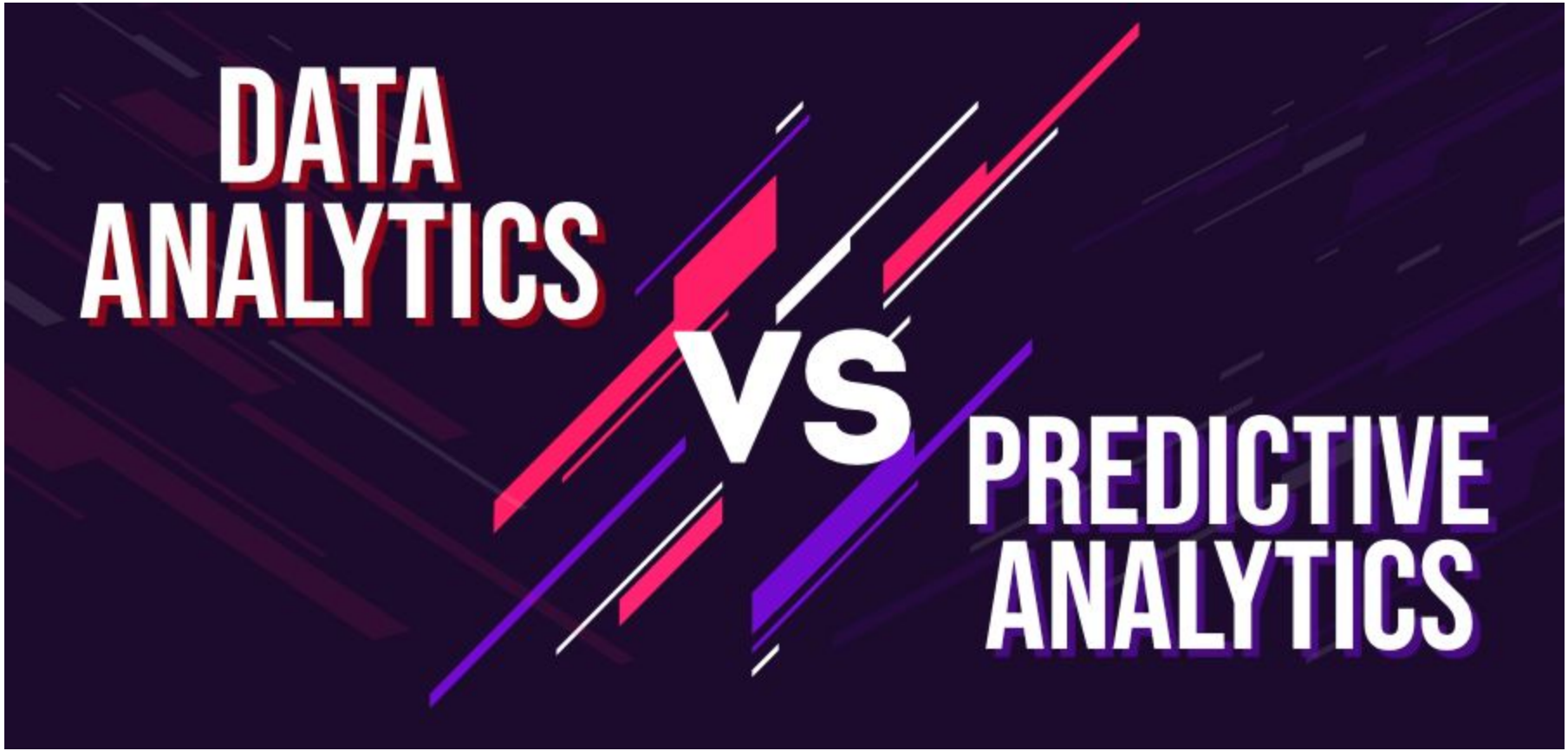
# Intro to Data Science

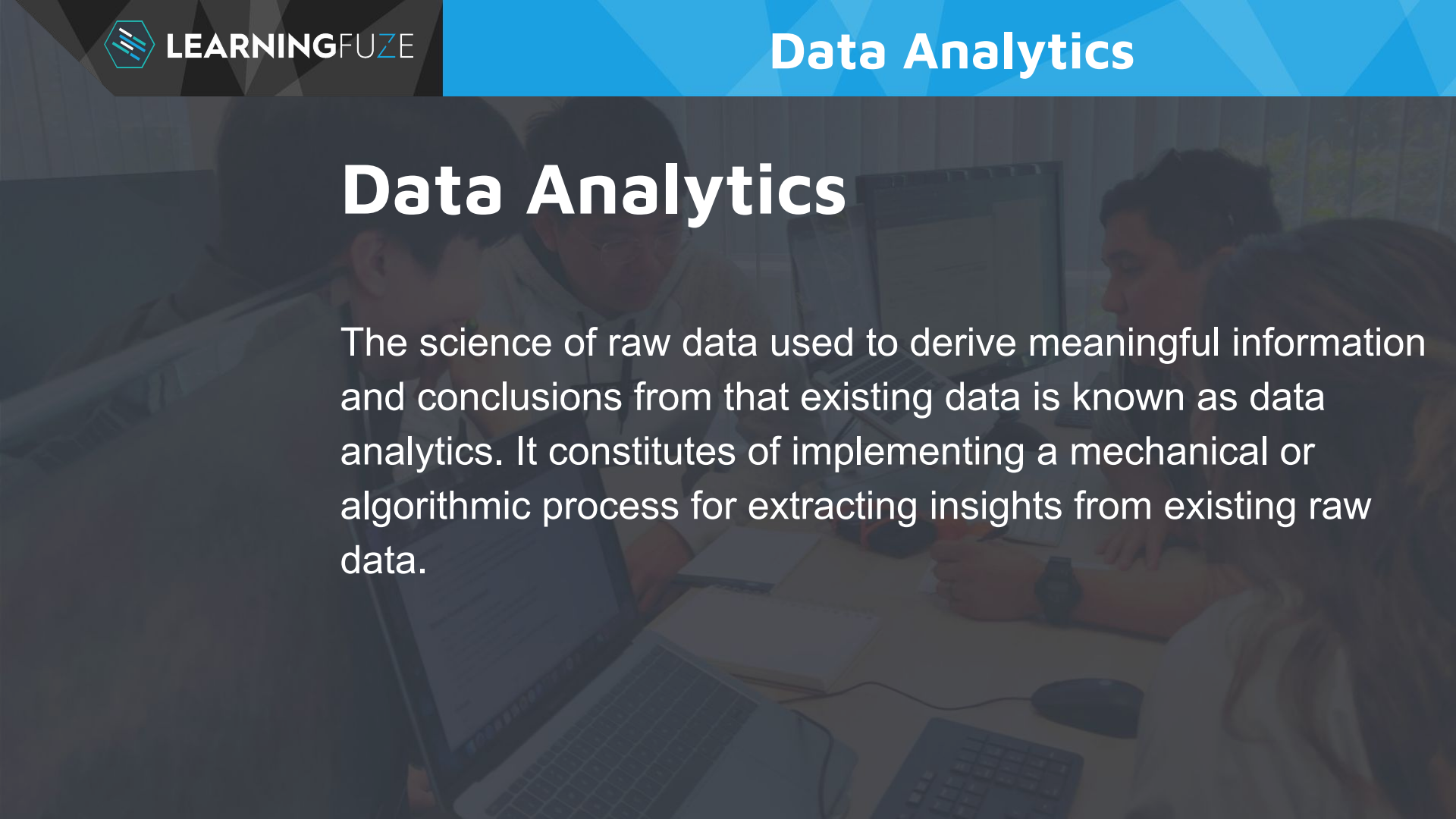
- Data Analytics
- Data Science
- Machine Learning (ML)
- Deep Learning (DL)
- Artificial Intelligence (AI)
- Big Data

