

Assignment 3 - Problem 2

Cleaning Datasets:

1. AnnualTicketSales:

No cleaning required.

Table:

	year	TICKETS_SOLD	TOTAL BOX OFFICE	TOTAL INFLATION ADJUSTED BOX OFFICE	AVERAGE TICKET PRICE
►	2021	42,37,74,881	\$3,881,777,912	\$3,881,777,912	\$9.16
	2020	22,36,38,958	\$2,048,534,616	\$2,048,534,616	\$9.16
	2019	1,22,85,41,629	\$11,253,443,955	\$11,253,444,050	\$9.16
	2018	1,31,15,36,128	\$11,948,096,650	\$12,013,670,952	\$9.11
	2017	1,22,56,39,761	\$10,993,991,460	\$11,226,860,216	\$8.97
	2016	1,30,25,56,378	\$11,267,115,924	\$11,931,416,424	\$8.65
	2015	1,32,33,56,776	\$11,155,900,636	\$12,121,948,075	\$8.43
	2014	1,25,74,02,920	\$10,272,985,008	\$11,517,810,744	\$8.17
	2013	1,33,91,68,926	\$10,887,446,341	\$12,266,787,382	\$8.13
	2012	1,38,09,21,942	\$10,992,141,616	\$12,649,244,986	\$7.96
	2011	1,28,29,15,168	\$10,173,519,704	\$11,751,502,955	\$7.93
	2010	1,32,85,49,021	\$10,482,254,025	\$12,169,509,032	\$7.89
	2009	1,41,85,67,388	\$10,639,257,284	\$12,994,051,137	\$7.50

2. HighestGrossers

GENRE column - blank to null

Table after cleaning:

	year	MOVIE	GENRE	MPAA RATING	DISTRIBUTOR	TOTAL FOR YEAR	TOTAL IN 2019 DOLLARS	TICKETS SOLD
►	1995	Batman Forever	Drama	PG-13	Warner Bros.	\$184,031,112	\$387,522,978	4,23,06,002
	1996	Independence Day	Adventure	PG-13	20th Century Fox	\$306,169,255	\$634,504,608	6,92,69,062
	1997	Men in Black	Adventure	PG-13	Sony Pictures	\$250,650,052	\$500,207,943	5,46,07,854
	1998	Titanic	Adventure	PG-13	Paramount Pictures	\$443,319,081	\$865,842,808	9,45,24,324
	1999	Star Wars Ep. I: The Phantom Menace	Adventure	PG	20th Century Fox	\$430,443,350	\$776,153,749	8,47,32,942
	2000	How the Grinch Stole Christmas	Adventure	PG	Universal	\$253,367,455	\$430,583,644	4,70,06,948
	2001	Harry Potter and the Sorcerer's Stone	Adventure	PG	Warner Bros.	\$300,404,434	\$486,166,890	5,30,74,988
	2002	Spider-Man	Adventure	PG-13	Sony Pictures	\$403,706,375	\$636,480,273	6,94,84,746
	2003	Finding Nemo	Adventure	G	Walt Disney	\$339,714,367	\$516,050,346	5,63,37,374
	2004	Shrek 2	Adventure	PG	Dreamworks SKG	\$441,226,247	\$650,826,473	7,10,50,925
	2005	Star Wars Ep. III: Revenge of the Sith	Action	PG-13	20th Century Fox	\$380,270,577	\$543,413,171	5,93,24,582
	2006	Pirates of the Caribbean: Dead Man's Chest	Action	PG-13	Walt Disney	\$423,315,812	\$591,995,851	6,46,28,368
	2007	Spider-Man 3	Adventure	PG-13	Sony Pictures	\$336,530,303	\$448,054,878	4,89,14,288

3. PopularCreativeTypes

MARKET SHARE column - removed '%'.

Table after cleaning:

	RANK	CREATIVE TYPES	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
►	1	Contemporary Fiction	7,442	\$96,203,727,036	\$12,927,133	40.46
	2	Kids Fiction	564	\$32,035,539,746	\$56,800,602	13.47
	3	Science Fiction	724	\$29,922,660,857	\$41,329,642	12.59
	4	Fantasy	759	\$21,724,062,575	\$28,621,953	9.14
	5	Super Hero	129	\$20,273,157,911	\$157,156,263	8.53
	6	Historical Fiction	1,487	\$18,521,260,744	\$12,455,454	7.79
	7	Dramatization	1,175	\$15,715,191,699	\$13,374,631	6.61
	8	Factual	2,467	\$2,960,327,207	\$1,199,970	1.25
	9	Multiple Creative Types	42	\$117,574,526	\$2,799,393	0.05

4. TopDistributors

MARKET SHARE column - removed '%'.
Table after cleaning:

	rank	DISTRIBUTORS	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
►	1	Walt Disney	588	\$40,472,424,278	\$68,830,654	17.02
	2	Warner Bros.	824	\$36,269,425,479	\$44,016,293	15.25
	3	Sony Pictures	747	\$29,113,002,302	\$38,973,229	12.24
	4	Universal	535	\$28,089,932,569	\$52,504,547	11.81
	5	20th Century Fox	525	\$25,857,839,756	\$49,253,028	10.88
	6	Paramount Pictures	493	\$24,361,425,304	\$49,414,656	10.25
	7	Lionsgate	426	\$9,631,837,781	\$22,609,948	4.05
	8	New Line	209	\$6,195,268,024	\$29,642,431	2.61
	9	Dreamworks SKG	77	\$4,278,649,271	\$55,566,874	1.8
	10	Miramax	385	\$3,836,019,208	\$9,963,686	1.61

5. TopGenres

MARKET SHARE column - removed '%'.
Table after cleaning:

	RANK	GENRES	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
►	1	Adventure	1,102	\$64,529,536,530	\$58,556,748	27.14
	2	Action	1,098	\$49,339,974,493	\$44,936,224	20.75
	3	Drama	5,479	\$35,586,177,269	\$6,495,013	14.97
	4	Comedy	2,418	\$33,687,992,318	\$13,932,172	14.17
	5	Thriller/Suspense	1,186	\$19,810,201,102	\$16,703,374	8.33
	6	Horror	716	\$13,430,378,699	\$18,757,512	5.65
	7	Romantic Comedy	630	\$10,480,124,374	\$16,635,118	4.41
	8	Musical	201	\$4,293,988,317	\$21,363,126	1.81
	9	Documentary	2,415	\$2,519,513,142	\$1,043,277	1.06
	10	Black Comedy	213	\$2,185,433,323	\$10,260,250	0.92

6. TopGrossingRatings

MARKET SHARE column - removed '%'.
Table after cleaning:

	RANK	MPAA RATINGS	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
▶	1	PG-13	3,243	\$113,524,789,243	\$35,006,102	47.75
	2	R	5,480	\$63,497,164,978	\$11,587,074	26.71
	3	PG	1,535	\$49,124,317,794	\$32,002,813	20.66
	4	G	395	\$9,572,240,391	\$24,233,520	4.03
	5	Not Rated	5,820	\$1,918,358,283	\$329,615	0.81
	6	NC-17	24	\$44,850,139	\$1,868,756	0.02
	7	Open	5	\$5,489,687	\$1,097,937	0
	8	GP	7	\$552,618	\$78,945	0

7. TopGrossingSources

MARKET SHARE column - removed '%'.

Table after cleaning:

	RANK	SOURCES	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
▶	1	Original screenplay	7,946	\$106,375,196,782	\$13,387,264	44.74
	2	Based on Fiction Book/Short Story	2,150	\$47,005,613,207	\$21,863,076	19.77
	3	Based on Comic/Graphic Novel	249	\$23,369,989,130	\$93,855,378	9.83
	4	Remake	328	\$12,832,659,970	\$39,123,963	5.4
	5	Based on Real Life Events	3,225	\$11,398,356,297	\$3,534,374	4.79
	6	Based on TV	231	\$11,305,006,312	\$48,939,421	4.75
	7	Based on Factual Book/Article	295	\$7,443,681,990	\$25,232,820	3.13
	8	Spin-Off	41	\$3,833,128,331	\$93,490,935	1.61
	9	Based on Folk Tale/Legend/Fairytale	78	\$3,406,118,495	\$43,668,186	1.43
	10	Based on Play	271	\$2,111,190,923	\$7,790,372	0.89

8. TopProductionMethods

MARKET SHARE column - removed '%'.

