RNN-based Dialogue Generation

Conditional Text-to-Image Synthesis

[3]: `!nvidia-smi`

```
Mon Apr1 18:16:20 2024
+-------------------------------------------------------------------
----------
--------+
| NVIDIA-SMI 546.26    Driver Version: 546.26    CUDA Version: 12.3
 |
|----------------------------------------+----------------------+----
----------
--------+
|   GPU    Name                TCC/WDDM  | Bus-Id    Disp.A | Volatile
Uncorr. ECC |
| Fan Temp  Perf         Pwr:Usage/Cap  |      Memory-Usage  | GPU-
Compute M. |
|                                       |                  |
MIG M. |
|========================================+======================+====
==========
========|
|   0NVIDIA GeForce RTX 3050 …WDDM| 00000000:01:00.0 Off |
N/A |
| N/A  56C   P8              3W / 60W |    0MiB / 4096MiB |    0%
Default |
|                                       |                  |
N/A |
+----------------------------------------+----------------------+----
----------
--------+

+-------------------------------------------------------------------
----------
--------+
|
Processes:
|
```

```
| GPU  GI  CI        PID  Type  Process name                          GPU
Memory |
|       ID  ID
Usage      |
|=================================================================
==========
=======|
|   0  N/A N/A    34488   C+G…inaries\Win64\EpicGamesLauncher.exe
N/A    | +-----------------------------------------------------------
-----------------
--------+
```

[24]: `!pip install transformers`

Requirement already satisfied: transformers in
c:\users\sdill\anaconda3\lib\site-packages (4.39.2)
Requirement already satisfied: filelock in
c:\users\sdill\anaconda3\lib\sitepackages (from transformers) (3.13.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in
c:\users\sdill\anaconda3\lib\site-packages (from transformers)
(0.22.2) Requirement already satisfied: numpy>=1.17 in
c:\users\sdill\anaconda3\lib\sitepackages (from transformers) (1.26.4)
Requirement already satisfied: packaging>=20.0 in
c:\users\sdill\anaconda3\lib\site-packages (from transformers) (23.1)
Requirement already satisfied: pyyaml>=5.1 in
c:\users\sdill\anaconda3\lib\sitepackages (from transformers) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in
c:\users\sdill\anaconda3\lib\site-packages (from transformers)
(2023.10.3) Requirement already satisfied: requests in
c:\users\sdill\anaconda3\lib\sitepackages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers<0.19,>=0.14 in
c:\users\sdill\anaconda3\lib\site-packages (from transformers)
(0.15.2) Requirement already satisfied: safetensors>=0.4.1 in
c:\users\sdill\anaconda3\lib\site-packages (from transformers) (0.4.2)
Requirement already satisfied: tqdm>=4.27 in
c:\users\sdill\anaconda3\lib\sitepackages (from transformers) (4.65.0)
Requirement already satisfied: fsspec>=2023.5.0 in
c:\users\sdill\anaconda3\lib\site-packages (from
huggingfacehub<1.0,>=0.19.3->transformers) (2023.10.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
c:\users\sdill\anaconda3\lib\site-packages (from huggingface-
hub<1.0,>=0.19.3->transformers) (4.9.0)
Requirement already satisfied: colorama in
c:\users\sdill\anaconda3\lib\sitepackages (from tqdm>=4.27-
>transformers) (0.4.6) Requirement already satisfied: charset-
normalizer<4,>=2 in

```
c:\users\sdill\anaconda3\lib\site-packages (from requests-
>transformers) (2.0.4) Requirement already satisfied: idna<4,>=2.5 in
c:\users\sdill\anaconda3\lib\site-packages (from requests-
>transformers) (3.4) Requirement already satisfied: urllib3<3,>=1.21.1
in
c:\users\sdill\anaconda3\lib\site-packages (from requests-
>transformers) (2.0.7) Requirement already satisfied:
certifi>=2017.4.17 in
c:\users\sdill\anaconda3\lib\site-packages (from requests-
>transformers) (2024.2.2)
```

[7]: `!pip install datasets`

```
Collecting datasets
  Downloading datasets-2.18.0-py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: filelock in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (3.13.1)
Requirement already satisfied: numpy>=1.17 in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (1.26.4)
Requirement already satisfied: pyarrow>=12.0.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets)
(14.0.2) Collecting pyarrow-hotfix (from datasets)
  Downloading pyarrow_hotfix-0.6-py3-none-any.whl.metadata (3.6 kB)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.3.7)
Requirement already satisfied: pandas in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (2.1.4)
Requirement already satisfied: requests>=2.19.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (4.65.0)
Collecting xxhash (from datasets)
  Downloading xxhash-3.4.1-cp311-cp311-win_amd64.whl.metadata (12 kB)
Collecting multiprocess (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: fsspec<=2024.2.0,>=2023.1.0 in
c:\users\sdill\anaconda3\lib\site-packages (from
fsspec[http]<=2024.2.0,>=2023.1.0->datasets) (2023.10.0)
Requirement already satisfied: aiohttp in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (3.9.3)
Requirement already satisfied: huggingface-hub>=0.19.4 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.22.2)
Requirement already satisfied: packaging in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (23.1)
Requirement already satisfied: pyyaml>=5.1 in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (6.0.1)
Requirement already satisfied: aiosignal>=1.1.2 in
```

c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets) (1.2.0) Requirement already satisfied: attrs>=17.3.0 in c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets) (23.1.0) Requirement already satisfied: frozenlist>=1.1.1 in c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets) (1.4.0) Requirement already satisfied: multidict<7.0,>=4.5 in c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets) (6.0.4) Requirement already satisfied: yarl<2.0,>=1.0 in c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets) (1.9.3) Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\sdill\anaconda3\lib\site-packages (from huggingfacehub>=0.19.4->datasets) (4.9.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0->datasets) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0->datasets)
(3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0->datasets) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0->datasets)
(2024.2.2)
Requirement already satisfied: colorama in c:\users\sdill\anaconda3\lib\site-packages (from tqdm>=4.62.1->datasets) (0.4.6)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets) (2.8.2) Requirement already satisfied: pytz>=2020.1 in c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets) (2023.3) Requirement already satisfied: six>=1.5 in c:\users\sdill\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.16.0)
Downloading datasets-2.18.0-py3-none-any.whl (510 kB)
   ------------------------------------- 0.0/510.5 kB ? eta -:--
   :----- ------------------------------- 61.4/510.5 kB 1.7
   MB/s eta 0:00:01

```
-------- ----------------------------- 112.6/510.5 kB 1.1 MB/s eta
   0:00:01
-------- ----------------------------- 122.9/510.5 kB 1.2 MB/s eta
   0:00:01
-------- ----------------------------- 122.9/510.5 kB 1.2 MB/s eta
   0:00:01
-------- ----------------------------- 122.9/510.5 kB 1.2 MB/s eta
   0:00:01
-------- ----------------------------- 122.9/510.5 kB 1.2 MB/s eta
   0:00:01
----------------- ----------------- 256.0/510.5 kB 787.7 kB/s eta
   0:00:01
-------------------- -------------- 307.2/510.5 kB 827.2 kB/s eta
   0:00:01
----------------------- ---------- 358.4/510.5 kB 857.5 kB/s eta
   0:00:01
----------------------- ---------- 358.4/510.5 kB 857.5 kB/s eta
   0:00:01
----------------------- ---------- 358.4/510.5 kB 857.5 kB/s eta
   0:00:01
----------------------- ---------- 358.4/510.5 kB 857.5 kB/s eta
   0:00:01
----------------------- ---------- 358.4/510.5 kB 857.5 kB/s eta
   0:00:01
--------------------------------- 501.8/510.5 kB 768.0 kB/s eta
0:00:01
----------------------------------- 510.5/510.5 kB 781.7 kB/s eta
   0:00:00
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
   ------------------------------------- 0.0/143.5 kB ? eta -:--:--
   ------------------------------------- 143.5/143.5 kB 8.3 MB/s eta
   0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
   ------------------------------------- 0.0/116.3 kB ? eta -:--:--
   ------------------------------------- 116.3/116.3 kB ? eta
   0:00:00
Downloading pyarrow_hotfix-0.6-py3-none-any.whl (7.9 kB)
Downloading xxhash-3.4.1-cp311-cp311-win_amd64.whl (29 kB)
Installing collected packages: xxhash, pyarrow-hotfix, dill,
multiprocess, datasets
  Attempting uninstall: dill
    Found existing installation: dill 0.3.7
    Uninstalling dill-0.3.7:
      Successfully uninstalled dill-0.3.7
Successfully installed datasets-2.18.0 dill-0.3.8 multiprocess-0.70.16
pyarrowhotfix-0.6 xxhash-3.4.1
```

```
[8]: !pip install datasets
```

Requirement already satisfied: datasets in
c:\users\sdill\anaconda3\lib\sitepackages (2.18.0)
Requirement already satisfied: filelock in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (3.13.1)
Requirement already satisfied: numpy>=1.17 in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (1.26.4)
Requirement already satisfied: pyarrow>=12.0.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (14.0.2)
Requirement already satisfied: pyarrow-hotfix in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.6)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (2.1.4)
Requirement already satisfied: requests>=2.19.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (4.65.0)
Requirement already satisfied: xxhash in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (3.4.1)
Requirement already satisfied: multiprocess in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2024.2.0,>=2023.1.0 in
c:\users\sdill\anaconda3\lib\site-packages (from
fsspec[http]<=2024.2.0,>=2023.1.0->datasets) (2023.10.0)
Requirement already satisfied: aiohttp in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (3.9.3)
Requirement already satisfied: huggingface-hub>=0.19.4 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets) (0.22.2)
Requirement already satisfied: packaging in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (23.1)
Requirement already satisfied: pyyaml>=5.1 in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets) (6.0.1)
Requirement already satisfied: aiosignal>=1.1.2 in
c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets)
(1.2.0) Requirement already satisfied: attrs>=17.3.0 in
c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets)
(23.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in
c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets)
(1.4.0) Requirement already satisfied: multidict<7.0,>=4.5 in
c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets)
(6.0.4) Requirement already satisfied: yarl<2.0,>=1.0 in

```
c:\users\sdill\anaconda3\lib\site-packages (from aiohttp->datasets)
(1.9.3) Requirement already satisfied: typing-extensions>=3.7.4.3 in
c:\users\sdill\anaconda3\lib\site-packages (from
huggingfacehub>=0.19.4->datasets) (4.9.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>datasets) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>datasets)
(3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>datasets) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>datasets)
(2024.2.2)
Requirement already satisfied: colorama in
c:\users\sdill\anaconda3\lib\site-
packages (from tqdm>=4.62.1->datasets) (0.4.6)
Requirement already satisfied: python-
dateutil>=2.8.2 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets)
(2.8.2) Requirement already satisfied: pytz>=2020.1 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets)
(2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->datasets)
(2023.3) Requirement already satisfied: six>=1.5 in
c:\users\sdill\anaconda3\lib\sitepackages (from python-
dateutil>=2.8.2->pandas->datasets) (1.16.0)
```

