1. Title column under Kaggler is more consistent than wiki. And we would use Kaggler title column.
2. For running time, Kaggler has more data and thus we would use Kaggler data. But if Kaggler data is 0, we will fill with wiki data.
3. For budget, Kaggler has more data and thus we would use Kaggler data. But if Kaggler data is 0, we will fill with wiki data.
4. For revenue, Kaggler has more data and thus we would use Kaggler data. But if Kaggler data is 0, we will fill with wiki data.
5. For release date, Kaggler has no none and thus we will just use Kaggler data and drop wiki data.
6. For original language, While the Wikipedia data has more information about multiple languages, the Kaggle data is already in a consistent and usable format. Parsing the Wikipedia data may create too many difficulties to make it worthwhile, though. We will use Kaggler data and drop wiki data.
7. For production company, Kaggler data is more consistent and thus we will use Kaggler data and drop wiki data.
8. Bad format budget data under Kaggler is small enough and thus we just drop it instead of parse it.
9. Duplicate imdb\_id under Kaggler is small and thus we can delete duplicate value and our data is still good.
10. For columns with null value that is larger than 10%, we will remove the columns for Kaggler data.