DS @ Flatiron School
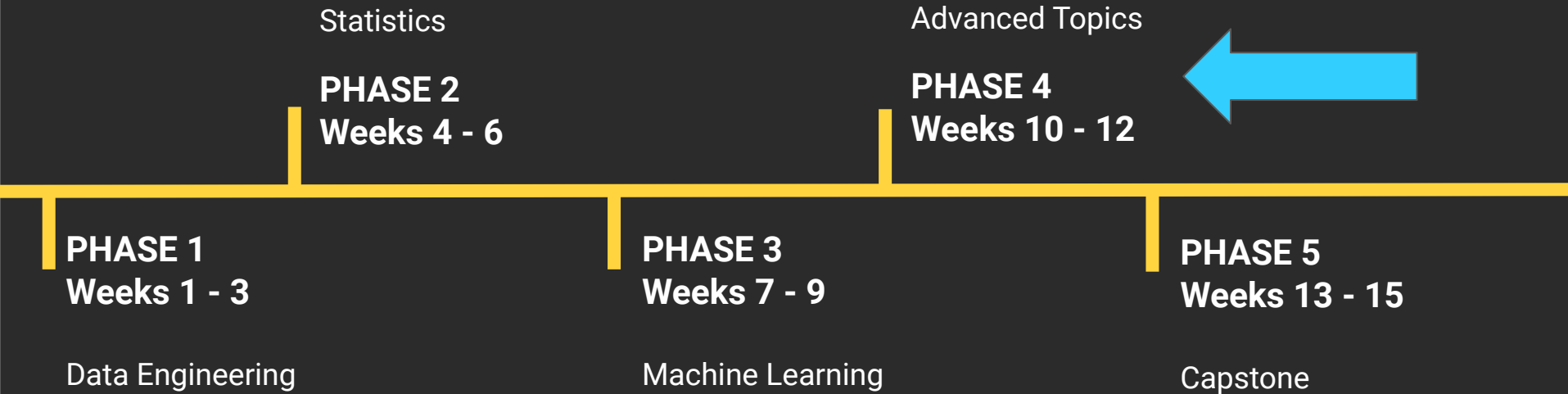# Phase 4 Expectations

Advanced Topics

Agenda

1. **What does student success look like in Phase 4?**

2. **Runway to Capstone**

3. **What can you expect to feel when going through Phase 4?**

Overall Phase Timeline

Statistics

PHASE 2
Weeks 4 - 6

Advanced Topics

PHASE 4
Weeks 10 - 12

PHASE 1
Weeks 1 - 3

Data Engineering

PHASE 3
Weeks 7 - 9

Machine Learning

PHASE 5
Weeks 13 - 15

Capstone

Tasting Menu of New Topics

Tools of Phase 4:

Unsupervised Learning

- ○ Dimensionality Reduction
  - i. PCA
  - ii. SVD
- ○ Clustering
  - i. Kmeans
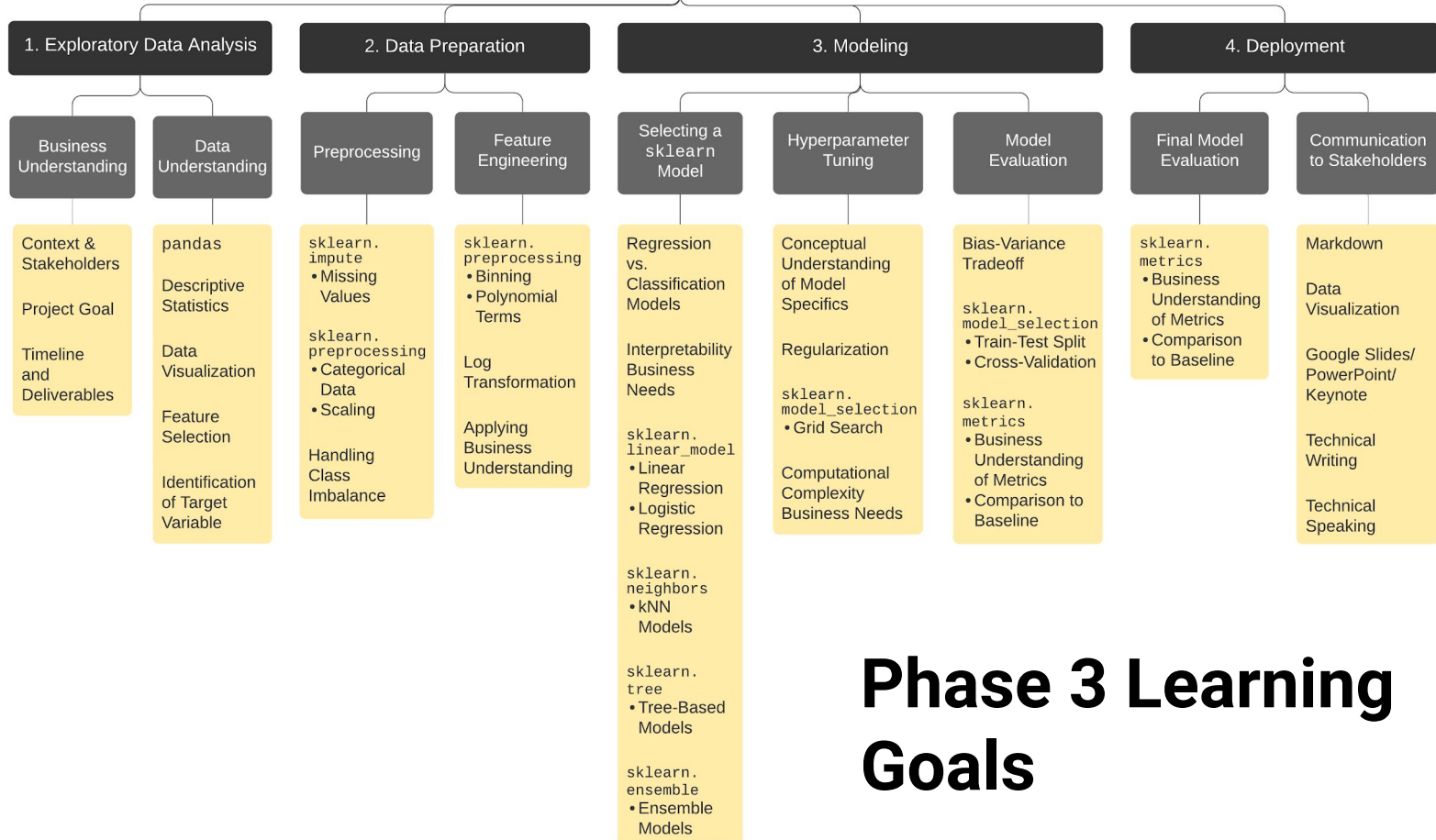 - ii. HAC

