

Lección 10: estructuración de datos

Lección en línea [Ver en Code Studio](#)

Propósito

En esta Lección, los estudiantes profundizan en la recopilación e interpretación de datos, incluida la limpieza y visualización de datos. Los estudiantes primero observan cómo la presentación de datos de diferentes maneras puede ayudar a las personas a comprenderlo mejor, y luego crean visualizaciones de sus propios datos. Utilizando los resultados de una encuesta de pizza preferida, los estudiantes deben decidir qué hacer con los datos que no encajan fácilmente en el esquema de visualización que han elegido. Finalmente, los estudiantes observan qué partes de este proceso pueden ser automatizadas por una computadora y cuáles necesitan un ser humano para tomar decisiones.

Esta Lección demuestra que los datos sin procesar deben interpretarse de alguna manera para ayudar a las personas a usarlos para tomar decisiones. Los estudiantes participan tanto en la visualización como en la limpieza de los datos, y ven cómo los datos pueden malinterpretarse si no se limpian adecuadamente. Los estudiantes también experimentan el trabajo con datos a mano y con herramientas computacionales, y ven cómo los datos deben estructurarse de maneras particulares para ser utilizados por una computadora.

Secuencia para el aprendizaje

- Conocimiento inicial (5 min)
- Visualización de datos (70 min)
- Transferencia del conocimiento (15 min)

Objetivos

Los estudiantes serán capaces de:

- Identificar y eliminar datos irrelevantes de un conjunto de datos.
- Crear un gráfico de barras basado en un conjunto de datos.
- Explicar por qué se debe limpiar un conjunto de datos antes de que una computadora pueda usarlos.

Recursos

¡Atención!
Por favor, haga una copia de cada documento que planea compartir con los estudiantes.

Para los profesores:

- [Datos de pizza \(descarga de csv\): recurso opcional](#)
- [Datos de pizza \(GSheets\) - Recurso opcional](#)
- [Estructurando Datos - Ejemplar](#)

Para los estudiantes:

- [Estructuración de datos - Guía de lecciones](#)

Estrategia de aprendizaje

Conocimiento inicial (5 min)

Dirija a los estudiantes a los [Niveles de Code Studio](#) correspondientes a la Lección.

Pida a los estudiantes que piensen por sí mismos por un momento, luego discutan sus respuestas con un compañero.

Discuta:

Haga que los estudiantes compartan sus respuestas para las preguntas en el pizarrón.

Observaciones

A veces, los datos “en bruto”, la forma en que se recopila la información por primera vez, se debe poner en una forma diferente para que los humanos y las computadoras puedan entender más fácilmente lo que significa.

Los estudiantes deben comprender que las diferentes formas de datos facilitan que las personas tomen decisiones. También deberían ver que a la gente a menudo le va mejor con imágenes, como el gráfico de barras, mientras que la computadora funciona mejor con números, como la tabla.

Visualización de datos (70 min)

Agrupe a los estudiantes en parejas y entregue a cada pareja [Estructuración de datos - Guía de lecciones](#).

Lea las instrucciones en conjunto, asegurándose de que los estudiantes comprendan el problema que están tratando de resolver (eligiendo un relleno de pizza para la fiesta de pizza). Se les pide a los estudiantes que creen el gráfico de barras para el conjunto de datos brutos proporcionados. Algunas de las respuestas no se relacionarán fácilmente en las elecciones dadas. Aliente a los estudiantes a usar su mejor juicio sobre las respuestas que son difíciles de poner en la tabla, y que estos desafíos son una parte normal del proceso de resolución de problemas de datos.

Discuta:

Después de que los estudiantes terminen de elaborar la tabla y llenen las preguntas de reflexión, Haga que compartan sus respuestas con la clase

Los estudiantes deben ver que hay varias formas en que las respuestas pueden ser difíciles de categorizar, ya sea que sean completamente irrelevantes, que no sean lo suficientemente específicas que no sean una opción dada. Ignore la ortografía por ahora, si los niños no la mencionan.