[9]: `!pip install rouge-score`

```
Collecting rouge-score
  Downloading rouge_score-0.1.2.tar.gz (17 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Collecting absl-py (from rouge-score)
  Using cached absl_py-2.1.0-py3-none-any.whl.metadata (2.3 kB)
Requirement already satisfied: nltk in
c:\users\sdill\anaconda3\lib\sitepackages (from rouge-score)
(3.8.1)
Requirement already satisfied: numpy in
c:\users\sdill\anaconda3\lib\sitepackages (from rouge-score) (1.26.4)
```

```
Requirement already satisfied: six>=1.14.0 in
c:\users\sdill\anaconda3\lib\sitepackages (from rouge-score) (1.16.0)
Requirement already satisfied: click in
c:\users\sdill\anaconda3\lib\sitepackages (from nltk->rouge-score)
(8.1.7)
Requirement already satisfied: joblib in
c:\users\sdill\anaconda3\lib\sitepackages (from nltk->rouge-score)
(1.2.0) Requirement already satisfied: regex>=2021.8.3 in
c:\users\sdill\anaconda3\lib\site-packages (from nltk->rouge-score)
(2023.10.3) Requirement already satisfied: tqdm in
c:\users\sdill\anaconda3\lib\sitepackages (from nltk->rouge-score)
(4.65.0)
Requirement already satisfied: colorama in
c:\users\sdill\anaconda3\lib\site-
packages (from click->nltk->rouge-score)
(0.4.6) Using cached absl_py-2.1.0-py3-none-
any.whl (133 kB) Building wheels for collected
packages: rouge-score
  Building wheel for rouge-score (setup.py): started
  Building wheel for rouge-score (setup.py): finished with status
  'done'
  Created wheel for rouge-score: filename=rouge_score-0.1.2-py3-none-
any.whl size=24970
sha256=59abb333cf238cdc3f890243e65f6665710810f8a6a7043e9e6b2804779dc1c
3
  Stored in directory:
  c:\users\sdill\appdata\local\pip\cache\wheels\1e\19\43\8a
442dc83660ca25e163e1bd1f89919284ab0d0c1475475148
Successfully built rouge-score
Installing collected packages: absl-py, rouge-score
Successfully installed absl-py-2.1.0 rouge-score-0.1.2
```

[10]: `!pip install py7zr`

```
Collecting py7zr
  Downloading py7zr-0.21.0-py3-none-any.whl.metadata (17 kB)
Collecting texttable (from py7zr)
  Downloading texttable-1.7.0-py2.py3-none-any.whl.metadata (9.8 kB)
Collecting pycryptodomex>=3.16.0 (from py7zr)
  Downloading pycryptodomex-3.20.0-cp35-abi3-win_amd64.whl.metadata
  (3.4 kB)
Collecting pyzstd>=0.15.9 (from py7zr)
  Downloading pyzstd-0.15.10-cp311-cp311-win_amd64.whl.metadata (8.1
  kB)
Collecting pyppmd<1.2.0,>=1.1.0 (from py7zr)
  Downloading pyppmd-1.1.0-cp311-cp311-win_amd64.whl.metadata (5.8 kB)
Collecting pybcj<1.1.0,>=1.0.0 (from py7zr)
```

```
  Downloading pybcj-1.0.2-cp311-cp311-win_amd64.whl.metadata (4.1 kB)
Collecting multivolumefile>=0.2.3 (from py7zr)
  Downloading multivolumefile-0.2.3-py3-none-any.whl.metadata (6.3 kB)
Collecting inflate64<1.1.0,>=1.0.0 (from py7zr)
  Downloading inflate64-1.0.0-cp311-cp311-win_amd64.whl.metadata (4.1
  kB)
Collecting brotli>=1.1.0 (from py7zr)
  Downloading Brotli-1.1.0-cp311-cp311-win_amd64.whl.metadata (5.6 kB)
Requirement already satisfied: psutil in
c:\users\sdill\anaconda3\lib\sitepackages (from py7zr) (5.9.0)
Downloading py7zr-0.21.0-py3-none-any.whl (67 kB)
   ------------------------------------- 0.0/67.6 kB ? eta -:--:----
   --------------------- -------------- 41.0/67.6 kB 991.0 kB/s eta
   0:00:01
   ------------------------------------- 67.6/67.6 kB 924.4 kB/s eta
0:00:00 Downloading Brotli-1.1.0-cp311-cp311-win_amd64.whl (357 kB) --
------------------------------------- 0.0/357.3 kB ? eta -:--:------
--- -------------------------- 81.9/357.3 kB 2.3 MB/s eta 0:00:01
   ------------ ------------------------- 112.6/357.3 kB 1.6 MB/s eta
   0:00:01
   --------------- --------------------- 143.4/357.3 kB 1.2 MB/s eta
   0:00:01
   --------------- --------------------- 143.4/357.3 kB 1.2 MB/s eta
   0:00:01
   --------------- --------------------- 143.4/357.3 kB 1.2 MB/s eta
   0:00:01
   --------------- --------------------- 143.4/357.3 kB 1.2 MB/s eta
   0:00:01
   --------------- --------------------- 143.4/357.3 kB 1.2 MB/s eta
   0:00:01
   ------------------------- ----------- 245.8/357.3 kB 684.7 kB/s eta
   0:00:01
   --------------------------- ------- 286.7/357.3 kB 707.1 kB/s eta
   0:00:01
   ---------------------------- ---- 317.4/357.3 kB 728.0 kB/s eta
   0:00:01
   ---------------------------- ---- 317.4/357.3 kB 728.0 kB/s eta
   0:00:01
   ---------------------------- ---- 317.4/357.3 kB 728.0 kB/s eta
   0:00:01
   ---------------------------- ---- 317.4/357.3 kB 728.0 kB/s eta
   0:00:01
   ------------------------------------ 357.3/357.3 kB 529.0 kB/s eta
   0:00:00
Downloading inflate64-1.0.0-cp311-cp311-win_amd64.whl (35 kB)
Downloading multivolumefile-0.2.3-py3-none-any.whl (17 kB)
```

```
Downloading pybcj-1.0.2-cp311-cp311-win_amd64.whl (24 kB)
Downloading pycryptodomex-3.20.0-cp35-abi3-win_amd64.whl (1.8 MB)
   ------------------------------------- 0.0/1.8 MB ? eta -:--
   :--- ------------------------------------- 0.1/1.8 MB 2.6 MB/s
   eta 0:00:01
   --- ------------------------------- 0.1/1.8 MB 1.7 MB/s eta
   0:00:01
   --- ------------------------------- 0.2/1.8 MB 1.5 MB/s eta
   0:00:02
   --- ------------------------------- 0.2/1.8 MB 1.5 MB/s eta
   0:00:02
   --- ------------------------------- 0.2/1.8 MB 1.5 MB/s eta
   0:00:02
   --- ------------------------------- 0.2/1.8 MB 1.5 MB/s eta
   0:00:02
   ---- ------------------------------- 0.2/1.8 MB 562.0 kB/s eta
   0:00:03
   ------ ------------------------------- 0.3/1.8 MB 863.3 kB/s eta
   0:00:02
   -------- ------------------------------- 0.4/1.8 MB 857.5 kB/s eta
   0:00:02
   -------- ------------------------------- 0.4/1.8 MB 867.0 kB/s eta
   0:00:02
   -------- ------------------------------- 0.4/1.8 MB 867.0 kB/s eta
   0:00:02
   -------- ------------------------------- 0.4/1.8 MB 867.0 kB/s eta
   0:00:02
   -------- ------------------------------- 0.4/1.8 MB 867.0 kB/s eta
   0:00:02
   ----------- ------------------------- 0.5/1.8 MB 766.8 kB/s eta
   0:00:02
   ------------ ------------------------- 0.5/1.8 MB 777.1 kB/s eta
   0:00:02
   ------------- ------------------------- 0.6/1.8 MB 823.1 kB/s eta
   0:00:02
   ------------- ------------------------- 0.6/1.8 MB 823.1 kB/s eta
   0:00:02
   ------------- ------------------------- 0.6/1.8 MB 823.1 kB/s eta
   0:00:02
   ------------- ------------------------- 0.6/1.8 MB 823.1 kB/s eta
   0:00:02
   ------------- ------------------------- 0.6/1.8 MB 823.1 kB/s eta
   0:00:02
   --------------- --------------------- 0.7/1.8 MB 760.9 kB/s eta
   0:00:02
```

```
----------------- -------------------- 0.8/1.8 MB 766.0 kB/s eta
0:00:02
----------------- ------------------ 0.9/1.8 MB 800.3 kB/s eta
0:00:02
----------------- ------------------ 0.9/1.8 MB 797.9 kB/s eta
0:00:02
----------------- ------------------ 0.9/1.8 MB 797.9 kB/s eta
0:00:02
----------------- ------------------ 0.9/1.8 MB 797.9 kB/s eta
0:00:02
----------------- ------------------ 0.9/1.8 MB 797.9 kB/s eta
0:00:02
-------------------- --------------- 1.0/1.8 MB 791.4 kB/s eta
0:00:01 --------------------- -------------- 1.1/1.8 MB 800.0
kB/s eta 0:00:01
----------------------- ------------- 1.2/1.8 MB 831.9 kB/s eta
0:00:01
----------------------- ------------- 1.2/1.8 MB 831.9 kB/s eta
0:00:01
----------------------- ------------- 1.2/1.8 MB 831.9 kB/s eta
0:00:01
----------------------- ------------- 1.2/1.8 MB 831.9 kB/s eta
0:00:01
-------------------------- --------- 1.3/1.8 MB 812.9 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
-------------------------- --------- 1.4/1.8 MB 823.7 kB/s eta
0:00:01
------------------------------- ---- 1.6/1.8 MB 797.0 kB/s eta
0:00:01
-------------------------------- --- 1.6/1.8 MB 803.8 kB/s eta
0:00:01
--------------------------------- -- 1.6/1.8 MB 812.7 kB/s eta
0:00:01
```

```
                 ------------------------------------ -- 1.6/1.8 MB 812.7 kB/s eta
    0:00:01
                 ------------------------------------ -- 1.6/1.8 MB 812.7 kB/s eta
    0:00:01
                 ------------------------------------ -- 1.6/1.8 MB 812.7 kB/s eta
    0:00:01
                 ------------------------------------ -- 1.6/1.8 MB 812.7 kB/s eta
    0:00:01
                 ------------------------------------ 1.8/1.8 MB 779.0 kB/s eta
    0:00:00
    Downloading pyppmd-1.1.0-cp311-cp311-win_amd64.whl (46 kB)
                 ------------------------------------ 0.0/46.1 kB ? eta -:--:--
                 ------------------------------------ 46.1/46.1 kB 1.2 MB/s eta
    0:00:00
    Downloading pyzstd-0.15.10-cp311-cp311-win_amd64.whl (245 kB)
                 ------------------------------------ 0.0/245.5 kB ? eta -:--:---
                 ------------------------------- --- 225.3/245.5 kB 4.6 MB/s eta
    0:00:01
                 ------------------------------------ 245.5/245.5 kB 3.7 MB/s eta
    0:00:00 Downloading texttable-1.7.0-py2.py3-none-any.whl (10 kB)
    Installing collected packages: texttable, brotli, pyzstd, pyppmd,
    pycryptodomex, pybcj, multivolumefile, inflate64, py7zr
      Attempting uninstall: brotli
        Found existing installation: Brotli 1.0.9
        Uninstalling Brotli-1.0.9:
          Successfully uninstalled Brotli-1.0.9
    Successfully installed brotli-1.1.0 inflate64-1.0.0 multivolumefile-
    0.2.3 py7zr-0.21.0 pybcj-1.0.2 pycryptodomex-3.20.0 pyppmd-1.1.0
    pyzstd-0.15.10
    texttable-1.7.0
```

```python
[11]: import pandas as pd
      import numpy as np
      from datasets import Dataset, load_metric
      import shutil

      # Data Visualization
      import plotly.express as px
      import plotly.graph_objs as go
```

```python
import plotly.subplots as sp
from plotly.subplots import make_subplots
import plotly.figure_factory as ff
import plotly.io as pio
from IPython.display import display
from plotly.offline import init_notebook_mode
init_notebook_mode(connected=True)

# Statistics & Mathematics
import scipy.stats as stats
import statsmodels.api as sm
from scipy.stats import shapiro, skew, anderson, kstest, gaussian_kde,spearmanr
import math

# Hiding warnings
import warnings
warnings.filterwarnings("ignore")
```

[25]: `!pip install evaluate`

```
Collecting evaluate
  Using cached evaluate-0.4.1-py3-none-any.whl.metadata (9.4 kB)
Requirement already satisfied: datasets>=2.0.0 in
c:\users\sdill\anaconda3\lib\site-packages (from evaluate) (2.18.0)
Requirement already satisfied: numpy>=1.17 in
c:\users\sdill\anaconda3\lib\sitepackages (from evaluate) (1.26.4)
Requirement already satisfied: dill in
c:\users\sdill\anaconda3\lib\sitepackages (from evaluate) (0.3.8)
Requirement already satisfied: pandas in
c:\users\sdill\anaconda3\lib\sitepackages (from evaluate) (2.1.4)
Requirement already satisfied: requests>=2.19.0 in
c:\users\sdill\anaconda3\lib\site-packages (from evaluate) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in
c:\users\sdill\anaconda3\lib\site-packages (from evaluate) (4.65.0)
Requirement already satisfied: xxhash in
c:\users\sdill\anaconda3\lib\sitepackages (from evaluate) (3.4.1)
Requirement already satisfied: multiprocess in
c:\users\sdill\anaconda3\lib\site-packages (from evaluate) (0.70.16)
Requirement already satisfied: fsspec>=2021.05.0
in c:\users\sdill\anaconda3\lib\site-packages
(from fsspec[http]>=2021.05.0->evaluate)
(2023.10.0) Requirement already satisfied:
huggingface-hub>=0.7.0 in
c:\users\sdill\anaconda3\lib\site-packages (from evaluate) (0.22.2)
```

Requirement already satisfied: packaging in
c:\users\sdill\anaconda3\lib\sitepackages (from evaluate) (23.1)
Collecting responses<0.19 (from evaluate)
  Using cached responses-0.18.0-py3-none-any.whl.metadata (29 kB)
Requirement already satisfied: filelock in
c:\users\sdill\anaconda3\lib\site-
packages (from datasets>=2.0.0->evaluate) (3.13.1)
Requirement already satisfied: pyarrow>=12.0.0 in
c:\users\sdill\anaconda3\lib\site-packages (from datasets>=2.0.0-
>evaluate) (14.0.2)
Requirement already satisfied: pyarrow-hotfix in
c:\users\sdill\anaconda3\lib\site-packages (from datasets>=2.0.0-
>evaluate)
(0.6)
Requirement already satisfied: aiohttp in
c:\users\sdill\anaconda3\lib\sitepackages (from datasets>=2.0.0-
>evaluate) (3.9.3)
Requirement already satisfied: pyyaml>=5.1 in
c:\users\sdill\anaconda3\lib\site-
packages (from datasets>=2.0.0->evaluate) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
c:\users\sdill\anaconda3\lib\site-packages (from
huggingfacehub>=0.7.0->evaluate) (4.9.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>evaluate) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>evaluate)
(3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>evaluate) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\sdill\anaconda3\lib\site-packages (from requests>=2.19.0-
>evaluate)
(2024.2.2)
Requirement already satisfied: colorama in
c:\users\sdill\anaconda3\lib\site-
packages (from tqdm>=4.62.1->evaluate) (0.4.6)
Requirement already satisfied: python-
dateutil>=2.8.2 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->evaluate)
(2.8.2) Requirement already satisfied: pytz>=2020.1 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->evaluate)
(2023.3.post1)

Requirement already satisfied: tzdata>=2022.1 in
c:\users\sdill\anaconda3\lib\site-packages (from pandas->evaluate)
(2023.3)
Requirement already satisfied:
aiosignal>=1.1.2 in
c:\users\sdill\anaconda3\lib\site-packages
(from aiohttp->datasets>=2.0.0->evaluate)
(1.2.0) Requirement already satisfied:
attrs>=17.3.0 in
c:\users\sdill\anaconda3\lib\site-packages
(from aiohttp->datasets>=2.0.0->evaluate)
(23.1.0) Requirement already satisfied:
frozenlist>=1.1.1 in
c:\users\sdill\anaconda3\lib\site-packages
(from aiohttp->datasets>=2.0.0->evaluate)
(1.4.0) Requirement already satisfied:
multidict<7.0,>=4.5 in
c:\users\sdill\anaconda3\lib\site-packages
(from aiohttp->datasets>=2.0.0->evaluate)
(6.0.4) Requirement already satisfied:
yarl<2.0,>=1.0 in
c:\users\sdill\anaconda3\lib\site-packages
(from aiohttp->datasets>=2.0.0->evaluate)
(1.9.3)
Requirement already satisfied: six>=1.5 in
c:\users\sdill\anaconda3\lib\site-
packages (from python-dateutil>=2.8.2->pandas->evaluate) (1.16.0)
Using cached evaluate-0.4.1-py3-none-any.whl (84 kB)
Using cached responses-0.18.0-py3-none-any.whl (38 kB)
Installing collected packages: responses, evaluate
Successfully installed evaluate-0.4.1 responses-0.18.0

```
[27]: !pip install textblob
```

Collecting textblob
  Downloading textblob-0.18.0.post0-py3-none-any.whl.metadata (4.5 kB)
Requirement already satisfied: nltk>=3.8 in
c:\users\sdill\anaconda3\lib\sitepackages (from textblob) (3.8.1)
Requirement already satisfied: click in
c:\users\sdill\anaconda3\lib\sitepackages (from nltk>=3.8->textblob)
(8.1.7)
Requirement already satisfied: joblib in
c:\users\sdill\anaconda3\lib\site-
packages (from nltk>=3.8->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3
in
c:\users\sdill\anaconda3\lib\site-packages (from nltk>=3.8->textblob)