Supervised Techniques of Phase 4:

- ○ Recommendation Systems*

- ○ Time Series forecasting

- ○ Natural Language Processing for classification

- ○ Neural Networks for image classification

More semi-supervised than fully supervised

Develop a Supervised Machine Learning Model using Tabular Data and `sklearn`

1. Exploratory Data Analysis

Business Understanding
- Context & Stakeholders
- Project Goal
- Timeline and Deliverables

Data Understanding
- `pandas`
- Descriptive Statistics
- Data Visualization
- Feature Selection
- Identification of Target Variable

2. Data Preparation

Preprocessing
```
sklearn.
impute
```
- Missing Values

```
sklearn.
preprocessing
```
- Categorical Data
- Scaling
- Handling Class Imbalance

Feature Engineering
```
sklearn.
preprocessing
```
- Binning
- Polynomial Terms
- Log Transformation
- Applying Business Understanding

3. Modeling

Selecting a `sklearn` Model
- Regression vs. Classification Models
- Interpretability Business Needs

```
sklearn.
linear_model
```
- Linear Regression
- Logistic Regression

```
sklearn.
neighbors
```
- kNN Models

```
sklearn.
tree
```
- Tree-Based Models

```
sklearn.
ensemble
```
- Ensemble Models

Hyperparameter Tuning
- Conceptual Understanding of Model Specifics
- Regularization

```
sklearn.
model_selection
```
- Grid Search
- Computational Complexity Business Needs

Model Evaluation
- Bias-Variance Tradeoff

```
sklearn.
model_selection
```
- Train-Test Split
- Cross-Validation

```
sklearn.
metrics
```
- Business Understanding of Metrics
- Comparison to Baseline

4. Deployment

Final Model Evaluation
```
sklearn.
metrics
```
- Business Understanding of Metrics
- Comparison to Baseline

Communication to Stakeholders
- Markdown
- Data Visualization
- Google Slides/ PowerPoint/ Keynote
- Technical Writing
- Technical Speaking

Phase 3 Learning Goals

Develop a Machine Learning Model

1. Exploratory Data Analysis

Business Understanding
- Context & Stakeholders
- Project Goal
- Timeline and Deliverables

Data Understanding
- pandas
- Descriptive Statistics
- Data Visualization
- Feature Selection
- Identification of Target Variable
- **Unsupervised Learning**
- **Big Data**

2. Data Preparation

Preprocessing
- Missing Values
- Categorical Data
- Scaling
- Handling Class Imbalance
- **Time Series Data**
- **Text Data**
- **Image Data**

Feature Engineering
- Binning
- Polynomial Terms
- Log Transformation
- Applying Business Understanding
- **Time Series Features**
- **Text and Document Features**
- **Image Features**

3. Modeling

Selecting a Model
- Regression vs. Classification Models
- Interpretability Business Needs
- sklearn Models
  - Linear
  - Neighbors
  - Tree
  - Ensemble
  - **Naive Bayes**
- **Recommender Models**
  - surprise
  - pyspark.ml
- **Time Series Models**
  - statsmodels
  - fbprophet
- **Neural Network Models**
  - tf.keras