# Intro to Data Science

Data Science & Artificial Intelligence ecosystem

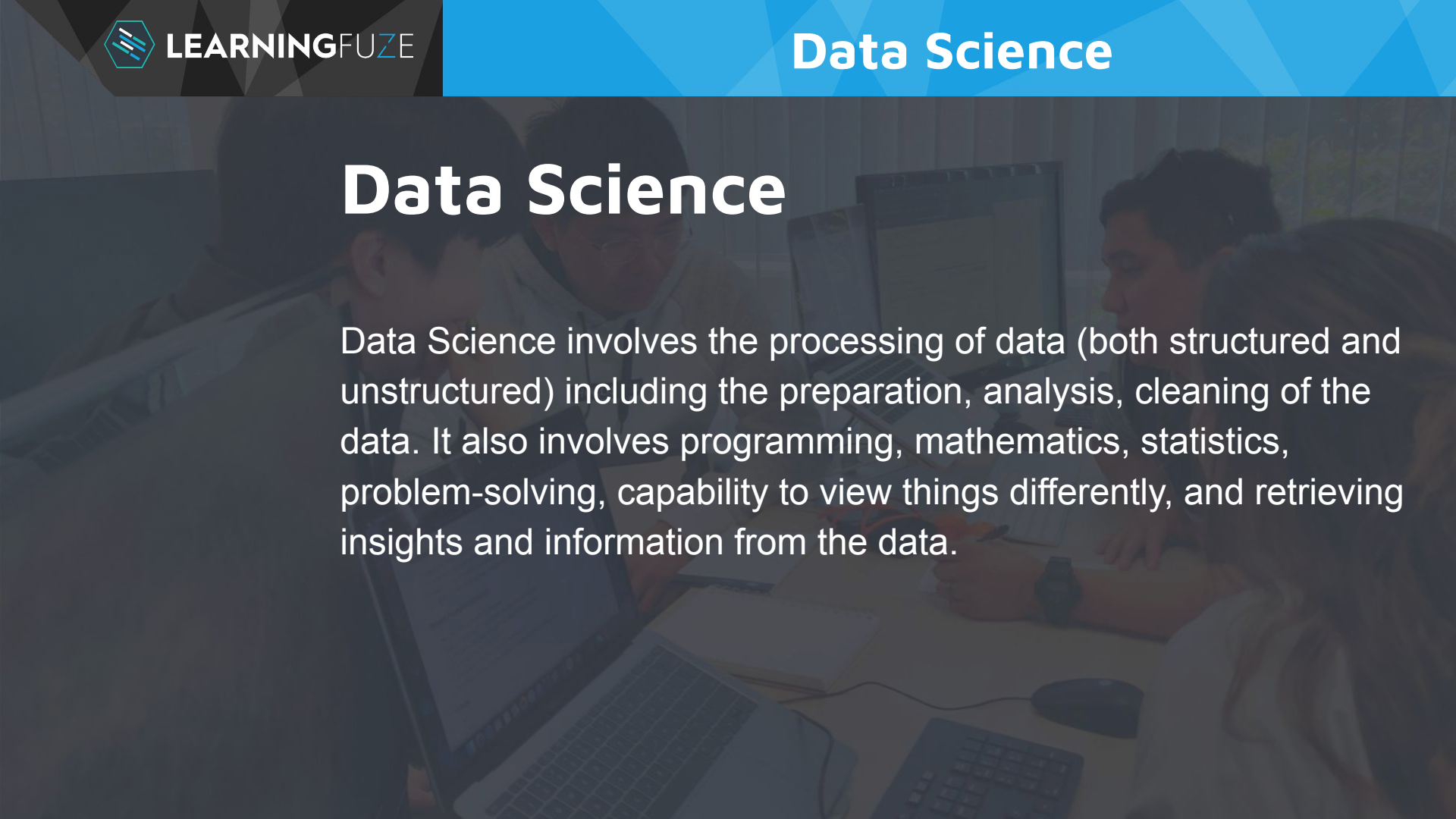



DATA
ANALYTICS
VS
PREDICTIVE
ANALYTICS



# Data Analytics

The science of raw data used to derive meaningful information and conclusions from that existing data is known as data analytics. It constitutes of implementing a mechanical or algorithmic process for extracting insights from existing raw data.



