



Introduction to regression models using R and Tidymodels

Follow along with these modules at
<https://aka.ms/learnlive-20220909A>



Join the chat at
<https://aka.ms/LearnLiveTV>



Eric Wanjau

Data Scientist and Researcher
Leeds Institute for Data Analytics
University of Leeds
@ericntay



Karani Wachira

Undergrad student
DeKUT R Community Lead
Dedan Kimathi University of Technology
@mrAfricanSeed

Say hi to our
moderator



Pascal Burume

Network and telecommunications
engineer, Microsoft Student
Ambassador

Youth Africa Technologies


LinkedIn: [/in/pascal-burume-909520218/](https://www.linkedin.com/in/pascal-burume-909520218/)

/learn

<https://aka.ms/learnlive-20220909A>

Complete interactive learning exercises, watch videos, and practice and apply your new skills.





Introduction to regression models by using R and tidymodels

1 hr 23 min • Module • 10 Units

★★★★★ 4.9 (12)

Beginner Developer Data Scientist Student Azure

Get an introduction to regression models. In machine learning, the goal of regression is to create a model that can predict a numeric, quantifiable value.

Learning objectives

In this module, you'll learn:

- When to use regression models.
- How to train and evaluate regression models by using the tidymodels framework.

[Start >](#) [⊕ Save](#)

1100 XP

Learning objectives



When to use regression models.



How to train and evaluate regression models by using the tidymodels framework.

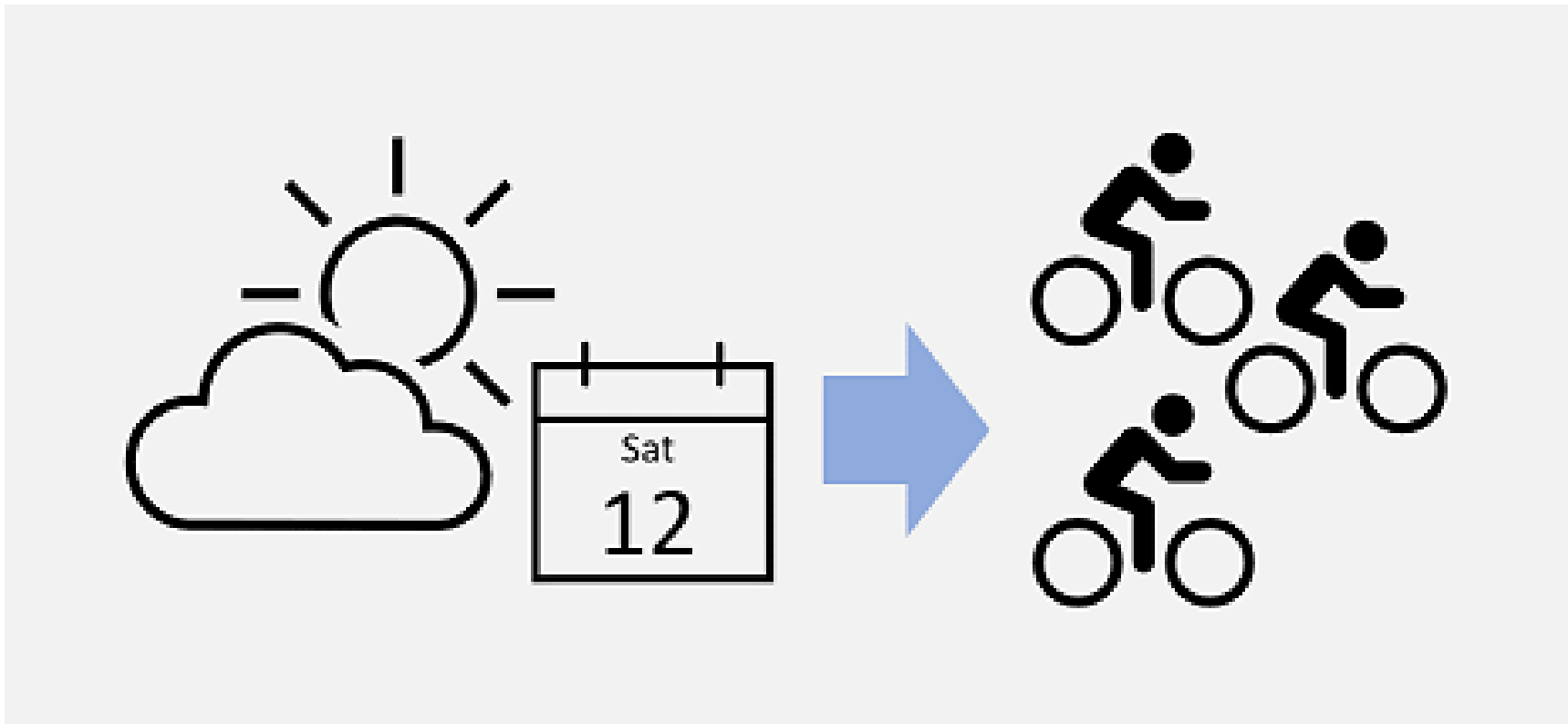
What is regression?