También puede completar esta Lección usando Hojas de cálculo de Google o Excel. Los archivos de hoja de cálculo relevantes están vinculados en el Nivel 3 de la Lección en línea o en el área de enlaces de recursos de este plan de Lección.

Observaciones

Podemos hacer el gráfico a mano, pero también es posible que la computadora lo haga por nosotros. Esto es especialmente útil cuando tienes muchos datos.

Envíe a los estudiantes al ejercicio 3, y pida que sigan las instrucciones en el nivel.

Pida a los estudiantes que discutan en parejas por qué la tabla se ve de la forma en que lo hace y luego compartan sus respuestas con la clase. ¿Por qué la computadora no pudo poner todo en la categoría correcta?

Cuando creamos nuestras tablas, sabíamos que necesitábamos dejar fuera algunas de las respuestas que no tenían sentido, y que algunas respuestas, como “pimientos” y “pimientos verdes”, realmente significaban lo mismo. También colocamos todo lo que había sido mal escrito en la categoría correcta. Las computadoras no saben cómo hacerlo, porque en realidad no entienden qué es un “pimiento”, o que una palabra mal escrita es igual a una palabra correctamente deletreada. Eso significa que tenemos que limpiar los datos antes de que la computadora pueda usarlos.

Diga a los estudiantes que crearán una nueva columna de datos “limpios” que será más fácil de interpretar para la computadora.

Los estudiantes deben tener en cuenta que la computadora usó todas las respuestas en la tabla, incluso las que fueron irrelevantes. También deberían tener en cuenta que las ortografías diferentes de la misma elección no se agruparon juntas.

Envíe estudiantes al [ejercicio 4](#).

Al hacer clic en un encabezado en la lista “Datos limpios” y editarlo o eliminarlo. Demuestre que cuando elimina / cambia las respuestas en la columna de limpieza de datos, el gráfico cambia automáticamente.

Pida a los estudiantes que, en parejas, limpien los datos hasta que solo se muestren las siete elecciones originales, y luego decidan qué topping de pizza es la mejor opción.

Al final, los estudiantes deben darse cuenta de que restringir las elecciones de un usuario mediante el uso de opciones múltiples en lugar de una respuesta de escritura hace que sea más fácil para una computadora usar los datos.

Observaciones

Esto fue mucho trabajo, y solo fueron unos cincuenta votos.

¿Cuánto tiempo crees que tomaría limpiar los datos para una encuesta nacional? ¿Puede pensar en alguna forma de asegurarse de que tenemos datos claros desde el principio, para que ahorrar todo este trabajo?

Permita que los estudiantes lo discutan por parejas y luego compártalos con la clase.

Cuando trabajamos con grandes cantidades de datos, queremos automatizar tanto como podamos el proceso de mejora de problemas. Debido a que las computadoras no pueden

hacer las mismas conexiones que las personas, eso significa que las personas tienen que ayudar a organizar los datos de manera que las computadoras puedan entenderlos. Eso significa limpiar los datos o recopilar datos de una manera que asegure que estén limpios cuando los obtengamos.

Transferencia del conocimiento (15 min)

Preguntar:

Pida a los estudiantes que reflexionen sobre el desarrollo de las cinco prácticas de Descubrimientos CS (Resolución de problemas, Persistencia, Creatividad, Colaboración, Comunicación). Elija una de las siguientes indicaciones según lo considere apropiado.

- Elija una de las cinco prácticas en las que cree que demostró crecimiento en esta Lección. Escriba algo que hizo que ejemplifique esta práctica.
- Elija una práctica en la que piense que puede seguir creciendo. ¿Qué le gustaría mejorar?

Elija una práctica que pensó que era especialmente importante para la Lección que completamos hoy. ¿Qué lo hizo tan importante?

Sugerencias para evaluar

Se sugiere el siguiente indicador para evaluar formativamente los aprendizajes:

- Presentan la información trabajada, formas de trabajarlo y conclusiones