Data Analyst Nanodegree Program

Project : Investigate a Dataset

Data set : tmdb-movies

Name Abdalla H. Ammar

By exploring that data set we finded that resalt for genres are most popular from year to year

ID	Year	Genera
0	1960	Thriller
1	1961	Animation
2	1962	Adventure
3	1963	Animation
4	1964	War
5	1965	Music
6	1966	Animation
7	1967	Animation
8	1968	Mystery
9	1969	Crime
10	1970	Animation
11	1971	Family
12	1972	Crime
13	1973	Animation
14	1974	Mystery
15	1975	Adventure
16	1976	Crime
17	1977	Science Fiction
18	1978	Music
19	1979	Action
20	1980	Science Fiction
21	1981	Adventure
22	1982	War
23	1983	Adventure
24	1984	Family
25	1985	Family
26	1986	Animation
27	1987	History
28	1988	Animation
29	1989	Animation
30	1990	Adventure
31	1991	Animation
32	1992	Animation
33	1993	Fantasy
34	1994	Crime
35	1995	Animation
36	1996	Crime
37	1997	Animation
38	1998	War
39	1999	Adventure

40	2000	War
41	2001	Fantasy
42	2002	Fantasy
43	2003	Fantasy
44	2004	Fantasy
45	2005	Fantasy
46	2006	Fantasy
47	2007	Fantasy
48	2008	Adventure
49	2009	Adventure
50	2010	Adventure
51	2011	Adventure
52	2012	Western
53	2013	Science Fiction
54	2014	Science Fiction
55	2015	Adventure

Conclusion

And we find that the most kinds of properties are associated with movies that have high revenues are

1- Runtime and it's very associated with revenues , When runtime between 70 to 200 Min we will gain a big revenues
2- Budget has a little +ve liner with revenues so it's associated
3- Years have a big positive associated with revenues except that in between 2008 to 2010 and I think that happened because Financial crisis of 2007–08 and maybe Netflix
4- Revenue and popularity have a very big associated and it's strong positive relation

Limitations

1- We have a many miss values in a many Important fields like ' company production'
2- That may reflect on our analyst
3- The data was good to prove what I want
4- The proplem like sum genras as a one field was a proplem but we dealed with it

Edit a cording to First Feedback

1- I ploted one plot according 3 variable

2- You told me that "**The analysis does not state or imply that one change causes another based** solely on a correlation"

But, in Runtime' AND Revenue plot that show that the revenue depend on run time, If it too long or too short the revenue will be small

3- I added titles and labels to all your plots.

Edit a cording to secound Feedback

1- Add a new type plot ‘bar’
2- Add Limitations to report
3- Add a few sentences to explain after plot.