(2023.10.3)
Requirement already satisfied: tqdm in
c:\users\sdill\anaconda3\lib\sitepackages (from nltk>=3.8->textblob)
(4.65.0)
Requirement already satisfied: colorama in
c:\users\sdill\anaconda3\lib\site-
packages (from click->nltk>=3.8->textblob) (0.4.6)
Downloading textblob-0.18.0.post0-py3-none-any.whl
(626 kB)
   ------------------------------------ 0.0/626.3 kB ? eta -:--
   :--- --------------------------------- 41.0/626.3 kB 653.6
   kB/s eta 0:00:01
   ----- ------------------------------- 81.9/626.3 kB 762.6 kB/s eta
   0:00:01
   ----- ------------------------------- 92.2/626.3 kB 751.6 kB/s eta
   0:00:01
   ----- ------------------------------- 92.2/626.3 kB 751.6 kB/s eta
   0:00:01
   ----- ------------------------------- 92.2/626.3 kB 751.6 kB/s eta
   0:00:01
   ----- ------------------------------- 92.2/626.3 kB 751.6 kB/s eta
   0:00:01
   ----------- ------------------------- 194.6/626.3 kB 562.0 kB/s eta
   0:00:01
   ------------- ----------------------- 225.3/626.3 kB 599.0 kB/s eta
   0:00:01
   ----------------- ------------------- 286.7/626.3 kB 655.2 kB/s eta
   0:00:01
   ------------------ ------------------ 297.0/626.3 kB 656.1 kB/s eta
   0:00:01
   ------------------ ------------------ 297.0/626.3 kB 656.1 kB/s eta
   0:00:01
   ------------------ ------------------ 297.0/626.3 kB 656.1 kB/s eta
   0:00:01
   ------------------ ------------------ 297.0/626.3 kB 656.1 kB/s eta
   0:00:01
   ----------------------- ------------- 409.6/626.3 kB 623.0 kB/s eta
   0:00:01
   ------------------------- ---------- 450.6/626.3 kB 640.9 kB/s eta
   0:00:01
   ------------------------- ---------- 460.8/626.3 kB 640.9 kB/s eta
   0:00:01
   ------------------------- ---------- 460.8/626.3 kB 640.9 kB/s eta
   0:00:01
   ------------------------- ---------- 460.8/626.3 kB 640.9 kB/s eta
   0:00:01

```
                      -------------------------- ---------- 460.8/626.3 kB 640.9 kB/s eta
        0:00:01 -------------------------- ---------- 460.8/626.3 kB 640.9
        kB/s eta 0:00:01
      ------------------------------- --- 573.4/626.3 kB 581.2 kB/s eta
                                              0:00:01
      ------------------------------------ 626.3/626.3 kB 606.4 kB/s eta
                                              0:00:00
    Installing collected packages: textblob
    Successfully installed textblob-0.18.0.post0
```

[28]:
```python
# Transformers
from transformers import BartTokenizer, BartForConditionalGeneration      #□
  □BERT Tokenizer and architecture
from transformers import Seq2SeqTrainer, Seq2SeqTrainingArguments          #□
  □These will help us to fine-tune our model
from transformers import pipeline                                          #□
  □Pipeline
from transformers import DataCollatorForSeq2Seq                           #□
  □DataCollator to batch the data
import torch                                                               #□
  □PyTorch
import evaluate                                                            #□
  □Hugging Face's library for model evaluation


# Other NLP libraries
from textblob import TextBlob                                             #□
  □This is going to help us fix spelling mistakes in texts
from sklearn.feature_extraction.text import TfidfVectorizer               #□
  □This is going to helps identify the most common terms in the corpus
import re                                                                  #□
  □This library allows us to clean text data
import nltk                                                                #□
  □Natural Language Toolkit
nltk.download('punkt')                                                     #□
  □This divides a text into a list of sentences
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\sdill\AppData\Roaming\nltk_data…
[nltk_data] Unzipping tokenizers\punkt.zip.
```

[28]: True

17

```python
[33]: pd.set_option('display.max_colwidth', 1000)
```

```python
[34]: seed = 42
      #paper_color =
      #bg_color =
      colormap = 'cividis'
      template = 'plotly_dark'
```

```python
[35]: if torch.cuda.is_available():
          print("GPU is available. \nUsing GPU")
          device = torch.device('cuda')
      else:
          print("GPU is not available. \nUsing CPU")
          device = torch.device('cpu')
```

```
GPU is not available.
Using CPU
```

```python
[36]: def display_feature_list(features, feature_type):

          '''
          This function displays the features within each list for each type
          of data
          '''

          print(f"\n{feature_type} Features: ")
          print(', '.join(features) if features else
          'None')

      def describe_df(df):
          """
          This function prints some basic info on the dataset
          and sets global variables for feature lists.
          """

          global categorical_features, continuous_features, binary_features
          categorical_features = [col for col in df.columns if df[col].dtype
          ==□
       □'object'] binary_features = [col for col in df.columns if
          df[col].nunique() <= 2 and□
       □df[col].dtype != 'object'] continuous_features = [col for col
       in df.columns if df[col].dtype !=□ □'object' and col not in
       binary_features]

          print(f"\n{type(df).__name__} shape: {df.shape}")
          print(f"\n{df.shape[0]:,.0f} samples")
          print(f"\n{df.shape[1]:,.0f} attributes")
```

```python
    print(f'\nMissing Data: \n{df.isnull().sum()}')
    print(f'\nDuplicates: {df.duplicated().sum()}')
    print(f'\nData Types: \n{df.dtypes}')

    #negative_valued_features = [col for col in df.columns if (df[col]
    < 0).
any()]
    #print(f'\nFeatures with Negative Values: {", ".
join(negative_valued_features) if negative_valued_features else
"None"}')

    display_feature_list(categorical_features, 'Categorical')
    display_feature_list(continuous_features, 'Continuous')
```

```
        display_feature_list(binary_features, 'Binary')

        print(f'\n{type(df).__name__} Head: \n')
        display(df.head(5))
        print(f'\n{type(df).__name__} Tail: \n')
        display(df.tail(5))
```

[37]:
```python
def histogram_boxplot(df,hist_color, box_color, height, width, legend, name):
    '''
    This function plots a Histogram and a Box Plot side by side

    Parameters:
    hist_color = The color of the histogram
    box_color = The color of the boxplots
    heigh and width = Image size
    legend = Either to display legend or not
    '''

    features = df.select_dtypes(include = [np.number]).columns.tolist()

    for feat in features:
        try:
            fig = make_subplots(
                rows=1,
                cols=2,
                subplot_titles=["Box Plot", "Histogram"],
                horizontal_spacing=0.2
            )

            density = gaussian_kde(df[feat])
            x_vals = np.linspace(min(df[feat]), max(df[feat]), 200)
            density_vals = density(x_vals)

            fig.add_trace(go.Scatter(x=x_vals, y = density_vals, mode = 'lines',
                                    fill = 'tozeroy', name="Density",
    ↪line_color=hist_color), row=1, col=2)
            fig.add_trace(go.Box(y=df[feat], name="Box Plot", boxmean=True,
    ↪line_color=box_color), row=1, col=1)

            fig.update_layout(title={'text': f'<b>{name} Word
    ↪Count<br><sup><i>    {feat}</i></sup></b>',
                                    'x': .025, 'xanchor': 'left'},
                            margin=dict(t=100),
                            showlegend=legend,
                            template = template,
                            #plot_bgcolor=bg_color,paper_bgcolor=paper_color,
                            height=height, width=width
```

```
                          )
            fig.update_yaxes(title_text=f"<b>Words</b>", row=1, col=1,␣
        ↪showgrid=False)
            fig.update_xaxes(title_text="", row=1, col=1, showgrid=False)

            fig.update_yaxes(title_text="<b>Frequency</b>", row=1,␣
        ↪col=2,showgrid=False)
            fig.update_xaxes(title_text=f"<b>Words</b>", row=1, col=2,␣
        ↪showgrid=False)

            fig.show()
            print('\n')
        except Exception as e:
            print(f"An error occurred: {e}")
```

```
[38]: def plot_correlation(df, title, subtitle, height, width, font_size):
          '''
          This function is resposible to plot a correlation map among features in the␣
      ↪dataset.

          Parameters:
          height = Define height
          width = Define width
          font_size = Define the font size for the annotations
          '''
          corr = np.round(df.corr(numeric_only = True), 2)
          mask = np.triu(np.ones_like(corr, dtype = bool))
          c_mask = np.where(~mask, corr, 100)

          c = []
          for i in c_mask.tolist()[1:]:
              c.append([x for x in i if x != 100])


          fig = ff.create_annotated_heatmap(z=c[::-1],
                                            x=corr.index.tolist()[:-1],
                                            y=corr.columns.tolist()[1:][::-1],
                                            colorscale = colormap)

          fig.update_layout(title = {'text': f"<b>{title} Heatmap<br><sup>  
      ↪  <i>{subtitle}</i></sup></b>",
                                     'x': .025, 'xanchor': 'left', 'y': .95},
                            margin = dict(t=210, l = 110),
                            yaxis = dict(autorange = 'reversed', showgrid = False),
                            xaxis = dict(showgrid = False),
```

```
                       template = template,
                       #plot_bgcolor=bg_color,paper_bgcolor=paper_color,
                       height = height, width = width)



        fig.add_trace(go.Heatmap(z = c[::-1],
                                 colorscale = colormap,
                                 showscale = True,
                                 visible = False))
        fig.data[1].visible = True

        for i in range(len(fig.layout.annotations)):
            fig.layout.annotations[i].font.size = font_size

        fig.show()
```

[39]:
```python
def compute_tfidf(df_column, ngram_range=(1,1), max_features=15):
    vectorizer = TfidfVectorizer(max_features=max_features,⎵
  ⎵stop_words='english', ngram_range=ngram_range)
    x = vectorizer.fit_transform(df_column.fillna(''))
    df_tfidfvect = pd.DataFrame(x.toarray(), columns=vectorizer.
  ⎵get_feature_names_out())
    return df_tfidfvect
```

[46]:
```python
train = pd.read_csv(r'C:\Users\sdill\Downloads\samsum-train.csv.zip')
```

[48]:
```python
mask = train['dialogue'].isnull() # Creating mask with null dialogues
filtered_train = train[mask] # filtering dataframe
filtered_train # Visualizing
```

[48]:
```
              id dialogue                                    summary
6054 13828807      NaNproblem with visualization of the content
```

[47]:
```python
describe_df(train)
```

```
DataFrame shape: (14732, 3)

14,732 samples

3 attributes

Missing Data:
id      0
dialogue    1
summary     0
dtype: int64
```

```
Duplicates: 0

Data Types:
id     object dialogue
 object summary
 object
dtype: object

Categorical
Features: id,
dialogue, summary

Continuous Features:
None

Binary Features:
None

DataFrame Head:


        id \
0  13818513
1  13728867
2  13681000
3  13730747
4  13728094
```

dialogue \
0

Amanda: I baked cookies. Do you want some?\r\nJerry: Sure!\r\nAmanda: I'll bring you tomorrow :-)

1

Olivia: Who are you voting for in this election? \r\nOliver: Liberals as always.\r\nOlivia: Me too!!\r\nOliver: Great

2

Tim: Hi, what's up?\r\nKim: Bad mood tbh, I was going to do lots of stuff but ended up procrastinating\r\nTim: What did you plan on doing?\r\nKim: Oh you know, uni stuff and unfucking my room\r\nKim: Maybe tomorrow I'll move my ass and do everything\r\nKim: We were going to defrost a fridge so instead of shopping I'll eat some defrosted veggies\r\nTim: For doing stuff I recommend Pomodoro technique where u use breaks for doing chores\r\nTim: It really helps\r\nKim: thanks, maybe I'll do that\r\nTim: I also like using post-its in kaban style

3

Edward: Rachel, I think I'm in ove with Bella..\r\nrachel: Dont say anything

else..\r\nEdward: What do you mean??\r\nrachel: Open your fu**ing door.. I'm outside

4 Sam: heyoverheard rick say something\r\nSam: i don't know what to do :-/

\r\nNaomi: what did he say??\r\nSam: he was talking on the phone with

someone\r\nSam: i don't know who\r\nSam: and he was telling them that he

wasn't very happy here\r\nNaomi: damn!!!\r\nSam: he was saying he doesn't like

being my roommate\r\nNaomi: wow, how do you feel about it?\r\nSam: i thought i

was a good rommate\r\nSam: and that we have a nice place\r\nNaomi: that's true man!!!\r\nNaomi: i used to love living with you before i moved in with me boyfriend\r\nNaomi: i don't know why he's saying that\r\nSam: what should i do?

??\r\nNaomi: honestly if it's bothering you that much you should talk to

him\r\nNaomi: see what's going on\r\nSam: i don't want to get in any kind of

confrontation though\r\nSam: maybe i'll just let it go\r\nSam: and see how it

goes in the future\r\nNaomi: it's your choice sam\r\nNaomi: if i were you i

would just talk to him and clear the air

	summary
0	Amanda baked cookies and will bring Jerry some tomorrow.
1	Olivia and Olivier are voting for liberals in this election.
2	Kim

may □

☐ try the pomodoro technique recommended by Tim to get more stuff done.

3

Edward thinks she is in love □

☐ with Bella. Rachel wants Edward to open his door. Rachel is outside.

4

Sam is confused, be

causeheoverheardRickcomplainingabouthimasa

roommate. Naomi thinks Sam should talk to Rick. Sam is not sure what to do.

```
DataFrame Tail:

              id  \
14727  13863028
14728  13828570
14729  13819050
14730  13828395
14731  13729017
```

```
                                                                    dialogue  \
14727

                                                                    Romeo:
       You are on my 'People you may know' list.\nGreta: Ah, maybe it is
       because of
       the changed number of somebody's?\nGreta: I don't know you?\nRomeo:
       This might be the beginning of a beautiful relationship\nRomeo: How
       about adding me on your friend list and talk a bit?\nGreta:
       No.\nRomeo: Okay I see.
14728  Theresa: <file_photo>\r\nTheresa: <file_photo>\r\nTheresa: Hey
Louise,
       how are u?\r\nTheresa: This is my workplace, they always give us so
       much food
       here \r\nTheresa: Luckily they also offer us yoga classes, so all
       the food
```

isn't much of a problem \r\nLouise: Hey!! \r\nLouise: Wow, that's awesome,

seems great Haha\r\nLouise: I'm good! Are you coming to visit Stockholm this

summer? \r\nTheresa: I don't think so :/ I need to prepare for Uni.. I will

probably attend a few lessons this winter\r\nLouise: Nice! Do you already know

which classes you will attend?\r\nTheresa: Yes, it will be psychology :) I

want to complete a few modules that I missed :)\r\nLouise: Very good! Is it at

the Uni in Prague?\r\nTheresa: No, it will be in my home town :)\r\nLouise: I

have so much work right now, but I will continue to work until the end of

summer, then I'm also back to Uni, on the 26th September!\r\nTheresa: You must

send me some pictures, so I can see where you live :) \r\nLouise: I will,…

14729
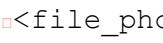
John: Every day some

bad news. Japan will hunt whales again\r\nErica: Yes, I've read this. It's

very upsetting\r\nJohn: Cruel Japanese\r\nFaith: I think this is a racist

remark. Because Island and Norways has never joined this international whaling

agreement\r\nErica: really? I haven't known, everybody is so outraged by

Japan\r\nFaith: sure, European hypocrisy \r\nJohn: not entirely. Scandinavians

don't use the nets that Japanese use, so Norway and Island kill much less

specimens that Japan will\r\nFaith: oh, it's much more complex than one may

expect\r\nJohn: True, but the truth is, that all of them should stop\r\nJohn:

and this decision is a step back\r\nFaith: yes, this is worrying\r\nErica: And

it seems that the most important whaling countries are out of the agreement

right now\r\nFaith: yes, seems so\r\nJohn: Just like USA leaving the Paris Agreement

14730   Jennifer: Dear Celia! How are you doing?\r\nJennifer: The afternoon with

the Collins was very pleasant, nice folks, but we missed you.\r\nJennifer: But
I appreciate your consideration for Peter.\r\nCelia: My dear Jenny! It turns
out that my decision not to come, though I wanted so much to see you again and
Peter and the Collins, was right. Yesterday it all developed into a full bore
cold. Sh…\r\nCelia: All symptoms like in a text book.\r\nCelia: Luckily it's
contagious only on the first 2, 3 days, so when we meet next week it should be
alright.\r\nCelia: Thanks for asking! Somehow for all of us Peter comes first
now.\r\nJennifer: That's too bad. Poor you…\r\nJennifer: I'll be driving to FR, do you want me to bring you sth? It's on my way.\r\nCelia: Thank you dear! I was at the pharmacy yesterday and had done my shopping the day before.
\r\nCelia: You'd better still stay away from me in case I'm still
contagious\r\nJennifer: Right. So I'll only leave a basket on your terrace. A…

14731 Georgia: are you ready for hotel hunting? We need to book something
finally for Lisbon\r\nJuliette: sure we can go on, show me what you
found\r\nGeorgia: <file_photo>\r\nJuliette: nah… it looks like an old lady's
room lol\r\nGeorgia: <file_photo>\r\nJuliette: that's better… but the bed
doesn't look very comfortable\r\nGeorgia: i kind of like it and it's really
close to the city center\r\nJuliette: show me the others please\r\nGeorgia:
<file_photo>\r\nJuliette: nah… this one sucks too, look at those horrible
curtains \r\nGeorgia: aff Julie you are such a princess\r\nJuliette: i just
want to be comfortable\r\nGeorgia: come on, stop whining you know we are on a budget\r\nJuliette: well hopefully we can find something that's decent right?
\r\nGeorgia: i did show you decent but you want a Marriott or something :/
\r\nJuliette: ok ok don't get angry\r\nGeorgia: we need to decide today, the
longer we wait the higher the prices get \r\nJuliette: ok how about we get the second one then?…

summary

14727

　Romeo is trying to get Greta to add him to her friend list but she refuses.

14728

yoga classes. Theresa won't go to visit Louise in Stockholm, because she will prepare for university psychology lessons. She'll be back at uni on 26th September.

14729

　　　　　Japan is going to hunt whales again. Island and Norway never stopped hunting them. The Scandinavians kill fewer whales than the Japanese.

14730

Jennifer as she is ill. She's working, but doesn't want to meet with Jennifer as it might be contagious. Jennifer will leave a basket with cookies on Celia's terrace.

14731

Georgia's choices. Juliette and Georgia decide on the second option presented by Georgia, but it has already been booked. Finally Georgia books the third hotel.

```
[49]: mask = train['dialogue'].isnull() # Creating mask with null
      dialogues filtered_train = train[mask] # filtering
      dataframe filtered_train # Visualizing
```

[49]:　　　　　id dialogue　　　　　　　　　　　summary

```
        6054  13828807      NaN  problem with visualization of the content
```

```
[52]: train = train.dropna()
```

```
[53]: categorical_features.remove('id')
```

```
[50]: df_text_lenght = pd.DataFrame() # Creating an empty dataframe
      for feat in categorical_features: # Iterating through features --> Dialogue & ⏎
        ↳Summary
          df_text_lenght[feat] = train[feat].apply(lambda x: len(str(x).split())) # ⏎
        ↳Counting words for each feature

      # Plotting histogram-boxplot
      histogram_boxplot(df_text_lenght,'#89c2e0', '#d500ff', 600, 1000, True, 'Train⏎
        ↳Dataset')
```

An error occurred: The data appears to lie in a lower-dimensional subspace of the space in which it is expressed. This has resulted in a singular data covariance matrix, which cannot be treated using the algorithms implemented in
`gaussian_kde`. Consider performing principle component analysis / dimensionality reduction and using `gaussian_kde` with the transformed data.
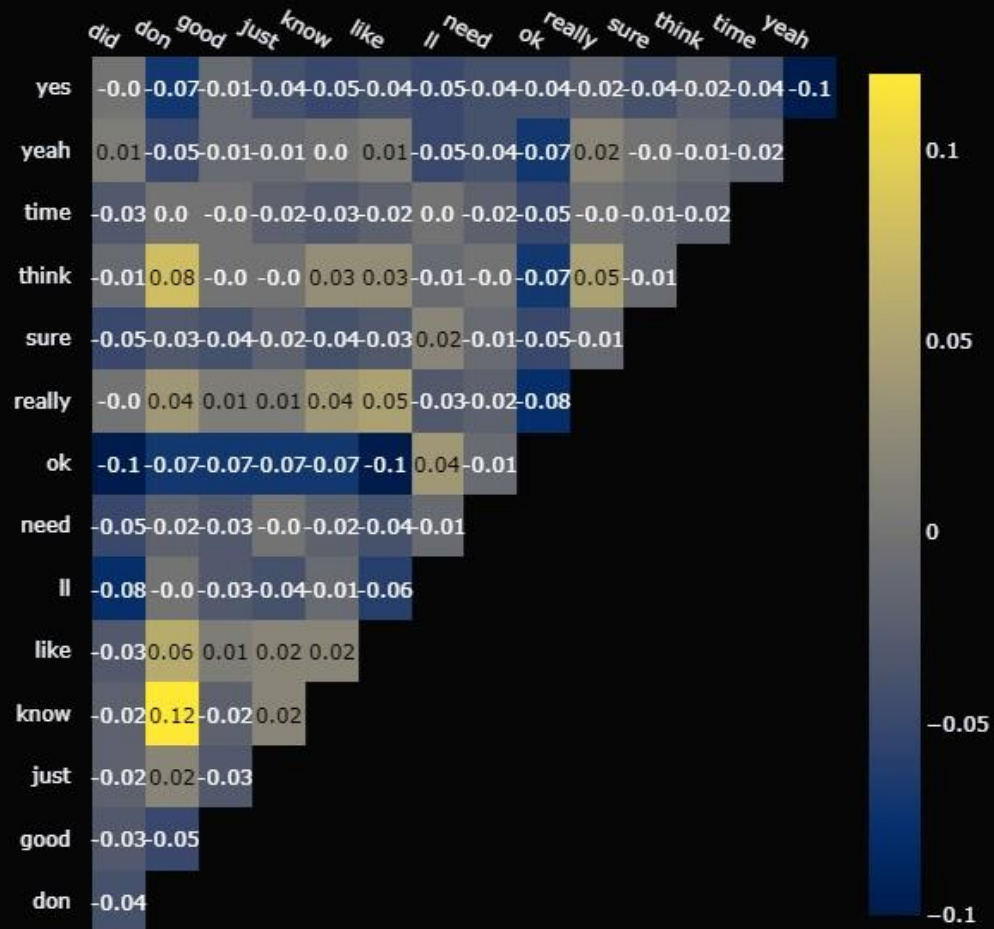
Train Dataset Word Count
dialogue

