Soccer Analytics

Tina Leard

Activity summary

This is a free resource for teachers and students, and is part of the [Callysto project](#), a federally-funded initiative to bring data science skills into Grade 5-12 classrooms in Canada.

In this activity, students will use an online Callysto notebook to learn how to explore, analyze and interpret data related to the [UEFA Champions League](#) 2020 - 2021.

Additional resources on soccer for student research

- Champions League [Skills Showcase](#) videos
- Champions League [Videos](#)
- [Football made Simple](#) YouTube channel for soccer tactics; Juventus' coach [Pirlo's Tactical Philosophy](#) provides insight to a game of possession
- [Friends of Tracking](#) provides in-depth videos of professional soccer analytics

Grade level

Grades 7–9

Before you start

1. This lesson plan allows you to use Callysto notebooks, mathematical modeling, and coding to understand data related to UEFA Championship League soccer. You can choose to let students work through the notebook with you. Or, you could introduce this topic separately and use the notebook for the interactive activities.
2. Students should know how to log in to the [Callysto Hub](#) as well as run a notebook prior to interacting with it. Teachers, to get started with Callysto notebooks and running material on the Callysto Hub, see our [Starter Kit](#).

Required materials

- A charged computer
- Access to the internet
- An installed internet browser, preferably Google Chrome, Safari, or Firefox
- A Google or Outlook email account.

Learning outcomes

Apply mathematical reasoning and critical thinking to make meaning of data in tables and data visualizations	
Use language connected to mathematics and data science, including dataset, visualization, and library	Notebooks 1-3
Interpret different sets of data presented in various ways by answering questions about the data and drawing conclusions	Notebooks 1-3
Recognize characteristics of quality and quantity of a large dataset, including filtered dataframes	Notebook 1
Locate Cartesian coordinates (quadrant I) on the soccer field	Notebook 1
Interpret the range of data	Notebook 1
Locate AB patterns in a time sequence visualization	Notebook 1
Recognize characteristics of quality and quantity of smaller datasets	Notebooks 2 & 3
Compare data on outliers	Notebook 3
Locate integers on the y-axis of a scatter plot	Notebook 3

Adapt code, decomposing and determining patterns in Python syntax, to manipulate, visualize, and understand data	
Adapt coding algorithms for one–variable graphs	Notebook 2
Adapt coding algorithms for two–variable graphs	Notebook 3
Use the Python library plotly to display graph titles and labels	Notebooks 2 & 3
Use the Python library plotly to represent data visually through bar graphs	Notebooks 2 & 3
Use the Python library plotly to represent data visually through a circle graph with counts translated to percentages	Notebook 2
Use the Python library plotly to represent data visually through scatter plots	Notebook 3
Use the Python library pandas to calculate mean, median, and range	Notebook 3
Use the Python library pandas to write a .csv file	Notebook 3

Background information

These soccer analytics activities use data from an anonymized professional game and statistics from [UEFA Champions League](#) 2020-2021 to create some data visualizations on soccer using [Python](#) code in a [Jupyter notebook](#).

There are three student notebooks:
1. [Interpreting visualizations](#) from a professional soccer anonymized dataset
2. [Reading data tables](#) from the Champions League to code bar graphs and a circle graph to explore goals
3. [Creating a .csv file](#) and coding scatter plots to explore relationships between ball possession and scoring

Those links to the notebooks will create a copy of the lesson plans repository in your Callysto Hub account, and open the (student or instructor) notebook. To log in to the Callysto Hub, you and your students will need a Google or Microsoft account. This can be a school division-provided account or a personal account. Callysto does not collect any personal information about accounts.

A Callysto [starter kit](#) is available for teachers as well as a [notebook presentation](#) on data literacy through sports.

Using the Notebooks

Although there is a learning progression from notebooks 1 to 3, the notebooks could be used separately to address targeted learning outcomes. The notebooks could be introduced through a class walkthrough, through group work, or as self-directed work.

Notebook 1

This notebook accesses a large dataset and explores several visualizations with interpretations. Prompts for noticing features of the tables and visualizations are present. As a next step, students could extend noticing to wondering and generate their own questions with answers.

Critical thinking prompts include…
- Continuity & Change. What stays the same and what changes over a period of time?
- Patterns & Trends. Do similar characteristics repeat?
- Interrelationships. What connections exist within and between systems?

The visualizations in this notebook could be used as real-life examples of Cartesian coordinates and time sequence patterning. If you're familiar with Python, Cartesian coordinates could be further explored by plotting types of play by team within a timeframe.

Notebook 2

This notebook accesses smaller datasets and provides opportunities to adapt existing code.

If students have some experience with coding, you may want to just provide a brief introduction and then let students work through the notebook on their own.

If data science and Python code in a Jupyter notebook are all new experiences for your students, you may prefer to provide direct instruction on the functioning of the notebook.

Python code is used to translate frequency counts into percentages in a circle graph. Further practice on converting to percentages and supporting proportional reasoning could be introduced, possibly through other student-selected or generated datasets.

Notebook 3

This notebook generates a .csv file of a smaller dataset and provides opportunities to adapt existing code.

An extension to this notebook could include determining the impact of adding or removing data from a dataset on a measure of central tendency. Students could write a new .csv file by removing teams, explaining their reasoning for removal. Then, they could define the new data frame within the code for the visualizations to explore different interpretations.

For the scatter plots, prompts for noticing features of the plots are present. As a next step, students could extend noticing to wondering and generate their own questions with answers. They could support their explanations with research on the teams.

Callysto Resources

For more information, you can check out our YouTube videos, online courses, or callysto.ca for learning modules, data visualization exercises, and lesson plans.

Sequencing

The Frog Jumping Problem

Removing outliers

Construction Times in Age of Empires
Data Science and Memes

Circle graphs

Twitch Popularity for Gamers

Line of best fit

Arctic Sea Ice Levels

Line of best fit and integers

Data Science and Climate Change

Scatter plots

Baseball Pitching Statistics