Group Members 1. 6310422053 เมธี ประเสริฐกิจพันธุ์ 2. 6310422031 ธนัตถ์กรณ์ ชื่นบรรลือสุข 3. 6310422042 พีรณัฐ เชื้อสวย **Load Dependencies** In []: !pip install pycaret Collecting pycaret Using cached pycaret-2.3.4-py3-none-any.whl (266 kB) Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.0.1) Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.5.0) Collecting yellowbrick>=1.0.1 Downloading yellowbrick-1.3.post1-py3-none-any.whl (271 kB) | 271 kB 6.7 MB/s 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 (from pycaret) (4.4.1) Collecting mlxtend>=0.17.0 Using cached mlxtend-0.19.0-py2.py3-none-any.whl (1.3 MB) Requirement already satisfied: scipy<=1.5.4 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.4.1) Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages (from pycaret) (7.6.5) Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.6.0) Collecting Boruta Downloading Boruta-0.3-py3-none-any.whl (56 kB) | 56 kB 4.5 MB/s Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.15.3) Collecting pyLDAvis Using cached pyLDAvis-3.3.1.tar.gz (1.7 MB) Installing build dependencies ... done Getting requirements to build wheel ... done Installing backend dependencies ... done Preparing wheel metadata ... done Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.11.2) Collecting scikit-learn==0.23.2 Downloading scikit_learn-0.23.2-cp37-cp37m-manylinux1_x86_64.whl (6.8 MB) 6.8 MB 33.8 MB/s Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.2.5) Collecting pyod Downloading pyod-0.9.3.tar.gz (111 kB) Collecting umap-learn Downloading umap-learn-0.5.1.tar.gz (80 kB) | 80 kB 10.1 MB/s Requirement already satisfied: numba<0.54 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.51.2) Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from pycaret) (3.2.2) Collecting mlflow Downloading mlflow-1.20.2-py3-none-any.whl (14.6 MB) | 14.6 MB 125 kB/s Requirement already satisfied: spacy<2.4.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (2.2.4) Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.5) Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.5.0) Collecting imbalanced-learn==0.7.0 Using cached imbalanced_learn-0.7.0-py3-none-any.whl (167 kB) Requirement already satisfied: numpy==1.19.5 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.19.5) Collecting lightgbm>=2.3.1 Using cached lightgbm-3.2.1-py3-none-manylinux1_x86_64.whl (2.0 MB) Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.17.3) Collecting pandas-profiling>=2.8.0 Using cached pandas_profiling-3.0.0-py2.py3-none-any.whl (248 kB) Collecting kmodes>=0.10.1 Downloading kmodes-0.11.0-py2.py3-none-any.whl (18 kB) Collecting threadpoolct1>=2.0.0 Downloading threadpoolctl-2.2.0-py3-none-any.whl (12 kB) Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycar Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pycaret) (1. Requirement already satisfied: setuptools>=34.4.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks>=0.17.0->pyca Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-packages (from gensim<4.0.0->pycaret) Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (0.7.5) Requirement already satisfied: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (5.1.0) Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from IPython->pyca ret) (1.0.18) Requirement already satisfied: pexpect in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (4.8.0) 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: 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->pycare Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (4.1 Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pyca Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pyca ret) (3.5.1) Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (5.1. 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->ipywidget s->pycaret) (5.3.5) Requirement already satisfied: wheel in /usr/local/lib/python3.7/dist-packages (from lightgbm>=2.3.1->pycaret) (0.37.0) Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (1. 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 m atplotlib->pycaret) (2.4.7) Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (0.10.0) Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->i pywidgets->pycaret) (2.6.0) Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets-> pycaret) (4.7.1) Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/dist-packages (from numba<0.54->py caret) (0.34.0) Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas->pycaret) (2018.9) Collecting visions[type_image_path] == 0.7.1 Downloading visions-0.7.1-py3-none-any.whl (102 kB) | 102 kB 58.8 MB/s Collecting PyYAML>=5.0.0 Downloading PyYAML-5.4.1-cp37-cp37m-manylinux1 x86 64.whl (636 kB) | 636 kB 42.3 MB/s 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) Collecting requests>=2.24.0 Downloading requests-2.26.0-py2.py3-none-any.whl (62 kB) Collecting phik>=0.11.1 Downloading phik-0.12.0-cp37-cp37m-manylinux2010_x86_64.whl (675 kB) | 675 kB 59.2 MB/s Collecting pydantic>=1.8.1 Downloading pydantic-1.8.2-cp37-cp37m-manylinux2014 x86 64.whl (10.1 MB) | 10.1 MB 19.0 MB/s Requirement already satisfied: tqdm>=4.48.2 in /usr/local/lib/python3.7/dist-packages (from pandas-profiling>=2.8.0->pyca ret) (4.62.2)Requirement already satisfied: jinja2>=2.11.1 in /usr/local/lib/python3.7/dist-packages (from pandas-profiling>=2.8.0->py caret) (2.11.3) Collecting tangled-up-in-unicode==0.1.0 Downloading tangled_up_in_unicode-0.1.0-py3-none-any.whl (3.1 MB) Collecting htmlmin>=0.1.12 Downloading htmlmin-0.1.12.tar.gz (19 kB) Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.7/dist-packages (from visions[type_image_path] ==0. 7.1->pandas-profiling>=2.8.0->pycaret) (2.6.3) Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.7/dist-packages (from visions[type image path] ==0. 7.1->pandas-profiling>=2.8.0->pycaret) (21.2.0) Requirement already satisfied: bottleneck in /usr/local/lib/python3.7/dist-packages (from visions[type_image_path] == 0.7.1 ->pandas-profiling>=2.8.0->pycaret) (1.3.2) Collecting multimethod==1.4 Downloading multimethod-1.4-py2.py3-none-any.whl (7.3 kB) Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from visions[type image path] == 0.7.1->pa ndas-profiling>=2.8.0->pycaret) (7.1.2) Collecting imagehash Downloading ImageHash-4.2.1.tar.gz (812 kB) | 812 kB 50.4 MB/s Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from jinja2>=2.11.1->pandas-pr ofiling>=2.8.0->pycaret) (2.0.1) Collecting scipy<=1.5.4 Downloading scipy-1.5.4-cp37-cp37m-manylinux1_x86_64.whl (25.9 MB) | 25.9 MB 1.0 kB/s Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from plotly>=4.4.1->pycaret) Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.0.0,>=1.0.4->IPyt hon->pycaret) (0.2.5) Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from pydantic>=1.8.1 ->pandas-profiling>=2.8.0->pycaret) (3.7.4.3) 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) (2.0.6) Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests>=2.24.0->panda s-profiling>=2.8.0->pycaret) (2021.5.30) Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests>=2.24.0->pa ndas-profiling>=2.8.0->pycaret) (1.24.3) Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests>=2.24.0->pandas-prof iling>=2.8.0->pycaret) (2.10) Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (7.4.0) Requirement already satisfied: murmurhash < 1.1.0, >=0.28.0 in /usr/local/lib/python3.7/dist-packages (from spacy < 2.4.0 -> pyc) aret) (1.0.5) Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycare Requirement already satisfied: blis<0.5.0, >=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) $\label{local_lib_python3.7} Requirement already satisfied: catalogue < 1.1.0, >= 0.0.7 in /usr/local/lib/python3.7/dist-packages (from spacy < 2.4.0-> pycar | pycar$ Requirement already satisfied: importlib-metadata>=0.20 in /usr/local/lib/python3.7/dist-packages (from catalogue<1.1.0,> =0.0.7->spacy<2.4.0->pycaret) (4.8.1) Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalo gue<1.1.0,>=0.0.7->spacy<2.4.0->pycaret) (3.5.0) 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) $\label{lem:reduced_reduced_reduced} Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnownedge$ bextension~=3.5.0->ipywidgets->pycaret) (0.12.1) Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbexten sion~=3.5.0->ipywidgets->pycaret) (1.8.0) $\label{local_lib_python3.7} Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextens) and the converting of the converting of$ ion~=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) 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_p ath] == 0.7.1 - pandas - profiling >= 2.8.0 - pycaret) (1.1.1)Collecting alembic <= 1.4.1 Downloading alembic-1.4.1.tar.gz (1.1 MB) | 1.1 MB 69.5 MB/s Collecting docker>=4.0.0 Downloading docker-5.0.2-py2.py3-none-any.whl (145 kB) | 145 kB 55.5 MB/s Collecting databricks-cli>=0.8.7 Downloading databricks-cli-0.15.0.tar.gz (56 kB) | 56 kB 5.2 MB/s Requirement already satisfied: sqlparse>=0.3.1 in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (0.4.2) Collecting prometheus-flask-exporter Downloading prometheus flask exporter-0.18.2.tar.gz (22 kB) Requirement already satisfied: Flask in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (1.1.4) Requirement already satisfied: entrypoints in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (0.3) Collecting querystring-parser Downloading querystring_parser-1.2.4-py2.py3-none-any.whl (7.9 kB) Requirement already satisfied: protobuf>=3.7.0 in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (3.17.3) Collecting gunicorn Downloading gunicorn-20.1.0-py3-none-any.whl (79 kB) | 79 kB 8.3 MB/s Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (7.1.2) Requirement already satisfied: cloudpickle in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (1.3.0) Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (1.4.23) Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (21.0) Collecting gitpython>=2.1.0 Downloading GitPython-3.1.24-py3-none-any.whl (180 kB) | 180 kB 59.1 MB/s Collecting Mako Downloading Mako-1.1.5-py2.py3-none-any.whl (75 kB) Collecting python-editor>=0.3 Downloading python editor-1.0.4-py3-none-any.whl (4.9 kB) Requirement already satisfied: tabulate>=0.7.7 in /usr/local/lib/python3.7/dist-packages (from databricks-cli>=0.8.7->mlf low->pycaret) (0.8.9) Collecting websocket-client>=0.32.0 Downloading websocket_client-1.2.1-py2.py3-none-any.whl (52 kB) Collecting gitdb<5,>=4.0.1 Downloading gitdb-4.0.7-py3-none-any.whl (63 kB) | 63 kB 1.9 MB/s Collecting smmap<5,>=3.0.1 Downloading smmap-4.0.0-py2.py3-none-any.whl (24 kB) Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.7/dist-packages (from sqlalchemy->mlflow->pycar et) (1.1.1)Requirement already satisfied: itsdangerous<2.0,>=0.24 in /usr/local/lib/python3.7/dist-packages (from Flask->mlflow->pyc aret) (1.1.0)Requirement already satisfied: Werkzeug<2.0,>=0.15 in /usr/local/lib/python3.7/dist-packages (from Flask->mlflow->pycare Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>= 4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (1.5.0) Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-packages (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->wid getsnbextension~=3.5.0->ipywidgets->pycaret) (0.7.1) Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widge tsnbextension~=3.5.0->ipywidgets->pycaret) (0.5.0) Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbconvert->notebook>=4.4.1->widgets nbextension~=3.5.0->ipywidgets->pycaret) (4.1.0) Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from 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-packages (from prometheus-flask-exporte r->mlflow->pycaret) (0.11.0) Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from pyLDAvis->pycaret) (0.16.0) Downloading pandas-1.3.3-cp37-cp37m-manylinux 2 17 x86 64.manylinux2014 x86 64.whl (11.3 MB) | 11.3 MB 34.8 MB/s Requirement already satisfied: numexpr in /usr/local/lib/python3.7/dist-packages (from pyLDAvis->pycaret) (2.7.3) Requirement already satisfied: sklearn in /usr/local/lib/python3.7/dist-packages (from pyLDAvis->pycaret) (0.0) Collecting funcy Downloading funcy-1.16-py2.py3-none-any.whl (32 kB) Collecting pvLDAvis Downloading pyLDAvis-3.3.0.tar.gz (1.7 MB) 1.7 MB 57.1 MB/s Installing build dependencies ... done Getting requirements to build wheel ... done Installing backend dependencies ... done Preparing wheel metadata \dots done WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'Pro tocolError('Connection aborted.', ConnectionResetError(104, 'Connection reset by peer'))': /packages/24/38/6d81eff34c84c9 158d3b7c846bff978ac88b0c2665548941946d3d591158/pyLDAvis-3.2.2.tar.gz Downloading pyLDAvis-3.2.2.tar.gz (1.7 MB) | 1.7 MB 6.8 MB/s Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (from pyod->pycaret) (0.10.2) Requirement already satisfied: patsy>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from statsmodels->pyod->pycaret) (0.5.1)Collecting pynndescent>=0.5 Downloading pynndescent-0.5.4.tar.gz (1.1 MB) | 1.1 MB 14.7 MB/s Building wheels for collected packages: htmlmin, imagehash, alembic, databricks-cli, prometheus-flask-exporter, pyLDAvis, pyod, 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=422e28190280983f9429eb908f3118b82 420b087609870f3484048aa1105f4b0 Stored in directory: /root/.cache/pip/wheels/70/e1/52/5b14d250ba868768823940c3229e9950d201a26d0bd3ee8655 Building wheel for imagehash (setup.py) ... done Created wheel for imagehash: filename=ImageHash-4.2.1-py2.py3-none-any.whl size=295207 sha256=ac9dbe564eaf831e8c4578813 8ee1ce46be33c4e49ad7e7daa9d6ad44fb872ff Stored in directory: /root/.cache/pip/wheels/4c/d5/59/5e3e297533ddb09407769762985d134135064c6831e29a914e Building wheel for alembic (setup.py) ... done Created wheel for alembic: filename=alembic-1.4.1-py2.py3-none-any.whl size=158172 sha256=94d30224408fbc1cc64f79ff0860f 52be48031a85f17187a7bc1ea3c1106862a Stored in directory: /root/.cache/pip/wheels/be/5d/0a/9e13f53f4f5dfb67cd8d245bb7cdffe12f135846f491a283e3 Building wheel for databricks-cli (setup.py) ... done Created wheel for databricks-cli: filename=databricks_cli-0.15.0-py3-none-any.whl size=105260 sha256=e03eb3301f7f992d2d 20bac3467cb77d7954b49e71055a660ed77bf9c9924fce Stored in directory: /root/.cache/pip/wheels/e7/ba/75/284f9a90ff7a010bb23b9798f2e9a19dd9fe619379c917bff4 Building wheel for prometheus-flask-exporter (setup.py) ... done Created wheel for prometheus-flask-exporter: filename=prometheus flask exporter-0.18.2-py3-none-any.whl size=17415 sha2 56=5505b948ae6f727badeba411135cedb787a804cc0747a469efbc5d3a3029f34e Stored in directory: /root/.cache/pip/wheels/6a/1e/1c/c765920cb92b2f0343d2dd8b481a407cee2823f9b4bbd2e52a Building wheel for pyLDAvis (setup.py) ... done Created wheel for pyLDAvis: filename=pyLDAvis-3.2.2-py2.py3-none-any.whl size=135618 sha256=a38ccf61f663962b824b69bb9a8 dafa7ea5b76d5d4670193a14fd83af5a0a5a7 Stored in directory: /root/.cache/pip/wheels/f8/b1/9b/560ac1931796b7303f7b517b949d2d31a4fbc512aad3b9f284 Building wheel for pyod (setup.py) ... done Created wheel for pyod: filename=pyod-0.9.3-py3-none-any.whl size=130719 sha256=0b6aac3cd3d21aec1f7ef74b48bcb27321a4f26 b495ac700c252d8a8d766f8e1 Stored in directory: /root/.cache/pip/wheels/ac/94/65/d29b87d4ad6defcec32686063e9bbaf5907a259eb2d8342174 Building wheel for umap-learn (setup.py) ... done Created wheel for umap-learn: filename=umap_learn-0.5.1-py3-none-any.whl size=76564 sha256=556eaf2865f9f0e399ab0e1560d1 b76a5e05a04b6c4225b15f7fd35923cd16d6 Stored in directory: /root/.cache/pip/wheels/01/e7/bb/347dc0e510803d7116a13d592b10cc68262da56a8eec4dd72f Building wheel for pynndescent (setup.py) ... done Created wheel for pynndescent: filename=pynndescent-0.5.4-py3-none-any.whl size=52373 sha256=1b87c1fe338df22c5826569b79 464d884d165a545d3983a83fcbc3d2d6686738 Stored in directory: /root/.cache/pip/wheels/d0/5b/62/3401692ddad12324249c774c4b15ccb046946021e2b581c043 Successfully built htmlmin imagehash alembic databricks-cli prometheus-flask-exporter pyLDAvis pyod umap-learn pynndescen Installing collected packages: threadpoolctl, tangled-up-in-unicode, smmap, scipy, multimethod, websocket-client, vision s, scikit-learn, requests, python-editor, Mako, imagehash, gitdb, querystring-parser, PyYAML, pynndescent, pydantic, prom etheus-flask-exporter, phik, htmlmin, gunicorn, gitpython, funcy, docker, databricks-cli, alembic, yellowbrick, umap-lear n, scikit-plot, pyod, pyLDAvis, pandas-profiling, mlxtend, mlflow, lightgbm, kmodes, imbalanced-learn, Boruta, pycaret Attempting uninstall: scipy Found existing installation: scipy 1.4.1 Uninstalling scipy-1.4.1: Successfully uninstalled scipy-1.4.1 Attempting uninstall: scikit-learn Found existing installation: scikit-learn 0.22.2.post1 Uninstalling scikit-learn-0.22.2.post1: Successfully uninstalled scikit-learn-0.22.2.post1 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: yellowbrick Found existing installation: yellowbrick 0.9.1 Uninstalling yellowbrick-0.9.1: Successfully uninstalled yellowbrick-0.9.1 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 0.4.3 Uninstalling imbalanced-learn-0.4.3: Successfully uninstalled imbalanced-learn-0.4.3 ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts. google-colab 1.0.0 requires requests~=2.23.0, but you have requests 2.26.0 which is incompatible. datascience 0.10.6 requires folium==0.2.1, but you have folium 0.8.3 which is incompatible. albumentations 0.1.12 requires imgaug<0.2.7,>=0.2.5, but you have imgaug 0.2.9 which is incompatible. Successfully installed Boruta-0.3 Mako-1.1.5 PyYAML-5.4.1 alembic-1.4.1 databricks-cli-0.15.0 docker-5.0.2 funcy-1.16 git db-4.0.7 gitpython-3.1.24 gunicorn-20.1.0 htmlmin-0.1.12 imagehash-4.2.1 imbalanced-learn-0.7.0 kmodes-0.11.0 lightgbm-3. 2.1 mlflow-1.20.2 mlxtend-0.19.0 multimethod-1.4 pandas-profiling-3.0.0 phik-0.12.0 prometheus-flask-exporter-0.18.2 pyLD Avis-3.2.2 pycaret-2.3.4 pydantic-1.8.2 pynndescent-0.5.4 pyod-0.9.3 python-editor-1.0.4 querystring-parser-1.2.4 request s-2.26.0 scikit-learn-0.23.2 scikit-plot-0.3.7 scipy-1.5.4 smmap-4.0.0 tangled-up-in-unicode-0.1.0 threadpoolctl-2.2.0 um ap-learn-0.5.1 visions-0.7.1 websocket-client-1.2.1 yellowbrick-1.3.post1 In []: import numpy as np import pandas as pd from pycaret.clustering import * **Load Data** In []: from google.colab import drive drive.mount('/content/drive') Mounted at /content/drive df = pd.read csv('/content/drive/MyDrive/Colab Notebooks/7105/Code Templates/Supermarket Data.csv') df['SHOP DATE'] = df['SHOP DATE'].apply(lambda x: pd.to datetime(str(x), format='%Y%m%d')) In []: df.tail(5) SHOP_WEEK SHOP_DATE SHOP_WEEKDAY SHOP_HOUR QUANTITY SPEND PROD_CODE PROD_CODE_10 956569 200617 2006-06-22 12 3 3.96 PRD0904997 CL00074 **956570** 200633 3.96 PRD0904997 CL00074 2006-10-13 **956571** 200617 CL00074 2006-06-22 5 18 3.96 PRD0904997 **956572** 200619 2006-07-06 3.96 PRD0904997 CL00074 956573 200635 CL00074 2006-10-23 3.96 PRD0904997 In []: for column in df.columns: print(f"{column}: {len(pd.unique(df[column]))}") SHOP_WEEK SHOP_DATE SHOP_WEEKDAY SHOP_HOUR QUANTITY SPEND \ <bound method DataFrame.info of</pre> 0 200732 2007-10-05 6 17 3 6.75 200733 2007-10-10 200741 2007-12-09 1 2.25 17 200731 2007-09-29 14 200737 2007-11-10 12 956569 200617 2006-06-22 5 3 3.96 20 956570 200633 2006-10-13 6 3 3.96 956571 200617 2006-06-22 3 3.96 956572 200619 2006-07-06 956573 200635 2006-10-23 3 3.96 PROD_CODE PROD_CODE_10 PROD_CODE_20 PROD_CODE_30 PROD_CODE_40 PRD0900001 CL00072 DEP00021 G00007 D00002
 O
 PRD0900001
 CL00072
 DEP00021
 G00007
 D00002

 1
 PRD0900001
 CL00072
 DEP00021
 G00007
 D00002

 2
 PRD0900001
 CL00072
 DEP00021
 G00007
 D00002

 3
 PRD0900001
 CL00072
 DEP00021
 G00007
 D00002

 4
 PRD0904997
 CL00074
 DEP00021
 G00007
 D00002

 956570
 PRD0904997
 CL00074
 DEP00021
 G00007
 D00002

 956571
 PRD0904997
 CL00074
 DEP00021
 G00007
 D00002

 956572
 PRD0904997
 CL00074
 DEP00021
 G00007
 D00002

 956573
 PRD0904997
 CL00074
 DEP00021
 G00007
 D00002
956573 PRD0904997 CL00074 DEP00021 G00007 D00002 CUST CODE CUST PRICE SENSITIVITY CUST LIFESTAGE BASKET ID \ CUST0000583261 YF 994107800547472 0 UM OF 994107900512001 CUST0000537317 MM YF 994108700468327 CUST0000472158 MM CUST0000099658 LA OF 994107700237811 4 NaN NaN 994108300002212 NaN NaN 994101100088778 NaN NaN 994102700099738 956571 CUST0000544241 LA YA 994101100506174 LA YF 994101300433650 956572 CUST0000423155 NaN NaN 994102900104676 BASKET SIZE BASKET PRICE SENSITIVITY BASKET TYPE \ MM Top Up 0 MM Full Shop MM Full Shop LA Full Shop MM Full Shop MM Top Up LA 956570 L Top Up LA Top Up 956571 956572 LA Full Shop MM Top Up BASKET_DOMINANT_MISSION STORE_CODE STORE_FORMAT STORE_REGION Grocery STORE00001 LS E02 0 Fresh STORE00001 LS LS Grocery STORE00001 E02 LS Mixed STORE00001 Fresh STORE00001 LS . . . LS Fresh STORE00002 W01 956569 LS Fresh STORE00002 956570 W01 Fresh STORE00002 LS 956571 956572 Fresh STORE00002 LS 956573 Fresh STORE00002 LS [956574 rows x 22 columns]> Prepare customer single view **Define features** Total visits = COUNT(DISTINCT BASKET ID) Ticket size = SUM(SPEND)/COUNT(DISTINCT BASKET ID) Total no. of SKUs Calculate features df['SHOP DATE MONTH'] = df['SHOP DATE'].apply(lambda x: f"{x.year} {x.month}") df['SHOP DATE MONTH'] 2007 10 2007 10 2007 12 2007 9 2007 11 956569 2006 6 956570 2006 10 956571 2006 6 956572 2006 7 956573 2006 10 Name: SHOP_DATE_MONTH, Length: 956574, dtype: object df cm = df[df['CUST CODE'].notnull()].groupby(by=['CUST CODE', 'SHOP DATE MONTH']).agg(TotalVisits=('BAS KET ID', 'nunique'), TotalSpend=('SPEND', 'sum')).reset index() In []: df vpm = df cm.groupby(by=['CUST CODE']).agg(AvgVisitsPerMonth=('TotalVisits', np.mean),ARPU=('TotalSpen d', np.mean)).reset index() df bdm = df[df['CUST CODE'].notnull()].groupby(by=['CUST CODE', 'BASKET ID']).agg(BasketDominantMode=("BASKET DOMINANT MISSION", lambda x: x.value counts().index[0])).reset index() df bdm = df bdm.groupby(by=['CUST CODE']).agg(BasketDominantMode=("BasketDominantMode", lambda x: x.valu e counts().index[0])).reset index() In []: ##prepare customer single view df csv = df groupby = df[df['CUST CODE'].notnull()].groupby(by=['CUST CODE']).agg(TotalSpend=('SPEND', 'sum'), TotalQuantity=('QUANTI TY', 'sum'), TotalVisits=('BASKET I D', 'nunique'), TotalSKUs=('PROD CODE' , 'nunique'), MedShopHour=('SHOP HOU R', 'median'), FirstDate=('SHOP DATE' , 'min'), LastDate=('SHOP DATE', 'max'),).reset index() In []: df_csv = pd.merge(df_csv, df_vpm, how='left', on='CUST_CODE') df csv = pd.merge(df csv, df bdm, how='left', on='CUST CODE') In []: df csv weekday = df groupby = df[(df['CUST CODE'].notnull())& (df["SHOP WEEKDAY"].isin([1,7]))].groupby(by=['CUST_CODE']).agg(CountShopWeekDay=('SHOP WEEK', 'count')).reset index().sort values(by=['CUST CODE', 'CountShopWeekDay'], ascending =[True,False]) df csv weekday total = df groupby = df[(df['CUST CODE'].notnull())].groupby(by=['CUST CODE']).agg(CountShopWeekDayTotal=('SHOP WEEK', 'count')).reset index().sort values(by=['CUST CODE', 'CountShopWeekDayTotal'], asce nding=[True, False]) df csv result = pd.merge(df csv weekday, df csv weekday total, how="left", on=["CUST CODE"]) df csv result["percentage weekday"] = df csv result["CountShopWeekDay"]/ df csv result["CountShopWeekD ayTotal"] In []: df csv = pd.merge(df csv, df csv result, how='left', on='CUST CODE') In []: df csv CUST CODE TotalSpend TotalQuantity TotalVisits TotalSKUs MedShopHour FirstDate LastDate AvgVisitsPerI 2007-01-2007-01-CUST0000000181 2.44 1.000000 2006-04-2008-07-CUST0000000369 959.33 220 189 16.0 7.857143 923 05 2007-07-2008-06-2 CUST0000000689 328.57 334 16 116 19.0 1.777778 22 23 2006-05-2006-07-CUST0000000998 5.95 1.500000 11.5 2008-06-2006-10-CUST0000001163 39.74 24 15.0 1.000000 2008-04-2006-05-1.578947 6095 CUST0000999593 453.58 206 395 15.0 2006-04-2008-05-6096 CUST0000999645 105.11 1.833333 80 2006-06-2006-06-6097 CUST0000999654 5.86 14.0 1.000000 25 2006-11-2008-06-6098 CUST0000999935 444.17 1.615385 16.0 2006-08-2008-04-**6099** CUST0000999936 17.67 21 6 13 12.0 1.200000 6100 rows × 14 columns In []: ##calculate ticket size df csv['TicketSize'] = df csv['TotalSpend']/df csv['TotalVisits'] In []: ##find max date in the dataset max date = df csv['LastDate'].max() In []: ##calculate total days of the relationship df csv['total days'] = (df csv['LastDate'] - df csv['FirstDate']).dt.days + 1 In []: ##calculate recency days df_csv['recency'] = (max_date - df_csv['LastDate']).dt.days In []: df csv['BasketSize'] = df csv['TotalQuantity']/df csv['TotalVisits'] In []: df csv = pd.get dummies(df csv, columns=["BasketDominantMode"]) df csv.head() CUST_CODE TotalSpend TotalQuantity TotalVisits TotalSKUs MedShopHour FirstDate LastDate **AvgVisitsPerMor** 2007-01-2007-01-1.000000 CUST0000000181 13.0 06 06 2006-04-2008-07-7.857143 CUST0000000369 220 189 16.0 2007-07-2008-06-2 CUST0000000689 328.57 334 16 116 1.777778 19.0 2006-05-2006-07-CUST0000000998 1.500000 11.5 07 2006-10-2008-06-4 CUST0000001163 39.74 24 1.000000 15.0 Cluster customers exp_clu = setup(data=df_csv, ignore_features=['CUST_CODE','FirstDate','LastDate','CountShopWeekDay','Cou ntShopWeekDayTotal'], normalize=True) **Description Value** session_id 2905 0 **Original Data** (6100, 22)Missing Values True **Numeric Features** 17 0 Categorical Features Ordinal Features False 5 6 High Cardinality Features False High Cardinality Method None Transformed Data 8 (6100, 17)**CPU Jobs** 9 -1 Use GPU False 10 Log Experiment False cluster-default-name