What is regression?

Supervised ML technique that works by establishing a relationship between variables in the data to predict a numeric, quantifiable value.

- **A training dataset.** You'll apply an algorithm that determines a function that encapsulates the relationship between the feature values and the known label values.
- **A validation or test dataset.** You can use it to evaluate the model by using it to generate predictions for the label and comparing them to the actual known label values.

What is regression: Predicting bike rentals





EDA



Data Budgeting

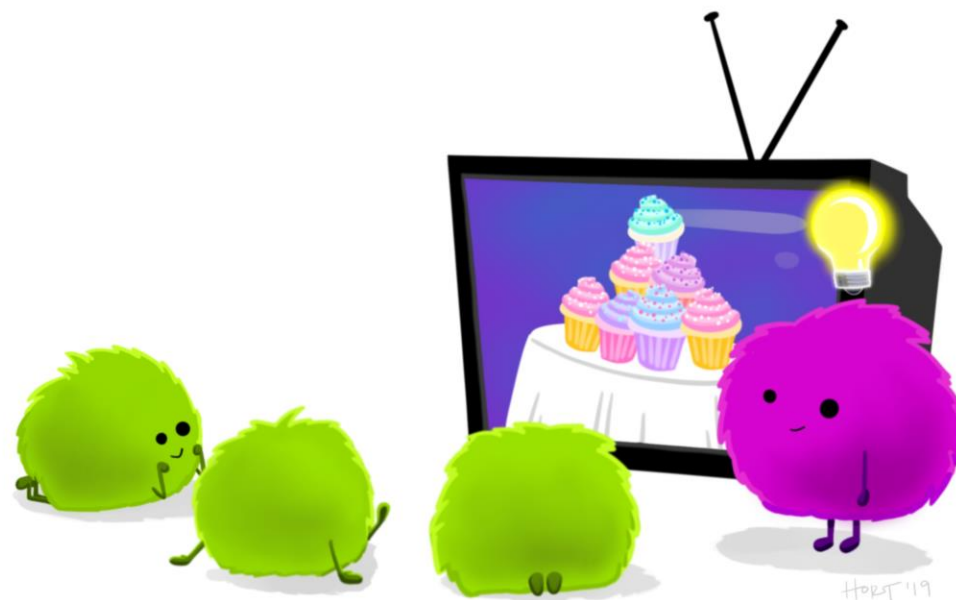


Fit and evaluate a linear regression model



Can we do better?

The plan: Predicting bike rentals

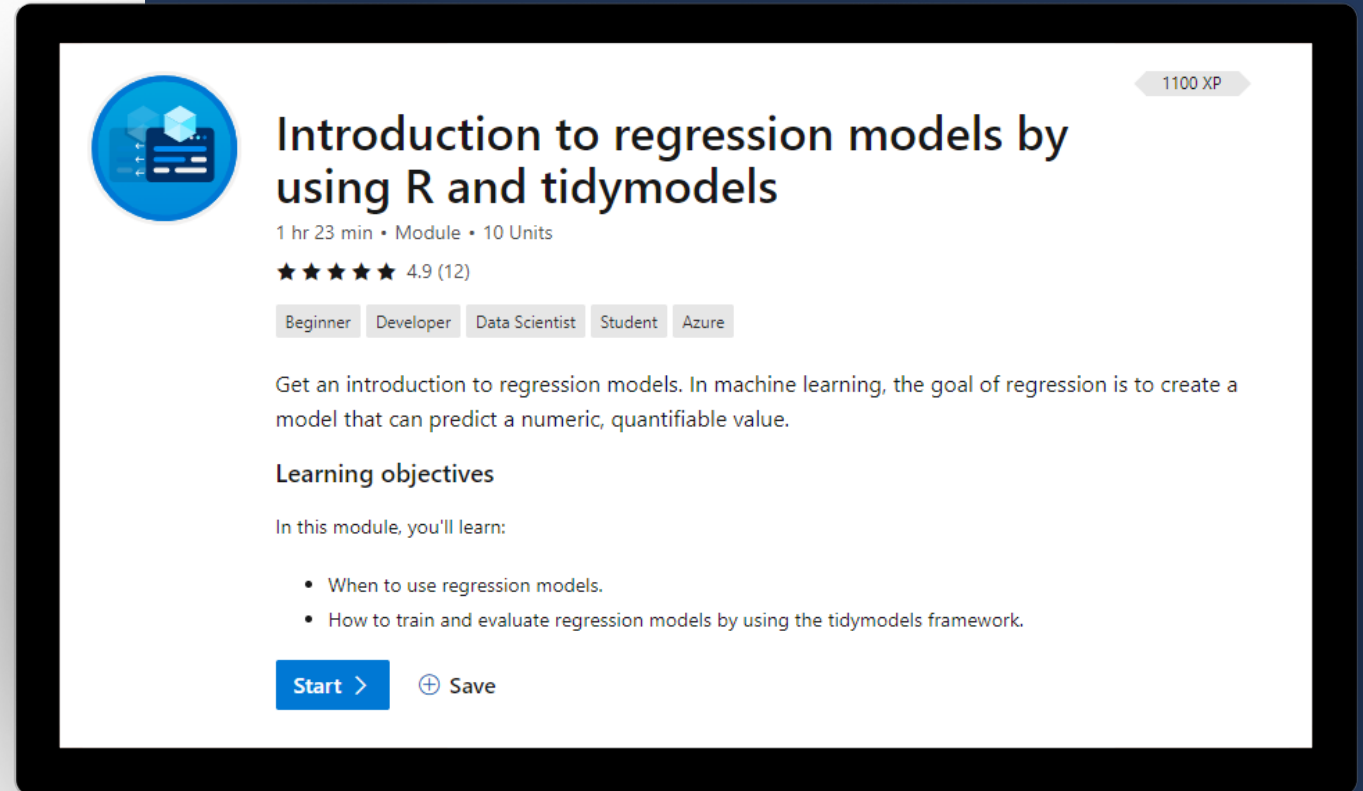


artwork by @allison_horst

/learn

<https://aka.ms/learnlive-20220909A>

Complete interactive learning exercises, watch videos, and practice and apply your new skills.



The image shows a digital tablet displaying a learning module card. The card has a blue header with a circular icon containing a document and a blue cube. The title 'Introduction to regression models by using R and tidymodels' is in bold black text. Below the title, it says '1 hr 23 min • Module • 10 Units' and '★★★★★ 4.9 (12)'. There are five tags: 'Beginner', 'Developer', 'Data Scientist', 'Student', and 'Azure'. The description reads: 'Get an introduction to regression models. In machine learning, the goal of regression is to create a model that can predict a numeric, quantifiable value.' Under 'Learning objectives', it lists: 'When to use regression models.' and 'How to train and evaluate regression models by using the tidymodels framework.' At the bottom, there is a blue 'Start >' button and a grey '+ Save' button. In the top right corner of the card, there is a grey badge that says '1100 XP'.

Introduction to regression models by using R and tidymodels

1 hr 23 min • Module • 10 Units

★★★★★ 4.9 (12)

Beginner Developer Data Scientist Student Azure

Get an introduction to regression models. In machine learning, the goal of regression is to create a model that can predict a numeric, quantifiable value.

Learning objectives

In this module, you'll learn:

- When to use regression models.
- How to train and evaluate regression models by using the tidymodels framework.

Start > **+ Save**

1100 XP

Knowledge Check



Test your knowledge in the chat



The tidymodels framework was used in R to train a regression model from a dataset of sales data. To evaluate the model to ensure it will predict accurately with new data, what should be done?

Question 1

Vote at



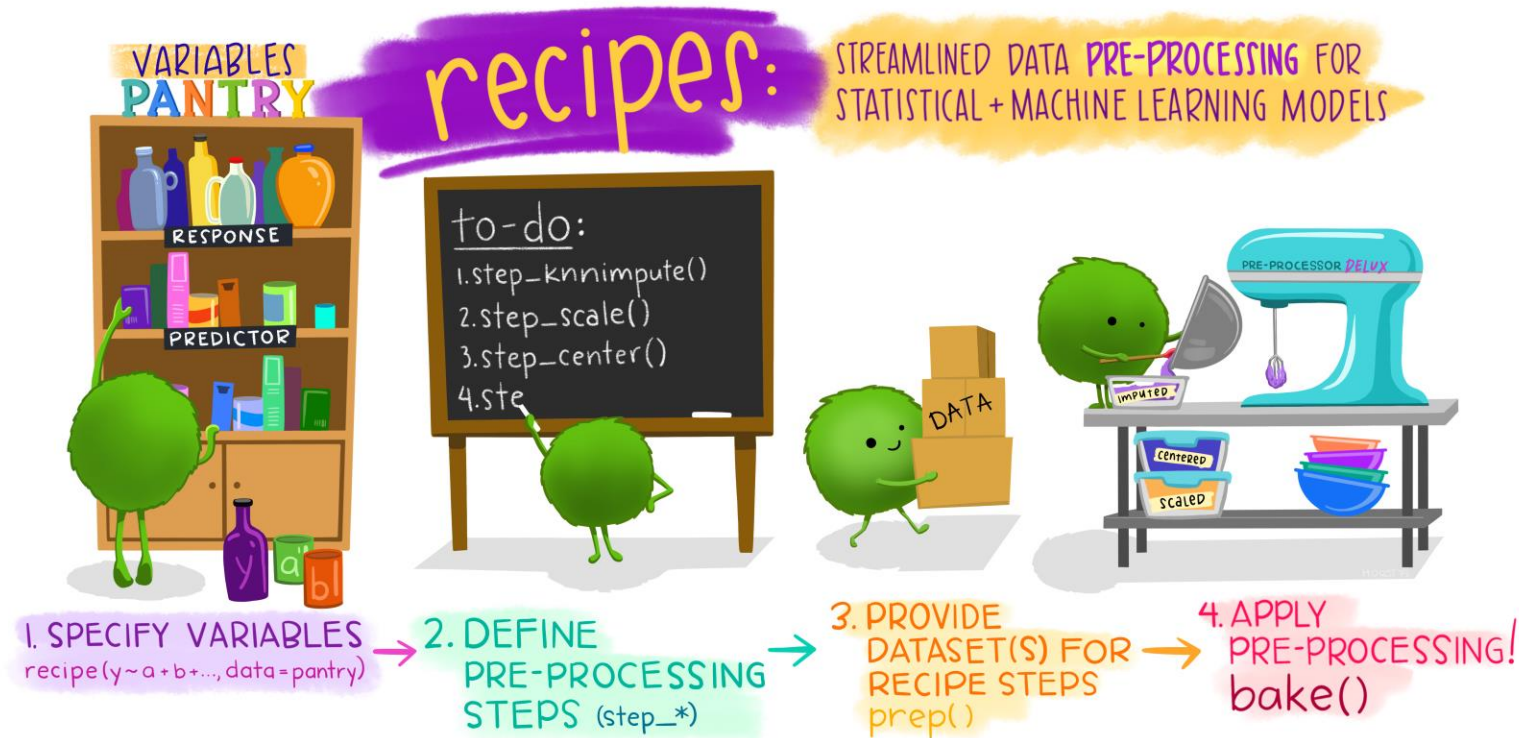
- A. Split the data randomly into two subsets. Use one subset to train the model and the other subset to evaluate it.
- B. Use all the data to train the model. Then use all the data to evaluate it.
- C. Train the model by using only the feature columns. Then evaluate it by using only the label column.

Vote at <https://aka.ms/polls>



Can we do
better?

Feature Engineering with recipes



Normalize data

- step_normalize()

Impute missing values

- step_impute_mean()

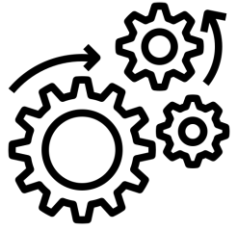
Try out a
different
model





Model Tuning

Some model parameters cannot be estimated directly from the training data.



Instead of learning these kinds of hyperparameters during model training, we tune them



Try different values



Measure performance



Pick the best



Let's pause for a sec...

- Typically we can't decide on which final model to use with the test set before first assessing model performance.

- There is a gap between our need to measure performance reliably and the data splits (training and testing) we have available

Resampling for evaluating model performance

Cross Validation

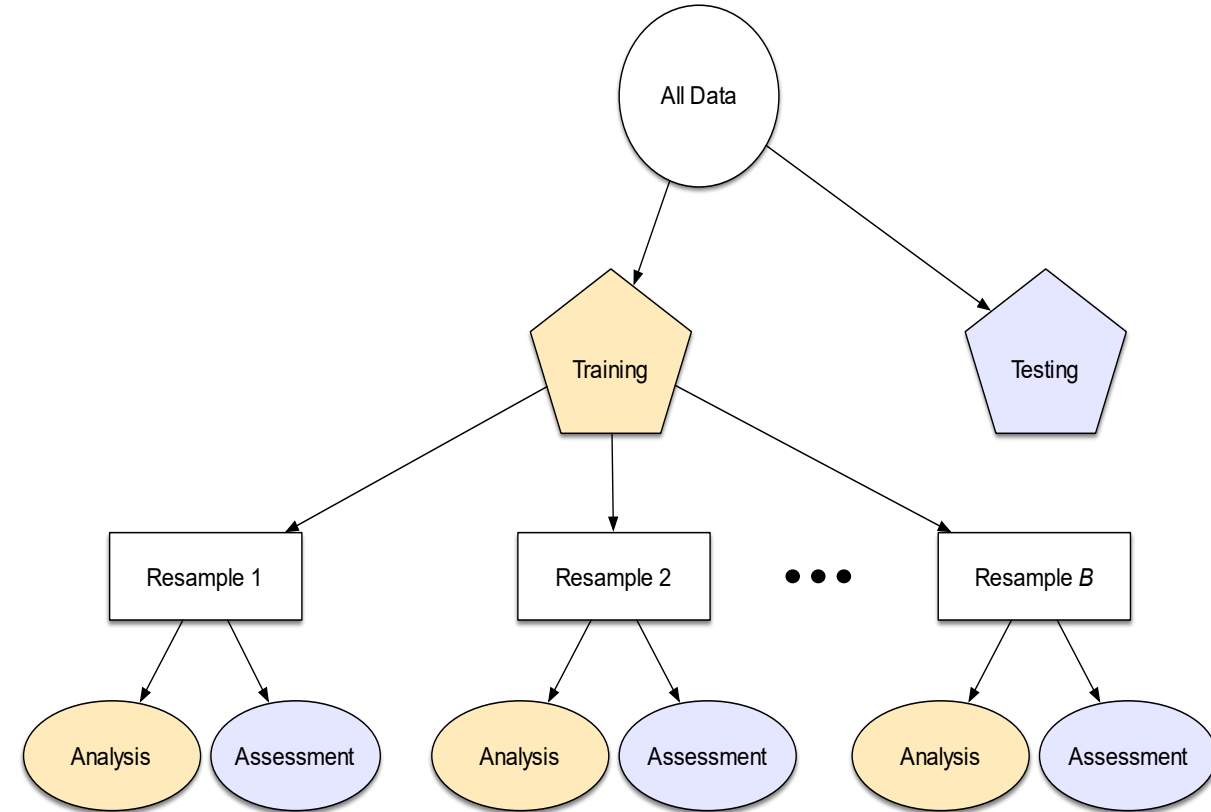


Image from Tidy Modeling with R

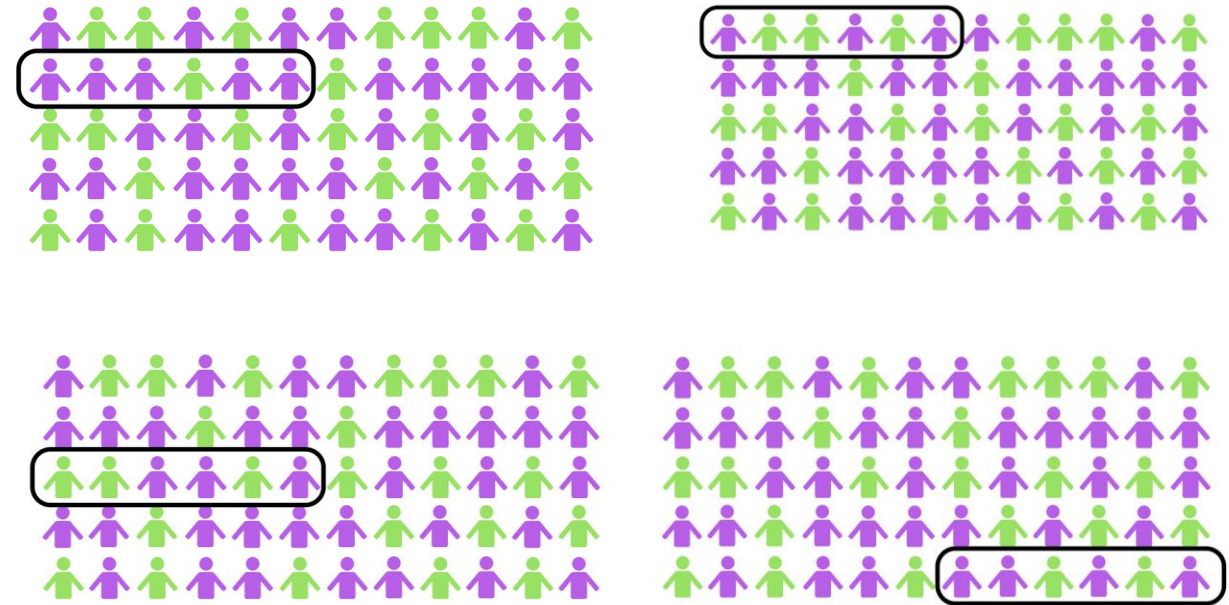


Image from Julia Silge's ML course

Knowledge Check



Test your knowledge in the chat

Question 2

Vote at



A regression model specification was created by using the `linear_reg()` function in the `tidymodels` `parsnip` package. What should be done to train the model?

- A. Call the `predict()` function and specify the model specification, formula, and data.
- B. Call the `recipe()` function and specify the model specification, formula, and data.
- C. Call the `fit()` function and specify the model specification, formula, and data.

Vote at <https://aka.ms/polls>



A regression model was trained by using the tidymodels framework. When it's evaluated with test data, the model achieves an R-squared metric of 0.95. What does this metric say about the model?

- A. The model is 95% accurate.
- B. The model explains most of the variance between predicted and actual values.
- C. On average, predictions are 0.95 higher than actual values.

Question 3

Vote at



Vote at <https://aka.ms/polls>

Summary



When to use regression models.

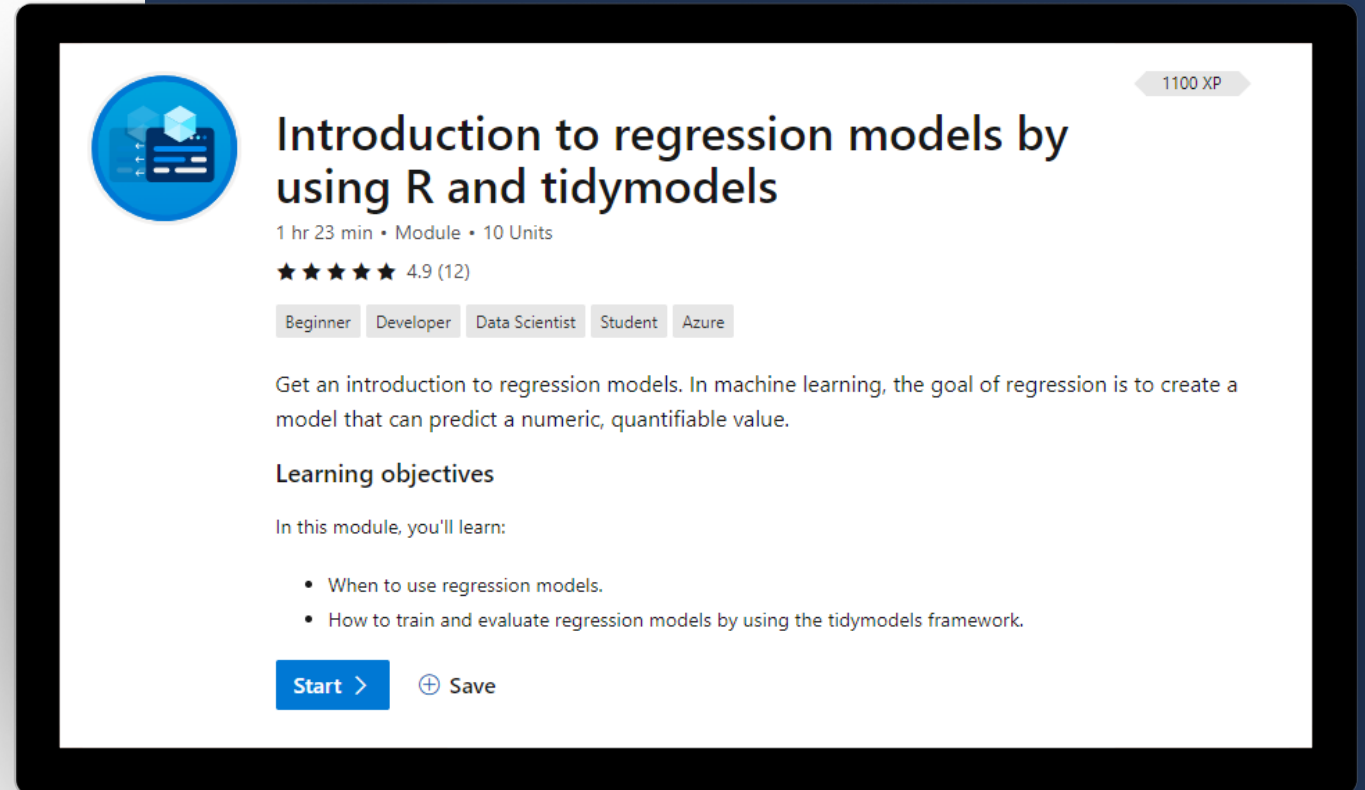


How to train and evaluate regression models by using the tidymodels framework.

/learn

<https://aka.ms/learnlive-20220909A>

Complete interactive learning exercises, watch videos, and practice and apply your new skills.



The image shows a digital tablet displaying a learning module card. The card has a blue circular icon on the left containing a server rack and a blue cube. To the right of the icon, the title 'Introduction to regression models by using R and tidymodels' is displayed in a large, bold, black font. Above the title, on the right side of the card, is a grey badge with the text '1100 XP'. Below the title, the text '1 hr 23 min • Module • 10 Units' is shown. Underneath that is a row of five stars, with the first four filled and the fifth half-filled, followed by the text '4.9 (12)'. Below the stars is a horizontal row of five grey buttons with rounded corners, labeled 'Beginner', 'Developer', 'Data Scientist', 'Student', and 'Azure'. The main body of the card contains a paragraph: 'Get an introduction to regression models. In machine learning, the goal of regression is to create a model that can predict a numeric, quantifiable value.' This is followed by a section header 'Learning objectives' and a sub-header 'In this module, you'll learn:'. Below this are two bullet points: '• When to use regression models.' and '• How to train and evaluate regression models by using the tidymodels framework.' At the bottom of the card, there is a blue button with the text 'Start >' and a grey button with a plus icon and the text 'Save'.

Introduction to regression models by using R and tidymodels

1 hr 23 min • Module • 10 Units

★★★★★ 4.9 (12)

Beginner Developer Data Scientist Student Azure

Get an introduction to regression models. In machine learning, the goal of regression is to create a model that can predict a numeric, quantifiable value.

Learning objectives

In this module, you'll learn:

- When to use regression models.
- How to train and evaluate regression models by using the tidymodels framework.

Start > ⊕ Save

Don't miss!


Introduction to classification models by using R and tidymodels

September 16, 2022 4:00PM - 5:30PM (Central Europe)

Continue your learning with the next module in this series!!

<https://aka.ms/learnlive-machine-learning-r-tidymodels-Ep3>





Introduction to classification models by using R and tidymodels

1 hr 22 min • Module • 10 Units

★★★★★ 4.7 (10)

Beginner Developer Data Scientist Student Azure

Classification is a form of machine learning in which you train a classification model to predict which category an item belongs to. In this module, you learn how to use the R programming language and tidymodels framework to train classification models.

Learning objectives

In this module, you'll learn:

- When to use classification.
- How to train and evaluate a classification model by using the tidymodels framework.

[Start >](#) [+ Save](#)

1100 XP