**Day 73 Goals: what you will make by the end of the day**

**Learn to Aggregate and Merge Data in Pandas while Analysing a Dataset of LEGO Pieces**



Today we're going to be diving deep into a dataset all about LEGO, which will help us answer a whole bunch of interesting questions about the history of the company, their product offering, and which LEGO set rules them all:

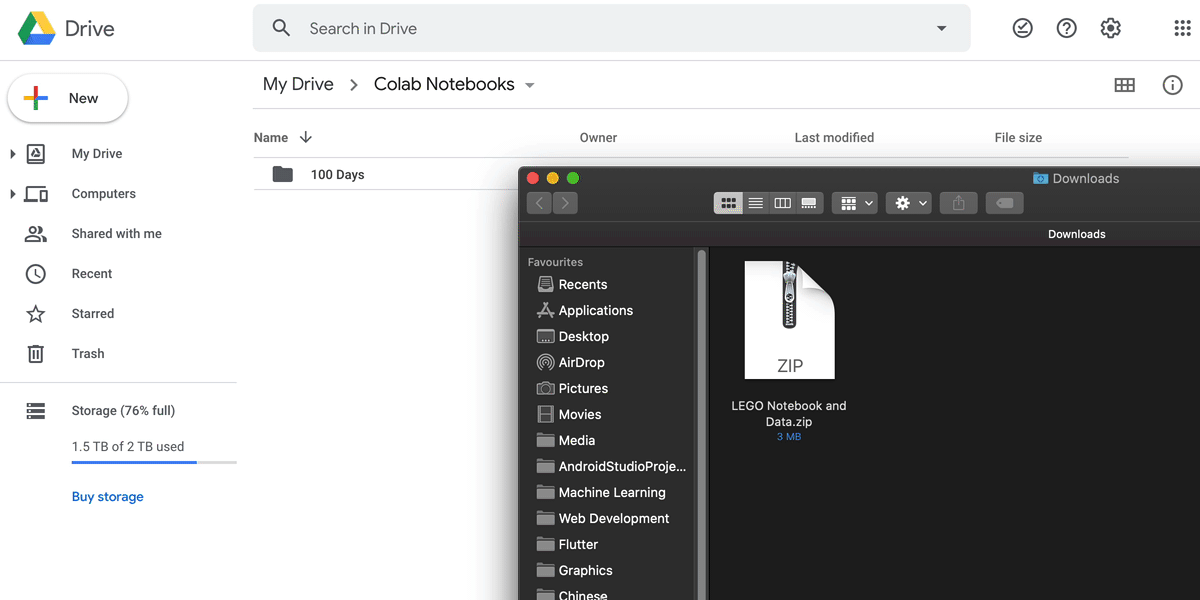
* What is the most enormous LEGO set ever created and how many parts did it have?
* In which year were the first LEGO sets released and how many sets did the company sell when it first launched?
* Which LEGO theme has the most sets? Is it Harry Potter, Ninjago, Friends or something else?
* When did the LEGO company really take-off based on its product offering? How many themes and sets did it release every year?
* Did LEGO sets grow in size and complexity over time? Do older LEGO sets tend to have more or fewer parts than newer sets?

**What you'll learn today**

* How to combine a Notebook with HTML Markup.
* Apply Python List slicing techniques to Pandas DataFrames.
* How to aggregate data using the .agg() function.
* How to create scatter plots, bar charts, and line charts with two axes in Matplotlib.
* Understand database schemas that are organised by primary and foreign keys.
* How to merge DataFrames that share a common key

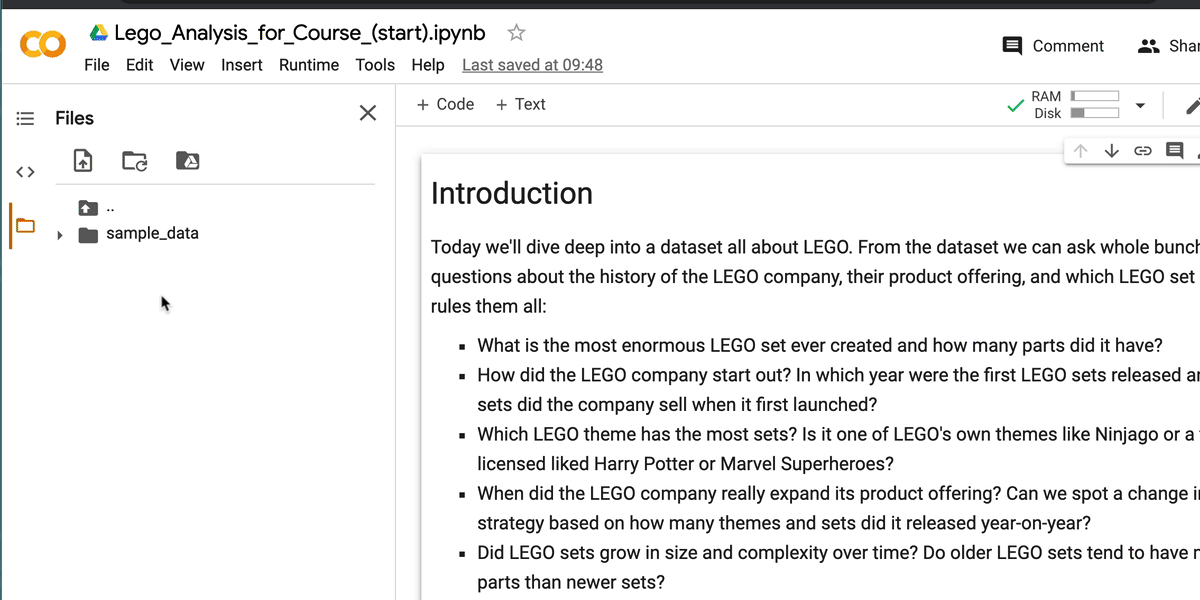
**Download and Open the Template Notebook**

Download the .zip file from this lesson. This contains the starter Notebook, some images, and the data. Unzip the file and add the Notebook to your Google Drive.



**Add the Data to your Project**

Right-click to create a new folder called data and add the .csv files to the data folder (sadly dragging and dropping entire folders won't work).



If you are running the Notebook locally using Jupyter, I've supplied the image assets for you as well. If you're using Google Colab, then you won't need the assets folder.