Hyperparameter Tuning
- Conceptual Understanding of Model Specifics
- Regularization
- Grid Search
- Computational Complexity Business Needs

Model Evaluation
- Bias-Variance Tradeoff
- Train-Test Split
- Cross-Validation
- Business Understanding of Metrics
- Comparison to Baseline

4. Deployment

Final Model Evaluation
- Business Understanding of Metrics
- Comparison to Baseline

Communication to Stakeholders
- Markdown
- Data Visualization
- Google Slides/ PowerPoint/ Keynote
- Technical Writing
- Technical Speaking

Phase 4 Learning Goals

Measuring Student Progress

1. **Blog Post**
    a. Due the **Tuesday** of the Second Week (May, 2nd)
    b. Topic: summarize a research paper/case study
        i. Alternative: write a tutorial (follow phase 3 prompt)
2. **Canvas & Code Challenge**
    a. Checkpoints
        i. Wed 26th: Unsupervised ML CP
        ii. Friday 28th: Time Series CP
        iii. Wed 3rd: NLP CP
        iv. Fri 5th: Tensorflow CP
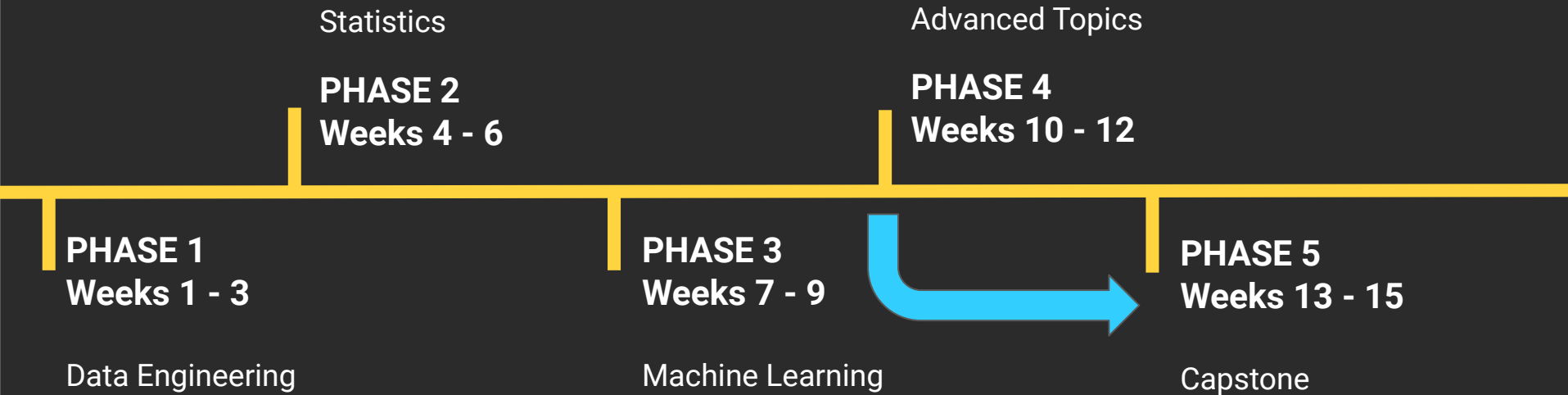   b. Code Challenge
        i. Thur 4th
3. **Project**
    a. 4th Group project - advanced topics

Phase 4

Pick a research paper/case study about data science and write a blog post to explain it to a non-technical business stakeholder. This can be a new paper from the past few months, or you can refer to this spreadsheet for a list of historical papers. Potential elements to include:

I. What was the original context of this paper when it was written?

II. Summary of the paper findings/outcomes

III. How can this paper inform your work as a junior data scientist?

IV. Why is this paper important/why does it matter to a non-technical business stakeholder?

Capstone Brainstorming

Answer the following questions for three (3) capstone ideas:
- What real-world problem would your project aim to solve?
- What data do you plan to use and what would be your target variable(s)
- What methods would you use to model the data, and what would be the final product(s)?

- Due **NEXT MONDAY** (May 1st)

- Why three? So you have backup options!

- Schedule 1:1 Check Ins in Weeks 1 & 2 & 3 to connect and discuss

- Full Capstone proposal due Day 1 of Phase 5 (May 15th)

Peeking Into Phase 5:

- **FSM**: Due Friday, Week 1 (May 19th)

- **MVP**: Due Friday, Week 2 (May 26th)

  - First chance to pass graduation requirements!

- **MEMORIAL DAY**: Monday, May 29th

- **CAPSTONE DUE**: Due Thursday, June 1st

  - Final presentation that afternoon

- **GRADUATION:** Friday, June 2nd

Feelings at the Start of Phase 4

Confident about what's going on...kind of

Frustrated by ongoing "beginner" mistakes

Overwhelmed by studying new concepts while fleshing out capstone.

Feelings at the End of Phase 4

Cautiously optimistic when things start to "click"

Struck by the reality of the bootcamp ending

Prepared to get started on capstone

Let's Do This