



In [1]:

```
import sys  
!{sys.executable} -m pip install pycaret
```

Requirement already satisfied: pycaret in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (2.1.2)

Requirement already satisfied: joblib in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.14.1)

Requirement already satisfied: gensim in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (3.8.3)

Requirement already satisfied: numpy>=1.17 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.18.1)

Requirement already satisfied: catboost>=0.23.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.24.1)

Requirement already satisfied: xgboost>=0.90 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.2.0)

Requirement already satisfied: yellowbrick>=1.0.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.1)

Requirement already satisfied: wordcloud in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.8.0)

Requirement already satisfied: seaborn in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.10.0)

Requirement already satisfied: pyod in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.8.3)

Requirement already satisfied: spacy in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (2.3.2)

Requirement already satisfied: IPython in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (7.12.0)

Requirement already satisfied: lightgbm>=2.3.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (3.0.0)

Requirement already satisfied: scikit-learn>=0.23 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.23.2)

Requirement already satisfied: cufflinks>=0.17.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.17.3)

Requirement already satisfied: pyLDAvis in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (2.1.2)

Requirement already satisfied: ipywidgets in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (7.5.1)

Requirement already satisfied: textblob in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.15.3)

Requirement already satisfied: pandas-profiling>=2.8.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (2.9.0)

Requirement already satisfied: datefinder>=0.7.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.7.1)

Requirement already satisfied: mlflow in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.11.0)

Requirement already satisfied: imbalanced-learn>=0.6.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.7.0)

Requirement already satisfied: pandas in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (1.0.1)

Requirement already satisfied: umap-learn in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.4.6)

Requirement already satisfied: nltk in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (3.4.5)

Requirement already satisfied: matplotlib in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (3.1.3)

Requirement already satisfied: mlxtend in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.17.3)

Requirement already satisfied: plotly>=4.4.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (4.10.0)

Requirement already satisfied: kmodes>=0.10.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pycaret) (0.10.2)

Requirement already satisfied: six>=1.5.0 in /Users/keirynhart/opt/a

anaconda3/lib/python3.7/site-packages (from gensim->pycaret) (1.14.0)  
Requirement already satisfied: scipy>=0.18.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from gensim->pycaret) (1.4.1)  
Requirement already satisfied: smart-open>=1.8.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from gensim->pycaret) (2.2.0)  
Requirement already satisfied: graphviz in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from catboost>=0.23.2->pycaret) (0.14.1)  
Requirement already satisfied: cycler>=0.10.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from yellowbrick>=1.0.1->pycaret) (0.10.0)  
Requirement already satisfied: pillow in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from wordcloud->pycaret) (7.0.0)  
Requirement already satisfied: suod in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyod->pycaret) (0.0.4)  
Requirement already satisfied: statsmodels in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyod->pycaret) (0.11.0)  
Requirement already satisfied: combo in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyod->pycaret) (0.1.1)  
Requirement already satisfied: numba>=0.35 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyod->pycaret) (0.48.0)  
Requirement already satisfied: thinc==7.4.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (7.4.1)  
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (4.50.0)  
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (3.0.2)  
Requirement already satisfied: blis<0.5.0,>=0.4.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (0.4.1)  
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (2.24.0)  
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (1.0.2)  
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (0.8.0)  
Requirement already satisfied: setuptools in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (46.0.0.post20200309)  
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (2.0.3)  
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (1.1.3)  
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (1.0.2)  
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from spacy->pycaret) (1.0.0)  
Requirement already satisfied: traitlets>=4.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (4.

3.3)

Requirement already satisfied: jedi>=0.10 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (0.14.1)

Requirement already satisfied: prompt-toolkit!=3.0.0,!3.0.1,<3.1.0,>=2.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (3.0.3)

Requirement already satisfied: pygments in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (2.5.2)

Requirement already satisfied: pexpect; sys\_platform != "win32" in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (4.8.0)

Requirement already satisfied: backcall in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (0.1.0)

Requirement already satisfied: appnope; sys\_platform == "darwin" in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (0.1.0)

Requirement already satisfied: pickleshare in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (0.7.5)

Requirement already satisfied: decorator in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from IPython->pycaret) (4.4.1)

Requirement already satisfied: threadpoolctl>=2.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from scikit-learn>=0.23->pycaret) (2.1.0)

Requirement already satisfied: colorlover>=0.2.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from cufflinks>=0.17.0->pycaret) (0.3.0)

Requirement already satisfied: Jinja2>=2.7.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (2.11.1)

Requirement already satisfied: pytest in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (5.3.5)

Requirement already satisfied: future in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (0.18.2)

Requirement already satisfied: funcy in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (1.15)

Requirement already satisfied: numexpr in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (2.7.1)

Requirement already satisfied: wheel>=0.23.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pyLDavis->pycaret) (0.34.2)

Requirement already satisfied: ipykernel>=4.5.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from ipywidgets->pycaret) (5.1.4)

Requirement already satisfied: nbformat>=4.2.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from ipywidgets->pycaret) (5.0.4)

Requirement already satisfied: widgetsnbextension~=3.5.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from ipywidgets->pycaret) (3.5.1)

Requirement already satisfied: htmlmin>=0.1.12 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (0.1.12)

Requirement already satisfied: visions[type\_image\_path]==0.5.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (0.5.0)

Requirement already satisfied: tangled-up-in-unicode>=0.0.6 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (0.0.6)

Requirement already satisfied: attrs>=19.3.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (19.3.0)

t/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (19.3.0)  
Requirement already satisfied: phik>=0.9.10 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (0.10.0)  
Requirement already satisfied: missingno>=0.4.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (0.4.2)  
Requirement already satisfied: confuse>=1.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pandas-profiling>=2.8.0->pycaret) (1.3.0)  
Requirement already satisfied: pytz in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from datefinder>=0.7.0->pycaret) (2019.3)  
Requirement already satisfied: python-dateutil>=2.4.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from datefinder>=0.7.0->pycaret) (2.8.1)  
Requirement already satisfied: regex>=2017.02.08 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from datefinder>=0.7.0->pycaret) (2020.9.27)  
Requirement already satisfied: gitpython>=2.1.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (3.1.8)  
Requirement already satisfied: click>=7.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (7.0)  
Requirement already satisfied: sqlparse in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (0.3.1)  
Requirement already satisfied: querystring-parser in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (1.2.4)  
Requirement already satisfied: pyyaml in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (5.3)  
Requirement already satisfied: alembic<=1.4.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (1.4.1)  
Requirement already satisfied: entrypoints in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (0.3)  
Requirement already satisfied: gunicorn; platform\_system != "Windows" in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (20.0.4)  
Requirement already satisfied: gorilla in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (0.3.0)  
Requirement already satisfied: protobuf>=3.6.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (3.13.0)  
Requirement already satisfied: cloudpickle in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (1.3.0)  
Requirement already satisfied: Flask in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (1.1.1)  
Requirement already satisfied: prometheus-flask-exporter in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (0.18.0)  
Requirement already satisfied: docker>=4.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (4.3.1)  
Requirement already satisfied: sqlalchemy<=1.3.13 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (1.3.13)  
Requirement already satisfied: azure-storage-blob>=12.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (12.5.0)

Requirement already satisfied: databricks-cli>=0.8.7 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from mlflow->pycaret) (0.12.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from matplotlib->pycaret) (1.1.0)

Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from matplotlib->pycaret) (2.4.6)

Requirement already satisfied: retrying>=1.3.3 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from plotly>=4.4.1->pycaret) (1.3.3)

Requirement already satisfied: patsy>=0.5 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from statsmodels->pyod->pycaret) (0.5.1)

Requirement already satisfied: llvmlite<0.32.0,>=0.31.0dev0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from numba>=0.35->pyod->pycaret) (0.31.0)

Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy->pycaret) (1.25.8)

Requirement already satisfied: certifi>=2017.4.17 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy->pycaret) (2019.11.28)

Requirement already satisfied: chardet<4,>=3.0.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy->pycaret) (3.0.4)

Requirement already satisfied: idna<3,>=2.5 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from requests<3.0.0,>=2.13.0->spacy->pycaret) (2.8)

Requirement already satisfied: importlib-metadata>=0.20; python\_version < "3.8" in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from catalogue<1.1.0,>=0.0.7->spacy->pycaret) (1.5.0)

Requirement already satisfied: ipython-genutils in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from traitlets>=4.2->IPython->pycaret) (0.2.0)

Requirement already satisfied: parso>=0.5.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from jedi>=0.10->IPython->pycaret) (0.5.2)

Requirement already satisfied: wcwidth in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->IPython->pycaret) (0.1.8)

Requirement already satisfied: ptyprocess>=0.5 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pexpect; sys\_platform != "win32"->IPython->pycaret) (0.6.0)

Requirement already satisfied: MarkupSafe>=0.23 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from jinja2>=2.7.2->pyLDAvis->pycaret) (1.1.1)

Requirement already satisfied: py>=1.5.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pytest->pyLDAvis->pycaret) (1.8.1)

Requirement already satisfied: packaging in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pytest->pyLDAvis->pycaret) (20.1)

Requirement already satisfied: more-itertools>=4.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pytest->pyLDAvis->pycaret) (8.2.0)

Requirement already satisfied: pluggy<1.0,>=0.12 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from pytest->pyLDAvis->pycaret) (0.13.1)

Requirement already satisfied: jupyter-client in /Users/keirynhart/o

pt/anaconda3/lib/python3.7/site-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (5.3.4)  
Requirement already satisfied: tornado>=4.2 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (6.0.3)  
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (3.2.0)  
Requirement already satisfied: jupyter-core in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (4.6.1)  
Requirement already satisfied: notebook>=4.4.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from widgetsnbextension~=3.5.0->ipywidgets->pycaret) (6.0.3)  
Requirement already satisfied: networkx>=2.4 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from visions[type\_image\_path]==0.5.0->pandas-profiling>=2.8.0->pycaret) (2.4)  
Requirement already satisfied: imagehash; extra == "type\_image\_path" in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from visions[type\_image\_path]==0.5.0->pandas-profiling>=2.8.0->pycaret) (4.1.0)  
Requirement already satisfied: gitdb<5,>=4.0.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from gitpython>=2.1.0->mlflow->pycaret) (4.0.5)  
Requirement already satisfied: Mako in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from alembic<=1.4.1->mlflow->pycaret) (1.1.3)  
Requirement already satisfied: python-editor>=0.3 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from alembic<=1.4.1->mlflow->pycaret) (1.0.4)  
Requirement already satisfied: Werkzeug>=0.15 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from Flask->mlflow->pycaret) (1.0.0)  
Requirement already satisfied: itsdangerous>=0.24 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from Flask->mlflow->pycaret) (1.1.0)  
Requirement already satisfied: prometheus-client in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from prometheus-flask-exporter->mlflow->pycaret) (0.7.1)  
Requirement already satisfied: websocket-client>=0.32.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from docker>=4.0.0->mlflow->pycaret) (0.57.0)  
Requirement already satisfied: azure-core<2.0.0,>=1.6.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from azure-storage-blob>=12.0->mlflow->pycaret) (1.8.1)  
Requirement already satisfied: cryptography>=2.1.4 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from azure-storage-blob>=12.0->mlflow->pycaret) (2.8)  
Requirement already satisfied: msrest>=0.6.10 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from azure-storage-blob>=12.0->mlflow->pycaret) (0.6.19)  
Requirement already satisfied: tabulate>=0.7.7 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from databricks-cli>=0.8.7->mlflow->pycaret) (0.8.7)  
Requirement already satisfied: tenacity>=6.2.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from databricks-cli>=0.8.7->mlflow->pycaret) (6.2.0)  
Requirement already satisfied: zipp>=0.5 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from importlib-metadata>=0.20; python\_version < "3.8"->catalogue<1.1.0,>=0.0.7->spacy->pycaret) (2.2.0)



Requirement already satisfied: pyzmq>=13 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from jupyter-client->ipykernel>=4.5.1->ipywidgets->pycaret) (18.1.1)

Requirement already satisfied: pyparsing>=2.4.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets->pycaret) (2.4.7)

Requirement already satisfied: terminado>=0.8.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (0.8.3)

Requirement already satisfied: Send2Trash in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (1.5.0)

Requirement already satisfied: nbconvert in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (5.6.1)

Requirement already satisfied: PyWavelets in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from imagehash; extra == "type\_image\_path"->visions[type\_image\_path]==0.5.0->pandas-profiling>=2.8.0->pycaret) (1.1.1)

Requirement already satisfied: smmap<4,>=3.0.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from gitdb<5,>=4.0.1->gitpython>=2.1.0->mlflow->pycaret) (3.0.4)

Requirement already satisfied: cffi!=1.11.3,>=1.8 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from cryptography>=2.1.4->azure-storage-blob>=12.0->mlflow->pycaret) (1.14.0)

Requirement already satisfied: isodate>=0.6.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from msrest>=0.6.10->azure-storage-blob>=12.0->mlflow->pycaret) (0.6.0)

Requirement already satisfied: requests-oauthlib>=0.5.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from msrest>=0.6.10->azure-storage-blob>=12.0->mlflow->pycaret) (1.3.0)

Requirement already satisfied: pandocfilters>=1.4.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (1.4.2)

Requirement already satisfied: testpath in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (0.4.4)

Requirement already satisfied: bleach in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (3.1.0)

Requirement already satisfied: mistune<2,>=0.8.1 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (0.8.4)

Requirement already satisfied: defusedxml in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (0.6.0)

Requirement already satisfied: pycparser in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from cffi!=1.11.3,>=1.8->cryptography>=2.1.4->azure-storage-blob>=12.0->mlflow->pycaret) (2.19)

Requirement already satisfied: oauthlib>=3.0.0 in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from requests-oauthlib>=0.5.0->msrest>=0.6.10->azure-storage-blob>=12.0->mlflow->pycaret) (3.1.0)

Requirement already satisfied: webencodings in /Users/keirynhart/opt/anaconda3/lib/python3.7/site-packages (from bleach->nbconvert->notebook>=4.4.1->widgets-nbextension~>3.5.0->ipywidgets->pycaret) (0.5.1)

In [3]:

```
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from pycaret.classification import *
```

In [4]:

```
train = pd.read_csv("/Users/keirynhart/Documents/Uni/Data 301/assignment 5/A5_data/train.csv")
test = pd.read_csv("/Users/keirynhart/Documents/Uni/Data 301/assignment 5/A5_data/test.csv")
```

In [24]:

```
train_labels = train['label']
test_labels = test["label"]
train_full = train.copy()
train = train.drop(['label'], axis = 1)
test_full = test.copy()
test = test.drop(['label'], axis = 1)
```

In [ ]:

```
test.head()
```

In [5]:

models( )

Out[5]:

	Name	Reference	Turbo
ID			
<b>lr</b>	Logistic Regression	sklearn.linear_model.LogisticRegression	True
<b>knn</b>	K Neighbors Classifier	sklearn.neighbors.KNeighborsClassifier	True
<b>nb</b>	Naive Bayes	sklearn.naive_bayes.GaussianNB	True
<b>dt</b>	Decision Tree Classifier	sklearn.tree.DecisionTreeClassifier	True
<b>svm</b>	SVM - Linear Kernel	sklearn.linear_model.SGDClassifier	True
<b>rbfsvm</b>	SVM - Radial Kernel	sklearn.svm.SVC	False
<b>gpc</b>	Gaussian Process Classifier	sklearn.gaussian_process.GPC	False
<b>mlp</b>	MLP Classifier	sklearn.neural_network.MLPClassifier	False
<b>ridge</b>	Ridge Classifier	sklearn.linear_model.RidgeClassifier	True
<b>rf</b>	Random Forest Classifier	sklearn.ensemble.RandomForestClassifier	True
<b>qda</b>	Quadratic Discriminant Analysis	sklearn.discriminant_analysis.QDA	True
<b>ada</b>	Ada Boost Classifier	sklearn.ensemble.AdaBoostClassifier	True
<b>gbc</b>	Gradient Boosting Classifier	sklearn.ensemble.GradientBoostingClassifier	True
<b>lda</b>	Linear Discriminant Analysis	sklearn.discriminant_analysis.LDA	True
<b>et</b>	Extra Trees Classifier	sklearn.ensemble.ExtraTreesClassifier	True
<b>xgboost</b>	Extreme Gradient Boosting	xgboost.readthedocs.io	True
<b>lightgbm</b>	Light Gradient Boosting Machine	github.com/microsoft/LightGBM	True
<b>catboost</b>	CatBoost Classifier	catboost.ai	True

In [7]:

```
clf = setup(data = train, target = 'label')
```

Setup Succesfully Completed!

	Description	Value
0	session_id	3241
1	Target Type	Binary
2	Label Encoded	negative: 0, positive: 1
3	Original Data	(6744, 1025)
4	Missing Values	False
5	Numeric Features	0
6	Categorical Features	1024
7	Ordinal Features	False
8	High Cardinality Features	False
9	High Cardinality Method	None
10	Sampled Data	(6744, 1025)
11	Transformed Train Set	(4720, 2048)
12	Transformed Test Set	(2024, 2048)
13	Numeric Imputer	mean
14	Categorical Imputer	constant
15	Normalize	False
16	Normalize Method	None
17	Transformation	False
18	Transformation Method	None
19	PCA	False
20	PCA Method	None
21	PCA Components	None
22	Ignore Low Variance	False
23	Combine Rare Levels	False
24	Rare Level Threshold	None
25	Numeric Binning	False
26	Remove Outliers	False
27	Outliers Threshold	None
28	Remove Multicollinearity	False
29	Multicollinearity Threshold	None
30	Clustering	False
31	Clustering Iteration	None
32	Polynomial Features	False
33	Polynomial Degree	None
34	Trigonometry Features	False
35	Polynomial Threshold	None
36	Group Features	False

	Description	Value
37	Feature Selection	False
38	Features Selection Threshold	None
39	Feature Interaction	False
40	Feature Ratio	False
41	Interaction Threshold	None
42	Fix Imbalance	False
43	Fix Imbalance Method	SMOTE

In [12]:

```
trials = compare_models(sort = 'AUC')
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
0	Extra Trees Classifier	0.9352	0.8698	0.3856	0.6940	0.4915	0.4603	0.4850	0.5902
1	CatBoost Classifier	0.9345	0.8674	0.3111	0.7487	0.4368	0.4086	0.4546	30.1576
2	Light Gradient Boosting Machine	0.9352	0.8604	0.3958	0.6878	0.4995	0.4677	0.4896	1.1767
3	Extreme Gradient Boosting	0.9352	0.8532	0.3906	0.6910	0.4946	0.4632	0.4869	34.3761
4	Gradient Boosting Classifier	0.9331	0.8455	0.3084	0.7261	0.4293	0.4000	0.4433	10.0004
5	Logistic Regression	0.9150	0.8218	0.3883	0.4862	0.4305	0.3853	0.3888	0.5021
6	K Neighbors Classifier	0.9314	0.8170	0.3545	0.6604	0.4561	0.4235	0.4492	0.9005
7	Random Forest Classifier	0.9309	0.8116	0.2928	0.7141	0.4104	0.3804	0.4256	0.1462
8	Linear Discriminant Analysis	0.9138	0.8049	0.4756	0.4776	0.4746	0.4279	0.4289	3.2726
9	Ada Boost Classifier	0.9271	0.8011	0.3012	0.6239	0.4015	0.3682	0.3980	3.2257
10	Naive Bayes	0.7125	0.7362	0.7017	0.1810	0.2876	0.1801	0.2451	0.1761
11	Decision Tree Classifier	0.8956	0.6785	0.4190	0.3820	0.3986	0.3417	0.3427	0.8523
12	Quadratic Discriminant Analysis	0.9176	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	2.5969
13	SVM - Linear Kernel	0.8890	0.0000	0.4526	0.4354	0.4049	0.3505	0.3690	1.2820
14	Ridge Classifier	0.9275	0.0000	0.3497	0.6119	0.4404	0.4051	0.4254	0.4044

After using the training set to evaluate a variety of different machine learning models we can see that based on AUC the top 3 sklearn classifiers are Extra Trees Classifier, Gradient boosting classifier, logistic regression, K Neighbours Classifier and Random forest classifier. because of this these are the 5 classifiers that I am going to continue with.

In [66]:

```
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, classification_report
from sklearn.model_selection import cross_val_predict
from sklearn.model_selection import cross_val_score
```

In [18]:

```
xt = ExtraTreesClassifier()
grad = GradientBoostingClassifier()
log = LogisticRegression()
knn = KNeighborsClassifier(n_neighbors=3)
rf = RandomForestClassifier()
classifier_list = [xt, grad, log, knn, rf]
```

In [60]:

```
from sklearn.decomposition import PCA

pca = PCA(n_components= 20)
train_pca = pca.fit(train).transform(train)

test_pca = pca.transform(test)

def classifiers(train, train_labels, test, test_labels):
    for i in classifier_list:

        i.fit(train_pca, train_labels)

        pred = i.predict(test_pca)
        pred1 = cross_val_predict(i, train, train_labels, cv = 10)

        print(accuracy_score(pred1, train_labels))
        print(classification_report(test_labels, pred))
```



In [61]:

```
classifiers(train, train_labels, test, test_labels)
```

0.9430604982206405

	precision	recall	f1-score	support
negative	0.94	0.99	0.96	2063
positive	0.76	0.26	0.39	185
accuracy			0.93	2248
macro avg	0.85	0.63	0.68	2248
weighted avg	0.92	0.93	0.92	2248

0.9337188612099644

	precision	recall	f1-score	support
negative	0.93	0.99	0.96	2063
positive	0.69	0.20	0.31	185
accuracy			0.93	2248
macro avg	0.81	0.60	0.64	2248
weighted avg	0.91	0.93	0.91	2248

0.9233392645314353

	precision	recall	f1-score	support
negative	0.93	1.00	0.96	2063
positive	0.73	0.10	0.18	185
accuracy			0.92	2248
macro avg	0.83	0.55	0.57	2248
weighted avg	0.91	0.92	0.90	2248

0.930011862396204

	precision	recall	f1-score	support
negative	0.95	0.98	0.97	2063
positive	0.65	0.46	0.54	185
accuracy			0.94	2248
macro avg	0.80	0.72	0.75	2248
weighted avg	0.93	0.94	0.93	2248

0.941429418742586

	precision	recall	f1-score	support
negative	0.94	0.99	0.96	2063
positive	0.75	0.25	0.38	185
accuracy			0.93	2248
macro avg	0.84	0.62	0.67	2248
weighted avg	0.92	0.93	0.92	2248

In [73]:

```
from sklearn.ensemble import StackingClassifier
from sklearn.model_selection import RepeatedStratifiedKFold
```

In [71]:

```
models = [
    ('xt', ExtraTreesClassifier()),
    ('grad', GradientBoostingClassifier()),
    ('log', LogisticRegression()),
    ('knn', KNeighborsClassifier(n_neighbors=3)),
    ('rf', RandomForestClassifier())
]
```

In [74]:

```
cv = RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=1)
```

In [76]:

```
for clf in models:
    scores = cross_val_score(clf[1], train, train_labels, scoring='accuracy', cv
=cv, n_jobs=-1)
    print('%s: %.4f (%.4f)' % (clf[0], np.mean(scores), np.std(scores)))
```

```
xt: 0.9427 (0.0048)
grad: 0.9346 (0.0058)
log: 0.9234 (0.0085)
knn: 0.9290 (0.0071)
rf: 0.9408 (0.0050)
```

In [85]:

```
def get_stacking():
    # define the base models
    level0 = list()
    level0.append(('xt', ExtraTreesClassifier()))
    level0.append(('grad', GradientBoostingClassifier()))
    level0.append(('log', LogisticRegression()))
    level0.append(('knn', KNeighborsClassifier(n_neighbors=3)))
    level0.append(('rf', RandomForestClassifier()))
    # define meta learner model
    level1 = ExtraTreesClassifier()
    # define the stacking ensemble
    model = StackingClassifier(estimators=level0, final_estimator=level1, cv=5)
    return model
```

In [ ]:

```
model = get_stacking()
print(cross_val_score(xt, test, test_labels, scoring="roc_auc", cv = 10).max())
```

In [ ]:

```
clf = StackingClassifier(estimators=models,
                        final_estimator=ExtraTreesClassifier())
scores = cross_val_score(model, train, train_labels, scoring='accuracy', cv=cv,
n_jobs=-1)
print('%s: %.4f (%.4f)' % (name, np.mean(scores), np.std(scores)))
```

In [ ]:

In [ ]: