Sustainable Data Science and Artificial Intelligence

Tutorial #3

|  |  |
| --- | --- |
| **People in group:** |  |

**Movie Dataset**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Movie Title** | **Genre** | **Year** | **Box Office**  **($M)** | **IMDB**  **Rating** | **Animated?** | **PG-13?** | **Sequel?** | **Lead**  **Actor**  **Famous?** |
| The Lion King | Animation | 1994 | 968 | 8.5 | Yes | No | Yes | Yes |
| Avengers: Endgame | Action | 2019 | 2797 | 8.4 | No | Yes | Yes | Yes |
| Inception | Sci-Fi | 2010 | 837 | 8.8 | No | Yes | No | Yes |
| Finding Nemo | Animation | 2003 | 940 | 8.2 | Yes | No | Yes | No |
| Titanic | Romance | 1997 | 2202 | 7.9 | No | No | No | Yes |
| The Dark Knight | Action | 2008 | 1004 | 9.0 | No | Yes | Yes | Yes |
| Toy Story | Animation | 1995 | 373 | 8.3 | Yes | No | Yes | No |
| Jurassic Park | Sci-Fi | 1993 | 1046 | 8.2 | No | Yes | Yes | Yes |
| The Matrix | Sci-Fi | 1999 | 467 | 8.7 | No | Yes | Yes | Yes |
| Shrek | Animation | 2001 | 488 | 7.9 | Yes | No | Yes | No |
| Black Panther | Action | 2018 | 1347 | 7.3 | No | Yes | Yes | Yes |
| Spider-Man | Action | 2002 | 825 | 7.4 | No | Yes | Yes | Yes |
| Frozen | Animation | 2013 | 1280 | 7.4 | Yes | No | Yes | No |
| Star Wars: A New Hope | Sci-Fi | 1977 | 775 | 8.6 | No | Yes | Yes | Yes |
| Beauty and the Beast | Romance | 1991 | 424 | 8.0 | Yes | No | Yes | No |
| Iron Man | Action | 2008 | 585 | 7.9 | No | Yes | Yes | Yes |

**Instructions**

Data Classification.

* Which movies made more than $1 billion?
* How many are sequels and how many are not?
* Does the impact of a presence of a Famous Lead Actor help make a $1 billion?

Feature Engineering:

* What other categories could be added to make the dataset more useful? e.g. could there be more than one Genre?
* Which feature or features are most useful for classification?

Projection:

* If you had to predict a new movie’s success, which attributes would matter the most?
* Are there any movies that are difficult to classify? Why?

**Submission Instructions**

Please submit in the following two ways

|  |
| --- |
| Please e-mail me your completed solution, and with filename:  *Familyname\_FirstName\_TU850\_1\_SDSaAI\_Tutorial3.docx*  e.g.  *Gordon\_Damian\_TU850\_1\_SDSaAI\_ Tutorial3.docx*  e-mail to [Damian.X.Gordon@TUDublin.ie](mailto:Damian.X.Gordon@TUDublin.ie) with subject heading as follows:  [DT850/1] SDSaAI Tutorial3 |
| Please submit into Brightspace in:   * Assessment   + Assignments     - Tutorial #3 |

For the email, please include the following message:

|  |
| --- |
| **Damian,**  **I am a student in your Sustainable DS & AI (CMPU1040) class in the BSc in DS & AI (DT850/1).**    **Please find attached Tutorial #3**  **Regards,**  **Your Name**  **Student Number.**  **DT580, BSc in DS & AI** |