Box Plot — Histogram

Train Dataset Word Count
summary
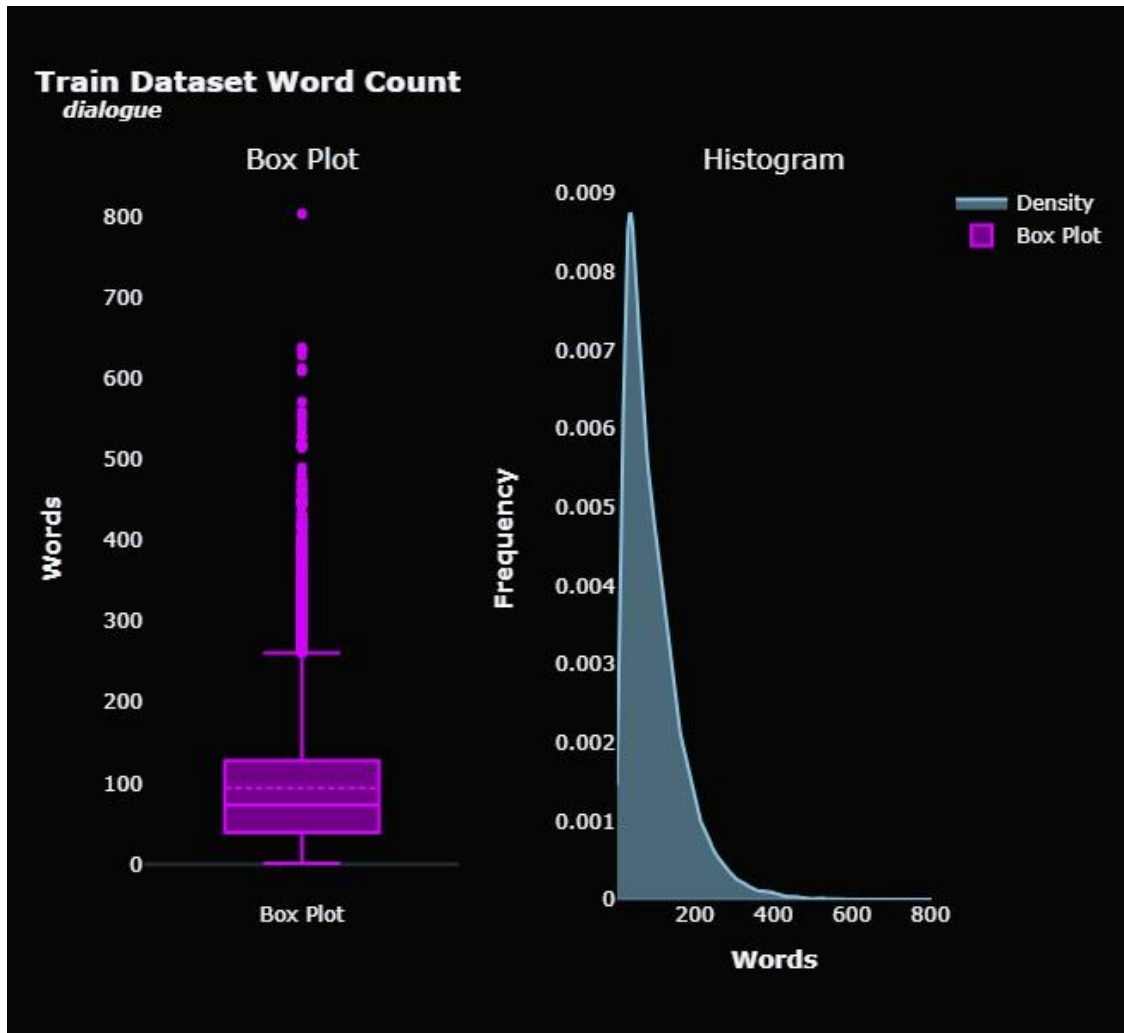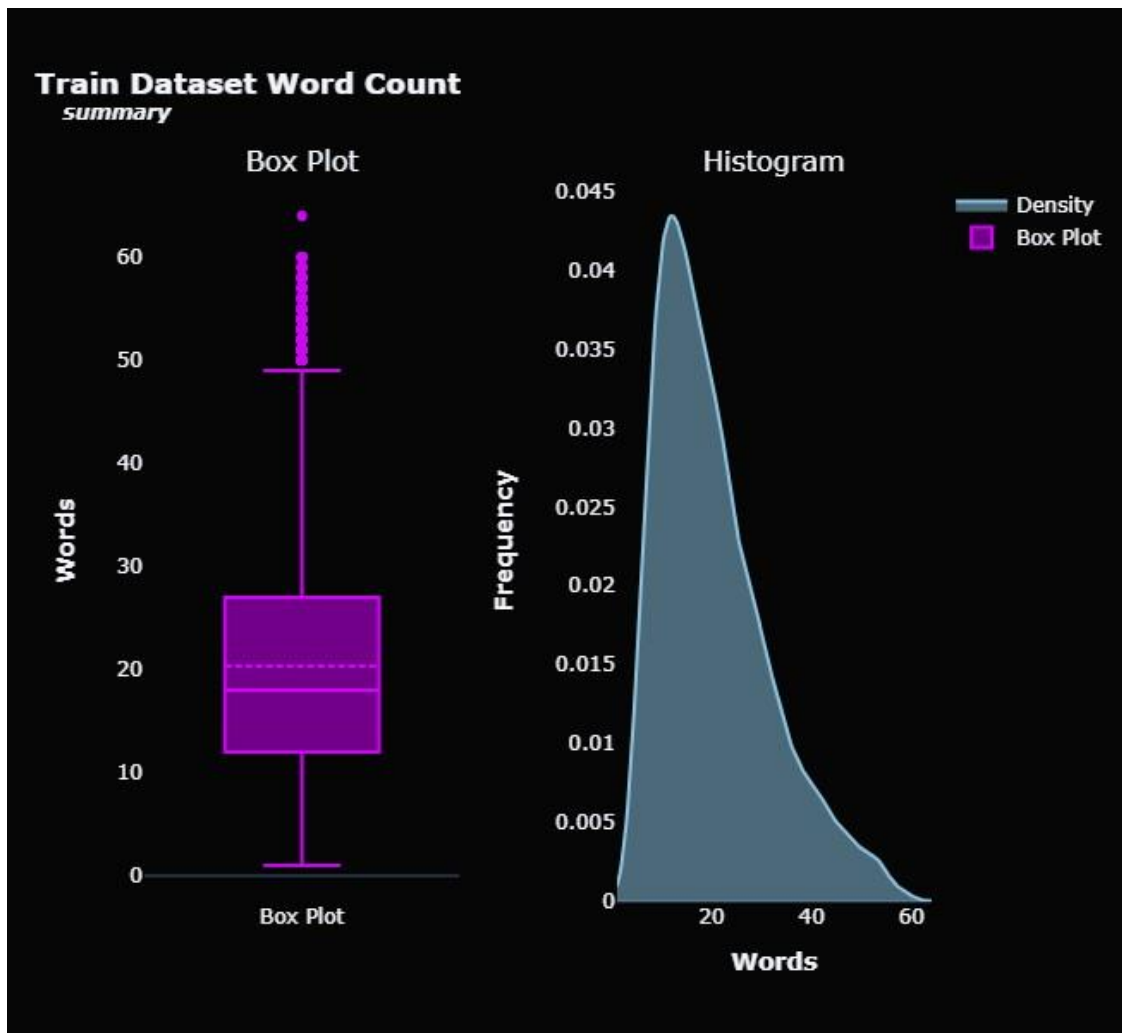
```
[54]: vectorizer = TfidfVectorizer(max_features = 15,stop_words =
'english') # Top 15□
      □terms x =
vectorizer.fit_transform(train['dialogue'])
df_tfidfvect = pd.DataFrame(x.toarray(),
columns=vectorizer.
      □get_feature_names_out()) plot_correlation(df_tfidfvect,
'Unigrams', 'Train - Dialogue', 800, 800, 12)
```

Unigrams Heatmap
Train - Dialogue

	did	don	good	just	know	like	ll	need	ok	really	sure	think	time	yeah
yes	-0.0	-0.07	-0.01	-0.04	-0.05	-0.04	-0.05	-0.04	-0.04	-0.02	-0.04	-0.02	-0.04	-0.1
yeah	0.01	-0.05	-0.01	-0.01	0.0	0.01	-0.05	-0.04	-0.07	0.02	-0.0	-0.01	-0.02	
time	-0.03	0.0	-0.0	-0.02	-0.03	-0.02	0.0	-0.02	-0.05	-0.0	-0.01	-0.02		
think	-0.01	0.08	-0.0	-0.0	0.03	0.03	-0.01	-0.0	-0.07	0.05	-0.01			
sure	-0.05	-0.03	-0.04	-0.02	-0.04	-0.03	0.02	-0.01	-0.05	-0.01				
really	-0.0	0.04	0.01	0.01	0.04	0.05	-0.03	-0.02	-0.08					
ok	-0.1	-0.07	-0.07	-0.07	-0.07	-0.1	0.04	-0.01						
need	-0.05	-0.02	-0.03	-0.0	-0.02	-0.04	-0.01							
ll	-0.08	-0.0	-0.03	-0.04	-0.01	-0.06								
like	-0.03	0.06	0.01	0.02	0.02									
know	-0.02	0.12	-0.02	0.02										
just	-0.02	0.02	-0.03											
good	-0.03	-0.05												
don	-0.04													

```
[55]: vectorizer = TfidfVectorizer(max_features = 15,stop_words =
'english') # Top 15␣
      ␣terms x =
      vectorizer.fit_transform(train['summary'].fillna(''))
      ) df_tfidfvect = pd.DataFrame(x.toarray(),
      columns=vectorizer.
```

```
 get_feature_names_out()) plot_correlation(df_tfidfvect,
'Unigrams', 'Train - Summary', 800, 800, 12)
```

Unigrams Heatmap
Train - Summary
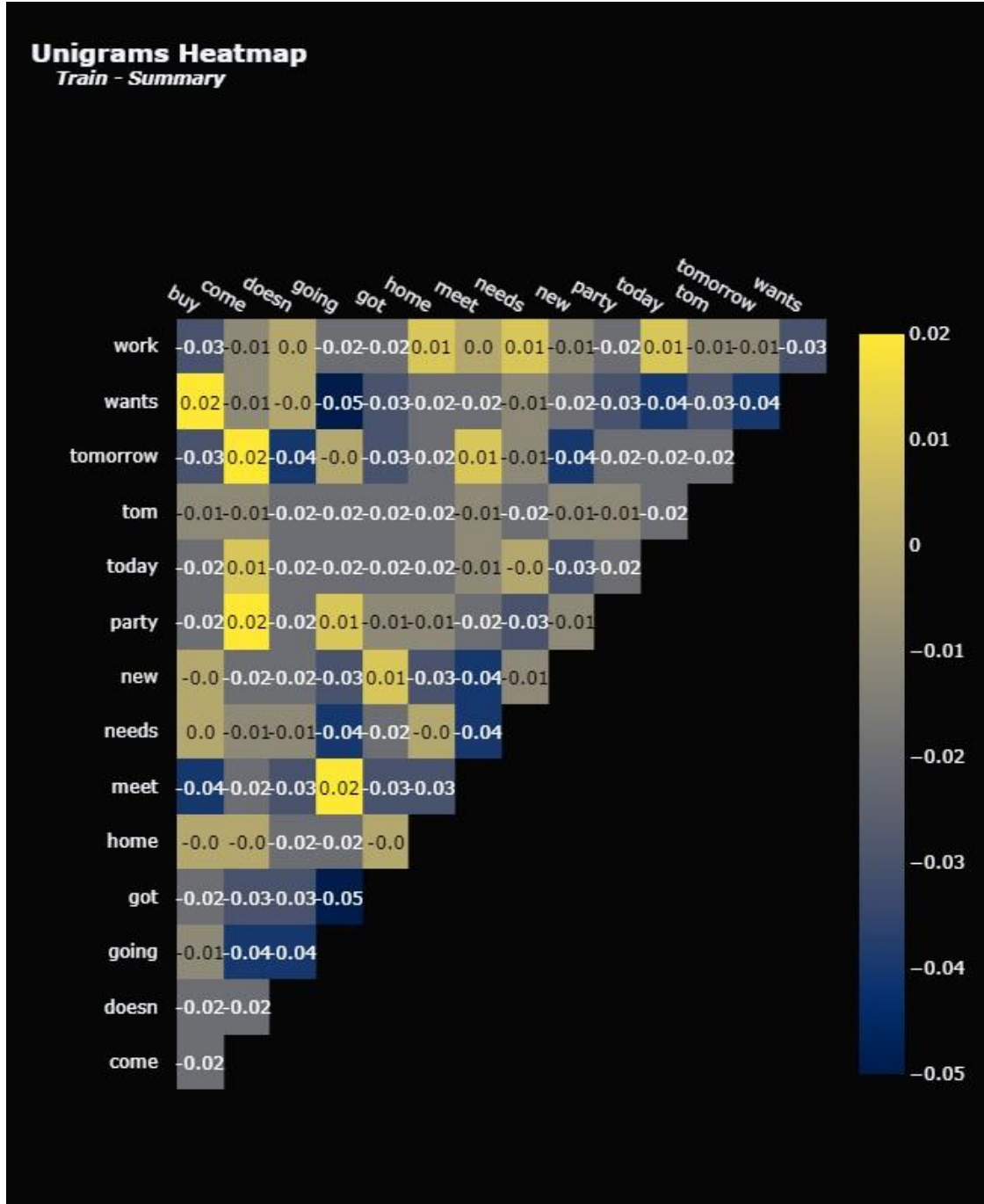

```
[57]: filtered_train = train[train['summary'].str.contains('15 minutes',
case=False,
     na=False)]
    filtered_train.head()
```

```
[57]:            id  \
     136    13827893
     428  13811484-1
     570    13818296
     1213 13682296-1
     1812    13820691


                                      dialogue \
     136
     Kate: I'm here <file_other>\r\nKate: there was no place in Red
     Lion\r\nSteven: hey! but it's quite far away\r\nKate: c'mon it's just
     10 min by bike!\r\nSteven: yes, but I'm not by bike\r\nKate:
     car?\r\nSteven: nope\r\nSteven: by foot :P :P \r\nSteven: anyway
     google maps says 15 min and I'm there:D\r\nKate: ok, w8in ^^ 428
     Jenny: Let's go out to eat.\r\nLucy: That sounds like fun.\r\nJenny:
     Where do you wanna go?\r\nLucy: Let me think a minute.\r\nJenny: I
     feel like Chinese.\r\nLucy: That sounds yummy.\r\nJenny: I know a
     good Chinese restaurant.\r\nLucy: How far away is it?\r\nJenny: It's
     only 10 minutes from my place.\r\nLucy: Do we have to book a
     table?\r\nJenny: Oh, no. We can walk right in.\r\nLucy: Cool. Will be
     in 15 minute. I'm really hungry!
     570
     Danielle: hey where RU?\r\nJuan: I told u I'd be late!\r\nDanielle:
     but it's been almost 45 mins!\r\nDanielle: <file_gif>\r\nJuan: I'll
     be there in 15 minutes\r\nJuan: <file_gif>
     1213  John: I know you will be outraged but I like to provoke you
     :P\r\nTyre: What is it?\r\nJohn: I talked to our neighbour today and
     I am really starting to think that religious people are just
     stupid.\r\nTyre: Gosh. You know it's a stupid claim.\r\nJohn: I know
     that there are some clever, religious individuals. But statistically
     religious people are stupid.\r\nTyre: It's not true. There are stupid
     religious people and clever ones, just like atheists.\r\nJohn: But
     most of academics are not religious.\r\nTyre: How do you know
     it?\r\nJohn: Experience but also some data I've seen.\r\nTyre: It's
     just not true.\r\nJohn: They are mostly people believing in things
     that have nothing to do with logic or reason: miracles, ghosts,
     witchcraft, just as our neighbour.\r\nTyre: I think it's only one
     part of them. There are theologians, people who actually know a lot
     about philosophy, logic etc.\r\nJohn: Yes, there are also people
     doing "scientifically" tarot, horoscopes and astrology.\r\nTyre: You
     ca…
     1812
     Madge: are you alive? xD\r\nDorothy: i'm still drunk\r\nMadge:
     xDDDDDDDD jeeez\r\nFelicia: I don't know…how much did i
     drink?\r\nMadge: like 10 rounds\r\nFelicia: SHIT \r\nFelicia: you
     gotta be kidding me ahahaha xDDDDDDDDDDDDDDDDDDD\r\nDorothy: of course
```

37

she is\r\nDorothy: it was at least 15\r\nFelicia:
;_____;\r\nFelicia: was nice to meet you girls…shame on me
as always\r\nDorothy: oh stop talking\r\nDorothy: just live the
moment B-)\r\nFelicia: how am i supossed to live the moment if i
don't remember the half o the night XD\r\nDorothy: well it happens :p
\r\nMadge: we gotta repeat it, i had a lot of fun :D\r\nDorothy: i'm
in. in 15min?\r\nFelicia: you're crazy ;-;

                              summary
136
Kate will meet with Steven in 15 minutes.
428   Jenny and Lucy are going to a Chinese restaurant to eat. They do
not need to book a table. Lucy will be at Jenny's in 15 minutes.
570
Juan is almost 45 minutes late. He'll be there in 15 minutes.
1213   John and Tyre's neighbour stopped John in the staircase and
talked about some miracles for 15 minutes. John thinks that religious
people are stupid. Tyre disagrees with this generalization.
1812   Dorothy is still intoxicated after at least 15 rounds of drink
yesterday and can't remember much of what happened. She would like to
meet her friends for a drink again in 15 minutes.

```
[58]: vectorizer = TfidfVectorizer(max_features = 15,stop_words =
       'english',ngram_range = (3,3)) # Top 15 terms x =
      vectorizer.fit_transform(train['dialogue'].fillna(''
      )) df_tfidfvect = pd.DataFrame(x.toarray(),
      columns=vectorizer.
       get_feature_names_out()) plot_correlation(df_tfidfvect,
      'Trigrams', 'Train - Dialogue', 800, 800, 12)
```
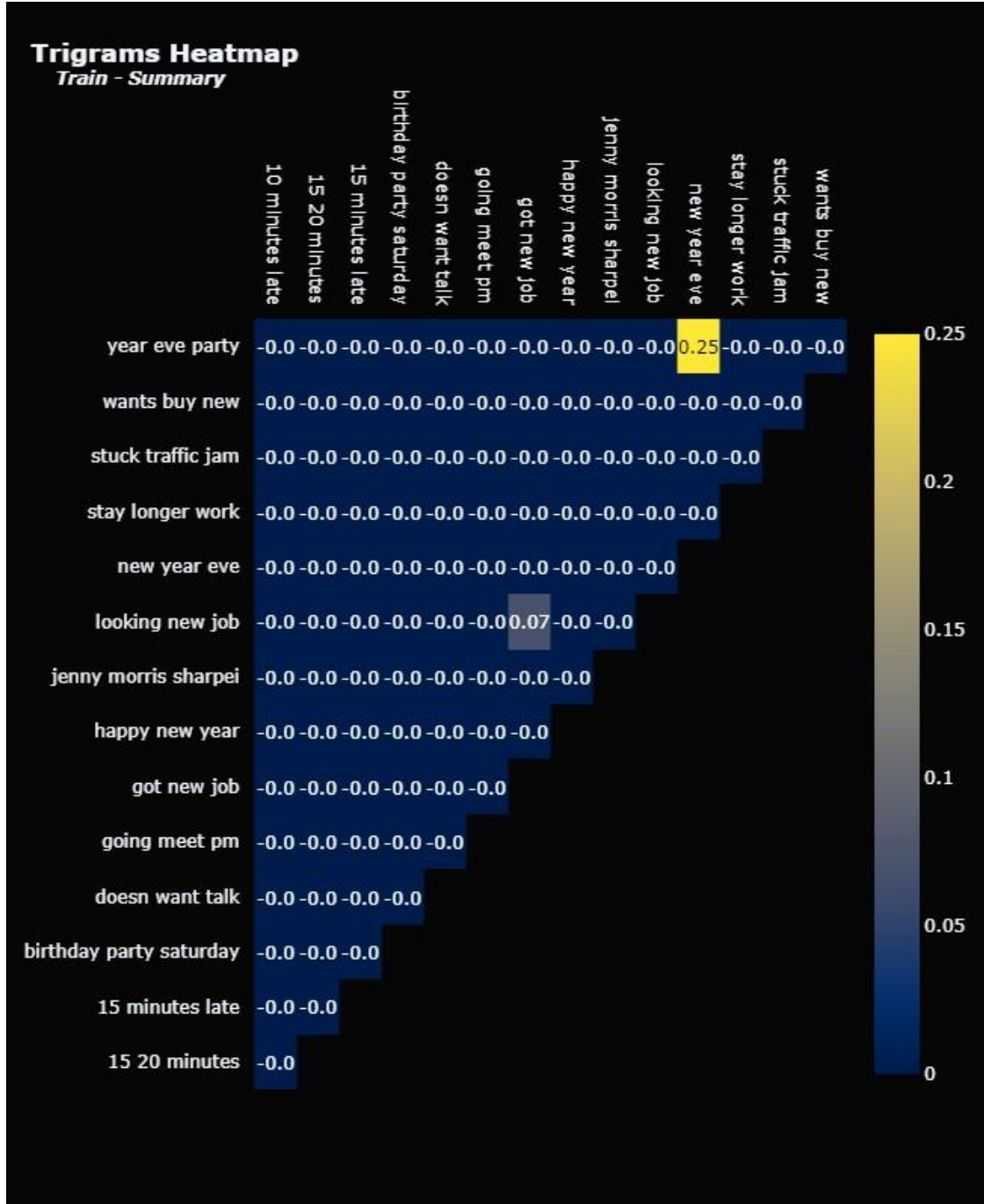
Trigrams Heatmap
Train - Dialogue
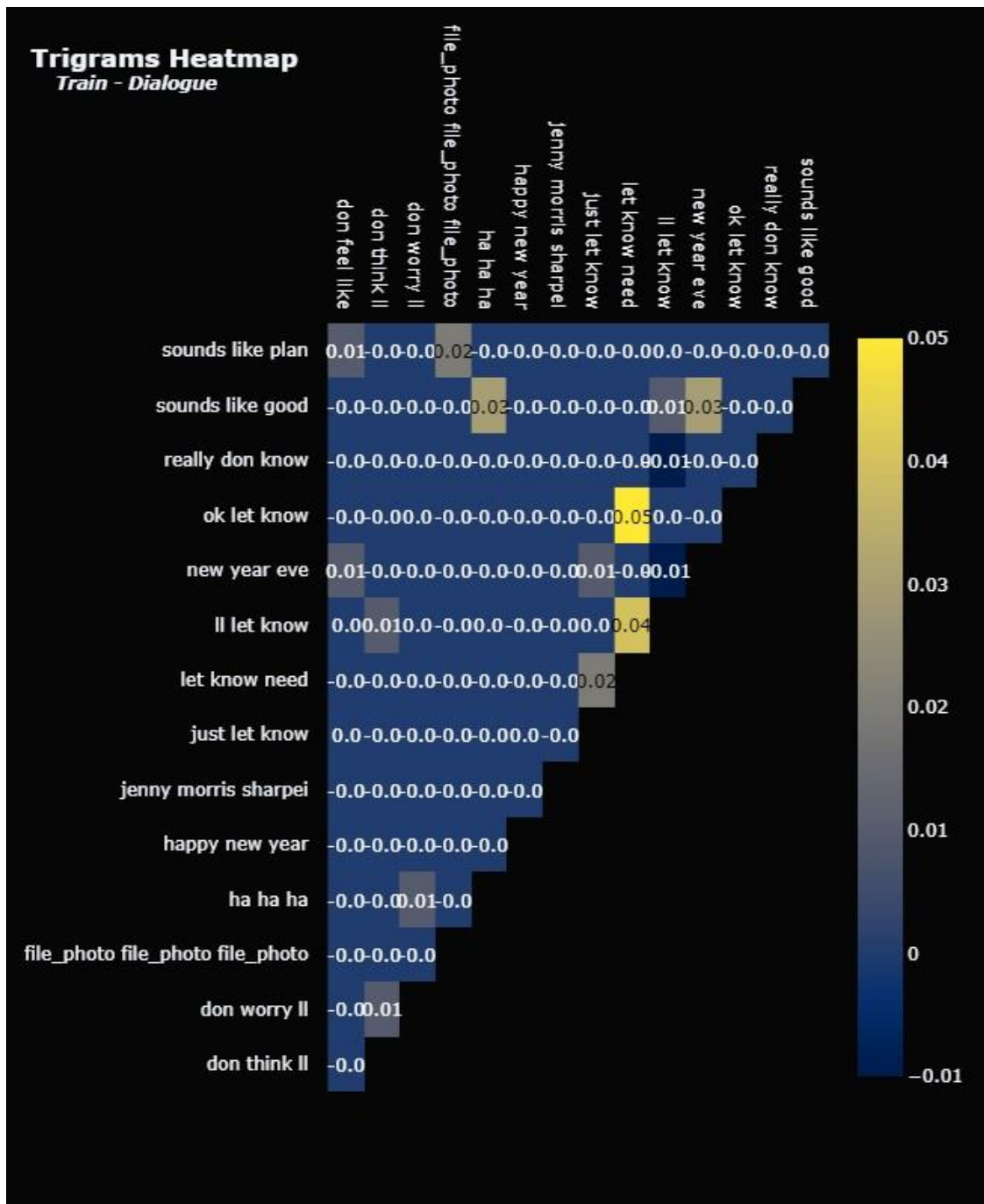
```
[59]: vectorizer = TfidfVectorizer(max_features = 15,stop_words =□
      □'english',ngram_range = (3,3)) # Top 15 terms x =
      vectorizer.fit_transform(train['summary'].fillna('')
      ) df_tfidfvect = pd.DataFrame(x.toarray(),
      columns=vectorizer.
```

```
 get_feature_names_out()) plot_correlation(df_tfidfvect,
'Trigrams', 'Train - Summary', 800, 800, 12)
```


Trigrams Heatmap
Train - Summary

```
[12]: !pip install --upgrade diffusers transformers -q
```

```
[17]: !pip install pycocotools
```

Requirement already satisfied: pycocotools in
c:\users\sdill\anaconda3\lib\sitepackages (2.0.7)
Requirement already satisfied: matplotlib>=2.1.0 in
c:\users\sdill\anaconda3\lib\site-packages (from pycocotools) (3.8.0)
Requirement already satisfied: numpy in
c:\users\sdill\anaconda3\lib\sitepackages (from pycocotools) (1.26.4)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools)
(0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools)
(4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools) (23.1)
Requirement already satisfied: pillow>=6.2.0 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools)
(10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\sdill\anaconda3\lib\site-packages (from matplotlib>=2.1.0-
>pycocotools) (2.8.2)
Requirement already satisfied: six>=1.5 in
c:\users\sdill\anaconda3\lib\sitepackages (from python-dateutil>=2.7-
>matplotlib>=2.1.0->pycocotools) (1.16.0)

```python
[18]: import time
      import keras_cv
      from tensorflow import keras
      import matplotlib.pyplot as plt
```

```python
[19]: model = keras_cv.models.StableDiffusion(img_width=512, img_height=512)
```

By using this model checkpoint, you acknowledge that its usage is
subject to the terms of the CreativeML Open RAIL-M license at

https://raw.githubusercontent.com/CompVis/stable-
diffusion/main/LICENSE

[20]:
```python
images = model.text_to_image("photograph of an astronaut in a dark alley with ⬚
    ⬚purple lights", batch_size=4)



def plot_images(images):
    plt.figure(figsize=(20, 20))
    for i in range(len(images)):
        ax = plt.subplot(1, len(images), i + 1)
        plt.imshow(images[i])
        plt.axis("off")


plot_images(images)
```
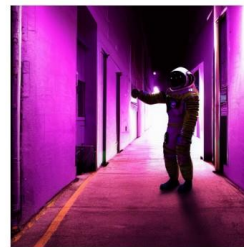
WARNING:tensorflow:From
C:\Users\sdill\anaconda3\Lib\sitepackages\keras\src\backend\tensorflow
\core.py:188: The name tf.placeholder is deprecated. Please use
tf.compat.v1.placeholder instead.

50/50   1366s 26s/step
Downloading data from
https://huggingface.co/fchollet/stablediffusion/resolve/main/kcv_decod
er.h5
198180272/198180272   16s 0us/step

[12]: