1

May 19, 2024

0.1

1. :
   - : , , - ,
     .
   - : ,
     .
2. :
   - .
   - .
3. :
   - .
   - .
4. :
   - .
   - , , (, .).
   - , .
5. :
   - .
   - , .
6. :
   - .
   - .
7. :
   - .
   - .

0.1.1 1. .

, mkrf_movies .

```
<div class="card">
    <div class="card-header" id="headingThree">
        <button class="btn btn-link collapsed" data-toggle="collapse" data-target="#collapseHi
    </div>
    <div id="collapseHint_0" class="collapse" aria-labelledby="headingThree" data-parent="#acc
        <div class="card-body">
```

, . .

```
    </div>
</div>
```

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
pd.set_option('display.float_format', '{:,.2f}'.format)
pd.set_option('chained_assignment', None)
```

[ ]:
```
#
movies_data = pd.read_csv('datasets\\mkrf_movies.csv')
shows_data = pd.read_csv('datasets\\mkrf_shows.csv')
```

[ ]:
```
display(movies_data)
movies_data.info()
```

```
                                                title   puNumber  \
0                                                      221048915
1                                                     111013716
2                                                        221038416
3                                                        221026916
4                                                        221030815
…                                                              …
7481            :                  126008019
7482                                                    121037819
7483             :               !  124010819
7484              .     "          "  124010619
7485          .     "            "  124010719

               show_start_date             type  \
0     2015-11-27T12:00:00.000Z
1     2016-09-13T12:00:00.000Z
2     2016-10-10T12:00:00.000Z
3     2016-06-10T12:00:00.000Z
4     2015-07-29T12:00:00.000Z
…                          …              …
7481  2019-12-23T12:00:00.000Z
7482  2019-12-24T12:00:00.000Z
7483  2019-12-28T12:00:00.000Z
7484  2019-12-30T12:00:00.000Z
7485  2019-12-30T12:00:00.000Z

                                           film_studio  \
0              ,          ,       …
1                                        "     "
2               ,           ,      …
3               ,           ,      …
```

```
4               ,            ,      …
…                                                  …
7481
7482                ,             …
7483                           ,
7484                    .,         "      "
7485                    .,         "      "

           production_country            director  \
0
1                                        .
2
3
4
…                         …                 …
7481
7482
7483                 -             .
7484           ,                .
7485           ,                .

                                                 producer  \
0                       ,         ,
1                                                      NaN
2            ,          ,     .   ,
3            ,          ,     .   ,
4            ,          ,     .   ,
…                                                     …
7481
7482          ,          ,        ,   …
7483
7484                                                .
7485                                                .

                               age_restriction  refundable_support  \
0                 «18+» -                                    NaN
1               «6+» -              6                        NaN
2                 «18+» -                                    NaN
3                 «18+» -                                    NaN
4                 «18+» -                                    NaN
…                                          …                   …
7481          «12+» -              12                        NaN
7482          «16+» -              16                        NaN
7483  «0+» -                                      NaN
7484  «0+» -                                      NaN
7485  «0+» -                                      NaN

      nonrefundable_support  budget financing_source ratings  \
```

```
0                       NaN     NaN             NaN     7.2
1                       NaN     NaN             NaN     6.6
2                       NaN     NaN             NaN     6.8
3                       NaN     NaN             NaN     6.8
4                       NaN     NaN             NaN     6.8
…                       …       …               …       …
7481                    NaN     NaN             NaN     NaN
7482                    NaN     NaN             NaN     5.4
7483                    NaN     NaN             NaN     NaN
7484                    NaN     NaN             NaN     NaN
7485                    NaN     NaN             NaN     6.3

                        genres
0           ,   ,
1                   ,
2           ,   ,
3           ,   ,
4           ,   ,
…                       …
7481                    NaN
7482                ,
7483                    NaN
7484                    NaN
7485                ,

[7486 rows x 15 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7486 entries, 0 to 7485
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   title                  7486 non-null   object
 1   puNumber               7486 non-null   object
 2   show_start_date        7486 non-null   object
 3   type                   7486 non-null   object
 4   film_studio            7468 non-null   object
 5   production_country     7484 non-null   object
 6   director               7477 non-null   object
 7   producer               6918 non-null   object
 8   age_restriction        7486 non-null   object
 9   refundable_support     332 non-null    float64
 10  nonrefundable_support  332 non-null    float64
 11  budget                 332 non-null    float64
 12  financing_source       332 non-null    object
 13  ratings                6519 non-null   object
 14  genres                 6510 non-null   object
dtypes: float64(3), object(12)
```

```
memory usage: 877.4+ KB
```

[ ]:
```
display(shows_data)
shows_data.info()
```

```
       puNumber     box_office
0     111000113       2,450.00
1     111000115      61,040.00
2     111000116 153,030,013.40
3     111000117  12,260,956.00
4     111000118 163,684,057.79
…         …             …
3153  224014814       1,530.00
3154  224021411       9,270.00
3155  224022410         400.00
3156  224023210         360.00
3157  224026410         400.00

[3158 rows x 2 columns]

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3158 entries, 0 to 3157
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   puNumber    3158 non-null   int64
 1   box_office  3158 non-null   float64
dtypes: float64(1), int64(1)
memory usage: 49.5 KB
```

[ ]:
```
#               puNumber
movies_data['puNumber'] = pd.to_numeric(movies_data['puNumber'],␣
 ↪errors='coerce')
shows_data['puNumber'] = pd.to_numeric(shows_data['puNumber'], errors='coerce')

#               puNumber   int
movies_data['puNumber'] = movies_data['puNumber'].astype('Int64')
shows_data['puNumber'] = shows_data['puNumber'].astype('Int64')
```

[ ]:
```
merged_data = pd.merge(movies_data, shows_data, on='puNumber', how='left')
```

- moviesdata showsdata.

1. ‘puNumber’ pd.to_numeric(). . errors=‘coerce’ , - , , NaN.
2. ‘puNumber’ ‘Int64’ astype(). “ ” - , (, NaN , NaN).

3. , , pd.merge(). ‘left’ , moviesdata, ‘puNumber’ showsdata . NaN. ‘puNumber’ , .

0.1.2 2.

2.1.

- , .

[ ]:
```
merged_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7486 entries, 0 to 7485
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   title                  7486 non-null   object
 1   puNumber               7485 non-null   Int64
 2   show_start_date        7486 non-null   object
 3   type                   7486 non-null   object
 4   film_studio            7468 non-null   object
 5   production_country     7484 non-null   object
 6   director               7477 non-null   object
 7   producer               6918 non-null   object
 8   age_restriction        7486 non-null   object
 9   refundable_support     332 non-null    float64
 10  nonrefundable_support  332 non-null    float64
 11  budget                 332 non-null    float64
 12  financing_source       332 non-null    object
 13  ratings                6519 non-null   object
 14  genres                 6510 non-null   object
 15  box_office             3158 non-null   float64
dtypes: Int64(1), float64(4), object(11)
memory usage: 943.2+ KB
```

[ ]:
```
#         show_start_date     datetime
merged_data['show_start_date'] = pd.to_datetime(merged_data['show_start_date'])

#             category
merged_data['type'] = merged_data['type'].astype('category')
merged_data['age_restriction'] = merged_data['age_restriction'].
 ↪astype('category')
merged_data['financing_source'] = merged_data['financing_source'].
 ↪astype('category')
#             ,        '%'
merged_data['ratings'] = merged_data['ratings'].str.replace('%', '').
 ↪astype(float)
```

```
merged_data['ratings'] = merged_data['ratings'].apply(lambda x: x / 10 if x >␣
↪10 else x)
```

1. show_start_date datetime ,
. , .
datetime, , ,
(,).
2. type category .
(. .), category
object ( ). .
3. financing_source age_restriction category. film_studio
production_country director producer genres
4. ratings .
. , , 10,
, 10, 10.
.

[ ]:
```
display(merged_data)
```

```
                                          title   puNumber  \
0                                          221048915
1                                         111013716
2                                            221038416
3                                            221026916
4                                            221030815
…                                           …        …
7481            :                  126008019
7482                                          121037819
7483             :               !  124010819
7484              .     "          "  124010619
7485          .     "            "  124010719

              show_start_date             type  \
0    2015-11-27 12:00:00+00:00
1    2016-09-13 12:00:00+00:00
2    2016-10-10 12:00:00+00:00
3    2016-06-10 12:00:00+00:00
4    2015-07-29 12:00:00+00:00
…                         …              …
7481 2019-12-23 12:00:00+00:00
7482 2019-12-24 12:00:00+00:00
7483 2019-12-28 12:00:00+00:00
7484 2019-12-30 12:00:00+00:00
7485 2019-12-30 12:00:00+00:00

                                          film_studio  \
0              ,          ,        …
```

```
1                                        "     "
2              ,            ,      …
3              ,            ,      …
4              ,            ,      …
…                                                 …
7481
7482               ,             …
7483                          ,
7484                  .,         "      "
7485                  .,         "      "

          production_country           director  \
0
1                                    .
2
3
4
…                       …                …
7481
7482
7483               -             .
7484          ,                 .
7485          ,                 .

                                                producer  \
0                    ,         ,
1                                                    NaN
2           ,         ,     .   ,
3           ,         ,     .   ,
4           ,         ,     .   ,
…                                                   …
7481
7482         ,          ,        ,   …
7483
7484                                               .
7485                                               .

                             age_restriction  refundable_support  \
0                «18+» -                              NaN
1               «6+» -             6                  NaN
2                «18+» -                              NaN
3                «18+» -                              NaN
4                «18+» -                              NaN
…                                      …                 …
7481          «12+» -            12                   NaN
7482          «16+» -            16                   NaN
7483  «0+» -                                    NaN
7484  «0+» -                                    NaN
```

```
7485  «0+» -                                    NaN

      nonrefundable_support  budget financing_source  ratings  \
0                       NaN     NaN              NaN     7.20
1                       NaN     NaN              NaN     6.60
2                       NaN     NaN              NaN     6.80
3                       NaN     NaN              NaN     6.80
4                       NaN     NaN              NaN     6.80
…                       …       …                …       …
7481                    NaN     NaN              NaN      NaN
7482                    NaN     NaN              NaN     5.40
7483                    NaN     NaN              NaN      NaN
7484                    NaN     NaN              NaN      NaN
7485                    NaN     NaN              NaN     6.30

                        genres  box_office
0            ,   ,              NaN
1                     ,             NaN
2           ,   ,               NaN
3           ,   ,               NaN
4           ,   ,               NaN
…                         …          …
7481                      NaN         NaN
7482                ,             NaN
7483                      NaN         NaN
7484                      NaN         NaN
7485                ,             NaN

[7486 rows x 16 columns]
```

[ ]:
```
merged_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7486 entries, 0 to 7485
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   title                 7486 non-null   object
 1   puNumber              7485 non-null   Int64
 2   show_start_date       7486 non-null   datetime64[ns, UTC]
 3   type                  7486 non-null   category
 4   film_studio           7468 non-null   object
 5   production_country    7484 non-null   object
 6   director              7477 non-null   object
 7   producer              6918 non-null   object
 8   age_restriction       7486 non-null   category
 9   refundable_support    332 non-null    float64
 10  nonrefundable_support  332 non-null    float64
```

```
 11  budget                 332 non-null    float64
 12  financing_source       332 non-null    category
 13  ratings                6519 non-null   float64
 14  genres                 6510 non-null   object
 15  box_office             3158 non-null   float64
dtypes: Int64(1), category(3), datetime64[ns, UTC](1), float64(5), object(6)
memory usage: 790.4+ KB
```

2.2.

- , .

[ ]:
```
#
temp = merged_data.copy() #
len(temp)
```

[ ]: 7486

[ ]:
```
#
merged_data['film_studio'].fillna('Unknown', inplace=True)
merged_data['production_country'] = merged_data['production_country'].
 ↪replace('2019', '')
merged_data['production_country'].fillna('Unknown', inplace=True)
merged_data['director'].fillna('Unknown', inplace=True)
merged_data['producer'].fillna('Unknown', inplace=True)
merged_data['genres'].fillna('Unknown', inplace=True)
#                          'puNumber'
merged_data['puNumber'].fillna(0, inplace=True)
```

```
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:2: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  merged_data['film_studio'].fillna('Unknown', inplace=True)
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:4: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.
```

```
For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  merged_data['production_country'].fillna('Unknown', inplace=True)
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:5: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  merged_data['director'].fillna('Unknown', inplace=True)
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:6: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  merged_data['producer'].fillna('Unknown', inplace=True)
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:7: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  merged_data['genres'].fillna('Unknown', inplace=True)
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\1364134147.py:9: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
```

```
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.
```

```
merged_data['puNumber'].fillna(0, inplace=True)
```

1. merged_data (, 'film_studio', 'production_country', 'director', 'producer', 'genres', 'puNumber') . 'Unknown' ('film_studio', 'production_country', 'director', 'producer', 'genres'), , .
2. 0 'puNumber', . , , .
3. refundable_support, nonrefundable_support, budget financing_source
4. box_office

[ ]:
```
display(merged_data)
```

```
                                          title   puNumber  \
0                                          221048915
1                                         111013716
2                                            221038416
3                                            221026916
4                                            221030815
…                                           …        …
7481            :                 126008019
7482                                          121037819
7483             :              !  124010819
7484              .     "         "  124010619
7485          .     "           "  124010719

               show_start_date             type  \
0    2015-11-27 12:00:00+00:00
1    2016-09-13 12:00:00+00:00
2    2016-10-10 12:00:00+00:00
3    2016-06-10 12:00:00+00:00
4    2015-07-29 12:00:00+00:00
…                        …              …
7481 2019-12-23 12:00:00+00:00
7482 2019-12-24 12:00:00+00:00
```

```
7483 2019-12-28 12:00:00+00:00
7484 2019-12-30 12:00:00+00:00
7485 2019-12-30 12:00:00+00:00

                                          film_studio  \
0              ,          ,        …
1                                      "     "
2              ,            ,      …
3              ,            ,      …
4              ,            ,      …
…                                                 …
7481
7482               ,             …
7483                          ,
7484                  .,        "      "
7485                  .,        "      "

          production_country           director  \
0
1                                    .
2
3
4
…                      …                …
7481
7482
7483               -            .
7484           ,                .
7485           ,                .

                                             producer  \
0                    ,         ,
1                                             Unknown
2           ,         ,     .   ,
3           ,         ,     .   ,
4           ,         ,     .   ,
…                                                …
7481
7482         ,         ,        ,   …
7483
7484                                           .
7485                                           .

                           age_restriction  refundable_support  \
0               «18+» -                            NaN
1              «6+» -            6                 NaN
2               «18+» -                            NaN
3               «18+» -                            NaN
```

```
4                «18+» -                                NaN
…                                      …                 …
7481         «12+» -            12                     NaN
7482         «16+» -            16                     NaN
7483  «0+» -                                      NaN
7484  «0+» -                                      NaN
7485  «0+» -                                      NaN

      nonrefundable_support  budget financing_source  ratings  \
0                       NaN     NaN               NaN     7.20
1                       NaN     NaN               NaN     6.60
2                       NaN     NaN               NaN     6.80
3                       NaN     NaN               NaN     6.80
4                       NaN     NaN               NaN     6.80
…                     …     …               …       …
7481                    NaN     NaN               NaN      NaN
7482                    NaN     NaN               NaN     5.40
7483                    NaN     NaN               NaN      NaN
7484                    NaN     NaN               NaN      NaN
7485                    NaN     NaN               NaN     6.30

                        genres  box_office
0           ,   ,              NaN
1                   ,             NaN
2           ,   ,             NaN
3           ,   ,             NaN
4           ,   ,             NaN
…                         …         …
7481                    Unknown         NaN
7482                ,             NaN
7483                    Unknown         NaN
7484                    Unknown         NaN
7485                ,             NaN

[7486 rows x 16 columns]
```

2.3.

- , . , .

[ ]:
```
#                          .
duplicates = merged_data.duplicated()

#            ,       .
if duplicates.any():
    print("          :")
    print(merged_data[duplicates])
else:
```

```
    print("            ")
```

[ ]:
```
duplicated_titles = merged_data[merged_data['title'].duplicated(keep=False)]
display(duplicated_titles)
```

```
                                                    title   puNumber  \
2                                                          221038416
3                                                          221026916
4                                                          221030815
12              /             . . …  221011415
26                                                         221012515
…                                                          …        …
7369                                                        111019319
7416                                  "         "  121035119
7438                                                         121035919
7450                                                       121036919
7474                                                        111022519

              show_start_date             type  \
2    2016-10-10 12:00:00+00:00
3    2016-06-10 12:00:00+00:00
4    2015-07-29 12:00:00+00:00
12   2015-04-03 12:00:00+00:00
26   2015-04-03 12:00:00+00:00
…                        …              …
7369 2019-11-07 12:00:00+00:00
7416 2019-12-13 12:00:00+00:00
7438 2019-12-05 12:00:00+00:00
7450 2019-12-09 12:00:00+00:00
7474 2019-12-23 12:00:00+00:00

                                          film_studio  \
2              ,           ,     …
3              ,           ,     …
4              ,           ,     …
12           ,             ,     …
26               3,          ,    …
…                                                 …
7369                                    "      "
7416           ,          ,     …
7438
7450         ,       ,       ,    …
7474         "           ",    "     2011"

          production_country           director  \
2
```

```
3
4
12
26       -      -
…                        …                 …
7369                                  .
7416                 ,
7438                                    -
7450
7474                                .

                                                  producer  \
2           ,          ,     .   ,
3           ,          ,     .   ,
4           ,          ,     .   ,
12           ,          ,       .   , …
26             ,       ,         ,  …
…                                                    …
7369   .    ,  .    ,    "            …
7416                 ,          ,
7438                    - ,      -  ,
7450
7474   .     ,  .    ,  .      ,  .

                      age_restriction  refundable_support  \
2         «18+» -                              NaN
3         «18+» -                              NaN
4         «18+» -                              NaN
12    «16+» -            16                      NaN
26        «18+» -                              NaN
…                                …                   …
7369  «16+» -            16                      NaN
7416  «16+» -            16                      NaN
7438      «18+» -                              NaN
7450      «18+» -                              NaN
7474  «12+» -            12         100,000,000.00

      nonrefundable_support         budget financing_source  ratings  \
2                       NaN            NaN              NaN     6.80
3                       NaN            NaN              NaN     6.80
4                       NaN            NaN              NaN     6.80
12                      NaN            NaN              NaN     8.10
26                      NaN            NaN              NaN     7.70
…                      …             …                …       …
7369                    NaN            NaN              NaN     7.50
7416                    NaN            NaN              NaN     7.90
7438                    NaN            NaN              NaN     6.50
7450                    NaN            NaN              NaN     6.80
```

```
7474         400,000,000.00 944,000,000.00                 5.70

                           genres  box_office
2                ,    ,             NaN
3                ,    ,             NaN
4                ,    ,             NaN
12                          ,            NaN
26               ,    ,               NaN
…                              …          …
7369           ,     ,             NaN
7416       ,       ,               NaN
7438                        ,     175,003.51
7450           ,     ,               NaN
7474                    ,              NaN

[1272 rows x 16 columns]
```

[ ]:
```
duplicated_titles = merged_data[merged_data['puNumber'].duplicated(keep=False)]
display(duplicated_titles)
```

```
                                                  title   puNumber  \
4638                                          221154310
4639      -                  /     …  221154310
5067                                               !  221054410
5068                                               !  221054410

              show_start_date            type  \
4638 2010-12-17 12:00:00+00:00
4639 2010-12-17 12:00:00+00:00
5067 2010-05-25 12:00:00+00:00
5068 2010-05-25 12:00:00+00:00

                                          film_studio  \
4638      ,          ,     1         …
4639            ,              ,    …
5067                 ,
5068          ,              ,  - …

                             production_country        director  \
4638
4639       -          -     -
5067                                  -
5068

                                       producer  \
4638                                ,
4639         ,     .    ,
5067                      ,
```

```
5068                    ,

                    age_restriction  refundable_support  \
4638  «16+» -            16                    NaN
4639  «16+» -            16                    NaN
5067  «16+» -            16                    NaN
5068  «12+» -            12                    NaN

      nonrefundable_support  budget financing_source  ratings  \
4638                    NaN     NaN              NaN     7.00
4639                    NaN     NaN              NaN     7.60
5067                    NaN     NaN              NaN     7.40
5068                    NaN     NaN              NaN     6.80

                     genres  box_office
4638             ,             NaN
4639     ,     ,            NaN
5067             ,               NaN
5068             ,               NaN
```

[ ]:
```
duplicated_titles = merged_data[merged_data[['puNumber', 'title']].
 ↪duplicated(keep=False)]

if duplicated_titles.empty:
    print("        ")
else:
    display(duplicated_titles)
```

[ ]:
```
#
a, b = len(temp), len(merged_data)
print(a, b, round((a-b)/a*100, 2))
```

```
7486 7486 0.0
```

[ ]:
```
display(merged_data)
```

```
                                          title   puNumber  \
0                                          221048915
1                                         111013716
2                                            221038416
3                                            221026916
4                                            221030815
…                                          …        …
7481            :                  126008019
7482                                        121037819
7483             :               !  124010819
7484              .     "          "  124010619
```

```
7485          .     "            "  124010719

               show_start_date             type  \
0    2015-11-27 12:00:00+00:00
1    2016-09-13 12:00:00+00:00
2    2016-10-10 12:00:00+00:00
3    2016-06-10 12:00:00+00:00
4    2015-07-29 12:00:00+00:00
…                          …              …
7481 2019-12-23 12:00:00+00:00
7482 2019-12-24 12:00:00+00:00
7483 2019-12-28 12:00:00+00:00
7484 2019-12-30 12:00:00+00:00
7485 2019-12-30 12:00:00+00:00

                                            film_studio  \
0              ,          ,        …
1                                      "     "
2              ,            ,      …
3              ,            ,      …
4              ,            ,      …
…                                                  …
7481
7482               ,             …
7483                          ,
7484                  .,         "      "
7485                  .,         "      "

           production_country           director  \
0
1                                    .
2
3
4
…                       …                …
7481
7482
7483               -             .
7484           ,                 .
7485           ,                 .

                                                producer  \
0                     ,         ,
1                                                Unknown
2           ,         ,     .   ,
3           ,         ,     .   ,
4           ,         ,     .   ,
…                                                   …
```

```
7481
7482         ,          ,        ,   …
7483
7484                                              .
7485                                              .

                            age_restriction  refundable_support  \
0                «18+» -                                    NaN
1               «6+» -             6                       NaN
2                «18+» -                                   NaN
3                «18+» -                                   NaN
4                «18+» -                                   NaN
…                                        …                   …
7481          «12+» -             12                       NaN
7482          «16+» -             16                       NaN
7483  «0+» -                                        NaN
7484  «0+» -                                        NaN
7485  «0+» -                                        NaN

      nonrefundable_support  budget financing_source  ratings  \
0                       NaN     NaN               NaN     7.20
1                       NaN     NaN               NaN     6.60
2                       NaN     NaN               NaN     6.80
3                       NaN     NaN               NaN     6.80
4                       NaN     NaN               NaN     6.80
…                     …     …               …       …
7481                    NaN     NaN               NaN      NaN
7482                    NaN     NaN               NaN     5.40
7483                    NaN     NaN               NaN      NaN
7484                    NaN     NaN               NaN      NaN
7485                    NaN     NaN               NaN     6.30

                         genres  box_office
0            ,   ,              NaN
1                      ,              NaN
2           ,    ,             NaN
3           ,    ,             NaN
4           ,    ,             NaN
…                          …          …
7481                    Unknown          NaN
7482                ,              NaN
7483                    Unknown          NaN
7484                    Unknown          NaN
7485                ,              NaN

[7486 rows x 16 columns]
```

2.4.

- , ;
- type.

```
<div class="card">
    <div class="card-header" id="headingThree">
        <button class="btn btn-link collapsed" data-toggle="collapse" data-target="#collapseHi
    </div>
    <div id="collapseHint_1" class="collapse" aria-labelledby="headingThree" data-parent="#acc
        <div class="card-body">
```

type , . « » – .str.strip.
, . : df['type'].str.strip()

```
    </div>
</div>
```

[ ]:
```
categorical_columns = merged_data.select_dtypes(include=['category']).columns

for column in categorical_columns:
    unique_values = merged_data[column].unique()
    unique_count = merged_data[column].nunique()

    print(f"    : {column}")
    print(f"           : {unique_values}")
    print(f"                : {unique_count}")
    print('---')
```

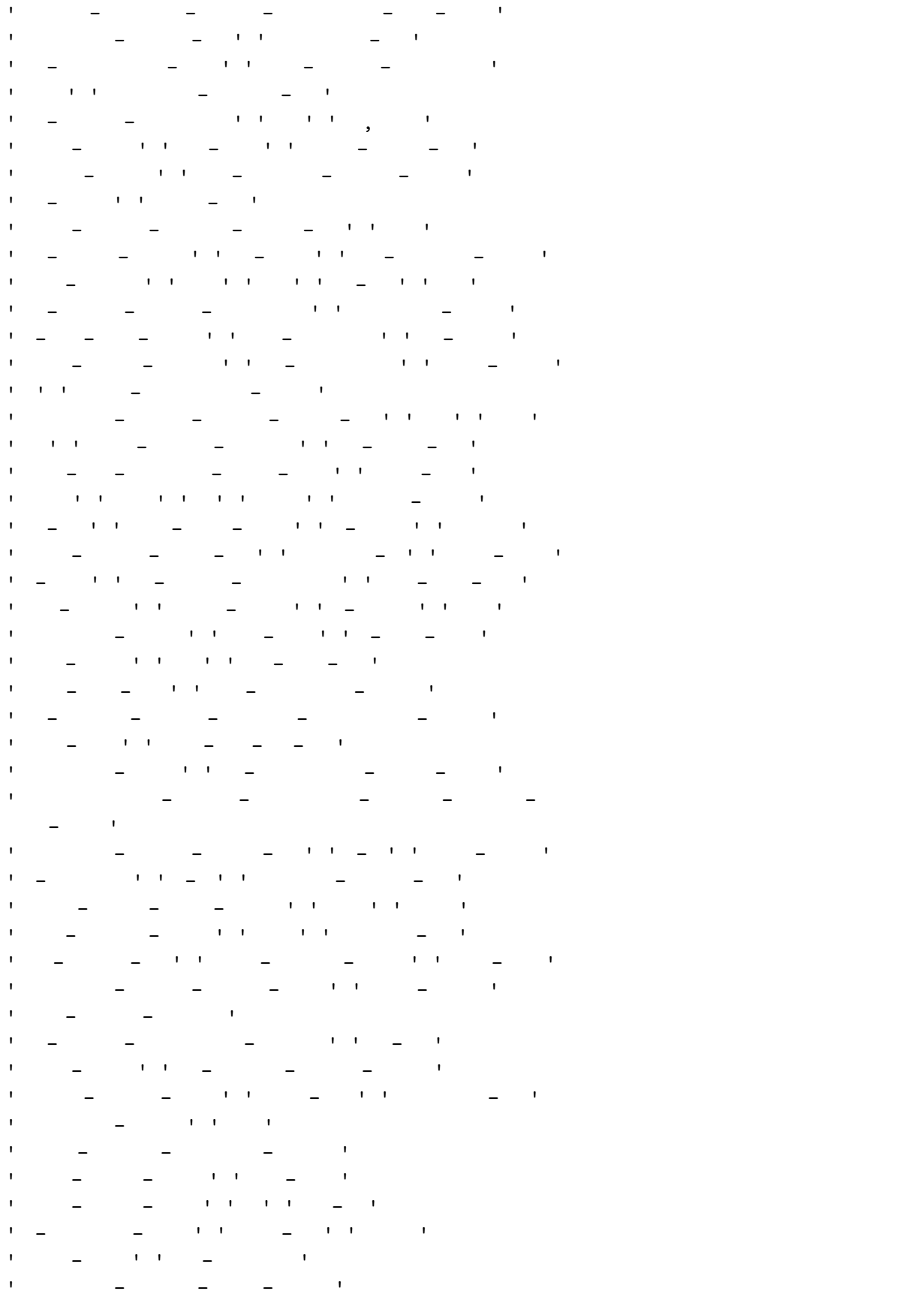
```
    : type
           : ['       ', '      ', '   ',
'       ', '   _     ', '        ', '       ',
'     -       ']
Categories (8, object): ['       ', '        ', '      ',
'       ', '     -       ', '   -     ', '   ',
'       ']
                : 8
---
    : age_restriction
           : ['«18+» -            ', '«6+» -           6
  ', '«12+» -           12   ', '«16+» -           16   ',
'«0+» -                  ']
Categories (5, object): ['«0+» -                  ', '«12+» -
       12   ', '«16+» -           16   ', '«18+» -
   ', '«6+» -           6   ']
                : 5
---
    : financing_source
           : [NaN, '           ', '     ', '
    ,      ']
```

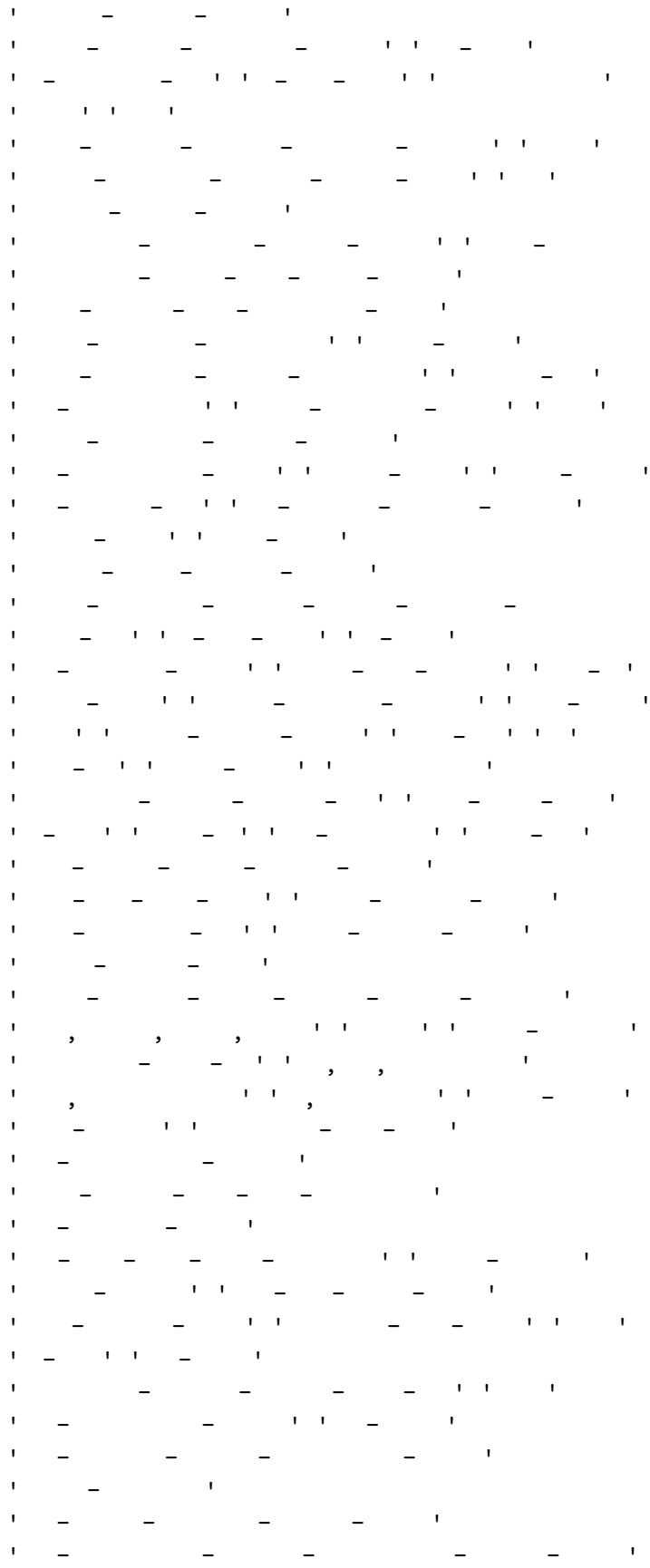
```
Categories (3, object): ['           ', '           ,
  ', '     ']
                : 3
---
```

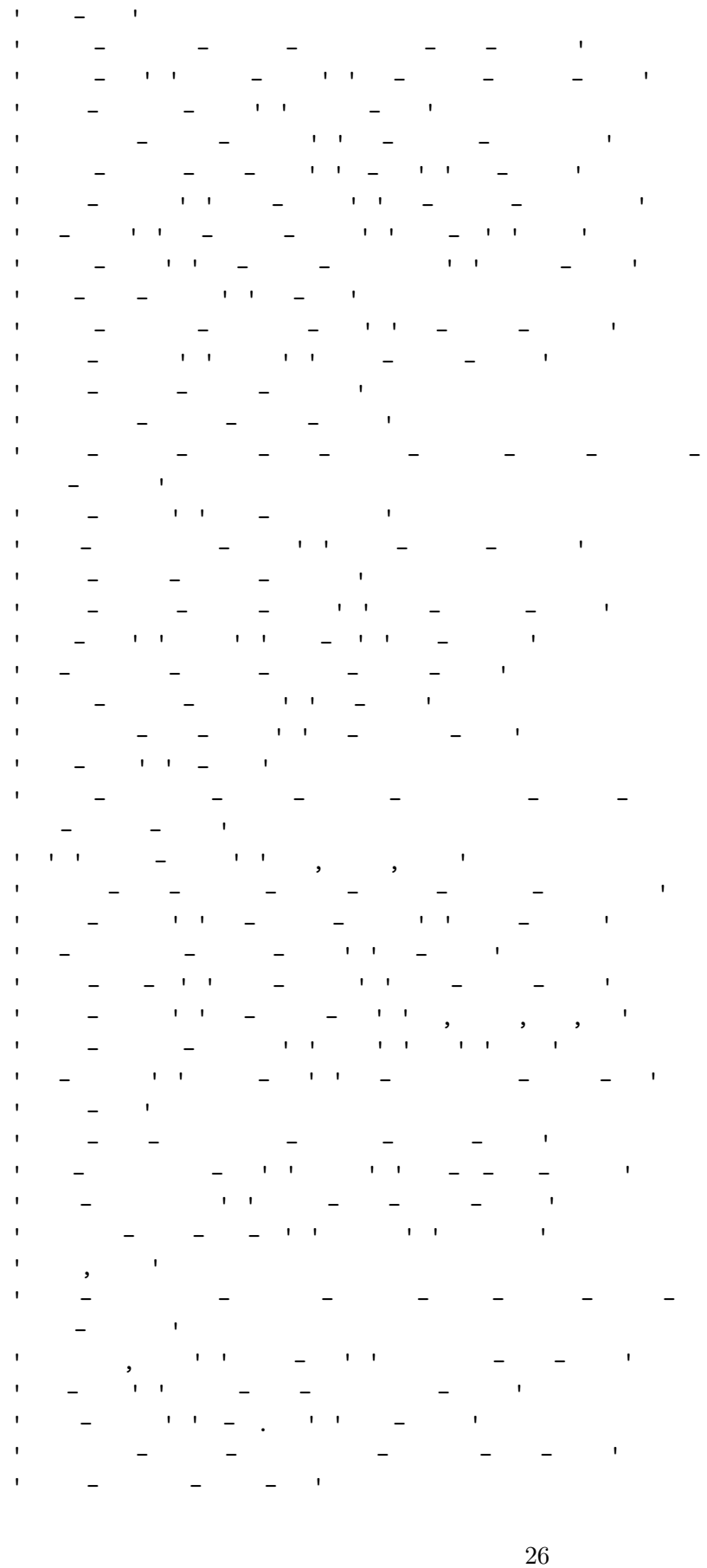
[ ]:
```
categorical_object_columns = merged_data.select_dtypes(include=['category',␣
↪'object']).columns

if len(categorical_object_columns) > 0:
    for column in categorical_object_columns:
        unique_values = merged_data[column].unique()
        print(f"\n    : {column}")
        print(f"                    : {unique_values}")
        print("-----------------------------------")
        #
        merged_data[column] = merged_data[column].str.replace(' - ', '-')
        merged_data[column] = merged_data[column].str.replace('-', ',')
        merged_data[column] = merged_data[column].str.replace(',', ', ')
        #
        merged_data[column] = merged_data[column].str.strip().str.lower()

        unique_values_after_fix = merged_data[column].unique()
        print(f"                    : {unique_values_after_fix}")

        print("-----------------------------------")
else:
    print("                              ")
```

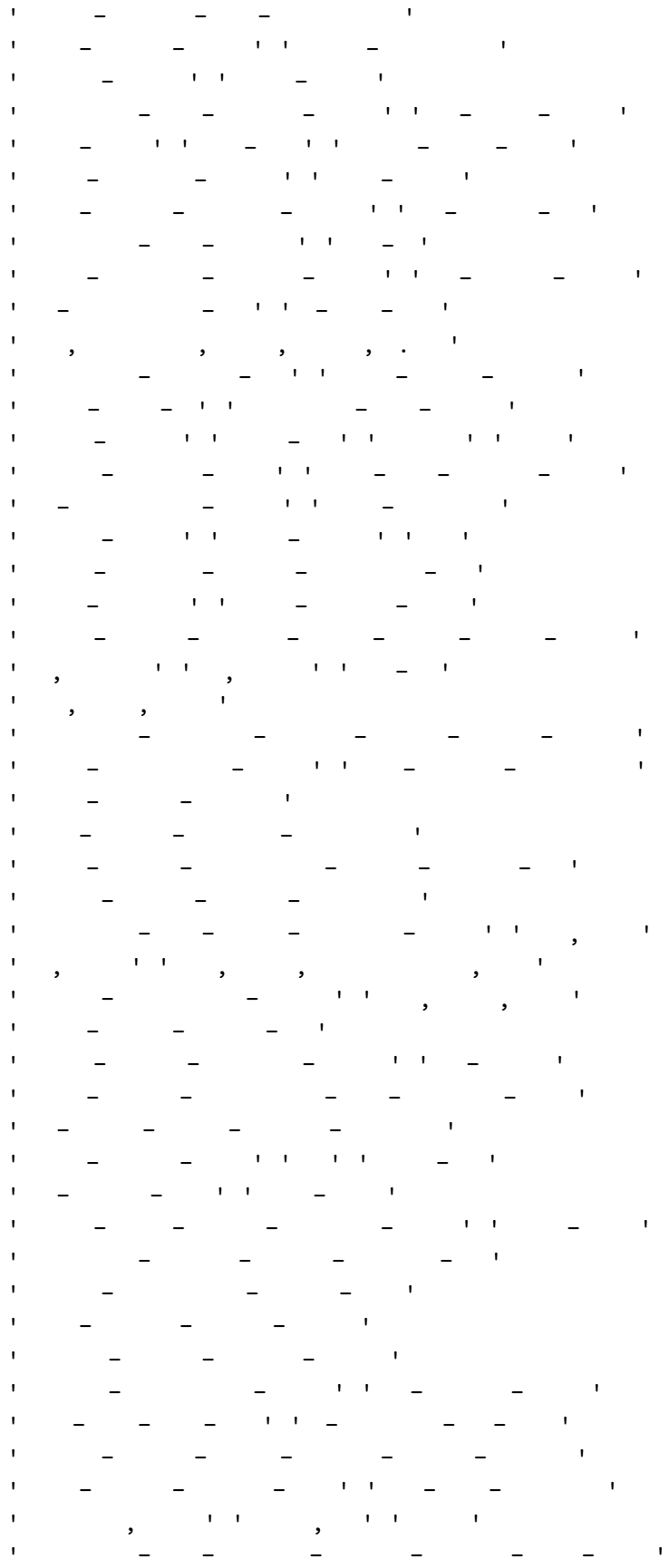
```
    : title
                    : ['         ' '           '
'       ' …
 '         :              !'
 '         .     "         "'
 '         .     "           "']
-----------------------------------
                     : ['         ' '
    ' '       ' …
 '         :              !'
 '         .     "         "'
 '         .     "           "']
-----------------------------------

    : type
                    : ['       ', '      ', '   ',
'       ', '   -     ', '        ', '       ',
'     -       ']
Categories (8, object): ['       ', '        ', '      ',
```

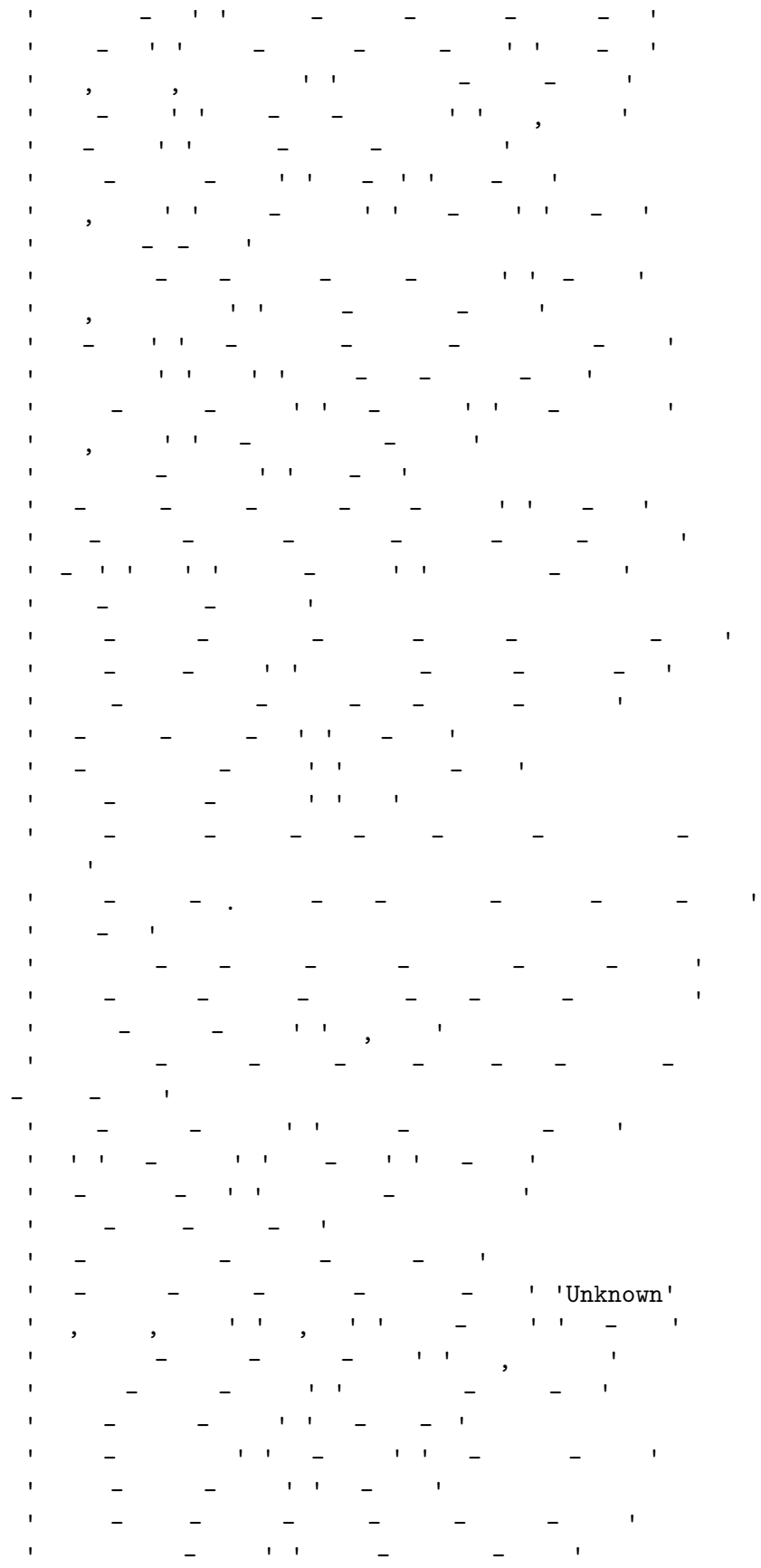
```
'        ', '     _        ', '   _     ', '   ',
'        ']
-----------------------------------
                      : ['        ' '       ' '    '
'        '
 '    ,       ' '      ,         ']
-----------------------------------

    : film_studio
                      : ['          ,          ,
    ,          '
 '       "     "'
 '           ,            ,             ,
   ,               ,          ,           '
 …
 '                ,                ,             ,
              ,        '
 '                 ,         '
 '         .,          "      "']
-----------------------------------
                      : ['          ,           ,
        ,             '
 '       "     "'
 '           ,             ,              ,
   ,                ,          ,             '
 …
 '                ,                  ,
       ,                ,        '
 '                   ,          '
 '         .,          "      "']
-----------------------------------

    : production_country
                      : ['  ' '   ' '    ' '   ,     '
'     _        '
 '         _     ' '       ' '   _      _     '
 '         ' '     _      _     ' '  _     '
 '   _          ' '        ' '   _      '
 '      _       _   ' '     ,       ,    ' '   ,     '
 '   ,     ,     ,    ,    ' '     '
 '        _ _     _  ' '   ,    ,          ,   '
 '   ,     ' '     ,    ' '   _      '
 '   _       _       _           '
 '      _       _       _          ' '  _    '
 '      _      _           _    ' '     _   '
 '     _         _     ' '    ' '  _      '
 '   ,        ' '  _    _         _     ' '   _     '
 '   _       _          ' '          _   '
```

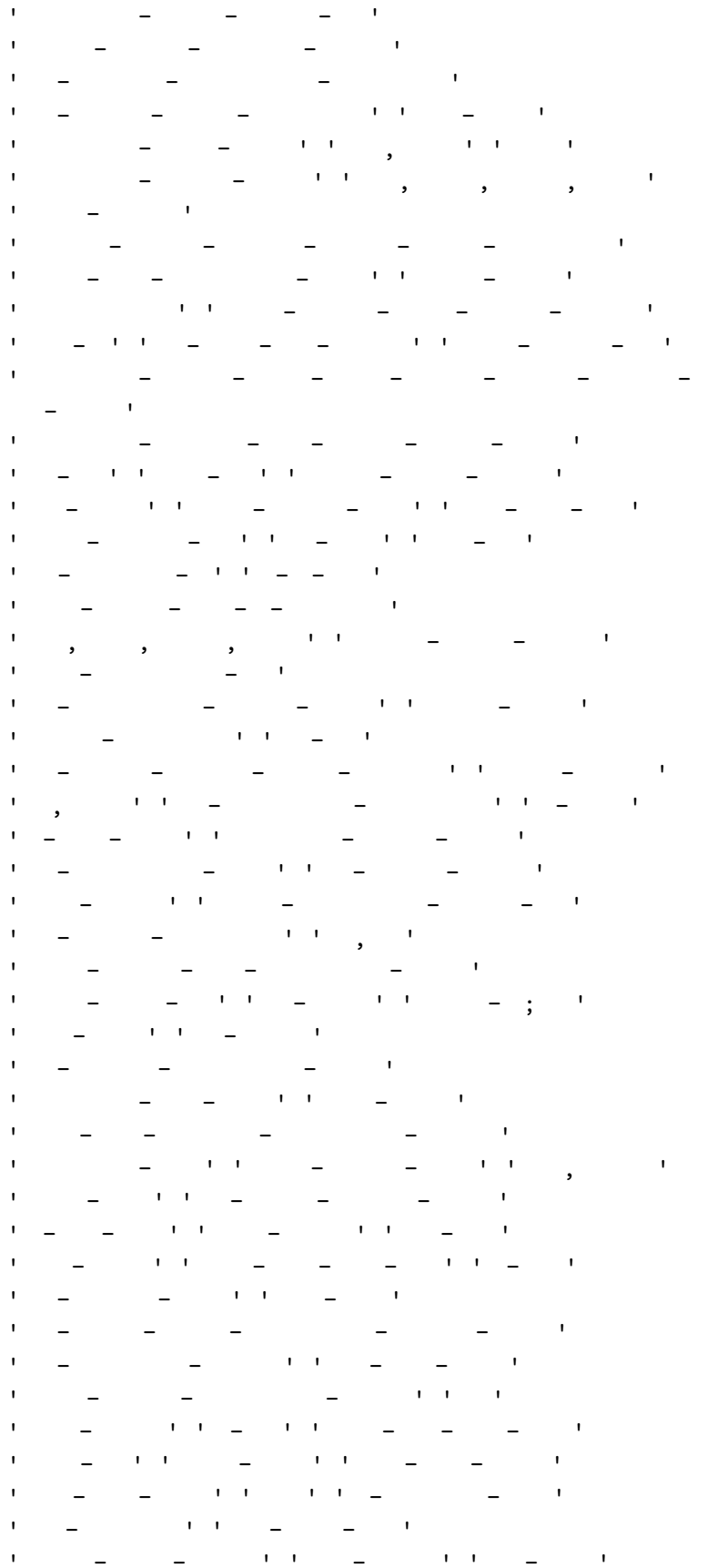


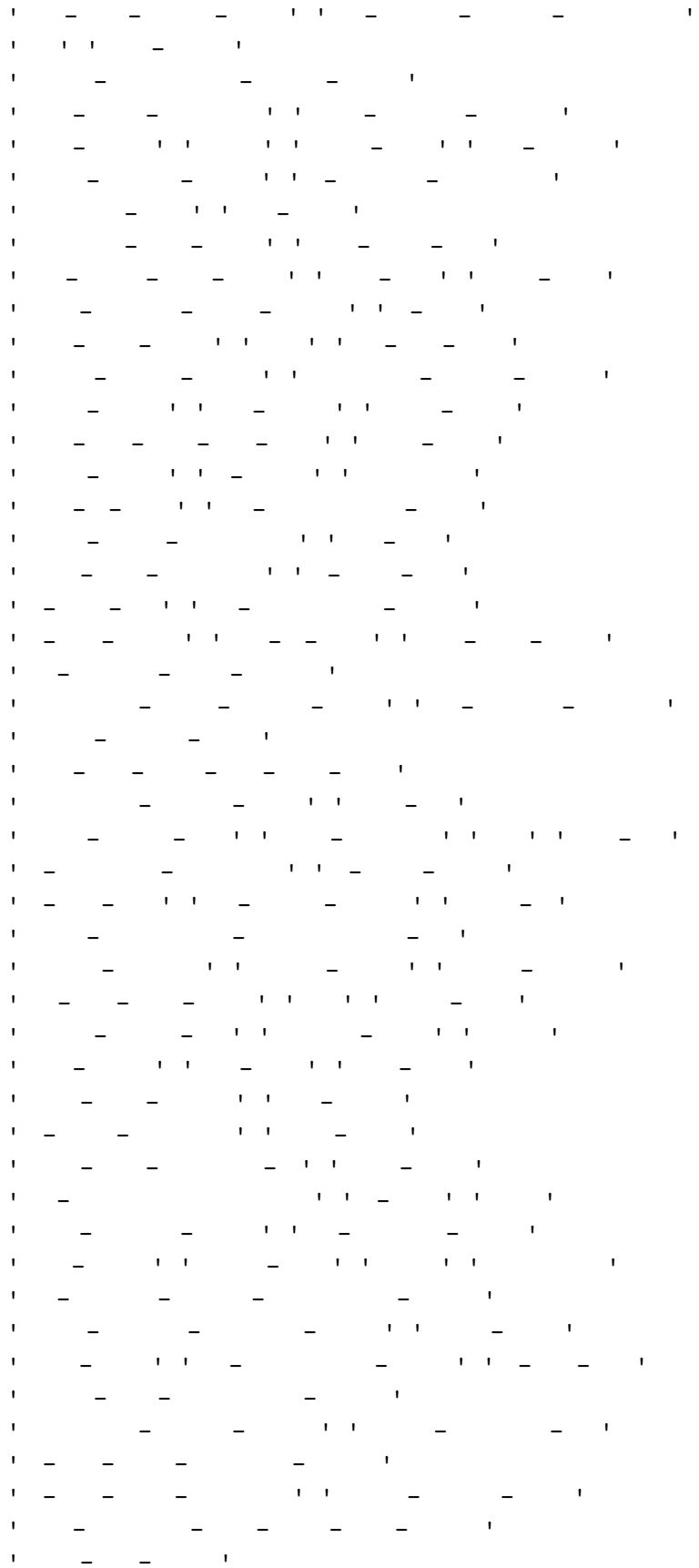


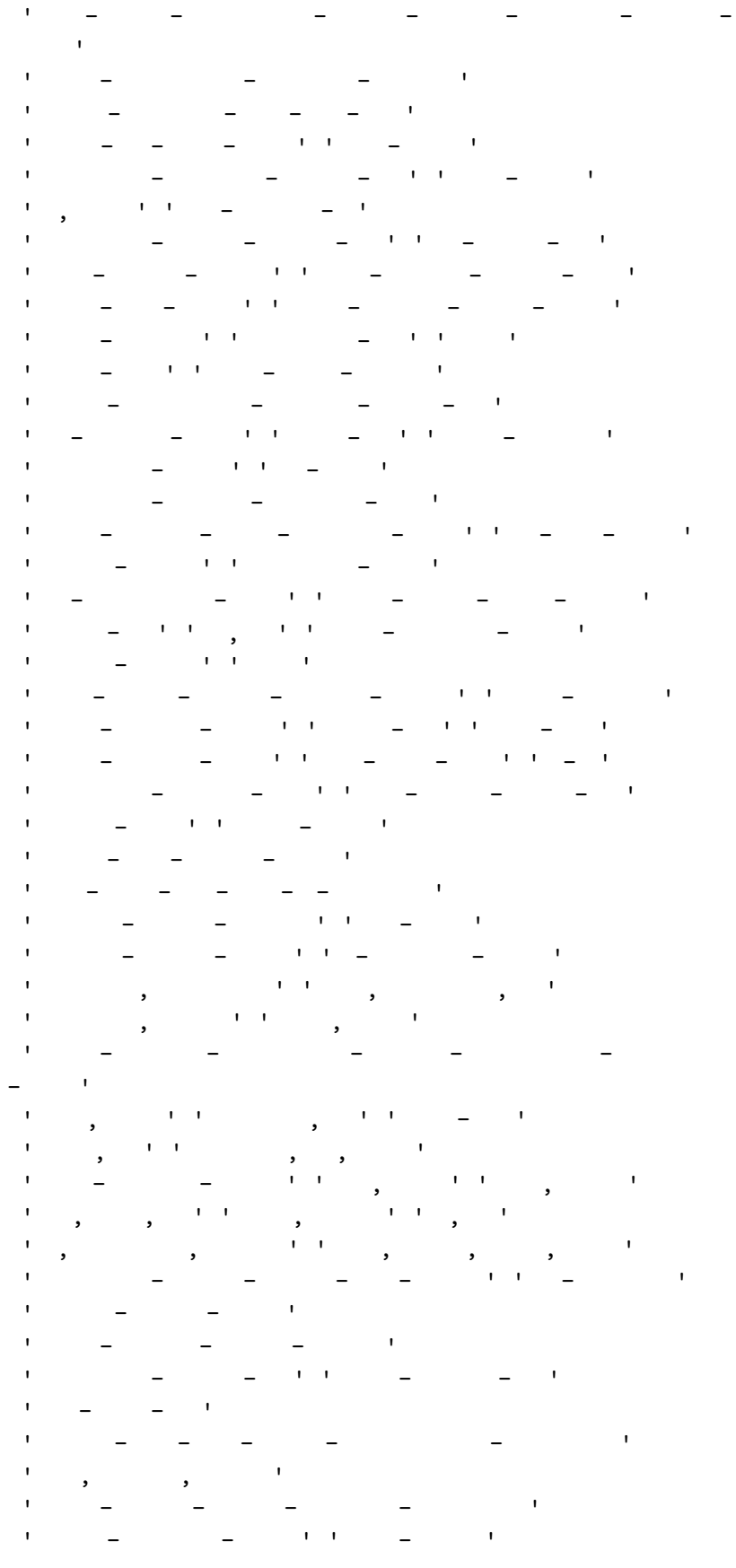


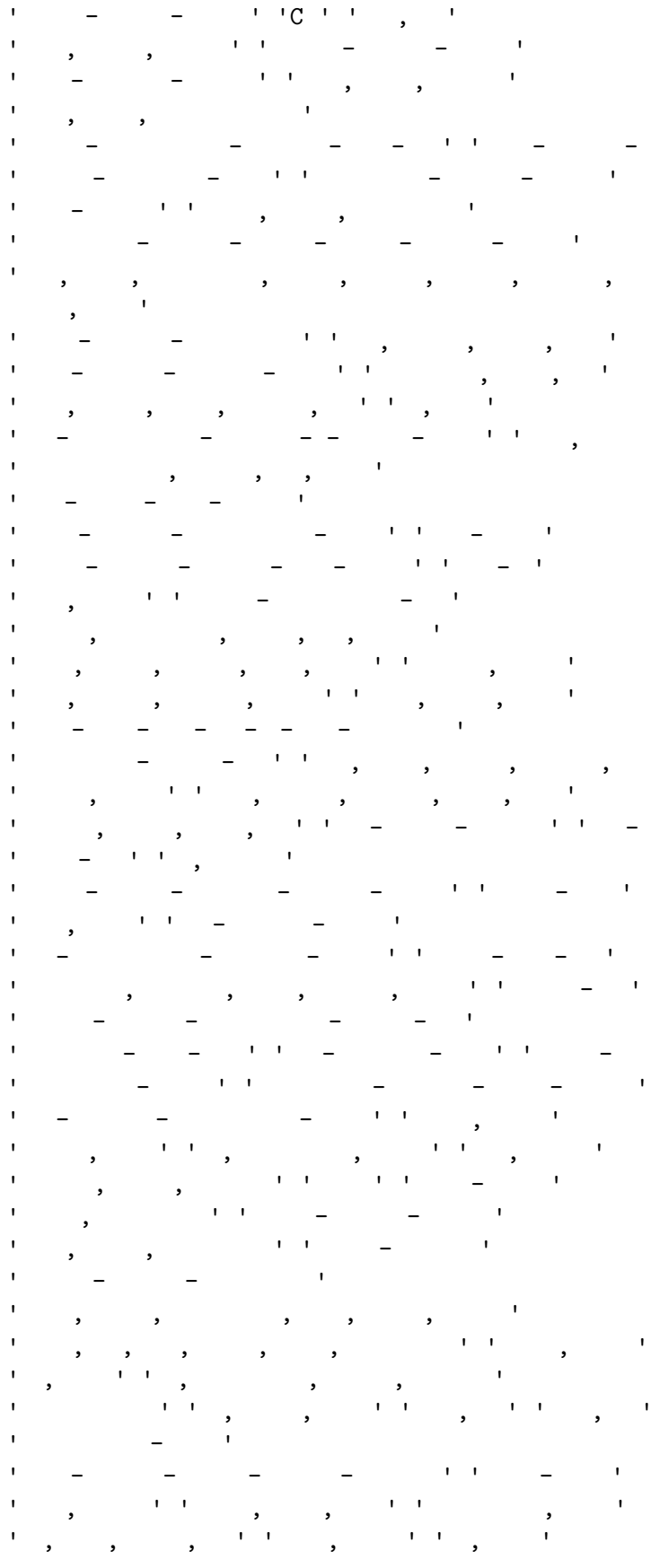


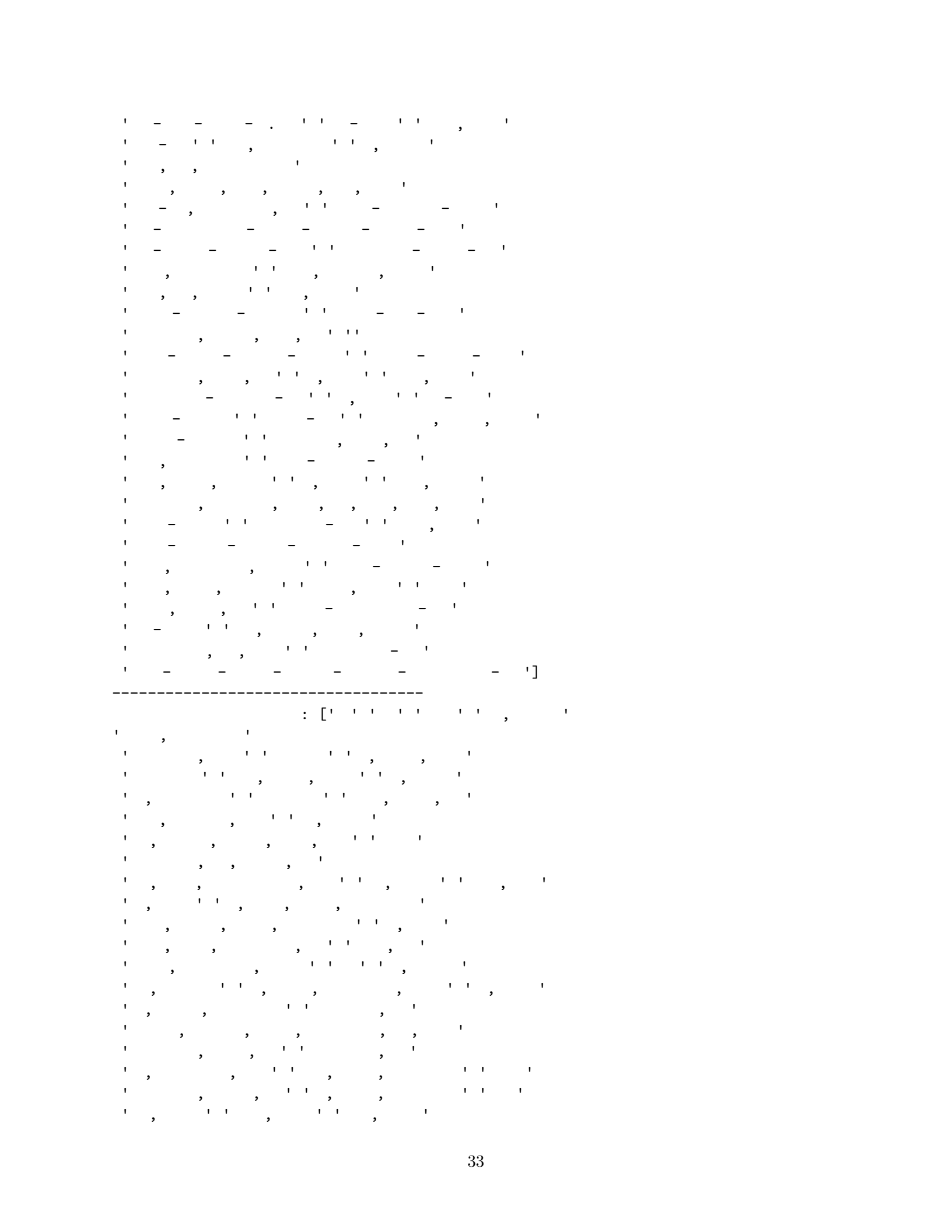

'Unknown'

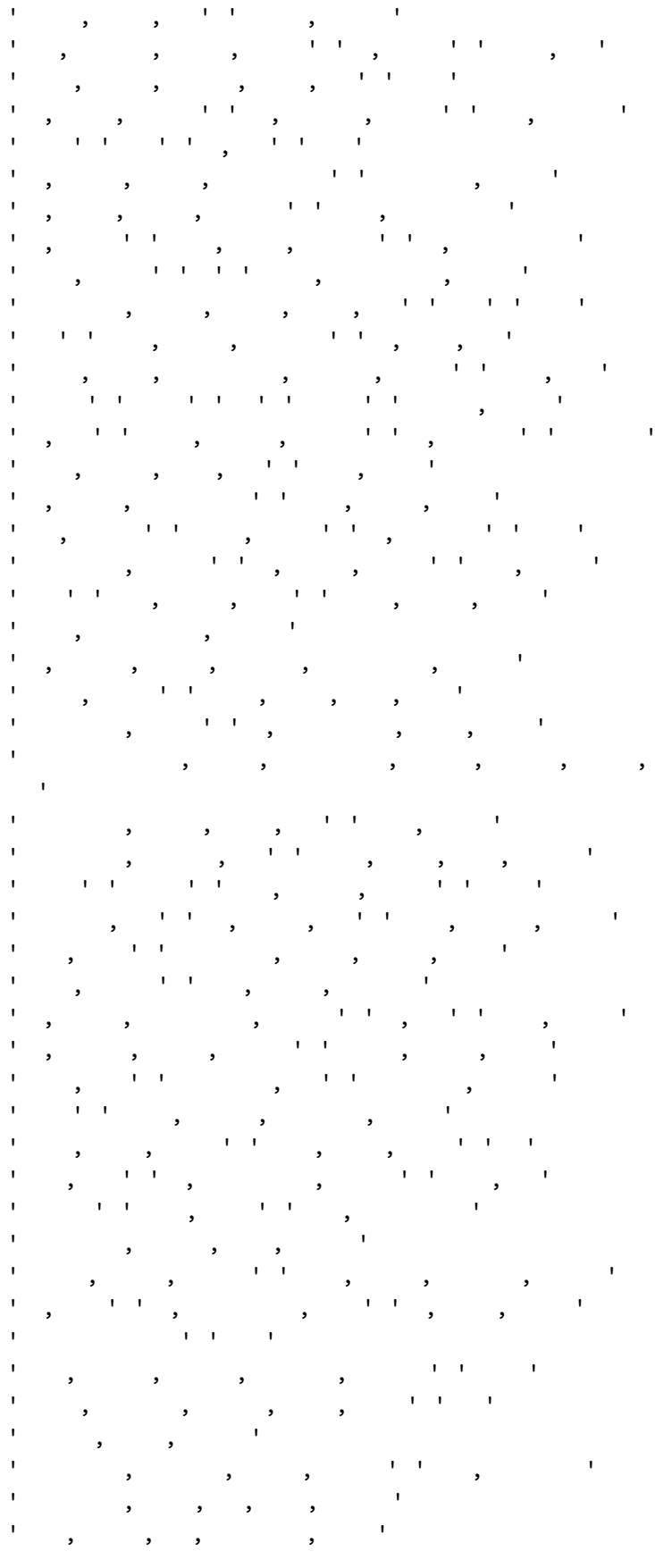


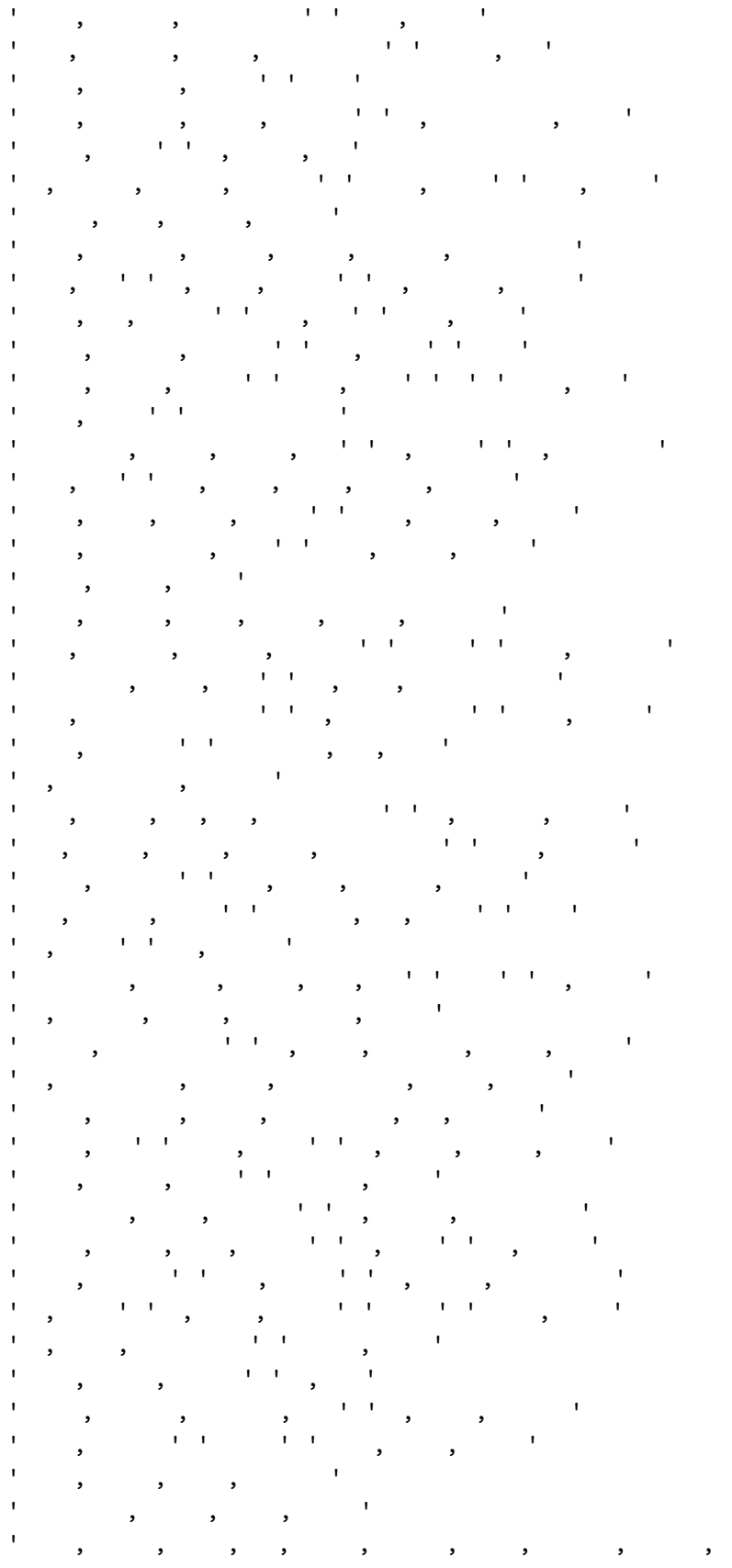


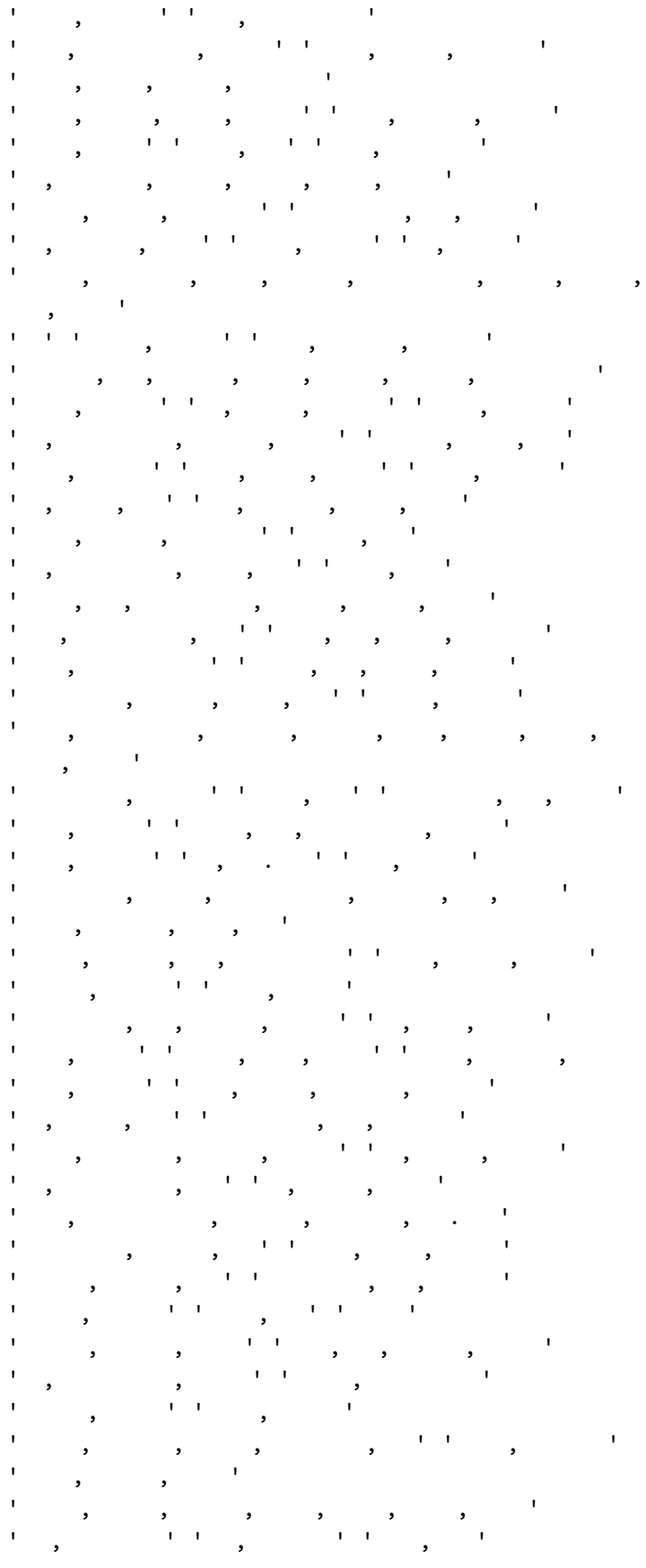


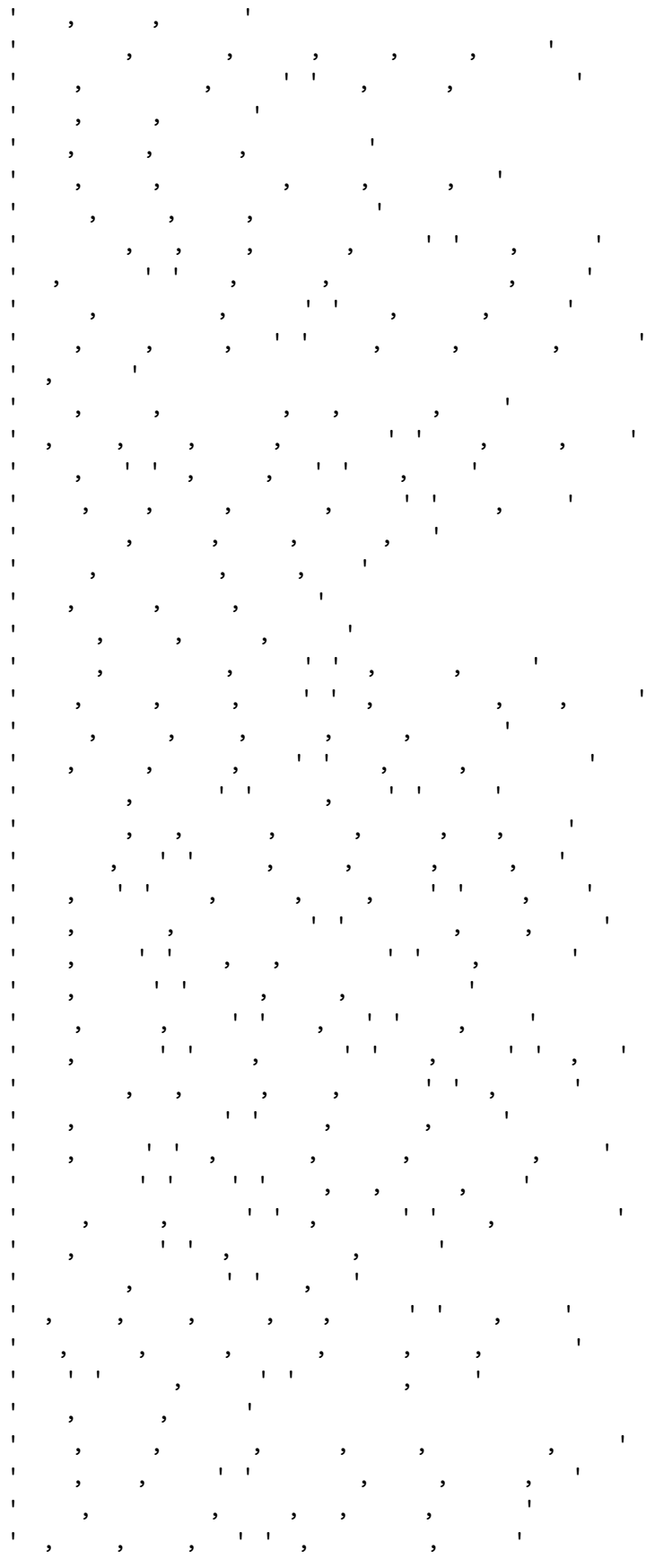




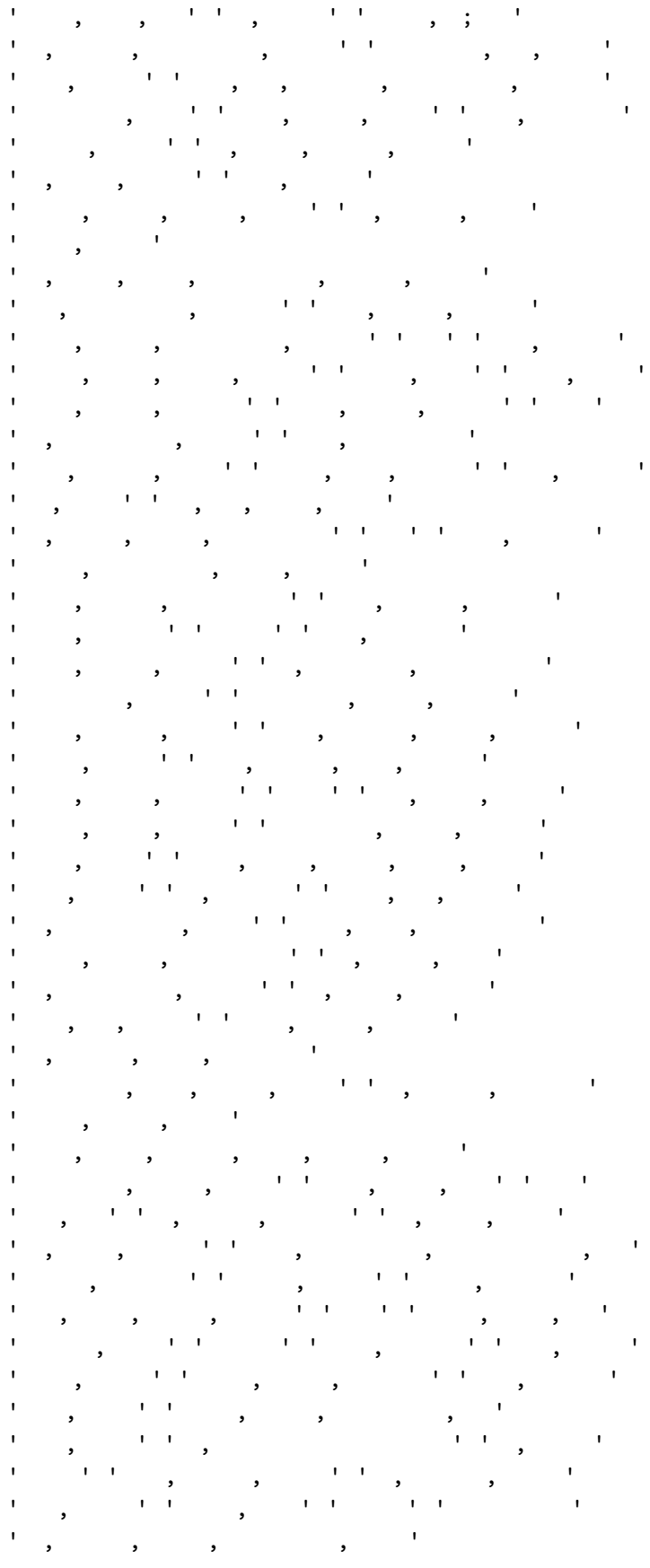


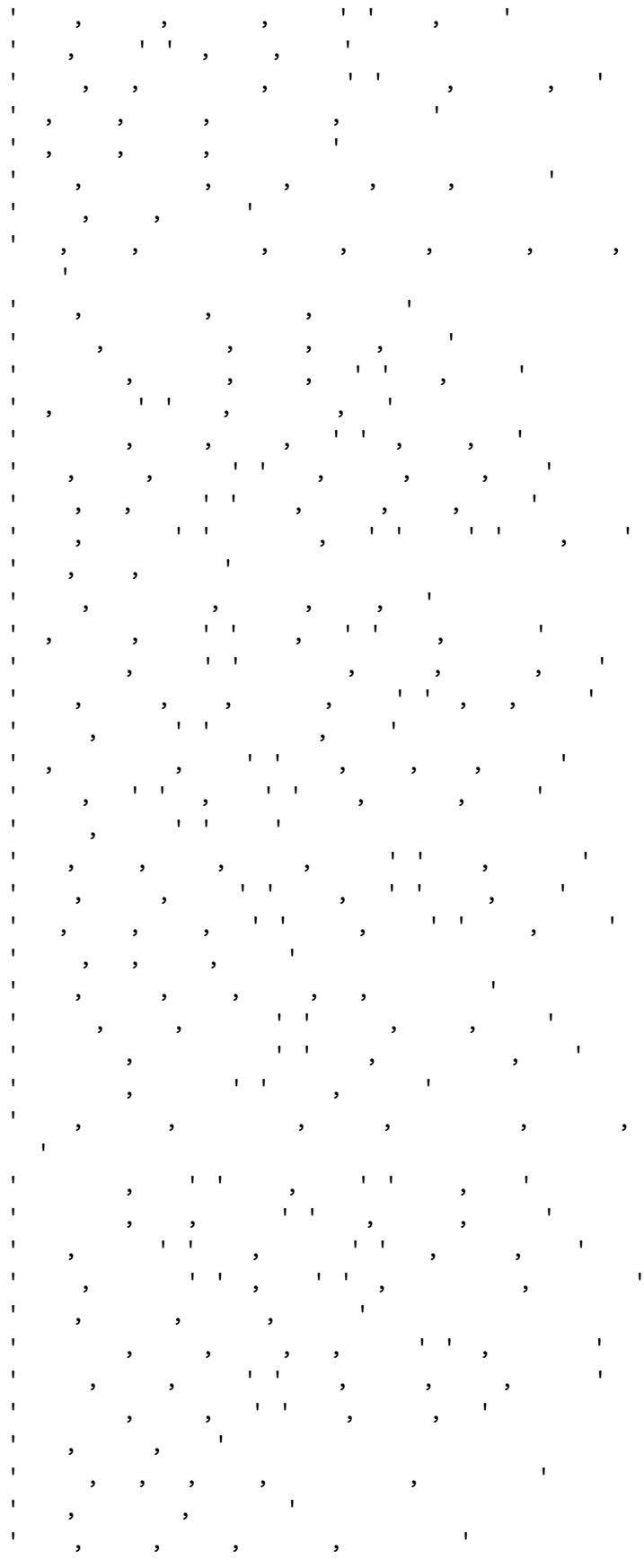


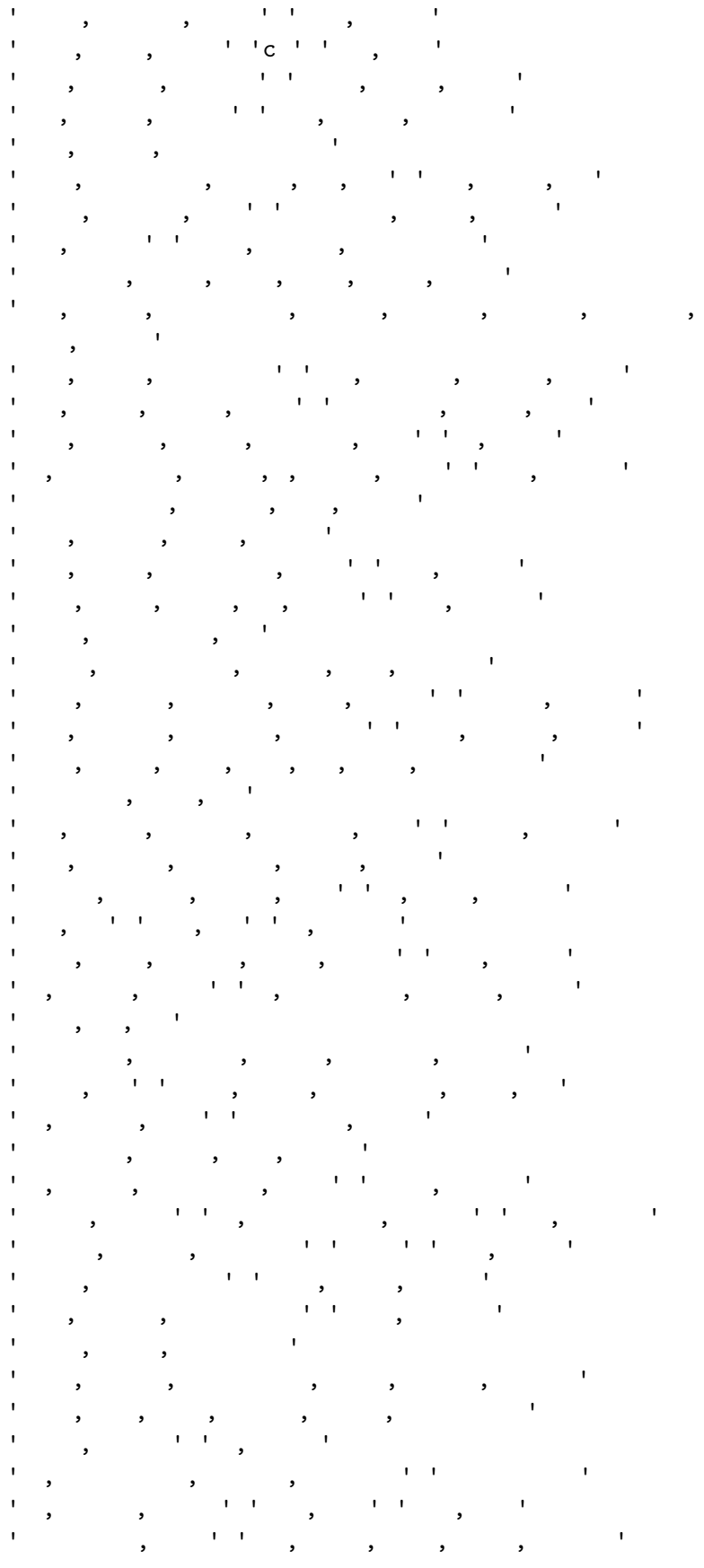


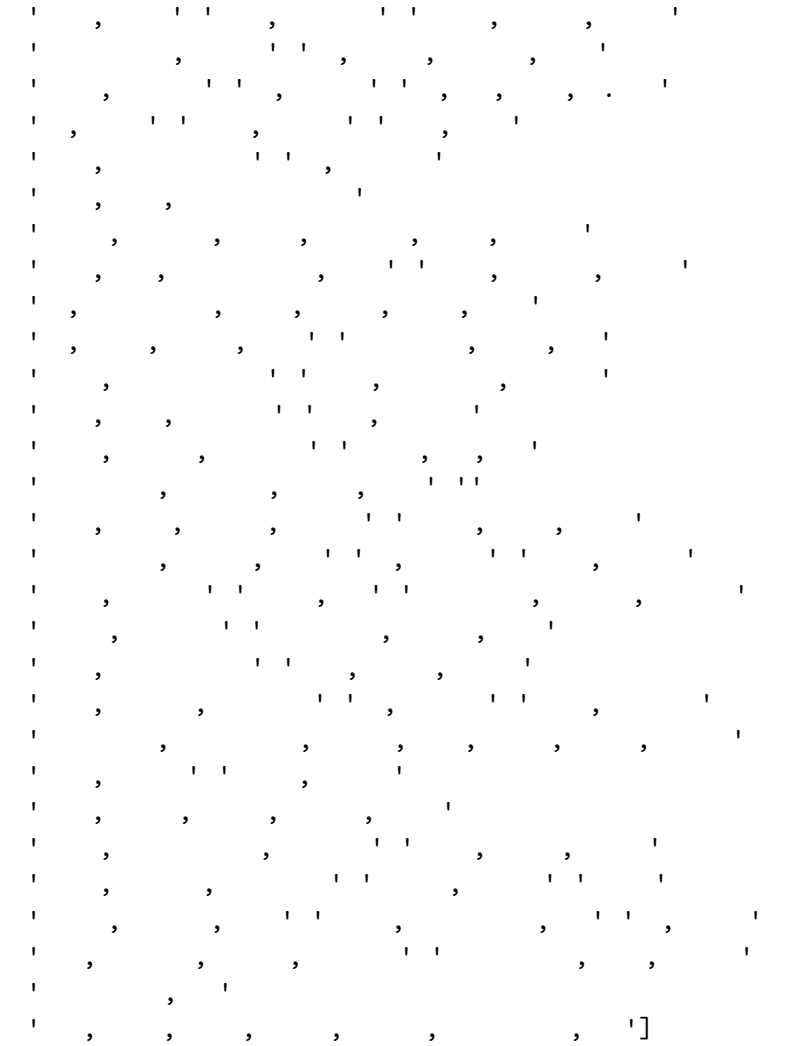


```
' , , '
' , , , , , , , '
' , , . , , , , , '
' , '
' , , , , , , '
' , , , , , , '
' , , ' ' , '
' , , , , , , , ,
 , '
' , , ' ' , , ' ' '
' , ' ' , ' ' , , '
' , ' ' , , , '
' , , , , '
' , , , , , ' 'unknown'
' , , ' ' , '
' , , , ' ' , '
' , , ' ' , , '
' , , ' ' , , ' ' , '
' , , ' ' , , ' ' , '
' , , , , , , '
' , ' ' , , '
' , , , '
' , , , '
' , , , '
' , , , ' ' , '
' , , ' ' , ' ' '
' , , '
' , , , ' ' , '
' , , , , , '
' , , , ' ' '
' , , , , ' ' , '
' , , , ' ' , , '
' , , , , , , , ,
 '
' , , , , , ' ' , '
' , ' ' , , ' ' , '
' , , ' ' , , ' ' , '
' , ' ' , , ' ' , , '
' , , , , '
' , , , ' ' , , '
' , , ' ' , , , '
' , ' ' , ' ' , '
' , , , , ' ' , '
' , ' ' , , ' ' , '
' , , ' ' , , '
' , , ' ' , , ' ' , '
' , , , ' ' , , '
' , ' ' , , , , '
```



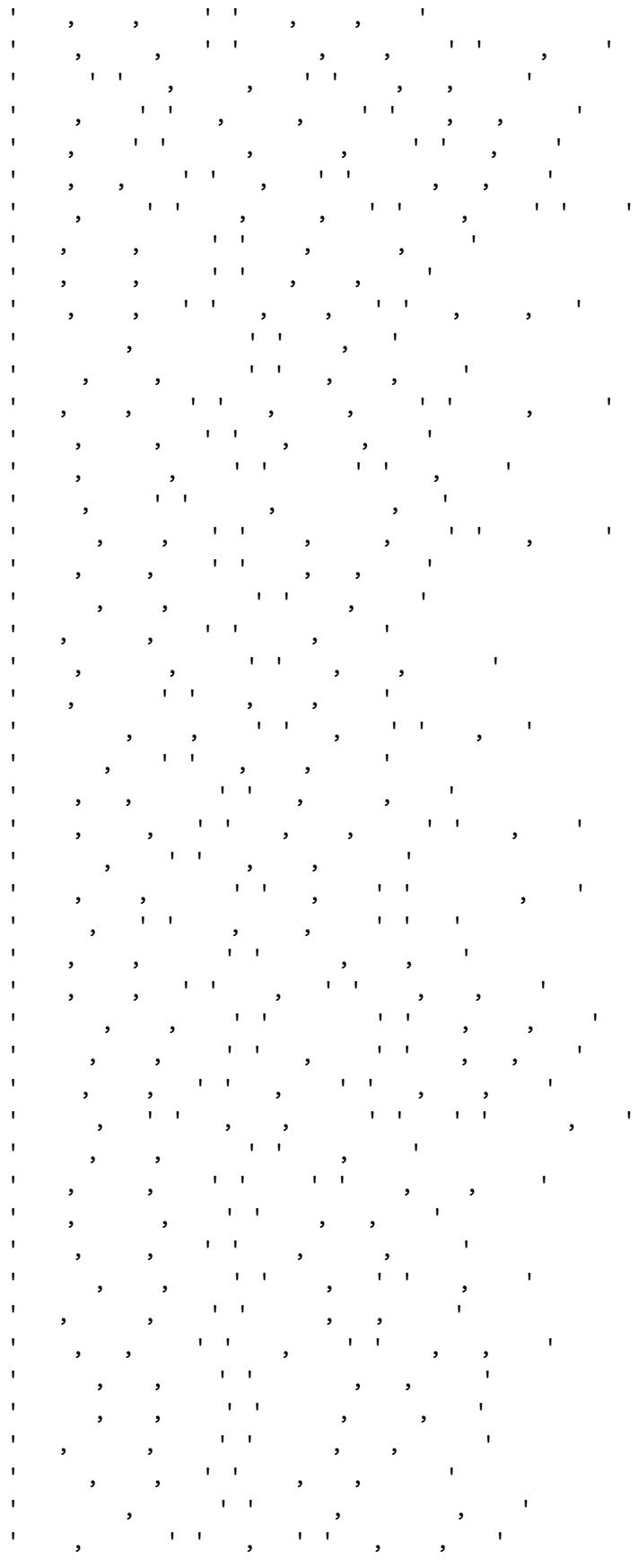


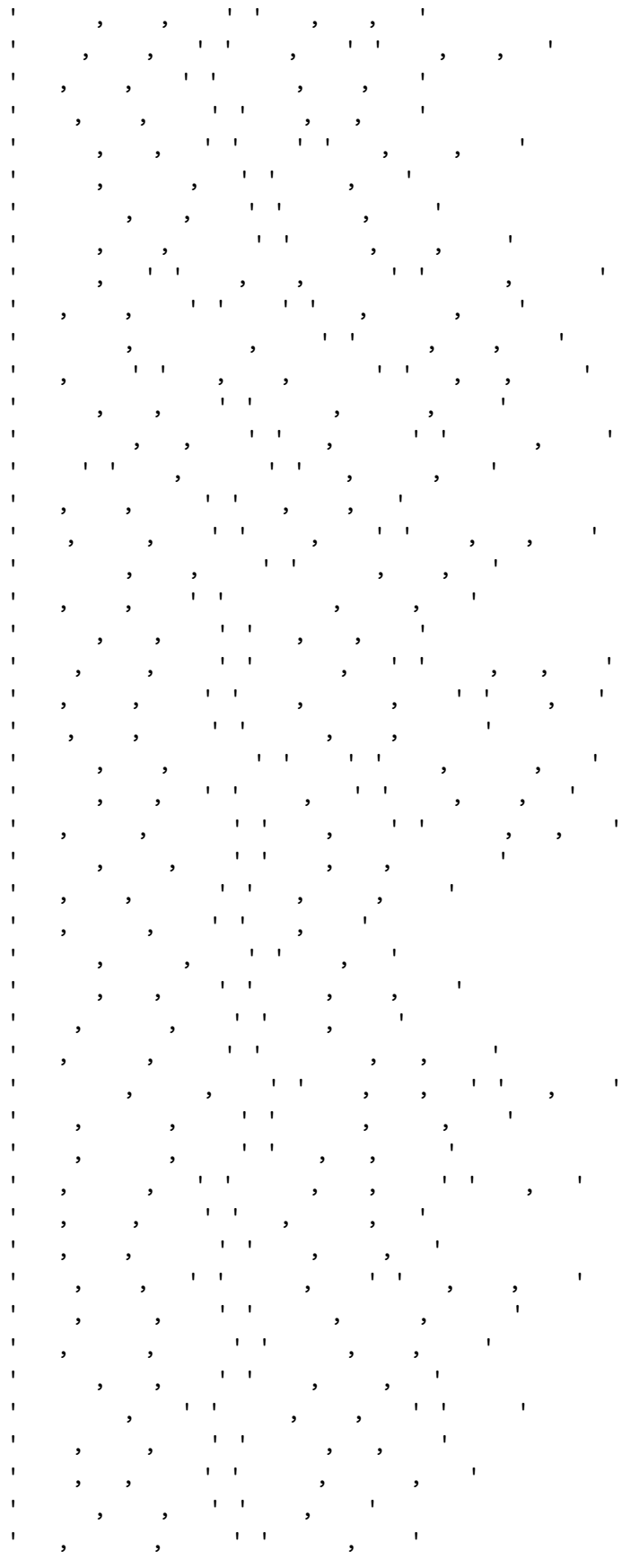


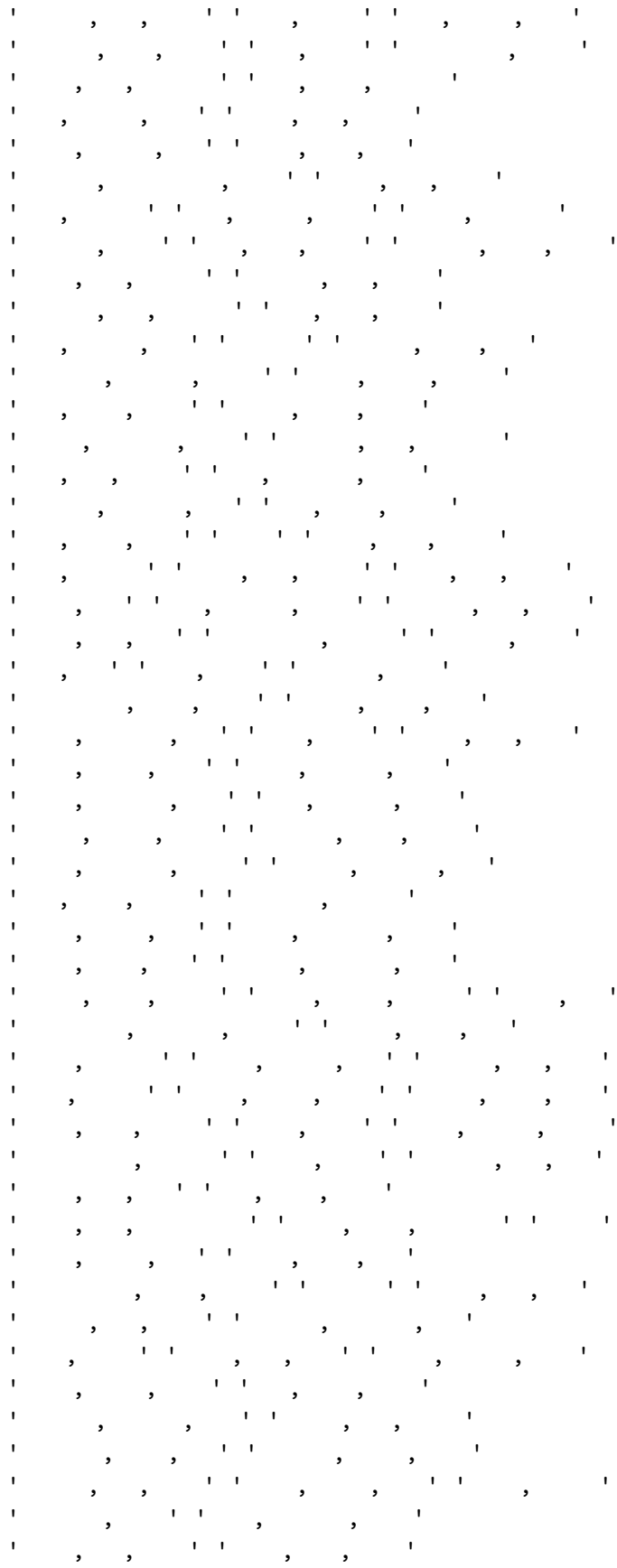


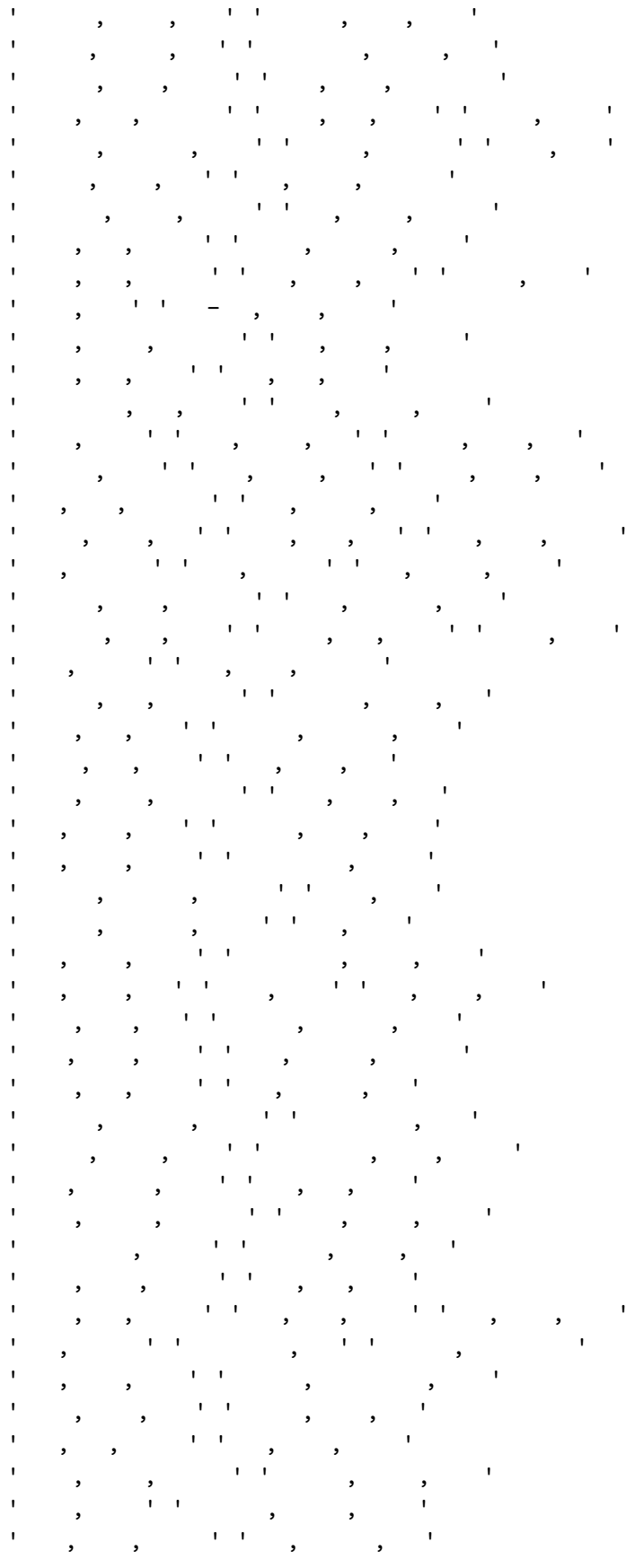
```
------------------------------------

    : director
                    : ['        ' ' .    ' '
      ' … '      '
 ' .    ' ' .    ']
------------------------------------
                     : ['        ' ' .    ' '
      ' … '      '
 ' .    ' ' .    ']
------------------------------------

    : producer
                    : ['        ,         ,        '
'Unknown'
 '       ,         ,     .   ,       ' …
 '       ,       '
 '        ,         ,        ,        ,      '
 '         ']
```

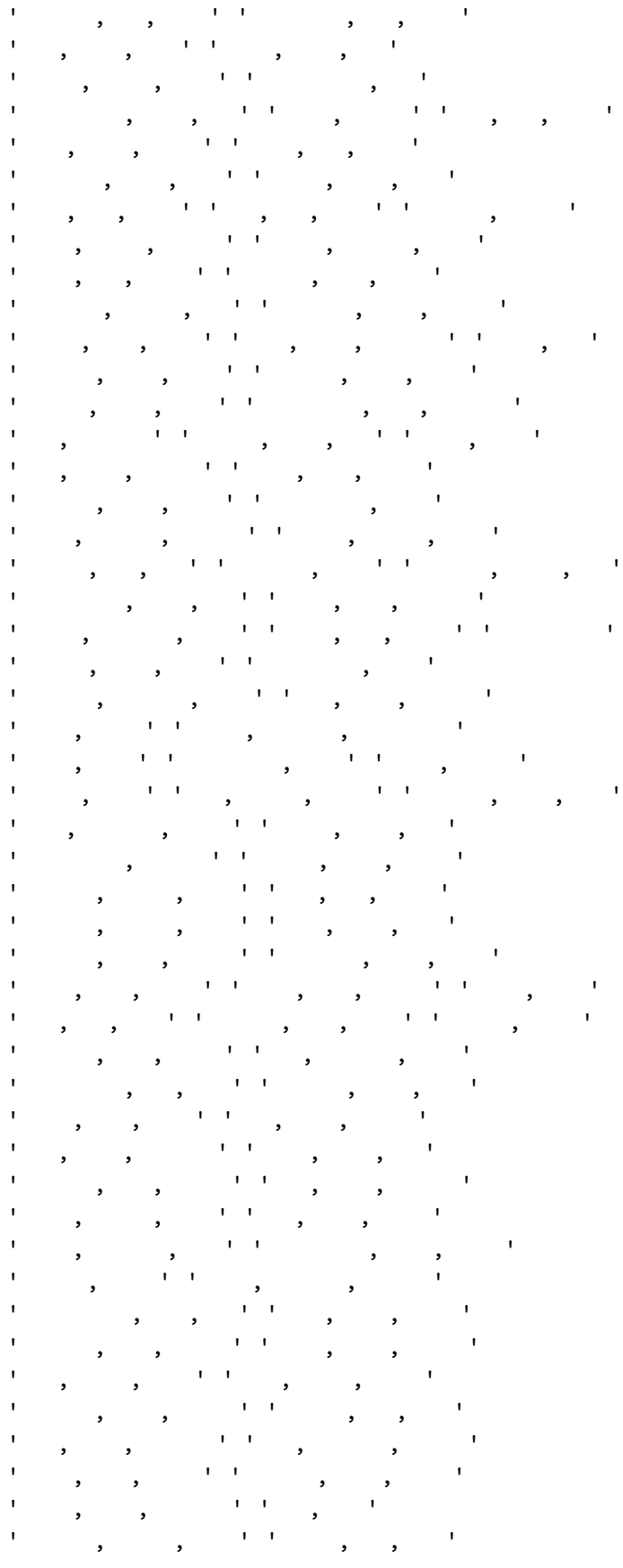
```
----------------------------------
                     : ['       ,          ,
   ' 'unknown'
 '       ,          ,      .   ,       ' …
 '       ,        '
 '        ,          ,         ,         ,       '
 '         ']
----------------------------------

    : age_restriction
                    : ['«18+» -            ', '«6+» -
        6   ', '«12+» -            12   ', '«16+» -
    16   ', '«0+» -                   ']
Categories (5, object): ['«0+» -                   ', '«12+» -
        12   ', '«16+» -            16   ', '«18+» -
   ', '«6+» -           6   ']
----------------------------------
                     : ['«18+»,            ' '«6+»,
        6   '
 '«12+»,            12   ' '«16+»,           16   '
 '«0+»,                    ']
----------------------------------

    : financing_source
                    : [NaN, '           ', '     ',
'            ,      ']
Categories (3, object): ['           ', '           ,
   ', '     ']
----------------------------------
                     : [nan '           ' '     '
 '            ,       ']
----------------------------------

    : genres
                    : ['    ,   ,     ' '   ,    '
'      ,   ,    '
 '   ' '      ,    ' '    ,     ,    ' '    ,     '
 '    ,    ,     ' '    ,   ,       ' '    ,   '
 '     ,   ,     ' '     ,      ' '    ,     ,    '
 '    ,      ' '    ,     ' '    ,      ,     '
 '     ,     ,     ' '    ' '    ,    ,     '
 '    ,     ' '    ,      ,     '
 '      ,         ,       ' '    ,     '
 '      ,      ,     ' '    ,    ,     '
 '    ,     ,      ' '    ,     ,     '
 '      ,   ,     ' '    ,     ' 'Unknown'
 '     ,      ,    ' '          ,    '
 '    ,     ,     ' '    ,     ,    ' '    ,    '
```

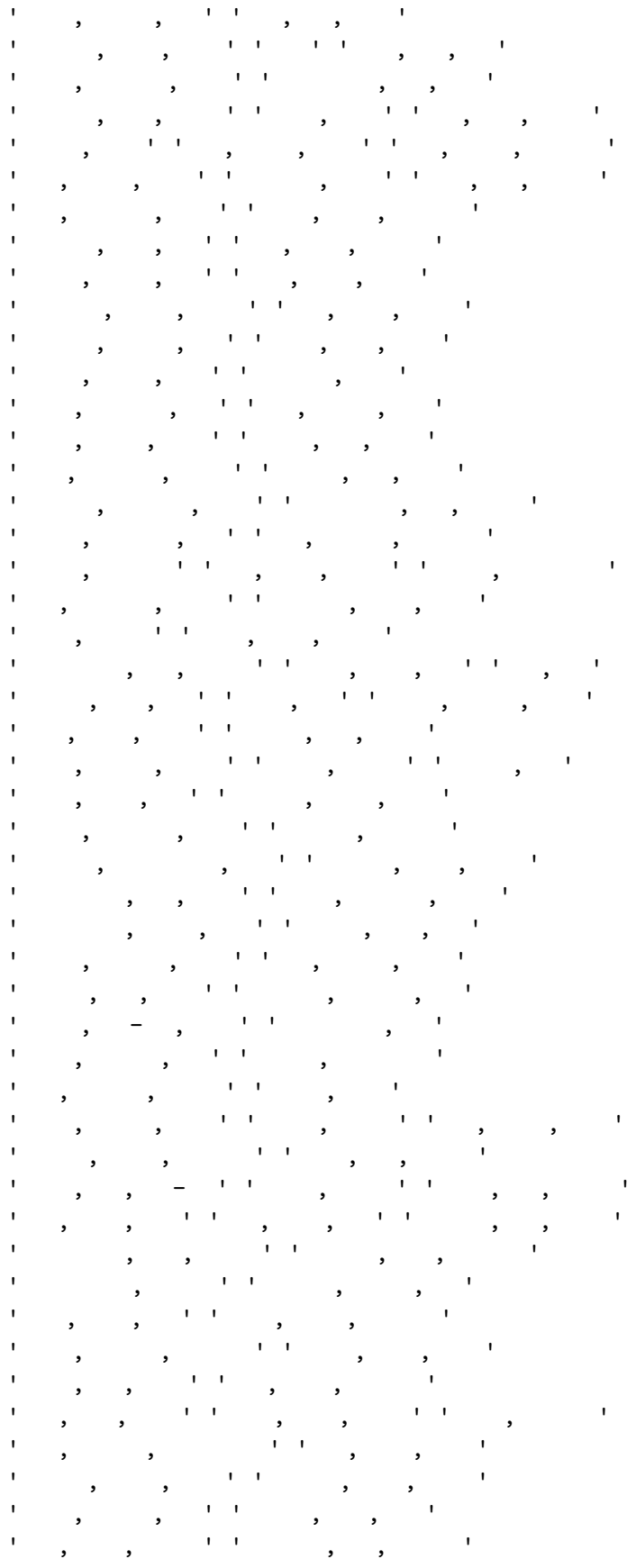


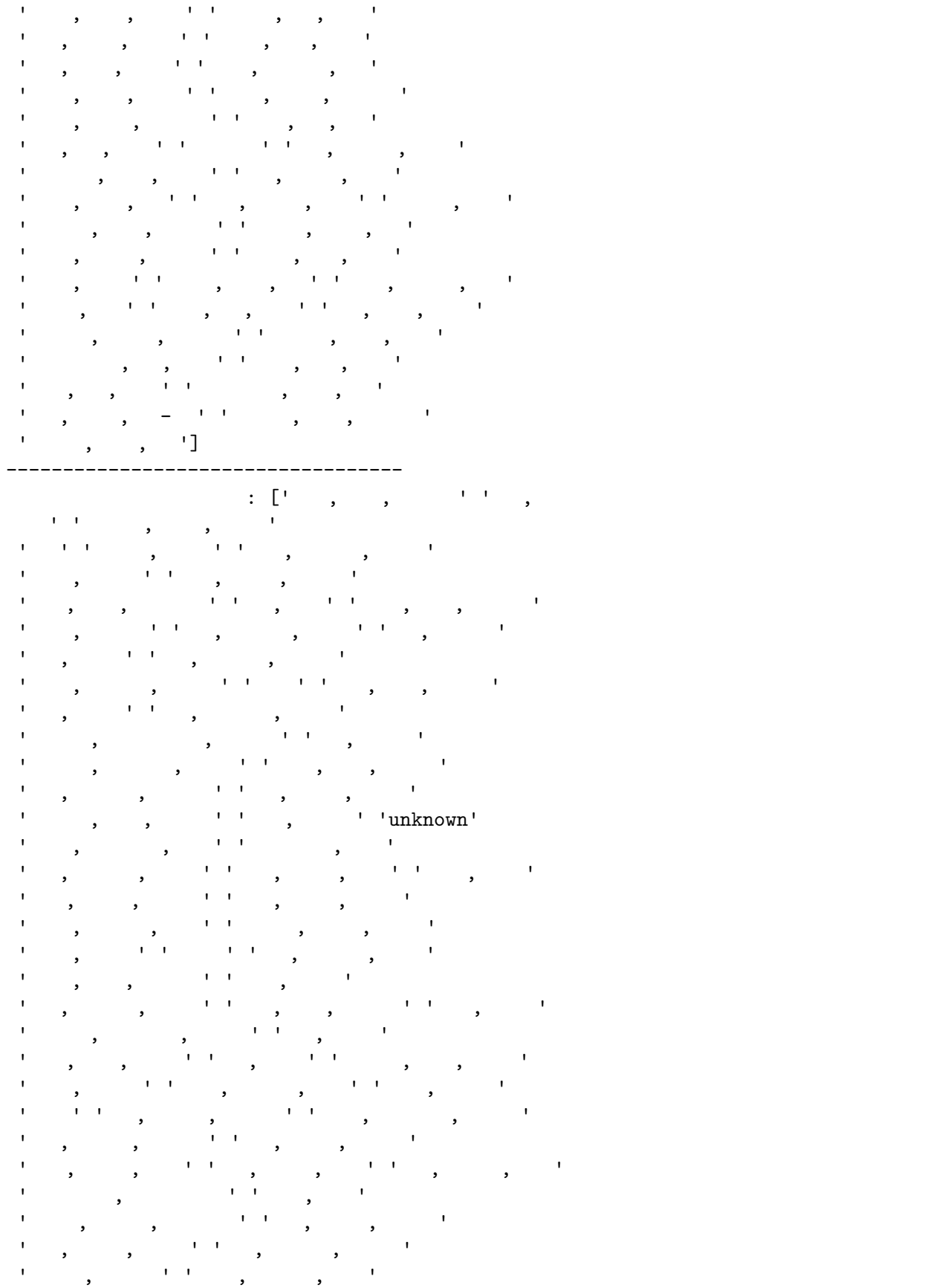


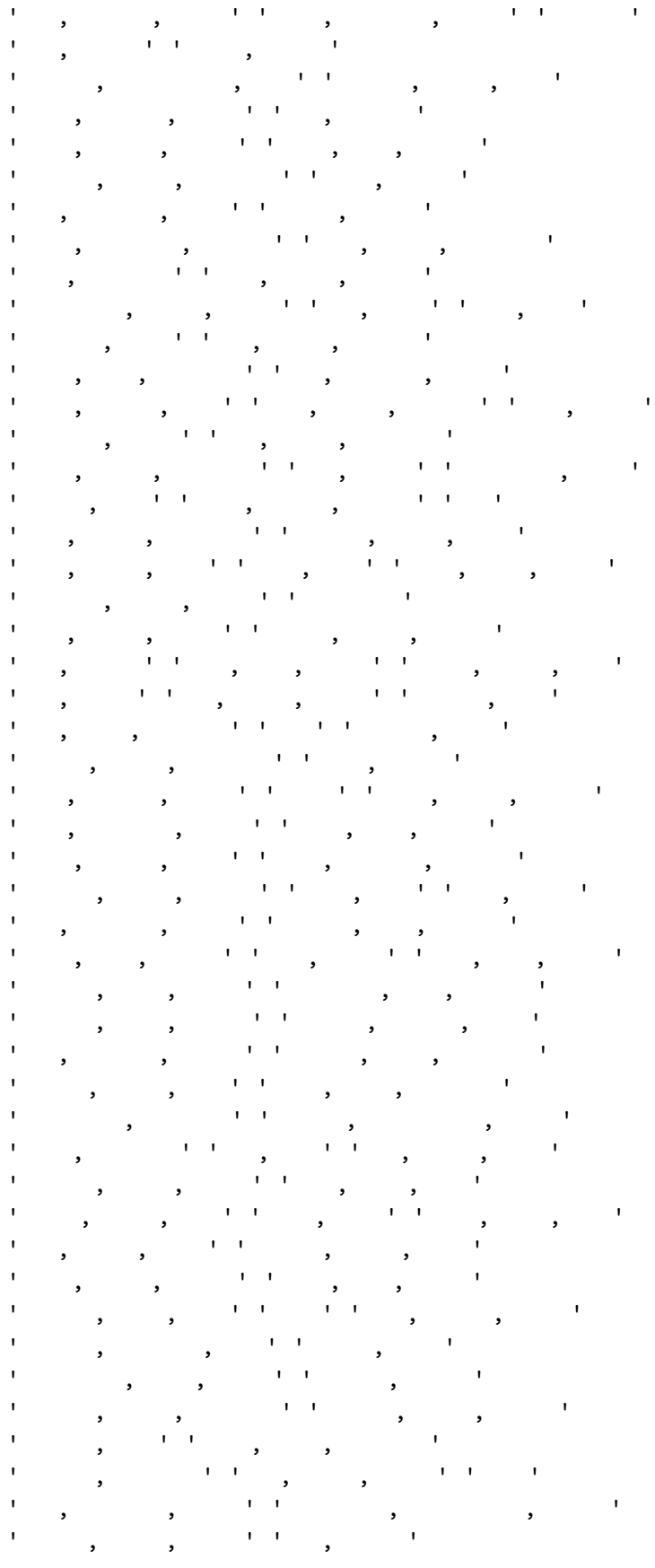


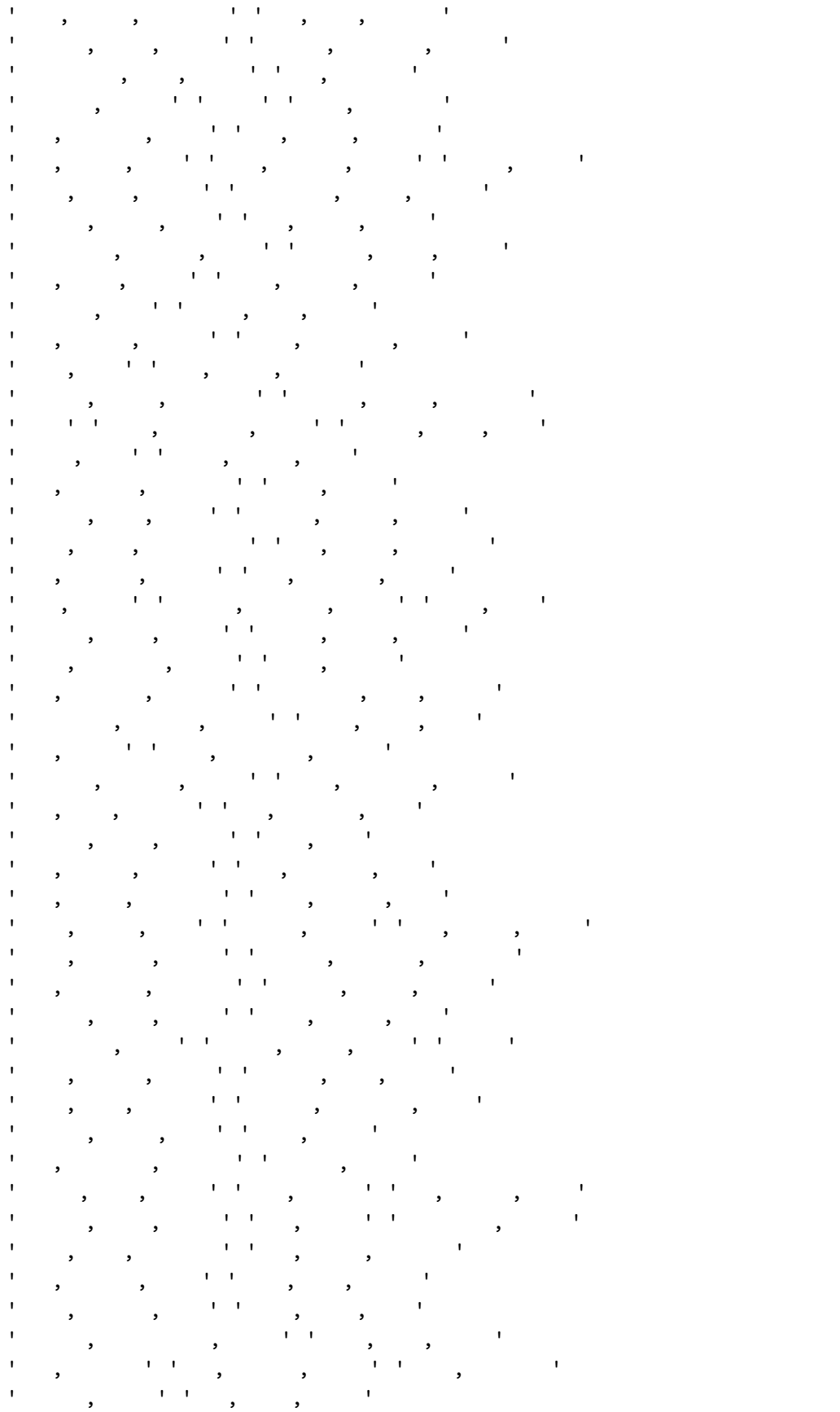


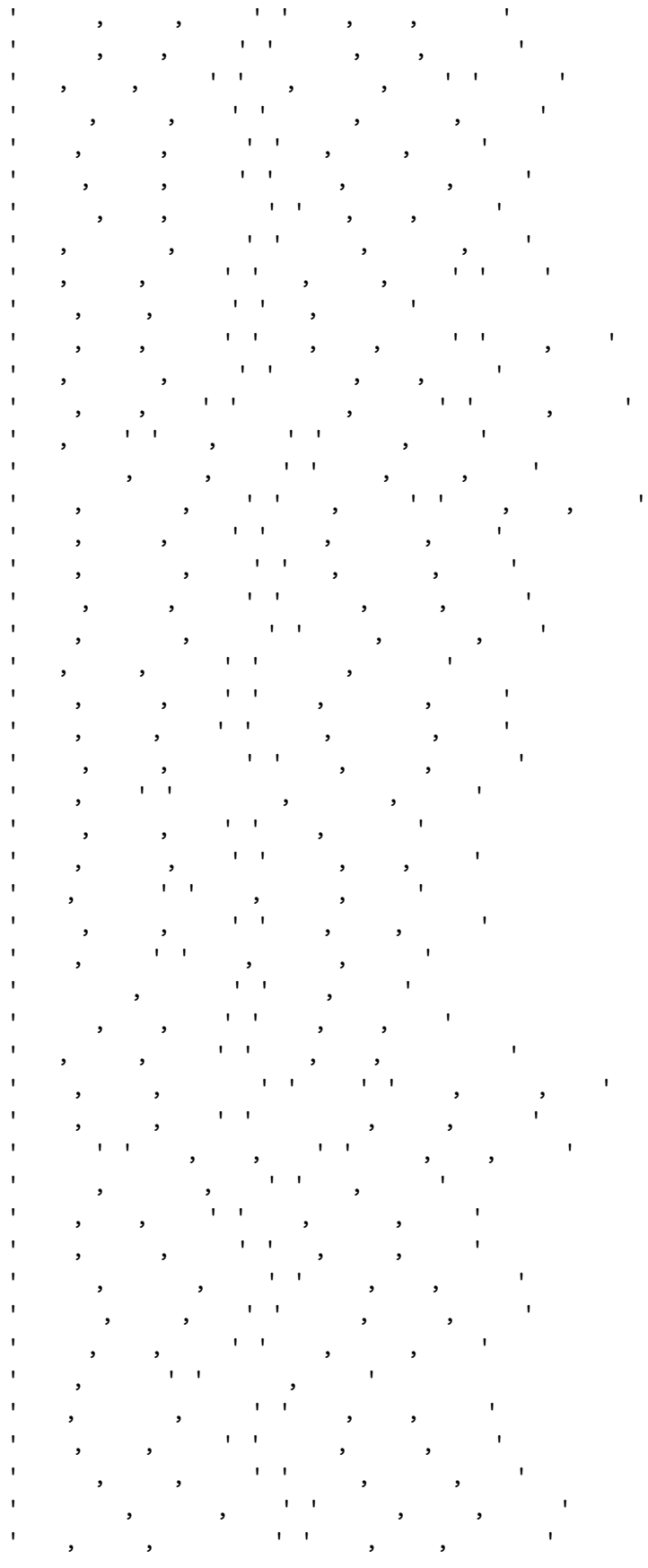


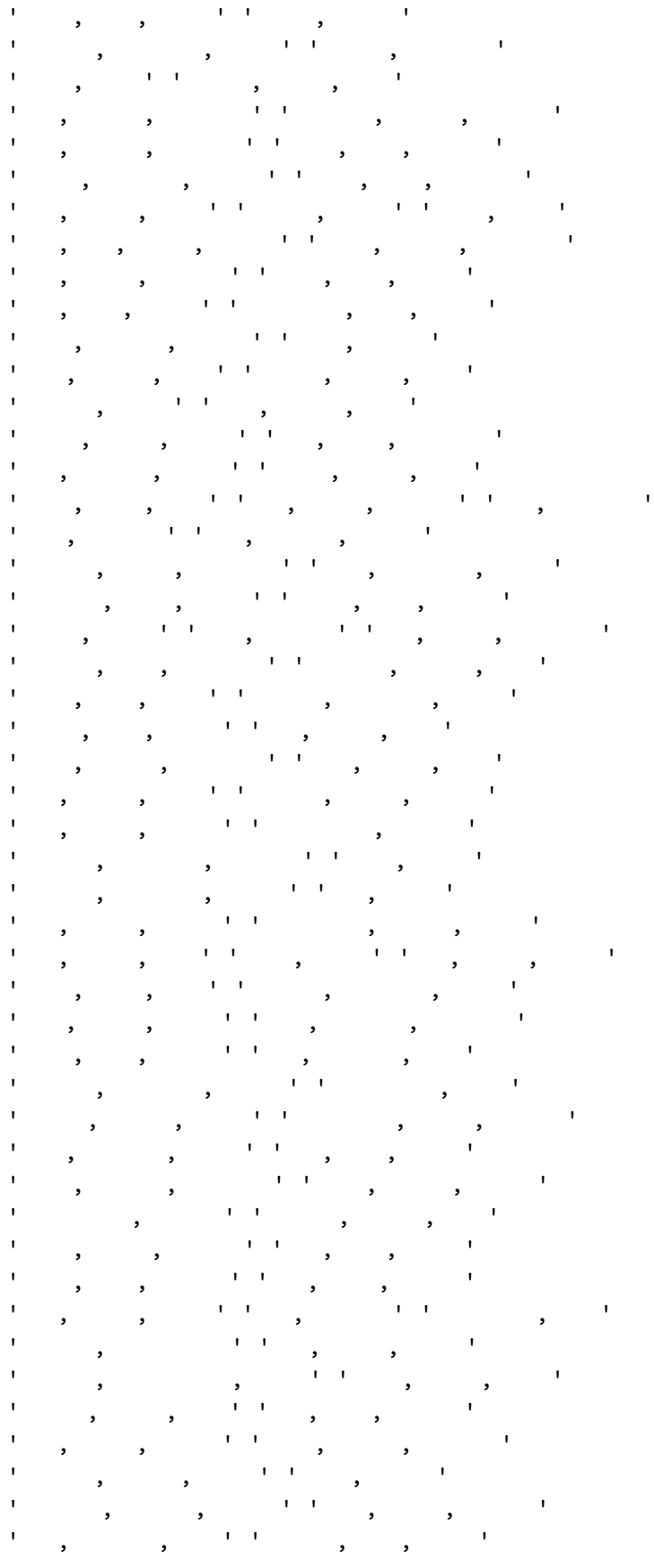


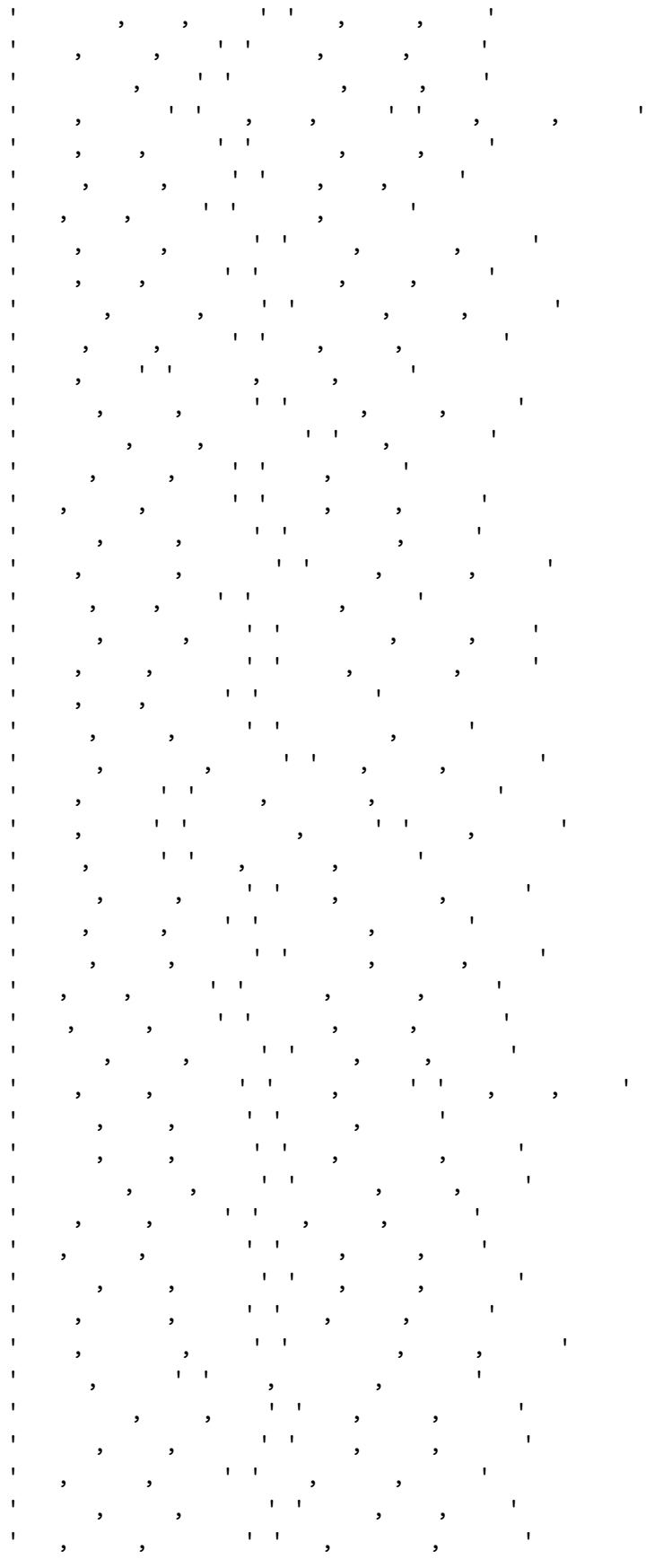


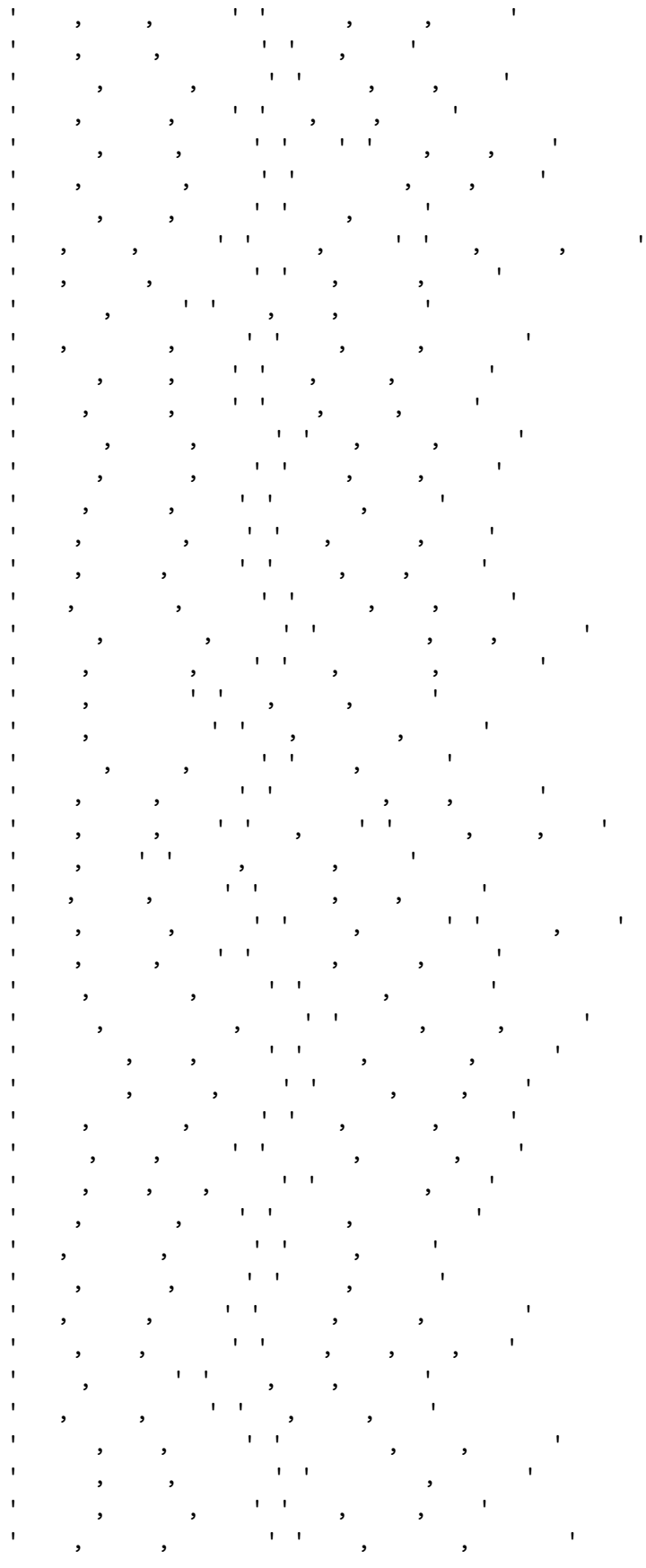


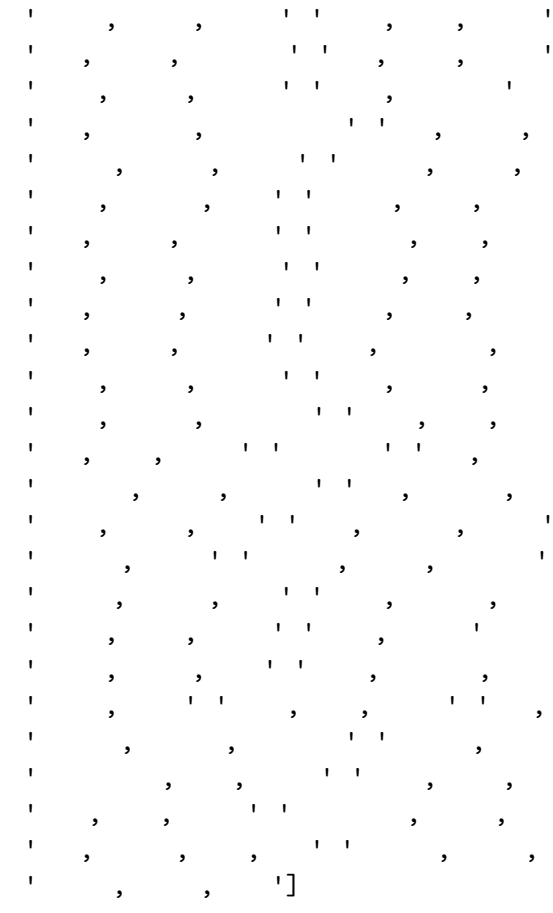












```
----------------------------------
```

, , ‘production_country’, ,

- . (“-”), (” - “), (‘,’), (”-“).

2.5.

- , . ?

```
<div class="card">
    <div class="card-header" id="headingThree">
        <button class="btn btn-link collapsed" data-toggle="collapse" data-target="#collapseHi
    </div>
    <div id="collapseHint_budget" class="collapse" aria-labelledby="headingThree" data-parent=
        <div class="card-body">
```

budget. : refundable_support nonrefundable_support?

```
    </div>
</div>
```

[ ]:
```
merged_data.describe()
```

[ ]:
```
             puNumber  refundable_support  nonrefundable_support  \
count        7,486.00              332.00                 332.00
mean   135,178,527.64       11,864,457.83          48,980,988.89
std     38,382,177.41       24,916,555.26          59,980,117.92
min             0.00                0.00                   0.00
25%   112,025,043.00                0.00          25,000,000.00
50%   121,015,464.50                0.00          30,000,000.00
75%   124,003,313.75       15,000,000.00          40,375,000.00
max   231,001,111.00      180,000,000.00         400,000,000.00

                budget  ratings       box_office
count           332.00 6,519.00         3,158.00
mean    127,229,716.68     6.49    76,478,696.16
std     188,588,333.12     1.11   240,353,122.82
min               0.00     1.00             0.00
25%      42,000,000.00     5.90        86,239.00
50%      68,649,916.00     6.60     2,327,987.55
75%     141,985,319.50     7.20    23,979,671.02
max   2,305,074,303.00     9.90 3,073,568,690.79
```

[ ]:
```
display(merged_data.query('budget <= 0'))
```

```
                     title   puNumber           show_start_date  \
2053                   14+  111023614 2014-12-19 12:00:00+00:00
2058               xless 2  111000415 2015-01-26 12:00:00+00:00
2472               (2015)  111017315 2015-09-28 12:00:00+00:00
2532                   .  111009615 2015-05-26 12:00:00+00:00
2615       .          114001715 2015-07-21 12:00:00+00:00
2684                  111010215 2015-06-08 12:00:00+00:00
2788                  111017115 2015-09-30 12:00:00+00:00
2874          ,      !  111004916 2016-03-01 12:00:00+00:00
3047                111008216 2016-04-29 12:00:00+00:00
3565             (2016)  111018116 2016-12-16 12:00:00+00:00
3675                 ,     114001817 2017-06-27 12:00:00+00:00
3739                       111004317 2017-06-07 12:00:00+00:00
5673                       111012117 2018-02-01 12:00:00+00:00
5806                      114001018 2018-05-31 12:00:00+00:00
6008                   114003018 2018-08-23 12:00:00+00:00
6323                       114006518 2018-11-08 12:00:00+00:00
6823                   114002419 2019-04-20 12:00:00+00:00

               type                                          film_studio  \
2053                                          "     "
2058             "       "   ",     "         …
2472             "          ",     "           "
2532                   "      ",     "           "
2615              "          ",     "           "
2684                              "        "     "
```

```
2788                    "      ",     "           "
2874                                                "     "
3047                    "      ",     "           "
3565                "     2011",      "           "
3675                                                "     "
3739                    "     ",      "           "
5673              "       "   ",      "          …
5806                                          "           "
6008                                          "           "
6323                                           "           "
6823                                          "           "

     production_country                              director  \
2053                                                  .
2058                                                 .
2472                                                 .
2532                                                .
2615                                                   .
2684                                                 .
2788
2874                                                 .
3047                                                    .
3565                                                 .
3675                                                   .
3739                                                .
5673                                                 .
5806                  .     ,   .    ,   .
6008                                                 .
6323                                                    .
6823                             .    ,   .

                                                      producer  \
2053                                  .    ,   .
2058                .    ,   .     ,   .
2472                                  "           "
2532                                               .
2615                                                .
2684                                 .     ,   .
2788             .     ,   .     ,   .
2874                                              "     "
3047             .     ,   .     ,   .
3565                                 "           "
3675                                              "     "
3739   .     ,   .     ,   .   ,   .   …
5673                .    ,   .     ,   .
5806                                 "           "
6008                                 "           "
6323     "           ",     "        …
```

```
6823                              "           "

                  age_restriction  refundable_support  \
2053  «16+»,            16                   0.00
2058  «16+»,            16                   0.00
2472  «12+»,            12          10,000,000.00
2532  «16+»,            16          20,000,000.00
2615    «6+»,            6          60,000,000.00
2684      «18+»,                               0.00
2788  «16+»,            16                   0.00
2874  «12+»,            12                   0.00
3047  «16+»,            16          19,000,000.00
3565  «12+»,            12                   0.00
3675    «6+»,            6          20,000,000.00
3739  «16+»,            16          25,000,000.00
5673  «16+»,            16          20,000,000.00
5806    «6+»,            6          10,000,000.00
6008    «6+»,            6          40,000,000.00
6323    «6+»,            6                   0.00
6823    «6+»,            6          60,000,000.00

      nonrefundable_support  budget                    financing_source  \
2053          23,000,000.00    0.00                ,
2058          75,000,000.00    0.00
2472          70,000,000.00    0.00
2532                  0.00     0.00
2615         100,000,000.00    0.00
2684          59,000,000.00    0.00
2788          35,000,000.00    0.00
2874           6,000,000.00    0.00
3047          51,000,000.00    0.00
3565         250,000,000.00    0.00
3675          20,000,000.00    0.00
3739          20,000,000.00    0.00
5673          40,000,000.00    0.00
5806          60,000,000.00    0.00
6008          40,000,000.00    0.00
6323         138,969,465.00    0.00
6823         120,000,000.00    0.00

      ratings                              genres       box_office
2053     6.60                                    10,234,016.10
2058     6.60                                     446,163,511.00
2472     7.80                 ,    ,         196,572,438.40
2532     6.20                        ,          67,418,974.80
2615     4.60          ,       ,          184,487,551.40
2684     6.30                          ,          1,957,738.51
2788     5.10                       ,           72,937,783.20
```

```
2874     4.30                                  15,362,931.43
3047     6.30                    ,           70,299,052.00
3565     5.60                                1,038,321,489.00
3675     5.20          ,     ,           55,108,014.76
3739     4.80                                  50,451,949.00
5673     5.80                        ,       225,196,400.03
5806     5.00                      ,        106,514,301.03
6008     5.20           ,     ,           83,366,328.66
6323      NaN                              unknown      1,334,699.40
6823     6.20        ,     ,           147,606,826.51
```

[ ]:
```
display(merged_data.query('budget == 0').count())
```

```
title                   17
puNumber                17
show_start_date         17
type                    17
film_studio             17
production_country      17
director                17
producer                17
age_restriction         17
refundable_support      17
nonrefundable_support   17
budget                  17
financing_source        17
ratings                 16
genres                  17
box_office              17
dtype: int64
```

financing_source , budget 0, refundable_support nonrefundable_support (17) 0

[ ]:
```
merged_data = merged_data[merged_data['budget'] != 0]
```

[ ]:
```
display(merged_data.query('box_office <= 0'))
```

```
                                                  title   puNumber  \
66                                                  111011013
237                                                   111007613
596                                                 111003012
914          /                    /  121027712
932                                                124002912
933                                                 124002812
976                                                 131000213
1126                                                  111006712
1387            /                 /  121014613
1559                     /       …  121002614
```

```
1772                                          ;)  111002414
1782                              ,            121031913
1978                                               121013014
2037                                             111007114
2078                      ,    . (          ).  111009614
2126                                               111002015
2219                                           121011614
2442                                               111013915
2685                                               111009815
3595                                                  111017116
3827                                    ,           111001610
3994                                              121008610
4066                                                 .  121020210
4790                              .                111005910

               show_start_date            type  \
66   2013-12-19 12:00:00+00:00
237  2013-10-18 12:00:00+00:00
596  2012-05-23 12:00:00+00:00
914  2012-11-07 12:00:00+00:00
932  2012-10-25 12:00:00+00:00
933  2012-10-25 12:00:00+00:00
976  2013-04-23 12:00:00+00:00
1126 2012-12-18 12:00:00+00:00
1387 2013-06-11 12:00:00+00:00
1559 2014-02-11 12:00:00+00:00
1772 2014-02-27 12:00:00+00:00
1782 2013-12-12 12:00:00+00:00
1978 2014-06-23 12:00:00+00:00
2037 2014-07-18 12:00:00+00:00
2078 2014-09-02 12:00:00+00:00
2126 2015-03-02 12:00:00+00:00
2219 2014-06-11 12:00:00+00:00
2442 2015-08-25 12:00:00+00:00
2685 2015-06-01 12:00:00+00:00
3595 2016-11-10 12:00:00+00:00
3827 2010-02-09 12:00:00+00:00
3994 2010-05-05 12:00:00+00:00
4066 2010-11-03 12:00:00+00:00
4790 2010-06-04 12:00:00+00:00

                                             film_studio  \
66                          "     ",…
237                                      "     "
596                                        "      "
914             ,          ,          ,…
932             ,            ,        …
933       ,                       , …
```

```
976                  ,
1126                                  "           "
1387
1559              ,            ,      …
1772      "                           "
1782                        .,        ,
1978                         ,
2037                         "         "       "
2078                                          "      "
2126                            "        " ",     "
2219                                  ,
2442                                       "     "
2685                                  "     "
3595
3827     "  ,     ,       ",      "        …
3994         ,          ,      4,         …
4066              ,          ,         …
4790                         ,    ,   …

                                  production_country  \
66                 ,       ,      ,     ,
237
596
914          ,       ,      ,     ,
932
933
976
1126
1387
1559                       ,      ,    ,
1772
1782
1978
2037
2078
2126
2219
2442
2685
3595
3827
3994                               ,
4066                                    ,
4790

                                                director  \
66
237                                             .
```

```
596                                          .
914
932
933
976
1126                                         .
1387
1559
1772                                           .
1782
1978                                         .
2037                                           .
2078                                           .
2126                                    .    ,
2219                                  ,
2442                              .       ,   .
2685                                         .
3595                                           .
3827  .     ,   .     ,   .     ,   .  …
3994
4066
4790                                          .

                                              producer  \
66                                             unknown
237                                            unknown
596                                   .    ,   .
914
932            ,             ,
933              ,        ,          , …
976
1126                                         .
1387                  ,           ,
1559            ,          ,          …
1772                                     .
1782                ,         ,
1978                  ,         ,
2037                                        .
2078                                         .
2126                                    .    ,
2219              ,         ,
2442                                           unknown
2685  .    ,   .     ,   .   ,   .     ,…
3595                                    .   ,   .
3827                                  .    ,   .
3994                  ,          ,
4066                                           unknown
4790                                          .
```

```
                          age_restriction  refundable_support  \
66            «6+»,           6                    NaN
237         «12+»,           12                    NaN
596         «16+»,           16                    NaN
914         «12+»,           12                    NaN
932         «12+»,           12                    NaN
933         «12+»,           12                    NaN
976         «12+»,           12                    NaN
1126        «16+»,           16                    NaN
1387        «12+»,           12                    NaN
1559        «16+»,           16                    NaN
1772          «6+»,           6                    NaN
1782        «16+»,           16                    NaN
1978        «16+»,           16                    NaN
2037        «12+»,           12                    NaN
2078        «12+»,           12                    NaN
2126        «16+»,           16                    NaN
2219        «12+»,           12                    NaN
2442  «0+»,                                 NaN
2685        «12+»,           12                    NaN
3595        «16+»,           16                    NaN
3827        «16+»,           16                    NaN
3994        «16+»,           16                    NaN
4066        «16+»,           16                    NaN
4790        «16+»,           16                    NaN

      nonrefundable_support  budget financing_source  ratings  \
66                      NaN     NaN              NaN     7.70
237                     NaN     NaN              NaN     8.20
596                     NaN     NaN              NaN     6.30
914                     NaN     NaN              NaN     6.50
932                     NaN     NaN              NaN     7.40
933                     NaN     NaN              NaN     7.70
976                     NaN     NaN              NaN     7.90
1126                    NaN     NaN              NaN     5.50
1387                    NaN     NaN              NaN     5.80
1559                    NaN     NaN              NaN     6.90
1772                    NaN     NaN              NaN      NaN
1782                    NaN     NaN              NaN     8.00
1978                    NaN     NaN              NaN     6.00
2037                    NaN     NaN              NaN     6.90
2078                    NaN     NaN              NaN      NaN
2126                    NaN     NaN              NaN     9.10
2219                    NaN     NaN              NaN     6.10
2442                    NaN     NaN              NaN     7.50
2685                    NaN     NaN              NaN      NaN
3595                    NaN     NaN              NaN     6.70
```

```
3827                    NaN      NaN              NaN     4.50
3994                    NaN      NaN              NaN     7.50
4066                    NaN      NaN              NaN      NaN
4790                    NaN      NaN              NaN     4.50

                              genres  box_office
66             ,       ,              0.00
237                                          0.00
596                                         0.00
914             ,      ,              0.00
932            ,       ,              0.00
933       ,        ,                 0.00
976                                        0.00
1126                  ,                0.00
1387            ,      ,              0.00
1559                                         0.00
1772                          unknown        0.00
1782      ,       ,                 0.00
1978                    ,              0.00
2037                                         0.00
2078                          unknown        0.00
2126                     ,                0.00
2219            ,     ,                0.00
2442                                         0.00
2685                          unknown        0.00
3595          ,      ,                0.00
3827                     ,              0.00
3994        ,       ,                0.00
4066                          unknown        0.00
4790                       ,                0.00
```

0 box_office, , Nan

[ ]:
```
#                    "box_office"   NaN (          )
merged_data.loc[merged_data['box_office'] == 0, 'box_office'] = np.nan
```

[ ]:
```
display(merged_data.query('(box_office <= 1000) & (ratings > 5)').
 ↪sort_values('box_office'))
```

```
                                  title   puNumber           show_start_date  \
151                                111006013 2013-10-18 12:00:00+00:00
2274                             .  111016714 2014-12-01 12:00:00+00:00
3917              22    :          121006410 2010-04-01 12:00:00+00:00
1180                    3d  121025012 2012-10-05 12:00:00+00:00
351                           124000905 2015-01-18 12:00:00+00:00
…                                  …        …                        …
4388                             111004710 2010-04-19 12:00:00+00:00
2002                              111007814 2014-07-22 12:00:00+00:00
1779                           .  111003713 2013-08-28 12:00:00+00:00
```

```
1776                              121007414 2014-04-21 12:00:00+00:00
2009                                111009214 2014-08-21 12:00:00+00:00

              type                                          film_studio  \
151                                          "     "
2274                                         "     "
3917
1180                ,           ,         ,    …
351                           ,         ,
…              …                                                  …
4388                                    "           "
2002                                               "     "
1779                                                  "   "
1776                          ,       ,       …
2009                      "                     "

                    production_country                          director  \
151                                                              .
2274                                                              .
3917
1180      ,     ,     ,
351                     ,
…                                  …                                 …
4388                                                          .
2002                                                            .
1779                                                            .
1776      ,     ,     ,               ,
2009                                                            .

                                              producer  \
151                                            unknown
2274                                           unknown
3917           ,         ,     ,
1180                           ,
351           ,         ,          ,  …
…                                                 …
4388   .  ,   .   ,   .     ,   .     ,   …
2002                                        .
1779                             .    ,   .
1776
2009                                         .

                          age_restriction  refundable_support  \
151          «12+»,            12                    NaN
2274  «0+»,                                        NaN
3917         «16+»,            16                    NaN
1180           «6+»,            6                    NaN
351   «0+»,                                        NaN
```

```
…                                    …                 …
4388         «16+»,            16                  NaN
2002             «18+»,                          NaN
1779             «18+»,                          NaN
1776         «16+»,            16                  NaN
2009         «16+»,            16                  NaN

      nonrefundable_support  budget financing_source  ratings  \
151                     NaN     NaN              NaN     8.10
2274                    NaN     NaN              NaN     7.10
3917                    NaN     NaN              NaN     7.10
1180                    NaN     NaN              NaN     5.60
351                     NaN     NaN              NaN     5.70
…                       …       …                …       …
4388                    NaN     NaN              NaN     5.80
2002                    NaN     NaN              NaN     6.70
1779                    NaN     NaN              NaN     5.40
1776                    NaN     NaN              NaN     6.90
2009                    NaN     NaN              NaN     7.90

                            genres  box_office
151           ,      ,              40.00
2274           ,    ,              50.00
3917             ,    ,             75.00
1180      ,        ,               80.00
351           ,     ,              100.00
…                                  …          …
4388                    ,              980.00
2002                  ,              1,000.00
1779                    ,            1,000.00
1776                  ,              1,000.00
2009                                 1,000.00

[187 rows x 16 columns]
```

1000 box_office ratings > 5, , Nan
ratings 5 box_office 1000

[ ]:
```
merged_data.loc[(merged_data['box_office'] <= 1000) & (merged_data['ratings'] >␣
↪5), ['box_office']] = np.nan
```

[ ]:
```
merged_data.describe()
```

[ ]:
```
            puNumber  refundable_support  nonrefundable_support  \
count       7,469.00              315.00                 315.00
mean  135,231,130.73       11,603,174.60          48,107,043.96
std    38,409,917.68       25,151,438.82          59,886,140.86
min             0.00                0.00                   0.00
```

```
25%   113,000,116.00                0.00          25,000,000.00
50%   121,015,518.00                0.00          30,000,000.00
75%   124,003,318.00       10,000,000.00          40,000,000.00
max   231,001,111.00      180,000,000.00         400,000,000.00

                budget  ratings       box_office
count           315.00 6,503.00         2,930.00
mean    134,096,082.35     6.49    81,483,380.41
std     191,226,039.63     1.11   247,854,967.91
min      14,462,464.00     1.00           120.00
25%      46,153,866.50     5.90       263,261.25
50%      73,379,554.00     6.60     3,044,403.00
75%     150,004,286.50     7.20    28,625,993.04
max   2,305,074,303.00     9.90 3,073,568,690.79
```

[ ]:
```
merged_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 7469 entries, 0 to 7485
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   title                  7469 non-null   object
 1   puNumber               7469 non-null   Int64
 2   show_start_date        7469 non-null   datetime64[ns, UTC]
 3   type                   7469 non-null   object
 4   film_studio            7469 non-null   object
 5   production_country     7469 non-null   object
 6   director               7469 non-null   object
 7   producer               7469 non-null   object
 8   age_restriction        7469 non-null   object
 9   refundable_support     315 non-null    float64
 10  nonrefundable_support  315 non-null    float64
 11  budget                 315 non-null    float64
 12  financing_source       315 non-null    object
 13  ratings                6503 non-null   float64
 14  genres                 7469 non-null   object
 15  box_office             2930 non-null   float64
dtypes: Int64(1), datetime64[ns, UTC](1), float64(5), object(9)
memory usage: 999.3+ KB
```

2.6.

- . .

[ ]:
```
#                'show_start_date'
col_position = merged_data.columns.get_loc('show_start_date')
```

```
#               'release_year'
merged_data['release_year'] = merged_data['show_start_date'].dt.year;

#               'release_year',              'show_start_date'
columns = list(merged_data.columns)
columns.insert(col_position + 1, columns.pop(columns.index('release_year')))
merged_data = merged_data[columns]
```

- : .

.

```
<div class="card">
    <div class="card-header" id="headingThree">
        <button class="btn btn-link collapsed" data-toggle="collapse" data-target="#collapseHi
    </div>
    <div id="collapseHint_2" class="collapse" aria-labelledby="headingThree" data-parent="#acc
        <div class="card-body">

                ,                          .                          .

    </div>
</div>
```

[ ]:
```
merged_data['director_name'] = merged_data['director'].apply(lambda full_name:␣
↪full_name.split(',')[0])
merged_data['primary_genre'] = merged_data['genres'].apply(lambda genre_list:␣
↪genre_list.split(',')[0])
```

[ ]:
```
display(merged_data)
```

```
                                          title   puNumber  \
0                                          221048915
1                                         111013716
2                                            221038416
3                                            221026916
4                                            221030815
…                                                …        …
7481            :                  126008019
7482                                          121037819
7483             :               !  124010819
7484              .     "          "  124010619
7485          .     "            "  124010719

               show_start_date  release_year            type  \
0    2015-11-27 12:00:00+00:00          2015
1    2016-09-13 12:00:00+00:00          2016
2    2016-10-10 12:00:00+00:00          2016
3    2016-06-10 12:00:00+00:00          2016
4    2015-07-29 12:00:00+00:00          2015
…                        …           …              …
```

```
7481 2019-12-23 12:00:00+00:00          2019
7482 2019-12-24 12:00:00+00:00          2019
7483 2019-12-28 12:00:00+00:00          2019
7484 2019-12-30 12:00:00+00:00          2019
7485 2019-12-30 12:00:00+00:00          2019

                                            film_studio  \
0              ,           ,        …
1                                        "     "
2              ,           ,      …
3              ,           ,      …
4              ,           ,      …
…                                                  …
7481
7482               ,             …
7483                        ,
7484                  .,          "      "
7485                  .,          "      "

           production_country           director  \
0
1                                      .
2
3
4
…                        …                …
7481
7482
7483                ,             .
7484          ,                  .
7485          ,                  .

                                               producer  \
0                  ,          ,
1                                             unknown
2          ,          ,      .   ,   …
3          ,          ,      .   ,   …
4          ,          ,      .   ,   …
…                                                 …
7481
7482         ,          ,         ,  …
7483
7484                                           .
7485                                           .

                          age_restriction  refundable_support  \
0               «18+»,                           NaN
1              «6+»,           6                 NaN
```

```
2                «18+»,                                NaN
3                «18+»,                                NaN
4                «18+»,                                NaN
…                                      …                    …
7481         «12+»,            12                      NaN
7482         «16+»,            16                      NaN
7483  «0+»,                                    NaN
7484  «0+»,                                    NaN
7485  «0+»,                                    NaN

      nonrefundable_support  budget financing_source  ratings  \
0                       NaN     NaN              NaN     7.20
1                       NaN     NaN              NaN     6.60
2                       NaN     NaN              NaN     6.80
3                       NaN     NaN              NaN     6.80
4                       NaN     NaN              NaN     6.80
…                       …       …                …       …
7481                    NaN     NaN              NaN      NaN
7482                    NaN     NaN              NaN     5.40
7483                    NaN     NaN              NaN      NaN
7484                    NaN     NaN              NaN      NaN
7485                    NaN     NaN              NaN     6.30

                          genres  box_office      director_name primary_genre
0           ,    ,               NaN
1                     ,              NaN           .
2           ,    ,              NaN
3           ,    ,              NaN
4           ,    ,              NaN
…                          …          …                …           …
7481                     unknown         NaN                       unknown
7482                   ,              NaN
7483                     unknown         NaN        .           unknown
7484                     unknown         NaN            .           unknown
7485                   ,              NaN           .

[7469 rows x 19 columns]
```

- , .

[ ]:
```
#
total_support = merged_data['refundable_support'] +␣
 ↪merged_data['nonrefundable_support']
merged_data['support_ratio'] = total_support / merged_data['budget']

#
merged_data['support_ratio'] = merged_data['support_ratio'].apply(lambda x:␣
 ↪f'{x:.2%}');
```

```
#
cols = list(merged_data.columns)

#               'budget'
budget_index = cols.index('budget')
cols = cols[:budget_index + 1] + ['support_ratio'] + cols[budget_index + 1:-1]

#
merged_data = merged_data[cols]
display(merged_data)
```

```
                                                 title   puNumber  \
0                                                221048915
1                                               111013716
2                                                  221038416
3                                                  221026916
4                                                  221030815
…                                                 …        …
7481            :                  126008019
7482                                              121037819
7483             :               !  124010819
7484              .     "          "  124010619
7485          .     "            "  124010719

              show_start_date  release_year            type  \
0    2015-11-27 12:00:00+00:00          2015
1    2016-09-13 12:00:00+00:00          2016
2    2016-10-10 12:00:00+00:00          2016
3    2016-06-10 12:00:00+00:00          2016
4    2015-07-29 12:00:00+00:00          2015
…                          …           …              …
7481 2019-12-23 12:00:00+00:00          2019
7482 2019-12-24 12:00:00+00:00          2019
7483 2019-12-28 12:00:00+00:00          2019
7484 2019-12-30 12:00:00+00:00          2019
7485 2019-12-30 12:00:00+00:00          2019

                                            film_studio  \
0              ,           ,       …
1                                        "     "
2              ,            ,      …
3              ,            ,      …
4              ,            ,      …
…                                                  …
7481
7482               ,             …
7483                         ,
```

```
7484                 .,         "      "
7485                 .,         "      "

           production_country           director  \
0
1                                      .
2
3
4
…                       …               …
7481
7482
7483               ,            .
7484          ,                 .
7485          ,                 .

                                                producer  \
0                  ,          ,
1                                                 unknown
2           ,          ,      .   ,   …
3           ,          ,      .   ,   …
4           ,          ,      .   ,   …
…                                                    …
7481
7482         ,          ,         ,  …
7483
7484                                               .
7485                                               .

                           age_restriction  refundable_support  \
0               «18+»,                                   NaN
1              «6+»,            6                        NaN
2               «18+»,                                   NaN
3               «18+»,                                   NaN
4               «18+»,                                   NaN
…                                       …                   …
7481         «12+»,            12                        NaN
7482         «16+»,            16                        NaN
7483  «0+»,                                           NaN
7484  «0+»,                                           NaN
7485  «0+»,                                           NaN

      nonrefundable_support  budget support_ratio financing_source  ratings  \
0                       NaN     NaN          nan%              NaN     7.20
1                       NaN     NaN          nan%              NaN     6.60
2                       NaN     NaN          nan%              NaN     6.80
3                       NaN     NaN          nan%              NaN     6.80
4                       NaN     NaN          nan%              NaN     6.80
```

```
…                    …     …          …                …      …
7481                 NaN   NaN        nan%             NaN    NaN
7482                 NaN   NaN        nan%             NaN    5.40
7483                 NaN   NaN        nan%             NaN    NaN
7484                 NaN   NaN        nan%             NaN    NaN
7485                 NaN   NaN        nan%             NaN    6.30

                        genres  box_office      director_name primary_genre
0            ,    ,               NaN
1                    ,               NaN           .
2            ,    ,              NaN
3            ,    ,              NaN
4            ,    ,              NaN
…                          …          …                …           …
7481                    unknown          NaN                       unknown
7482                  ,              NaN
7483                    unknown          NaN        .           unknown
7484                    unknown          NaN            .           unknown
7485                  ,              NaN           .

[7469 rows x 20 columns]
```

, , - .
( ) , . .

, ,
‘production_country’, ,
- . (“-”), (” - “),
(‘,’), (”-“).

, ,
.

, .

0.1.3 3.

- , . , .
, . , .
, .

```
[ ]: #
film_counts = merged_data.groupby('release_year').size().
  ↪reset_index(name='film_count')

#
films_with_box_office_count = merged_data[merged_data['box_office'].notna()].
  ↪groupby('release_year').size().
  ↪reset_index(name='films_with_box_office_count')
```

```
#
merged_counts = pd.merge(film_counts, films_with_box_office_count,␣
 ↪on='release_year', how='left')

#
merged_counts['box_office_ratio'] =␣
 ↪merged_counts['films_with_box_office_count'] / merged_counts['film_count']
```

[ ]:
```
plt.figure(figsize=(10, 6))
plt.plot(merged_counts['release_year'], merged_counts['box_office_ratio'])
plt.xlabel('      ')
plt.ylabel('                              ')
plt.title('                                        ')

for i, value in enumerate(merged_counts['box_office_ratio']):
    plt.text(merged_counts['release_year'][i], value, f'{value:.2f}%',␣
 ↪ha='center')

plt.show()
display(merged_counts)
```

	release_year	film_count	films_with_box_office_count	box_office_ratio
0	2010	985	67	0.07
1	2011	622	82	0.13

2	2012	593	99	0.17
3	2013	630	136	0.22
4	2014	806	227	0.28
5	2015	699	451	0.65
6	2016	815	513	0.63
7	2017	503	355	0.71
8	2018	887	471	0.53
9	2019	929	529	0.57

:

1. , , , 503 2017 985 2010 . , 2014 2019 929 .
2. 2010 2017 , 0.07 0.71. .
3. 2015 . , .
4. 2017 , .
5. , 2019 (0.57), .

, , .

- , . ? ?

[ ]:
```
plt.figure(figsize=(20, 6))
plt.plot(merged_counts['release_year'], merged_data.
 ↪groupby('release_year')['box_office'].sum())
plt.xlabel('      ')
plt.ylabel('        ')
plt.title('                    ')

for i, value in enumerate(merged_data.groupby('release_year')['box_office'].
 ↪sum()):
    plt.text(merged_counts['release_year'][i], value, f'{value/1000000:.2f}
 ↪\u20BD', ha='center')

plt.show()

#
box_office_table = pd.DataFrame(merged_data.
 ↪groupby('release_year')['box_office'].sum())
box_office_table.columns = ['        ']
box_office_table.index.name = '      '
```

```
#
display(box_office_table)
```


```
2010            2,415,119.00
2011           14,090,485.00
2012            6,945,533.00
2013           29,779,401.20
2014        7,434,695,443.10
2015       38,527,825,699.09
2016       46,742,312,299.48
2017       48,458,147,253.75
2018       49,251,991,405.20
2019       48,278,101,953.08
```

[ ]:
```
min_box_office_year = merged_data.groupby('release_year')['box_office'].sum().
 ↪idxmin()
max_box_office_year = merged_data.groupby('release_year')['box_office'].sum().
 ↪idxmax()

min_box_office_sum = merged_data.groupby('release_year')['box_office'].sum().
 ↪min()
max_box_office_sum = merged_data.groupby('release_year')['box_office'].sum().
 ↪max()

print("                    :", min_box_office_year, "        :",␣
 ↪min_box_office_sum, "\u20BD")
print("                     :", max_box_office_year, "        :",␣
 ↪max_box_office_sum, "\u20BD")
```

```
                  : 2010         : 2415119.0
                   : 2018         : 49251991405.2
```

:

1. 2010 2014 . 2014 7,4 ,
, .
2. 2014 .
.
3. 2015 , 5 2014 .
4. 2017 : 48-49 ,
.

, 2010 2015
.

- .

[ ]:
```
#
pivot_table = pd.pivot_table(merged_data, values='box_office',␣
↪index='release_year', aggfunc=[np.mean, np.median])

#
display(pivot_table)
```

```
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\2573536348.py:2: FutureWarning:
The provided callable <function mean at 0x0000013A8A2E8400> is currently using
DataFrameGroupBy.mean. In a future version of pandas, the provided callable will
be used directly. To keep current behavior pass the string "mean" instead.
  pivot_table = pd.pivot_table(merged_data, values='box_office',
index='release_year', aggfunc=[np.mean, np.median])
C:\Users\ivano\AppData\Local\Temp\ipykernel_4564\2573536348.py:2: FutureWarning:
The provided callable <function median at 0x0000013AA041F100> is currently using
DataFrameGroupBy.median. In a future version of pandas, the provided callable
will be used directly. To keep current behavior pass the string "median"
instead.
  pivot_table = pd.pivot_table(merged_data, values='box_office',
index='release_year', aggfunc=[np.mean, np.median])
```

	mean	median
	box_office	box_office
release_year		
2010	36,046.55	3,878.00
2011	171,835.18	5,555.00
2012	70,156.90	10,310.00
2013	218,966.19	6,497.50
2014	32,751,962.30	123,190.00
2015	85,427,551.44	5,034,342.40
2016	91,115,618.52	4,034,973.50
2017	136,501,823.25	9,919,415.00
2018	104,568,983.88	8,751,895.75

2019 91,262,952.65 4,601,033.60

, 2010 2019 .
, .
2014 , (-).
, .
, , .
, , ,
, .

- , («6+», «12+», «16+», «18+» . .)
2015 2019 ? ?
? , , .

[ ]:
```
#
merged_data['age_restriction'] = merged_data['age_restriction'].str.split(',').
 ↪str[0]#                2015   2019
filtered_data = merged_data[(merged_data['release_year'] >= 2015) &␣
 ↪(merged_data['release_year'] <= 2019)]

#
age_restriction_totals = filtered_data.groupby('age_restriction')['box_office'].
 ↪sum().reset_index(name='total_box_office')

#
age_restriction_max = age_restriction_totals.sort_values(by='total_box_office',␣
 ↪ascending=False).head(1)

print("                                  :")
display(age_restriction_totals)
print("\n              ,                          :")
display(age_restriction_max)
```

:

	age_restriction	total_box_office
0	«0+»	809,076,715.97
1	«12+»	59,369,189,769.50
2	«16+»	75,102,265,523.60
3	«18+»	40,757,657,233.61
4	«6+»	55,220,189,367.92

, :

	age_restriction	total_box_office
2	«16+»	75,102,265,523.60

, «16+»
, 75,102,265,523.60. ,

. , «16+»
,

[ ]:
```
#
year_age_restriction_totals = filtered_data.groupby(['release_year',␣
↪'age_restriction'])['box_office'].sum().reset_index(name='total_box_office')

#
print("                                        :")
display(year_age_restriction_totals)
```

:

	release_year	age_restriction	total_box_office
0	2015	«0+»	379,054,358.37
1	2015	«12+»	13,419,509,570.51
2	2015	«16+»	10,781,600,451.27
3	2015	«18+»	5,430,350,628.93
4	2015	«6+»	8,517,310,690.01
5	2016	«0+»	150,228,358.67
6	2016	«12+»	11,150,762,103.96
7	2016	«16+»	16,594,249,189.74
8	2016	«18+»	6,793,929,218.87
9	2016	«6+»	12,053,143,428.24
10	2017	«0+»	229,598,930.00
11	2017	«12+»	7,851,427,660.67
12	2017	«16+»	18,694,590,951.06
13	2017	«18+»	9,651,495,581.02
14	2017	«6+»	12,031,034,131.00
15	2018	«0+»	32,449,002.11
16	2018	«12+»	14,267,291,660.69
17	2018	«16+»	16,278,405,946.93
18	2018	«18+»	8,760,085,501.15
19	2018	«6+»	9,913,759,294.32
20	2019	«0+»	17,746,066.82
21	2019	«12+»	12,680,198,773.67
22	2019	«16+»	12,753,418,984.60
23	2019	«18+»	10,121,796,303.64
24	2019	«6+»	12,704,941,824.35

, :

1. , «0+», «6+», «12+», «16+» «18+»,
.
2. 2015 «12+» , 13,419,509,570.51
. , , , .
3. 2016 «16+» , 16,594,249,189.74 .
, , , .

4. 2017 «18+» , 9,651,495,581.02 .
, .

5. «0+» ,
, - .

6. , 2015 2019 ,
. .

, .
“18+” “16+” , “12+” “6+”
. , .
, , , , .
. , .

0.1.4 4. ,

— . ,
. , , .

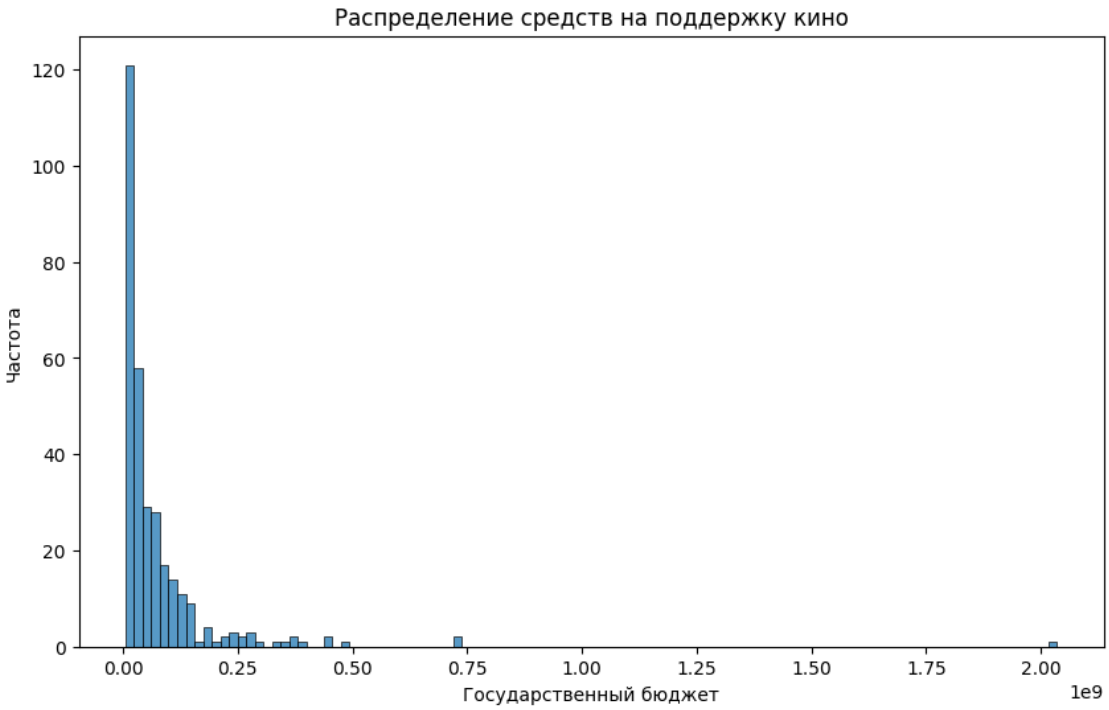
[ ]:
```
merged_data['gos_budget'] = merged_data['budget']␣
↪-(merged_data['refundable_support'] + merged_data['nonrefundable_support'])
```

[ ]:
```
merged_data2 = merged_data.dropna(subset=['financing_source'])
df = merged_data2[['title', 'budget', 'gos_budget', 'support_ratio',␣
↪'box_office', 'ratings', 'primary_genre', 'type', 'release_year',␣
↪'age_restriction']]
display(df)
```

```
                          title         budget     gos_budget support_ratio  \
1281                        79,411,900.00  52,911,900.00        33.37%
1448                      37,142,857.00  11,142,857.00        70.00%
1498                         176,023,490.00  68,175,545.00        61.27%
1524                            40,574,140.00  12,574,140.00        69.01%
1792                          40,015,122.00  15,015,122.00        62.48%
…                         …             …             …           …
7464                       980,000,000.00 480,000,000.00        51.02%
7466                  4 190,000,000.00  90,000,000.00        52.63%
7474                         944,000,000.00 444,000,000.00        52.97%
7476                          46,154,000.00  16,154,000.00        65.00%
7478        ( )           150,147,502.00  50,147,502.00        66.60%

        box_office  ratings primary_genre            type  release_year  \
1281     365,353.60     8.10                                 2013
1448      28,140.00      NaN       unknown                      2014
1498  19,957,031.50     5.30                                 2013
1524      55,917.50     4.20                                  2014
```

```
1792     232,100.00     6.40                                  2014
…              …      …           …               …            …
7464 717,703,185.53     6.00                                  2019
7466 501,069,235.00     6.70                                  2019
7474            NaN     5.70                                  2019
7476            NaN     5.90                                  2019
7478            NaN     4.50                                  2019

     age_restriction
1281           «16+»
1448           «12+»
1498           «12+»
1524           «16+»
1792           «16+»
…              …
7464           «12+»
7466            «6+»
7474           «12+»
7476           «12+»
7478           «12+»

[315 rows x 10 columns]
```

[ ]:
```
#
plt.figure(figsize=(10, 6))
sns.histplot(df['gos_budget'])
plt.title('                        ')
plt.xlabel('             ')
plt.ylabel('    ')
plt.show()

#
total_gos_budget = df['gos_budget'].sum()
avg_gos_budget = df['gos_budget'].mean()

print('                              :')
print('                 :', total_gos_budget)
print('                    :', avg_gos_budget)
```



:

: 23431547092.0

: 74385863.78412698

. 23 431 547 092.0 . 74 385 863.78 . ,

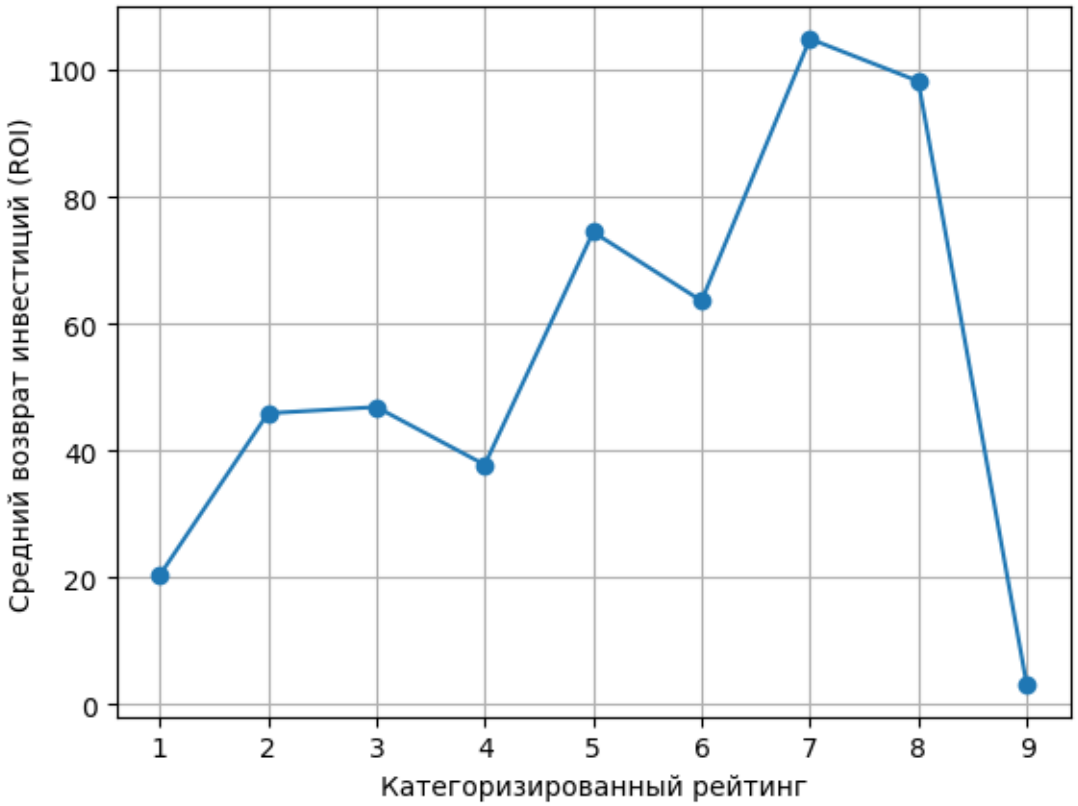
.

[ ]:
```
df['roi'] = df['box_office'] / df['budget'] *100
```

[ ]:
```
#
df['rounded_ratings'] = df['ratings'][~df['ratings'].isna()].apply(round)
```

[ ]:
```
mean_roi = df.groupby('rounded_ratings')['roi'].mean()
plt.plot(mean_roi.index, mean_roi.values, marker='o')
plt.xlabel("                ")
plt.ylabel("                 (ROI)")
plt.grid(True)
plt.show()
#
outputtable = pd.DataFrame({'    ': mean_roi.index, '     ROI': mean_roi.
 ↪values})

#
display(outputtable)
```

```
#                          (ROI)
avg_roi = df['roi'].mean()
print(f"                (ROI): {avg_roi}")

#                           (ROI)
max_roi_movie = df.loc[df['roi'].idxmax(), 'title']
print(f"                          (ROI): {max_roi_movie}")

#                          (ROI)
min_roi_movie = df.loc[df['roi'].idxmin(), 'title']
print(f"                          (ROI): {min_roi_movie}")
```



		ROI
0	1.00	20.33
1	2.00	45.90
2	3.00	46.88
3	4.00	37.78
4	5.00	74.45
5	6.00	63.60
6	7.00	104.89
7	8.00	98.21

8 9.00 3.19

(ROI): 76.33099908958417
(ROI):
(ROI): ,

ROI :

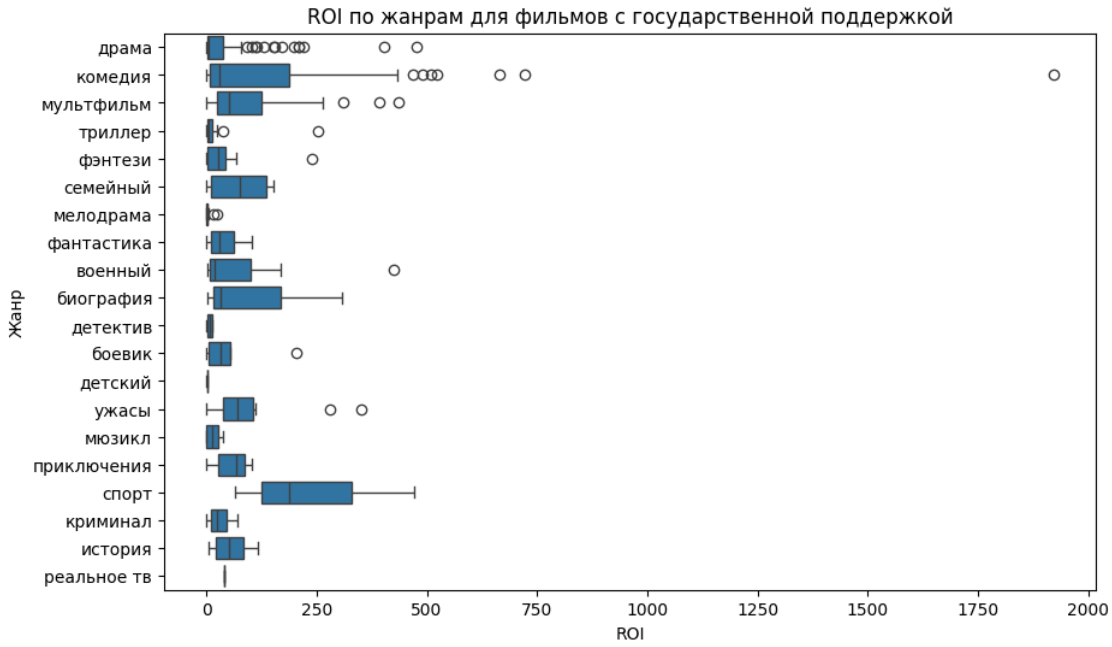
1) ROI 76.33%. , 11.99%
. , , .

2) (ROI) 7 104.89%.
.

3) (ROI) 9 3.14%.
ROI.

, .
, . ,
ROI , , , , .
.

[ ]:
```
#              primary_genre,             "unknown"
filtered_df = df[df['primary_genre'] != "unknown"]

# ROI
plt.figure(figsize=(10, 6))
sns.boxplot(x='roi', y='primary_genre',␣
 ↪data=filtered_df[filtered_df['gos_budget'] != 0])
plt.title('ROI                                  ')
plt.xlabel('ROI')
plt.ylabel('   ')
plt.show()

#       ROI
genre_avg_roi = filtered_df[filtered_df['gos_budget'] != 0].
 ↪groupby('primary_genre')['roi'].mean()

print('      ROI                              :')
print(genre_avg_roi)
```



ROI :

```
primary_genre
          113.67
            59.30
           85.73
            7.14
             2.17
            41.07
           56.13
          152.75
           31.75
            4.22
          99.79
            15.56
         57.89
          40.90
           75.25
            240.83
            26.18
            105.08
          38.79
            42.76
Name: roi, dtype: float64
```

ROI , .

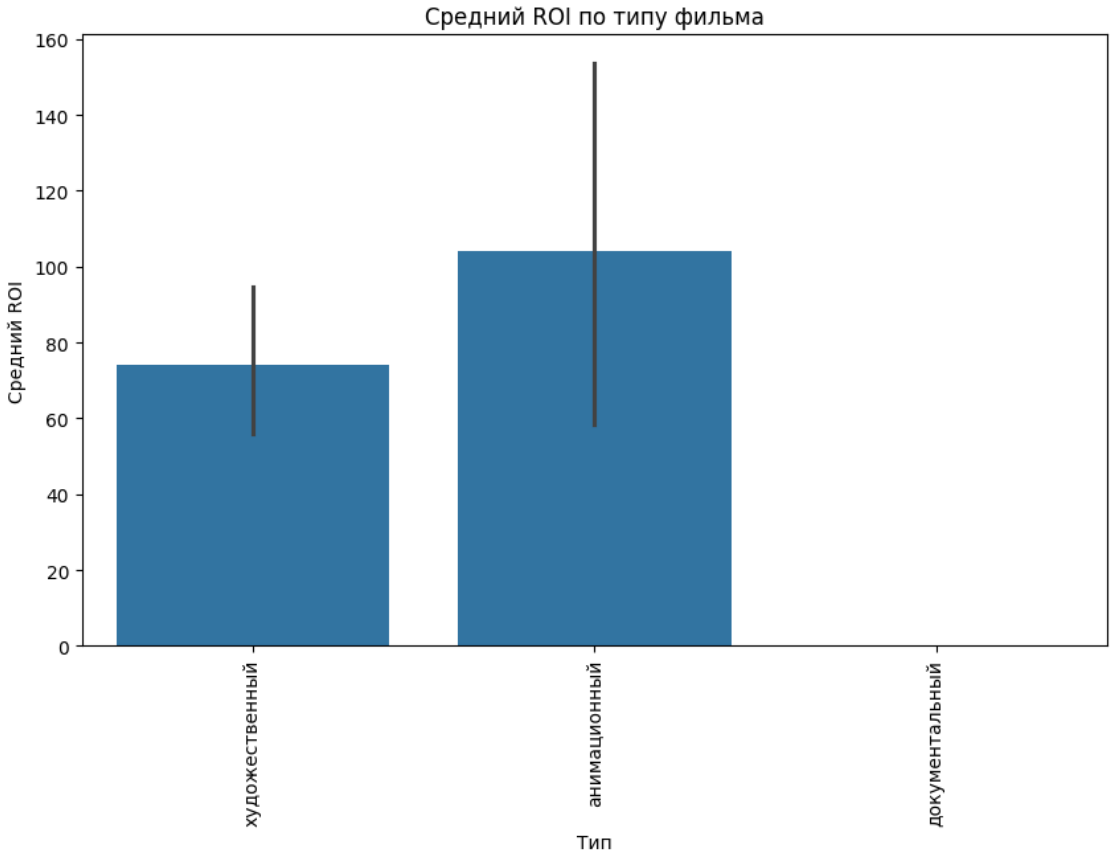
1. (ROI 152.75), (ROI 240.83), (ROI

105.08) (ROI 99.79).
.

2. , , (ROI 113.67), (ROI 85.73), (ROI 56.13), (ROI 57.89), (ROI 42.76). ROI , .
3. (ROI 41.07), (ROI 31.75), (ROI 40.90), (ROI 26.18) (ROI 38.79) ROI. , , .
4. (ROI 7.14), (ROI 2.17), (ROI 4.22) (ROI 15.56) ROI. , , , .

, , , - , , , , , , , , , , , , , .

[ ]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x='type', y='roi', data=df)
plt.title('     ROI          ')
plt.xlabel('  ')
plt.ylabel('     ROI')
plt.xticks(rotation=90)
plt.show();

type_roi = df.groupby('type')['roi'].mean()
print('     ROI          :')
print(type_roi);
```



ROI :

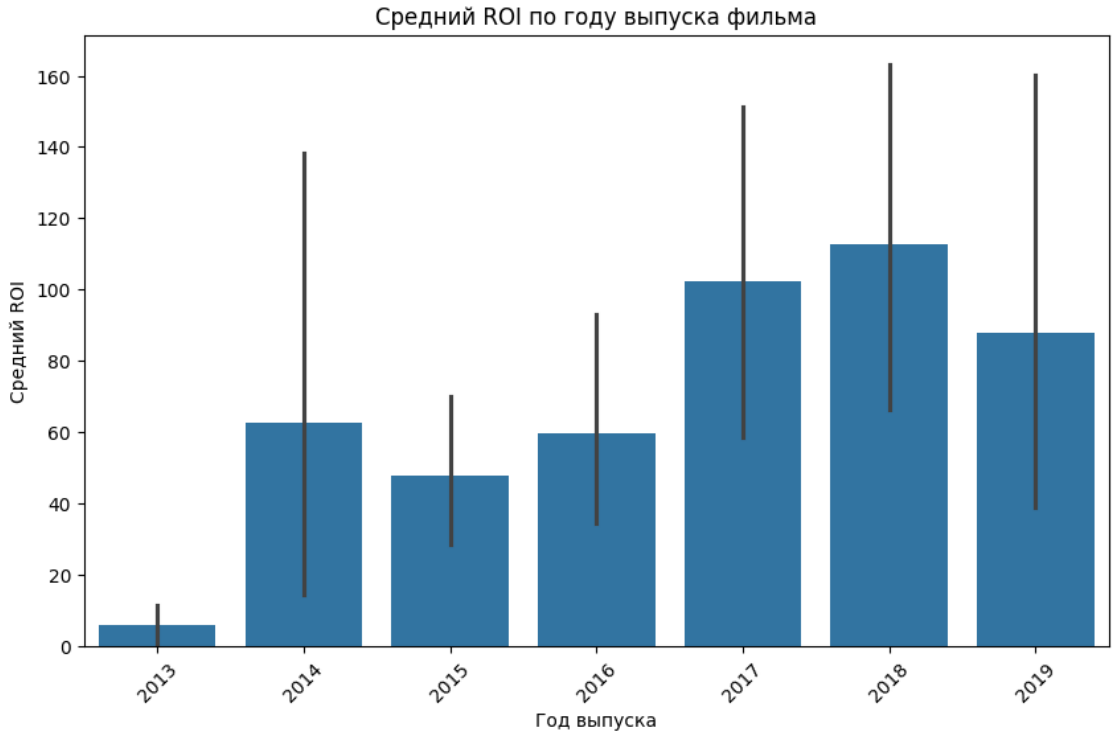
```
type
          103.94
            NaN
          73.94
Name: roi, dtype: float64
```

(ROI) :

1. ROI, 103.94%. , , ,
, . (
,), ,
.
2. ROI . :
, .
3. ROI 73.94%. , , ,
, , .

[ ]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x='release_year', y='roi', data=df)
```

```
plt.title('     ROI              ')
plt.xlabel('       ')
plt.ylabel('     ROI')
plt.xticks(rotation=45)
plt.show();

release_year_roi = df.groupby('release_year')['roi'].mean()
print('     ROI              :')
print(release_year_roi);
```




```
    ROI              :
release_year
2013     5.90
2014    62.54
2015    47.83
2016    59.86
2017   102.23
2018   112.92
2019    87.80
Name: roi, dtype: float64
```

(ROI) :

1. , 2013 2018 ROI.

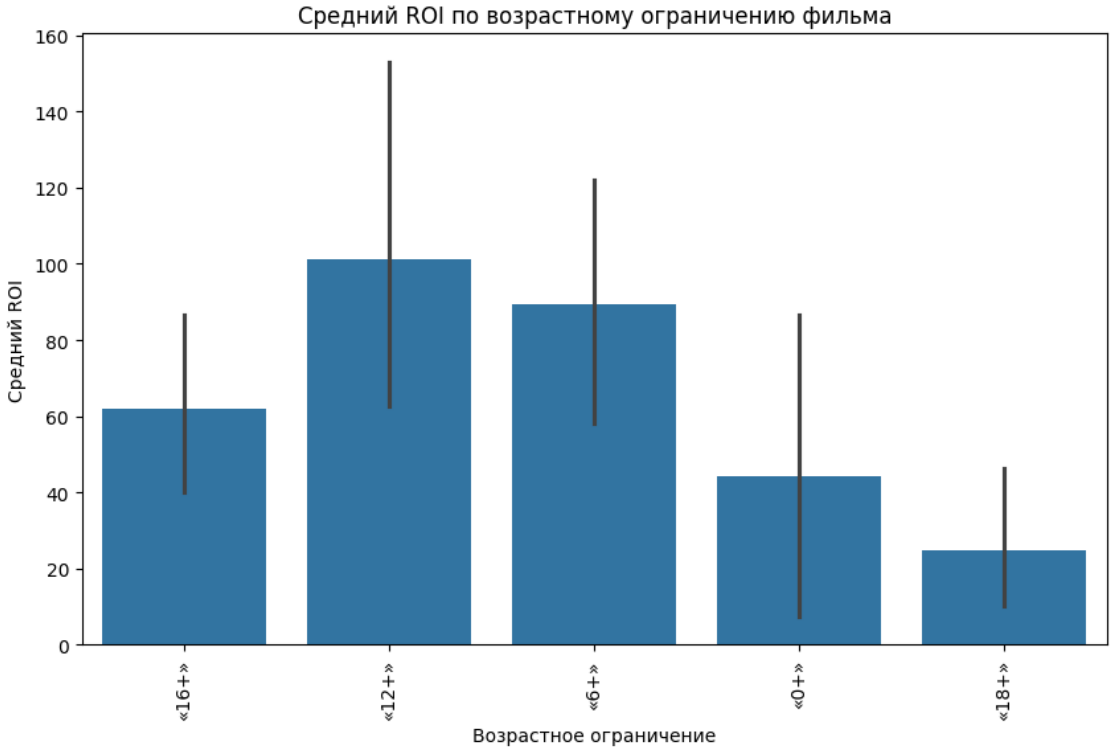
, , , .

2. 2013 ROI 5.90%.
.

3. 2018 ROI, 112.92%.
, , .

4. 2019 87.80%, ROI ,
.

[ ]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x='age_restriction', y='roi', data=df)
plt.title('     ROI                    ')
plt.xlabel('             ')
plt.ylabel('     ROI')
plt.xticks(rotation=90)
plt.show();

age_restriction_roi = df.groupby('age_restriction')['roi'].mean()
print('     ROI                    :')
print(age_restriction_roi);
```



ROI :
age_restriction

```
«0+»     44.17
«12+»   101.13
«16+»    61.85
«18+»    24.95
«6+»     89.52
Name: roi, dtype: float64
```

, Return on Investment (ROI)
“12+”, 101,13%. ,
.

ROI “6+”, 89,52%. - “16+”
ROI 61,85%. “18+”, ROI 24,95%.

, , “6+” “12+” ,
.

0.1.5 5.

: 1)
: moviesdata showsdata. ,
, .

2) , ‘puNumber’ , ,
‘Int64’. ,
.

3) ‘left’, moviesdata
showsdata. ,
.

4) , , , ,
, . ,
.

5) , , ,
, , .

:

, , .
, 2017 . 2019
.

, , 2010 2019 .
, , .
. “12+” “16+”
, , . ,
“18+” .

, ,
. 2017

.

.

ROI:

,

, . ROI , ,

, . , , , ,

, . ROI, ,

, . , 2013 2018

ROI, .

:

, :

1) : - ,

. , , .

2) : . ,

.

3) ROI: , , , ROI.

, , ,

.

, . , - ,

, .

.