

Lesson 12: Numerical Data in AI Lab

45 minutes

Overview

In this lesson, students will be introduced to numerical data which represents a range of values. Students are presented with a scenario where every feature and label is represented with numerical data, and they learn to use the new data visualization tools within AI Lab to help find patterns.

Question of the Day: How can we use AI Lab to predict numerical data?

Standards

Full Course Alignment

AI4K12 National Guidelines 2021

- ▶ **BI-3** - Computers can learn from data

CSTA K-12 Computer Science Standards (2017)

- ▶ **AP** - Algorithms & Programming
- ▶ **DA** - Data & Analysis

Agenda

Warm Up (5 minutes)

Journal

Activity (35 minutes)

Categorical vs Numerical

Investigating Data

Feature #1: Antelopes

Feature #2: People

Feature #3: Day of the Month

Feature #4: Temperature

Training a Model

Wrap Up (5 minutes)

Journal

Objectives

Students will be able to:

- Compare and contrast categorical data versus numerical data
- Use data visualizations to find patterns in numerical data

Preparation

- Check the "**Teacher's Lounge**" forum for verified teachers to find additional strategies or resources shared by fellow teachers

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the teachers

- **Numerical Data in AI Lab** - Slides

▼ Make a Copy

For the students

- **Numerical Data in AI Lab** - Activity Guide ▼ Make a Copy
- **Numerical Data in AI Lab** - Video
- **Using Data with Numerical Features** - Resource

Vocabulary

- **Numerical Data** - data that can be counted or measured

Teaching Guide

Warm Up (5 minutes)

Journal

 **Prompt:** *Brianna and Mikayla are movie critics who use different systems to recommend movies. Brianna recommends movies as either “Go see it!” Or “Don’t bother”. Mikayla recommends movies on a scale from 1-10, such as 7.2 or 4.1. How are these two systems similar? How are they different?*

Have students journal individually first, then share with a partner before inviting students to a full-class discussion.

Discussion Goal: Students may notice that you can tell “good movies” from “bad movies” in both systems, but the scale system allows a lot more possible values and it’s easier to compare movies based on their values. Students may notice that Brianna’s system uses categorical data with two categories, but they may struggle to define Mikayla’s system and initially say it is also categorical but with many more categories.

Remarks

Both of these systems help us make decisions, but in different ways. Brianna’s system looks similar to the type of recommendations we’ve seen so far because it simplifies the data into just two categories. Mikayla’s system is a little different, using a range of values to make a recommendation - this is called numerical data. Today, we’ll learn how to use numerical data in AI Lab to make new kinds of recommendations and predictions.

 **Question of the Day:** How can we use AI Lab to predict numerical data?

Activity (35 minutes)

Categorical vs Numerical (25 minutes)

 **Vocabulary:**

- **Numerical Data:** data that can be counted or measured

 **Discuss:** *What are other examples of Numerical Data?*

Have students discuss with a partner before having several students share with the full group. Keep a list of responses at the front of the room

Discussion Goal: If students seem stuck on coming up with other examples, remind them that we sometimes apply categories in order to simplify our data. For example, “tall” and “short” are simplified categories for something we could represent numerically - our height.

A few examples students may think of:

- Age, Weight, Height
- Cost, or anything money related
- Rating systems, similar to the warm-up
- Anytime you are counting “how much” of something

Remarks

This is a great list! I can tell that there’s numerical data all around us! Today, we’re going to take a look at numerical data in AI Lab and how we can use it to train a model.

Investigating Data

 **Video:** Show the [Numerical Data in AI Lab](#) video, which outlines how numerical data can be used in AI Lab and how accuracy is calculated.

Teaching Tip

To encourage active engagement and reflection, use one or more of the strategies discussed in the [Guide to Curriculum Videos](#).

 **Code Studio:** Have students log into Code Studio and open the first level. Students will spend most of the lesson exploring data in the first panel and recording their results on the activity guide.

 **Distribute:** Pass out [Numerical Data in AI Lab](#) to each student.



1

Safari Model

Feature #1: Antelopes

 **Do This:** Have students click on the antelopes column. This column represents how many antelopes were seen in the park on a given day. Look at the graph that appears in AI Lab, which lets you compare the antelope data with the lion data

Teaching Tip

Model Reading Graphs: It's a good idea to model the first graph with the class and fill in the activity guide together, especially since students are still learning how to interpret graphs in their other classes. As students practice this skill, the goal is to become confident identifying relationships in data and discerning if a pattern really exists, or if the data has a random relationship that isn't good for predictions.

Discuss:

- *If there are a low number of antelopes in the park, what does that mean for how many lions could be in the park?*
- *If there are a high number of antelopes in the park, what does that mean for how many lions could be in the park?*
- *Why do you think this is?*

Discussion Goal: Students should notice the less antelopes there are, the less lions there are and vice versa. They may imagine this has to do with predator / prey relationships - the lions eat the antelope, so if there are less antelope, there are less lions. Students should record their responses on their activity guide, even if the class discussed the answers together.

Feature #2: People

 **Do This:** Have students click on the people column. This column represents how many people were seen in the park on a given day. Have students answer the questions on their activity guide first before discussing as a group.

Circulate: Check in with students and help them interpret the graphs. Encourage students to use sentence starters like "When the number of people are... then the number of lions are...".

Discuss:

- *If there are a low number of people in the park, what does that mean for how many lions could be in the park?*
- *If there are a high number of people in the park, what does that mean for how many lions could be in the park?*
- *Why do you think this is?*

Content Corner

Associations: The antelope graph represents a positive association and the people graph represents a negative association. which are a part of the **Common Core 8th grade math standards**. Students don't need to know these terms to be successful in this unit, so we do not recommend using this vocabulary unless it directly supports their study in other classes.

Discussion Goal: Students should notice the less people there are, the more lions there are and vice versa. They may imagine this has to do with natural behavior and they may think back to their own experiences visiting zoos or wildlife - if there are more strangers in the park, they are less likely to come out. The opposite may also be true - the less people around, the more they may roam free.

Feature #3: Day of the Month

 **Do This:** Have students click on the dayOfMonth column. This column represents what day of the month you went to the park. Have students answer the questions in their activity guide

Discuss:

- *What happens if you visit on a day early in the month? How many lions do you think you'll see?*
- *What happens if you visit on a day late in the month? How many lions do you think you'll see?*

Discussion Goal: Students may get a little stumped with this one because there is no pattern in this data. You might see a lot of lions early in the month and you might also see a lot of lions late in the month, and vice versa. Students may imagine this is because the day of the month doesn't change the lions behaviors, especially compared to some of the other features.

Feature #4: Temperature

 **Do This:** Have students click on the temperature column. This column represents the weather that day and how hot or cold it was. Have students answer the questions on their activity guide before discussing as a class.

Discuss:

- *If there are a low temperature in the park, what does that mean for how many lions could be in the park?*
- *If there are a high temperature in the park, what does that mean for how many lions could be in the park?*
- *Why do you think this is?*

Discussion Goal: Students should notice that both high temperatures and low temperatures mean you won't see very many lions. Instead, midrange temperatures lead you to seeing a lot of lions. This may be because lions won't come out in extreme temperatures and instead prefer nicer weather. The same can also be said for human beings - we avoid extreme weather.

Remarks

We can use AI Lab to train a machine learning model to predict how many lions we'll see when visiting the park. We want to make sure we use features that have a relationship with our lions. Based on the ones we've seen so far, which column would not make a good feature?

 **Discuss:** Which graph would not make a good feature?

Discussion Goal: Students should explain that the dayOfMonth column is not a good candidate because the data appears random. Instead, the other columns have a relationship with the label that they can describe, almost like a story within the data.

Training a Model (10 minutes)

 **Do This:** Continue to explore the data by clicking on the remaining features in the dataset. Record your observations on your activity guide.

Circulate: Check in with students as they explore data, making sure to check with any students who appeared to be struggling to read graphs during the previous exercises. Ask students to explain why they think certain columns could be good features.

 **Do This:** Using our investigation, train a model with 80% accuracy.

Teaching Tip

80% Accuracy: Students may struggle to find a model that is at least 80% accurate. This is by design, so they can really experiment with which features to use in their model. One example that will satisfy these requirements is a model using the features [trees, overgrowthPercent, antelopes, temperature].

Order Matters: Students may discover that the order they select their features can sometimes matter - for example, choosing "temperature, antelopes" may end up with different accuracy than "antelopes, temperature". This is not a vital topic for using AI Lab fluently, and happens more often in these early levels because of the smaller dataset sizes. If students ask, one way to think about it is that the first feature represents how AI Bot first tries to separate the data before continuing on to the other features. Therefore: the stronger the relationship is with the first feature you pick, the stronger the patterns AI Bot will find.

Assessment Opportunity

Formative Assessment: Because this level requires 80% accuracy to continue, completing this level can help determine how successful students are with the objectives from this lesson.

Code Studio: Students who finish training their model can import into App Lab and begin customizing their app. They won't have enough time to truly finish their app, but the next lesson focuses more on App Lab where they will be able to customize their apps more completely.

 2

Safari App

Wrap Up (5 minutes)

Journal

 **Prompt:** What is one way categorical data and numerical data are similar? What is one way they are different?

Discussion Goal: Student answers should feel similar to the definitions of these two terms. Both categorical and numerical data represent data, but categorical data can be separated into discrete categories while numerical data is represented along a continuum. Students may also provide examples of categorical or numerical data to help describe their answers.



This work is available under a **Creative Commons License (CC BY-NC-SA 4.0)**.

If you are interested in licensing Code.org materials for commercial purposes **contact us**.