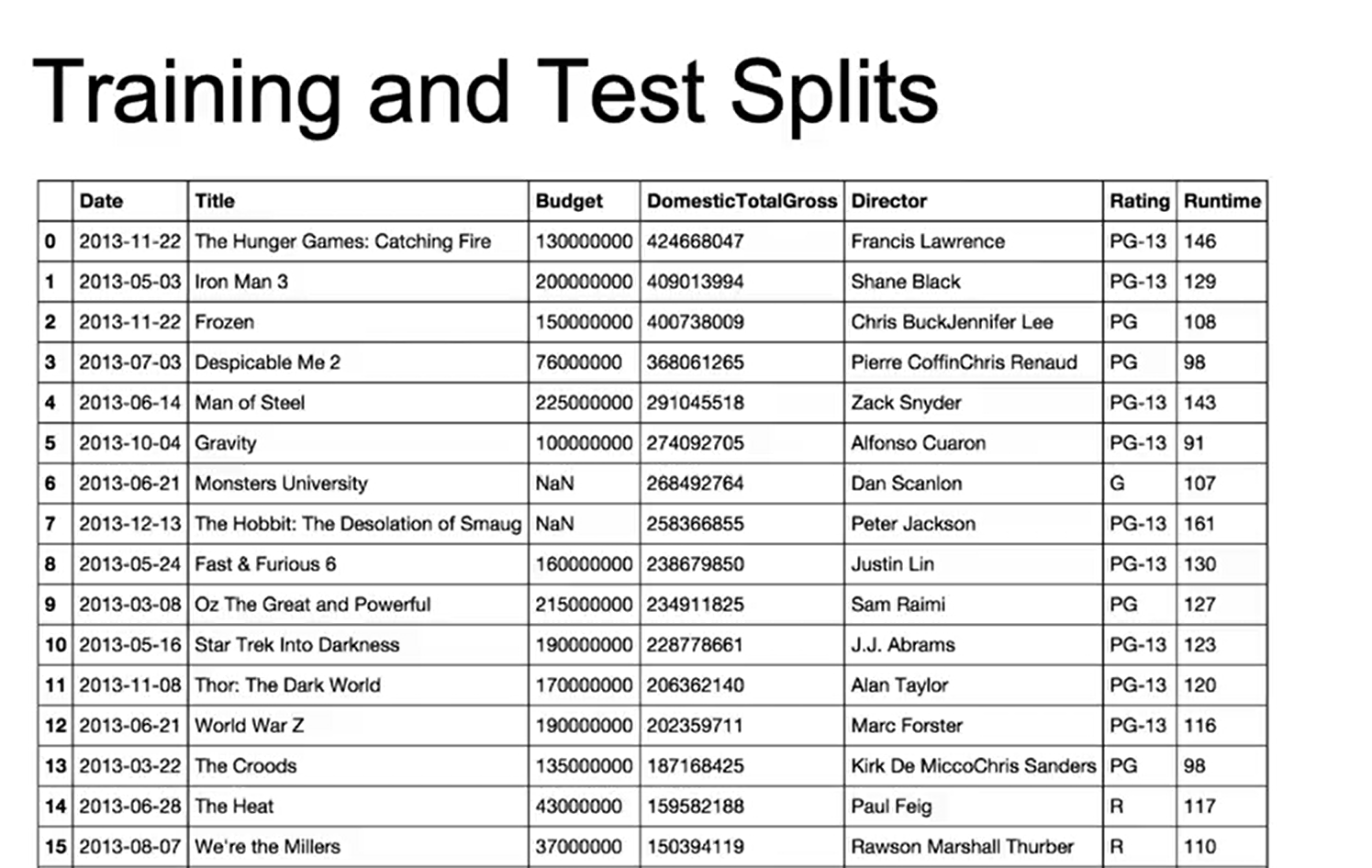
**MOOC 2-MODULE 2**

[**Supervised Machine Learning: Regression**](https://www.coursera.org/learn/supervised-machine-learning-regression/home/welcome)

