

Lesson 18: Survey Data in AI Lab

45 minutes

Overview

This is the third in a five-day sequence of lessons that prepare students for the final project. In this lesson, students learn how to view survey data in Google Sheets and save the data to their computer as a csv file. Then, they upload the saved data to AI Lab and examine the survey results from one of the students to train a model using their data. Then, students use Google Sheets to examine data from another student where the data has errors and then try to fix the errors. The steps students take in this lesson are identical to the steps students will take in their final project, and the problem-solving strategies they develop will help them overcome challenges in their own final project.

Question of the Day: How can I import data into AI Lab to train a machine learning model?

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
- ▶ **IC** - Impacts of Computing

Agenda

Warm Up (5 minutes)

Survey

Activity (35 minutes)

Wrap Up (5 minutes)

Objectives

Students will be able to:

- Analyze and clean data in Google Sheets
- Download a csv file from Google Sheets
- Upload a csv file to AI Lab

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

- **Survey Data in AI Lab** - Slides

▼ Make a Copy

For the students

- **Cleaning Survey Data** - Video
- **Isaac's Survey Data - Student Resource**
- **Kim's Survey Data - Student Resource**
- **View and Download Survey Data** - Video

Teaching Guide

Warm Up (5 minutes)

Survey

 **Code Studio:** Have students log into Code Studio. The first level links to a survey that students can take.

 1

Kim's Survey

 **Prompt:** *What are two questions on this survey that you think will be related? You might use a sentence like "I think people who answer ... would probably want to join club ..."*

Discussion Goal: It's most important to get students into Code Studio and participating in the survey, since they will use the data from this survey in the lesson. The prompt helps prime students to look for patterns when they get into AI Lab. As students share responses, there are no right or wrong answers - instead, take note of the ideas that students have and be ready to reference them later when using AI Lab to analyze the data.

Remarks

Yesterday we helped our team create a survey - today we're going to look at the data and use AI Lab to start training a model. Kim has already been collecting data, so we'll start by learning how to import their data into AI Lab and see what patterns we can find.

 **Question of the Day:** How can I import data into AI Lab to train a machine learning model?

Activity (35 minutes)

 **Code Studio:** Have students continue to the next level in code studio. This level has students click a link to make a copy of a Google Sheet that represents the survey data from this form.

 2

Kim's Survey Data

 Teaching Tip

Which Row Is Mine?: Even though students just filled out this form, their data won't be represented here. Instead, this is a snapshot of the data from different students that is used just for this lesson. This means the size of the dataset won't match the number of students who just took the survey, and if students try to reverse-engineer the data to find their particular entry, they won't be able to.

 **Video:** Show students the [View and Download Survey Data](#) video in the alides. This video shoes students how to view data in Google Sheets, save their data as a CSV, and upload to AI Lab.

 Teaching Tip

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

 **Do This:** Rename the headings for each column to something more descriptive. Aim to have one or two words in each column heading. Then save the file as a csv to your computer.

Teaching Tip

Data Overload: Looking at all this data at once can be overwhelming! Luckily, students only need to do two things on this screen:

- Type in the first row of the table to rename the headers
- Save the file as a CSV

Students don't need to explore the data or try to discover patterns - that's what AI Lab is for!

Circulate: Check in with students as they complete this process. Encourage students to check in with each other as well to help get more familiar with Google Sheets, especially the process of saving as a csv.

 **Code Studio:** Have students continue to the next level in code studio. In this level, students will upload their data to AI Lab and train a model from the data.

3

AI Lab and Kim's Data

 **Do This:** Follow the instructions at each stage in AI Lab. Over the course of this level, your goals are:

- Train a model with at least 70% accuracy that uses the least amount of features
- Fill out the model card and save your model

Circulate: Monitor students as they work through AI Lab. Focus on how students are selecting their features, ensuring they are balancing both accuracy and simplicity. Encourage students to try multiple features, and to ask questions that make sense in the context of having a club recommended.

As students get close to completing the model cards, regroup and share the next slide.

 **Display:** Show students the slide showing the model card screen.

Remarks

Since we uploaded our own data, we have some new fields to fill out on our model card! We need to describe where this data came from, and we need to describe each of the columns we're using as features and label. This is a great place to write the question from the survey!

 **Do This:** Have students complete the model card. Have students use the questions from the survey for the description of each column, and the information from the slide for the data description.

Remarks

This is great - we were able to help Kim create a machine learning model for her data! But - Kim isn't the only person collecting data! Let's also check in with Isaac and see how they're doing

 **Display:** Show the next slide, which shows Isaac and some of their data

 **Discuss:** *What do you notice about Isaac's data?*

Discussion Goal: Students should notice that the numerical data has a large range that doesn't make sense, and they should notice that the categorical data includes extra values that weren't supposed to be a part of the survey.

 **Discuss:** *Why do you think the data ended up like this?*

Discussion Goal: This discussion should focus on two areas: how Isaac setup his form, and how he collected his data.

- If Isaac wasn't careful with his form from yesterday, he could have accidentally allowed people to type in their own responses or leave answers blank which is why the categorical data has issues. This can be a good opportunity to reinforce yesterday's lesson and how important it is to setup the form correctly to make sure students collect accurate data.
- If Isaac forced people to take his survey that didn't want to, they may not have taken it seriously and given fake answers like we see with the numerical data. Students may contrast this with what Kim did, where she asked people to participate rather than forced them to.

Remarks

Now that we understand why Isaac's data might look like this, let's see if we can help him clean up his data.

 **Code Studio:** Have students continue to the next level in code studio. In this level, students will download Isaac's data and be given instructions for cleaning the data.

 **4**

Isaac's Survey Data

 **Video:** Show students the [Cleaning Survey Data](#) video in the slides. This video guides students through how to clean up data in a spreadsheet so it can be used in AI Lab.

Teaching Tip

Don't Repeat Mistakes: Ideally, students won't need this skill in later lessons when they are creating their own surveys. Hopefully they will be able to use the template correctly, which will avoid the issues Isaac is facing right now. With that in mind, it's okay if students don't master the ability to edit spreadsheets - instead, emphasize how important it is to make sure your data is being collected in a controlled way that avoids this kind of situation.

 **Do This:** Fix the errors in Isaac's data so it will work with AI Lab. Once you've fixed the errors, download the file as a CSV and upload to AI Lab to verify that the data works correctly again.

 **5**

AI Lab and Isaac's Data

Assessment Opportunity

Save as CSV: This exercise is another opportunity for students to practice saving their data as a csv file and uploading to AI Lab. If students don't fix all of the errors in their first try, they may need to re-download and re-upload the CSV file multiple times. This may feel tedious now, but becoming familiar with this process now will help save students time and frustration during their final project.

There are 6 intentional errors in Isaac's data. As students work, feel free to offer this as a hint to students.

Teaching Tip

Saving a Model: Students aren't able to create a model from Isaac's data. If students ask, it's because the

sample size is too small and Isaac should collect more data before feeling comfortable training a model from it.

Remarks

Getting our survey data prepared for AI Lab is an important step. Kim had her data ready to go because her form was setup correctly, but Isaac needed some help cleaning his data. This is an important lesson to keep in mind next week when you'll be preparing your own survey data.

Wrap Up (5 minutes)

 **Prompt:** *What is a situation where you think it's okay to make a change to your survey data to help clean it up? What's a situation where you think it's not okay?*

Discussion Goal: Based on today's activity, students may think it's okay to adjust data when it's clearly a typo that is similar to another answer. Students may think it's not okay to completely change an answer to something else. In general, there's a lot of gray area in these situations, which is another reason it's important to have human's involved in machine learning to use their judgement with these decisions.



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**.