

Group Members

- 1.ณพกษ พิมพสาร 6220422059
- 2.อมร ตยาดี 6220422058
- 3.วีรพัชร จินตนไชยรัตน์ 6220422071
- 4.มินตรา โคจิพันธุ 6220422057

```
In [1]: !pip install pycaret

Collecting pycaret
  Downloading pycaret-2.3.5-py3-none-any.whl (288 kB)
  |██████████| 288 kB 5.2 MB/s
Requirement already satisfied: wordcloud in /usr/local/lib/python3.7/dist-packages
(from pycaret) (1.5.0)
Collecting pandas-profiling>=2.8.0
  Downloading pandas_profiling-3.1.0-py2.py3-none-any.whl (261 kB)
  |██████████| 261 kB 41.6 MB/s
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (from
pycaret) (3.2.5)
Collecting umap-learn
  Downloading umap-learn-0.5.2.tar.gz (86 kB)
  |██████████| 86 kB 5.5 MB/s
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.7/dist-packages
(from pycaret) (7.6.5)
Requirement already satisfied: plotly>=4.4.1 in /usr/local/lib/python3.7/dist-packages
(from pycaret) (4.4.1)
Collecting pyod
  Downloading pyod-0.9.5.tar.gz (113 kB)
  |██████████| 113 kB 47.5 MB/s
Collecting mlxtend>=0.17.0
  Downloading mlxtend-0.19.0-py2.py3-none-any.whl (1.3 MB)
  |██████████| 1.3 MB 18.5 MB/s
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages
(from pycaret) (3.2.2)
Collecting scikit-learn==0.23.2
  Downloading scikit_learn-0.23.2-cp37-cp37m-manylinux1_x86_64.whl (6.8 MB)
  |██████████| 6.8 MB 20.3 MB/s
Requirement already satisfied: gensim<4.0.0 in /usr/local/lib/python3.7/dist-packages
(from pycaret) (3.6.0)
Requirement already satisfied: numpy==1.19.5 in /usr/local/lib/python3.7/dist-packages
(from pycaret) (1.19.5)
Requirement already satisfied: spacy<2.4.0 in /usr/local/lib/python3.7/dist-packages
(from pycaret) (2.2.4)
Collecting mlflow
  Downloading mlflow-1.22.0-py3-none-any.whl (15.5 MB)
  |██████████| 15.5 MB 28.4 MB/s
Collecting imbalanced-learn==0.7.0
  Downloading imbalanced_learn-0.7.0-py3-none-any.whl (167 kB)
  |██████████| 167 kB 56.2 MB/s
Requirement already satisfied: scipy<=1.5.4 in /usr/local/lib/python3.7/dist-packages
(from pycaret) (1.4.1)
Collecting lightgbm>=2.3.1
  Downloading lightgbm-3.3.1-py3-none-manylinux1_x86_64.whl (2.0 MB)
  |██████████| 2.0 MB 33.8 MB/s
Requirement already satisfied: textblob in /usr/local/lib/python3.7/dist-packages
(from pycaret) (0.15.3)
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (from
pycaret) (1.1.5)
```

```
Requirement already satisfied: yellowbrick>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.3.post1)
Collecting pyLDAvis
    Downloading pyLDAvis-3.3.1.tar.gz (1.7 MB)
    |██████████| 1.7 MB 42.9 MB/s
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Installing backend dependencies ... done
  Preparing wheel metadata ... done
Collecting kmodes>=0.10.1
    Downloading kmodes-0.11.1-py2.py3-none-any.whl (19 kB)
Collecting Boruta
    Downloading Boruta-0.3-py3-none-any.whl (56 kB)
    |██████████| 56 kB 3.0 MB/s
Requirement already satisfied: seaborn in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.11.2)
Requirement already satisfied: cufflinks>=0.17.0 in /usr/local/lib/python3.7/dist-packages (from pycaret) (0.17.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from pycaret) (1.1.0)
Requirement already satisfied: IPython in /usr/local/lib/python3.7/dist-packages (from pycaret) (5.5.0)
Collecting scikit-plot
    Downloading scikit_plot-0.3.7-py3-none-any.whl (33 kB)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn==0.23.2->pycaret) (3.0.0)
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: pexpect in /usr/local/lib/python3.7/dist-packages (from IPython->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: traitlets>=4.2 in /usr/local/lib/python3.7/dist-packages (from IPython->pycaret) (5.1.1)
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: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (1.0.2)
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: ipython-genutils~=0.2.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: widgetsnbextension~=3.5.0 in /usr/local/lib/python3.7/dist-packages (from ipywidgets->pycaret) (3.5.2)
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: tornado>=4.0 in /usr/local/lib/python3.7/dist-packages (from ipykernel>=4.5.1->ipywidgets->pycaret) (5.1.1)
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) (1.3.2)
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) (3.0.6)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (4.9.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-packages (from nbformat>=4.2.0->ipywidgets->pycaret) (2.6.0)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas->pycaret) (2018.9)
Collecting pydantic>=1.8.1
  Downloading pydantic-1.8.2-cp37-cp37m-manylinux2014_x86_64.whl (10.1 MB)
  [██████████] 10.1 MB 41.2 MB/s
Collecting requests>=2.24.0
  Downloading requests-2.26.0-py2.py3-none-any.whl (62 kB)
  [██████████] 62 kB 777 kB/s
Collecting joblib
  Downloading joblib-1.0.1-py3-none-any.whl (303 kB)
  [██████████] 303 kB 55.9 MB/s
Collecting multimethod>=1.4
  Downloading multimethod-1.6-py3-none-any.whl (9.4 kB)
Collecting phik>=0.11.1
  Downloading phik-0.12.0-cp37-cp37m-manylinux2010_x86_64.whl (675 kB)
  [██████████] 675 kB 50.9 MB/s
Requirement already satisfied: tqdm>=4.48.2 in /usr/local/lib/python3.7/dist-packages (from pandas-profiling>=2.8.0->pycaret) (4.62.3)
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 visions[type_image_path]==0.7.4
  Downloading visions-0.7.4-py3-none-any.whl (102 kB)
  [██████████] 102 kB 11.7 MB/s
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 46.7 MB/s
Collecting htmlmin>=0.1.12
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
Requirement already satisfied: markupsafe~=2.0.1 in /usr/local/lib/python3.7/dist-packages (from pandas-profiling>=2.8.0->pycaret) (2.0.1)
Collecting PyYAML>=5.0.0
  Downloading PyYAML-6.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_12_x86_64.manylinux2010_x86_64.whl (596 kB)
  [██████████] 596 kB 47.9 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)
Requirement already satisfied: attrs>=19.3.0 in /usr/local/lib/python3.7/dist-packages (from visions[type_image_path]==0.7.4->pandas-profiling>=2.8.0->pycaret) (21.2.0)
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.7/dist-packages (from 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 visions[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 27.8 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: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from plotly>=4.4.1->pycaret) (1.3.3)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-toolkit<2.0.0,>=1.0.4->IPython->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) (3.10.0.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (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-packages (from requests>=2.24.0->pandas-profiling>=2.8.0->pycaret) (1.24.3)
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) (2.0.8)
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: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (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-packages (from spacy<2.4.0->pycaret) (1.0.5)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (1.0.0)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (1.1.3)
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: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (2.0.6)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (0.4.1)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy<2.4.0->pycaret) (0.8.2)
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) (4.8.2)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy<2.4.0->pycaret) (3.6.0)
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: 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: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (0.12.1)
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)
Collecting docker>=4.0.0
  Downloading docker-5.0.3-py2.py3-none-any.whl (146 kB)
  [██████████] 146 kB 45.3 MB/s
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (1.3.0)
Collecting querystring-parser
  Downloading querystring_parser-1.2.4-py2.py3-none-any.whl (7.9 kB)
Collecting gitpython>=2.1.0
  Downloading GitPython-3.1.24-py3-none-any.whl (180 kB)
  [██████████] 180 kB 52.0 MB/s
Collecting prometheus-flask-exporter
  Downloading prometheus_flask_exporter-0.18.6-py3-none-any.whl (17 kB)
Requirement already satisfied: Flask in /usr/local/lib/python3.7/dist-packages (from mlflow->pycaret) (1.1.4)
```

```
Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.7/dist-packages  
(from mlflow->pycaret) (7.1.2)  
Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.7/dist-packages  
(from mlflow->pycaret) (1.4.27)  
Collecting alembic<=1.4.1  
  Downloading alembic-1.4.1.tar.gz (1.1 MB)  
 |██████████| 1.1 MB 46.1 MB/s  
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 7.4 MB/s  
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages  
(from mlflow->pycaret) (21.3)  
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  
(from mlflow->pycaret) (0.3)  
Collecting databricks-cli>=0.8.7  
  Downloading databricks-cli-0.16.2.tar.gz (58 kB)  
 |██████████| 58 kB 5.6 MB/s  
Collecting Mako  
  Downloading Mako-1.1.6-py2.py3-none-any.whl (75 kB)  
 |██████████| 75 kB 3.7 MB/s  
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->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 2.0 MB/s  
Collecting gitdb<5,>=4.0.1  
  Downloading gitdb-4.0.9-py3-none-any.whl (63 kB)  
 |██████████| 63 kB 1.2 MB/s  
Collecting smmap<6,>=3.0.1  
  Downloading smmap-5.0.0-py3-none-any.whl (24 kB)  
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-packages  
(from Flask->mlflow->pycaret) (1.0.1)  
Requirement already satisfied: itsdangerous<2.0,>=0.24 in /usr/local/lib/python3.7/dist-packages  
(from Flask->mlflow->pycaret) (1.1.0)  
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: testpath in /usr/local/lib/python3.7/dist-packages (from  
nbconvert->notebook>=4.4.1->widgetsnbextension~=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->widgetsnbextension~=3.5.0->ipywidgets->pycaret) (4.1.  
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-packages  
(from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets->pycaret)  
(0.8.4)  
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-exporter->mlflow->pycaret) (0.12.0)  
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from  
pyLDAvis->pycaret) (0.16.0)
```

```
Collecting funcy
  Downloading funcy-1.16-py2.py3-none-any.whl (32 kB)
Collecting pyLDAvis
  Downloading pyLDAvis-3.3.0.tar.gz (1.7 MB)
    [██████████] | 1.7 MB 42.4 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 41.2 MB/s
Requirement already satisfied: numexpr in /usr/local/lib/python3.7/dist-packages (from pyLDAvis->pycaret) (2.7.3)
Requirement already satisfied: numba>=0.35 in /usr/local/lib/python3.7/dist-packages (from pyod->pycaret) (0.51.2)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.7/dist-packages (from pyod->pycaret) (0.10.2)
Requirement already satisfied: llvmlite<0.35,>=0.34.0.dev0 in /usr/local/lib/python3.7/dist-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 (from statsmodels->pyod->pycaret) (0.5.2)
Collecting pynndescent>=0.5
  Downloading pynndescent-0.5.5.tar.gz (1.1 MB)
    [██████████] | 1.1 MB 39.8 MB/s
Building wheels for collected packages: htmlmin, imagehash, alembic, databricks-cli, 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=b0d511cf021a1b051aa31000d35aa12638e48d8b2a456f51606cf33384d1a53c
  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=10cb9b79b09651008dd85fc7addf9085e46a29cadff3310f3da7727cab1d40ca
  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=36e7acc3160923d394c6fc6a501c1d27f3d4f22e79651d7f08ee35873c767f7b
  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-0.16.2-py3-none-any.whl size=106811 sha256=0f6376d716f842f80f5816f288832c3053432a882dfcdfa289dcea34a6918632
  Stored in directory: /root/.cache/pip/wheels/f4/5c/ed/e1ce20a53095f63b27b4964abba03e59cf3472822addf7d29
  Building wheel for pyLDAvis (setup.py) ... done
  Created wheel for pyLDAvis: filename=pyLDAvis-3.2.2-py2.py3-none-any.whl size=135618 sha256=927cf25652c39c73b1fd26da6e386b0b0fcc19656ff0b727c52e096518acd67a
  Stored in directory: /root/.cache/pip/wheels/f8/b1/9b/560ac1931796b7303f7b517b949d2d31a4fb512aad3b9f284
  Building wheel for pyod (setup.py) ... done
  Created wheel for pyod: filename=pyod-0.9.5-py3-none-any.whl size=132699 sha256=ad04046253ae8a0cf2eba182ea6eee703f14e3119dc1960d458218c050d0dd03
  Stored in directory: /root/.cache/pip/wheels/3d/bb/b7/62b60fb451b33b0df1ab8006697fba7a6a49709a629055cf77
  Building wheel for umap-learn (setup.py) ... done
  Created wheel for umap-learn: filename=umap_learn-0.5.2-py3-none-any.whl size=82709 sha256=c011d887f76eb83b8a17124a079806ed4b03087900bceb88ceb1790af612c2ec
  Stored in directory: /root/.cache/pip/wheels/84/1b/c6/aaaf68a748122632967cef4dffef68224eb16798b6793257d82
  Building wheel for pynndescent (setup.py) ... done
  Created wheel for pynndescent: filename=pynndescent-0.5.5-py3-none-any.whl size=52603 sha256=a6440c49006783ac702cae3e36269d5702f1bcfab9cd42ee8cb4ac1bde74c01c
```

```
Stored in directory: /root/.cache/pip/wheels/af/e9/33/04db1436df0757c42fda8ea6796d  
7a8586e23c85fac355f476  
Successfully built htmlmin imagehash alembic databricks-cli pyLDAvis pyod umap-learn  
pynndescent  
Installing collected packages: tangled-up-in-unicode, smmap, scipy, multimethod, job  
lib, websocket-client, visions, scikit-learn, requests, python-editor, Mako, imageha  
sh, gitdb, querystring-parser, PyYAML, pynndescent, pydantic, prometheus-flask-expo  
rter, phik, htmlmin, gunicorn, gitpython, funcy, docker, databricks-cli, alembic, uma  
p-learn, scikit-plot, pyod, pyLDAvis, pandas-profiling, mlxtend, mlflow, lightgbm, k  
modes, 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: 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 1.0.1  
Uninstalling scikit-learn-1.0.1:  
    Successfully uninstalled scikit-learn-1.0.1  
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 packag  
es that are installed. This behaviour is the source of the following dependency con  
flicts.  
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 incomp  
atible.  
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-1.1.6 PyYAML-6.0 alembic-1.4.1 databricks-cli  
-0.16.2 docker-5.0.3 funcy-1.16 gitdb-4.0.9 gitpython-3.1.24 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-  
3.3.1 mlflow-1.22.0 mlxtend-0.19.0 multimethod-1.6 pandas-profiling-3.1.0 phik-0.12.  
0 prometheus-flask-exporter-0.18.6 pyLDAvis-3.2.2 pycaret-2.3.5 pydantic-1.8.2 pynn  
descent-0.5.5 pyod-0.9.5 python-editor-1.0.4 querystring-parser-1.2.4 requests-2.26.0  
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 websocket-client-1.2.3
```

```
In [2]: import pandas as pd  
from pycaret.clustering import *
```

```
In [3]: from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive

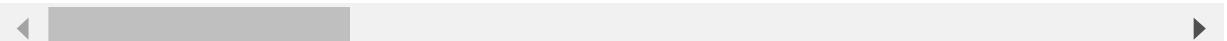
```
In [4]: df = pd.read_csv('drive/MyDrive/Colab Notebooks/CRM - Customer Segmentation/Copy of
```

```
In [5]: df['SHOP_DATE'] = df['SHOP_DATE'].apply(lambda x: pd.to_datetime(str(x), format='%Y%
```

```
In [6]: df.tail(5)
```

```
Out[6]:
```

	SHOP_WEEK	SHOP_DATE	SHOP_WEEKDAY	SHOP_HOUR	QUANTITY	SPEND	PROD_CODE
956569	200617	2006-06-22	5	12	3	3.96	PRD0904997
956570	200633	2006-10-13	6	20	3	3.96	PRD0904997
956571	200617	2006-06-22	5	18	3	3.96	PRD0904997
956572	200619	2006-07-06	5	19	3	3.96	PRD0904997
956573	200635	2006-10-23	2	21	3	3.96	PRD0904997



Prepare customer single view

```
In [7]: ##prepare customer single view  
df_csv = df[df['CUST_CODE'].notnull()].groupby(by=['CUST_CODE']).agg(TotalSpend=( 'SP  
To  
To  
To  
Fi  
La  
)
```

```
In [8]: ##calculate ticket size  
df_csv['TicketSize'] = df_csv['TotalSpend']/df_csv['TotalVisits']
```

```
In [9]: df_csv['TicketVolumn'] = df_csv['TotalQuantity']/df_csv['TotalVisits']
```

```
In [10]: ##calculate total days of the relationship  
df_csv['total_days'] = (df_csv['LastDate'] - df_csv['FirstDate']).dt.days + 1
```

```
In [11]: ##find max date in the dataset  
max_date = df_csv['LastDate'].max()  
  
##calculate recency days  
df_csv['recency'] = (max_date - df_csv['LastDate']).dt.days
```

In [12]:

```
df_csv.head(20)
```

Out[12]:

	CUST_CODE	TotalSpend	TotalVisits	TotalSKUs	TotalQuantity	FirstDate	LastDate	TicketSi
0	CUST0000000181	2.44	1	1	1	2007-01-06	2007-01-06	2.4400
1	CUST0000000369	959.33	220	189	923	2006-04-16	2008-07-05	4.3605
2	CUST0000000689	328.57	16	116	334	2007-07-22	2008-06-23	20.5356
3	CUST0000000998	5.95	3	4	6	2006-05-04	2006-07-07	1.9833
4	CUST0000001163	39.74	4	24	34	2006-10-22	2008-06-22	9.9350
5	CUST0000001194	2493.83	103	403	1705	2006-04-22	2008-06-20	24.2119
6	CUST0000002257	1.99	1	1	1	2006-07-21	2006-07-21	1.9900
7	CUST0000002499	0.34	1	1	1	2006-04-14	2006-04-14	0.3400
8	CUST0000002605	1.80	6	1	6	2006-10-08	2008-06-12	0.3000
9	CUST0000002637	124.63	15	53	126	2006-08-03	2008-05-30	8.3086
10	CUST0000002638	15.59	1	4	6	2007-04-14	2007-04-14	15.5900
11	CUST0000002813	145.02	12	73	118	2007-07-31	2008-04-12	12.0850
12	CUST0000003028	10.61	6	7	8	2007-09-16	2008-01-04	1.7683
13	CUST0000003044	32.17	2	21	37	2007-01-13	2007-01-24	16.0850
14	CUST0000003252	3.83	2	3	5	2007-10-11	2008-04-04	1.9150
15	CUST0000003287	254.10	13	122	227	2006-05-02	2008-05-10	19.5461
16	CUST0000003323	6.31	2	5	7	2007-01-01	2007-08-29	3.1550
17	CUST0000003667	6.98	1	2	2	2007-01-11	2007-01-11	6.9800
18	CUST0000004088	900.30	82	239	912	2006-04-13	2008-06-29	10.9792
19	CUST0000004507	2.35	1	2	2	2007-06-17	2007-06-17	2.3500



Cluster customers

In [13]:

```
exp_clu = setup(data=df_csv, ignore_features=['CUST_CODE', 'FirstDate', 'LastDate'],
```

	Description	Value
0	session_id	5721
1	Original Data	(6100, 11)
2	Missing Values	False
3	Numeric Features	8
4	Categorical Features	0
5	Ordinal Features	False
6	High Cardinality Features	False
7	High Cardinality Method	None
8	Transformed Data	(6100, 8)
9	CPU Jobs	-1
10	Use GPU	False
11	Log Experiment	False
12	Experiment Name	cluster-default-name
13	USI	cfa4
14	Imputation Type	simple
15	Iterative Imputation Iteration	None
16	Numeric Imputer	mean
17	Iterative Imputation Numeric Model	None
18	Categorical Imputer	mode
19	Iterative Imputation Categorical Model	None
20	Unknown Categoricals Handling	least_frequent
21	Normalize	True
22	Normalize Method	zscore
23	Transformation	False
24	Transformation Method	None
25	PCA	False
26	PCA Method	None
27	PCA Components	None
28	Ignore Low Variance	False
29	Combine Rare Levels	False
30	Rare Level Threshold	None
31	Numeric Binning	False

ID	Description	Value
32	Remove Outliers	False
33	Outliers Threshold	None
34	Remove Multicollinearity	False
35	Multicollinearity Threshold	None
36	Remove Perfect Collinearity	False
37	Clustering	False
38	Clustering Iteration	None
39	Polynomial Features	False
40	Polynomial Degree	None
41	Trigonometry Features	False
42	Polynomial Threshold	None
43	Group Features	False
44	Feature Selection	False
45	Feature Selection Method	classic
46	Features Selection Threshold	None
47	Feature Interaction	False
48	Feature Ratio	False
49	Interaction Threshold	None

In [14]:

```
models()
```

Out[14]:

ID	Name	Reference
kmeans	K-Means Clustering	sklearn.cluster._kmeans.KMeans
ap	Affinity Propagation	sklearn.cluster._affinity_propagation.Affinity...
meanshift	Mean Shift Clustering	sklearn.cluster._mean_shift.MeanShift
sc	Spectral Clustering	sklearn.cluster._spectral.SpectralClustering
hclust	Agglomerative Clustering	sklearn.cluster._agglomerative.AgglomerativeCl...
dbscan	Density-Based Spatial Clustering	sklearn.cluster._dbscan.DBSCAN
optics	OPTICS Clustering	sklearn.cluster._optics.OPTICS
birch	Birch Clustering	sklearn.cluster._birch.Birch
kmodes	K-Modes Clustering	kmodes.kmodes.KModes

In [15]:

```
get_metrics()
```

Out[15]:

ID	Name	Display Name	Score Function	Scorer	Target
silhouette	Silhouette	Silhouette	<function silhouette_score at 0x7f7e29cb3170>	make_scorer(silhouette_score)	p
chs	Calinski-Harabasz	Calinski-Harabasz	<function calinski_harabasz_score at 0x7f7e29c...	make_scorer(calinski_harabasz_score)	p
db	Davies-Bouldin	Davies-Bouldin	<function davies_bouldin_score at 0x7f7e29cb34d0>	make_scorer(davies_bouldin_score)	p
hs	Homogeneity Score	Homogeneity	<function homogeneity_score at 0x7f7e29f2b200>	make_scorer(homogeneity_score)	p
ari	Rand Index	Rand Index	<function adjusted_rand_score at 0x7f7e29f26f80>	make_scorer(adjusted_rand_score)	p
cs	Completeness Score	Completeness	<function completeness_score at 0x7f7e29f2b290>	make_scorer(completeness_score)	p

Compare model performance

In [16]:

```
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	-0.1801	128.5355	9.0214	0	0	0

In [17]:

```
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'], colo
```

Out[17]:

model	Silhouette	Calinski-Harabasz	Davies-Bouldin	Homogeneity	Rand Index	Completeness
sc	0.915900	186.239700	0.058500	0	0	0
birch	0.449700	1793.673300	0.859200	0	0	0

	Silhouette	Calinski-Harabasz	Davies-Bouldin	Homogeneity	Rand Index	Completeness
model						
kmeans	0.409700	3504.455200	0.900400	0	0	0
hclust	0.364900	2832.447700	0.974800	0	0	0
dbscan	0.142900	193.608200	1.807800	0	0	0
ap	0.000000	0.000000	0.000000	0	0	0
kmodes	-0.180100	128.535500	9.021400	0	0	0

Spectral Clustering Clustering

In [18]:

```
sc = create_model('sc')
```

	Silhouette	Calinski-Harabasz	Davies-Bouldin	Homogeneity	Rand Index	Completeness
0	0.9159	186.2397	0.0585	0	0	0

In [19]:

```
plot_model(sc)
```

KMeans Clustering

In [20]:

```
kmeans = create_model('kmeans', num_clusters = 5)
```

	Silhouette	Calinski-Harabasz	Davies-Bouldin	Homogeneity	Rand Index	Completeness
0	0.4107	3416.4862	0.9069	0	0	0

In [21]:

```
print(kmeans)
```

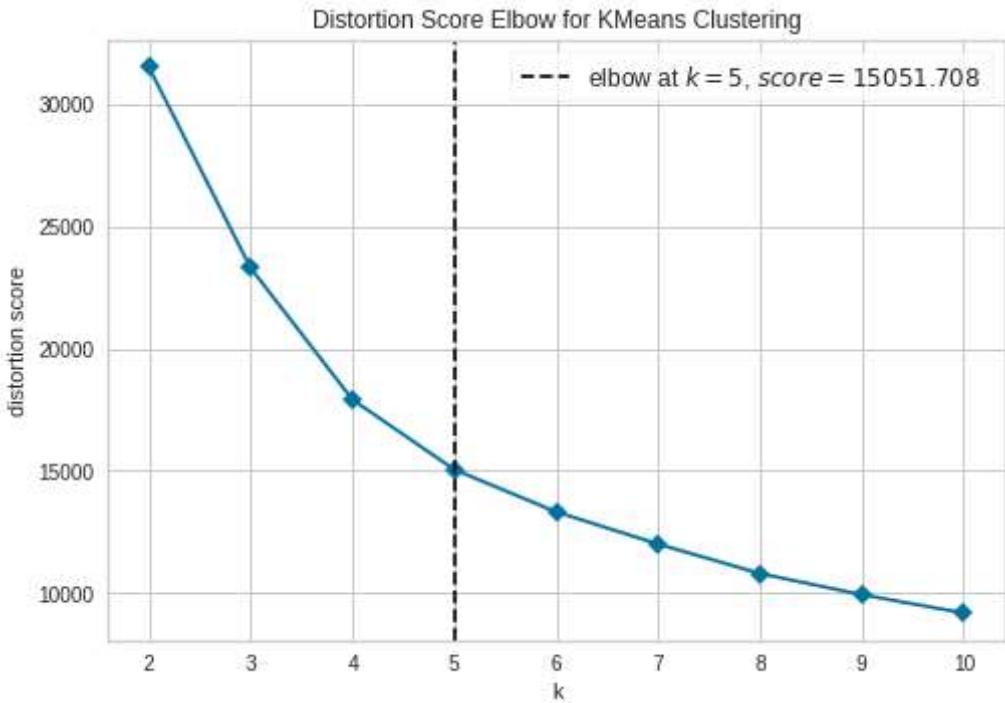
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=5, n_init=10, n_jobs=-1, precompute_distances='deprecated',
       random_state=5721, tol=0.0001, verbose=0)
```

In [22]:

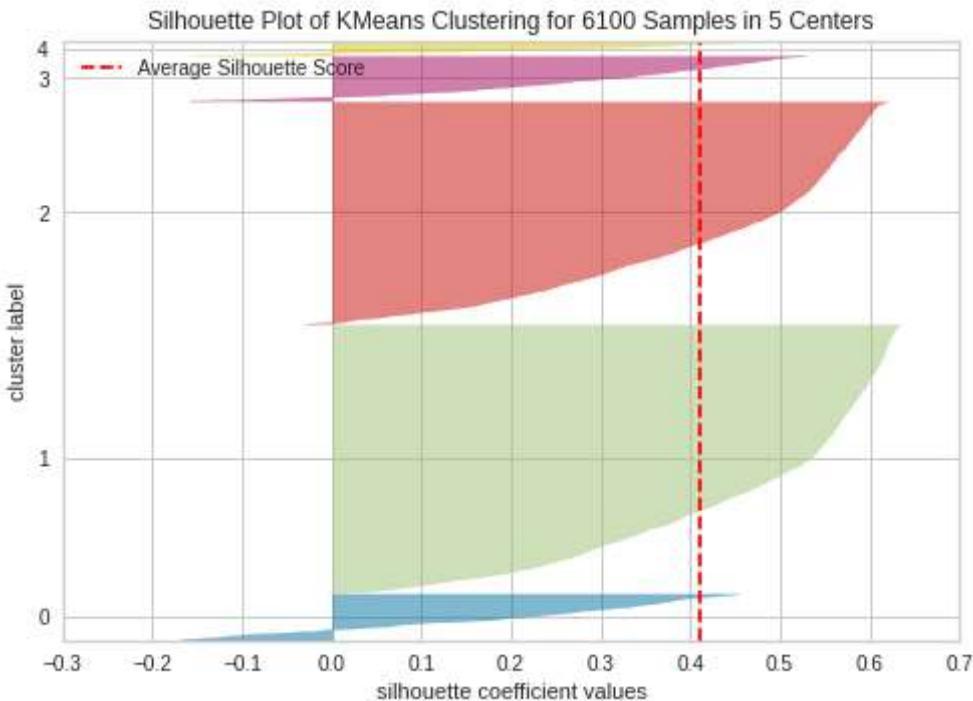
```
plot_model(kmeans)
```

In [23]:

```
plot_model(kmeans, plot = 'elbow')
```



```
In [24]: plot_model(kmeans, plot = 'silhouette')
```



```
In [29]: df_csv = df_csv.drop(columns=['CUST_CODE', 'FirstDate', 'LastDate'])

df_csv['cluster'] = kmeans.labels_
df_csv
```

Out[29]:

	TotalSpend	TotalVisits	TotalSKUs	TotalQuantity	TicketSize	TicketVolumn	total_days	recen
0	2.44	1	1	1	2.440000	1.000000	1	5
1	959.33	220	189	923	4.360591	4.195455	812	
2	328.57	16	116	334	20.535625	20.875000	338	

	TotalSpend	TotalVisits	TotalSKUs	TotalQuantity	TicketSize	TicketVolumn	total_days	recen
3	5.95	3	4	6	1.983333	2.000000	65	7
4	39.74	4	24	34	9.935000	8.500000	610	
...
6095	453.58	30	206	395	15.119333	13.166667	698	
6096	105.11	11	46	91	9.555455	8.272727	740	
6097	5.86	1	4	6	5.860000	6.000000	1	7
6098	444.17	21	127	343	21.150952	16.333333	555	
6099	17.67	6	13	21	2.945000	3.500000	624	

6100 rows × 9 columns

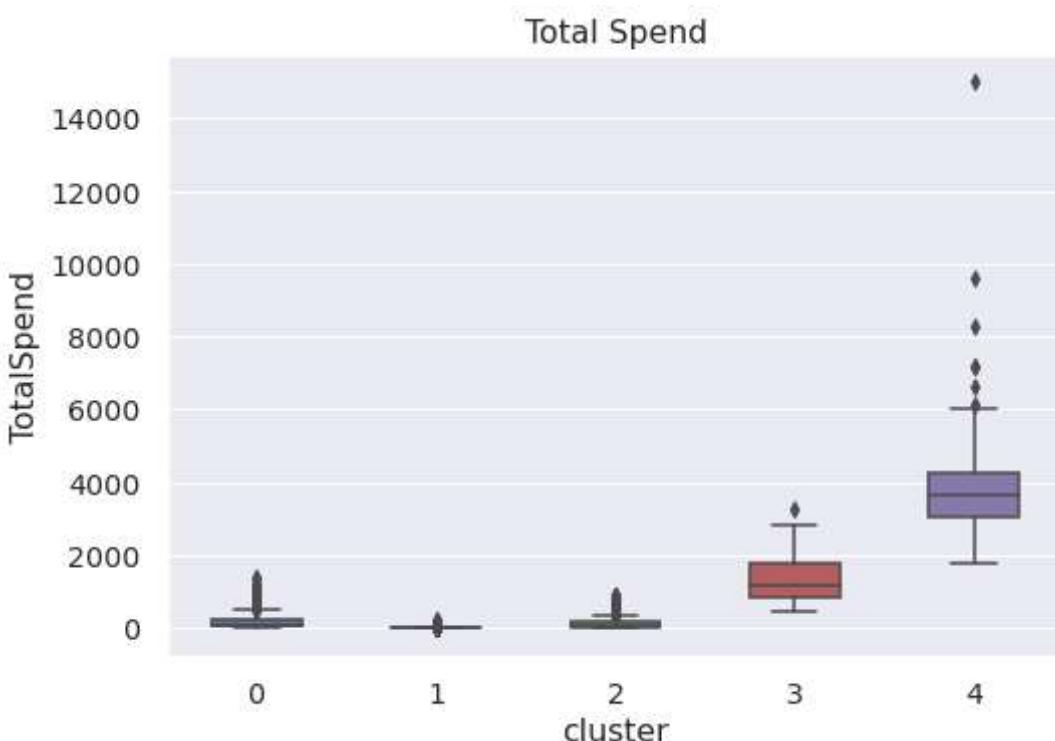


In [40]:

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.set(font_scale = 1.3)
sns.boxplot(x="cluster", y="TotalSpend", data=df_csv, width=0.5)
plt.title("Total Spend")
```

Out[40]:

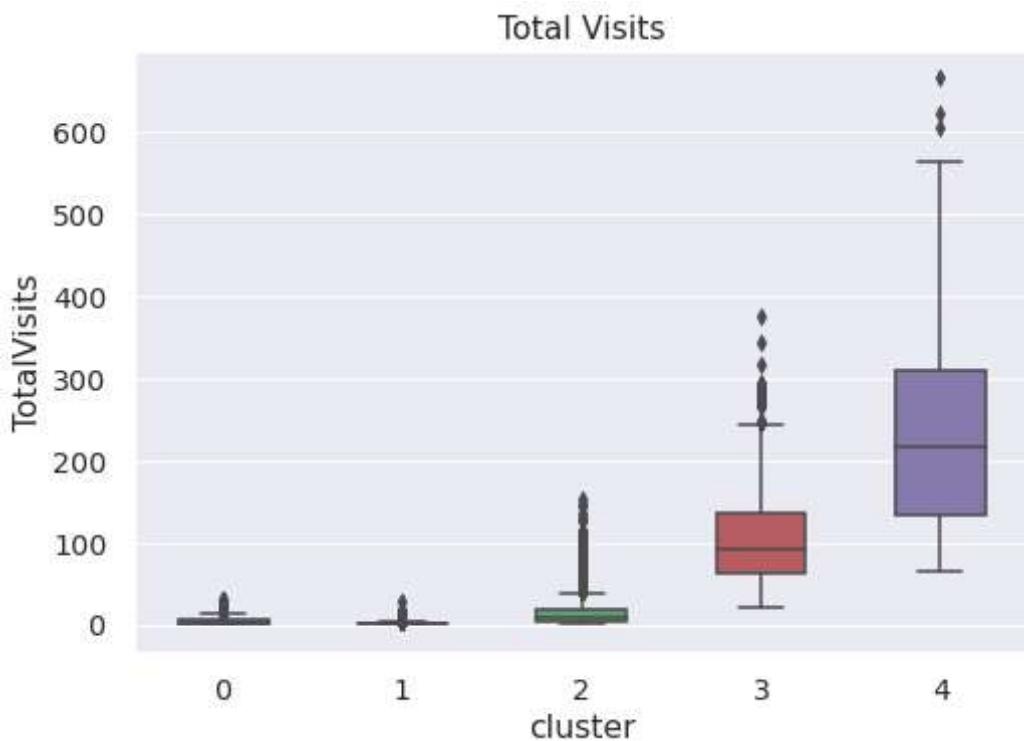


In [41]:

```
sns.set(font_scale = 1.3)
sns.boxplot(x="cluster", y="TotalVisits", data=df_csv, width=0.5)
plt.title("Total Visits")
```

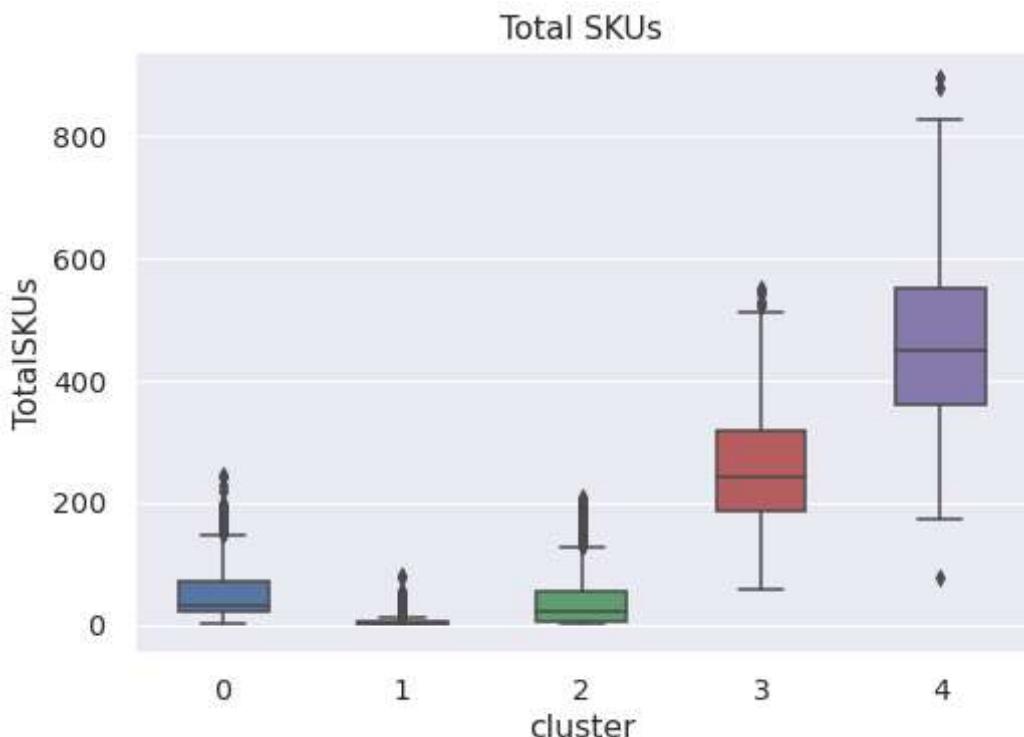
Out[41]:

Text(0.5, 1.0, 'Total Visits')



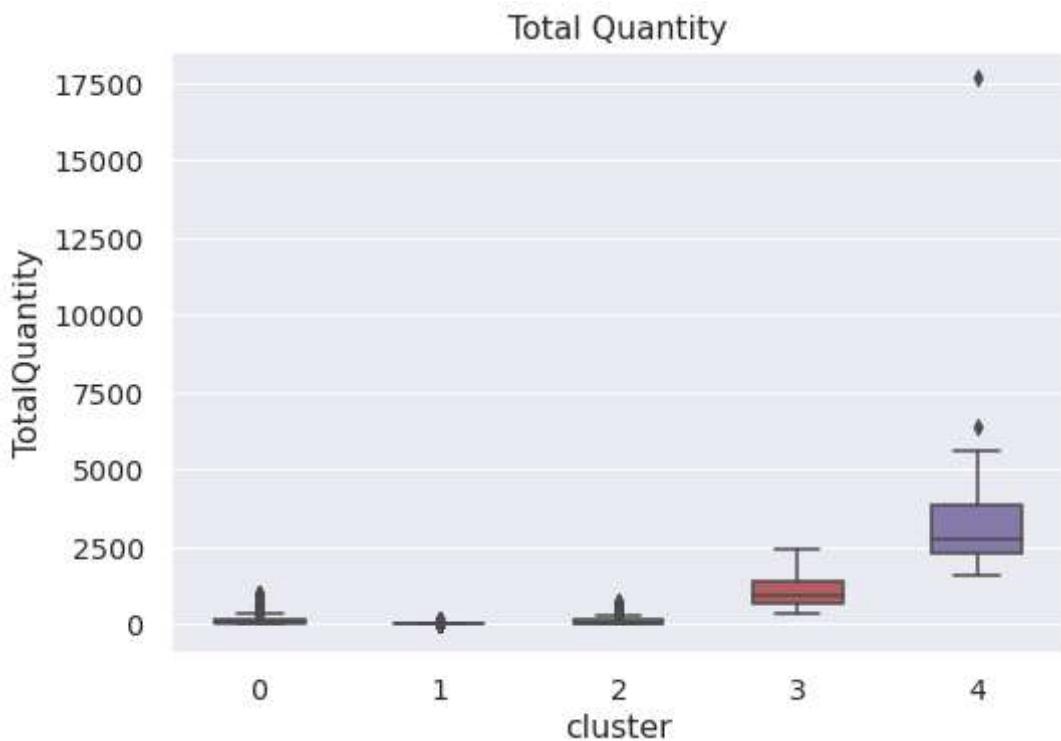
```
In [45]: sns.set(font_scale = 1.3)
sns.boxplot(x="cluster", y="TotalSKUs", data=df_csv, width=0.5)
plt.title("Total SKUs")
```

Out[45]: Text(0.5, 1.0, 'Total SKUs')



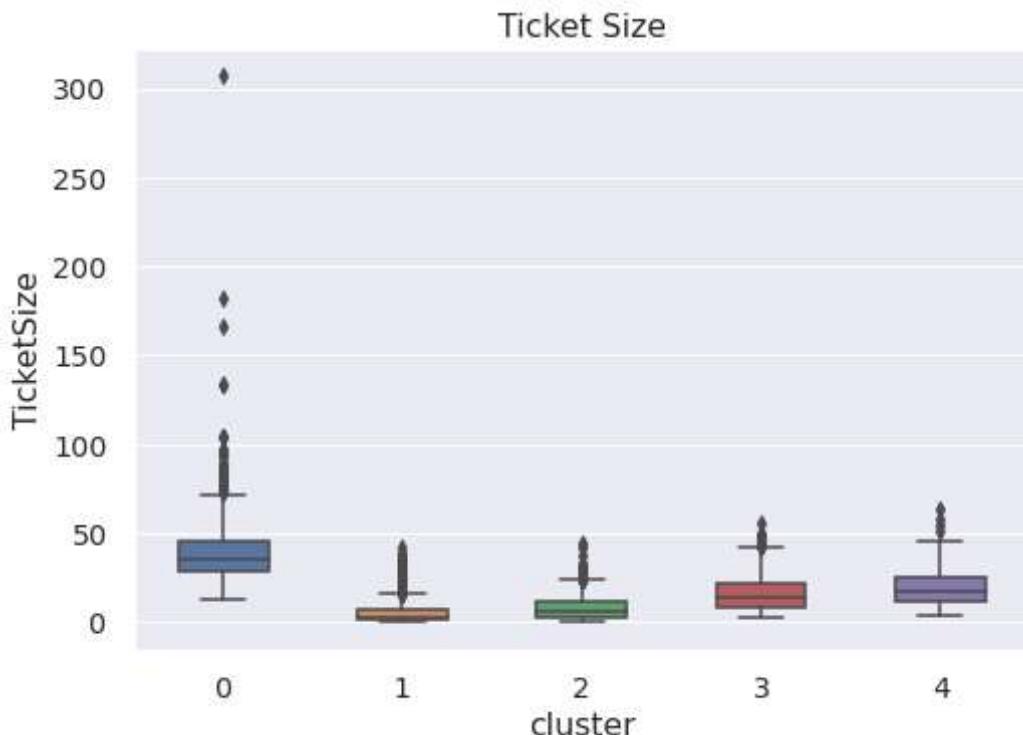
```
In [46]: sns.set(font_scale = 1.3)
sns.boxplot(x="cluster", y="TotalQuantity", data=df_csv, width=0.5)
plt.title("Total Quantity")
```

Out[46]: Text(0.5, 1.0, 'Total Quantity')



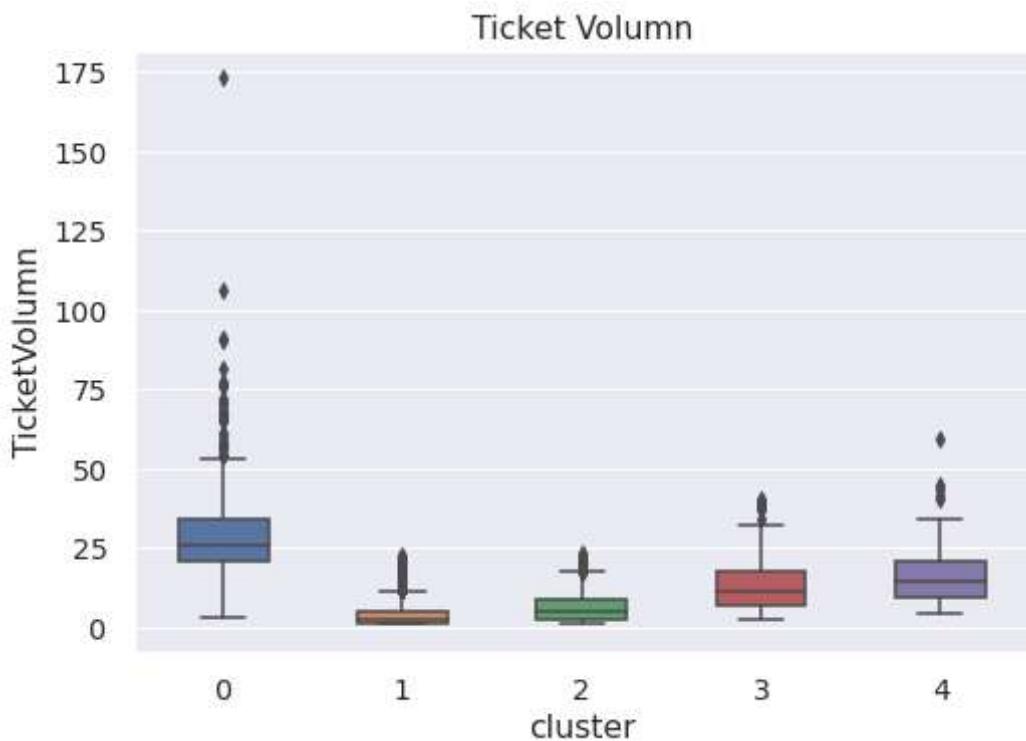
```
In [47]:  
sns.set(font_scale = 1.3)  
sns.boxplot(x="cluster", y="TicketSize", data=df_csv, width=0.5)  
plt.title("Ticket Size")
```

Out[47]: Text(0.5, 1.0, 'Ticket Size')



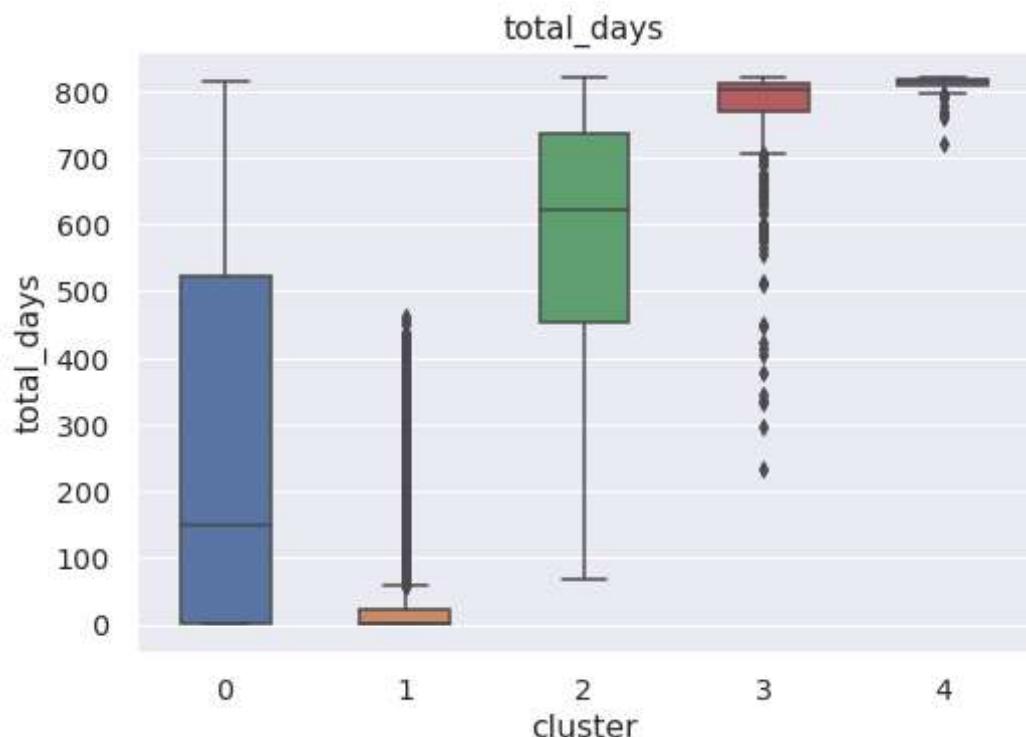
```
In [48]:  
sns.set(font_scale = 1.3)  
sns.boxplot(x="cluster", y="TicketColumn", data=df_csv, width=0.5)  
plt.title("Ticket Volumn")
```

Out[48]: Text(0.5, 1.0, 'Ticket Volumn')



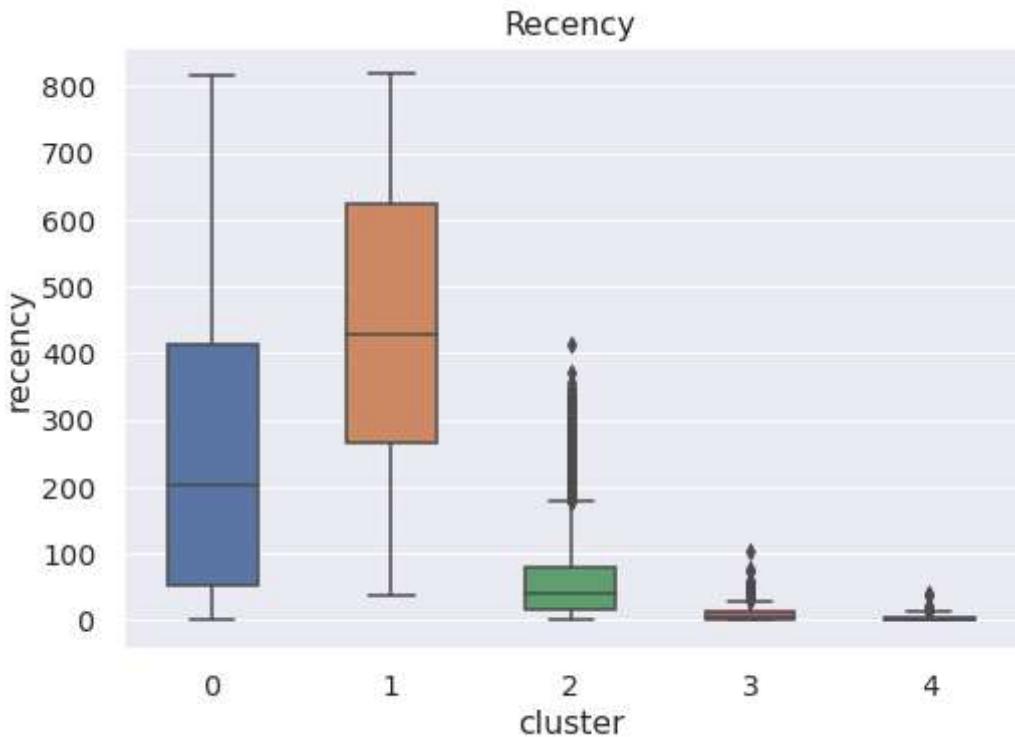
```
In [49]:  
sns.set(font_scale = 1.3)  
sns.boxplot(x="cluster", y="total_days", data=df_csv, width=0.5)  
plt.title("total_days")
```

Out[49]: Text(0.5, 1.0, 'total_days')



```
In [50]:  
sns.set(font_scale = 1.3)  
sns.boxplot(x="cluster", y="recency", data=df_csv, width=0.5)  
plt.title("Recency")
```

Out[50]: Text(0.5, 1.0, 'Recency')



Interpret

Cluster 0

ชีอกลุ่ม > เพื่อนสนิทที่อยู่ต่างประเทศ นานๆ เจอ กันที่แต่เจอทีกคุยๆ รา

เนื่องจาก ticket size และ ticket volume สูงกว่ากลุ่มนี้ หมายความว่าเป็นลูกค้าที่ชอบซื้อสินค้าราคาไม่แพง และซื้อครั้งละเยอะๆ ขณะที่จำนวนครั้งที่มาซื้อก็ไม่ปอยแสดงให้เห็นว่าอาจจะเป็นลูกค้าที่ชอบตุนของหรือซื้อเฉพาะของใช้จำเป็นซึ่งจะซื้อเป็นรอบๆ

Action

1. ทำโปรโมชั่นให้เพื่อเพิ่มจำนวนครั้งที่จะมาซื้อของที่ร้าน เช่น หากใช้จ่ายครบ 500 บาท เป็นจำนวน 8 ครั้งภายใน 1 เดือน คุณจะได้ส่วนลด 50% ในการซื้อของรอบหน้า

Cluster 1

ชีอกลุ่ม > เพื่อนที่ไม่เคยกับเราแล้ว เจอ กันทีกตามงานสังคมในกลุ่มเพื่อน เช่นงานแต่ง

เนื่องจากครั้งล่าสุดที่ลูกค้ามาเก็บนานมาก แฉมยังซื้อของน้อย ยอดก็ไม่เยอะเมื่อเทียบกับกลุ่มนี้ ดังนั้นจึงบอกได้ว่ากลุ่มนี้เป็นลูกค้าขาจรที่บังเอิญและผ่านร้านมาซื้อของติดมือnidหน่อย

Action

1. "ไม่ทำอะไร เพราะเปลือง cost เราไปลงทุนกับลูกค้ากลุ่มนี้ดีกว่า"

Cluster 2

ชีอกลุ่ม > เพื่อนที่คบกันเพราะผลประโยชน์ เช่น เป็น business partner

เป็นลูกค้าที่มีความสัมพันธ์กับร้านเราระดับกลางๆ เนื่องจาก recency ค่อนข้างต่ำ แต่มี outlier ที่ไปทางค่าสูงอยู่มาก และลูกค้ากลุ่มนี้มีค่าจำนวนครั้งที่มาซื้อ และยอดการซื้อไม่ได้แตกต่างจากกลุ่มอื่นมากนักแต่อยู่ในระดับกลางๆ จากทั้ง 5 กลุ่ม

Action

1. ส่งโปรโมชั่นส่วนลดไปให้เพื่อให้เก็บลับมาซื้อของที่ร้านปอยช์น รวมถึงทำ cross selling และ upselling กับสินค้าต่างๆ ให้กับลูกค้ากลุ่มนี้ เพื่อให้ลูกค้ากลุ่มนี้ loyal กับเรามากขึ้น เพราะเป็นกลุ่มที่มี potential ที่จะเป็นลูกค้าประจำเราสูง

Cluster 3

ชื่อกลุ่ม > เพื่อนสนิทที่เพิ่งรู้จักกันได้ไม่นาน และยังคุยกันอยู่เรื่อยๆ

เป็นลูกค้าประจำที่มาปอยฯ และใช้ชื่อของเยอะ ใช้จ่ายค่อนข้างสูง

Action

1. ทำ Upselling เพื่อให้ลูกค้ากลุ่มนี้ใช้จ่ายสูงขึ้น

Cluster 4

ชื่อกลุ่ม > เพื่อนตายที่คบกันนานนานมาก เป็นเพื่อนแท้ที่หายาก

เป็นลูกค้าแฟนพันธ์แท้ของร้าน เพราะมาบ่อยมาก และชื่อของเยอะ ชื่อที่ยอดก็เยอะ เป็นลูกค้าที่ร้านนานมากๆ

Action

1. จัดกิจกรรมทาง online และ offline ให้เฉพาะลูกค้ากลุ่มนี้เพื่อ maintain relationship ปัจจุบันไว้