1.0 Common commands

1.1 Mount drive

In [ ]:

```
#Mount Google drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount
("/content/drive", force_remount=True).
```

1.2 Install packages

In [ ]:

```
#Install packages
!pip install pycaret
!pip install --upgrade pycaret
!pip install pyod
!pip install --upgrade pyod
```

```
Collecting pycaret
  Downloading pycaret-2.3.6-py3-none-any.whl (301 kB)
     |████████████████████████████████| 301 kB 10.6 MB/s
Collecting pyLDAvis
  Downloading pyLDAvis-3.3.1.tar.gz (1.7 MB)
     |████████████████████████████████| 1.7 MB 42.2 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
    Preparing wheel metadata ... done
Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages (fr
om pycaret) (3.6.0)
Collecting mlflow
  Downloading mlflow-1.23.1-py3-none-any.whl (15.6 MB)
     |████████████████████████████████| 15.6 MB 603 kB/s
Collecting kmodes>=0.10.1
  Downloading kmodes-0.11.1-py2.py3-none-any.whl (19 kB)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (from
pycaret) (1.5.0)
Collecting yellowbrick>=1.0.1
  Downloading yellowbrick-1.4-py3-none-any.whl (274 kB)
     |████████████████████████████████| 274 kB 23.5 MB/s
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages (from
pycaret) (7.6.5)
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from py
caret) (5.5.0)
Collecting scikit-plot
  Downloading scikit_plot-0.3.7-py3-none-any.whl (33 kB)
Requirement already satisfied: plotly>=4.4.1 in /usr/local/lib/python3.7/dist-packages (f
rom pycaret) (5.5.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from pycar
et) (3.2.5)
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (from p
ycaret) (0.15.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from pyc
aret) (1.3.5)
Collecting mlxtend>=0.17.0
  Downloading mlxtend-0.19.0-py2.py3-none-any.whl (1.3 MB)
     |████████████████████████████████| 1.3 MB 36.7 MB/s
Collecting lightgbm>=2.3.1
  Downloading lightgbm-3.3.2-py3-none-manylinux1_x86_64.whl (2.0 MB)
```

```
     |████████████████████████████████| 2.0 MB 9.7 MB/s
Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-package
s (from pycaret) (0.17.3)
Requirement already satisfied: spacy<2.4.0 in /usr/local/lib/python3.7/dist-packages (fro
m pycaret) (2.2.4)
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from py
caret) (0.11.2)
Collecting Boruta
  Downloading Boruta-0.3-py3-none-any.whl (56 kB)
     |████████████████████████████████| 56 kB 3.5 MB/s
Collecting umap-learn
  Downloading umap-learn-0.5.2.tar.gz (86 kB)
     |████████████████████████████████| 86 kB 2.7 MB/s
Requirement already satisfied: pyyaml<6.0.0 in /usr/local/lib/python3.7/dist-packages (fr
om pycaret) (3.13)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from
pycaret) (3.2.2)
Collecting pandas-profiling>=2.8.0
  Downloading pandas_profiling-3.1.0-py2.py3-none-any.whl (261 kB)
     |████████████████████████████████| 261 kB 35.8 MB/s
Collecting imbalanced-learn==0.7.0
  Downloading imbalanced_learn-0.7.0-py3-none-any.whl (167 kB)
     |████████████████████████████████| 167 kB 46.9 MB/s
Requirement already satisfied: scipy<=1.5.4 in /usr/local/lib/python3.7/dist-packages (fr
om pycaret) (1.4.1)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pyc
aret) (1.1.0)
Collecting pyod
  Downloading pyod-0.9.7.tar.gz (114 kB)
     |████████████████████████████████| 114 kB 41.8 MB/s
Collecting scikit-learn==0.23.2
  Downloading scikit_learn-0.23.2-cp37-cp37m-manylinux1_x86_64.whl (6.8 MB)
     |████████████████████████████████| 6.8 MB 31.3 MB/s
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (f
rom imbalanced-learn==0.7.0->pycaret) (1.21.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-pack
ages (from scikit-learn==0.23.2->pycaret) (3.1.0)
Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-package
s (from cufflinks>=0.17.0->pycaret) (0.3.0)
Requirement already satisfied: setuptools>=34.4.1 in /usr/local/lib/python3.7/dist-packag
es (from cufflinks>=0.17.0->pycaret) (57.4.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from
cufflinks>=0.17.0->pycaret) (1.15.0)
Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-package
s (from gensim<4.0.0->pycaret) (5.2.1)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from I
Python->pycaret) (2.6.1)
Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-package
s (from IPython->pycaret) (0.8.1)
Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packages (
from IPython->pycaret) (5.1.1)
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/d
ist-packages (from IPython->pycaret) (1.0.18)
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from IP
ython->pycaret) (4.8.0)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from
IPython->pycaret) (4.4.2)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (fro
m IPython->pycaret) (0.7.5)
Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist
-packages (from ipywidgets->pycaret) (3.5.2)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages
(from ipywidgets->pycaret) (4.10.1)
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages
(from ipywidgets->pycaret) (5.1.3)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist
-packages (from ipywidgets->pycaret) (1.0.2)
Requirement already satisfied: ipython-genutils~=0.2.0 in /usr/local/lib/python3.7/dist-p
ackages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: tornado>=4.0 in /usr/local/lib/python3.7/dist-packages (fr
om ipykernel>=4.5.1->ipywidgets->pycaret) (5.1.1)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (
```

```
from ipykernel>=4.5.1->ipywidgets->pycaret) (5.3.5)
Requirement already satisfied: wheel in /usr/local/lib/python3.7/dist-packages (from ligh
tgbm>=2.3.1->pycaret) (0.37.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-package
s (from matplotlib->pycaret) (1.3.2)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pack
ages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (fr
om matplotlib->pycaret) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib
/python3.7/dist-packages (from matplotlib->pycaret) (3.0.7)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-p
ackages (from nbformat>=4.2.0->ipywidgets->pycaret) (4.3.3)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (fr
om nbformat>=4.2.0->ipywidgets->pycaret) (4.9.2)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packag
es (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (4.11.1)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-package
s (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (3.10.0.2)
Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.7/dis
t-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (5.4.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /usr/loca
l/lib/python3.7/dist-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets-
>pycaret) (0.18.1)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-packages (f
rom jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (21.4.0)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/dist-packages (fro
m importlib-resources>=1.4.0->jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycar
et) (3.7.0)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (fr
om pandas->pycaret) (2018.9)
Collecting pydantic>=1.8.1
  Downloading pydantic-1.9.0-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1
0.9 MB)
     |████████████████████████████████| 10.9 MB 47.4 MB/s
Collecting htmlmin>=0.1.12
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
Collecting tangled-up-in-unicode==0.1.0
  Downloading tangled_up_in_unicode-0.1.0-py3-none-any.whl (3.1 MB)
     |████████████████████████████████| 3.1 MB 52.9 MB/s
Collecting pyyaml<6.0.0
  Downloading PyYAML-5.4.1-cp37-cp37m-manylinux1_x86_64.whl (636 kB)
     |████████████████████████████████| 636 kB 59.3 MB/s
Requirement already satisfied: markupsafe~=2.0.1 in /usr/local/lib/python3.7/dist-package
s (from pandas-profiling>=2.8.0->pycaret) (2.0.1)
Collecting multimethod>=1.4
  Downloading multimethod-1.7-py3-none-any.whl (9.5 kB)
Collecting joblib
  Downloading joblib-1.0.1-py3-none-any.whl (303 kB)
     |████████████████████████████████| 303 kB 59.9 MB/s
Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.7/dist-packages (
from pandas-profiling>=2.8.0->pycaret) (2.11.3)
Collecting phik>=0.11.1
  Downloading phik-0.12.0-cp37-cp37m-manylinux2010_x86_64.whl (675 kB)
     |████████████████████████████████| 675 kB 61.2 MB/s
Collecting visions[type_image_path]==0.7.4
  Downloading visions-0.7.4-py3-none-any.whl (102 kB)
     |████████████████████████████████| 102 kB 9.2 MB/s
Collecting requests>=2.24.0
  Downloading requests-2.27.1-py2.py3-none-any.whl (63 kB)
     |████████████████████████████████| 63 kB 1.0 MB/s
Requirement already satisfied: tqdm>=4.48.2 in /usr/local/lib/python3.7/dist-packages (fr
om pandas-profiling>=2.8.0->pycaret) (4.62.3)
Requirement already satisfied: missingno>=0.4.2 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (0.5.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.7/dist-packages (f
rom visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (2.6.3)
Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from vis
ions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (7.1.2)
Collecting imagehash
  Downloading ImageHash-4.2.1.tar.gz (812 kB)
     |████████████████████████████████| 812 kB 15.0 MB/s
```

```
Collecting scipy<=1.5.4
  Downloading scipy-1.5.4-cp37-cp37m-manylinux1_x86_64.whl (25.9 MB)
     |████████████████████████████████| 25.9 MB 1.5 MB/s
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.7/dist-packages
(from plotly>=4.4.1->pycaret) (8.0.1)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from pr
ompt-toolkit<2.0.0,>=1.0.4->IPython->pycaret) (0.2.5)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.7/dist-packages (fr
om requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.10)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.7/dist-pac
kages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (1.24.3)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packag
es (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.7/dist
-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.0.12)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packa
ges (from spacy<2.4.0->pycaret) (2.0.6)
Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om spacy<2.4.0->pycaret) (7.4.0)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packag
es (from spacy<2.4.0->pycaret) (1.1.3)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
kages (from spacy<2.4.0->pycaret) (3.0.6)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-packa
ges (from spacy<2.4.0->pycaret) (1.0.5)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packag
es (from spacy<2.4.0->pycaret) (0.4.1)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist-p
ackages (from spacy<2.4.0->pycaret) (1.0.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/dist
-packages (from spacy<2.4.0->pycaret) (1.0.6)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pack
ages (from spacy<2.4.0->pycaret) (0.9.0)
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages
(from widgetsnbextension~=3.5.0->ipywidgets->pycaret) (5.3.1)
Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages
(from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.13.1)
Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (5.6.1)
Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1.8.0)
Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from
jupyter-client->ipykernel>=4.5.1->ipywidgets->pycaret) (22.3.0)
Requirement already satisfied: ptyprocess in /usr/local/lib/python3.7/dist-packages (from
terminado>=0.8.1->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.7.0
)
Collecting yellowbrick>=1.0.1
  Downloading yellowbrick-1.3.post1-py3-none-any.whl (271 kB)
     |████████████████████████████████| 271 kB 68.0 MB/s
Collecting numpy>=1.13.3
  Downloading numpy-1.19.5-cp37-cp37m-manylinux2010_x86_64.whl (14.8 MB)
     |████████████████████████████████| 14.8 MB 234 kB/s
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.7/dist-packages (from
imagehash->visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (1.2.0)
Requirement already satisfied: entrypoints in /usr/local/lib/python3.7/dist-packages (fro
m mlflow->pycaret) (0.4)
Requirement already satisfied: protobuf>=3.7.0 in /usr/local/lib/python3.7/dist-packages
(from mlflow->pycaret) (3.17.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (21.3)
Collecting querystring-parser
  Downloading querystring_parser-1.2.4-py2.py3-none-any.whl (7.9 kB)
Collecting gitpython>=2.1.0
  Downloading GitPython-3.1.27-py3-none-any.whl (181 kB)
     |████████████████████████████████| 181 kB 41.6 MB/s
Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (7.1.2)
Collecting gunicorn
  Downloading gunicorn-20.1.0-py3-none-any.whl (79 kB)
     |████████████████████████████████| 79 kB 7.3 MB/s
Requirement already satisfied: Flask in /usr/local/lib/python3.7/dist-packages (from mlfl
ow->pycaret) (1.1.4)
```

```
Requirement already satisfied: sqlparse>=0.3.1 in /usr/local/lib/python3.7/dist-packages
(from mlflow->pycaret) (0.4.2)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.7/dist-packages (fro
m mlflow->pycaret) (1.3.0)
Collecting alembic
  Downloading alembic-1.7.6-py3-none-any.whl (210 kB)
     |████████████████████████████████| 210 kB 52.7 MB/s
Collecting databricks-cli>=0.8.7
  Downloading databricks-cli-0.16.4.tar.gz (58 kB)
     |████████████████████████████████| 58 kB 5.3 MB/s
Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (1.4.31)
Collecting prometheus-flask-exporter
  Downloading prometheus_flask_exporter-0.18.7-py3-none-any.whl (17 kB)
Collecting docker>=4.0.0
  Downloading docker-5.0.3-py2.py3-none-any.whl (146 kB)
     |████████████████████████████████| 146 kB 48.3 MB/s
Requirement already satisfied: tabulate>=0.7.7 in /usr/local/lib/python3.7/dist-packages
(from databricks-cli>=0.8.7->mlflow->pycaret) (0.8.9)
Collecting websocket-client>=0.32.0
  Downloading websocket_client-1.2.3-py3-none-any.whl (53 kB)
     |████████████████████████████████| 53 kB 2.1 MB/s
Collecting gitdb<5,>=4.0.1
  Downloading gitdb-4.0.9-py3-none-any.whl (63 kB)
     |████████████████████████████████| 63 kB 1.6 MB/s
Collecting smmap<6,>=3.0.1
  Downloading smmap-5.0.0-py3-none-any.whl (24 kB)
Collecting Mako
  Downloading Mako-1.1.6-py2.py3-none-any.whl (75 kB)
     |████████████████████████████████| 75 kB 3.8 MB/s
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.7/dist-packages
(from sqlalchemy->mlflow->pycaret) (1.1.2)
Requirement already satisfied: itsdangerous<2.0,>=0.24 in /usr/local/lib/python3.7/dist-p
ackages (from Flask->mlflow->pycaret) (1.1.0)
Requirement already satisfied: Werkzeug<2.0,>=0.15 in /usr/local/lib/python3.7/dist-packa
ges (from Flask->mlflow->pycaret) (1.0.1)
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-pack
ages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1
.5.0)
Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from n
bconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.5.0)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.7.1)
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-package
s (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.8.
4)
Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbc
onvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (4.1.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (fr
om bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0
.5.1)
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.7/dist-package
s (from prometheus-flask-exporter->mlflow->pycaret) (0.13.1)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from pyL
DAvis->pycaret) (0.16.0)
Requirement already satisfied: sklearn in /usr/local/lib/python3.7/dist-packages (from py
LDAvis->pycaret) (0.0)
Requirement already satisfied: numexpr in /usr/local/lib/python3.7/dist-packages (from py
LDAvis->pycaret) (2.8.1)
Collecting funcy
  Downloading funcy-1.17-py2.py3-none-any.whl (33 kB)
Collecting pyLDAvis
  Downloading pyLDAvis-3.3.0.tar.gz (1.7 MB)
     |████████████████████████████████| 1.7 MB 44.0 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
    Preparing wheel metadata ... done
  Downloading pyLDAvis-3.2.2.tar.gz (1.7 MB)
     |████████████████████████████████| 1.7 MB 57.6 MB/s
Requirement already satisfied: numba>=0.35 in /usr/local/lib/python3.7/dist-packages (fro
m pyod->pycaret) (0.51.2)
```

```
Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (fro
m pyod->pycaret) (0.10.2)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/di
st-packages (from numba>=0.35->pyod->pycaret) (0.34.0)
Requirement already satisfied: patsy>=0.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod->pycaret) (0.5.2)
Collecting pynndescent>=0.5
  Downloading pynndescent-0.5.6.tar.gz (1.1 MB)
     |████████████████████████████████| 1.1 MB 44.1 MB/s
Building wheels for collected packages: htmlmin, imagehash, databricks-cli, pyLDAvis, pyo
d, umap-learn, pynndescent
  Building wheel for htmlmin (setup.py) ... done
  Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27098 sha256=0
5b7b3ec66a7671d88392614a63f8cb9aa79f4ac41d437d9e25a9e6efc93e29d
  Stored in directory: /root/.cache/pip/wheels/70/e1/52/5b14d250ba868768823940c3229e9950d
201a26d0bd3ee8655
  Building wheel for imagehash (setup.py) ... done
  Created wheel for imagehash: filename=ImageHash-4.2.1-py2.py3-none-any.whl size=295206
sha256=6234356ed982878eeaa1cd3165b2e5798ff8407438424458c32842103858e39f
  Stored in directory: /root/.cache/pip/wheels/4c/d5/59/5e3e297533ddb09407769762985d13413
5064c6831e29a914e
  Building wheel for databricks-cli (setup.py) ... done
  Created wheel for databricks-cli: filename=databricks_cli-0.16.4-py3-none-any.whl size=
106877 sha256=6f53c9cb572d20bbd1bfe597d602ee4c2ca1cf3f3753b2d03b2a2cb6e0871b5e
  Stored in directory: /root/.cache/pip/wheels/a2/a1/6d/fa1d22ea25ed8593887437fe1c7e00f6e
f307fc240ccd4dc5c
  Building wheel for pyLDAvis (setup.py) ... done
  Created wheel for pyLDAvis: filename=pyLDAvis-3.2.2-py2.py3-none-any.whl size=135617 sh
a256=cacb745a4398951229b34641908b134805d507e79a4b5efb53efd15044a9e1c5
  Stored in directory: /root/.cache/pip/wheels/f8/b1/9b/560ac1931796b7303f7b517b949d2d31a
4fbc512aad3b9f284
  Building wheel for pyod (setup.py) ... done
  Created wheel for pyod: filename=pyod-0.9.7-py3-none-any.whl size=136279 sha256=b7a75ff
75e27ac702e4b3daf2c6639b2c531a78226789b2db18b6ada7587153d
  Stored in directory: /root/.cache/pip/wheels/ce/14/ae/60cbb36511e59bc12f8f0883805f586db
3b315972b54865d33
  Building wheel for umap-learn (setup.py) ... done
  Created wheel for umap-learn: filename=umap_learn-0.5.2-py3-none-any.whl size=82708 sha
256=51b7f59d16cd97b0f51d242a23e7d05d0d675c1df44ad0b44e8585885727fe58
  Stored in directory: /root/.cache/pip/wheels/84/1b/c6/aaf68a748122632967cef4dffef68224e
b16798b6793257d82
  Building wheel for pynndescent (setup.py) ... done
  Created wheel for pynndescent: filename=pynndescent-0.5.6-py3-none-any.whl size=53943 s
ha256=6aa828c653deb1ee304f24a9b23ad3993e3ec2a008aa1a561c61e40bd78eb916
  Stored in directory: /root/.cache/pip/wheels/03/f1/56/f80d72741e400345b5a5b50ec3d929aca
581bf45e0225d5c50
Successfully built htmlmin imagehash databricks-cli pyLDAvis pyod umap-learn pynndescent
Installing collected packages: numpy, tangled-up-in-unicode, smmap, scipy, multimethod, j
oblib, websocket-client, visions, scikit-learn, requests, Mako, imagehash, gitdb, queryst
ring-parser, pyyaml, pynndescent, pydantic, prometheus-flask-exporter, phik, htmlmin, gun
icorn, gitpython, funcy, docker, databricks-cli, alembic, yellowbrick, umap-learn, scikit
-plot, pyod, pyLDAvis, pandas-profiling, mlxtend, mlflow, lightgbm, kmodes, imbalanced-le
arn, Boruta, pycaret
  Attempting uninstall: numpy
    Found existing installation: numpy 1.21.5
    Uninstalling numpy-1.21.5:
      Successfully uninstalled numpy-1.21.5
  Attempting uninstall: scipy
    Found existing installation: scipy 1.4.1
    Uninstalling scipy-1.4.1:
      Successfully uninstalled scipy-1.4.1
  Attempting uninstall: joblib
    Found existing installation: joblib 1.1.0
    Uninstalling joblib-1.1.0:
      Successfully uninstalled joblib-1.1.0
  Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.0.2
    Uninstalling scikit-learn-1.0.2:
      Successfully uninstalled scikit-learn-1.0.2
  Attempting uninstall: requests
    Found existing installation: requests 2.23.0
    Uninstalling requests-2.23.0:
```

```
      Successfully uninstalled requests-2.23.0
  Attempting uninstall: pyyaml
    Found existing installation: PyYAML 3.13
    Uninstalling PyYAML-3.13:
      Successfully uninstalled PyYAML-3.13
  Attempting uninstall: pandas-profiling
    Found existing installation: pandas-profiling 1.4.1
    Uninstalling pandas-profiling-1.4.1:
      Successfully uninstalled pandas-profiling-1.4.1
  Attempting uninstall: mlxtend
    Found existing installation: mlxtend 0.14.0
    Uninstalling mlxtend-0.14.0:
      Successfully uninstalled mlxtend-0.14.0
  Attempting uninstall: lightgbm
    Found existing installation: lightgbm 2.2.3
    Uninstalling lightgbm-2.2.3:
      Successfully uninstalled lightgbm-2.2.3
  Attempting uninstall: imbalanced-learn
    Found existing installation: imbalanced-learn 0.8.1
    Uninstalling imbalanced-learn-0.8.1:
      Successfully uninstalled imbalanced-learn-0.8.1
ERROR: pip's dependency resolver does not currently take into account all the packages th
at are installed. This behaviour is the source of the following dependency conflicts.
tensorflow 2.8.0 requires tf-estimator-nightly==2.8.0.dev2021122109, which is not install
ed.
tensorflow 2.8.0 requires numpy>=1.20, but you have numpy 1.19.5 which is incompatible.
google-colab 1.0.0 requires requests~=2.23.0, but you have requests 2.27.1 which is incom
patible.
datascience 0.10.6 requires folium==0.2.1, but you have folium 0.8.3 which is incompatibl
e.
albumentations 0.1.12 requires imgaug<0.2.7,>=0.2.5, but you have imgaug 0.2.9 which is i
ncompatible.
Successfully installed Boruta-0.3 Mako-1.1.6 alembic-1.7.6 databricks-cli-0.16.4 docker-5
.0.3 funcy-1.17 gitdb-4.0.9 gitpython-3.1.27 gunicorn-20.1.0 htmlmin-0.1.12 imagehash-4.2
.1 imbalanced-learn-0.7.0 joblib-1.0.1 kmodes-0.11.1 lightgbm-3.3.2 mlflow-1.23.1 mlxtend
-0.19.0 multimethod-1.7 numpy-1.19.5 pandas-profiling-3.1.0 phik-0.12.0 prometheus-flask-
exporter-0.18.7 pyLDAvis-3.2.2 pycaret-2.3.6 pydantic-1.9.0 pynndescent-0.5.6 pyod-0.9.7
pyyaml-5.4.1 querystring-parser-1.2.4 requests-2.27.1 scikit-learn-0.23.2 scikit-plot-0.3
.7 scipy-1.5.4 smmap-5.0.0 tangled-up-in-unicode-0.1.0 umap-learn-0.5.2 visions-0.7.4 web
socket-client-1.2.3 yellowbrick-1.3.post1

Requirement already satisfied: pycaret in /usr/local/lib/python3.7/dist-packages (2.3.6)
Requirement already satisfied: scipy<=1.5.4 in /usr/local/lib/python3.7/dist-packages (fr
om pycaret) (1.5.4)
Requirement already satisfied: mlflow in /usr/local/lib/python3.7/dist-packages (from pyc
aret) (1.23.1)
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (from p
ycaret) (0.15.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from pyc
aret) (1.3.5)
Requirement already satisfied: plotly>=4.4.1 in /usr/local/lib/python3.7/dist-packages (f
rom pycaret) (5.5.0)
Requirement already satisfied: Boruta in /usr/local/lib/python3.7/dist-packages (from pyc
aret) (0.3)
Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-package
s (from pycaret) (0.17.3)
Requirement already satisfied: pandas-profiling>=2.8.0 in /usr/local/lib/python3.7/dist-p
ackages (from pycaret) (3.1.0)
Requirement already satisfied: pyLDAvis in /usr/local/lib/python3.7/dist-packages (from p
ycaret) (3.2.2)
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from py
caret) (5.5.0)
Requirement already satisfied: umap-learn in /usr/local/lib/python3.7/dist-packages (from
pycaret) (0.5.2)
Requirement already satisfied: pyyaml<6.0.0 in /usr/local/lib/python3.7/dist-packages (fr
om pycaret) (5.4.1)
Requirement already satisfied: spacy<2.4.0 in /usr/local/lib/python3.7/dist-packages (fro
m pycaret) (2.2.4)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (from
pycaret) (1.5.0)
Requirement already satisfied: pyod in /usr/local/lib/python3.7/dist-packages (from pycar
et) (0.9.7)
Requirement already satisfied: plotly in /usr/local/lib/python3.7/dist-packages (from pyca
```

Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.2.5)
Requirement already satisfied: yellowbrick>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.3.post1)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.2.2)
Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.6.0)
Requirement already satisfied: scikit-learn==0.23.2 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.23.2)
Requirement already satisfied: imbalanced-learn==0.7.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.7.0)
Requirement already satisfied: lightgbm>=2.3.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.3.2)
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.6.5)
Requirement already satisfied: mlxtend>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.19.0)
Requirement already satisfied: kmodes>=0.10.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.11.1)
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.11.2)
Requirement already satisfied: scikit-plot in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.3.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.1)
Requirement already satisfied: numpy>=1.13.3 in /usr/local/lib/python3.7/dist-packages (from imbalanced-learn==0.7.0->pycaret) (1.19.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.23.2->pycaret) (3.1.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (1.15.0)
Requirement already satisfied: setuptools>=34.4.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (57.4.0)
Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (0.3.0)
Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-packages (from gensim<4.0.0->pycaret) (5.2.1)
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (1.0.18)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (2.6.1)
Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.8.1)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.7.5)
Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (4.8.0)
Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (4.4.2)
Requirement already satisfied: ipython-genutils~=0.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (4.10.1)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (1.0.2)
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (5.1.3)
Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (3.5.2)
Requirement already satisfied: tornado>=4.0 in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (5.1.1)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (5.3.5)
Requirement already satisfied: wheel in /usr/local/lib/python3.7/dist-packages (from lightgbm>=2.3.1->pycaret) (0.37.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (3.0.7)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages

```
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-package
s (from matplotlib->pycaret) (1.3.2)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pack
ages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (fr
om nbformat>=4.2.0->ipywidgets->pycaret) (4.9.2)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-p
ackages (from nbformat>=4.2.0->ipywidgets->pycaret) (4.3.3)
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packag
es (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (4.11.1)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.7/dist-packages (f
rom jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (21.4.0)
Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.7/dis
t-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (5.4.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-package
s (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (3.10.0.2)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /usr/loca
l/lib/python3.7/dist-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets-
>pycaret) (0.18.1)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/dist-packages (fro
m importlib-resources>=1.4.0->jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycar
et) (3.7.0)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (fr
om pandas->pycaret) (2018.9)
Requirement already satisfied: missingno>=0.4.2 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (0.5.0)
Requirement already satisfied: visions[type_image_path]==0.7.4 in /usr/local/lib/python3.
7/dist-packages (from pandas-profiling>=2.8.0->pycaret) (0.7.4)
Requirement already satisfied: tangled-up-in-unicode==0.1.0 in /usr/local/lib/python3.7/d
ist-packages (from pandas-profiling>=2.8.0->pycaret) (0.1.0)
Requirement already satisfied: pydantic>=1.8.1 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (1.9.0)
Requirement already satisfied: htmlmin>=0.1.12 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (0.1.12)
Requirement already satisfied: multimethod>=1.4 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (1.7)
Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.7/dist-packages (
from pandas-profiling>=2.8.0->pycaret) (2.11.3)
Requirement already satisfied: requests>=2.24.0 in /usr/local/lib/python3.7/dist-packages
(from pandas-profiling>=2.8.0->pycaret) (2.27.1)
Requirement already satisfied: tqdm>=4.48.2 in /usr/local/lib/python3.7/dist-packages (fr
om pandas-profiling>=2.8.0->pycaret) (4.62.3)
Requirement already satisfied: phik>=0.11.1 in /usr/local/lib/python3.7/dist-packages (fr
om pandas-profiling>=2.8.0->pycaret) (0.12.0)
Requirement already satisfied: markupsafe~=2.0.1 in /usr/local/lib/python3.7/dist-package
s (from pandas-profiling>=2.8.0->pycaret) (2.0.1)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.7/dist-packages (f
rom visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (2.6.3)
Requirement already satisfied: imagehash in /usr/local/lib/python3.7/dist-packages (from
visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (4.2.1)
Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from vis
ions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (7.1.2)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.7/dist-packages
(from plotly>=4.4.1->pycaret) (8.0.1)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from pr
ompt-toolkit<2.0.0,>=1.0.4->IPython->pycaret) (0.2.5)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packag
es (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.7/dist-pac
kages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (1.24.3)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.7/dist-packages (fr
om requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.10)
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.7/dist
-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.0.12)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pack
ages (from spacy<2.4.0->pycaret) (0.9.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packa
ges (from spacy<2.4.0->pycaret) (2.0.6)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
kages (from spacy<2.4.0->pycaret) (3.0.6)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-packa
ges (from spacy<2.4.0->pycaret) (1.0.5)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packag
```

```
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packag
es (from spacy<2.4.0->pycaret) (1.1.3)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/dist
-packages (from spacy<2.4.0->pycaret) (1.0.6)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist-p
ackages (from spacy<2.4.0->pycaret) (1.0.0)
Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om spacy<2.4.0->pycaret) (7.4.0)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packag
es (from spacy<2.4.0->pycaret) (0.4.1)
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages
(from widgetsnbextension~=3.5.0->ipywidgets->pycaret) (5.3.1)
Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1.8.0)
Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages
(from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.13.1)
Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from
notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (5.6.1)
Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from
jupyter-client->ipykernel>=4.5.1->ipywidgets->pycaret) (22.3.0)
Requirement already satisfied: ptyprocess in /usr/local/lib/python3.7/dist-packages (from
terminado>=0.8.1->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.7.0
)
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.7/dist-packages (from
imagehash->visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (1.2.0)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.7/dist-packages (fro
m mlflow->pycaret) (1.3.0)
Requirement already satisfied: sqlparse>=0.3.1 in /usr/local/lib/python3.7/dist-packages
(from mlflow->pycaret) (0.4.2)
Requirement already satisfied: entrypoints in /usr/local/lib/python3.7/dist-packages (fro
m mlflow->pycaret) (0.4)
Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (1.4.31)
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (21.3)
Requirement already satisfied: Flask in /usr/local/lib/python3.7/dist-packages (from mlfl
ow->pycaret) (1.1.4)
Requirement already satisfied: gitpython>=2.1.0 in /usr/local/lib/python3.7/dist-packages
(from mlflow->pycaret) (3.1.27)
Requirement already satisfied: docker>=4.0.0 in /usr/local/lib/python3.7/dist-packages (f
rom mlflow->pycaret) (5.0.3)
Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.7/dist-packages (from
mlflow->pycaret) (7.1.2)
Requirement already satisfied: protobuf>=3.7.0 in /usr/local/lib/python3.7/dist-packages
(from mlflow->pycaret) (3.17.3)
Requirement already satisfied: querystring-parser in /usr/local/lib/python3.7/dist-packag
es (from mlflow->pycaret) (1.2.4)
Requirement already satisfied: gunicorn in /usr/local/lib/python3.7/dist-packages (from m
lflow->pycaret) (20.1.0)
Requirement already satisfied: databricks-cli>=0.8.7 in /usr/local/lib/python3.7/dist-pac
kages (from mlflow->pycaret) (0.16.4)
Requirement already satisfied: prometheus-flask-exporter in /usr/local/lib/python3.7/dist
-packages (from mlflow->pycaret) (0.18.7)
Requirement already satisfied: alembic in /usr/local/lib/python3.7/dist-packages (from ml
flow->pycaret) (1.7.6)
Requirement already satisfied: tabulate>=0.7.7 in /usr/local/lib/python3.7/dist-packages
(from databricks-cli>=0.8.7->mlflow->pycaret) (0.8.9)
Requirement already satisfied: websocket-client>=0.32.0 in /usr/local/lib/python3.7/dist-
packages (from docker>=4.0.0->mlflow->pycaret) (1.2.3)
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.7/dist-packages
(from gitpython>=2.1.0->mlflow->pycaret) (4.0.9)
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.7/dist-packages
(from gitdb<5,>=4.0.1->gitpython>=2.1.0->mlflow->pycaret) (5.0.0)
Requirement already satisfied: Mako in /usr/local/lib/python3.7/dist-packages (from alemb
ic->mlflow->pycaret) (1.1.6)
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.7/dist-packages
(from sqlalchemy->mlflow->pycaret) (1.1.2)
Requirement already satisfied: Werkzeug<2.0,>=0.15 in /usr/local/lib/python3.7/dist-packa
ges (from Flask->mlflow->pycaret) (1.0.1)
Requirement already satisfied: itsdangerous<2.0,>=0.24 in /usr/local/lib/python3.7/dist-p
ackages (from Flask->mlflow->pycaret) (1.1.0)
Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from n
```

```
bconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.5.0)
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-package
s (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.8.
4)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from
nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.7.1)
Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbc
onvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (4.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-pack
ages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1
.5.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (fr
om bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0
.5.1)
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.7/dist-package
s (from prometheus-flask-exporter->mlflow->pycaret) (0.13.1)
Requirement already satisfied: numexpr in /usr/local/lib/python3.7/dist-packages (from py
LDAvis->pycaret) (2.8.1)
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from pyL
DAvis->pycaret) (0.16.0)
Requirement already satisfied: funcy in /usr/local/lib/python3.7/dist-packages (from pyLD
Avis->pycaret) (1.17)
Requirement already satisfied: numba>=0.35 in /usr/local/lib/python3.7/dist-packages (fro
m pyod->pycaret) (0.51.2)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (fro
m pyod->pycaret) (0.10.2)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/di
st-packages (from numba>=0.35->pyod->pycaret) (0.34.0)
Requirement already satisfied: patsy>=0.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod->pycaret) (0.5.2)
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.7/dist-packages
(from umap-learn->pycaret) (0.5.6)
Requirement already satisfied: pyod in /usr/local/lib/python3.7/dist-packages (0.9.7)
Requirement already satisfied: scipy>=1.3.1 in /usr/local/lib/python3.7/dist-packages (fr
om pyod) (1.5.4)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from pyod)
(1.15.0)
Requirement already satisfied: numpy>=1.13 in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (1.19.5)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pyo
d) (1.0.1)
Requirement already satisfied: numba>=0.35 in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (0.51.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from
pyod) (3.2.2)
Requirement already satisfied: scikit-learn>=0.20.0 in /usr/local/lib/python3.7/dist-pack
ages (from pyod) (0.23.2)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (0.10.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from
numba>=0.35->pyod) (57.4.0)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/di
st-packages (from numba>=0.35->pyod) (0.34.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-pack
ages (from scikit-learn>=0.20.0->pyod) (3.1.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib
/python3.7/dist-packages (from matplotlib->pyod) (3.0.7)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pack
ages (from matplotlib->pyod) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-package
s (from matplotlib->pyod) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (fr
om matplotlib->pyod) (0.11.0)
Requirement already satisfied: patsy>=0.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod) (0.5.2)
Requirement already satisfied: pandas>=0.19 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod) (1.3.5)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (fr
om pandas>=0.19->statsmodels->pyod) (2018.9)
Requirement already satisfied: pyod in /usr/local/lib/python3.7/dist-packages (0.9.7)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (0.10.2)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from pyod)
```

```
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from pyod)
(1.15.0)
Requirement already satisfied: numba>=0.35 in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (0.51.2)
Requirement already satisfied: numpy>=1.13 in /usr/local/lib/python3.7/dist-packages (fro
m pyod) (1.19.5)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pyo
d) (1.0.1)
Requirement already satisfied: scipy>=1.3.1 in /usr/local/lib/python3.7/dist-packages (fr
om pyod) (1.5.4)
Requirement already satisfied: scikit-learn>=0.20.0 in /usr/local/lib/python3.7/dist-pack
ages (from pyod) (0.23.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from
pyod) (3.2.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (from
numba>=0.35->pyod) (57.4.0)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/di
st-packages (from numba>=0.35->pyod) (0.34.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-pack
ages (from scikit-learn>=0.20.0->pyod) (3.1.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (fr
om matplotlib->pyod) (0.11.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib
/python3.7/dist-packages (from matplotlib->pyod) (3.0.7)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-pack
ages (from matplotlib->pyod) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-package
s (from matplotlib->pyod) (1.3.2)
Requirement already satisfied: pandas>=0.19 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod) (1.3.5)
Requirement already satisfied: patsy>=0.4.0 in /usr/local/lib/python3.7/dist-packages (fr
om statsmodels->pyod) (0.5.2)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (fr
om pandas>=0.19->statsmodels->pyod) (2018.9)
```

1.3 Import libraries

In [ ]:

```
#Import libraries
import pandas as pd

import numpy as np

from scipy.stats import uniform
from pycaret.classification import *

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.ensemble import GradientBoostingClassifier

from pyod.models.cblof import CBLOF

import pickle
```

```
/usr/local/lib/python3.7/dist-packages/distributed/config.py:20: YAMLLoadWarning: calling
yaml.load() without Loader=... is deprecated, as the default Loader is unsafe. Please rea
d https://msg.pyyaml.org/load for full details.
  defaults = yaml.load(f)
```

1.4 Import data

In [ ]:

```
#Function named dataframe_optimizer is defined. This will reduce space consumption by dat
aframes.
```

```
#Credit - https://www.kaggle.com/rinnqd/reduce-memory-usage and
#https://www.analyticsvidhya.com/blog/2021/04/how-to-reduce-memory-usage-in-python-pandas
/
def dataframe_optimizer(df):
  '''This is a dataframe optimizer'''
  start_mem=np.round(df.memory_usage().sum()/1024**2,2)
  for col in df.columns:
    col_type=df[col].dtype
    if col_type!=object:
      c_min=df[col].min()
      c_max=df[col].max()
      if str(col_type)[:3]=='int':
        if c_min>np.iinfo(np.int8).min and c_max<np.iinfo(np.int8).max:
            df[col]=df[col].astype(np.int8)
        elif c_min>np.iinfo(np.int16).min and c_max<np.iinfo(np.int16).max:
            df[col]=df[col].astype(np.int16)
        elif c_min>np.iinfo(np.int32).min and c_max<np.iinfo(np.int32).max:
            df[col]=df[col].astype(np.int32)
        elif c_min>np.iinfo(np.int64).min and c_max<np.iinfo(np.int64).max:
            df[col]=df[col].astype(np.int64)
      else:
        if c_min>np.finfo(np.float16).min and c_max<np.finfo(np.float16).max:
            df[col]=df[col].astype(np.float16)
        elif c_min>np.finfo(np.float32).min and c_max<np.finfo(np.float32).max:
            df[col]=df[col].astype(np.float32)
        else:
            df[col]=df[col].astype(np.float64)
  end_mem=np.round(df.memory_usage().sum()/1024**2,2)
  return df
```

All the relevant files can be accessed through the following link:
https://drive.google.com/drive/folders/1evFZRwFWh4zkR9CiT46llB9PlaXFLfLA?usp=sharing

In [ ]:

```
#Read application_train
application_train = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project
/Data/application_train.csv'))

#Read application_test
application_test = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/
Data/application_test.csv'))

#Read bureau
bureau = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/burea
u.csv'))

#Read previous_application
previous_application = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Proj
ect/Data/previous_application.csv'))
```

In [ ]:

```
#Print the shape of imported data
print(application_train.shape)
print(application_test.shape)
print(bureau.shape)
print(previous_application.shape)
```

```
(307511, 122)
(48744, 121)
(1716428, 17)
(1670214, 37)
```

2.0 Feature Engineering and data merger

2.1 Create 3 new ratios from existing columns of application_train and application_test

application_test

In [ ]:

```
#Add columns titled DEBT_INCOME_RATIO to application_train
application_train['DEBT_INCOME_RATIO'] = application_train['AMT_ANNUITY']/application_tra
in['AMT_INCOME_TOTAL']

#Add columns titled LOAN_VALUE_RATIO to application_train
application_train['LOAN_VALUE_RATIO'] = application_train['AMT_CREDIT']/application_trai
n['AMT_GOODS_PRICE']

#Add columns titled LOAN_INCOME_RATIO to application_train
application_train['LOAN_INCOME_RATIO'] = application_train['AMT_CREDIT']/application_tra
in['AMT_INCOME_TOTAL']
```

2.2 Merge bureau with application_train

In [ ]:

```
#Create a dataframe with numerical columns of bureau
bureau_numerical = bureau.select_dtypes(exclude=object)
#Create a dataframe with categorical columns of bureau
bureau_categorical = bureau.select_dtypes(include=object)
```

In [ ]:

```
#Merge numerical features from bureau to application_train
bureau_numerical_merge = bureau_numerical.groupby(by=['SK_ID_CURR']).median().reset_inde
x()
application_train_bureau = application_train.merge(bureau_numerical_merge, on='SK_ID_CUR
R', how='left', suffixes=('', '_BUREAU'))

#Merge categorical features from bureau to application_train
bureau_categorical['SK_ID_CURR'] = bureau['SK_ID_CURR']
bureau_categorical_merge = bureau_categorical.groupby(by=['SK_ID_CURR']).agg(lambda x:x.
value_counts().index[0] if len(x.value_counts()) != 0 else '').reset_index()
application_train_bureau = application_train_bureau.merge(bureau_categorical_merge, on='S
K_ID_CURR', how='left', suffixes=('', '_BUREAU'))

#Drop SK_ID_BUREAU
application_train_bureau = application_train_bureau.drop(columns = ['SK_ID_BUREAU'])

#Shape of application and bureau data combined
print('The shape of application_train and bureau data merged: ', application_train_bureau
.shape)
```

```
The shape of application_train and bureau data merged:  (307511, 140)
```

In [ ]:

```
#Save the dataframes into CSV files for future use
bureau_numerical_merge.to_csv('bureau_numerical_merge.csv', index = False)
bureau_categorical_merge.to_csv('bureau_categorical_merge.csv', index = False)
```

2.3 Merge previous_application with application_train_bureau

In [ ]:

```
#Create a dataframe with numerical columns of previous_application
previous_application_numerical = previous_application.select_dtypes(exclude=object)
#Create a dataframe with categorical columns of previous_application
previous_application_categorical = previous_application.select_dtypes(include=object)
```

In [ ]:

```
#Merge numerical features from previous_application to application_train_bureau
previous_numerical_merge = previous_application_numerical.groupby(by=['SK_ID_CURR']).mea
```

```
n().reset_index()
application_train_bureau_previous = application_train_bureau.merge(previous_numerical_mer
ge, on='SK_ID_CURR', how='left', suffixes=('', '_PREVIOUS'))

#Merge categorical features from previous_application to application_train_bureau
previous_application_categorical['SK_ID_CURR'] = bureau['SK_ID_CURR']
previous_categorical_merge = previous_application_categorical.groupby(by=['SK_ID_CURR'])
.agg(lambda x:x.value_counts().index[0] if len(x.value_counts()) != 0 else '').reset_ind
ex()
application_train_bureau_previous = application_train_bureau_previous.merge(previous_cate
gorical_merge, on='SK_ID_CURR', how='left', suffixes=('', '_PREVIOUS'))

#Drop SK_ID_PREV
application_train_bureau_previous = application_train_bureau_previous.drop(columns = ['SK
_ID_PREV'])

#Shape of application_train_bureau and previous_application data combined
print('The shape of application_train_bureau and previous_application data merged: ', app
lication_train_bureau_previous.shape)
```

```
The shape of application_train_bureau and previous_application data merged:  (307511, 175
)
```

In [ ]:

```
#Save the dataframes into CSV files for future use
previous_numerical_merge.to_csv('previous_numerical_merge.csv', index = False)
previous_categorical_merge.to_csv('previous_categorical_merge.csv', index = False)
```

2.4 Prepare train data and save it and it's column names

In [ ]:

```
#Final train data ready for preprocessing
train_data = application_train_bureau_previous.drop(columns=['SK_ID_CURR'])
```

In [ ]:

```
#Save the dataframes into CSV files for future use
train_data.to_csv('train_data.csv', index = False)
```

3.0 Training and Pipeline using Pycaret

In [ ]:

```
#Save the list of columns in train_data
file = open('columns_query_data.pkl', 'wb')
pickle.dump(list(application_test.columns), file)
file.close()
```

3.1 Fetch data

In [ ]:

```
#Read train_data
train_data_full = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/D
ata/train_data.csv'))
```

3.2 Train model using 100% data

In [ ]:

```
#Make lists of numerical and categorical columns
columns_numerical = list(train_data_full.select_dtypes(exclude=object).columns)
```

```
columns_numerical.remove('TARGET')
columns_categorical = list(train_data_full.select_dtypes(include=object).columns)
```

In [ ]:

```
#Set up data for feeding to model
data = setup(data=train_data_full, target="TARGET", categorical_features=columns_categori
cal, numeric_features=columns_numerical, train_size=0.9,
             numeric_imputation='median', normalize=True, remove_outliers=True, data_spl
it_stratify=True, feature_selection=True, feature_selection_threshold=0.35)
```

	Description	Value
0	session_id	202
1	Target	TARGET
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(307511, 174)
5	Missing Values	True
6	Numeric Features	138
7	Categorical Features	35
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(262921, 199)
12	Transformed Test Set	(30752, 199)
13	Shuffle Train-Test	True
14	Stratify Train-Test	True
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	False
20	Experiment Name	clf-default-name
21	USI	ba25
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	median
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent
29	Normalize	True
30	Normalize Method	zscore
31	Transformation	False
32	Transformation Method	None
33	PCA	False
34	PCA Method	None
35	PCA Components	None
36	Ignore Low Variance	False

	Description	Value
37	Combine Rare Levels	False
38	Rare Level Threshold	None
39	Numeric Binning	False
40	Remove Outliers	True
41	Outliers Threshold	0.05
42	Remove Multicollinearity	False
43	Multicollinearity Threshold	None
44	Remove Perfect Collinearity	True
45	Clustering	False
46	Clustering Iteration	None
47	Polynomial Features	False
48	Polynomial Degree	None
49	Trignometry Features	False
50	Polynomial Threshold	None
51	Group Features	False
52	Feature Selection	True
53	Feature Selection Method	classic
54	Features Selection Threshold	0.35
55	Feature Interaction	False
56	Feature Ratio	False
57	Interaction Threshold	None
58	Fix Imbalance	False
59	Fix Imbalance Method	SMOTE

In [ ]:

```
#Train model
model = create_model('lightgbm')
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9191	0.7646	0.0225	0.5333	0.0432	0.0369	0.0971
1	0.9187	0.7574	0.0239	0.4722	0.0455	0.0380	0.0920
2	0.9189	0.7671	0.0206	0.5057	0.0397	0.0335	0.0896
3	0.9188	0.7645	0.0230	0.4804	0.0439	0.0367	0.0913
4	0.9195	0.7615	0.0248	0.5889	0.0477	0.0414	0.1090
5	0.9197	0.7699	0.0234	0.6410	0.0452	0.0397	0.1119
6	0.9193	0.7699	0.0239	0.5667	0.0459	0.0396	0.1042
7	0.9199	0.7694	0.0272	0.6444	0.0522	0.0459	0.1209
8	0.9190	0.7729	0.0220	0.5165	0.0423	0.0359	0.0940
9	0.9192	0.7617	0.0300	0.5378	0.0568	0.0487	0.1128
Mean	0.9192	0.7659	0.0241	0.5487	0.0462	0.0396	0.1023
SD	0.0004	0.0046	0.0026	0.0578	0.0047	0.0044	0.0104

In [ ]:

```
#Tune model
tuned_model = tune_model(model)
```

Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9194	0.7658	0.0370	0.5524	0.0694	0.0598	0.1277
1	0.9193	0.7563	0.0361	0.5347	0.0677	0.0580	0.1233
2	0.9188	0.7671	0.0361	0.4873	0.0672	0.0567	0.1157
3	0.9189	0.7627	0.0347	0.5034	0.0649	0.0551	0.1160
4	0.9193	0.7557	0.0380	0.5400	0.0710	0.0609	0.1273
5	0.9194	0.7697	0.0361	0.5540	0.0678	0.0584	0.1263
6	0.9187	0.7718	0.0323	0.4825	0.0606	0.0510	0.1087
7	0.9195	0.7679	0.0403	0.5513	0.0751	0.0648	0.1330
8	0.9188	0.7666	0.0398	0.4913	0.0737	0.0623	0.1223
9	0.9195	0.7601	0.0450	0.5486	0.0832	0.0718	0.1401
Mean	0.9192	0.7644	0.0376	0.5246	0.0701	0.0599	0.1240
SD	0.0003	0.0052	0.0033	0.0282	0.0059	0.0054	0.0086

In [ ]:

```
#Save best model and store it in Google Drive for future use
save_model(tuned_model, "model")
```

```
Transformation Pipeline and Model Successfully Saved
```

Out[ ]:

```
(Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=['NAME_CONTRACT_TYPE',
                                                             'CODE_GENDER',
                                                             'FLAG_OWN_CAR',
                                                             'FLAG_OWN_REALTY',
                                                             'NAME_TYPE_SUITE',
                                                             'NAME_INCOME_TYPE',
                                                             'NAME_EDUCATION_TYPE',
                                                             'NAME_FAMILY_STATUS',
                                                             'NAME_HOUSING_TYPE',
                                                             'OCCUPATION_TYPE',
                                                             'WEEKDAY_APPR_PROCESS_START
',
                                                             'ORGANIZATION_TYPE',
                                                             'FONDKAPREMONT_MODE',
                                                             'HOUSETYPE_MO...
                                      boosting_type='gbdt', class_weight=None,
                                      colsample_bytree=1.0, feature_fraction=1.0,
                                      importance_type='split', learning_rate=0.1,
                                      max_depth=-1, min_child_samples=76,
                                      min_child_weight=0.001, min_split_gain=0.2,
                                      n_estimators=160, n_jobs=-1, num_leaves=90,
                                      objective=None, random_state=8025,
                                      reg_alpha=0.15, reg_lambda=4, silent='warn',
                                      subsample=1.0, subsample_for_bin=200000,
                                      subsample_freq=0)]],
          verbose=False), 'model.pkl')
```

Remarks: This model is not getting successfully deployed on Heroku. The model size is 172 MB. We shall try with 50% and 25% of total data points in case size of the model is an issue.

3.3 Train model using 25% data

In [ ]:

```
#Using train test split to extract 25% of train data with stratification
train_data, X_test, y_train, y_test = train_test_split( train_data_full, train_data_full
['TARGET'],
```

```
                                                    test_size=0.75, random_state=42,
stratify=train_data_full['TARGET'])
```

In [ ]:

```
#Make lists of numerical and categorical columns
columns_numerical = list(train_data.select_dtypes(exclude=object).columns)
columns_numerical.remove('TARGET')
columns_categorical = list(train_data.select_dtypes(include=object).columns)
```

In [ ]:

```
#Set up data for feeding to model
data = setup(data=train_data, target="TARGET", categorical_features=columns_categorical,
numeric_features=columns_numerical, train_size=0.9,
             numeric_imputation='median', normalize=True, remove_outliers=True, data_spl
it_stratify=True, feature_selection=True, feature_selection_threshold=0.35)
```

	Description	Value
0	session_id	771
1	Target	TARGET
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(76877, 174)
5	Missing Values	True
6	Numeric Features	138
7	Categorical Features	35
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(65729, 194)
12	Transformed Test Set	(7688, 194)
13	Shuffle Train-Test	True
14	Stratify Train-Test	True
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	False
20	Experiment Name	clf-default-name
21	USI	1d1c
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	median
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent
29	Normalize	True
30	Normalize Method	zscore
31	Transformation	False

	Description	Value
32	Transformation Method	None
33	PCA	False
34	PCA Method	None
35	PCA Components	None
36	Ignore Low Variance	False
37	Combine Rare Levels	False
38	Rare Level Threshold	None
39	Numeric Binning	False
40	Remove Outliers	True
41	Outliers Threshold	0.05
42	Remove Multicollinearity	False
43	Multicollinearity Threshold	None
44	Remove Perfect Collinearity	True
45	Clustering	False
46	Clustering Iteration	None
47	Polynomial Features	False
48	Polynomial Degree	None
49	Trignometry Features	False
50	Polynomial Threshold	None
51	Group Features	False
52	Feature Selection	True
53	Feature Selection Method	classic
54	Features Selection Threshold	0.35
55	Feature Interaction	False
56	Feature Ratio	False
57	Interaction Threshold	None
58	Fix Imbalance	False
59	Fix Imbalance Method	SMOTE

In [ ]:

```
#Train model
model = create_model('lightgbm')
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9192	0.7522	0.0264	0.5000	0.0501	0.0423	0.1006
1	0.9185	0.7470	0.0132	0.3684	0.0255	0.0200	0.0568
2	0.9178	0.7449	0.0132	0.3043	0.0253	0.0187	0.0486
3	0.9197	0.7539	0.0188	0.5882	0.0365	0.0316	0.0948
4	0.9191	0.7524	0.0245	0.4815	0.0466	0.0391	0.0944
5	0.9189	0.7572	0.0263	0.4828	0.0499	0.0419	0.0981
6	0.9171	0.7357	0.0188	0.3030	0.0354	0.0262	0.0578
7	0.9180	0.7488	0.0169	0.3600	0.0323	0.0252	0.0632
8	0.9195	0.7581	0.0263	0.5600	0.0503	0.0433	0.1085
9	0.9194	0.7335	0.0301	0.5161	0.0569	0.0485	0.1100
Mean	0.9187	0.7484	0.0215	0.4464	0.0409	0.0337	0.0833
SD	0.0008	0.0079	0.0057	0.0988	0.0107	0.0101	0.0225

In [ ]:

```
#Tune model
tuned_model = tune_model(model)
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9197	0.7564	0.0132	0.6364	0.0258	0.0226	0.0835
1	0.9191	0.7562	0.0113	0.4615	0.0221	0.0183	0.0622
2	0.9198	0.7587	0.0245	0.5909	0.0470	0.0409	0.1085
3	0.9189	0.7573	0.0188	0.4545	0.0362	0.0299	0.0795
4	0.9198	0.7504	0.0207	0.6111	0.0401	0.0350	0.1020
5	0.9198	0.7650	0.0226	0.6316	0.0436	0.0382	0.1087
6	0.9188	0.7444	0.0150	0.4444	0.0291	0.0239	0.0698
7	0.9195	0.7572	0.0132	0.6364	0.0258	0.0226	0.0834
8	0.9191	0.7617	0.0113	0.5000	0.0221	0.0186	0.0657
9	0.9200	0.7410	0.0264	0.6087	0.0505	0.0441	0.1148
Mean	0.9194	0.7548	0.0177	0.5576	0.0342	0.0294	0.0878
SD	0.0004	0.0071	0.0053	0.0777	0.0101	0.0091	0.0184

In [ ]:

```
#Save best model and store it in Google Drive for future use
save_model(tuned_model, "model")
```

```
Transformation Pipeline and Model Successfully Saved
```

Out[ ]:

```
(Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=['NAME_CONTRACT_TYPE',
                                                             'CODE_GENDER',
                                                             'FLAG_OWN_CAR',
                                                             'FLAG_OWN_REALTY',
                                                             'NAME_TYPE_SUITE',
                                                             'NAME_INCOME_TYPE',
                                                             'NAME_EDUCATION_TYPE',
                                                             'NAME_FAMILY_STATUS',
                                                             'NAME_HOUSING_TYPE',
                                                             'OCCUPATION_TYPE',
                                                             'WEEKDAY_APPR_PROCESS_START
',
                                                             'ORGANIZATION_TYPE',
                                                             'FONDKAPREMONT_MODE',
                                                             'HOUSETYPE_MO...
                                      boosting_type='gbdt', class_weight=None,
                                      colsample_bytree=1.0, feature_fraction=1.0,
                                      importance_type='split', learning_rate=0.1,
                                      max_depth=-1, min_child_samples=36,
                                      min_child_weight=0.001, min_split_gain=0.7,
                                      n_estimators=190, n_jobs=-1, num_leaves=4,
                                      objective=None, random_state=771,
                                      reg_alpha=0.005, reg_lambda=5, silent='warn',
                                      subsample=1.0, subsample_for_bin=200000,
                                      subsample_freq=0)]],
          verbose=False), 'model.pkl')
```

3.4 Create Pipeline and predict

In [ ]:

```
#Import saved data and pickle files
```

```
#Import saved data and pickle files
bureau_numerical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/bureau_numerical_merge.csv'))
bureau_categorical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/bureau_categorical_merge.csv'))
previous_numerical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/previous_numerical_merge.csv'))
previous_categorical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/previous_categorical_merge.csv'))
filename = open('/content/drive/MyDrive/AI_ML_Project/Data/columns_train_data.pkl', 'rb')
columns = pickle.load(filename)
filename.close()
tuned_model = load_model('/content/drive/MyDrive/AI_ML_Project/Data/model')
```

```
Transformation Pipeline and Model Successfully Loaded
```

In [ ]:

```
#Read query data point(s)
query = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/application_test.csv'))
```

In [ ]:

```
#Define a function to create a pipeline for prediction
def inference(query):

  #Add columns titled DEBT_INCOME_RATIO to application_train
  query['DEBT_INCOME_RATIO'] = query['AMT_ANNUITY']/query['AMT_INCOME_TOTAL']

  #Add columns titled LOAN_VALUE_RATIO to application_train
  query['LOAN_VALUE_RATIO'] = query['AMT_CREDIT']/query['AMT_GOODS_PRICE']

  #Add columns titled LOAN_INCOME_RATIO to application_train
  query['LOAN_INCOME_RATIO'] = query['AMT_CREDIT']/query['AMT_INCOME_TOTAL']

  #Merge numerical features from bureau to query
  query_bureau = query.merge(bureau_numerical_merge, on='SK_ID_CURR', how='left', suffixes=('', '_BUREAU'))

  #Merge categorical features from bureau to query
  query_bureau = query_bureau.merge(bureau_categorical_merge, on='SK_ID_CURR', how='left', suffixes=('', '_BUREAU'))

  #Drop SK_ID_BUREAU
  query_bureau = query_bureau.drop(columns = ['SK_ID_BUREAU'])

  #Shape of query and bureau data combined
  print('The shape of query and bureau data merged: ', query_bureau.shape)

  #Merge numerical features from previous_application to query_bureau
  query_bureau_previous = query_bureau.merge(previous_numerical_merge, on='SK_ID_CURR', how='left', suffixes=('', '_PREVIOUS'))

  #Merge categorical features from previous_application to query_bureau
  query_bureau_previous = query_bureau_previous.merge(previous_categorical_merge, on='SK_ID_CURR', how='left', suffixes=('', '_PREVIOUS'))

  #Drop SK_ID_PREV and SK_ID_CURR
  query_bureau_previous = query_bureau_previous.drop(columns = ['SK_ID_PREV'])

  #Shape of query_bureau and previous_application data combined
  print('The shape of query_bureau and previous_application data merged: ', query_bureau_previous.shape)

  #Drop SK_ID_PREV and SK_ID_CURR
  query_bureau_previous = query_bureau_previous.drop(columns = ['SK_ID_CURR'])

  missing_columns = set(list(columns)) - set(['TARGET']) - set(list(query_bureau_previous.columns))
  if len(missing_columns) != 0:
    print("Please enter values for all columns")
```

```
  else:
    predictions = predict_model(tuned_model, query_bureau_previous)
    return predictions
```

In [ ]:

```
#Show predictions
query_prediction = inference(query)
query_prediction
```

```
The shape of query and bureau data merged:  (48744, 139)
The shape of query_bureau and previous_application data merged:  (48744, 174)
```

Out[ ]:

	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOT
0	Cash loans	F	N	Y	0	13500
1	Cash loans	M	N	Y	0	9900
2	Cash loans	M	Y	Y	0	20250
3	Cash loans	F	N	Y	2	31500
4	Cash loans	M	Y	N	1	18000
...	...	...	...	...	...	
48739	Cash loans	F	N	Y	0	12150
48740	Cash loans	F	N	N	2	15750
48741	Cash loans	F	Y	Y	1	20250
48742	Cash loans	M	N	N	0	22500
48743	Cash loans	F	Y	N	0	13500

48744 rows × 175 columns

3.5 Conclusion

1. Pycaret is a very convenient tool for data pre-processing and model training. A one line code for data set-up does all the pre-processing. Further one line code can prepare a model and another line can tune it.
2. Model created by Pycaret is huge in size. Model created using 100% train_data was 173MB in size. Predictions are being made in Google colaboratory using saved models. However, deployment on Heroku throws an error. So models were trained using 50% and 25% data thinking that error in deployment was due to size of the model. Error persisted with models prepared using 50% and 25% data.
3. It was decided to switch to Sklearn and try to deploy a model created by using Sklearn. This is implemented in the next sub-section.

4.0 Training and Pipeline using Sklearn

4.1 Define list of column names for use in pipeline

In [ ]:

```
#Read application_test
application_test = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/
Data/application_test.csv'))

#Read columns from application_test
columns_input = list(application_test.columns)

#Save columns_input
file = open('columns_input.pkl', 'wb')
pickle.dump(columns_input, file)
file.close()
```

4.2 Data Preparation

In [ ]:

```
#Read train_data
train_data = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/t
rain_data.csv'))
```

In [ ]:

```
#Make lists of numerical and categorical columns
y_train = train_data['TARGET']
X_train_numerical = train_data.select_dtypes(exclude=object).drop(columns=['TARGET'])
X_train_categorical = train_data.select_dtypes(include=object)
columns_numerical = X_train_numerical.columns
columns_categorical = X_train_categorical.columns
```

4.3 Imputation and scaling of numerical data

In [ ]:

```
#Imputation of missing data
imputer = SimpleImputer(missing_values=np.nan, strategy='median')
imputer.fit(X_train_numerical)
X_train_numerical_imputed = imputer.transform(X_train_numerical)
```

In [ ]:

```
#Save imputer
file = open('imputer.pkl', 'wb')
pickle.dump(imputer, file)
file.close()
```

In [ ]:

```
#Scaling of data
scaler = StandardScaler()
scaler.fit(X_train_numerical_imputed)
X_train_numerical_imputed_scaled = scaler.transform(X_train_numerical_imputed)
X_train_numerical_imputed_scaled_df = pd.DataFrame(data = X_train_numerical_imputed_scale
d, columns = columns_numerical)
```

In [ ]:

```
#Save scaler
file = open('scaler.pkl', 'wb')
pickle.dump(scaler, file)
file.close()
```

4.4 One hot encoding of categorical data

In [ ]:

```
#Imputation of missing data
imputer_constant = SimpleImputer(strategy='constant', fill_value='missing_vale')
imputer_constant.fit(X_train_categorical)
X_train_categorical_imputed = imputer_constant.transform(X_train_categorical)
```

In [ ]:

```
#Save imputer_constant
file = open('imputer_constant.pkl', 'wb')
pickle.dump(imputer_constant, file)
file.close()
```

In [ ]:

```
#One hot encoding of categorical data
ohe = OneHotEncoder(handle_unknown='ignore')
ohe.fit(X_train_categorical_imputed)
X_train_categorical_imputed_ohe = ohe.transform(X_train_categorical_imputed)
columns_ohe = ohe.get_feature_names(input_features=columns_categorical)
X_train_categorical_imputed_ohe_df = pd.DataFrame(data = X_train_categorical_imputed_ohe.
toarray(), columns = list(columns_ohe))
```

In [ ]:

```
#Save ohe columns
file = open('columns_ohe.pkl', 'wb')
pickle.dump(columns_ohe, file)
file.close()
```

In [ ]:

```
#Save ohe
file = open('ohe.pkl', 'wb')
pickle.dump(ohe, file)
file.close()
```

4.5 Define train data with all columns

In [ ]:

```
#Define train data with all columns
X_train_all_columns = pd.concat([X_train_numerical_imputed_scaled_df, X_train_categorical
_imputed_ohe_df], axis = 1)
```

4.6 Outlier removal

In [ ]:

```
#Define outlier detector and fit it to X_train_all_columns with contamination = 0.05
clf = CBLOF(contamination=0.05, check_estimator=False, random_state=42)
clf.fit(X_train_all_columns)
scores_pred = clf.decision_function(X_train_all_columns) * -1

#Predict the datapoints as outlier or inlier
outlier_prediction = clf.predict(X_train_all_columns)
inliers = len(outlier_prediction) - np.count_nonzero(outlier_prediction)
outliers = np.count_nonzero(outlier_prediction == 1)
```

In [ ]:

```
#Remove outliers
X_train_all_columns_outlier_label = X_train_all_columns.copy()
X_train_all_columns_outlier_label['outlier'] = outlier_prediction.tolist()
X_y_train_all_columns_outlier_label = pd.concat([X_train_all_columns_outlier_label, y_tra
in], axis = 1)
X_y_train_final_outlier_removed = X_y_train_all_columns_outlier_label[X_y_train_all_colum
ns_outlier_label['outlier'] != 1]
```

```
X_train = X_y_train_final_outlier_removed.drop(columns = ['TARGET', 'outlier'])
y_train = X_y_train_final_outlier_removed['TARGET']
```

4.7 Feature Selection

In [ ]:

```
#Define model for feature slecetion
model_feature_slection = GradientBoostingClassifier(random_state=0).fit(X_train, y_train
)
```

In [ ]:

```
#Select features
feature_importance = pd.DataFrame(model_feature_slection.feature_importances_, index=X_tr
ain.columns, columns=['importance']).sort_values('importance', ascending=False)
selected_features = list(feature_importance['importance'].head(175).index)
```

In [ ]:

```
#Save selected columns
file = open('selected_features.pkl', 'wb')
pickle.dump(selected_features, file)
file.close()
```

4.8 Train model

In [ ]:

```
#Define model
model = GradientBoostingClassifier(random_state=0).fit(X_train[selected_features], y_tra
in)
```

In [ ]:

```
#Save model
file = open('model.pkl', 'wb')
pickle.dump(model, file)
file.close()
```

4.9 Create Pipeline and predict

In [ ]:

```
#Import saved data and pickle files
bureau_numerical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Pr
oject/Data/bureau_numerical_merge.csv'))
bureau_categorical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_
Project/Data/bureau_categorical_merge.csv'))
previous_numerical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_
Project/Data/previous_numerical_merge.csv'))
previous_categorical_merge = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_M
L_Project/Data/previous_categorical_merge.csv'))
filename = open('/content/drive/MyDrive/AI_ML_Project/Data/columns_input.pkl', 'rb')
columns_input = pickle.load(filename)
filename.close()
filename1 = open('/content/drive/MyDrive/AI_ML_Project/Data/model.pkl', 'rb')
model = pickle.load(filename1)
filename1.close()
filename2 = open('/content/drive/MyDrive/AI_ML_Project/Data/imputer.pkl', 'rb')
imputer = pickle.load(filename2)
filename2.close()
filename3 = open('/content/drive/MyDrive/AI_ML_Project/Data/scaler.pkl', 'rb')
scaler = pickle.load(filename3)
filename3.close()
filename4 = open('/content/drive/MyDrive/AI_ML_Project/Data/imputer_constant.pkl', 'rb')
```

```
imputer_constant = pickle.load(filename4)
filename4.close()
filename5 = open('/content/drive/MyDrive/AI_ML_Project/Data/ohe.pkl', 'rb')
ohe = pickle.load(filename5)
filename5.close()
filename6 = open('/content/drive/MyDrive/AI_ML_Project/Data/selected_features.pkl', 'rb')
selected_features = pickle.load(filename6)
filename6.close()
filename7 = open('/content/drive/MyDrive/AI_ML_Project/Data/columns_ohe.pkl', 'rb')
columns_ohe = pickle.load(filename7)
filename7.close()
```

In [ ]:

```
#Define a function to create a pipeline for prediction
def inference(query):
  #Add columns titled DEBT_INCOME_RATIO, LOAN_VALUE_RATIO & LOAN_INCOME_RATIO to a copy o
f query data
  query_with_additinal_features = query.copy()
  query_with_additinal_features['DEBT_INCOME_RATIO'] = query_with_additinal_features['AMT
_ANNUITY']/query_with_additinal_features['AMT_INCOME_TOTAL']
  query_with_additinal_features['LOAN_VALUE_RATIO'] = query_with_additinal_features['AMT_
CREDIT']/query_with_additinal_features['AMT_GOODS_PRICE']
  query_with_additinal_features['LOAN_INCOME_RATIO'] = query_with_additinal_features['AMT
_CREDIT']/query_with_additinal_features['AMT_INCOME_TOTAL']

  #Merge numerical features from bureau to query data
  query_bureau = query_with_additinal_features.merge(bureau_numerical_merge, on='SK_ID_C
URR', how='left', suffixes=('', '_BUREAU'))

  #Merge categorical features from bureau to query data
  query_bureau = query_bureau.merge(bureau_categorical_merge, on='SK_ID_CURR', how='left
', suffixes=('', '_BUREAU'))

  #Drop SK_ID_BUREAU
  query_bureau = query_bureau.drop(columns = ['SK_ID_BUREAU'])

  #Shape of query and bureau data combined
  #print('The shape of query and bureau data merged: ', query_bureau.shape)

  #Merge numerical features from previous_application to query_bureau
  query_bureau_previous = query_bureau.merge(previous_numerical_merge, on='SK_ID_CURR',
how='left', suffixes=('', '_PREVIOUS'))

  #Merge categorical features from previous_application to query_bureau
  query_bureau_previous = query_bureau_previous.merge(previous_categorical_merge, on='SK
_ID_CURR', how='left', suffixes=('', '_PREVIOUS'))

  #Drop SK_ID_PREV
  query_bureau_previous = query_bureau_previous.drop(columns = ['SK_ID_PREV'])

  #Shape of query_bureau and previous_application data combined
  #print('The shape of query_bureau and previous_application data merged: ', query_bureau
_previous.shape)

  #Drop SK_ID_CURR
  query_bureau_previous = query_bureau_previous.drop(columns = ['SK_ID_CURR'])

  query_numerical = query_bureau_previous.select_dtypes(exclude=object)
  query_categorical = query_bureau_previous.select_dtypes(include=object)

  columns_numerical = query_numerical.columns
  columns_categorical = query_categorical.columns

  query_numerical_imputed_scaled_df = imputer.transform(query_numerical)
  query_numerical_imputed_scaled_df = scaler.transform(query_numerical_imputed_scaled_df)
  query_numerical_imputed_scaled_df = pd.DataFrame(data = query_numerical_imputed_scaled_
df, columns = columns_numerical)

  query_categorical_imputed_ohe_df = imputer_constant.transform(query_categorical)
  query_categorical_imputed_ohe_df = ohe.transform(query_categorical_imputed_ohe_df)
```

```
  query_categorical_imputed_ohe_df = pd.DataFrame(data = query_categorical_imputed_ohe_df.toarray(), columns = list(columns_ohe))

  query_data_all_features = pd.concat([query_numerical_imputed_scaled_df, query_categorical_imputed_ohe_df], axis = 1)
  query_data = query_data_all_features[selected_features]

  predictions = model.predict(query_data)
  return predictions

  #missing_columns = set(list(columns)) - set(['TARGET']) - set(list(query_bureau_previous.columns))
  #if len(missing_columns) != 0:
  #  print("Please enter values for all columns")
  #else:
  #  predictions = predict_model(tuned_model, query_bureau_previous)
  #  return predictions
```

In [ ]:

```
#Read query data point(s)
query = dataframe_optimizer(pd.read_csv('/content/drive/MyDrive/AI_ML_Project/Data/application_test.csv'))

columns_query = list(query.columns)

if columns_query == columns_input:
  query_prediction = inference(query)
  query_data_with_prediction = query.copy()
  query_data_with_prediction['LABEL'] = query_prediction
  conditions = [(query_data_with_prediction['LABEL'] == 0), (query_data_with_prediction['LABEL'] == 1)]
  values = ['NO', 'YES']
  query_data_with_prediction['DEFAULT TENDENCY'] = np.select(conditions, values)
  query_data_with_prediction = query_data_with_prediction.drop(columns = ['LABEL'])
  display(query_data_with_prediction)
else:
  print("Query columns do not match the columns of required format. Please input in the given format.")
```

	SK_ID_CURR	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT
0	100001	Cash loans	F	N	Y	0	
1	100005	Cash loans	M	N	Y	0	
2	100013	Cash loans	M	Y	Y	0	
3	100028	Cash loans	F	N	Y	2	
4	100038	Cash loans	M	Y	N	1	
...	...	...	...	...	...	...	...
48739	456221	Cash loans	F	N	Y	0	
48740	456222	Cash loans	F	N	N	2	
48741	456223	Cash loans	F	Y	Y	1	
48742	456224	Cash loans	M	N	N	0	
48743	456250	Cash loans	F	Y	N	0	

48744 rows × 122 columns

4.10 Conclusion

1. **Number of lines of code increases significantly when compared to Pycaret.**
2. **Models created are significantly smaller in size compared to models created by Pycaret.**
3. **Deployment on Heroku did not throw any error.**
4. **Deployment can be accessed from https://deployment-0.herokuapp.com/.**