**I.Training and Test Splits**

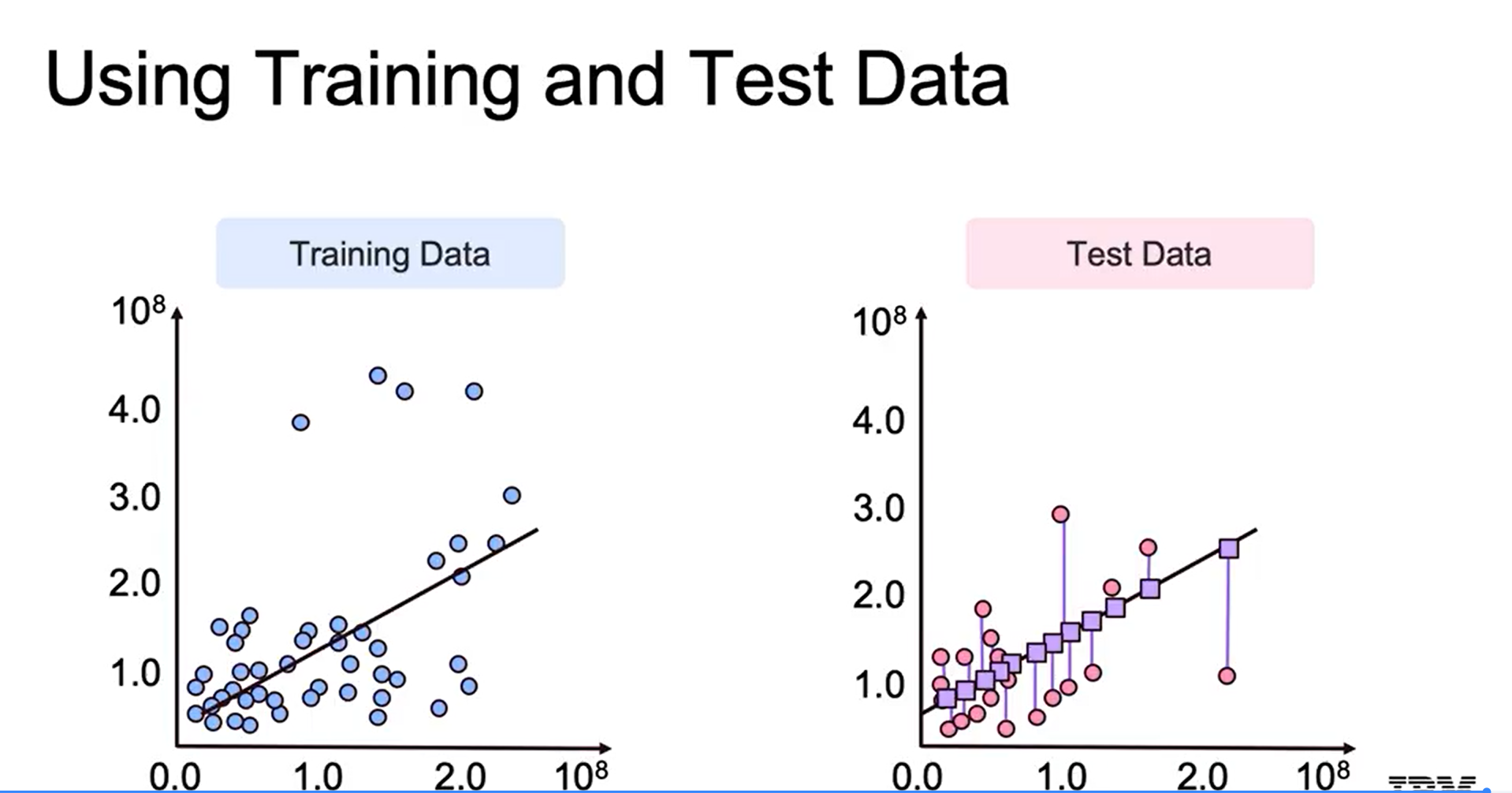
