

When I start my cleaning data I found:

1- I make description and found "budget" and "revenue" have zero value in min, 25% and 50% so i filter more than zero.

2- Use histogram to understand data:

a- Right-skewed: budget, budget_adj, id, popularity, revenue, revenue_adj and vote_count

b- Left-skewed: release_year

c- symmetric: vote_average

3- No duplicated value

4- Some missing value so i deleted

5- "release_year", "release_date" and "runtime" I change type to datetime

After that i start to answer questions. and my Conclusions are drama is the most popular genres. we have positive correlation between budget, popularity and revenue. That mean if budget or popularity increase, we have high revenue