Table after cleaning:

	RANK	PRODUCTION METHODS	MOVIES	TOTAL GROSS	AVERAGE GROSS	MARKET SHARE
▶	1	Live Action	14,613	\$179,637,201,848	\$12,292,972	75.56
	2	Animation/Live Action	264	\$30,346,622,254	\$114,949,327	12.76
	3	Digital Animation	365	\$23,920,180,508	\$65,534,741	10.06
	4	Hand Animation	164	\$2,960,497,487	\$18,051,814	1.25
	5	Stop-Motion Animation	37	\$676,490,120	\$18,283,517	0.28
	6	Multiple Production Methods	26	\$43,728,300	\$1,681,858	0.02
	7	Rotoscoping	4	\$8,468,385	\$2,117,096	0

9. WideReleasesCount

No cleaning required.

Table:

	year	WARNER BROS	WALT DISNEY	20TH CENTURY FOX	PARAMOUNT PICTURES	SONY PICTURES	UNIVERSAL	TOTAL MAJOR 6	TOTAL OTHER STUDIOS
▶	2021	17	7	0	4	16	17	61	38
	2020	5	3	1	3	9	13	34	23
	2019	18	10	11	9	18	21	87	44
	2018	19	10	11	10	16	20	86	58
	2017	18	8	14	10	16	15	81	50
	2016	17	12	16	12	16	22	95	46
	2015	22	11	18	9	13	20	93	33
	2014	17	12	17	10	17	15	88	37
	2013	17	8	15	8	14	16	78	42
	2012	16	11	15	13	18	17	90	42
	2011	20	13	15	13	21	19	101	35
	2010	20	12	18	12	17	17	96	30
	2009	25	14	20	10	21	21	111	30

Code Output Screenshot:

```
How many transactions do you want to start?
2
First Thread created!
Second Thread created!
Since the first update query is from second transaction, locking 2nd transaction..
Starting first transaction..
Starting second transaction..
Transaction 1 Select query running without lock Select TICKETS_SOLD from annuالتicketsales where YEAR=2011;
Cant execute first transaction query since lock is enforced on other transaction!
Transaction 2 Select query running without lock Select TICKETS_SOLD from annuالتicketsales where YEAR=2011;
Second transaction queries executing since lock is enforced on second transaction!
Transaction 2 query executed: Update annuالتicketsales set TICKETS_SOLD = 0 where YEAR=2011;
Released lock of transaction 2
Notifying other transactions..
Transaction 2 Select query running without lock Select TICKETS_SOLD from annuالتicketsales where YEAR=2011;
Second transaction queries executing since lock is enforced on second transaction!
Transaction 2 query executed: Update annuالتicketsales set TICKETS_SOLD = 0 where YEAR=2011;
Released lock of transaction 2
Notifying other transactions..
Concurrent transaction with 2-phase locking protocol completed!

Process finished with exit code 0
```

Code explanation:

I have coded in such a manner that the 2-phase locking protocol is implemented. First, I created 2 threads to run parallelly. The locks are implemented with a static integer value – if the value of lock is 0 this means there are no locks on any transactions. Similarly, if the value of lock is 1 or 2 then it means the lock is acquired by transaction 1 or transaction 2 respectively. At first, I locked the second transaction since the update query was fired in this transaction before the first transaction. Now the code will begin starting both the threads – the select query from transaction 1 will run successfully no matter where the lock is but the next update query will not be executed of transaction 1 since the lock is acquired by transaction 2. Now the transaction 2 queries will be executed and once both the queries

of transaction 2 is executed and committed, the lock is released from transaction 2 and the system notifies the first transaction that the lock is released and now the lock is acquired by transaction 1 which makes it ready to execute its update query.