Sign In

Register

1/1 point



Activity overview

1.

By now, you've learned a lot about different data types and data structures. In this activity, you will work with datasets from Kaggle, an online community of

Every data analyst has a data community that they rely on for help, support, and inspiration. Kaggle can help you build your own data community. Kaggle has millions of users in all stages of their data career, from beginners to data scientists with decades of experience. The Kaggle community brings

people together to develop their data analysis skills, share datasets and interactive notebooks, and collaborate on solving real-life data problems.

people passionate about data. To start this activity, you'll create a Kaggle account, set up a profile, and explore Kaggle notebooks.

Check out this <u>brief introductory video</u> 'L' to learn more about Kaggle.

By the time you complete this activity, you will be able to use many of Kaggle's key features. This will enable you to create notebooks and browse data, which is important for completing and sharing data projects in your career as a data analyst.

To get started, follow these steps to create a Kaggle account.

Note: Kaggle frequently updates its user interface. The latest changes may not be reflected in the screenshots, but the principles in this activity remain the same. Adapting to changes in software updates is an essential skill for data analysts, and we encourage you to practice troubleshooting. You can

Create a Kaggle account

also reach out to your community of learners on the discussion forum for help. 1. Go to kaggle.com ☑

kaggle

Rachit Shukla . Follow

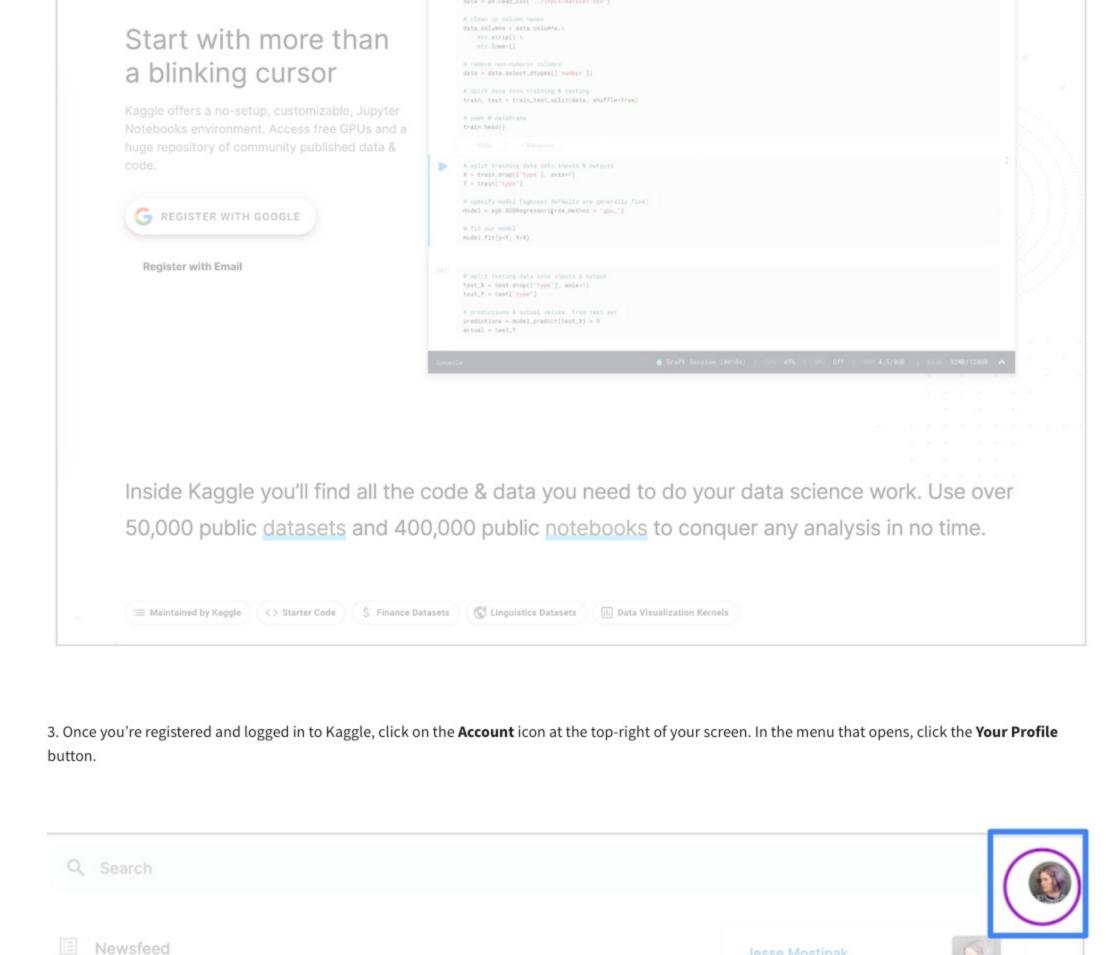
Next Gen Gaming Laptops

2. Click on the **Register** button at the top-right of the Kaggle homepage. You can register with your Google credentials or your personal email address.

Discuss

Notebooks

Datasets



Marília Prata commented on 1 Competitions Contributor

Jesse Mostipak

Datasets Expert

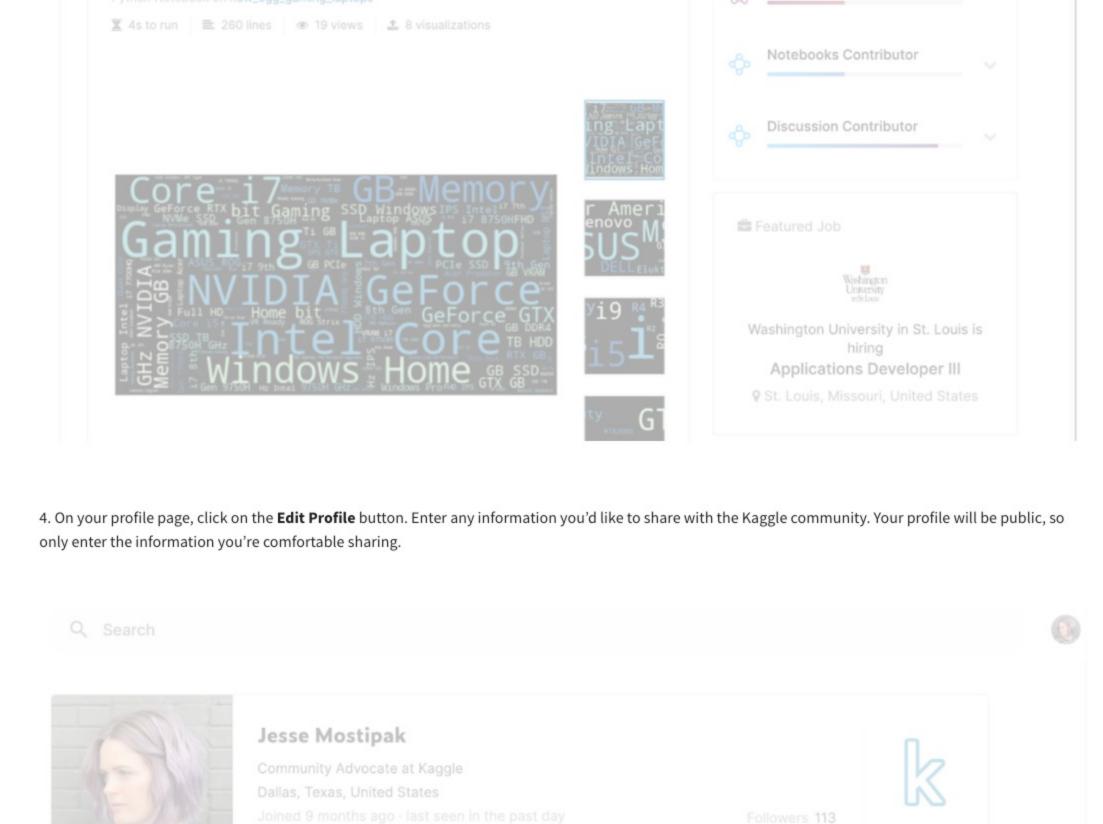
Edit Profile

= Filters

See all (277)