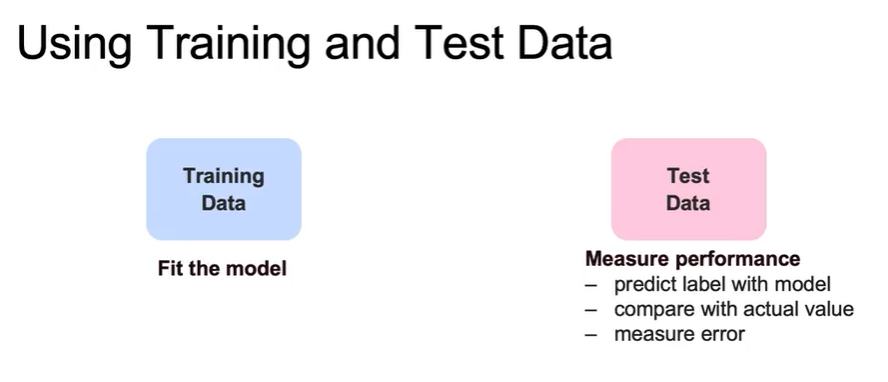
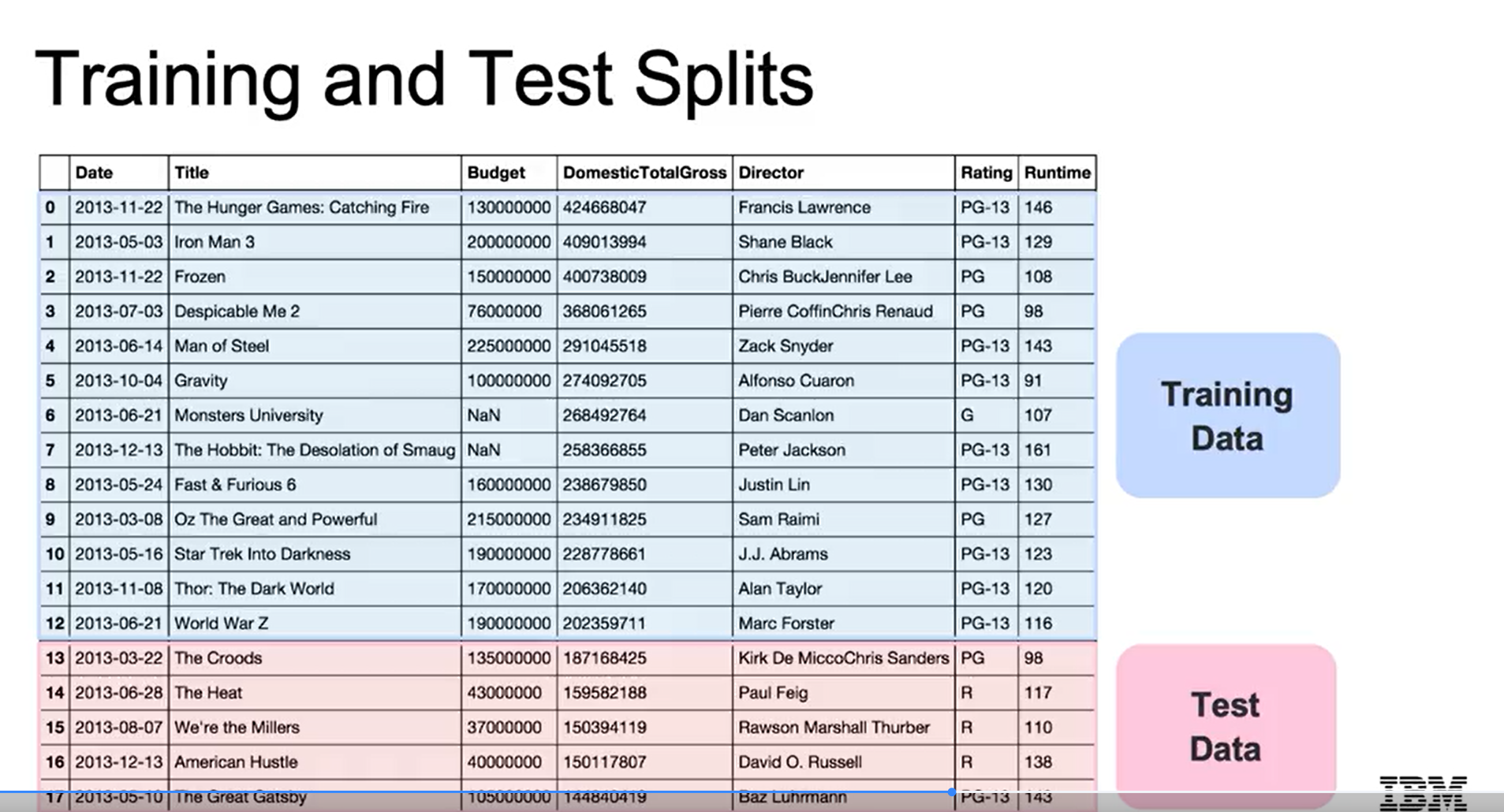
### Data Splitting

* Divide dataset into training set and testing set.
* Purpose: Evaluate how well the model performs on unseen data.
* A holdout set ensures proper assessment of model generalization.



### Cross-Validation

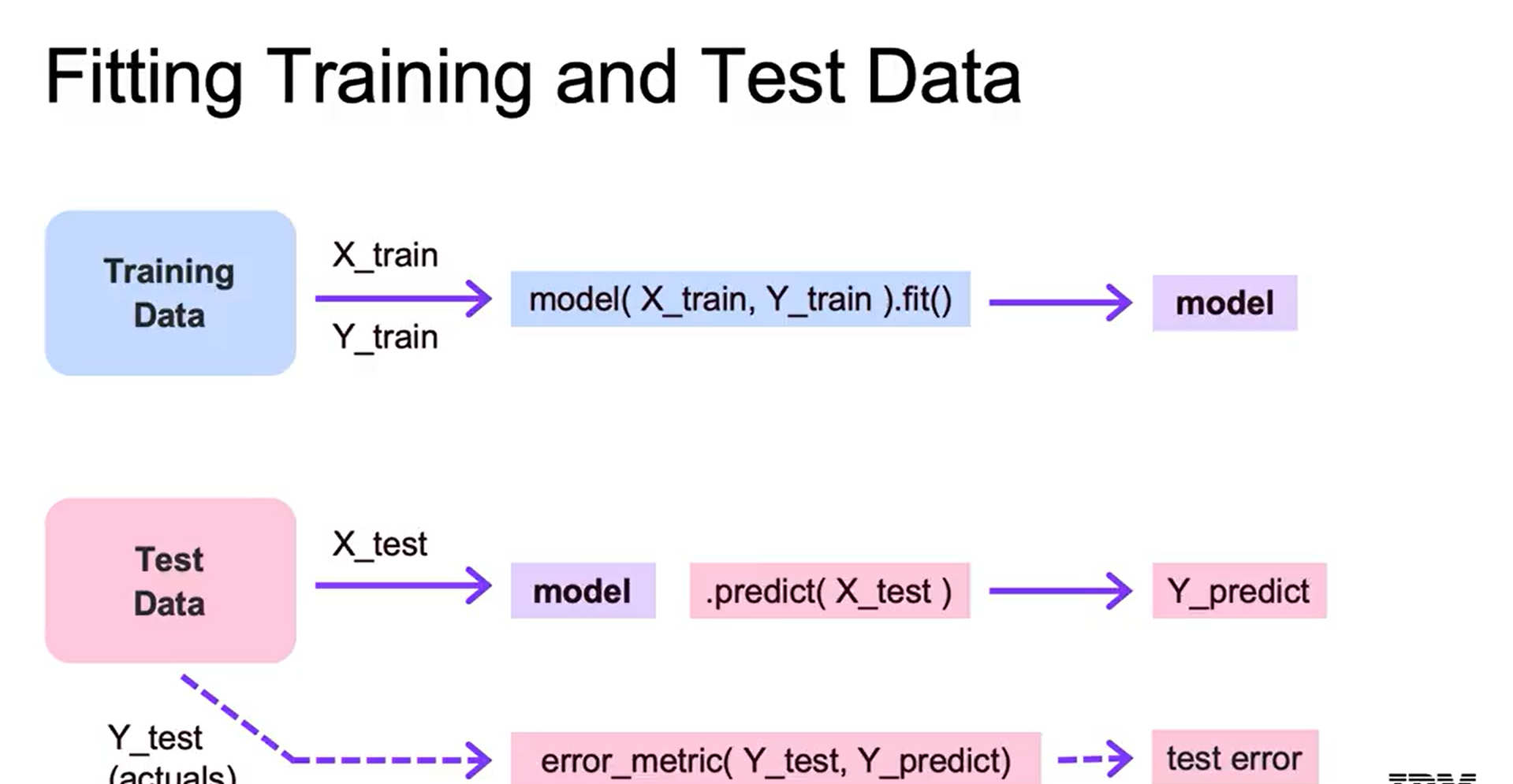
* Splits data into multiple subsets for repeated training and testing.
* Provides a more robust estimate of model performance.
* Reduces bias from relying on a single train-test split.



### Model Complexity and Error

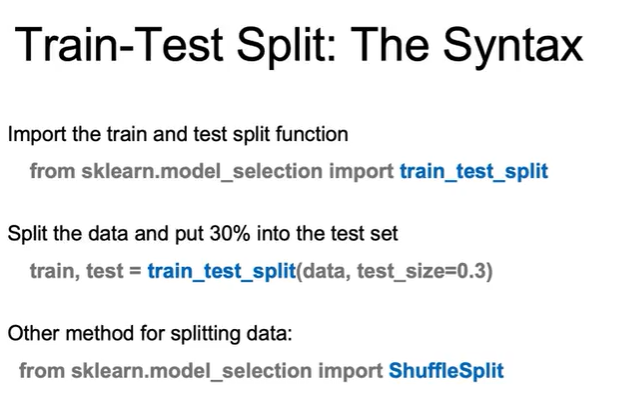
* Bias-Variance Tradeoff:  
  + Too simple → high bias, underfitting.
  + Too complex → high variance, overfitting.
* Goal: Find the optimal complexity that balances training and test error.

## **Training Data and Model Fitting**

* **Training data**:  
  
* The model is **fitted** on training data to learn the parameters that define the relationship between features and the outcome variable.

## Data Splitting Techniques

* Train-Test Split (scikit-learn):  
  + Splits dataset into training and testing sets.
  + Option to specify percentage of test set (e.g., 20%).
* Shuffle Split:  
  + Creates multiple random splits for more robust evaluation.
* Stratified Shuffle Split:  
  + Preserves the distribution of outcome variables (important in classification problems).



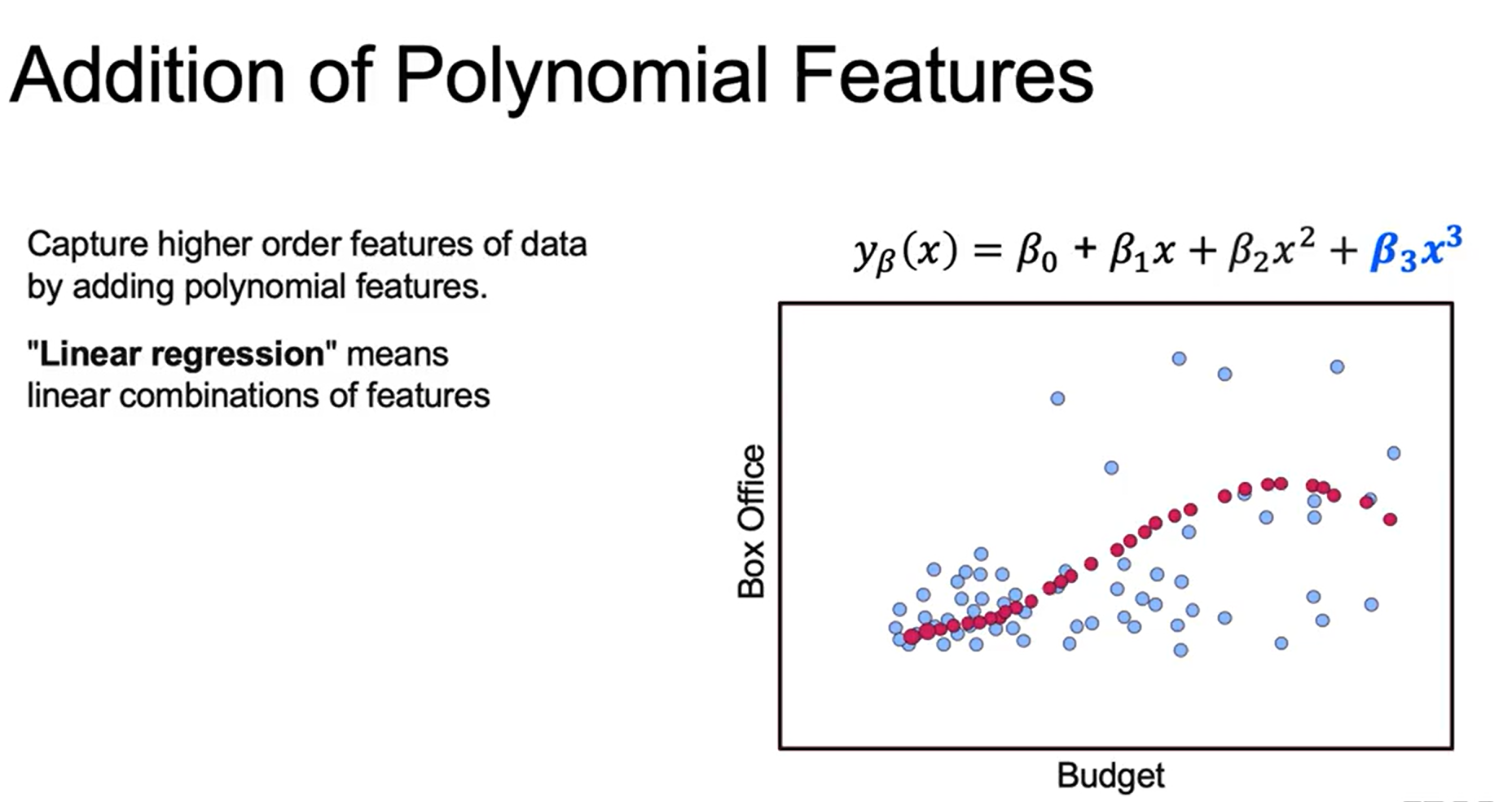
**II.Polynomial Regression**

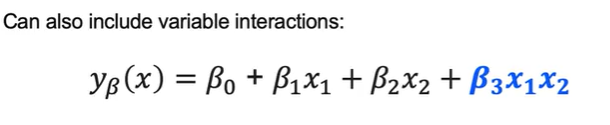
Extending Linear Regression

* Discusses the importance of adjusting linear regression to create more complex models for better predictions.
* Introduces polynomial features to capture nonlinear relationships while still using linear regression.

Feature Engineering

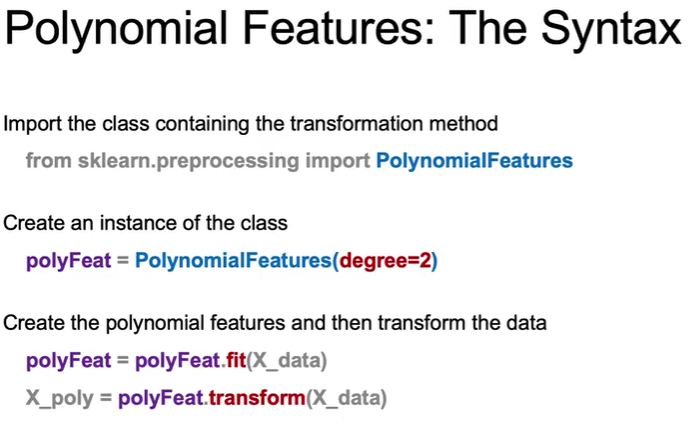
* Explains how to create new features from existing ones, such as interaction terms and polynomial transformations.
* Emphasizes the need for feature selection techniques to identify which features positively impact model performance.





Model Evaluation Framework

* Highlights the importance of understanding the bias-variance trade-off in model evaluation.
* Mentions that the concepts learned will apply to various regression and classification models throughout the course.

****