

# データサイエンス概論 第一回レポート

201821636 村松 直哉

平成 30 年 11 月 19 日

このレポートには，一部の人が不快になる表現が多分に含まれております．

## 1 重回帰分析

本レポートでは，Sexualitics<sup>\*1</sup>の xHamster<sup>\*2</sup>に関するデータセット<sup>\*3</sup>を利用する．

データセットの構造を表 1 に示す．ただし表が大きくなりすぎるため，一部を省略している．

---

<sup>\*1</sup> <http://sexualitics.github.io/>

<sup>\*2</sup> <https://xhamster.com/>

<sup>\*3</sup> <http://pornstudies.sexualitics.org/data/xhamster.csv.tar.gz>

表 1 xHamster データセット

id	upload_date	title	channels	description	nb_views	nb_votes	nb_comments	runtime	uploader
378466	2010-06-29	girl riding black cock	['BBW', (省略)]	(省略)	17262	65	11	120	(省略)
478576	2010-11-07	masturbation	['Masturbation']	(省略)	953	3	NA	15	(省略)
287146	2010-02-12	sexy horny booty dance	['Babes', (省略)]	(省略)	6060	11	3	163	(省略)
1583074	2012-11-18	its rather big	['Amateur', (省略)]	NA	64413	55	4	146	(省略)
1246688	2012-06-01	dildo	['Men']	NA	247	1	NA	51	(省略)
1450532	2012-09-14	tribute to me	['Men']	(省略)	428	18	128	146	(省略)
1450531	2012-09-14	(省略)	['Men']	NA	199	1	1	57	(省略)
1450537	2012-09-14	(省略)	['Double Penetration', (省略)]	NA	33481	194	14	1137	(省略)
1450536	2012-09-14	x036	['BDSM', (省略)]	Part 3/3	104039	72	7	619	(省略)
1450535	2012-09-14	boy and mature woman part1	['Matures', (省略)]	NA	400167	171	3	240	(省略)
1450534	2012-09-12	(省略)	['Amateur', (省略)]	(省略)	39276	37	4	780	(省略)
968454	2011-12-22	(省略)	['Brunettes', (省略)]	NA	20864	40	3	404	(省略)
1246681	2012-06-01	amante tetona 01	['Amateur', (省略)]	Amateur	8921	16	1	127	(省略)
1246682	2012-06-01	british slut victoria	['Amateur', (省略)]	NA	180901	125	14	1505	(省略)
968457	2011-12-22	my cock	['Men']	my cock	868	5	3	173	(省略)
968450	2011-12-22	masturbation in black dress	['Men']	(省略)	845	4	4	362	(省略)
968451	2011-12-22	cuckold cleans it up 2	['Amateur', (省略)]	(省略)	353342	497	48	747	(省略)
778288	2011-07-29	(省略)	['Amateur', (省略)]	(省略)	310578	269	24	1059	(省略)
1246687	2012-05-30	school time orgy	['Big Boobs', (省略)]	(省略)	50202	74	8	185	(省略)
978809	2011-12-31	sexy gir	['Hardcore']	NA	1234	16	5	284	(省略)
978802	2012-01-03	denise young bbw threesomes	['BBW', (省略)]	bbw porno	63186	150	16	1090	(省略)
978806	2012-01-03	blonde with pigtails fucking 2	['Amateur', (省略)]	NA	3775	8	1	199	(省略)
978807	2011-12-31	up bra	['Amateur', (省略)]	NA	337309	182	8	96	(省略)
1698377	2013-01-21	strokin i	['Men']	BBC	87	1	1	23	(省略)
1698375	2013-01-21	amateur cellphone sex	['Amateur', 'Hardcore']	NA	6954	18	2	174	(省略)
			中略						
89377	2008-12-06	fisting	['Amateur']	NA	122575	206	3	289	NA

今回の分析では，説明変数として「nb\_views（その動画の視聴回数）」，「nb\_comments（動画に対して投稿されたコメント数）」，「runtime（動画の時間 (秒))」を利用した．目的変数として「nb\_votes（動画の平均評価値）」を利用した．この値は，動画に対して 1 ユーザーにつき，良かったか悪かったかを 1 票投票できる．nb\_votes は，良かった時を 1，悪かった時を  $-1$  として，全投票の合計値である．