

ML-Dev-Hackathon-2022-GPCET

Untitled1.ipynb - Colaboratory

colab.research.google.com/drive/1KV1j9cH0IE5F_JLejoSvblvEhe46PtHi

CommentShare

RAMDisk

Editing

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_dataStartup.csv

```
import pandas as pd

dataset=pd.read_csv('Startup.csv')
print(dataset)
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	134615.46	147198.87	127716.82	California	156122.51
7	130298.13	145530.06	323876.68	Florida	155752.60
8	120542.52	148718.95	311613.29	New York	152211.77
9	123334.88	108679.17	304981.62	California	149759.96
10	101913.08	110594.11	229160.95	Florida	146121.95
11	100671.96	91790.61	249744.55	California	144259.40
12	93863.75	127320.38	249839.44	Florida	141585.52
13	91992.39	135495.07	252664.93	California	134307.35
14	119943.24	156547.42	256512.92	Florida	132602.65
15	114523.61	122616.84	261776.23	New York	129917.04
16	78013.11	121597.55	264346.06	California	126992.93
17	94657.16	145077.58	282574.31	New York	125370.37
18	91749.16	114175.79	294919.57	Florida	124266.90
19	86419.70	153514.11	0.00	New York	122776.86
20	76253.86	113867.30	298664.47	California	118474.03
21	78389.47	153773.43	299737.29	New York	111313.02
22	73994.56	122782.75	303319.26	Florida	110352.25
23	67532.53	105751.03	304768.73	Florida	108733.99
24	77044.01	00701.24	140574.01	New York	100553.04

0s completed at 10:39 PM

Startup

Show all

Type here to search

22:4022-03-2022

Files

- ..
- sample_data
- Startup.csv

Disk 68.20 GB available

```
+ Code + Text
X=dataset.iloc[:, :-1].values
Y=dataset.iloc[:, :-1].values
X
array([[165349.2, 136897.8, 471784.1, 'New York'],
       [162597.7, 151377.59, 443898.53, 'California'],
       [153441.51, 101145.55, 407934.54, 'Florida'],
       [144372.41, 118671.85, 383199.62, 'New York'],
       [142107.34, 91391.77, 366168.42, 'Florida'],
       [131876.9, 99814.71, 362861.36, 'New York'],
       [134615.46, 147198.87, 127716.82, 'California'],
       [130298.13, 145530.06, 323876.68, 'Florida'],
       [120542.52, 148718.95, 311613.29, 'New York'],
       [123334.88, 108679.17, 304981.62, 'California'],
       [101913.08, 110594.11, 229160.95, 'Florida'],
       [100671.96, 91790.61, 249744.55, 'California'],
       [93863.75, 127320.38, 249839.44, 'Florida'],
       [91992.39, 135495.07, 252664.93, 'California'],
       [119943.24, 156547.42, 256512.92, 'Florida'],
       [114523.61, 122616.84, 261776.23, 'New York'],
       [78013.11, 121597.55, 264346.06, 'California'],
       [94657.16, 145077.58, 282574.31, 'New York'],
       [91749.16, 114175.79, 294919.57, 'Florida'],
       [86419.7, 153514.11, 0.0, 'New York'],
       [76253.86, 113867.3, 298664.47, 'California'],
       [78389.47, 153773.43, 299737.29, 'New York'],
       [73994.56, 122782.75, 303319.26, 'Florida'],
       [67532.53, 105751.03, 304768.73, 'Florida'],
       [77044.01, 99281.34, 140574.81, 'New York'],
       [64664.71, 139553.16, 137962.62, 'California'],
       [75328.87, 144135.98, 134050.07, 'Florida']])
```

0s completed at 10:40 PM

Files



+ Code + Text

sample_data
Startup.csv

```
from sklearn.compose import ColumnTransformer  
from sklearn.preprocessing import OneHotEncoder  
  
ct=ColumnTransformer(transformers=[('encoder',OneHotEncoder(),[3])],remainder='passthrough')  
X=ct.fit_transform(X)  
print(X)
```

```
[[0.0 0.0 1.0 165349.2 136897.8 471784.1]  
 [1.0 0.0 0.0 162597.7 151377.59 443898.53]  
 [0.0 1.0 0.0 153441.51 101145.55 407934.54]  
 [0.0 0.0 1.0 144372.41 118671.85 383199.62]  
 [0.0 1.0 0.0 142107.34 91391.77 366168.42]  
 [0.0 0.0 1.0 131876.9 99814.71 362861.36]  
 [1.0 0.0 0.0 134615.46 147198.87 127716.82]  
 [0.0 1.0 0.0 130298.13 145530.06 323876.68]  
 [0.0 0.0 1.0 120542.52 148718.95 311613.29]  
 [1.0 0.0 0.0 123334.88 108679.17 304981.62]  
 [0.0 1.0 0.0 101913.08 110594.11 229160.95]  
 [1.0 0.0 0.0 100671.96 91790.61 249744.55]  
 [0.0 1.0 0.0 93863.75 127320.38 249839.44]  
 [1.0 0.0 0.0 91992.39 135495.07 252664.93]  
 [0.0 1.0 0.0 119943.24 156547.42 256512.92]  
 [0.0 0.0 1.0 114523.61 122616.84 261776.23]  
 [1.0 0.0 0.0 78013.11 121597.55 264346.06]  
 [0.0 0.0 1.0 94657.16 145077.58 282574.31]  
 [0.0 1.0 0.0 91749.16 114175.79 294919.57]  
 [0.0 0.0 1.0 86419.7 153514.11 0.0]  
 [1.0 0.0 0.0 76253.86 113867.3 298664.47]  
 [0.0 0.0 1.0 78389.47 153773.43 299737.29]  
 [0.0 1.0 0.0 73994.56 122782.75 303319.26]  
 [0.0 1.0 0.0 67537.53 105751.02 301760.72]
```

0s completed at 10:42 PM

Startup

Show all

Type here to search



ML-Dev-Hackathon-2022-GPCE? x

Untitled1.ipynb - Colaboratory x

+

colab.research.google.com/drive/1KV1j9cH0IE5F_JLejoSvblvEhe46PtHi#scrollTo=K6IEgearm65T

Colab logo

Untitled1.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings User

RAM Disk

Editing

Files

sample_data

Startup.csv

68.20 GB available

0s completed at 10:41 PM

Startup

Type here to search

Taskbar icons

22:41 22-03-2022

Y

array([192261.83, 191792.06, 191050.39, 182901.99, 166187.94, 156991.12, 156122.51, 155752.6 , 152211.77, 149759.96, 146121.95, 144259.4 , 141585.52, 134307.35, 132602.65, 129917.04, 126992.93, 125370.37, 124266.9 , 122776.86, 118474.03, 111313.02, 110352.25, 108733.99, 108552.04, 107404.34, 105733.54, 105008.31, 103282.38, 101004.64, 99937.59, 97483.56, 97427.84, 96778.92, 96712.8 , 96479.51, 90708.19, 89949.14, 81229.06, 81005.76, 78239.91, 77798.83, 71498.49, 69758.98, 65200.33, 64926.08, 49490.75, 47559.73, 35673.41, 14681.4])

[4]

NameError Traceback (most recent call last)
 <ipython-input-4-b5fec669aca1> in <module>()
----> 1 X

NameError: name 'X' is not defined

SEARCH STACK OVERFLOW

Files

sample_data
Startup.csv

Disk 68.20 GB available

```
+ Code + Text
```

```
from sklearn.model_selection import train_test_split

X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2)

print(X_train.shape)
print(X_test.shape)
```

(40, 6)
(10, 6)

```
from sklearn.linear_model import LinearRegression

regressor=LinearRegression()
regressor.fit(X_train,Y_train)
```

LinearRegression()

ML-Dev-Hackathon-2022-GPCET

Untitled1.ipynb - Colaboratory

+

colab.research.google.com/drive/1KV1j9cHOIE5F_JLejoSvblvEhe46PtHi#scrollTo=Cn8_vDSunz4-

Untitled1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

sample_dataStartup.csv

+ Code + Text

RAM Disk

Editing

[10] Y_pred=regressor.predict(X_test)

from sklearn.metrics import r2_score
print(r2_score(Y_pred,Y_test))

0.8974082996331627

print(regressor.predict([[0.0, 0.0, 1.0, 86418.7, 153514.11, 0.0]]))

[119031.89986111]

0s completed at 10:44 PM

Show all

Startup

Type here to search

22:44
22-03-2022