

K-means Project Tutorial

[Edit on Github](#)

- In this small project you will use the k-means algorithm to segment houses based on their coordinates and median income.

How to start this project

You will not be forking this time, please take some time to read this instructions:

1. Create a new repository based on [machine learning project](#) by [clicking here](#).
2. Open the recently created repository on Gitpod by using the [Gitpod button extension](#).
3. Once Gitpod VSCode has finished opening you start your project following the Instructions below.

How to deliver this project

Once you are finished creating your clustering project, make sure to commit your changes, push to your repository and go to 4Geeks.com to upload the repository link.



Instructions

House clustering

We will create 6 housing clusters based only on their 'latitude','longitude' and their 'medincome' column.

Dataset links:

<https://raw.githubusercontent.com/4GeeksAcademy/k-means-project-tutorial/main/housing.csv>

Step 1:

Install and import the necessary libraries: pandas, sklearn and seaborn.

Step 2:

Load the housing dataset and take a look at the first rows. Then create a new dataframe with only the 'latitude','longitude' and 'medincome' column to create our clusters.

Step 3:

Instantiate the kmeans algorithm. Then, create a new 'cluster' feature in your dataset and predict the cluster by fitting the 3 columns you have. You can view the k-means

documentation to implement it: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>.

Step 4:

Convert your new 'cluster' column to 'category' type.

Step 5:

Use seaborn's relplot to visualize your new clusters.

Step 6:

As always, use your notebook to experiment and make sure you are getting the results you want.

Use your app.py file to save your defined steps, pipelines or functions in the right order.

In your README file write a brief summary.

Solution guide:

https://github.com/4GeeksAcademy/k-means-project-tutorial/blob/main/solution_guide.ipynb