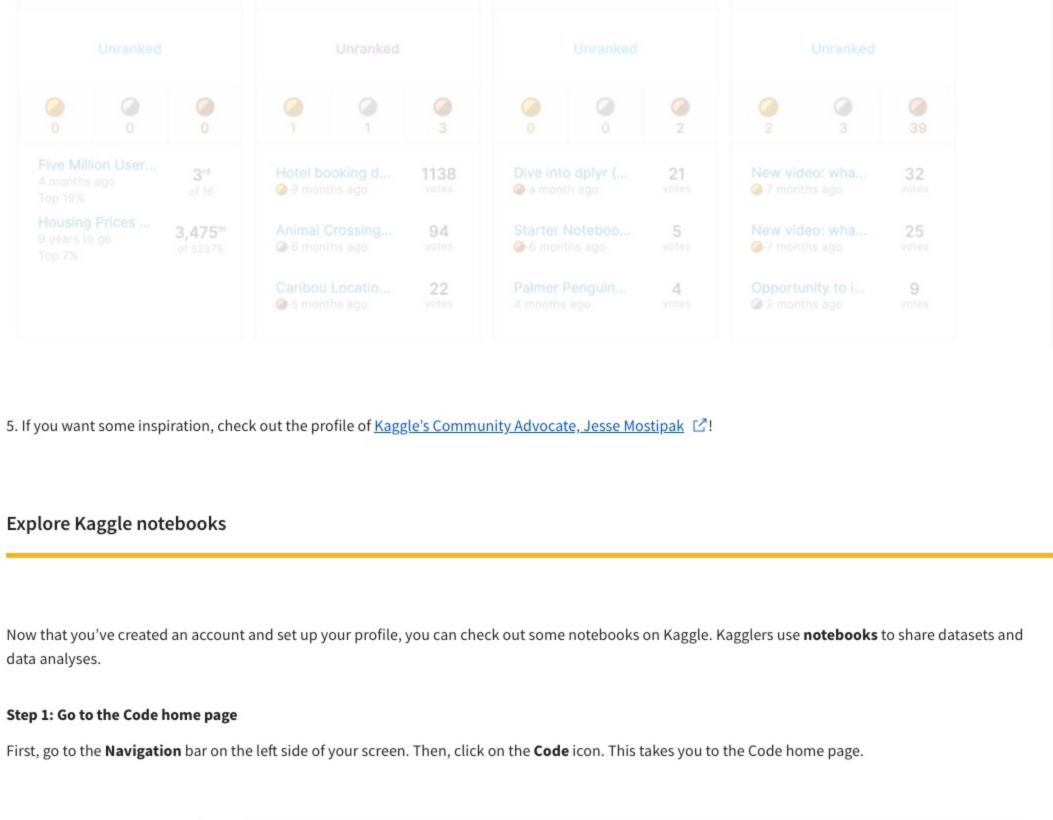
TabularMarch21 DAE

starter CV-Inference



Code Discussion Followers Notifications

https://www.jessemaegan.com/



Code Compete Explore and run machine learning code with Kaggle Notebooks. Find help in the Documentation. Data Code Your work

Finance

Random Forest

SCL 2021 Data Science -

Part 1 (Data Cleaning)

GPU

TPU

Competition notebook

Sales Cluster & Analysis

3 Share

© Save Version

input (62.92 KB)

/kaggle/working

output

Settings

Language

Environment

Accelerator

Internet

Code Help

>

(5)

+ Add data

M palmer-archipelago-antarctica-p...

Preferences

None

Search for examples of how to do things

●On

Q Find Code Help

Updated a day ago

Q Search public notebooks

Trending

Titanic_new

On the Code home page, you'll notice links to notebooks created by other Kagglers.

Updated 3 hours ago

Q Search

kaggle

Communities

Dive into dplyr (tutorial... Tidying up Among Us ...

View Active Events

Step 2: Review Kaggler contributions

More

Recently Viewed

Home

```
To begin, feel free to scroll through the list and click on notebooks that interest you. As you explore, you may come across unfamiliar terms and new
information: That's fine! Kagglers come from diverse backgrounds and focus on different areas of data analysis, data science, machine learning, and deep
learning.
Step 3: Narrow your search
Once you're familiar with the Code home page, you can narrow your search results by typing a word in the search bar or by using the filter feature.
For example, type Beginner in the search bar to show notebooks tagged as beginner-friendly. Or, click on the Filter icon, the triangle shape on the right side
of the search bar. You can filter results by tags, programming language, output, and other options. Filter to Datasets to show notebooks that use one of the
tens of thousands of public datasets available on Kaggle.
Step 4: Review suggested notebooks
If you're looking for specific suggestions, check out the following notebooks:
     gganimate [ ] [ by Meg Risdal
     Writing Hamilton Lyrics with TensorFlow/R ☐ by Ana Sofia Uzsoy
     <u>Dive into dplyr (tutorial #1)</u> ✓ by Jesse Mostipak
Spend some time checking out a couple of notebooks to get an idea of the work that Kagglers share online—and that you'll be able to create by the time
you've finished this course!
Edit a notebook
```

Now, take a look at a specific notebook: Dive into dplyr (tutorial #1) 🔀 by Jesse Mostipak. Follow these steps to learn how to edit notebooks:

There are a lot of amazing packages in the Tidyverse, but dplyr is hands-down my absolute favorite package. I use dplyr when I'm

cleaning and exploring my dataset, and what I particularly love is that after I get a good handle on my dataset with dplyr, I can feed

This tutorial is for anyone interested in learning the basics of the dplyr package. We'll be focusing on data exploration and

manipulation, building off of the examples in the dplyr package documentation using the Palmer Penguins dataset.

2. Click on the Copy and Edit button at the top-right to make a copy of the notebook in your account. Now, the notebook appears in Edit mode. Edit mode

Draft Session (0m)

1. Click on the link to open up the notebook. It contains the dataset you'll work with later on.

the various manipulations I've creating into the ggplot2 package for visualization.

Show apps)ive into dplyr (tutorial #1) Draft saved View Run Add-ons Help Ø ▶ Run All Φ

<>

0

lets you make changes to the notebook if you want.

Introduction: why dplyr?

By the end of this notebook, you'll be able to:

. Demonstrate what each of the main five dplyr functions does

Use the pipe operator %>% to chain together multiple dplyr functions

What I've learned I still have questions about

My analytical workflow We won't be covering all of the steps in my workflow in this tutorial, but in general I follow these steps:

1. Set up the programming environment by loading packages

2. Import my data 3. Check out my data 4. Explore my data 5. Model my data

6. Communicate what I've learned

This notebook is private. If you want to share your work, you can choose to make it public. When you copy and edit another Kaggler's work, always make

Dive into dplyr (tutorial #1) Draft saved

PAL0708 Adelie Penguin PAL0708

sliders.

Confirmation and reflection

Correct

✓ Correct

work.

Which statements are true about the two penguin datasets in the **Dive into dplyr (tutorial #1)** notebook? Select all that apply. In **penguins_lter.csv**, the highest value in the column Sample Number is 152.

datasets, you used an interactive notebook's data viewing feature. Going forward, you can use interactive notebooks to examine and describe data. This is an important skill that will help you complete data projects in the future. In penguins_lter.csv, the column Individual ID cannot be sorted.

This is an important skill that will help you complete data projects in the future.

In both datasets, the number of columns is the same. penguins_size.csv has 7 columns. Correct The **penguins_size.csv** has 7 columns. In **penguins_lter.csv**, the highest value in the column Sample Number is 152. To learn about the penguin datasets, you used an interactive notebook's data viewing feature. Going forward, you can use interactive notebooks to examine and describe data.

The **penguins_size.csv** has 7 columns. In **penguins_lter.csv**, the highest value in the column Sample Number is 152. To learn about the penguin

Using all of the information you learned while exploring in Kaggle, how would you thoroughly describe these datasets to someone else? How do you think sharing interactive notebooks online can help you develop your data analysis skills?

Data analysts use a variety of resources to complete data analysis projects. For instance, an analyst could use Kaggle notebooks to host projects in a portfolio. This is important for practicing and demonstrating your skills, as well as getting feedback from more experienced data analysts on your work.

two penguins datasets. In the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:

2. In this activity, you've learned a lot about data types and data structures. Using what you've learned so far, consider your experience with datasets and the

Congratulations on completing this hands-on activity! You worked with notebooks and used different datasets in Kaggle. A strong response would include that online resources like Kaggle help data analysts accomplish many important tasks. Beyond that, consider the following: Data analysts use a variety of resources to complete data analysis projects. For instance, an analyst could use Kaggle notebooks to host projects in a

meaningful changes to the notebook before publishing it. That way, you're not misrepresenting someone else's work as your own. Take a moment to explore the Edit mode of the notebook. Some of this may seem unfamiliar—and that's just fine. By the end of this course, you'll know how to create a notebook like this from scratch! Working with datasets in notebooks Now, you can check out the data! In this notebook, you'll find the data in a box labeled Data at the top-right of your screen. In the box, there's an input folder with the title: palmer-

1. Click on this title. Two .csv files appear: penguins_lter.csv and penguins_size.csv. Click on one of them. At the bottom of the notebook, you'll now find an

archipelago-antarctica-penguin-data. Follow these instructions to explore the datasets and learn more about the data within them:

interactive data table with all the information from the dataset. © Save Version 13 Share $|\langle$ File Edit View Run Add-ons Help ● Draft Session (1m) Ø + D DD Run All + Add data input (62.92 KB) Φ - kt palmer-archipelago-antarctica-p... ↑ ↓ B × I penguins_Iter.csv Introduction: why dplyr? penguins_size.csv output 🗆 <> There are a lot of amazing packages in the Tidyverse, but dplyr is hands-down my absolute favorite package. I use dplyr when I'm cleaning ▶ □ /kaggle/working and exploring my dataset, and what I particularly love is that after I get a good handle on my dataset with dplyr, I can feed the various manipulations I've creating into the ggplot2 package for visualization. Settings Ø This tutorial is for anyone interested in learning the basics of the dplyr package. We'll be focusing on data exploration and manipulation, building off of the examples in the dplyr package documentation using the Palmer Penguins dataset. Environment Preferences By the end of this notebook, you'll be able to: Accelerator Demonstrate what each of the main five dplyr functions does **3**0n . Use the pipe operator %>% to chain together multiple dplyr functions Code Help 吕, Q Find Code Help Copy 10 of 17 columns v ../input/palmer-archipelago-antarctica-penguin-data/penguins_lter.csv ₹ # Sample Nu... = A Species A Region A Individual ID = ✓ Clutch Co... = A Island A Stage ☐ Date Egg # Culmen Le... =

PAL0708 Adult, 1 Egg Stage 11/11/07 39.1 N1A2 11/11/07 39.5 Torgersen Adult, 1 Egg Stage 11/16/07 Adult, 1 Egg Stage N2A1 Torgersen Adult, 1 Egg Stage Torgersen PAL0708 Adult, 1 Egg Stage N3A1 11/16/07 36.7 (Pygoscelis Adelie Penguin PAL0708 Adult, 1 Egg Stage N3A2 11/16/07 39.3 PAL0708 Adelie Penguin Adult, 1 Egg Stage N4A1 11/15/07 38.9 adeliae) PAL0708 Adelle Penguin Amvers Torgersen Adult, 1 Egg Stage N4A2 39.2 2. Click on the other .csv file. This opens a second tab with the second dataset. Take a moment to check out each dataset. 4. Change the range of the quantitative data in the rows by clicking on the horizontal bars to the right of the column name and changing the position of the 5. Click on the button that says **10 of 17 columns** to change the columns that are visible in the table. In the dropdown menu, there's a checkmark next to the name of each column that appears in the table. Checking or unchecking one of these boxes will change what data is presented. Congratulations! You've explored several ways to interact with the dataset. This will help you get familiar with the Kaggle interface. You can save the notebook you worked in for future reference. Coming up, you'll learn more about other ways you can use Kaggle.

Search for examples of how to do things

Congratulations on completing this hands-on activity! You worked with notebooks and used different datasets in Kaggle. A strong response would include that online resources like Kaggle help data analysts accomplish many important tasks. Beyond that, consider the following:

1/1 point

portfolio. This is important for practicing and demonstrating your skills, as well as getting feedback from more experienced data analysts on your