Experiment Name 13 USI a657 Imputation Type 14 simple Iterative Imputation Iteration None Numeric Imputer 16 mean Iterative Imputation Numeric Model 17 None Categorical Imputer 18 mode Iterative Imputation Categorical Model 19 None Unknown Categoricals Handling least_frequent 20 Normalize 21 True Normalize Method zscore Transformation False Transformation Method None PCA False 25 26 PCA Method None PCA Components 27 None Ignore Low Variance 28 False Combine Rare Levels False Rare Level Threshold None **Numeric Binning** 32 Remove Outliers False Outliers Threshold None Remove Multicollinearity False Multicollinearity Threshold None 35 Remove Perfect Collinearity 36 False Clustering False 37 Clustering Iteration 38 None Polynomial Features False 39 Polynomial Degree 40 None 41 **Trignometry Features** False Polynomial Threshold None **Group Features** False 43 Feature Selection False Feature Selection Method classic Features Selection Threshold None Feature Interaction False Feature Ratio False Interaction Threshold None models() Name Reference ID K-Means Clustering sklearn.cluster._kmeans.KMeans kmeans Affinity Propagation sklearn.cluster._affinity_propagation.Affinity... ap Mean Shift Clustering meanshift sklearn.cluster._mean_shift.MeanShift Spectral Clustering sklearn.cluster._spectral.SpectralClustering SC hclust Agglomerative Clustering sklearn.cluster._agglomerative.AgglomerativeCl... **Density-Based Spatial Clustering** sklearn.cluster._dbscan.DBSCAN dbscan optics **OPTICS Clustering** sklearn.cluster._optics.OPTICS birch Birch Clustering sklearn.cluster._birch.Birch kmodes K-Modes Clustering kmodes.kmodes.KModes get metrics() Needs Greater Display **Score Function** Name Scorer Target Args is Ground Name **Truth** Better ID <function silhouette Silhouette Silhouette silhouette score at make_scorer(silhouette_score) {} True False pred 0x7fe1d3b8bcb0> Calinski-Calinskichs calinski_harabasz_score make_scorer(calinski_harabasz_score) {} False True Harabasz Harabasz at 0x7fe1d3b... <function Davies-Daviesdavies_bouldin_score at make_scorer(davies_bouldin_score) db {} True False pred Bouldin Bouldin 0x7fe1d3b90050> Homogeneity homogeneity_score at Homogeneity pred {} True hs make_scorer(homogeneity_score) True Score 0x7fe1d3bead40> <function ari Rand Index Rand Index adjusted rand score at {} True True make_scorer(adjusted_rand_score) pred 0x7fe1d3beab00> <function Completeness True completeness score at make_scorer(completeness_score) Score 0x7fe1d3beadd0> Compare model performance metrics = [] for model in models().index: if model in ['meanshift', 'optics']: continue create model (model) metric result = pull() metric result['model'] = model metrics.append(metric result) Silhouette Calinski-Harabasz Davies-Bouldin Homogeneity Rand Index Completeness **0** -0.052 70.1559 8.1948 cluster metrics = pd.concat(metrics) cluster metrics.set index("model", inplace=True) cluster metrics.sort values(by='Silhouette', ascending=False, inplace=True) cluster metrics.style.highlight max(subset=['Silhouette', 'Calinski-Harabasz'], color = 'green', axis = 0).highlight min(subset=['Davies-Bouldin'], color = 'green', axis = 0) Silhouette Calinski-Harabasz Davies-Bouldin Homogeneity Rand Index Completeness model 152.003600 0 0 SC 0.337000 1567.560000 1.209700 birch 0.248600 0 kmeans 1.276900 0.207500 1800.405500 1.468600 hclust 0 0 0.171800 302.175000 1.303300 ap -0.052000 70.155900 8.194800 kmodes dbscan -0.408700 30.982000 1.564000 **Spectral Clustering Clustering** sc = create model('sc') Silhouette Calinski-Harabasz Davies-Bouldin Homogeneity Rand Index Completeness 0.8998 152.0036 0.0687

