



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

Start with more than
a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter
Notebooks environment. Access GPUs at no cost
to you and a huge repository of community
published data & code.



REGISTER WITH GOOGLE

Register with Email

```
Predict Malicious Websites: XGBoost Draft saved
File Edit Insert Run View Help

This kernel has an XGBoost model that predicts whether a website is malicious or not.

In [ ]:
import numpy as np
import pandas as pd
import xgboost as xgb

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.utils.multiclass import unique_labels

data = pd.read_csv("../input/dataset.csv")

# clean up column names
data.columns = data.columns.\
    str.strip().\
    str.lower()

# remove non-numeric columns
data = data.select_dtypes(['number'])

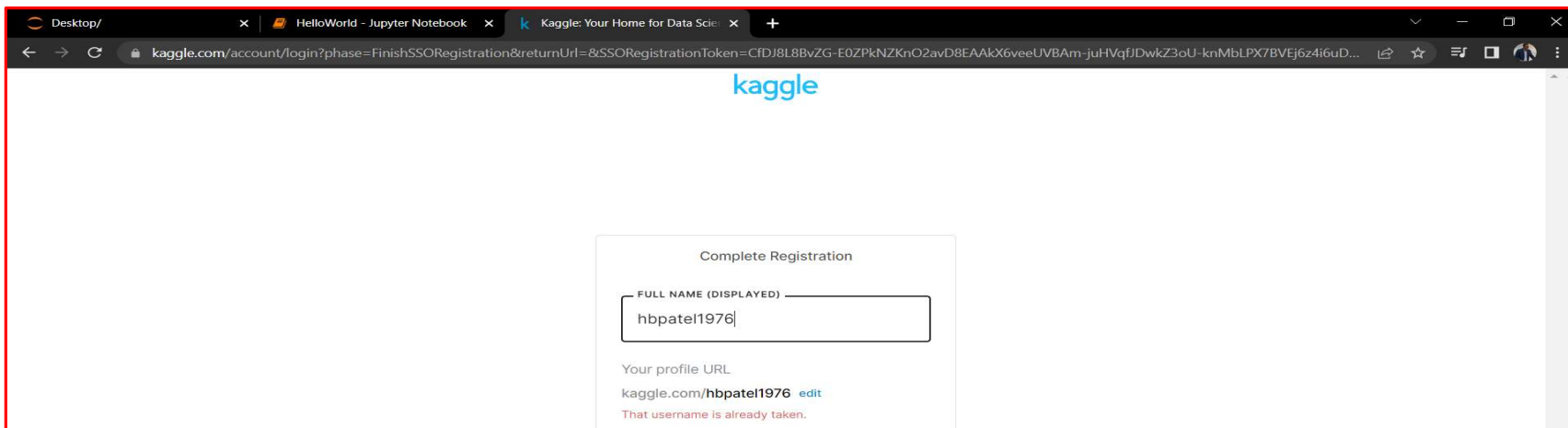
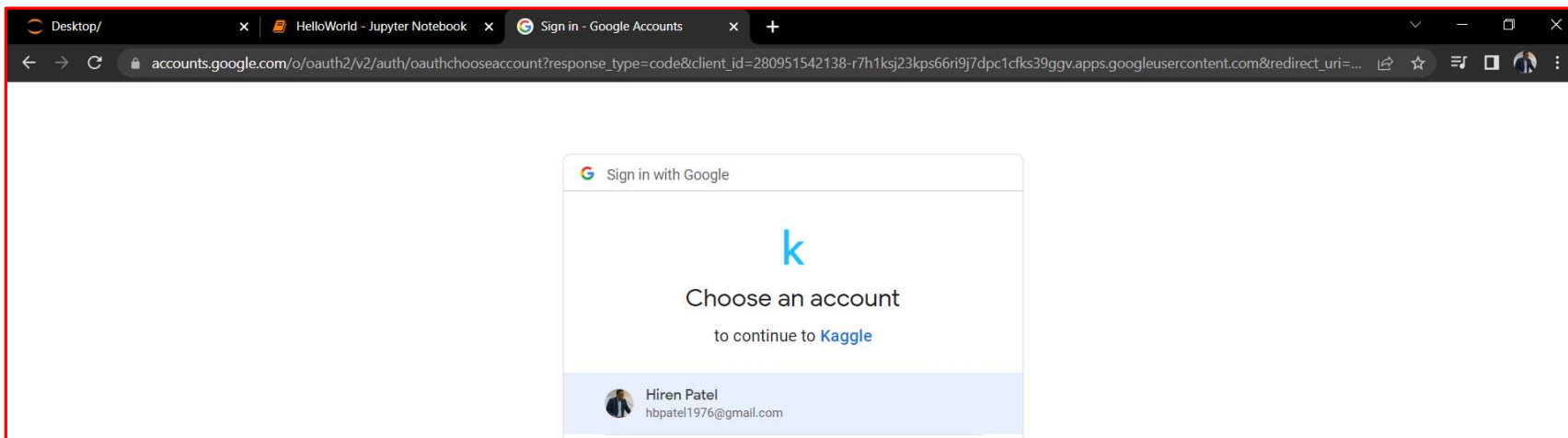
# split data into training & testing
train, test = train_test_split(data, shuffle=True)

# peek @ dataframe
train.head()
```



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in





Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

Desktop/ x HelloWorld - Jupyter Notebook x Kaggle: Your Home for Data Science x +

kaggle.com

Search

Welcome, hbpatel1976!
Kaggle is the place to learn data science and build a portfolio

How to start: Choose a focus for today
Help us make relevant suggestions for you

Learn to compete on Kaggle
Improve and test your skills
Get started →

Take a short course
Our courses are the fastest way to learn data science
Get started →

Browse inspiring data and code
Improve your data science projects
Get started →

View Active Events

27°C Cloudy 11:09 AM 8/17/2022



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

Desktop/ x HelloWorld - Jupyter Notebook x Search | Kaggle x +

kaggle.com/search?q=video

video game sales

View all results for "video game sales"

video game sales with ratings

video game sales 2020

Datasets View all 46

Video Game Sales 6 years ago • 1 file (csv) • 390 kB • 4676

Video Game Sales with Ratings 6 years ago • 2 MB • 913

Notebooks View all 845

EDA - VIDEO GAME SALES 2 years ago • 3s to run • R • 372

Video Game Sales Exploratory Data Analysis 3 years ago • 24s to run • Python • 150

Comments View all 495

Reply to Video Game Sales 2 years ago • General • 8

Reply to Video Game Sales 2 years ago • General • 3



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

The screenshot shows the Kaggle website interface. The browser tabs at the top include 'Desktop/', 'HelloWorld - Jupyter Notebook', and 'Video Game Sales | Kaggle'. The address bar shows 'kaggle.com/datasets/gregorut/videogamesales'. The left sidebar contains navigation links: 'Create', 'Home', 'Competitions', 'Datasets', 'Code', 'Discussions', 'Courses', 'More', 'Your Work', 'RECENTLY VIEWED' (with 'Video Game Sales' listed), and 'View Active Events'. The main content area displays the 'Video Game Sales' dataset page. It features a search bar, a user profile for 'GREGORYSMITH' (updated 6 years ago), and a 'Download (390 kB)' button. The dataset title 'Video Game Sales' is prominently displayed, followed by the description 'Analyze sales data from more than 16,500 games.' Below this, there are tabs for 'Data', 'Code (909)', 'Discussion (35)', and 'Metadata'. The 'About Dataset' section states: 'This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com. Fields include: Rank - Ranking of overall sales, Name - The games name, Platform - Platform of the games release (i.e. PC,PS4, etc.), Year - Year of the game's release, Genre - Genre of the game, Publisher - Publisher of the game, NA_Sales - Sales in North America (in millions)'. On the right side, there are sections for 'Usability' (5.88), 'License' (Unknown), and 'Expected update frequency' (Not specified). The bottom of the screen shows a Windows taskbar with the date '8/17/2022' and time '11:11 AM'.



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

Desktop/ x HelloWorld - Jupyter Notebook x Video Game Sales | Kaggle x +

kaggle.com/datasets/gregorut/videogamesales

kaggle

+ Create

Home

Competitions

Datasets

Code

Discussions

Courses

More

Your Work

RECENTLY VIEWED

Video Game Sales

View Active Events

Search

Video Game Sales

Data Code (909) Discussion (35) Metadata

Games Video Games

4686 New Notebook Download (390 kB)

vgsales.csv (1.36 MB)

11493 unique values

16.6k

10 of 11 columns

# Rank	Name	Platform	Year	Genre	Platform
1	Wii Sports	Wii	2006	Sports	Nin
2	Super Mario Bros.	NES	1985	Platform	Nin
3	Mario Kart Wii	Wii	2008	Racing	Nin
4	Wii Sports Resort	Wii	2009	Sports	Nin
5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nin
6	Tetris	GB	1989	Puzzle	Nin

Data Explorer

Version 2 (1.36 MB)

vgsales.csv

vgsales.csv.zip

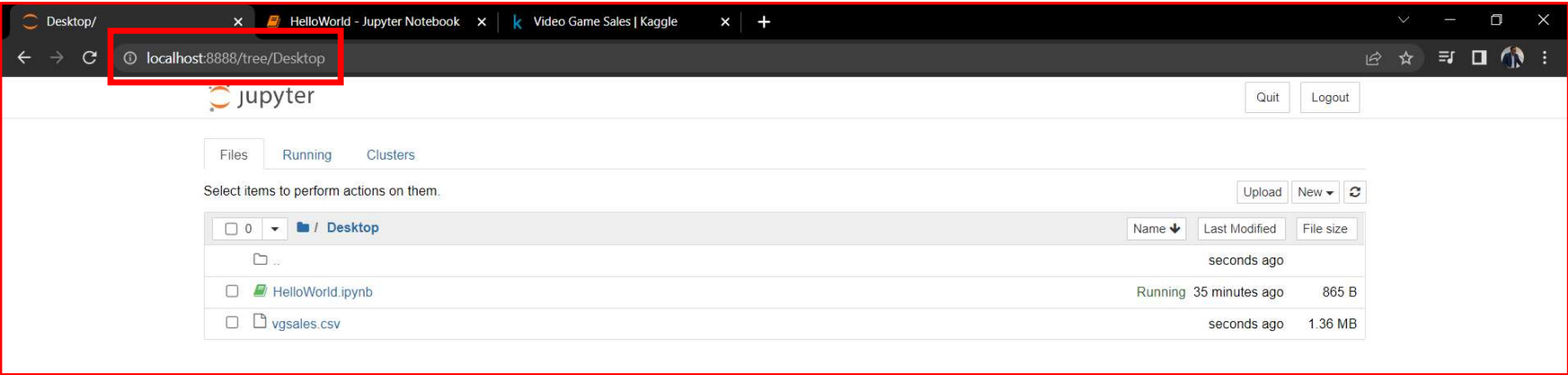
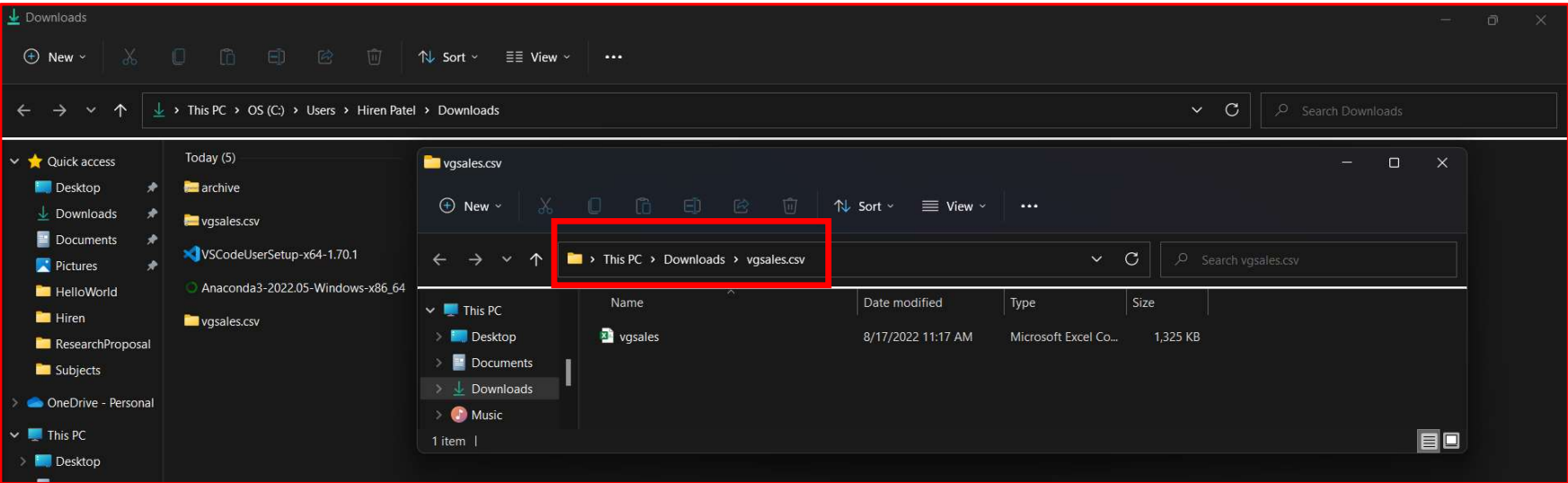
27°C Cloudy

ENG IN 11:16 AM 8/17/2022



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in





Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

Desktop/ x HelloWorld - Jupyter Notebook x Video Game Sales | Kaggle x +

localhost:8888/notebooks/Desktop/HelloWorld.ipynb

jupyter HelloWorld Last Checkpoint: an hour ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run

```
In [2]: import pandas as pd
df = pd.read_csv('vgsales.csv')
df
```

Out[2]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37
...
16593	16596	Woody Woodpecker in Crazy Castle 5	GBA	2002.0	Platform	Kemco	0.01	0.00	0.00	0.00	0.01
16594	16597	Men in Black II: Alien Escape	GC	2003.0	Shooter	Infogrames	0.01	0.00	0.00	0.00	0.01
16595	16598	SCORE International Baja 1000: The Official Game	PS2	2008.0	Racing	Activision	0.00	0.00	0.00	0.00	0.01
16596	16599	Know How 2	DS	2010.0	Puzzle	7G//AMES	0.00	0.01	0.00	0.00	0.01
16597	16600	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.00	0.01

16598 rows x 11 columns

In []:



Importing/Loading Dataset (CSV) in Jupyter

www.hbpatel.in

```
In [3]: import pandas as pd
df = pd.read_csv('vgsales.csv')
df.shape
```

```
Out[3]: (16598, 11)
```

```
In [4]: df.describe()
```

```
Out[4]:
```

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16598.000000	16327.000000	16598.000000	16598.000000	16598.000000	16598.000000	16598.000000
mean	8300.605254	2006.406443	0.264667	0.146652	0.077782	0.048063	0.537441
std	4791.853933	5.828981	0.816683	0.505351	0.309291	0.188588	1.555028
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4151.250000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8300.500000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12449.750000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.470000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

```
In [5]: df.values
```

```
Out[5]: array([[1, 'Wii Sports', 'Wii', ..., 3.77, 8.46, 82.74],
 [2, 'Super Mario Bros.', 'NES', ..., 6.81, 0.77, 40.24],
 [3, 'Mario Kart Wii', 'Wii', ..., 3.79, 3.31, 35.82],
 ...,
 [16598, 'SCORE International Baja 1000: The Official Game', 'PS2',
 ..., 0.0, 0.0, 0.01],
 [16599, 'Know How 2', 'DS', ..., 0.0, 0.0, 0.01],
 [16600, 'Spirits & Spells', 'GBA', ..., 0.0, 0.0, 0.01]],
 dtype=object)
```