Importing Libraries

import numpy as np
import pandas as pd

Out[6]: (4803, 4)

Importing Dataset

```
In [2]: movies = pd.read csv('tmdb 5000 movies.csv')
           credits = pd.read csv('tmdb 5000 credits.csv')
In [3]:
          movies.head()
                  budget
                                  genres
                                                                            homepage
                                                                                               id keywords original language
                                                                                                                                    original title
                                                                                                                                                     overview
                                                                                                        [{"id":
                                                                                                                                                         In the
                                [{"id": 28,
                                                                                                        1463,
                                                                                                                                                          22nd
                                  "name":
                                                                                                      "name":
                                                                                                                                                     century, a
           0 237000000
                                "Action"},
                                                           http://www.avatarmovie.com/
                                                                                           19995
                                                                                                                                 en
                                                                                                                                            Avatar
                                                                                                                                                     paraplegic
                                                                                                       "culture
                                {"id": 12,
                                                                                                      clash"},
                                                                                                                                                      Marine is
                                                                                                       {"id":...
                                                                                                                                                           di...
                                                                                                                                                       Captain
                                                                                                    [{"id": 270,
                                                                                                                                                     Barbossa.
                                [{"id": 12,
                                                                                                                                     Pirates of the
                                                                                                      "name":
                                                                                                                                                           long
                                 "name":
                                                                                                                                        Caribbean:
              300000000
                                                                                             285
                                                                                                     "ocean"}.
                                             http://disney.go.com/disneypictures/pirates/
                                                                                                                                                       believed
                                                                                                                                en
                            "Adventure"},
                                                                                                                                        At World's
                                                                                                    {"id": 726,
                                                                                                                                                          to be
                             {"id": 14, "...
                                                                                                                                              End
                                                                                                         "na...
                                                                                                                                                         dead,
                                                                                                                                                           ha...
                                                                                                                                                      A cryptic
                                [{"id": 28,
                                                                                                   [{"id": 470.
                                                                                                                                                      message
                                  "name":
                                                                                                      "name":
                                                                                                                                                          from
           2 245000000
                                "Action"},
                                           http://www.sonypictures.com/movies/spectre/ 206647
                                                                                                       "spy"},
                                                                                                                                 en
                                                                                                                                           Spectre
                                                                                                                                                        Bond's
                                {"id": 12,
                                                                                                    {"id": 818,
                                                                                                                                                           past
                                  "nam...
                                                                                                      "name...
                                                                                                                                                     sends him
                                                                                                                                                            0...
                                                                                                   [{"id": 849,
                                [{"id": 28,
                                                                                                                                                      Following
                                                                                                      "name":
                                  "name":
                                                                                                                                                      the death
                                                                                                          "dc
                                                                                                                                         The Dark
              250000000
                                "Action"},
                                                                                           49026
                                                     http://www.thedarkknightrises.com/
                                                                                                                                                      of District
                                                                                                                                en
                                                                                                     comics"},
                                                                                                                                      Knight Rises
                                {"id": 80,
                                                                                                                                                       Attorney
                                                                                                         {"id":
                                  "nam...
                                                                                                                                                       Harve...
                                                                                                       853,...
                                                                                                                                                          John
                                [{"id": 28,
                                                                                                   [{"id": 818,
                                                                                                                                                     Carter is a
                                 "name":
                                                                                                      "name":
                                                                                                                                                          war-
              260000000
                                                                                           49529
                                                                                                                                       John Carter
                                "Action"},
                                                    http://movies.disney.com/john-carter
                                                                                                    "based on
                                                                                                                                                         weary,
                                {"id": 12,
                                                                                                      novel"},
                                                                                                                                                        former
                                                                                                       {"id":...
                                                                                                                                                        military
                                  "nam...
                                                                                                                                                           ca...
In [4]:
          movies.shape
           (4803, 20)
Out[4]:
In [5]:
           credits.head()
Out[5]:
              movie_id
                                                           title
                                                                                                       cast
                                                                                                                                                      crew
                                                                   [{"cast_id": 242, "character": "Jake Sully",
                                                                                                                [{"credit_id": "52fe48009251416c750aca23",
           0
                  19995
                                                        Avatar
                                                                    [{"cast_id": 4, "character": "Captain Jack
                                                                                                                [{"credit_id": "52fe4232c3a36847f800b579",
                           Pirates of the Caribbean: At World's
           1
                    285
                                                           End
                                                                                                     Spa...
                                                                                                                                                      "de...
                                                                   [{"cast_id": 1, "character": "James Bond",
                                                                                                               [{"credit_id": "54805967c3a36829b5002c41",
                 206647
           2
                                                       Spectre
                                                                                                                                                      "de...
                                                                                                                [{"credit id": "52fe4781c3a36847f81398c3",
                                                                  [{"cast id": 2, "character": "Bruce Wayne /
                                        The Dark Knight Rises
           3
                  49026
                                                                                                       Ва...
                                                                                                                                                      "de...
                                                                   [{"cast_id": 5, "character": "John Carter",
                                                                                                                [{"credit_id": "52fe479ac3a36847f813eaa3",
           4
                  49529
                                                   John Carter
In [6]:
          credits.shape
```

Merging the dataframes on the basis of 'title'

```
In [7]: movies = movies.merge(credits, on = 'title')
In [8]: movies.head()
Out[8]:
                budaet
                                                                    homepage
                                                                                    id keywords original_language original_title
                                                                                                                                    overview
                              aenres
                                                                                                                                        In the
                                                                                            [{"id":
                            [{"id": 28,
                                                                                            1463.
                                                                                                                                        22nd
                              "name":
                                                                                          "name":
                                                                                                                                    century, a
          0 237000000
                            "Action"},
                                                    http://www.avatarmovie.com/
                                                                                                                            Avatar
                                                                                           "culture
                                                                                                                                    paraplegic
                             {"id": 12,
                                                                                           clash"},
                                                                                                                                     Marine is
                              "nam...
                                                                                           {"id":...
                                                                                                                                      Captain
                                                                                        [{"id": 270,
                                                                                                                                    Barbossa,
                            [{"id": 12,
                                                                                                                      Pirates of the
                                                                                          "name":
                                                                                                                                         long
                                                                                                                        Caribbean:
                             "name":
             300000000
                                                                                   285
                                                                                         "ocean"},
                                        http://disney.go.com/disneypictures/pirates/
                                                                                                                                     believed
                        "Adventure"},
                                                                                                                         At World's
                                                                                         {"id": 726,
                                                                                                                                        to be
                                                                                                                              Fnd
                          {"id": 14. "..
                                                                                             "na...
                                                                                                                                        dead,
                                                                                                                                         ha...
                                                                                                                                     A cryptic
                            [{"id": 28,
                                                                                        [{"id": 470,
                                                                                                                                     message
                              "name":
                                                                                           "name":
                                                                                                                                         from
          2 245000000
                            "Action"},
                                      http://www.sonypictures.com/movies/spectre/ 206647
                                                                                           "spy"},
                                                                                                                  en
                                                                                                                           Spectre
                                                                                                                                       Bond's
                             {"id": 12,
                                                                                         {"id": 818,
                                                                                                                                         past
                                                                                                                                    sends him
                              "nam...
                                                                                          "name...
                                                                                                                                          0...
                                                                                        [{"id": 849,
                                                                                                                                    Following
                            [{"id": 28,
                                                                                           "name":
                              "name":
                                                                                                                                    the death
                                                                                              "dc
                                                                                                                          The Dark
          3 250000000
                            "Action"}
                                               http://www.thedarkknightrises.com/
                                                                                49026
                                                                                                                                    of District
                                                                                         comics"},
                                                                                                                      Knight Rises
                             {"id": 80,
                                                                                                                                     Attorney
                                                                                             {"id":
                                                                                                                                      Harve...
                              "nam...
                                                                                           853,...
                                                                                                                                        John
                            [{"id": 28,
                                                                                        [{"id": 818,
                                                                                                                                    Carter is a
                             "name":
                                                                                          "name":
                                                                                                                                        war-
          4 260000000
                            "Action"},
                                              http://movies.disney.com/john-carter
                                                                                49529
                                                                                        "based on
                                                                                                                       John Carter
                                                                                                                                       weary,
                             {"id": 12,
                                                                                          novel"},
                                                                                                                                       former
                              "nam...
                                                                                           {"id":...
                                                                                                                                      military
                                                                                                                                         ca...
         5 rows × 23 columns
In [9]: movies.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 4809 entries, 0 to 4808
        Data columns (total 23 columns):
         #
             Column
                                        Non-Null Count Dtype
         0
              budget
                                        4809 non-null
                                                           int64
                                        4809 non-null
                                                           object
              genres
         2
              homepage
                                        1713 non-null
                                                           object
         3
              id
                                        4809 non-null
                                                           int64
         4
              keywords
                                        4809 non-null
                                                           object
              original language
                                        4809 non-null
                                                           object
         6
              original_title
                                        4809 non-null
                                                           object
              overview
                                        4806 non-null
                                                           object
         8
              popularity
                                        4809 non-null
                                                           float64
              production companies 4809 non-null
                                                           object
         10
              production_countries 4809 non-null
                                                           object
         11
              release date
                                        4808 non-null
                                                           object
                                        4809 non-null
         12
              revenue
                                                           int64
         13
             runtime
                                        4807 non-null
                                                           float64
                                        4809 non-null
         14
              spoken languages
                                                           object
         15
              status
                                        4809 non-null
                                                           object
                                        3965 non-null
         16
             tagline
                                                           object
              title
                                        4809 non-null
                                                           object
                                        4809 non-null
                                                           float64
         18
              vote_average
         19
                                        4809 non-null
              vote_count
                                                           int64
         20
              movie id
                                        4809 non-null
                                                           int64
         21
             cast
                                        4809 non-null
                                                           object
                                        4809 non-null
         22 crew
                                                           object
        dtypes: float64(3), int64(5), object(15)
        memory usage: 864.2+ KB
```

Extracting relevant columns

```
#genres
#id
#keywords
#title
#overview
#cast
#crew
movies = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]

In [11]: movies.head()

Out[11]: movie_id title overview genres keywords cast crew
In the 22nd [["id": 28, [["id": 1463] [["cast_id": 242] [["credit_id": 242] [["cre
```

]:	n	novie_id	title	overview	genres	keywords	cast	crew
	0	19995	Avatar	In the 22nd century, a paraplegic Marine is di	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 1463, "name": "culture clash"}, {"id":	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "de
	1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[{"id": 12, "name": "Adventure"}, {"id": 14, "	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "de
	2	206647	Spectre	A cryptic message from Bond's past sends him o	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 470, "name": "spy"}, {"id": 818, "name	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "de
	3	49026	The Dark Knight Rises	Following the death of District Attorney Harve	[{"id": 28, "name": "Action"}, {"id": 80, "nam	[{"id": 849, "name": "dc comics"}, {"id": 853,	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "de
	4	49529	John Carter	John Carter is a war-weary, former military ca	[{"id": 28, "name": "Action"}, {"id": 12, "nam	[{"id": 818, "name": "based on novel"}, {"id":	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "de

Dealing with missing and duplicate data

```
In [12]: movies.isnull().sum()
Out[12]: movie_id
         title
         overview
                    3
         genres
         keywords
                    0
         cast
         crew
                     0
         dtype: int64
In [13]: movies.dropna(inplace = True)
In [14]: movies.isnull().sum()
Out[14]: movie id
                    0
         title
         overview
                    0
                   0
         genres
         keywords
                   0
                    0
         cast
         dtype: int64
In [15]: movies.duplicated().sum()
Out[15]: 0
```

Extracting 'genres'

```
In [16]: movies.iloc[0].genres
Out[16]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
In [17]: # Convert this
#'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "na
```

```
# ['Action', 'Adventure', 'Fantasy', 'Science Fiction']
In [18]: import ast
In [19]:
           def convert(obj):
                  L = []
                  for i in ast.literal eval(obj):
                       L.append(i['name'])
                  return L
In [20]: movies['genres'] = movies['genres'].apply(convert)
           movies.head()
In [21]:
Out[21]:
                movie id
                                       title
                                                       overview
                                                                            genres
                                                                                            keywords
                                                                                                                      cast
                                                                                                                                                       crew
                                                     In the 22nd
                                                                            [Action,
                                                                                           [{"id": 1463,
                                                                                                           [{"cast_id": 242
                                                                                                                                                 [{"credit_id":
                                                      century, a
                                                                         Adventure.
                                                                                                          "character": "Jake
            0
                    19995
                                     Avatar
                                                                                       "name": "culture
                                                                                                                              "52fe48009251416c750aca23",
                                               paraplegic Marine
                                                                  Fantasy, Science
                                                                                        clash"}, {"id":...
                                                                                                                 Sully", "...
                                                                                                                                                        "de...
                                                          is di
                                                                            Fiction1
                                                                                            [{"id": 270,
                                                                                                              [{"cast_id": 4,
                              Pirates of the
                                              Captain Barbossa,
                                                                                                                                                 [{"credit_id":
                                                                                                                "character":
                                                                        [Adventure.
                                                                                               "name":
            1
                      285
                              Caribbean: At
                                              long believed to be
                                                                                                                               "52fe4232c3a36847f800b579",
                                                                                         "ocean"}, {"id":
                                                                                                              "Captain Jack
                                                                    Fantasy, Action]
                               World's End
                                                      dead, ha...
                                                                                                                                                        "de...
                                                                                             726, "na...
                                                                                                                     Spa...
                                                                                                              [{"cast_id": 1,
                                                                                            [{"id": 470,
                                                                                                                                                 [{"credit_id":
                                              A cryptic message
                                                                            [Action.
                                                                                        "name": "spy"},
                                                                                                                "character":
            2
                  206647
                                    Spectre
                                                from Bond's past
                                                                         Adventure,
                                                                                                                             "54805967c3a36829b5002c41",
                                                                                             {"id": 818.
                                                                                                             "James Bond".
                                                                                                                                                        "de...
                                                   sends him o...
                                                                             Crime]
                                                                                               "name...
                                                                                                                       "cr...
                                                                                                              [{"cast_id": 2,
                                                                                            [{"id": 849,
                                                   Following the
                                                                                                                                                 [{"credit_id":
                                                                                        "name": "dc
comics"}, {"id":
                                  The Dark
                                                                     [Action, Crime,
                                                                                                                "character"
            3
                   49026
                                                 death of District
                                                                                                                               "52fe4781c3a36847f81398c3",
                                                                                                            "Bruce Wavne /
                               Knight Rises
                                                                    Drama, Thriller]
                                                Attorney Harve...
                                                                                                                                                        "de...
                                                                                                853,...
                                                                                                                      Ba...
                                                                                            [{"id": 818,
                                                 John Carter is a
                                                                            [Action,
                                                                                                              [{"cast_id": 5,
                                                                                                                                                 [{"credit_id":
                                                                                        "name": "based
                   49529
                                John Carter
                                              war-weary, former
                                                                         Adventure
                                                                                                          "character": "John
                                                                                                                              "52fe479ac3a36847f813eaa3",
                                                                                            on novel"},
                                                    military ca...
                                                                    Science Fiction]
                                                                                                               Carter", "c...
                                                                                                                                                        "de...
                                                                                                {"id":...
```

Extracting 'keywords'

```
In [22]: movies['keywords'] = movies['keywords'].apply(convert)
In [23]:
             movies.head()
Out[23]:
                movie_id
                                        title
                                                        overview
                                                                              genres
                                                                                                keywords
                                                                                                                          cast
                                                                                                                                                            crew
                                                       In the 22nd
                                                                              [Action,
                                                                                                             [{"cast_id": 242,
"character": "Jake
                                                                                                                                                     [{"credit_id":
                                                                                            sculture clash.
                                                                           Adventure,
                                                        century, a
                                                                                                                                  "52fe48009251416c750aca23",
             0
                    19995
                                      Avatar
                                                                                        future, space war,
                                                paraplegic Marine
                                                                             Fantasy,
                                                                                                                     Sully", "...
                                                                                            space colon...
                                                                                                                                                            "de
                                                            is di...
                                                                     Science Fiction]
                                                                                              [ocean, drug
                                                                                                                 [{"cast_id": 4,
                               Pirates of the
                                               Captain Barbossa,
                                                                                                                                                     [{"credit id":
                                                                          [Adventure,
                                                                                             abuse, exotic
                                                                                                                   "character":
                                                  long believed to
                       285
                              Caribbean: At
                                                                                                                                   "52fe4232c3a36847f800b579",
             1
                                                                     Fantasy, Action]
                                                                                         island, east india
                                                                                                                 "Captain Jack
                                World's End
                                                    be dead. ha...
                                                                                                                                                            "de...
                                                                                                                         Spa...
                                                                                                                 [{"cast_id": 1,
                                                                                           [spy, based on
                                               A cryptic message
                                                                                                                                                     [{"credit id":
                                                                              [Action.
                                                                                             novel, secret
                                                                                                                   "character":
                   206647
                                                                                                                                 "54805967c3a36829b5002c41",
                                     Spectre
                                                 from Bond's past
                                                                           Adventure,
                                                                                            agent, sequel,
                                                                                                                "James Bond",
                                                                               Crime1
                                                                                                                                                            "de...
                                                    sends him o...
                                                                                                      mi...
                                                                                                                          "cr...
                                                                                                                 [{"cast_id": 2
                                                    Following the
                                                                                         [dc comics, crime
                                                                                                                                                     [{"credit id":
                                   The Dark
                                                                       [Action, Crime,
                                                                                                                   "character":
             3
                    49026
                                                  death of District
                                                                                          fighter, terrorist,
                                                                                                                                   "52fe4781c3a36847f81398c3",
                                Knight Rises
                                                                      Drama, Thriller]
                                                                                                               "Bruce Wayne /
                                                 Attorney Harve...
                                                                                                                                                            "de...
                                                                                                 secret i...
                                                                                                                          Ва...
                                                                                                                 [{"cast_id": 5,
                                                  John Carter is a
                                                                              [Action,
                                                                                         [based on novel,
                                                                                                                                                     [{"credit_id":
                                                                                                                   "character":
                    49529
                                 John Carter
                                                                                                                                  "52fe479ac3a36847f813eaa3".
                                               war-weary, former
                                                                           Adventure.
                                                                                         mars, medallion.
                                                                                                                 "John Carter",
                                                      military ca...
                                                                     Science Fiction]
                                                                                                                                                             "de...
                                                                                            space travel...
                                                                                                                           "c...
```

Extracting top 3 'cast' of the movie

```
break
In [25]: movies['cast'] = movies['cast'].apply(convert2)
In [26]: movies.head()
Out[26]:
                                      title
                movie id
                                                     overview
                                                                                           kevwords
                                                                                                                      cast
                                                                                                                                                       crew
                                                                          aenres
                                                    In the 22nd
                                                                          [Action,
                                                                                       [culture clash.
                                                                                                        [Sam Worthington,
                                                                                                                                                [{"credit_id":
                                                     century, a
                                                                       Adventure.
                                                                                        future, space
            0
                   19995
                                    Avatar
                                                                                                             Zoe Saldana,
                                                                                                                             "52fe48009251416c750aca23",
                                             paraplegic Marine
                                                                         Fantasy.
                                                                                          war, space
                                                                                                        Sigourney Weaver]
                                                                                                                                                       "de...
                                                                  Science Fiction1
                                                         is di
                                                                                              colon
                                                       Captain
                                                                                         [ocean, drug
                             Pirates of the
                                                                                                            [Johnny Depp,
                                                                                                                                                [{"credit id":
                                                Barbossa, long
                                                                      [Adventure.
                                                                                        abuse, exotic
                      285
                             Caribbean: At
                                                                                                           Orlando Bloom,
                                                                                                                              "52fe4232c3a36847f800b579",
                                                 believed to be
                                                                  Fantasy, Action]
                                                                                    island, east india
                               World's End
                                                                                                           Keira Knightley]
                                                                                                                                                       "de...
                                                    dead ha
                                                      A cryptic
                                                                                      [spy, based on
                                                                          [Action.
                                                                                                             [Daniel Craig,
                                                                                                                                                [{"credit id":
                                                 message from
                                                                                        novel, secret
                                                                                                                             "54805967c3a36829b5002c41",
            2
                  206647
                                   Spectre
                                                                       Adventure,
                                                                                                          Christoph Waltz
                                                   Bond's past
                                                                                       agent, sequel,
                                                                           Crime]
                                                                                                            Léa Seydoux]
                                                                                                                                                       "de...
                                                 sends him o...
                                                                                                 mi...
                                                  Following the
                                                                                    [dc comics, crime
                                                                                                           [Christian Bale,
                                                                                                                                                [{"credit id":
                                 The Dark
                                                                   [Action, Crime,
            3
                   49026
                                               death of District
                                                                                     fighter, terrorist.
                                                                                                            Michael Caine
                                                                                                                              "52fe4781c3a36847f81398c3",
                              Knight Rises
                                                                  Drama, Thriller]
                                              Attorney Harve...
                                                                                                            Gary Oldman]
                                                                                                                                                       "de...
                                                                                            secret i...
                                               John Carter is a
                                                                          [Action,
                                                                                     [based on novel,
                                                                                                            [Taylor Kitsch,
                                                                                                                                                [{"credit_id":
                                                                                                             Lynn Collins,
                   49529
                               John Carter war-weary, former
                                                                       Adventure.
                                                                                    mars. medallion.
                                                                                                                              "52fe479ac3a36847f813eaa3",
                                                                  Science Fiction]
                                                   military ca...
                                                                                                        Samantha Mortonl
                                                                                                                                                       "de...
                                                                                       space travel...
```

Extracting Director name from the 'crew'

else:

```
In [27]: def fetch director(obj):
                 L = []
                 for i in ast.literal eval(obj):
                      if i['job'] == 'Director':
                           L.append(i['name'])
                           break
                 return L
In [28]: movies['crew'] = movies['crew'].apply(fetch_director)
In [29]:
           movies.head()
Out[29]:
               movie_id
                                        title
                                                           overview
                                                                                   genres
                                                                                                       keywords
                                                                                                                                     cast
                                                                                                                                                   crew
                                                In the 22nd century, a
                                                                        [Action, Adventure,
                                                                                             [culture clash, future,
                                                                                                                    [Sam Worthington, Zoe
                                                                                                                                                 [James
            0
                   19995
                                      Avatar
                                                 paraplegic Marine is
                                                                         Fantasy, Science
                                                                                                space war, space
                                                                                                                       Saldana, Sigourney
                                                                                                                                               Cameron]
                                                                                   Fiction]
                                                                                                          colon...
                                                                                                                                  Weaver]
                                Pirates of the
                                                   Captain Barbossa.
                                                                                              [ocean, drug abuse,
                                                                               [Adventure,
                                                                                                                    [Johnny Depp, Orlando
                                                                                                                                                  [Gore
            1
                     285
                               Caribbean: At
                                                  long believed to be
                                                                                                exotic island, east
                                                                           Fantasy, Action]
                                                                                                                    Bloom, Keira Knightley]
                                                                                                                                               Verbinski]
                                 World's End
                                                          dead, ha...
                                                                                                          india ...
                                                                                                                             [Daniel Craig,
                                                                                            [spy, based on novel.
                                                   A cryptic message
                                                                        [Action, Adventure,
                                                                                                                                                   [Sam
                  206647
                                     Spectre
                                                    from Bond's past
                                                                                             secret agent, sequel,
                                                                                                                      Christoph Waltz, Léa
                                                                                   Crime]
                                                                                                                                                Mendes]
                                                       sends him o..
                                                                                                             mi...
                                                                                                                                 Seydoux]
                                               Following the death of
                                                                                                [dc comics, crime
                             The Dark Knight
                                                                            [Action, Crime,
                                                                                                                   [Christian Bale, Michael
                                                                                                                                            [Christopher
            3
                   49026
                                                     District Attorney
                                                                                                  fighter, terrorist.
                                       Rises
                                                                           Drama, Thriller]
                                                                                                                      Caine, Gary Oldman]
                                                                                                                                                 Nolan]
                                                             Harve..
                                                                                                        secret i...
                                                John Carter is a war-
                                                                                                 [based on novel,
                                                                                                                       [Taylor Kitsch, Lynn
                                                                        [Action, Adventure
                                                                                                                                                [Andrew
                   49529
                                 John Carter
                                                weary, former military
                                                                                                 mars, medallion.
                                                                                                                        Collins, Samantha
                                                                           Science Fiction]
                                                                                                                                                Stanton]
                                                                                                                                  Morton1
                                                                ca...
                                                                                                   space travel...
```

Converting string type 'overview' into list type 'overview'

```
In [30]: movies['overview'][0]
Out[30]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes to rn between following orders and protecting an alien civilization.'
In [31]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
In [32]: movies.head()
```

crew	cast	keywords	genres	overview	title	movie_id	Out[32]:
[James Cameron]	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[culture clash, future, space war, space colon	[Action, Adventure, Fantasy, Science Fiction]	[In, the, 22nd, century,, a, paraplegic, Marin	Avatar	19995	
[Gore Verbinski]	[Johnny Depp, Orlando Bloom, Keira Knightley]	[ocean, drug abuse, exotic island, east india	[Adventure, Fantasy, Action]	[Captain, Barbossa,, long, believed, to, be, d	Pirates of the Caribbean: At World's End	285	
[Sam Mendes]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[spy, based on novel, secret agent, sequel, mi	[Action, Adventure, Crime]	[A, cryptic, message, from, Bond's, past, send	Spectre	206647	
[Christopher Nolan]	[Christian Bale, Michael Caine, Gary Oldman]	[dc comics, crime fighter, terrorist, secret i	[Action, Crime, Drama, Thriller]	[Following, the, death, of, District, Attorney	The Dark Knight Rises	49026	
[Andrew Stanton]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[based on novel, mars, medallion, space travel	[Action, Adventure, Science Fiction]	[John, Carter, is, a, war-weary,, former, mili	John Carter	49529	

Dealing with Whitespaces

Spectre

```
movies['keywords'] = movies['keywords'].apply(lambda x : [i.replace(" ", "") for i in x])
            movies['cast'] = movies['cast'].apply(lambda x : [i.replace(" ", "") for i in x])
movies['crew'] = movies['crew'].apply(lambda x : [i.replace(" ", "") for i in x])
In [34]: movies.head()
Out[34]:
                movie id
                                          title
                                                          overview
                                                                                  genres
                                                                                                       keywords
                                                                                                                                      cast
                                                                                                                                                           crew
                                                      [In, the, 22nd,
                                                                       [Action, Adventure,
                                                                                             [cultureclash, future,
                                                                                                                         [SamWorthington,
            0
                    19995
                                       Avatar
                                                                                 Fantasy.
                                                                                                                              ZoeSaldana,
                                                                                                                                               [JamesCameron]
                                                        centurv.. a.
                                                                                                       spacewar.
                                                 paraplegic, Marin...
                                                                           ScienceFiction1
                                                                                                                        SigourneyWeaver]
                                                                                                  spacecolony, ...
                                 Pirates of the
                                                          [Captain,
                                                                                              [ocean, drugabuse,
                                                                                                                             [JohnnyDepp,
                                                                              [Adventure.
                                                                                                                            OrlandoBloom,
                      285
                                Caribbean: At
                                                    Barbossa,, long,
                                                                                                     exoticisland,
                                                                                                                                                [GoreVerbinski]
                                                                         Fantasy, Action]
                                  World's End
                                                believed, to, be, d...
                                                                                                   eastindiatrad...
                                                                                                                            KeiraKnightley]
```

[Action, Adventure,

In [33]: movies['genres'] = movies['genres'].apply(lambda x : [i.replace(" ", "") for i in x])

[A, cryptic,

send...

message, from,

Bond's, past,

[Following, the, [dccomics, [ChristianBale, The Dark Knight [Action, Crime, 3 49026 death, of, District, crimefighter, terrorist, MichaelCaine, [ChristopherNolan] Drama, Thriller] Rises Attorney... GaryOldman] secretiden... [John, Carter, is, a, [basedonnovel, [TaylorKitsch, [Action, Adventure, 49529 John Carter war-weary,, former, mars, medallion, LynnCollins, [AndrewStanton] ScienceFiction1 SamanthaMorton] spacetravel. p...

Crime]

[spy, basedonnovel,

secretagent, sequel,

mi6. ...

[DanielCraig,

LéaSeydoux]

[SamMendes]

ChristophWaltz,

```
In [35]: movies.info()
```

206647

2

```
<class 'pandas.core.frame.DataFrame'>
Index: 4806 entries, 0 to 4808
Data columns (total 7 columns):
               Non-Null Count Dtype
 #
    Column
 0
     movie_id 4806 non-null
                               int64
     title
               4806 non-null
                               object
     overview
               4806 non-null
                               object
               4806 non-null
                               object
     genres
 4
     keywords
               4806 non-null
                                object
               4806 non-null
     cast
                               object
               4806 non-null
     crew
                               object
dtypes: int64(1), object(6)
memory usage: 429.4+ KB
```

Creating a new coloumn as 'tags'

```
In [36]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
In [37]: movies.head()
```

tags	crew	cast	keywords	genres	overview	title	movie_id	Out[37]:
[In, the, 22nd, century,, a, paraplegic, Marin	[JamesCameron]	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[cultureclash, future, spacewar, spacecolony,	[Action, Adventure, Fantasy, ScienceFiction]	[In, the, 22nd, century,, a, paraplegic, Marin	Avatar	19995	0
[Captain, Barbossa,, long, believed, to, be, d	[GoreVerbinski]	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[ocean, drugabuse, exoticisland, eastindiatrad	[Adventure, Fantasy, Action]	[Captain, Barbossa,, long, believed, to, be, d	Pirates of the Caribbean: At World's End	285	1
[A, cryptic, message, from, Bond's, past, send	[SamMendes]	[DanielCraig, ChristophWaltz, LéaSeydoux]	[spy, basedonnovel, secretagent, sequel, mi6,	[Action, Adventure, Crime]	[A, cryptic, message, from, Bond's, past, send	Spectre	206647	2
[Following, the, death, of, District, Attorney	[ChristopherNolan]	[ChristianBale, MichaelCaine, GaryOldman]	[dccomics, crimefighter, terrorist, secretiden	[Action, Crime, Drama, Thriller]	[Following, the, death, of, District, Attorney	The Dark Knight Rises	49026	3
[John, Carter, is, a, war- weary,, former, mili	[AndrewStanton]	[TaylorKitsch, LynnCollins, SamanthaMorton]	[basedonnovel, mars, medallion, spacetravel, p	[Action, Adventure, ScienceFiction]	[John, Carter, is, a, war-weary,, former, mili	John Carter	49529	4

Creating new datframe

In [38]:	new	<pre>u_df = movies[['movie_id', 'title', 'tags']]</pre>				
in [39]:	new_df.head()					
ut[39]:	ı	novie_id	title	tags		
	0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin		
	1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d		
	2	206647	Spectre	[A,cryptic,message,from,Bond's,past,send		
	3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney		
	4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili		

Converting list type 'tags' into string type 'tags'

```
In [40]: new_df['tags'] = new_df['tags'].apply(lambda x :" ".join(x))
         C:\Users\Uttam\AppData\Local\Temp\ipykernel_9988\4095421860.py:1: SettingWithCopyWarning:
         A value is trying to be set on a copy of a slice from a DataFrame.
         Try using .loc[row_indexer,col_indexer] = value instead
         See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
         rning-a-view-versus-a-copy
           new_df['tags'] = new_df['tags'].apply(lambda x :" ".join(x))
In [41]: new_df.head()
Out[41]:
             movie_id
                                                     title
          0
                19995
                                                   Avatar
                                                              In the 22nd century, a paraplegic Marine is di...
                  285 Pirates of the Caribbean: At World's End
                                                            Captain Barbossa, long believed to be dead, ha...
          2
               206647
                                                  Spectre A cryptic message from Bond's past sends him o...
                                      The Dark Knight Rises
                49026
                                                             Following the death of District Attorney Harve...
                49529
                                               John Carter
                                                             John Carter is a war-weary, former military ca...
In [42]: new df["tags"][0]
```

Out[42]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes to rn between following orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction cultu reclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi ma rine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'

Converting 'tags' into lowercase

```
In [43]: new_df['tags'] = new_df['tags'].apply(lambda x : x.lower())
         C:\Users\Uttam\AppData\Local\Temp\ipykernel 9988\3679624391.py:1: SettingWithCopyWarning:
         A value is trying to be set on a copy of a slice from a DataFrame.
         Try using .loc[row indexer,col indexer] = value instead
         See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
         rning-a-view-versus-a-copy
           new df['tags'] = new df['tags'].apply(lambda x : x.lower())
In [44]: new_df.head()
Out[44]:
             movie_id
                                                      title
                                                                                                 tags
                19995
                                                    Avatar
                                                              in the 22nd century, a paraplegic marine is di...
                  285 Pirates of the Caribbean: At World's End
                                                            captain barbossa, long believed to be dead, ha...
          2
               206647
                                                   Spectre a cryptic message from bond's past sends him o...
                                                               following the death of district attorney harve...
           3
                49026
                                      The Dark Knight Rises
                49529
                                               John Carter
                                                              john carter is a war-weary, former military ca...
```

Stemming words in 'tags'

Simplifying words to their most basic form is called stemming. It reduces the words like "chocolates", "chocolatey", and "choco" to the root word, "chocolate".

```
In [45]: import nltk
In [46]: from nltk.stem.porter import PorterStemmer
         ps = PorterStemmer()
In [47]: def stem(text):
             y = []
             for i in text.split():
                 y.append(ps.stem(i))
             return " ".join(y)
In [48]: new df['tags'].apply(stem)
Out[48]:
                  in the 22nd century, a parapleg marin is dispa...
         1
                  captain barbossa, long believ to be dead, ha c...
                  a cryptic messag from bond' past send him on a...
         2
         3
                  follow the death of district attorney harvey d...
                  john carter is a war-weary, former militari ca...
          4804
                  el mariachi just want to play hi guitar and ca...
         4805
                  a newlyw couple' honeymoon is upend by the arr...
          4806
                  "signed, sealed, delivered" introduc a dedic q...
         4807
                  when ambiti new york attorney sam is sent to s...
          4808
                  ever sinc the second grade when he first saw h...
         Name: tags, Length: 4806, dtype: object
```

Text Vectorization of coloumn 'tags' without stop words

Stop words are a set of commonly used words in a language. Examples of stop words in English are "a," "the," "is," "are," etc.

```
In [49]: from sklearn.feature_extraction.text import CountVectorizer
    cv = CountVectorizer(max_features = 5000, stop_words = 'english')
In [50]: vectors = cv.fit_transform(new_df['tags']).toarray()
In [51]: vectors
```

```
Out[51]: array([[0, 0, 0, ..., 0, 0, 0],
                  [0, 0, 0, \ldots, 0, 0, 0],
                  [0, 0, 0, \ldots, 0, 0, 0],
                  [0, 0, 0, ..., 0, 0, 0],
                  [0, 0, 0, ..., 0, 0, 0],
[0, 0, 0, ..., 0, 0, 0]], dtype=int64)
In [52]: vectors[0]
Out[52]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
In [53]: a = cv.get feature names out()
In [54]: for feature name in a[10:20]:
              print(feature_name)
         17
         18
         18th
         19
         1930s
         1940s
         1950
         1950s
         1960s
         1970s
In [55]: print(len(a))
```

Measuring Cosine similarity

It is a way to measure how similar two vectors are in an inner product space.

Movie Recommendation Funtion

```
In [59]: def recommend(movie):
    movie_index = new_df[new_df['title'] == movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)), reverse=True, key=lambda x:x[1])[1:6]

    for i in movies_list:
        print(new_df.iloc[i[0]].title)

In [60]: recommend('Avatar')

Titan A.E.
Small Soldiers
Independence Day
Ender's Game
Aliens vs Predator: Requiem
```

Sending list of movies to the website

```
In [61]: import pickle
In [62]: pickle.dump(new_df,open('movies.pkl', 'wb'))
In [63]: pickle.dump(similarity,open('similarity.pkl', 'wb'))
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js