Задача 1: IMDb

Итак, есть дата-сеты https://datasets.imdbws.com/ с фильмами, сериалами, ТВ-шоу и всем, что с ними может быть связано из IMDb и доки к ним: http://www.imdb.com/interfaces/. Твоим заданием будет скачать их и разобраться: дата-сеты с какими данными есть, за что отвечает каждая колонка в них, как они связаны и что можно получить имея эти данные?

Как только ты разобрался, пора бы заняться БИГ ДАТОЙ и превратить это все во что-то реально полезное:

1. Найди топ-100 фильмов(да-да, именно фильмов, а не сериалов или ТВ-шоу), заметь что фильмы с рейтингом 9,9 и со 100 голосами мы не можем считать популярными, пусть если за фильм проголосовало хотя бы 100 000 человек - он популярный
    a. за все время,
    b. за последние 10 лет,
    c. фильмы которые были популярны в 60х годах прошлого века.

    Пример аутпута первого задания:

    ```
    +---------+-----------------------------------------+--------+-------------+---------+
    |tconst   |primaryTitle                             |numVotes|averageRating|startYear|
    +---------+-----------------------------------------+--------+-------------+---------+
    |tt0111161|The Shawshank Redemption                 |2118893 |9.3          |1994     |
    |tt5813916|The Mountain II                          |101048  |9.2          |2016     |
    |tt5813916|The Mountain II                          |101048  |9.2          |2016     |
    |tt0068646|The Godfather                            |1454828 |9.2          |1972     |
    |tt0068646|The Godfather                            |1454828 |9.2          |1972     |
    |tt5813916|The Mountain II                          |101048  |9.2          |2016     |
    |tt0468569|The Dark Knight                          |2084449 |9.0          |2008     |
    |tt0071562|The Godfather: Part II                   |1011697 |9.0          |1974     |
    |tt0071562|The Godfather: Part II                   |1011697 |9.0          |1974     |
    |tt0468569|The Dark Knight                          |2084449 |9.0          |2008     |
    |tt0468569|The Dark Knight                          |2084449 |9.0          |2008     |
    |tt0167260|The Lord of the Rings: The Return of the King|1507969 |8.9          |2003     |
    |tt0050083|12 Angry Men                             |603066  |8.9          |1957     |
    |tt0108052|Schindler's List                         |1099364 |8.9          |1993     |
    |tt0110912|Pulp Fiction                             |1658087 |8.9          |1994     |
    |tt0167260|The Lord of the Rings: The Return of the King|1507969 |8.9          |2003     |
    |tt0108052|Schindler's List                         |1099364 |8.9          |1993     |
    |tt0108052|Schindler's List                         |1099364 |8.9          |1993     |
    |tt0167260|The Lord of the Rings: The Return of the King|1507969 |8.9          |2003     |
    |tt0110912|Pulp Fiction                             |1658087 |8.9          |1994     |
    +---------+-----------------------------------------+--------+-------------+---------+
    ```

2. Лучшие фильмы нам известны, а что если ты захочешь посмотреть, ну, скажем лучший триллер всех времен и народов. Так вот нужно найти по топ-10 фильмов каждого жанра (результат должен быть в одном файлике =))

    Пример аутупта:

    ```
    +---------+-----------------------------------+---------+-------+-------------+--------+
    |tconst   |primaryTitle                       |startYear|genre  |averageRating|numVotes|
    +---------+-----------------------------------+---------+-------+-------------+--------+
    |tt0068646|The Godfather                      |1972     |Crime  |9.2          |1454828 |
    |tt0071562|The Godfather: Part II             |1974     |Crime  |9.0          |1011697 |
    |tt0468569|The Dark Knight                    |2008     |Crime  |9.0          |2084449 |
    |tt0110912|Pulp Fiction                       |1994     |Crime  |8.9          |1658087 |
    |tt0099685|Goodfellas                         |1990     |Crime  |8.7          |915187  |
    |tt0317248|City of God                        |2002     |Crime  |8.6          |650070  |
    |tt0110413|Léon: The Professional             |1994     |Crime  |8.6          |935941  |
    |tt0120689|The Green Mile                     |1999     |Crime  |8.6          |1027281 |
    |tt0114369|Se7en                              |1995     |Crime  |8.6          |1300945 |
    |tt0114814|The Usual Suspects                 |1995     |Crime  |8.6          |914904  |
    |tt0109830|Forrest Gump                       |1994     |Romance|8.8          |1628871 |
    |tt0118799|Life Is Beautiful                  |1997     |Romance|8.6          |559173  |
    |tt0034583|Casablanca                         |1942     |Romance|8.5          |480778  |
    |tt0021749|City Lights                        |1931     |Romance|8.5          |149576  |
    |tt0119217|Good Will Hunting                  |1997     |Romance|8.3          |774968  |
    |tt0338013|Eternal Sunshine of the Spotless Mind|2004     |Romance|8.3          |831843  |
    |tt0045152|Singin' in the Rain                |1952     |Romance|8.3          |199267  |
    |tt0052357|Vertigo                            |1958     |Romance|8.3          |324017  |
    |tt0053604|The Apartment                      |1960     |Romance|8.3          |146002  |
    |tt0211915|Amélie                             |2001     |Romance|8.3          |656694  |
    +---------+-----------------------------------+---------+-------+-------------+--------+
    ```

3. А теперь усложним задачу. Нужно найти все то же самое, но только для каждого десятилетия с сейчас до 1950х (тоже в одном файле)
   Пример аутупта:

```
+---------+---------------------------------+---------+--------+-------------+--------+-----------+
|tconst   |primaryTitle                     |startYear|genre   |averageRating|numVotes|yearRange  |
+---------+---------------------------------+---------+--------+-------------+--------+-----------+
|tt1853728|Django Unchained                 |2012     |Western |8.4          |1222787 |2010 - 2020|
|tt1403865|True Grit                        |2010     |Western |7.6          |286195  |2010 - 2020|
|tt2404435|The Magnificent Seven            |2016     |Western |6.9          |169905  |2010 - 2020|
|tt1210819|The Lone Ranger                  |2013     |Western |6.4          |209882  |2010 - 2020|
|tt2557490|A Million Ways to Die in the West|2014     |Western |6.1          |161345  |2010 - 2020|
|tt5813916|The Mountain II                  |2016     |War     |9.2          |101048  |2010 - 2020|
|tt1255953|Incendies                        |2010     |War     |8.3          |125855  |2010 - 2020|
|tt0816442|The Book Thief                   |2013     |War     |7.6          |119698  |2010 - 2020|
|tt2713180|Fury                             |2014     |War     |7.6          |385332  |2010 - 2020|
|tt1568911|War Horse                        |2011     |War     |7.2          |137248  |2010 - 2020|
|tt0989757|Dear John                        |2010     |War     |6.3          |129581  |2010 - 2020|
|tt1345836|The Dark Knight Rises            |2012     |Thriller|8.4          |1397978 |2010 - 2020|
|tt1832382|A Separation                     |2011     |Thriller|8.3          |199516  |2010 - 2020|
|tt3170832|Room                             |2015     |Thriller|8.2          |321079  |2010 - 2020|
|tt3011894|Wild Tales                       |2014     |Thriller|8.1          |153034  |2010 - 2020|
|tt1130884|Shutter Island                   |2010     |Thriller|8.1          |1015753 |2010 - 2020|
|tt2267998|Gone Girl                        |2014     |Thriller|8.1          |768955  |2010 - 2020|
|tt2265171|The Raid 2                       |2014     |Thriller|8.0          |105331  |2010 - 2020|
|tt2084970|The Imitation Game               |2014     |Thriller|8.0          |626124  |2010 - 2020|
|tt0947798|Black Swan                       |2010     |Thriller|8.0          |652953  |2010 - 2020|
+---------+---------------------------------+---------+--------+-------------+--------+-----------+
```

4. Представь, что ты собрался снимать фильм и необходимо подобрать актерский состав. Твоей задачей будет выбрать самых востребованых актеров, будем считать, что актер востребованый, если он снимался в топовых фильмах и не один раз
   Пример аутупта:

```
+--------------------+
|         primaryName|
+--------------------+
|       John Travolta|
|           Al Pacino|
|         Gary Sinise|
|       Robert Duvall|
|   Robert Downey Jr.|
|          Bedii Akin|
|     Lorraine Bracco|
|        Bruce Willis|
|        Ken Watanabe|
|       Orlando Bloom|
|         Chris Evans|
|         Elijah Wood|
|      Robert De Niro|
|   Samuel L. Jackson|
|         Sally Field|
|           Ozan Agaç|
|       Ralph Fiennes|
|Joseph Gordon-Levitt|
|          Ellen Page|
|          James Caan|
+--------------------+
```

5. Ну и на последок, найди топ-5 фильмов по рейтингу у каждого режиссера
   Пример аутупта:

primaryName	primaryTitle	startYear	averageRating	numVotes
Matt Damon	Manchester by the Sea	2016	7.8	217102
Matt Damon	Stolen Summer	2002	6.5	2575
Danny DeVito	Gattaca	1997	7.8	262410
Danny DeVito	Freedom Writers	2007	7.5	62340
Danny DeVito	Erin Brockovich	2000	7.3	159864
Danny DeVito	Out of Sight	1998	7.0	80208
Danny DeVito	Reality Bites	1994	6.6	42458
Kevin Loader	Nowhere Boy	2009	7.1	33025
Kevin Loader	Venus	2006	7.1	12012
Kevin Loader	The History Boys	2006	6.8	19528
Kevin Loader	The Mother	2003	6.8	3792
Kevin Loader	Ferrari: Race to Immortality	2017	6.8	677
Christopher Webb Young	Life Sentence	2004	5.9	14
Michael Huens	Model Minority	2012	7.6	880
Michael Huens	Not Again!	1996	5.0	20
Karim Abouobayd	Sara	2013	6.6	17
Karim Abouobayd	Un Marocain à Paris	2012	6.0	36
Liisa Akimof	Onni von Sopanen	2006	6.6	124
Arnold Albert	The Man I Love	1947	6.9	661
William Aldridge	Interdevochka	1989	7.1	894

Основной упор в этом задании дается на:

- инструменты и методы обработки больших объемов данных
- считывать поврежденные данные и правильно их обрабатывать
- грамотно организовать обработку данных
- научиться правильно определять порядок операций при работе с большими данными

P.S. Результаты, полученые при выполнении заданий, сохранять в CSV формате