# Data Science

Data Science involves the processing of data (both structured and unstructured) including the preparation, analysis, cleaning of the data. It also involves programming, mathematics, statistics, problem-solving, capability to view things differently, and retrieving insights and information from the data.

LEARNINGFUZE

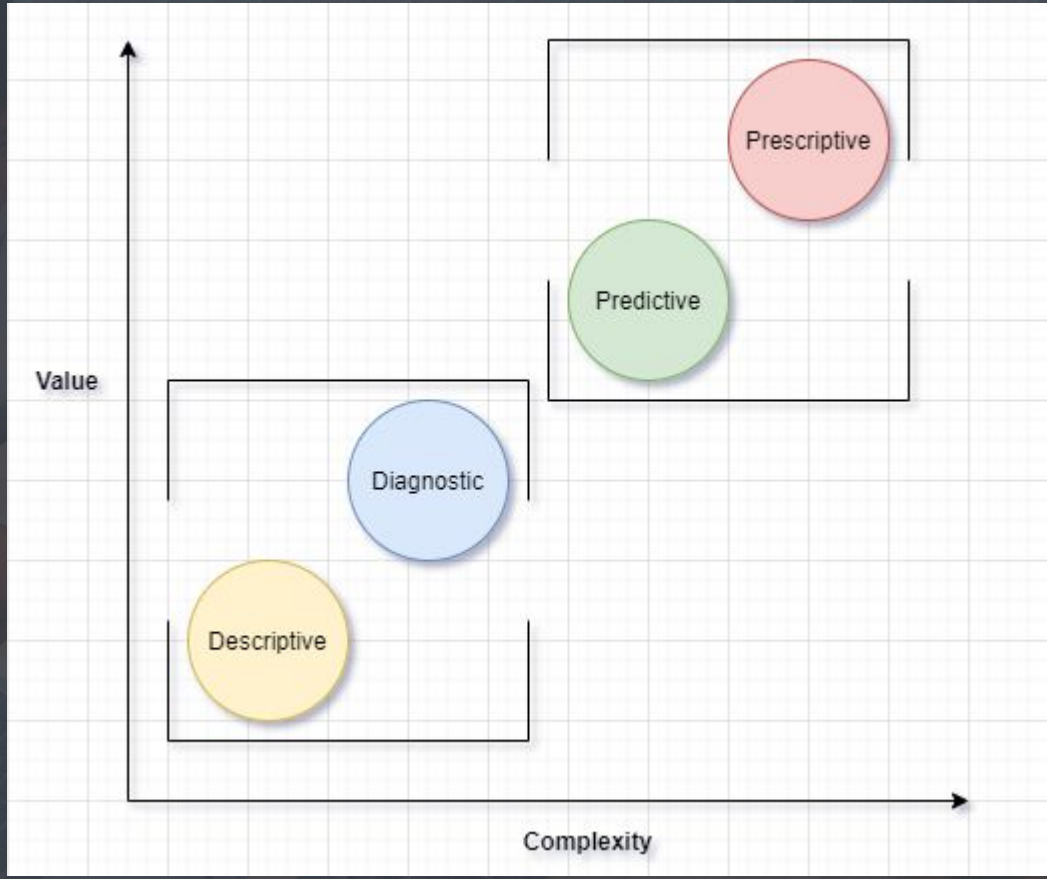
# Analytics value chain

**Descriptive:**
What happened

**Diagnostic:**
Why did it happen

**Predictive:**
What will happen

**Prescriptive:**
What do we need to do

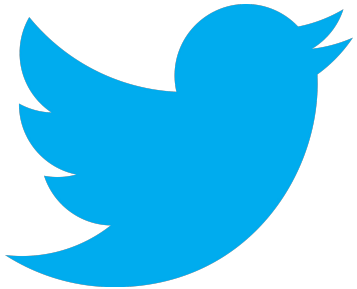



Google

amazon

NETFLIX

TARGET



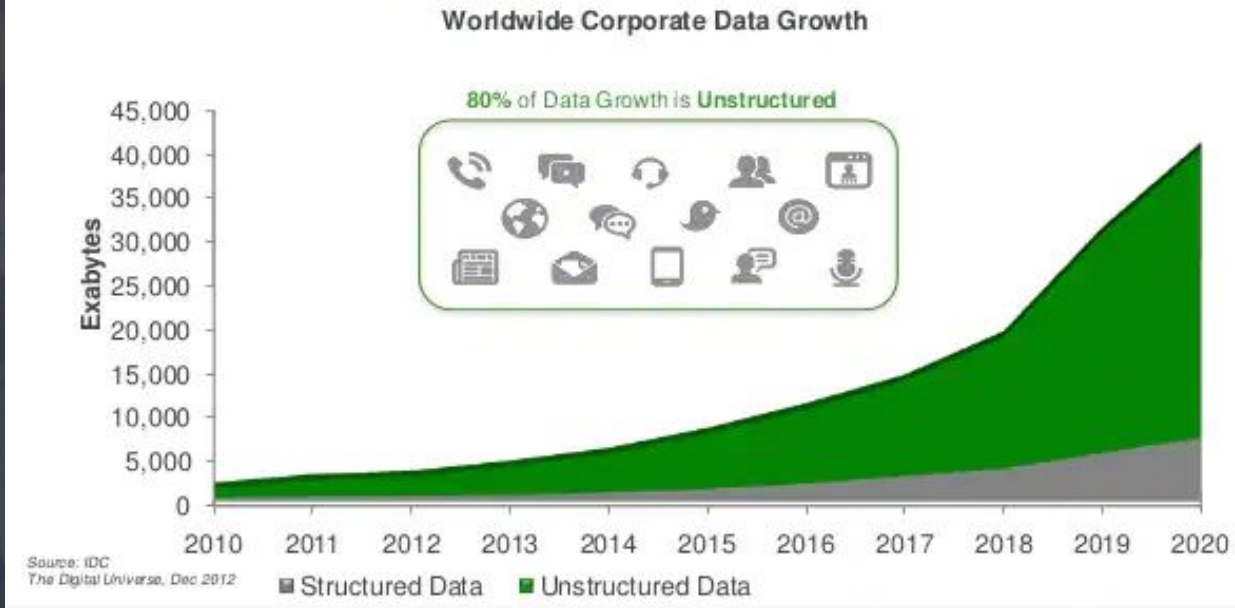

Zillow

LEARNINGFUZE

# Data Growth

## MASSIVE GROWTH IN UNSTRUCTURED CONTENT

RECOMMIND



# Big Data

*Big data is high-volume, and high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*

# Four V's of Big Data

- Volume - amount of data is large
- Velocity - data is generated or processed quickly
- Variety - data takes many different forms
- Veracity - data accurately represents reality

# Data Science use case

Hired by real estate company to build a predictive model to estimate price of a house

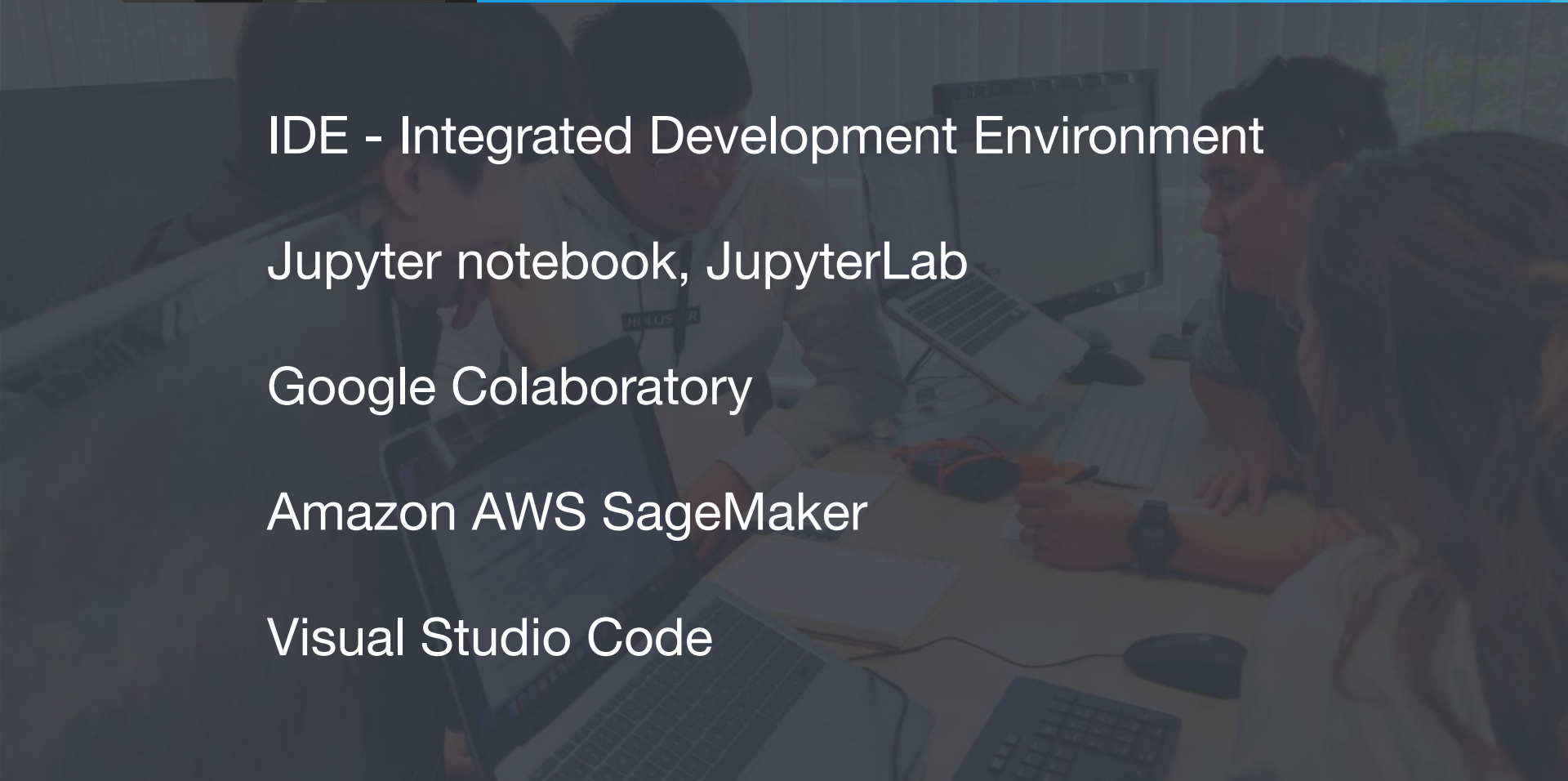
How do we go about that?

# Data Science Approach

- Business goal
- Data collection - Data gathering
- Data analysis and understanding
- Data preparation
- Modeling and model evaluation
- Business goal evaluation
- Implementation

# Data Science Tools

# Coding Environment



IDE - Integrated Development Environment

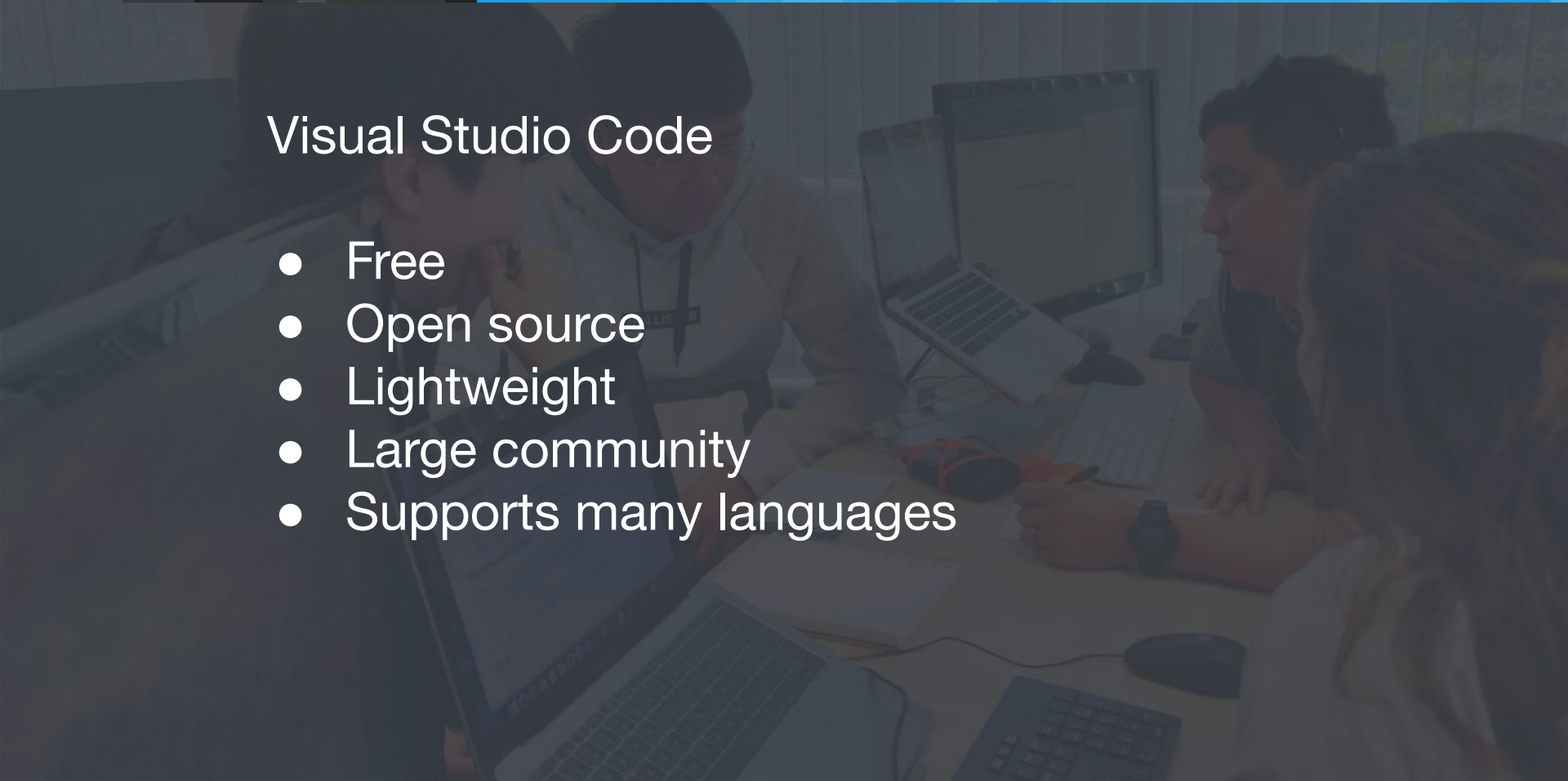
Jupyter notebook, JupyterLab

Google Colaboratory

Amazon AWS SageMaker

Visual Studio Code

# VS Code



## Visual Studio Code

- Free
- Open source
- Lightweight
- Large community
- Supports many languages