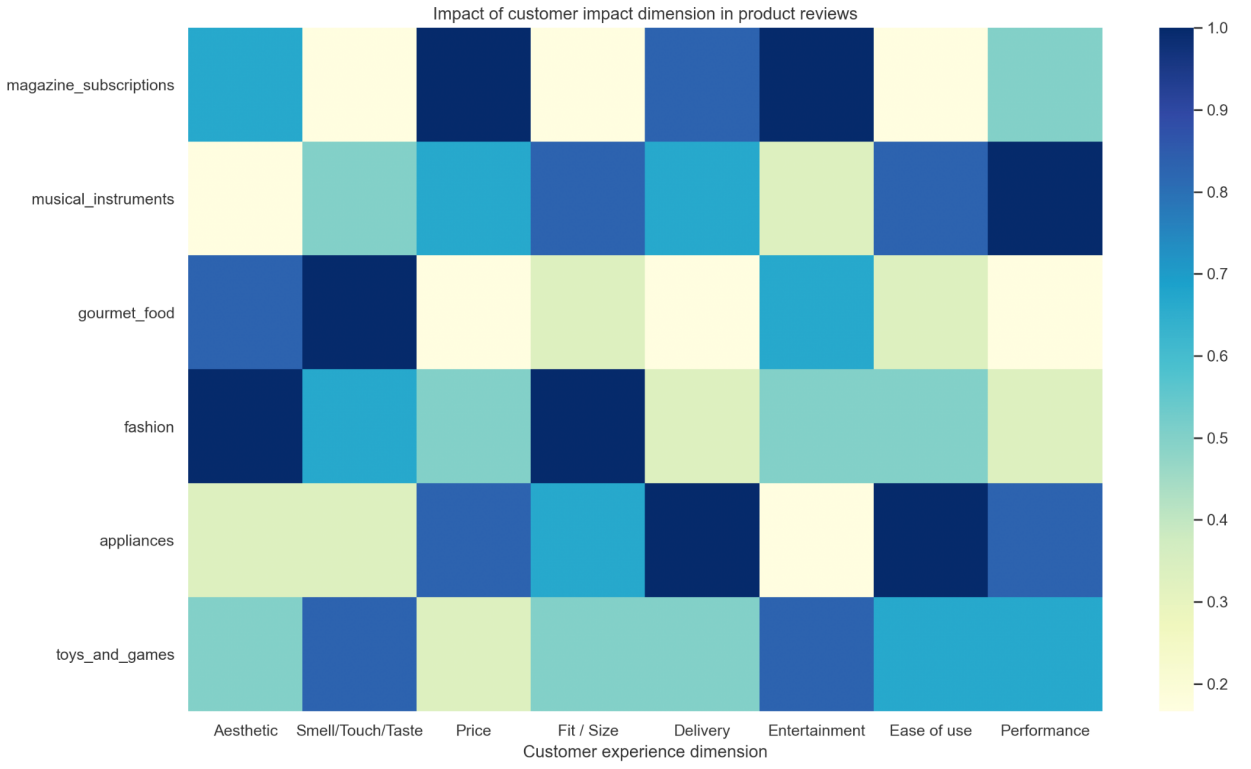
All source code is stored in the 'src' folder. The execution of commands occurs from the root of the directory in the run_logit.py and run_topic_model.py files. All commands (including creation of a Docker instance to run a Flask app) have been aggregated to simple commands in a Makefile.

Output generated when running run_logit.py (with the make train_models command)

```
(py3) [23:14:58] louisgenereux:Independent project git:(main) $ make train_models
python run_logit.py
- Generating results for: musical_instruments
Number of reviews: 369776
Total words: 18531520
Acc:  0.91264  Prec:  [0.912 0.913]  Rec:  [0.913 0.912]  f1:  [0.913 0.913]
- Generating results for: magazine_subscriptions
Number of reviews: 32502
Total words: 1671134
Acc:  0.88586  Prec:  [0.881 0.891]  Rec:  [0.894 0.877]  f1:  [0.888 0.884]
- Generating results for: fashion
Number of reviews: 343594
Total words: 10050838
Acc:  0.91229  Prec:  [0.908 0.917]  Rec:  [0.916 0.908]  f1:  [0.912 0.912]
- Generating results for: appliances
Number of reviews: 160720
Total words: 8073989
Acc:  0.91159  Prec:  [0.907 0.916]  Rec:  [0.916 0.907]  f1:  [0.912 0.911]
- Generating results for: gourmet_food
Number of reviews: 211900
Total words: 8882375
Acc:  0.90779  Prec:  [0.905 0.911]  Rec:  [0.911 0.905]  f1:  [0.908 0.908]
- Generating results for: toys_and_games
Number of reviews: 244002
Total words: 12524412
Acc:  0.92271  Prec:  [0.921 0.925]  Rec:  [0.925 0.921]  f1:  [0.923 0.923]
- Results stored
```

Output generated when running run_topic.py (with the make topic_models *command*)

```
(py3) [23:22:05] louisgenereux:Independent project git:(main) $ make topic_models
python run_topic_model.py
- pretrained glove embeddings read in
- logistic regression parameters summary read in
- Synonym of dimension descriptors identified, embeddings stored
- All logit parameters evaluated (proximity to dimensions)
- Parameter proximity to dimension aggregated at product level (weighted by parameter importance
+ Scaled aggregated proximity:
         product_category  Aesthetic  Smell/Touch/Taste     Price  Fit / Size  Delivery  Entertainment  Ease of use  Performance
0  magazine_subscriptions   1.000000          -1.000000  1.000000   -1.000000  0.468475       1.000000    -1.000000    -0.694483
1     musical_instruments  -0.709852          -0.189461  0.504980    0.346558  0.139698      -0.646852     1.000000     0.654940
2                 fashion   0.686525          -0.004652  0.207048    1.000000 -0.539372      -0.106521    -0.110442    -0.402950
3            gourmet_food   0.788818           1.000000 -1.000000   -0.685127 -1.000000       0.419228    -0.008094    -1.000000
4          toys_and_games   0.508925           0.073464 -0.307433   -0.841882 -0.502338       0.618093     0.488142    -0.174514
5              appliances  -1.000000          -0.557539  0.974156   -0.519674  1.000000      -1.000000     0.395151     1.000000
+ Heatmap created
+ Product comparison created
```



Output generated when running tests (with the python -m pytest command)

```
(py3) [2:01:24] louisgenereux:Independent project git:(main) $ python -m pytest
============================================================================= test session starts ==============================================================================
platform darwin -- Python 3.7.9, pytest-5.4.2, py-1.10.0, pluggy-0.13.1
rootdir: /Users/louisgenereux/Desktop/Term 4/Text_analytics/Project/Independent project
collected 4 items

test/test_pre_process.py ..                                                                                                                                          [ 50%]
test/test_summary_transformations.py ..                                                                                                                              [100%]

============================================================================== 4 passed in 2.15s ===============================================================================
```

Output generated when creating the app Docker image (with the make app_image command)

```
(py3) [23:33:25] louisgenereux:Independent project git:(main) $ make app_image
docker build -f app/Dockerfile_app -t app_lcg:latest .
[+] Building 34.7s (14/14) FINISHED
 => [internal] load build definition from Dockerfile_app
 => => transferring dockerfile: 41B
 => [internal] load .dockerignore
 => => transferring context: 2B
 => [internal] load metadata for docker.io/library/ubuntu:18.04
 => [internal] load build context
 => => transferring context: 41.97kB
 => [1/9] FROM docker.io/library/ubuntu:18.04@sha256:0fedbd5bd9fb72089c7bbca476949e10593cebed9b1fb9edf5b79dbbacddd7d6
 => CACHED [2/9] RUN apt-get update -y && apt-get install -y python3-pip python3-dev git gcc dos2unix g++
 => CACHED [3/9] COPY ./requirements.txt /app/requirements.txt
 => CACHED [4/9] WORKDIR /app
 => CACHED [5/9] RUN pip3 install --upgrade pip
 => CACHED [6/9] RUN pip3 install -r requirements.txt
 => [7/9] COPY . /app
 => [8/9] RUN dos2unix app/boot.sh && apt-get --purge remove -y dos2unix
 => [9/9] RUN chmod +x app/boot.sh
 => exporting to image
 => => exporting layers
 => => writing image sha256:31d8a6905d5ca94df4643868f175e79800f8dbd38da9cecb1fe12878d3559926
 => => naming to docker.io/library/app_lcg:latest
```

MSiA 414: Natural Language Processing Independent Project

Output after launching Docker app (with the make run_app command)

The app can be accessed from http://127.0.0.1:5000/

NLP INDEPENDENT PROJECT: Assessing the importance of user-defined dimensions in product reviews

This tool allows users to compare the relative importance of various user-experience dimensions across different product categories

User input

Please enter two product categories.

Type product type 1

Type product type 2

Visualize