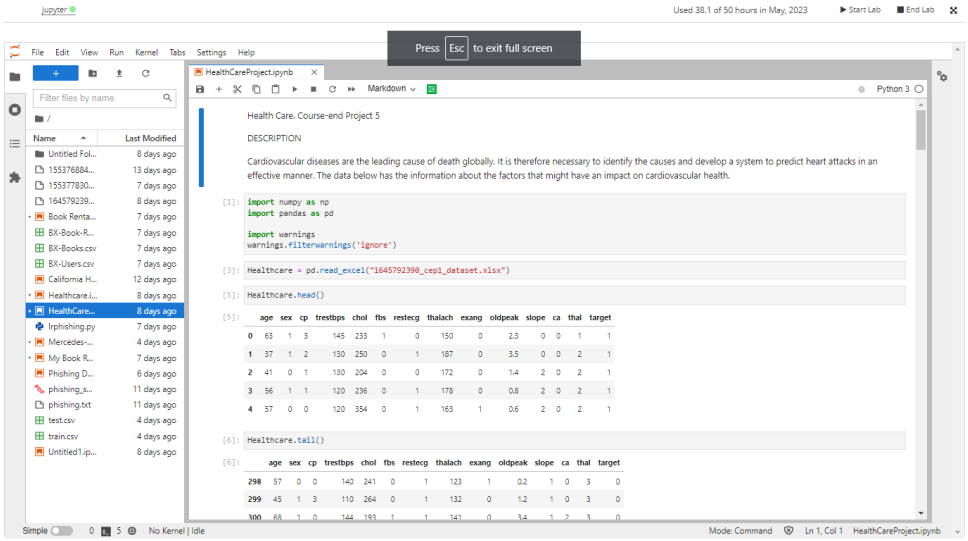
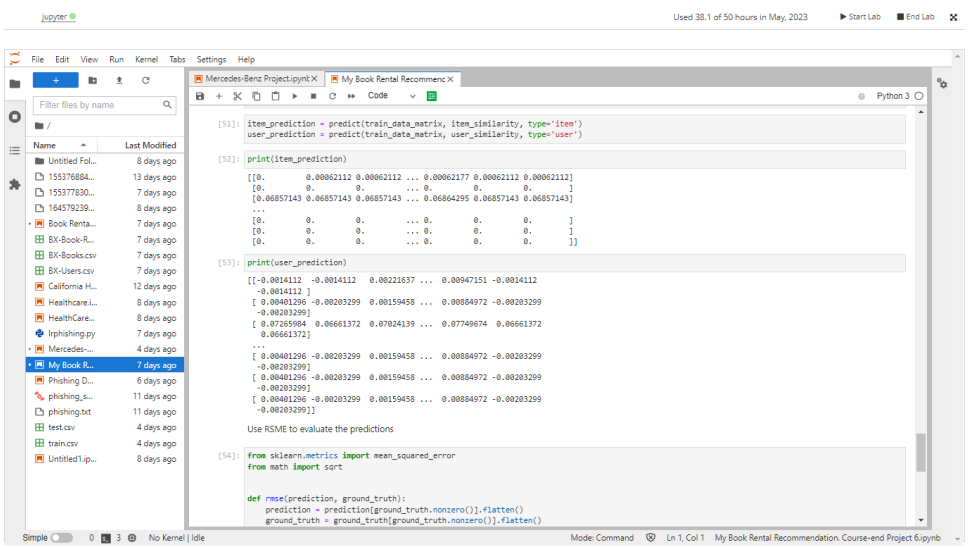
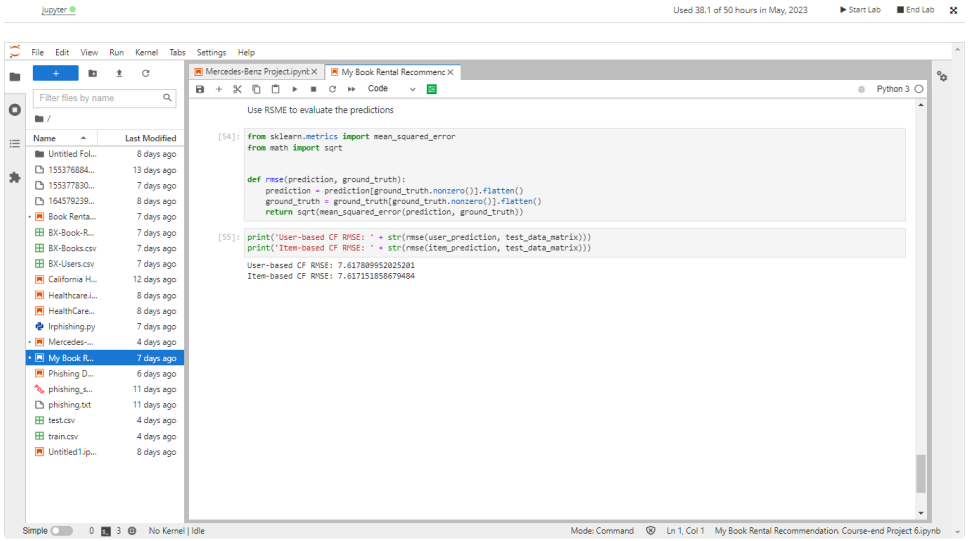
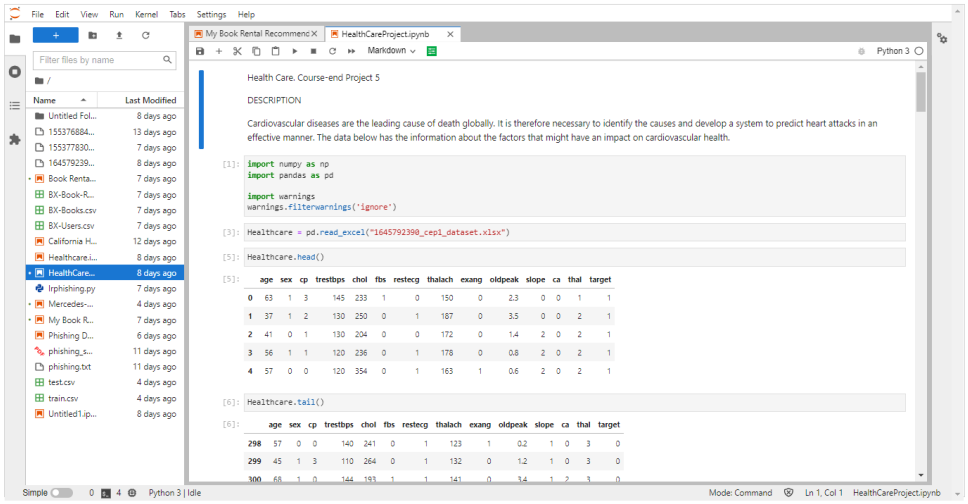
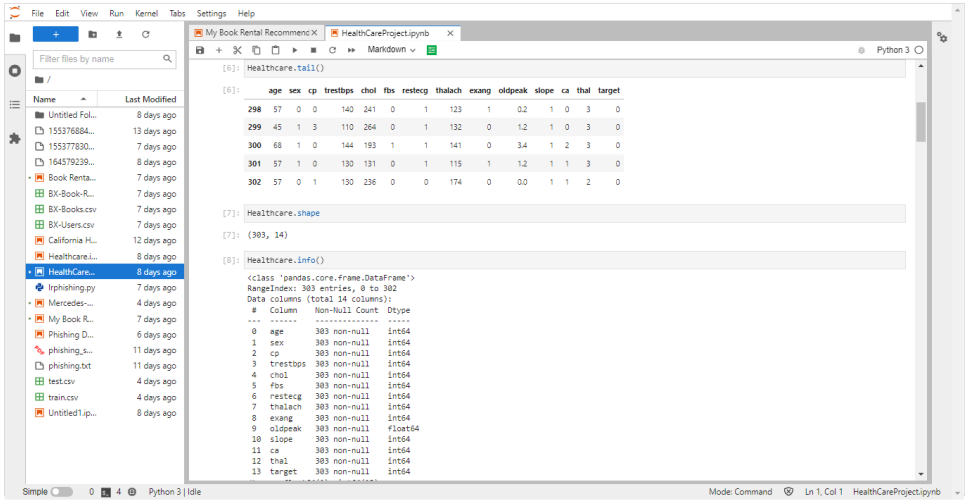
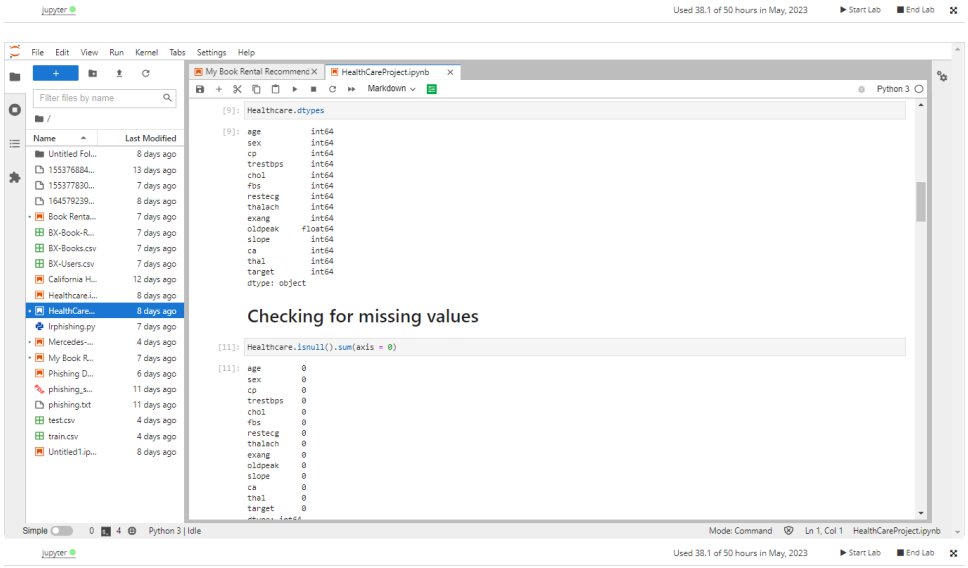

Jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
Press Esc to exit full screen
File Edit View Run Kernel Tabs Settings Help
HealthCareProject.ipynb
Markdown
Python 3
Health Care. Course-end Project 5
DESCRIPTION
Cardiovascular diseases are the leading cause of death globally. It is therefore necessary to identify the causes and develop a system to predict heart attacks in an effective manner. The data below has the information about the factors that might have an impact on cardiovascular health.
[1]: import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
[3]: Healthcare = pd.read_excel("1645792390_cep1_dataset.xlsx")
[5]: Healthcare.head()
[5]:
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
0 63 1 3 145 233 1 0 150 0 2.3 0 0 1 1
1 37 1 2 130 250 0 1 187 0 3.5 0 0 2 1
2 41 0 1 130 204 0 0 172 0 1.4 2 0 2 1
3 56 1 1 120 236 0 1 178 0 0.8 2 0 2 1
4 57 0 0 120 354 0 1 163 1 0.6 2 0 2 1
[6]: Healthcare.tail()
[6]:
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
298 57 0 0 140 241 0 1 123 1 0.2 1 0 3 0
299 45 1 3 110 264 0 1 132 0 1.2 1 0 3 0
Simple 0 5 No Kernel | Idle
Mode: Command Ln 1, Col 1 HealthCareProject.ipynb

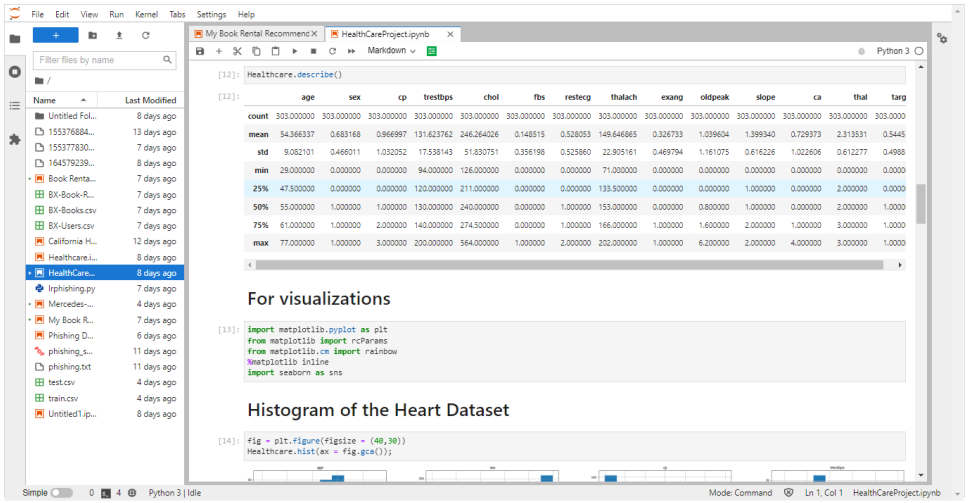

Jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
Mercedes-Benz Project.ipynb
My Book Rental Recommenc
Code
Python 3
[51]: item_prediction = predict(train_data_matrix, item_similarity, type='item')
user_prediction = predict(train_data_matrix, user_similarity, type='user')
[52]: print(item_prediction)
[[0. 0.00062112 0.00062112 ... 0.00062177 0.00062112 0.00062112]
[0. 0. 0. ... 0. 0. 0.]
[0.06857143 0.06857143 0.06857143 ... 0.06864295 0.06857143 0.06857143]
...
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]
[0. 0. 0. ... 0. 0. 0.]]
[53]: print(user_prediction)
[[-0.0014112 -0.0014112 0.00221637 ... 0.00947151 -0.0014112
-0.0014112]
[0.00401296 -0.00203299 0.00159458 ... 0.00884972 -0.00203299
-0.00203299]
[0.07265984 0.06661372 0.07024139 ... 0.07749674 0.06661372
0.06661372]
...
[0.00401296 -0.00203299 0.00159458 ... 0.00884972 -0.00203299
-0.00203299]
[0.00401296 -0.00203299 0.00159458 ... 0.00884972 -0.00203299
-0.00203299]
[0.00401296 -0.00203299 0.00159458 ... 0.00884972 -0.00203299
-0.00203299]]
Use RSME to evaluate the predictions
[54]: from sklearn.metrics import mean_squared_error
from math import sqrt
def rmse(prediction, ground_truth):
prediction = prediction[ground_truth.nonzero()].flatten()
ground_truth = ground_truth[ground_truth.nonzero()].flatten()
Simple 0 3 No Kernel | Idle
Mode: Command Ln 1, Col 1 My Book Rental Recommendation. Course-end Project 6.ipynb

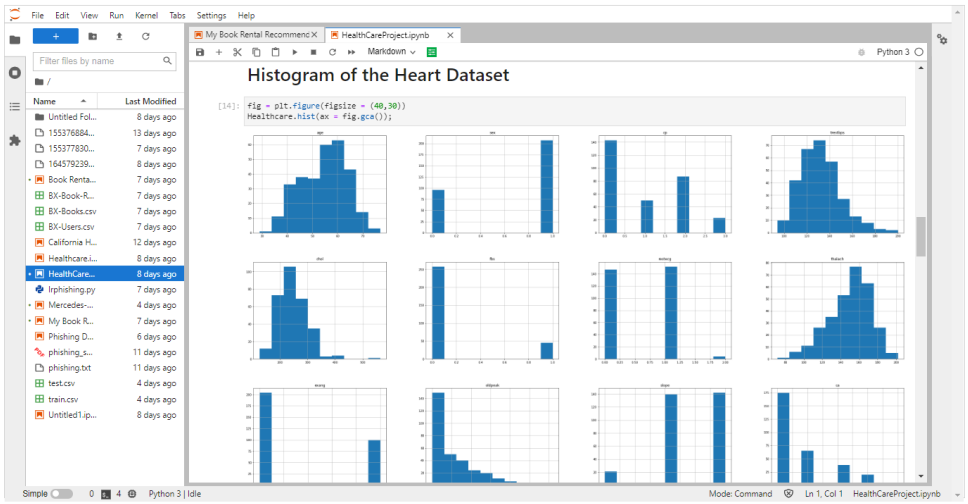

jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
Mercedes-Benz Project.ipynb
My Book Rental Recommenc
Code
Python 3
Use RSME to evaluate the predictions
[54]:
from sklearn.metrics import mean_squared_error
from math import sqrt
def rmse(prediction, ground_truth):
prediction = prediction[ground_truth.nonzero()].flatten()
ground_truth = ground_truth[ground_truth.nonzero()].flatten()
return sqrt(mean_squared_error(prediction, ground_truth))
[55]:
print('User-based CF RMSE: ' + str(rmse(user_prediction, test_data_matrix)))
print('Item-based CF RMSE: ' + str(rmse(item_prediction, test_data_matrix)))
User-based CF RMSE: 7.617809952025201
Item-based CF RMSE: 7.617151858679484
Simple
No Kernel | Idle
Mode: Command
Ln 1, Col 1
My Book Rental Recommendation. Course-end Project 6.ipynb

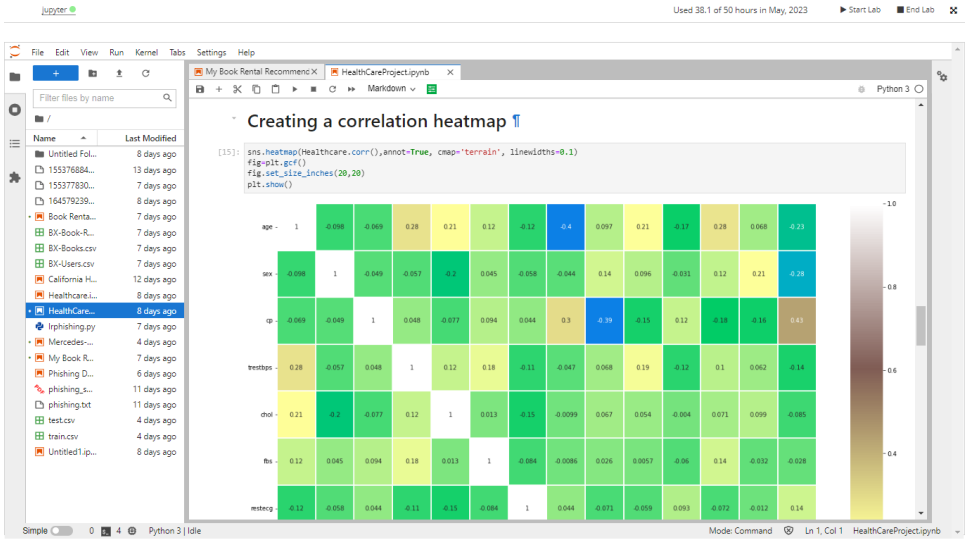

jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Health Care. Course-end Project 5
DESCRIPTION
Cardiovascular diseases are the leading cause of death globally. It is therefore necessary to identify the causes and develop a system to predict heart attacks in an effective manner. The data below has the information about the factors that might have an impact on cardiovascular health.
[1]:
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
[3]:
Healthcare = pd.read_excel("1645792390_cep1_dataset.xlsx")
[5]:
Healthcare.head()
[5]:
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
0 63 1 3 145 233 1 0 150 0 2.3 0 0 1 1
1 37 1 2 130 250 0 1 187 0 3.5 0 0 2 1
2 41 0 1 130 204 0 0 172 0 1.4 2 0 2 1
3 56 1 1 120 236 0 1 178 0 0.8 2 0 2 1
4 57 0 0 120 354 0 1 163 1 0.6 2 0 2 1
[6]:
Healthcare.tail()
[6]:
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
298 57 0 0 140 241 0 1 123 1 0.2 1 0 3 0
299 45 1 3 110 264 0 1 132 0 1.2 1 0 3 0
300 68 1 0 144 193 1 1 141 0 3.4 1 2 3 0
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb

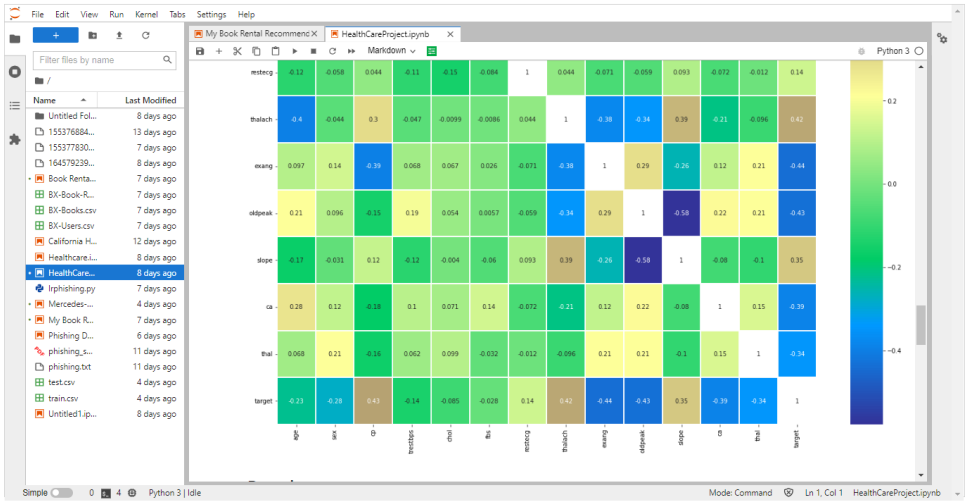

jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
[6]:
Healthcare.tail()
[6]:
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
298 57 0 0 140 241 0 1 123 1 0.2 1 0 3 0
299 45 1 3 110 264 0 1 132 0 1.2 1 0 3 0
300 68 1 0 144 193 1 1 141 0 3.4 1 2 3 0
301 57 1 0 130 131 0 1 115 1 1.2 1 1 3 0
302 57 0 1 130 236 0 0 174 0 0.0 1 1 2 0
[7]:
Healthcare.shape
[7]:
(303, 14)
[8]:
Healthcare.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
Column Non-Null Count Dtype
0 age 303 non-null int64
1 sex 303 non-null int64
2 cp 303 non-null int64
3 trestbps 303 non-null int64
4 chol 303 non-null int64
5 fbs 303 non-null int64
6 restecg 303 non-null int64
7 thalach 303 non-null int64
8 exang 303 non-null int64
9 oldpeak 303 non-null float64
10 slope 303 non-null int64
11 ca 303 non-null int64
12 thal 303 non-null int64
13 target 303 non-null int64
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb

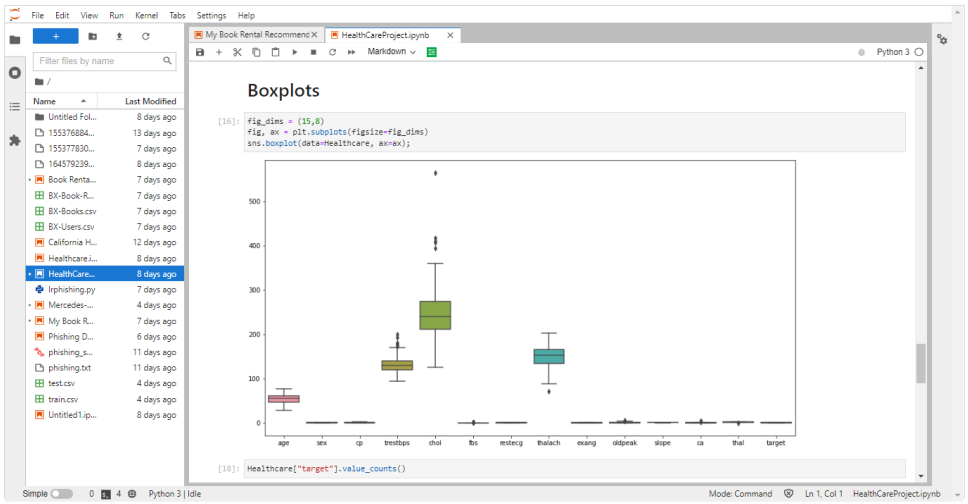

Jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Filter files by name
Name
Last Modified
[9]: Healthcare.dtypes
age int64
sex int64
cp int64
trestbps int64
chol int64
fbs int64
restecg int64
thalach int64
exang int64
oldpeak float64
slope int64
ca int64
thal int64
target int64
dtype: object
Checking for missing values
[11]: Healthcare.isnull().sum(axis = 0)
age 0
sex 0
cp 0
trestbps 0
chol 0
fbs 0
restecg 0
thalach 0
exang 0
oldpeak 0
slope 0
ca 0
thal 0
target 0
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb


Jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Filter files by name
Name
Last Modified
[12]: Healthcare.describe()
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal targ
count 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.000000 303.0000
mean 54.366337 0.683168 0.966997 131.623762 246.264026 0.148515 0.528053 149.646865 0.326733 1.039604 1.399340 0.729373 2.313531 0.5445
std 9.082101 0.466011 1.032052 17.538143 51.830751 0.356198 0.525860 22.905161 0.469794 1.161075 0.616226 1.022606 0.612277 0.4988
min 29.000000 0.000000 0.000000 94.000000 126.000000 0.000000 0.000000 71.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.0000
25% 47.500000 0.000000 0.000000 120.000000 211.000000 0.000000 0.000000 133.500000 0.000000 0.000000 1.000000 0.000000 2.000000 0.0000
50% 55.000000 1.000000 1.000000 130.000000 240.000000 0.000000 1.000000 153.000000 0.000000 0.800000 1.000000 0.000000 2.000000 1.0000
75% 61.000000 1.000000 2.000000 140.000000 274.500000 0.000000 1.000000 166.000000 1.000000 1.600000 2.000000 1.000000 3.000000 1.0000
max 77.000000 1.000000 3.000000 200.000000 564.000000 1.000000 2.000000 202.000000 1.000000 6.200000 2.000000 4.000000 3.000000 1.0000
For visualizations
[13]: import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
%matplotlib inline
import seaborn as sns
Histogram of the Heart Dataset
[14]: fig = plt.figure(figsize = (40,30))
Healthcare.hist(ax = fig.gca());
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb

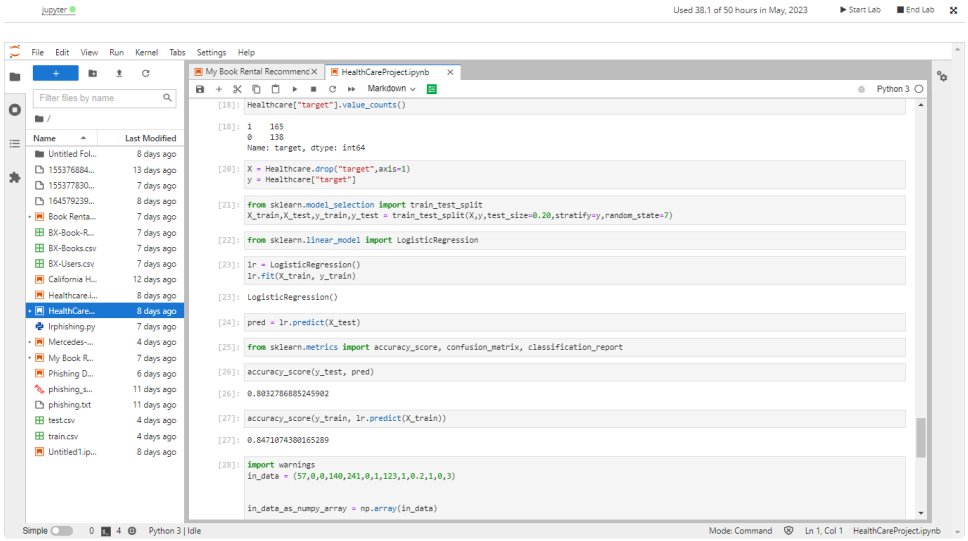

Jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Filter files by name
Name
Last Modified
Histogram of the Heart Dataset
[14]: fig = plt.figure(figsize = (40,30))
Healthcare.hist(ax = fig.gca());
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb

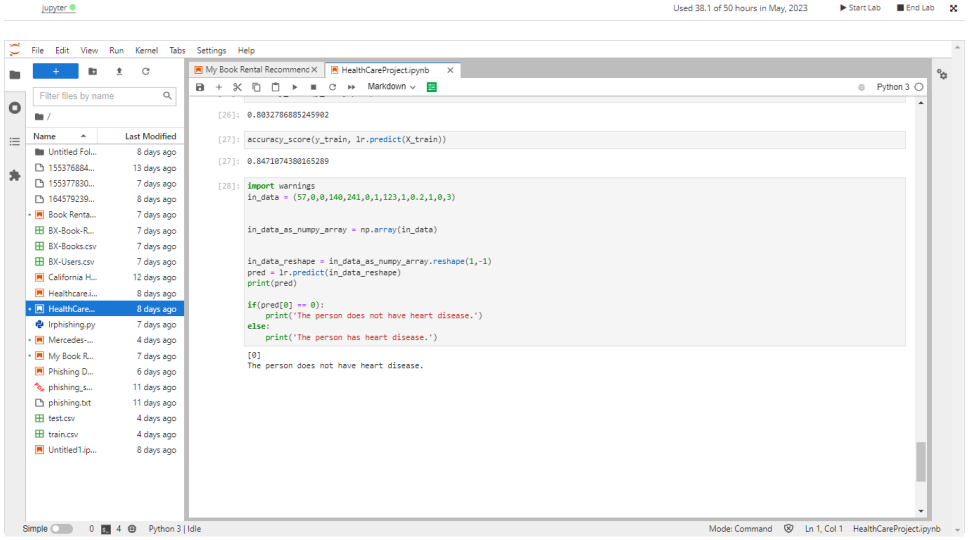

Creating a correlation heatmap ¶
[15]:
sns.heatmap(Healthcare.corr(),annot=True, cmap='terrain', linewidths=0.1)
fig=plt.gcf()
fig.set_size_inches(20,20)
plt.show()





Boxplots
[16]:
fig_dims = (15,8)
fig, ax = plt.subplots(figsize=fig_dims)
sns.boxplot(data=Healthcare, ax=ax);
[18]:
Healthcare["target"].value_counts()


jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Filter files by name
Name
Last Modified
Untitled Fol... 8 days ago
155376884... 13 days ago
155377830... 7 days ago
164579239... 8 days ago
Book Renta... 7 days ago
BX-Book-R... 7 days ago
BX-Books.csv 7 days ago
BX-Users.csv 7 days ago
California H... 12 days ago
Healthcare.i... 8 days ago
HealthCare... 8 days ago
Irphishing.py 7 days ago
Mercedes-... 4 days ago
My Book R... 7 days ago
Phishing D... 6 days ago
phishing_s... 11 days ago
phishing.txt 11 days ago
test.csv 4 days ago
train.csv 4 days ago
Untitled1.ip... 8 days ago
[18]: Healthcare["target"].value_counts()
[18]: 1 165
0 138
Name: target, dtype: int64
[20]: X = Healthcare.drop("target",axis=1)
y = Healthcare["target"]
[21]: from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20,stratify=y,random_state=7)
[22]: from sklearn.linear_model import LogisticRegression
[23]: lr = LogisticRegression()
lr.fit(X_train, y_train)
[23]: LogisticRegression()
[24]: pred = lr.predict(X_test)
[25]: from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
[26]: accuracy_score(y_test, pred)
[26]: 0.8032786885245902
[27]: accuracy_score(y_train, lr.predict(X_train))
[27]: 0.8471074380165289
[28]: import warnings
in_data = (57,0,0,140,241,0,1,123,1,0.2,1,0,3)
in_data_as_numpy_array = np.array(in_data)
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb


jupyter
Used 38.1 of 50 hours in May, 2023
Start Lab
End Lab
File Edit View Run Kernel Tabs Settings Help
My Book Rental Recommenc
HealthCareProject.ipynb
Markdown
Python 3
Filter files by name
Name
Last Modified
Untitled Fol... 8 days ago
155376884... 13 days ago
155377830... 7 days ago
164579239... 8 days ago
Book Renta... 7 days ago
BX-Book-R... 7 days ago
BX-Books.csv 7 days ago
BX-Users.csv 7 days ago
California H... 12 days ago
Healthcare.i... 8 days ago
HealthCare... 8 days ago
Irphishing.py 7 days ago
Mercedes-... 4 days ago
My Book R... 7 days ago
Phishing D... 6 days ago
phishing_s... 11 days ago
phishing.txt 11 days ago
test.csv 4 days ago
train.csv 4 days ago
Untitled1.ip... 8 days ago
[26]: 0.8032786885245902
[27]: accuracy_score(y_train, lr.predict(X_train))
[27]: 0.8471074380165289
[28]: import warnings
in_data = (57,0,0,140,241,0,1,123,1,0.2,1,0,3)
in_data_as_numpy_array = np.array(in_data)
in_data_reshape = in_data_as_numpy_array.reshape(1,-1)
pred = lr.predict(in_data_reshape)
print(pred)
if(pred[0] == 0):
print('The person does not have heart disease.')
else:
print('The person has heart disease.')
[0]
The person does not have heart disease.
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
HealthCareProject.ipynb