In [64]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
!pip install xgboost
from xgboost import XGBClassifier



from sklearn.preprocessing import StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer


!pip install nbconvert[webpdf]
!playwright install
```

```
Requirement already satisfied: xgboost in c:\users\pablo\appdata\local\abf certifica
te\python\python\lib\site-packages (3.1.2)
Requirement already satisfied: numpy in c:\users\pablo\appdata\local\abf certificate
\python\python\lib\site-packages (from xgboost) (1.26.1)
Requirement already satisfied: scipy in c:\users\pablo\appdata\local\abf certificate
\python\python\lib\site-packages (from xgboost) (1.11.3)
Requirement already satisfied: nbconvert[webpdf] in c:\users\pablo\appdata\local\abf
certificate\python\python\lib\site-packages (7.16.6)
Requirement already satisfied: beautifulsoup4 in c:\users\pablo\appdata\local\abf ce
rtificate\python\python\lib\site-packages (from nbconvert[webpdf]) (4.14.2)
Requirement already satisfied: bleach!=5.0.0 in c:\users\pablo\appdata\local\abf cer
tificate\python\python\lib\site-packages (from bleach[css]!=5.0.0->nbconvert[webpd
f]) (6.2.0)
Requirement already satisfied: defusedxml in c:\users\pablo\appdata\local\abf certif
icate\python\python\lib\site-packages (from nbconvert[webpdf]) (0.7.1)
Requirement already satisfied: jinja2>=3.0 in c:\users\pablo\appdata\local\abf certi
ficate\python\python\lib\site-packages (from nbconvert[webpdf]) (3.1.6)
Requirement already satisfied: jupyter-core>=4.7 in c:\users\pablo\appdata\local\abf
certificate\python\python\lib\site-packages (from nbconvert[webpdf]) (5.9.1)
Requirement already satisfied: jupyterlab-pygments in c:\users\pablo\appdata\local\a
bf certificate\python\python\lib\site-packages (from nbconvert[webpdf]) (0.3.0)
Requirement already satisfied: markupsafe>=2.0 in c:\users\pablo\appdata\local\abf c
ertificate\python\python\lib\site-packages (from nbconvert[webpdf]) (3.0.3)
Requirement already satisfied: mistune<4,>=2.0.3 in c:\users\pablo\appdata\local\abf
certificate\python\python\lib\site-packages (from nbconvert[webpdf]) (3.1.4)
Requirement already satisfied: nbclient>=0.5.0 in c:\users\pablo\appdata\local\abf c
ertificate\python\python\lib\site-packages (from nbconvert[webpdf]) (0.10.2)
Requirement already satisfied: nbformat>=5.7 in c:\users\pablo\appdata\local\abf cer
tificate\python\python\lib\site-packages (from nbconvert[webpdf]) (5.10.4)
Requirement already satisfied: packaging in c:\users\pablo\appdata\local\abf certifi
cate\python\python\lib\site-packages (from nbconvert[webpdf]) (25.0)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\pablo\appdata\local
\abf certificate\python\python\lib\site-packages (from nbconvert[webpdf]) (1.5.1)
Requirement already satisfied: pygments>=2.4.1 in c:\users\pablo\appdata\local\abf c
ertificate\python\python\lib\site-packages (from nbconvert[webpdf]) (2.19.2)
Requirement already satisfied: traitlets>=5.1 in c:\users\pablo\appdata\local\abf ce
rtificate\python\python\lib\site-packages (from nbconvert[webpdf]) (5.14.3)
Collecting playwright (from nbconvert[webpdf])
  Downloading playwright-1.57.0-py3-none-win_amd64.whl.metadata (3.5 kB)
Requirement already satisfied: webencodings in c:\users\pablo\appdata\local\abf cert
ificate\python\python\lib\site-packages (from bleach!=5.0.0->bleach[css]!=5.0.0->nbc
onvert[webpdf]) (0.5.1)
Requirement already satisfied: tinycss2<1.5,>=1.1.0 in c:\users\pablo\appdata\local
\abf certificate\python\python\lib\site-packages (from bleach[css]!=5.0.0->nbconvert
[webpdf]) (1.4.0)
Requirement already satisfied: platformdirs>=2.5 in c:\users\pablo\appdata\local\abf
certificate\python\python\lib\site-packages (from jupyter-core>=4.7->nbconvert[webpd
f]) (4.5.0)
Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\pablo\appdata\loca
l\abf certificate\python\python\lib\site-packages (from nbclient>=0.5.0->nbconvert[w
ebpdf]) (8.6.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\pablo\appdata\loca
l\abf certificate\python\python\lib\site-packages (from jupyter-client>=6.1.12->nbcl
ient>=0.5.0->nbconvert[webpdf]) (2.9.0.post0)
Requirement already satisfied: pyzmq>=23.0 in c:\users\pablo\appdata\local\abf certi
ficate\python\python\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0
```

```
->nbconvert[webpdf]) (27.1.0)
Requirement already satisfied: tornado>=6.2 in c:\users\pablo\appdata\local\abf cert
ificate\python\python\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.
0->nbconvert[webpdf]) (6.5.2)
Requirement already satisfied: fastjsonschema>=2.15 in c:\users\pablo\appdata\local
\abf certificate\python\python\lib\site-packages (from nbformat>=5.7->nbconvert[webp
df]) (2.21.2)
Requirement already satisfied: jsonschema>=2.6 in c:\users\pablo\appdata\local\abf c
ertificate\python\python\lib\site-packages (from nbformat>=5.7->nbconvert[webpdf])
(4.25.1)
Requirement already satisfied: attrs>=22.2.0 in c:\users\pablo\appdata\local\abf cer
tificate\python\python\lib\site-packages (from jsonschema>=2.6->nbformat>=5.7->nbcon
vert[webpdf]) (25.4.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\pabl
o\appdata\local\abf certificate\python\python\lib\site-packages (from jsonschema>=2.
6->nbformat>=5.7->nbconvert[webpdf]) (2025.9.1)
Requirement already satisfied: referencing>=0.28.4 in c:\users\pablo\appdata\local\a
bf certificate\python\python\lib\site-packages (from jsonschema>=2.6->nbformat>=5.7-
>nbconvert[webpdf]) (0.37.0)
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\pablo\appdata\local\abf ce
rtificate\python\python\lib\site-packages (from jsonschema>=2.6->nbformat>=5.7->nbco
nvert[webpdf]) (0.28.0)
Requirement already satisfied: six>=1.5 in c:\users\pablo\appdata\local\abf certific
ate\python\python\lib\site-packages (from python-dateutil>=2.8.2->jupyter-client>=6.
1.12->nbclient>=0.5.0->nbconvert[webpdf]) (1.17.0)
Requirement already satisfied: typing-extensions>=4.4.0 in c:\users\pablo\appdata\lo
cal\abf certificate\python\python\lib\site-packages (from referencing>=0.28.4->jsons
chema>=2.6->nbformat>=5.7->nbconvert[webpdf]) (4.15.0)
Requirement already satisfied: soupsieve>1.2 in c:\users\pablo\appdata\local\abf cer
tificate\python\python\lib\site-packages (from beautifulsoup4->nbconvert[webpdf])
(2.8)
Collecting pyee<14,>=13 (from playwright->nbconvert[webpdf])
  Downloading pyee-13.0.0-py3-none-any.whl.metadata (2.9 kB)
Collecting greenlet<4.0.0,>=3.1.1 (from playwright->nbconvert[webpdf])
  Downloading greenlet-3.3.0-cp312-cp312-win_amd64.whl.metadata (4.2 kB)
Downloading playwright-1.57.0-py3-none-win_amd64.whl (36.6 MB)
   ---------------------------------------- 0.0/36.6 MB ? eta -:--:--
   ---------- ----------------------------- 9.4/36.6 MB 58.8 MB/s eta 0:00:01
   -------------------------------- -------- 28.6/36.6 MB 75.5 MB/s eta 0:00:01
   ---------------------------------------- - 35.4/36.6 MB 62.4 MB/s eta 0:00:01
   ---------------------------------------  36.4/36.6 MB 57.9 MB/s eta 0:00:01
   ---------------------------------------- 36.6/36.6 MB 42.2 MB/s  0:00:00
Downloading greenlet-3.3.0-cp312-cp312-win_amd64.whl (301 kB)
Downloading pyee-13.0.0-py3-none-any.whl (15 kB)
Installing collected packages: pyee, greenlet, playwright

   ---------------------------------------- 0/3 [pyee]
   ---------------------------------------- 0/3 [pyee]
   ---------------------------------------- 0/3 [pyee]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
   ------------ --------------------------- 1/3 [greenlet]
```

```
   ------------- -------------------------- 1/3 [greenlet]
   ------------- -------------------------- 1/3 [greenlet]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   -------------------------- ------------- 2/3 [playwright]
   ---------------------------------------- 3/3 [playwright]

Successfully installed greenlet-3.3.0 playwright-1.57.0 pyee-13.0.0
Downloading Chromium 143.0.7499.4 (playwright build v1200) from https://cdn.playwrig
ht.dev/dbazure/download/playwright/builds/chromium/1200/chromium-win64.zip
```

```
|                                                                                |
0% of 169.8 MiB
|â– â– â– â– â– â– â– â–
|  10% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  20% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  30% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                                 |  40% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                         |  50%
of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  60% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  70% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                 |  80% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  90% of 169.8 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of 1
69.8 MiB
Chromium 143.0.7499.4 (playwright build v1200) downloaded to C:\Users\pablo\AppData
\Local\ms-playwright\chromium-1200
Downloading Chromium Headless Shell 143.0.7499.4 (playwright build v1200) from http
s://cdn.playwright.dev/dbazure/download/playwright/builds/chromium/1200/chromium-hea
dless-shell-win64.zip
|                                                                                |
0% of 107.2 MiB
|â– â– â– â– â– â– â– â–
|  10% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  20% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  30% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                                 |  40% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                         |  50%
of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  60% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  70% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                 |  80% of 107.2 MiB
```

```
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  90% of 107.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of 1
07.2 MiB
Chromium Headless Shell 143.0.7499.4 (playwright build v1200) downloaded to C:\Users
\pablo\AppData\Local\ms-playwright\chromium_headless_shell-1200
Downloading Firefox 144.0.2 (playwright build v1497) from https://cdn.playwright.de
v/dbazure/download/playwright/builds/firefox/1497/firefox-win64.zip
|                                                                                |
0% of 107.1 MiB
|â– â– â– â– â– â– â– â–
|  10% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  20% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  30% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                                 |  40% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                         |  50%
of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  60% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  70% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                 |  80% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  90% of 107.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of 1
07.1 MiB
Firefox 144.0.2 (playwright build v1497) downloaded to C:\Users\pablo\AppData\Local
\ms-playwright\firefox-1497
Downloading Webkit 26.0 (playwright build v2227) from https://cdn.playwright.dev/dba
zure/download/playwright/builds/webkit/2227/webkit-win64.zip
|                                                                                |
0% of 58.2 MiB
|â– â– â– â– â– â– â– â–
|  10% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  20% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  30% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                                 |  40% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                         |  50%
```

```
of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  60% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  70% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                 |  80% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  90% of 58.2 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of 5
8.2 MiB
Webkit 26.0 (playwright build v2227) downloaded to C:\Users\pablo\AppData\Local\ms-p
laywright\webkit-2227
Downloading FFMPEG playwright build v1011 from https://cdn.playwright.dev/dbazure/do
wnload/playwright/builds/ffmpeg/1011/ffmpeg-win64.zip
|                                                                                 |
1% of 1.3 MiB
|â– â– â– â– â– â– â– â–
|  11% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  21% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  30% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                      |  40% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                          |  50%
of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  60% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  70% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                 |  80% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  90% of 1.3 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of
1.3 MiB
FFMPEG playwright build v1011 downloaded to C:\Users\pablo\AppData\Local\ms-playwrig
ht\ffmpeg-1011
Downloading Winldd playwright build v1007 from https://cdn.playwright.dev/dbazure/do
wnload/playwright/builds/winldd/1007/winldd-win64.zip
|â– â– â– â– â– â– â– â–
|  10% of 0.1 MiB
```

```
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  22% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  35% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â–                                                 |  48% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â–                                          |  54%
of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
|  67% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
                         |  79% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â–                  |  85% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–         |  98% of 0.1 MiB
|â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â–
 â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– â– | 100% of
0.1 MiB
Winldd playwright build v1007 downloaded to C:\Users\pablo\AppData\Local\ms-playwrig
ht\winldd-1007
```

```
(node:16856) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardize
d and prone to errors that have security implications. Use the WHATWG URL API instea
d. CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
(node:14980) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardize
d and prone to errors that have security implications. Use the WHATWG URL API instea
d. CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
(node:14940) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardize
d and prone to errors that have security implications. Use the WHATWG URL API instea
d. CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
(node:16332) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardize
d and prone to errors that have security implications. Use the WHATWG URL API instea
d. CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
(node:21748) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardize
d and prone to errors that have security implications. Use the WHATWG URL API instea
d. CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
(node:2140) [DEP0169] DeprecationWarning: `url.parse()` behavior is not standardized
and prone to errors that have security implications. Use the WHATWG URL API instead.
CVEs are not issued for `url.parse()` vulnerabilities.
(Use `node --trace-deprecation ...` to show where the warning was created)
```

In [2]:

```python
df = pd.read_csv('No_show_data.csv')
df
```

Out[2]:

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Ne
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62	
1	5.589980e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z	56	
2	4.262960e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z	62	M
3	8.679510e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z	8	
4	8.841190e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z	56	
...	...	...	...	...	...	...	
110522	2.572130e+12	5651768	F	2016-05-03T09:15:35Z	2016-06-07T00:00:00Z	56	
110523	3.596270e+12	5650093	F	2016-05-03T07:27:33Z	2016-06-07T00:00:00Z	51	
110524	1.557660e+13	5630692	F	2016-04-27T16:03:52Z	2016-06-07T00:00:00Z	21	
110525	9.213490e+13	5630323	F	2016-04-27T15:09:23Z	2016-06-07T00:00:00Z	38	
110526	3.775120e+14	5629448	F	2016-04-27T13:30:56Z	2016-06-07T00:00:00Z	54	

110527 rows × 14 columns

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110527 entries, 0 to 110526
Data columns (total 14 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   PatientId       110527 non-null  float64
 1   AppointmentID   110527 non-null  int64  
 2   Gender          110527 non-null  object 
 3   ScheduledDay    110527 non-null  object 
 4   AppointmentDay  110527 non-null  object 
 5   Age             110527 non-null  int64  
 6   Neighbourhood   110527 non-null  object 
 7   Scholarship     110527 non-null  int64  
 8   Hipertension    110527 non-null  int64  
 9   Diabetes        110527 non-null  int64  
 10  Alcoholism      110527 non-null  int64  
 11  Handcap         110527 non-null  int64  
 12  SMS_received    110527 non-null  int64  
 13  No-show         110527 non-null  object 
dtypes: float64(1), int64(8), object(5)
memory usage: 11.8+ MB
```

In [4]:
```
print(df['No-show'].value_counts())
```

```
No-show
No     88208
Yes    22319
Name: count, dtype: int64
```

In [ ]:
```

```

In [5]:
```
#standardize columns
df.columns = [c.strip().lower().replace('-','_') for c in df.columns]

#Target: no_show = 1/showed up = 0
df['no_show_flag'] = df['no_show'].map({'Yes': 1 , 'No': 0 })

#Drop NA

df = df.dropna(subset = ['appointmentday', 'scheduledday'])

df.no_show_flag.value_counts(normalize = True)
```

Out[5]:
```
no_show_flag
0    0.798067
1    0.201933
Name: proportion, dtype: float64
```

In [ ]:
```

```

In [6]:
```
df['appointmentday'] = pd.to_datetime(df['appointmentday']).dt.tz_localize(None)
df['scheduledday']   = pd.to_datetime(df['scheduledday']).dt.tz_localize(None)
```

```
# Days between scheduling and appointment
df['days_wait'] = (df['appointmentday'] - df['scheduledday']).dt.days

# Appointment weekday (Monday = 0)
df['appt_weekday'] = df['appointmentday'].dt.weekday

# Appointment Hour
df['appt_hour'] = df['appointmentday'].dt.hour

# Age Buckets
df['age_bucket'] = pd.cut(
    df['age'],
    bins=[-1, 12, 25, 40, 60, 120],
    labels=['child', 'youth', 'young_adult', 'adult', 'senior']
)

drop_cols = ['patientid', 'appointmentday', 'scheduledday', 'no_show']
df_model = df.drop(columns=[c for c in drop_cols if c in df.columns])


df_model
```

Out[6]:

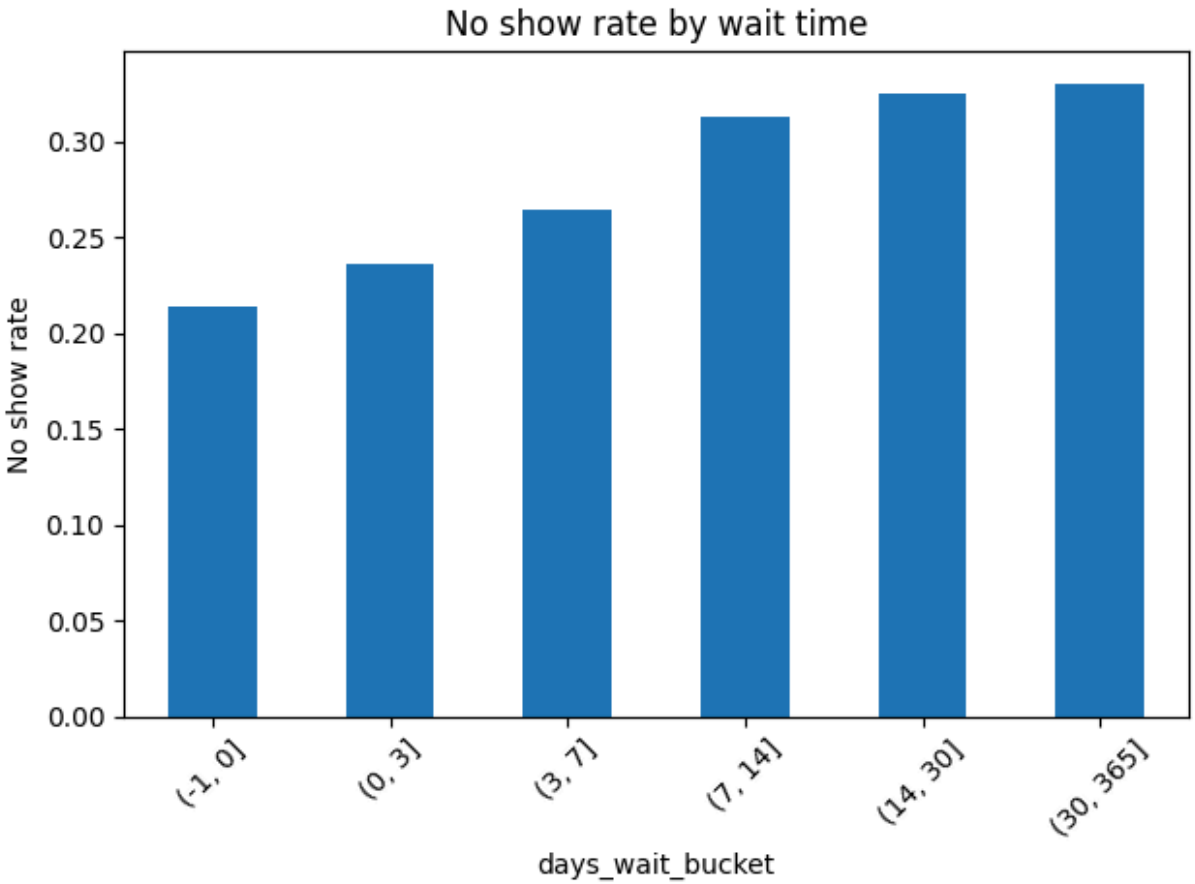
	appointmentid	gender	age	neighbourhood	scholarship	hipertension	diabetes
0	5642903	F	62	JARDIM DA PENHA	0	1	0
1	5642503	M	56	JARDIM DA PENHA	0	0	0
2	5642549	F	62	MATA DA PRAIA	0	0	0
3	5642828	F	8	PONTAL DE CAMBURI	0	0	0
4	5642494	F	56	JARDIM DA PENHA	0	1	1
...	...	...	...	...	...	...	...
110522	5651768	F	56	MARIA ORTIZ	0	0	0
110523	5650093	F	51	MARIA ORTIZ	0	0	0
110524	5630692	F	21	MARIA ORTIZ	0	0	0
110525	5630323	F	38	MARIA ORTIZ	0	0	0
110526	5629448	F	54	MARIA ORTIZ	0	0	0

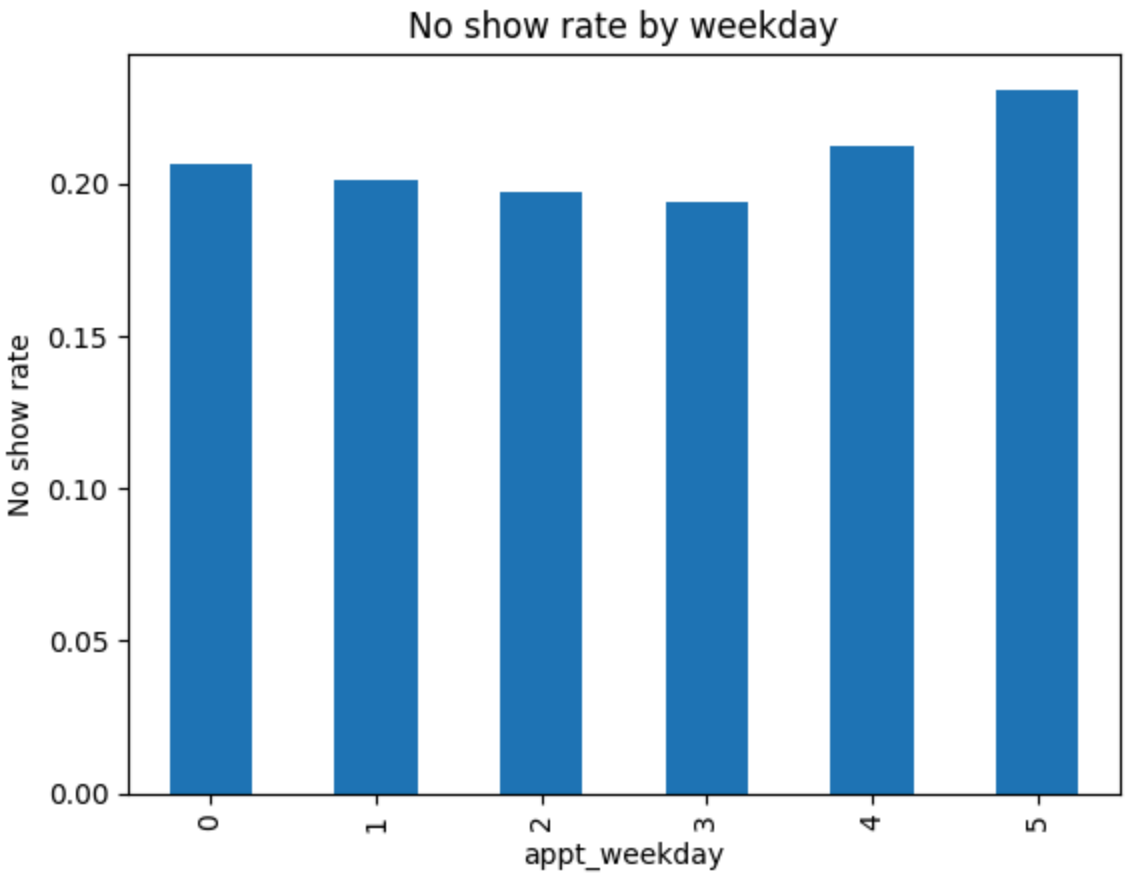
110527 rows × 15 columns

In [ ]:

In [7]:
```
# No_show rate by days_wait bucket
df_model['days_wait_bucket'] = pd.cut(df_model['days_wait'], bins=[-1, 0, 3, 7, 14,

no_show_by_wait = df_model.groupby('days_wait_bucket')['no_show_flag'].mean()

print(no_show_by_wait)

no_show_by_wait.plot(kind='bar')
plt.ylabel('No show rate')
plt.title('No show rate by wait time')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# No show by weekday
no_show_by_weekday = df_model.groupby('appt_weekday')['no_show_flag'].mean()

no_show_by_weekday.plot(kind='bar')
plt.ylabel('No show rate')
plt.title('No show rate by weekday')
plt.show()
```

```
C:\Users\pablo\AppData\Local\Temp\ipykernel_20840\663844668.py:4: FutureWarning: The
default of observed=False is deprecated and will be changed to True in a future vers
ion of pandas. Pass observed=False to retain current behavior or observed=True to ad
opt the future default and silence this warning.
  no_show_by_wait = df_model.groupby('days_wait_bucket')['no_show_flag'].mean()
days_wait_bucket
(-1, 0]      0.213505
(0, 3]       0.235697
(3, 7]       0.264706
(7, 14]      0.312254
(14, 30]     0.325212
(30, 365]    0.330208
Name: no_show_flag, dtype: float64
```


No show rate by wait time
No show rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
(-1, 0]
(0, 3]
(3, 7]
(7, 14]
(14, 30]
(30, 365]
days_wait_bucket


No show rate by weekday
No show rate
0.00
0.05
0.10
0.15
0.20
0
1
2
3
4
5
appt_weekday

In [ ]:

In [8]:
```
## Train/Test Split

target = 'no_show_flag'

numeric_feature = ['age', 'days_wait', 'appt_weekday', 'appt_hour', 'scholarship',
numeric_feature = [f for f in numeric_feature if f in df_model.columns]


categorical_features = ['gender', 'neighbourhood', 'age_bucket']
categorical_features = [f for f in categorical_features if f in df_model.columns]

x = df_model[numeric_feature + categorical_features]
y = df_model[target]

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_s
```

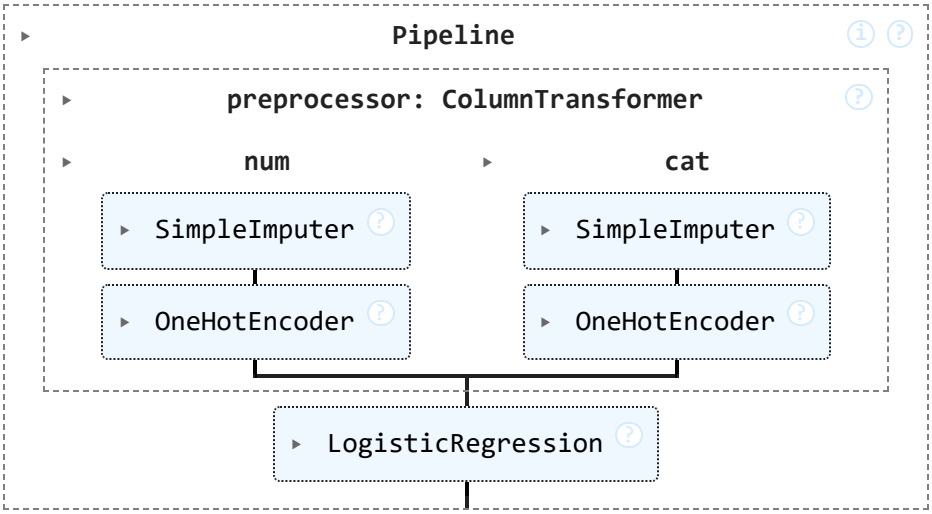
In [ ]:
```
```

In [16]:
```
from sklearn.preprocessing import OneHotEncoder

#Numeric pipeline

numeric_transformer = Pipeline(steps = [
    ('imputer',SimpleImputer(strategy = 'most_frequent')),
    ('encoder',OneHotEncoder(handle_unknown = 'ignore'))
])


categorical_transformer = Pipeline(steps = [
    ('imputer',SimpleImputer(strategy = 'most_frequent')),
    ('encoder',OneHotEncoder(handle_unknown = 'ignore'))
])



preprocessor = ColumnTransformer(
    transformers = [
        ('num', numeric_transformer, numeric_feature),
        ('cat', categorical_transformer, categorical_features)
    ]
)
```

In [ ]:
```
```

In [17]:
```
#Logistic Regression

log_reg_clf = Pipeline(steps = [
    ('preprocessor', preprocessor),
    ('model', LogisticRegression(max_iter = 200, n_jobs =-1))
])

log_reg_clf.fit(x_train, y_train)
```

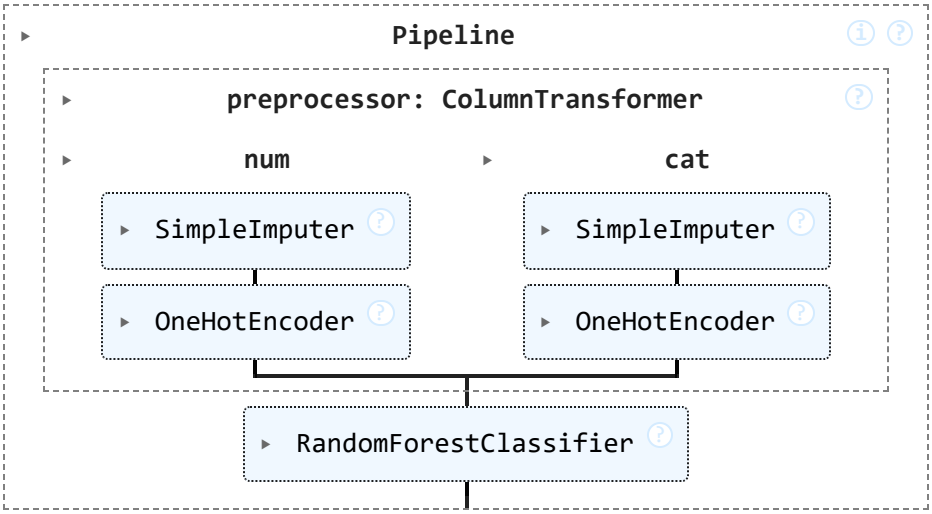
Out[17]:



In [ ]:

In [18]:
```
#Random Forest

rf_clf = Pipeline(steps = [
    ('preprocessor', preprocessor),
    ('model', RandomForestClassifier(
        n_estimators = 300,
        max_depth = None,
        random_state = 42,
        n_jobs = -1))])

rf_clf.fit(x_train, y_train)
```

Out[18]:



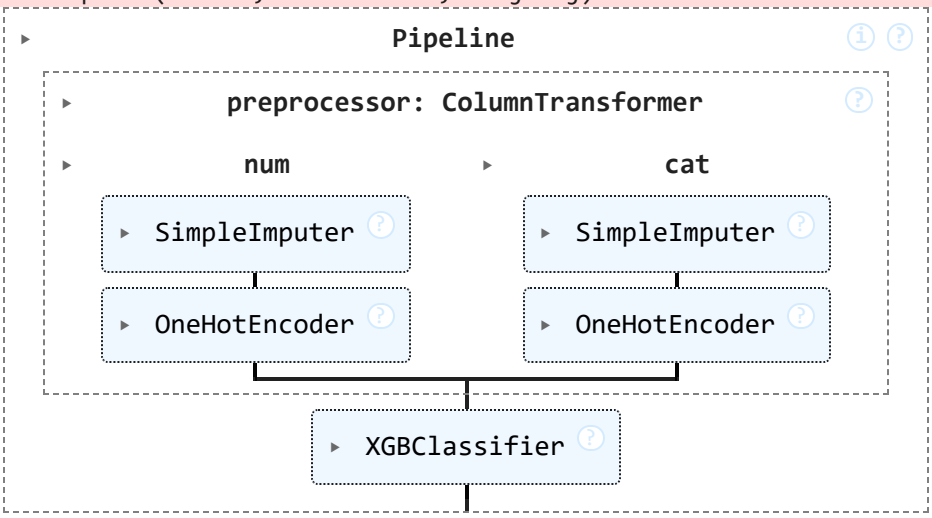
In [ ]:

In [33]:
```
#xgboost

xgb_clf = Pipeline(steps = [
    ('preprocessor', preprocessor),
    ('model', XGBClassifier(
        n_estimators = 400,
```

```
        max_depth = 5,
        subsample = 0.8,
        colsample_bytree = 0.8,
        random_state = 42,
        n_jobs = -1,
        eval_metric = 'logloss',
        use_label_encoder = False))
])

xgb_clf.fit(x_train,y_train)
```

```
C:\Users\pablo\AppData\Local\ABF Certificate\python\python\Lib\site-packages\xgboost
\training.py:199: UserWarning: [21:16:13] WARNING: C:\actions-runner\_work\xgboost\x
gboost\src\learner.cc:790:
Parameters: { "use_label_encoder" } are not used.

  bst.update(dtrain, iteration=i, fobj=obj)
```

Out[33]:



In [ ]:

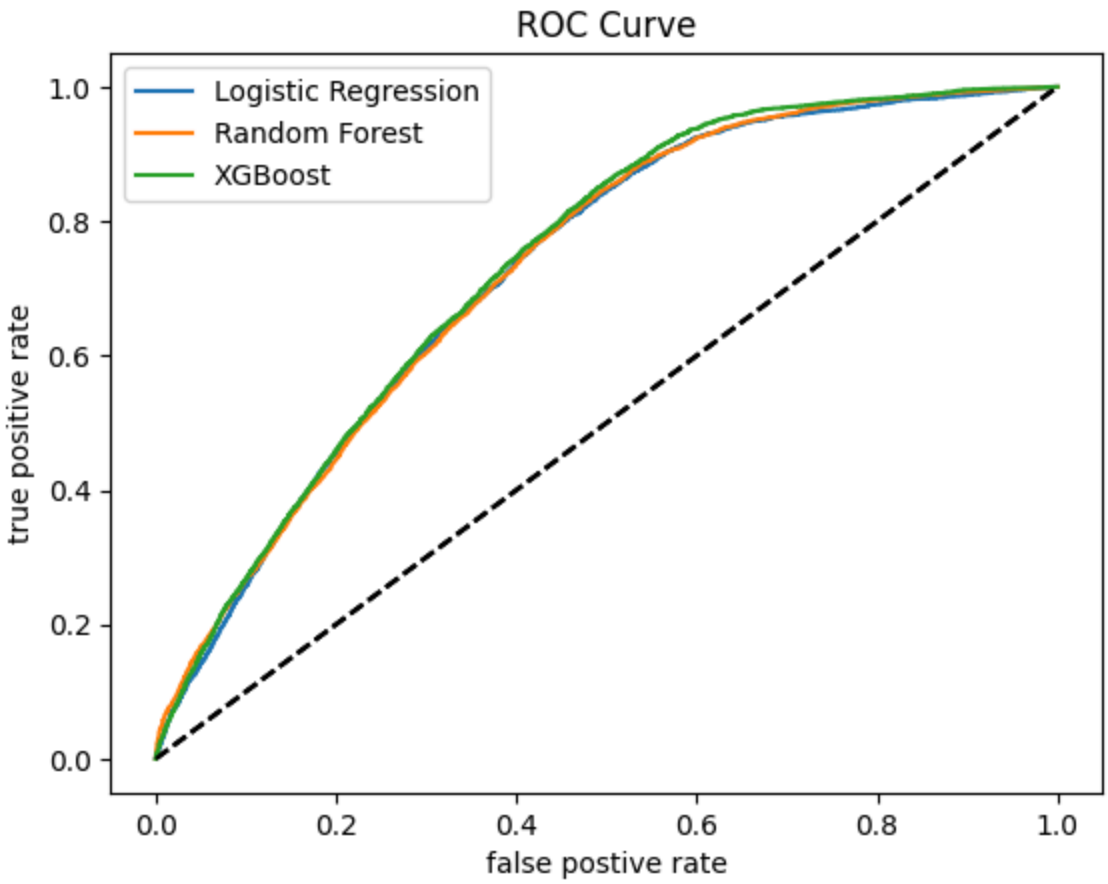
In [34]:
```
#Evaluation Function

def evaluate_model(name, model, x_test, y_test):
    y_pred = model.predict(x_test)
    y_prob = model.predict_proba(x_test)[:,1]

    acc = accuracy_score(y_test, y_pred)
    prec = precision_score(y_test, y_pred)
    rec = recall_score (y_test, y_pred)
    f1 = f1_score (y_test, y_pred)
    auc = roc_auc_score (y_test, y_prob)


    print(f'=== {name}===')
    print(f'Accuracy:{acc:.3f}')
    print(f'Precision:{prec:.3f}')
    print(f'Recall:{rec:.3f}')
    print(f'F1 Score: {f1:.3f}')
    print(f'ROC AUC: {auc:.3f}')
```

```
    print()

    #Confusion Matrix

    cm = confusion_matrix(y_test, y_pred)
    print('Confusion Matrix:')
    print(cm)
    print('-'*40)


    #ROC CURVE
    fpr, tpr, thresholds = roc_curve(y_test, y_prob)
    plt.plot(fpr, tpr, label = name)
    plt.plot ([0,1], [0,1], linestyle = '--', color = 'black')
    plt.xlabel('false postive rate')
    plt.ylabel('true positive rate')
    plt.title('ROC Curve')

    plt.legend()
```

In [ ]:

In [35]:
```
plt.figsize=(7,5)

evaluate_model('Logistic Regression', log_reg_clf, x_test, y_test)
evaluate_model('Random Forest', rf_clf, x_test, y_test)
evaluate_model('XGBoost', xgb_clf, x_test, y_test)

plt.show()
```

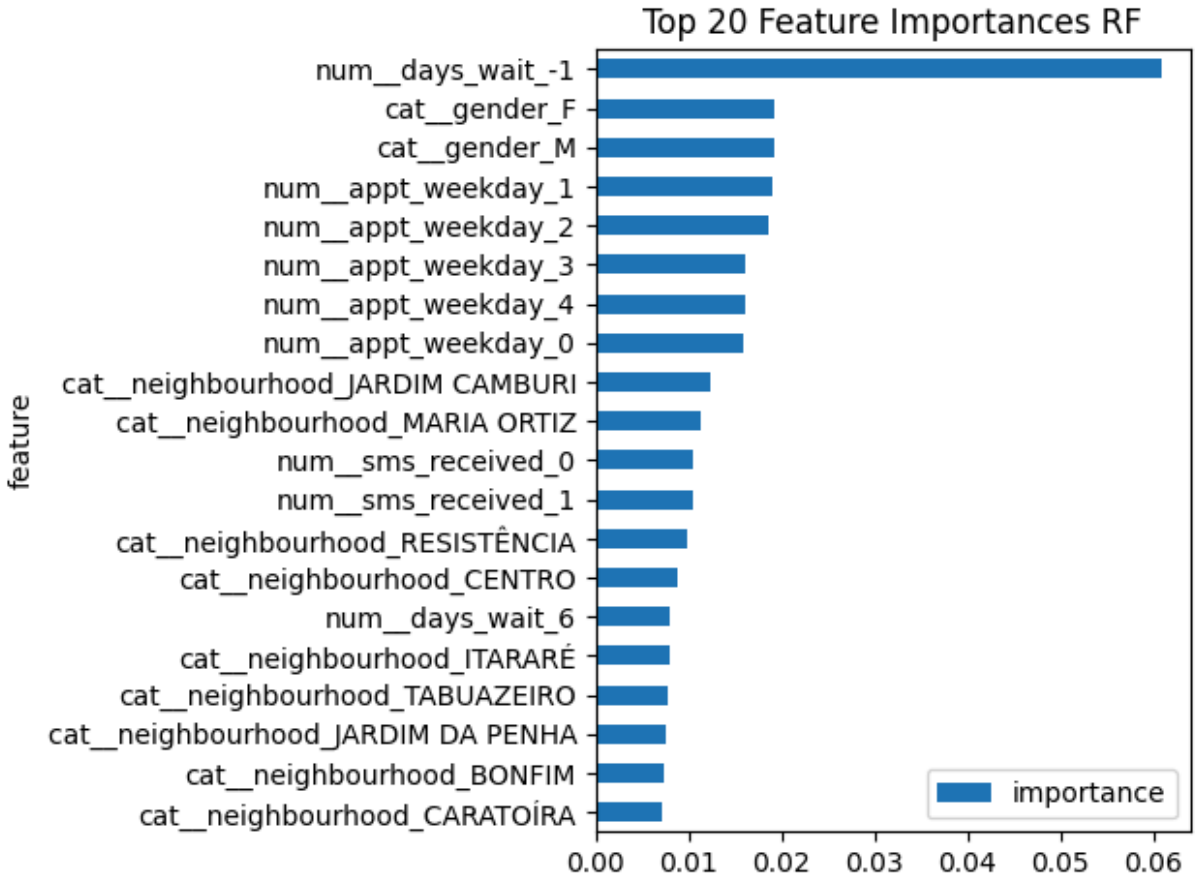
```
=== Logistic Regression===
Accuracy:0.799
Precision:0.500
Recall:0.019
F1 Score: 0.036
ROC AUC: 0.727

Confusion Matrix:
[[17586    83]
 [ 4354    83]]
----------------------------------------
=== Random Forest===
Accuracy:0.794
Precision:0.461
Recall:0.160
F1 Score: 0.237
ROC AUC: 0.730

Confusion Matrix:
[[16839   830]
 [ 3728   709]]
----------------------------------------
=== XGBoost===
Accuracy:0.796
Precision:0.460
Recall:0.094
F1 Score: 0.155
ROC AUC: 0.736

Confusion Matrix:
[[17182   487]
 [ 4022   415]]
----------------------------------------
```



In [ ]:

In [42]:
```
# All feature names in the correct order
all_feature_names = preproc.get_feature_names_out()

# Importances from RF
importances = rf_model.feature_importances_

fi_rf = (
    pd.DataFrame({
        'feature': all_feature_names,
        'importance': importances
    })
    .sort_values('importance', ascending=False)  # NOTE: bool, not string
)

print(fi_rf.head(20))

# Plot top 20
fi_rf.head(20).plot(kind='barh', x='feature', y='importance')
plt.gca().invert_yaxis()
plt.title('Top 20 Feature Importances RF')
plt.tight_layout()  # typo fixed
plt.show()
```

	feature	importance
105	num__days_wait_-1	0.060926
253	cat__gender_F	0.019242
254	cat__gender_M	0.019180
234	num__appt_weekday_1	0.018985
235	num__appt_weekday_2	0.018591
236	num__appt_weekday_3	0.016168
237	num__appt_weekday_4	0.016019
233	num__appt_weekday_0	0.015904
293	cat__neighbourhood_JARDIM CAMBURI	0.012247
298	cat__neighbourhood_MARIA ORTIZ	0.011192
251	num__sms_received_0	0.010467
252	num__sms_received_1	0.010461
314	cat__neighbourhood_RESISTÊNCIA	0.009832
265	cat__neighbourhood_CENTRO	0.008829
112	num__days_wait_6	0.007981
291	cat__neighbourhood_ITARARÉ	0.007863
333	cat__neighbourhood_TABUAZEIRO	0.007770
294	cat__neighbourhood_JARDIM DA PENHA	0.007470
263	cat__neighbourhood_BONFIM	0.007417
264	cat__neighbourhood_CARATOÍRA	0.007213



In [ ]:

In [63]:
```
#Prediction Function

def predict_no_show_prob (model, sample_dict):
    sample_df = pd.DataFrame ([sample_dict])
    prob = model.predict_proba(sample_df)[0,1]
```

```
    return prob

patient_example = {
    'age' : 15,
    'days_wait': 7,
    'appt_weekday': 0,
    'appt_hour' : 14,
    'scholarship': 0,
    'hypertension': 0,
    'diabetes': 0,
    'alcoholism': 0,
    'handcap':0,
    'sms_received':0,
    'gender': 'M',
    'neighbourhood': (),
    'age_bucket': 'young_adult'
}


prob = predict_no_show_prob(rf_clf, patient_example)

print(f'Predicited no show probability: {prob:.2%}')
```

```
Predicited no show probability: 36.22%
```

In [ ]:

In [ ]: