

Home

Healthcare Appointment No-Sho

+

localhost:8888/notebooks/Healthcare%20Appointment%20No-Show%20Prediction.ipynb?

jupyter Healthcare Appointment No-Show Prediction Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help

Trusted

JupyterLab Python [conda env:base] *

HEALTHCARE APPOINTMENT NO-SHOW PRECTION

```
[9]: import pandas as pd
import numpy as np
from datetime import datetime

# Visualization
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

# Machine Learning
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, roc_auc_score
```

Import the Dataset

```
[11]: df = pd.read_csv('KaggleV2-May-2016.csv')
```

```
[13]: df
```

7

Search

ENG US 17:39 04-07-2025

Home

Healthcare Appointment No-Sho

+

localhost:8888/notebooks/Healthcare%20Appointment%20No-Show%20Prediction.ipynb?

Jupyter Healthcare Appointment No-Show Prediction Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help

Trusted

JupyterLab Python [conda env:base]

Convert dates

```
[17]: df['ScheduledDay'] = pd.to_datetime(df['ScheduledDay'])
df['ScheduledDay']

[17]: 0      2016-04-29 18:38:08+00:00
1      2016-04-29 16:08:27+00:00
2      2016-04-29 16:19:04+00:00
3      2016-04-29 17:29:31+00:00
4      2016-04-29 16:07:23+00:00
...
110522 2016-05-03 09:15:35+00:00
110523 2016-05-03 07:27:33+00:00
110524 2016-04-27 16:03:52+00:00
110525 2016-04-27 15:09:23+00:00
110526 2016-04-27 13:30:56+00:00
Name: ScheduledDay, Length: 110527, dtype: datetime64[ns, UTC]

[19]: df['AppointmentDay'] = pd.to_datetime(df['AppointmentDay'])
df['AppointmentDay']

[19]: 0      2016-04-29 00:00:00+00:00
1      2016-04-29 00:00:00+00:00
2      2016-04-29 00:00:00+00:00
3      2016-04-29 00:00:00+00:00
4      2016-04-29 00:00:00+00:00
...
110522 2016-05-07 00:00:00+00:00
110523 2016-05-07 00:00:00+00:00
110524 2016-05-07 00:00:00+00:00
110525 2016-05-07 00:00:00+00:00
110526 2016-05-07 00:00:00+00:00
Name: AppointmentDay, Length: 110527, dtype: datetime64[ns, UTC]

[21]: df

[21]: PatientID AppointmentID Gender ScheduledDay AppointmentDay Age Neighbourhood Scholarship Hipertension Diabetes Alcoholism Handcap SM
```

7

Search

ENG US 17:40 04-07-2025

Home

Healthcare Appointment No-Sho

+

localhost:8888/notebooks/Healthcare%20Appointment%20No-Show%20Prediction.ipynb?

Jupyter Healthcare Appointment No-Show Prediction Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help Trusted

JupyterLab Python [conda env:base] *

Calculate waiting time (days between scheduling and appointment)

```
[23]: # Feature engineering
df['WaitingDays'] = (df['AppointmentDay'] - df['ScheduledDay']).dt.days
df['WaitingDays']
```

```
[23]: 0      -1
      1      -1
      2      -1
      3      -1
      4      -1
      ..
110522  34
110523  34
110524  40
110525  40
110526  40
Name: WaitingDays, Length: 110527, dtype: int64
```

```
[25]: df
```

```
[25]:
```

| | PatientId | AppointmentID | Gender | ScheduledDay | AppointmentDay | Age | Neighbourhood | Scholarship | Hipertension | Diabetes | Alcoholism | Handcap | SM |
|-----|--------------|---------------|--------|---------------------------|---------------------------|-----|-------------------|-------------|--------------|----------|------------|---------|-----|
| 0 | 2.987250e+13 | 5642903 | F | 2016-04-29 18:38:08+00:00 | 2016-04-29 00:00:00+00:00 | 62 | JARDIM DA PENHA | 0 | 1 | 0 | 0 | 0 | |
| 1 | 5.589978e+14 | 5642503 | M | 2016-04-29 16:08:27+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 0 | 0 | 0 | 0 | |
| 2 | 4.262962e+12 | 5642549 | F | 2016-04-29 16:19:04+00:00 | 2016-04-29 00:00:00+00:00 | 62 | MATA DA PRAIA | 0 | 0 | 0 | 0 | 0 | |
| 3 | 8.679512e+11 | 5642828 | F | 2016-04-29 17:29:31+00:00 | 2016-04-29 00:00:00+00:00 | 8 | PONTAL DE CAMBURI | 0 | 0 | 0 | 0 | 0 | |
| 4 | 8.841186e+12 | 5642494 | F | 2016-04-29 16:07:23+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 1 | 1 | 0 | 0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Search

ENG US

17:40
04-07-2025

Extract day of the week (0=Monday, 6=Sunday)

```
[27]: df['AppointmentDayOfWeek'] = df['AppointmentDay'].dt.dayofweek
df['AppointmentDayOfWeek']
```

```
[27]: 0      4
      1      4
      2      4
      3      4
      4      4
      ..
110522 1
110523 1
110524 1
110525 1
110526 1
Name: AppointmentDayOfWeek, Length: 110527, dtype: int32
```

```
[29]: df
```

| | PatientId | AppointmentID | Gender | ScheduledDay | AppointmentDay | Age | Neighbourhood | Scholarship | Hipertension | Diabetes | Alcoholism | Handcap | SM |
|--------|--------------|---------------|--------|------------------------------|------------------------------|-----|----------------------|-------------|--------------|----------|------------|---------|----|
| 0 | 2.987250e+13 | 5642903 | F | 2016-04-29 18:38:08+00:00 | 2016-04-29 00:00:00+00:00 | 62 | JARDIM DA PENHA | 0 | 1 | 0 | 0 | 0 | |
| 1 | 5.589970e+14 | 5642503 | M | 2016-04-29 16:08:27+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 0 | 0 | 0 | 0 | |
| 2 | 4.262962e+12 | 5642549 | F | 2016-04-29 16:19:04+00:00 | 2016-04-29 00:00:00+00:00 | 62 | MATA DA PRAIA | 0 | 0 | 0 | 0 | 0 | |
| 3 | 8.679512e+11 | 5642028 | F | 2016-04-29 17:29:31+00:00 | 2016-04-29 00:00:00+00:00 | 8 | PONTAL DE CAMBURI | 0 | 0 | 0 | 0 | 0 | |
| 4 | 8.841180e+12 | 5642494 | F | 2016-04-29 16:07:23+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 1 | 1 | 0 | 0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 110522 | 2.572134e+12 | 5651768 | F | 2016-05-03 09:15:35+00:00 | 2016-06-07 00:00:00+00:00 | 56 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |

Home

Healthcare Appointment No-Sho

+

localhost:8888/notebooks/Healthcare%20Appointment%20No-Show%20Prediction.ipynb?

Healthcare Appointment No-Sho Prediction Last Checkpoint: 2 days ago

File Edit View Run Kernel Settings Help

Trusted

JupyterLab Python [conda env:base] *

Convert "No-show" to binary (1 for no-show, 0 for showed)

```
[31]: df['NoShow'] = df['No-show'].apply(lambda x: 1 if x == 'Yes' else 0)
df['NoShow']
```

```
[31]: 0      0
1      0
2      0
3      0
4      0
...
110522 0
110523 0
110524 0
110525 0
110526 0
Name: NoShow, Length: 110527, dtype: int64
```

```
[33]: df
```

```
[33]:
```

| | PatientId | AppointmentID | Gender | ScheduledDay | AppointmentDay | Age | Neighbourhood | Scholarship | Hipertension | Diabetes | Alcoholism | Handcap | SM |
|--------|--------------|---------------|--------|------------------------------|------------------------------|-----|----------------------|-------------|--------------|----------|------------|---------|----|
| 0 | 2.987250e+13 | 5642903 | F | 2016-04-29 18:38:08+00:00 | 2016-04-29 00:00:00+00:00 | 62 | JARDIM DA PENHA | 0 | 1 | 0 | 0 | 0 | |
| 1 | 5.589978e+14 | 5642503 | M | 2016-04-29 16:08:27+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 0 | 0 | 0 | 0 | |
| 2 | 4.262962e+12 | 5642549 | F | 2016-04-29 16:19:04+00:00 | 2016-04-29 00:00:00+00:00 | 62 | MATA DA PRAIA | 0 | 0 | 0 | 0 | 0 | |
| 3 | 8.679512e+11 | 5642828 | F | 2016-04-29 17:29:31+00:00 | 2016-04-29 00:00:00+00:00 | 8 | PONTAL DE CAMBURI | 0 | 0 | 0 | 0 | 0 | |
| 4 | 8.841186e+12 | 5642494 | F | 2016-04-29 16:07:23+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 1 | 1 | 0 | 0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 110522 | 2.572134e+12 | 5651768 | F | 2016-05-03 00:00:00+00:00 | 2016-06-07 00:00:00+00:00 | 56 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |

7

Search

ENG US 17:40 04-07-2025

Save the cleaned data for Power BI

```
[39]: df.to_csv('processed_appointments.csv', index=False)

[41]: processed_appointment = pd.read_csv('processed_appointments.csv')
processed_appointment

[41]:
```

| | PatientId | AppointmentID | Gender | ScheduledDay | AppointmentDay | Age | Neighbourhood | Scholarship | Hipertension | Diabetes | Alcoholism | Handcap | SM |
|--------|--------------|---------------|--------|---------------------------|---------------------------|-----|-------------------|-------------|--------------|----------|------------|---------|-----|
| 0 | 2.987250e+13 | 5642903 | F | 2016-04-29 18:38:08+00:00 | 2016-04-29 00:00:00+00:00 | 62 | JARDIM DA PENHA | 0 | 1 | 0 | 0 | 0 | |
| 1 | 5.589978e+14 | 5642503 | M | 2016-04-29 16:08:27+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 0 | 0 | 0 | 0 | |
| 2 | 4.262962e+12 | 5642549 | F | 2016-04-29 16:19:04+00:00 | 2016-04-29 00:00:00+00:00 | 62 | MATA DA PRAIA | 0 | 0 | 0 | 0 | 0 | |
| 3 | 8.679512e+11 | 5642828 | F | 2016-04-29 17:29:31+00:00 | 2016-04-29 00:00:00+00:00 | 8 | PONTAL DE CAMBURI | 0 | 0 | 0 | 0 | 0 | |
| 4 | 8.841186e+12 | 5642494 | F | 2016-04-29 16:07:23+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 1 | 1 | 0 | 0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 110522 | 2.572134e+12 | 5651768 | F | 2016-05-03 09:15:35+00:00 | 2016-06-07 00:00:00+00:00 | 56 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |
| 110523 | 3.596266e+12 | 5650093 | F | 2016-05-03 07:27:33+00:00 | 2016-06-07 00:00:00+00:00 | 51 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |
| 110524 | 1.557663e+13 | 5630692 | F | 2016-04-27 16:03:52+00:00 | 2016-06-07 00:00:00+00:00 | 21 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |
| 110525 | 9.213493e+13 | 5630323 | F | 2016-04-27 15:09:23+00:00 | 2016-06-07 00:00:00+00:00 | 38 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |
| 110526 | 3.775115e+14 | 5629448 | F | 2016-04-27 13:30:56+00:00 | 2016-06-07 00:00:00+00:00 | 54 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 | |

jupyter

Healthcare Appointment No-Show Prediction

Last Checkpoint: 2 days ago

File

Edit

View

Run

Kernel

Settings

Help

Trusted

Code

JupyterLab

Python [conda env:base]

| | | | | | | | | | | | | |
|--------|--------------|---------|-----|------------------------------|------------------------------|-----|----------------------|-----|-----|-----|-----|-----|
| 2 | 4.262962e+12 | 5642549 | F | 16:19:04+00:00 | 00:00:00+00:00 | 62 | MAIA DA PRAIA | 0 | 0 | 0 | 0 | 0 |
| 3 | 8.679512e+11 | 5642828 | F | 2016-04-29 17:29:31+00:00 | 2016-04-29 00:00:00+00:00 | 8 | PONTAL DE CAMBURI | 0 | 0 | 0 | 0 | 0 |
| 4 | 8.841186e+12 | 5642494 | F | 2016-04-29 16:07:23+00:00 | 2016-04-29 00:00:00+00:00 | 56 | JARDIM DA PENHA | 0 | 1 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 110522 | 2.572134e+12 | 5651768 | F | 2016-05-03 09:15:35+00:00 | 2016-06-07 00:00:00+00:00 | 56 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 |
| 110523 | 3.596266e+12 | 5650093 | F | 2016-05-03 07:27:33+00:00 | 2016-06-07 00:00:00+00:00 | 51 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 |
| 110524 | 1.557663e+13 | 5630692 | F | 2016-04-27 16:03:52+00:00 | 2016-06-07 00:00:00+00:00 | 21 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 |
| 110525 | 9.213493e+13 | 5630323 | F | 2016-04-27 15:09:23+00:00 | 2016-06-07 00:00:00+00:00 | 38 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 |
| 110526 | 3.775115e+14 | 5629448 | F | 2016-04-27 13:30:56+00:00 | 2016-06-07 00:00:00+00:00 | 54 | MARIA ORTIZ | 0 | 0 | 0 | 0 | 0 |

110527 rows × 17 columns

[2]:

import os

print(os.getcwd())

D:\Healthcare Appointment prediction