Predicting Mobile Game Success

MileStone1

TEAM ID 19

Step (1) Feature Selection:

Exclude all unique strings.

Almost Empty

Unique Strings   Label   Assume all empty to be 0   Unique String

URL	ID	Name	Subtitle	Icon URL	Average User Rating	User Rating Count	Price	In-app Purchases	Description	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
https	##	Sudoku		https	4	3553	2.99		Join over 21,	Mighty Migh	4+	DA, NL, EN,	15853568	Games	Games, Strategy, F	11-07-08	30-05-17
https	##	Reversi		https	3.5	284	1.99		The classic g	Kiss The Mac	4+	EN	12328960	Games	Games, Strategy, E	11-07-08	17-05-18
https	##	Morocco		https	3	8376	0		Play the clas	Bayou Game	4+	EN	674816	Games	Games, Board, Str	11-07-08	05-09-17
https	##	Sudoku (Free)		https	3.5	190394	0		Top 100 free	Mighty Migh	4+	DA, NL, EN,	21552128	Games	Games, Strategy, F	23-07-08	30-05-17
https	##	Senet Deluxe		https	3.5	28	2.99		"Senet Delux	RoGame Sof	4+	DA, NL, EN,	34689024	Games	Games, Strategy, E	18-07-08	22-07-18
https	##	Sudoku	Original	https	3	47	0	1.99	Sudoku will t	OutOfTheBit	4+	EN	48672768	Games	Games, Entertainr	30-07-08	29-04-19
https	##	Gravitation		https	2.5	35	0		"Gravitation	Robert Farnu	4+		6328320	Games	Games, Entertainr	30-07-08	14-11-13
https	##	Colony		https	2.5	125	0.99		"50 levels of	Chris Haynes	4+	EN	64333824	Games	Games, Strategy, E	03-08-08	03-10-18
https	##	Carte		https	2.5	44	0		"Jeu simple c	Jean-Franco	4+	FR	2657280	Games	Games, Strategy, E	03-08-08	23-11-17
https	##	"Barrels O' Fun		https	2.5	184	0		Barrels O\u2	BesqWare	4+	EN	1466515	Games	Games, Casual, Str	01-08-08	01-08-08
https	##	Quaddraxx		https://is5-ssl.mzstatic.con			0		Quaddraxx-L	H2F Informa	4+	EN	3089867	Games	Games, Entertainr	11-08-08	30-09-08
https	##	Lumen Lite		https	3.5	5072	0		"The objectiv	Bridger Max	4+	EN	7086403	Games	Games, Puzzle, Str	18-08-08	22-11-08
https	##	BubblePop		https	3	526	0		Are you read	TMSOFT	4+	EN	845008	Games	Games, Strategy, E	22-08-08	25-07-09
https	##	Marple		https	3.5	989	0.99		AWARDED "I	Mikko Kanka	4+	EN	3643392	Games	Games, Puzzle, Str	28-08-08	05-05-19
https	##	Tetravex Lite		https	2.5	2358	0		Play the clas	Futrell Softw	4+	EN	731525	Games	Games, Puzzle, Str	27-08-08	21-10-08
https	##	Awele/Oware -		https	3	112	0	0.99	Awele/Owar	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Strategy, E	31-08-08	06-04-15
https	##	Awele/Oware -		https	3	112	0	0.99	Awele/Owar	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Strategy, E	31-08-08	06-04-15
https	##	Chess Game		https	3	504	0		"How about	Memention	4+	EN	444163	Games	Games, Board, Str	02-09-08	07-10-09
https	##	Catcha Mouse		https	3.5	95971	0	1.99, 0.99,	*More than	Meme, Inc	4+	EN	26820692	Games	Games, Puzzle, Str	19-02-09	26-04-13
https	##	Cool S Sudoku,		https	3.5	357	0	0.99, 0.99,	You can now	Splash Softw	4+	EN	52236288	Games	Games, Strategy, F	24-09-08	26-04-19
https	##	Mind the Corne		https	3	7	0.99		"Mind the Cc	Zentropy Sof	4+	EN	1530880	Games	Games, Strategy, E	19-09-08	26-10-16
https	##	All You Can Eat		https	2	30	0.99		"'All You Can	Armin Heinri	9+	EN	4341760	Entertain	Entertainment, Ga	24-09-08	16-03-19
https	##	Ane Rouge		https://is2-ssl.mzstatic.con			0		Your Goal is	Jean-Franco	4+	EN, FR	1190912	Games	Games, Puzzle, Ed	27-09-08	08-03-17

Output:

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	Mighty Migh	4+	DA, NL, EN, FI,	15853568	Games	Games, Strate	11-07-08	30-05-17
3.5	284	1.99	0	Kiss The Mac	4+	EN	12328960	Games	Games, Strate	11-07-08	17-05-18
3	8376	0	0	Bayou Game	4+	EN	674816	Games	Games, Board	11-07-08	05-09-17
3.5	190394	0	0	Mighty Migh	4+	DA, NL, EN, FI,	21552128	Games	Games, Strate	23-07-08	30-05-17
3.5	28	2.99	0	RoGame Sof	4+	DA, NL, EN, FR,	34689024	Games	Games, Strate	18-07-08	22-07-18
3	47	0	1.99	OutOfTheBit	4+	EN	48672768	Games	Games, Enter	30-07-08	29-04-19
2.5	35	0	0	Robert Farnu	4+		6328320	Games	Games, Enter	30-07-08	14-11-13
2.5	125	0.99	0	Chris Haynes	4+	EN	64333824	Games	Games, Strate	03-08-08	03-10-18
2.5	44	0	0	Jean-Franco	4+	FR	2657280	Games	Games, Strate	03-08-08	23-11-17
2.5	184	0	0	BesqWare	4+	EN	1466515	Games	Games, Casua	01-08-08	01-08-08
		0	0	H2F Informa	4+	EN	3089867	Games	Games, Enter	11-08-08	30-09-08
3.5	5072	0	0	Bridger Max	4+	EN	7086403	Games	Games, Puzzle	18-08-08	22-11-08
3	526	0	0	TMSOFT	4+	EN	845008	Games	Games, Strate	22-08-08	25-07-09
3.5	989	0.99	0	Mikko Kanka	4+	EN	3643392	Games	Games, Puzzle	28-08-08	05-05-19
2.5	2358	0	0	Futrell Softw	4+	EN	731525	Games	Games, Puzzle	27-08-08	21-10-08
3	112	0	0.99	SOLILAB	4+	EN, FR, DE, IT, I	1.23E+08	Games	Games, Strate	31-08-08	06-04-15
3	112	0	0.99	SOLILAB	4+	EN, FR, DE, IT, I	1.23E+08	Games	Games, Strate	31-08-08	06-04-15
3	504	0	0	Memention	4+	EN	444163	Games	Games, Board	02-09-08	07-10-09
3.5	95971	0	1.99, 0.99, 1	Meme, Inc	4+	EN	26820692	Games	Games, Puzzle	19-02-09	26-04-13
3.5	357	0	0.99, 0.99, 0	Splash Softw	4+	EN	52236288	Games	Games, Strate	24-09-08	26-04-19
3	7	0.99	0	Zentropy Sof	4+	EN	1530880	Games	Games, Strate	19-09-08	26-10-16
2	30	0.99	0	Armin Heinri	9+	EN	4341760	Entertainn	Entertainmen	24-09-08	16-03-19
		0	0	Jean-Franco	4+	EN, FR	1190912	Games	Games, Puzzle	27-09-08	08-03-17

Step (2) Data Cleaning

1. Remove all rows with missing data in Label column.

Label   Empty Fields

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	Mighty Migh	4+	DA, NL, EN, FI,	15853568	Games	Games, Strate	11-07-08	30-05-17
3.5	284	1.99	0	Kiss The Mac	4+	EN	12328960	Games	Games, Strate	11-07-08	17-05-18
3	8376	0	0	Bayou Game	4+	EN	674816	Games	Games, Board	11-07-08	05-09-17
3.5	190394	0	0	Mighty Migh	4+	DA, NL, EN, FI,	21552128	Games	Games, Strate	23-07-08	30-05-17
3.5	28	2.99	0	RoGame Sof	4+	DA, NL, EN, FR,	34689024	Games	Games, Strate	18-07-08	22-07-18
3	47	0	1.99	OutOfTheBit	4+	EN	48672768	Games	Games, Enter	30-07-08	29-04-19
2.5	35	0	0	Robert Farni	4+		6328320	Games	Games, Enter	30-07-08	14-11-13
2.5	125	0.99	0	Chris Haynes	4+	EN	64333824	Games	Games, Strate	03-08-08	03-10-18
2.5	44	0	0	Jean-Franco	4+	FR	2657280	Games	Games, Strate	03-08-08	23-11-17
2.5	184	0	0	BesqWare	4+	EN	1466515	Games	Games, Casua	01-08-08	01-08-08
		0	0	H2F Informa	4+	EN	3089867	Games	Games, Enter	11-08-08	30-09-08
3.5	5072	0	0	Bridger Max	4+	EN	7086403	Games	Games, Puzzle	18-08-08	22-11-08
3	526	0	0	TMSOFT	4+	EN	845008	Games	Games, Strate	22-08-08	25-07-09
3.5	989	0.99	0	Mikko Kanka	4+	EN	3643392	Games	Games, Puzzle	28-08-08	05-05-19
2.5	2358	0	0	Futrell Softw	4+	EN	731525	Games	Games, Puzzle	27-08-08	21-10-08
3	112	0	0.99	SOLILAB	4+	EN, FR, DE, IT, I	1.23E+08	Games	Games, Strate	31-08-08	06-04-15
3	112	0	0.99	SOLILAB	4+	EN, FR, DE, IT, I	1.23E+08	Games	Games, Strate	31-08-08	06-04-15
3	504	0	0	Memention	4+	EN	444163	Games	Games, Board	02-09-08	07-10-09
3.5	95971	0	1.99, 0.99, 1	Meme, Inc	4+	EN	26820692	Games	Games, Puzzle	19-02-09	26-04-13
3.5	357	0	0.99, 0.99, 0	Splash Softw	4+	EN	52236288	Games	Games, Strate	24-09-08	26-04-19
3	7	0.99	0	Zentropy Sof	4+	EN	1530880	Games	Games, Strate	19-09-08	26-10-16
2	30	0.99	0	Armin Heinri	9+	EN	4341760	Entertainn	Entertainmen	24-09-08	16-03-19
		0	0	Jean-Franco	4+	EN, FR	1190912	Games	Games, Puzzle	27-09-08	08-03-17

Output:

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	Mighty Mig	4+	DA, NL, EN,	15853568	Games	Games, Str	11-07-08	30-05-17
3.5	284	1.99	0	Kiss The Ma	4+	EN	12328960	Games	Games, Str	11-07-08	17-05-18
3	8376	0	0	Bayou Gam	4+	EN	674816	Games	Games, Bo	11-07-08	05-09-17
3.5	190394	0	0	Mighty Mig	4+	DA, NL, EN,	21552128	Games	Games, Str	23-07-08	30-05-17
3.5	28	2.99	0	RoGame Sc	4+	DA, NL, EN,	34689024	Games	Games, Str	18-07-08	22-07-18
3	47	0	1.99	OutOfTheB	4+	EN	48672768	Games	Games, En	30-07-08	29-04-19
2.5	35	0	0	Robert Farr	4+		6328320	Games	Games, En	30-07-08	14-11-13
2.5	125	0.99	0	Chris Hayne	4+	EN	64333824	Games	Games, Str	03-08-08	03-10-18
2.5	44	0	0	Jean-Franc	4+	FR	2657280	Games	Games, Str	03-08-08	23-11-17
2.5	184	0	0	BesqWare	4+	EN	1466515	Games	Games, Ca	01-08-08	01-08-08
3.5	5072	0	0	Bridger Ma	4+	EN	7086403	Games	Games, Pu	18-08-08	22-11-08
3	526	0	0	TMSOFT	4+	EN	845008	Games	Games, Str	22-08-08	25-07-09
3.5	989	0.99	0	Mikko Kank	4+	EN	3643392	Games	Games, Pu	28-08-08	05-05-19
2.5	2358	0	0	Futrell Soft	4+	EN	731525	Games	Games, Pu	27-08-08	21-10-08
3	112	0	0.99	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Str	31-08-08	06-04-15
3	112	0	0.99	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Str	31-08-08	06-04-15
3	504	0	0	Memention	4+	EN	444163	Games	Games, Bo	02-09-08	07-10-09
3.5	95971	0	1.99, 0.99,	Meme, Inc	4+	EN	26820692	Games	Games, Pu	19-02-09	26-04-13
3.5	357	0	0.99, 0.99,	Splash Soft	4+	EN	52236288	Games	Games, Str	24-09-08	26-04-19
3	7	0.99	0	Zentropy Sc	4+	EN	1530880	Games	Games, Str	19-09-08	26-10-16
2	30	0.99	0	Armin Hein	9+	EN	4341760	Entertainn	Entertainn	24-09-08	16-03-19
4	637	0.99	0	TwoDesk Sc	4+	EN	22873088	Games	Games, Pu	07-10-08	20-05-17
3.5	1563	0	0	Quicksand I	4+	EN	6550985	Games	Games, Str	02-10-08	02-10-08

Step (3) Data Cleaning

2. Encoding: Give weights to strings.
   - Give weights to categorial strings based on label column.
   - Value of 'Year' for Date Strings.

Categorial Strings   Date Strings

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	Mighty Mig	4+	DA, NL, EN,	15853568	Games	Games, Str	11-07-08	30-05-17
3.5	284	1.99	0	Kiss The Ma	4+	EN	12328960	Games	Games, Str	11-07-08	17-05-18
3	8376	0	0	Bayou Gam	4+	EN	674816	Games	Games, Bo	11-07-08	05-09-17
3.5	190394	0	0	Mighty Mig	4+	DA, NL, EN,	21552128	Games	Games, Str	23-07-08	30-05-17
3.5	28	2.99	0	RoGame Sc	4+	DA, NL, EN,	34689024	Games	Games, Str	18-07-08	22-07-18
3	47	0	1.99	OutOfTheB	4+	EN	48672768	Games	Games, En	30-07-08	29-04-19
2.5	35	0	0	Robert Farr	4+		6328320	Games	Games, En	30-07-08	14-11-13
2.5	125	0.99	0	Chris Hayne	4+	EN	64333824	Games	Games, Str	03-08-08	03-10-18
2.5	44	0	0	Jean-Franc	4+	FR	2657280	Games	Games, Str	03-08-08	23-11-17
2.5	184	0	0	BesqWare	4+	EN	1466515	Games	Games, Ca	01-08-08	01-08-08
3.5	5072	0	0	Bridger Ma	4+	EN	7086403	Games	Games, Pu	18-08-08	22-11-08
3	526	0	0	TMSOFT	4+	EN	845008	Games	Games, Str	22-08-08	25-07-09
3.5	989	0.99	0	Mikko Kank	4+	EN	3643392	Games	Games, Pu	28-08-08	05-05-19
2.5	2358	0	0	Futrell Soft	4+	EN	731525	Games	Games, Pu	27-08-08	21-10-08
3	112	0	0.99	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Str	31-08-08	06-04-15
3	112	0	0.99	SOLILAB	4+	EN, FR, DE,	1.23E+08	Games	Games, Str	31-08-08	06-04-15
3	504	0	0	Mementior	4+	EN	444163	Games	Games, Bo	02-09-08	07-10-09
3.5	95971	0	1.99, 0.99,	Meme, Inc	4+	EN	26820692	Games	Games, Pu	19-02-09	26-04-13
3.5	357	0	0.99, 0.99,	Splash Soft	4+	EN	52236288	Games	Games, Str	24-09-08	26-04-19
3	7	0.99	0	Zentropy Sc	4+	EN	1530880	Games	Games, Str	19-09-08	26-10-16
2	30	0.99	0	Armin Hein	9+	EN	4341760	Entertainn	Entertainn	24-09-08	16-03-19
4	637	0.99	0	TwoDesk Sc	4+	EN	22873088	Games	Games, Pu	07-10-08	20-05-17
3.5	1563	0	0	Quicksand I	4+	EN	6550985	Games	Games. Str	02-10-08	02-10-08

Output:

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	4	4.01125		69.0165	15853568	4.036432	12.173	8	17
3.5	284	1.99	0	3.5	4.01125	4.028574	12328960	4.036432	11.91386	8	18
3	8376	0	0	3.75	4.01125	4.028574	674816	4.036432	11.91386	8	17
3.5	190394	0	0	4	4.01125		69.0165	21552128	4.036432	12.173	8	17
3.5	28	2.99	0	3.5	4.01125	61.11307	34689024	4.036432	15.92107	8	18
3	47	0	1.99	3.75	4.01125	4.028574	48672768	4.036432	16.19213	8	19
2.5	35	0	0	2.5	4.01125		6328320	4.036432	16.19213	8	13
2.5	125	0.99	0	2.5	4.01125	4.028574	64333824	4.036432	11.91386	8	18
2.5	44	0	0	2.5	4.01125	3.7	2657280	4.036432	15.93299	8	17
2.5	184	0	0	2.5	4.01125	4.028574	1466515	4.036432	12.22784	8	8
3.5	5072	0	0	3.5	4.01125	4.028574	7086403	4.036432	12.173	8	8
3	526	0	0	4	4.01125	4.028574	845008	4.036432	12.08678	8	9
3.5	989	0.99	0	3.5	4.01125	4.028574	3643392	4.036432	12.173	8	19
2.5	2358	0	0	3	4.01125	4.028574	731525	4.036432	16.19213	8	8
3	112	0	0.99	3	4.01125	20.32397	1.23E+08	4.036432	11.91386	8	15
3	112	0	0.99	3	4.01125	20.32397	1.23E+08	4.036432	11.91386	8	15
3	504	0	0	3	4.01125	4.028574	444163	4.036432	11.91386	8	9
3.5	95971	0	2.132857	7.351064	4.01125	4.028574	26820692	4.036432	12.173	9	13
3.5	357	0	0.99	3.5	4.01125	4.028574	52236288	4.036432	16.19213	8	19
3	7	0.99	0	3	4.01125	4.028574	1530880	4.036432	15.93299	8	16
2	30	0.99	0	2	4.062102	4.028574	4341760	3.821429	15.91121	8	19
4	637	0.99	0	4	4.01125	4.028574	22873088	4.036432	16.19213	8	17
3.5	1563	0	0	3.5	4.01125	4.028574	6550985	4.036432	12.173	8	8

Step (4) Data Cleaning:

3. Filling empty cells in each feature by the average values in this feature.   Empty Cell

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	4	4.01125	69.0165	15853568	4.036432	12.173	8	17
3.5	284	1.99	0	3.5	4.01125	4.028574	12328960	4.036432	11.91386	8	18
3	8376	0	0	3.75	4.01125	4.028574	674816	4.036432	11.91386	8	17
3.5	190394	0	0	4	4.01125	69.0165	21552128	4.036432	12.173	8	17
3.5	28	2.99	0	3.5	4.01125	61.11307	34689024	4.036432	15.92107	8	18
3	47	0	1.99	3.75	4.01125	4.028574	48672768	4.036432	16.19213	8	19
2.5	35	0	0	2.5	4.01125		6328320	4.036432	16.19213	8	13
2.5	125	0.99	0	2.5	4.01125	4.028574	64333824	4.036432	11.91386	8	18
2.5	44	0	0	2.5	4.01125	3.7	2657280	4.036432	15.93299	8	17
2.5	184	0	0	2.5	4.01125	4.028574	1466515	4.036432	12.22784	8	8
3.5	5072	0	0	3.5	4.01125	4.028574	7086403	4.036432	12.173	8	8
3	526	0	0	4	4.01125	4.028574	845008	4.036432	12.08678	8	9
3.5	989	0.99	0	3.5	4.01125	4.028574	3643392	4.036432	12.173	8	19
2.5	2358	0	0	3	4.01125	4.028574	731525	4.036432	16.19213	8	8
3	112	0	0.99	3	4.01125	20.32397	1.23E+08	4.036432	11.91386	8	15
3	112	0	0.99	3	4.01125	20.32397	1.23E+08	4.036432	11.91386	8	15
3	504	0	0	3	4.01125	4.028574	444163	4.036432	11.91386	8	9
3.5	95971	0	2.132857	7.351064	4.01125	4.028574	26820692	4.036432	12.173	9	13
3.5	357	0	0.99	3.5	4.01125	4.028574	52236288	4.036432	16.19213	8	19
3	7	0.99	0	3	4.01125	4.028574	1530880	4.036432	15.93299	8	16
2	30	0.99	0	2	4.062102	4.028574	4341760	3.821429	15.91121	8	19
4	637	0.99	0	4	4.01125	4.028574	22873088	4.036432	16.19213	8	17
3.5	1563	0	0	3.5	4.01125	4.028574	6550985	4.036432	12.173	8	8

Output:

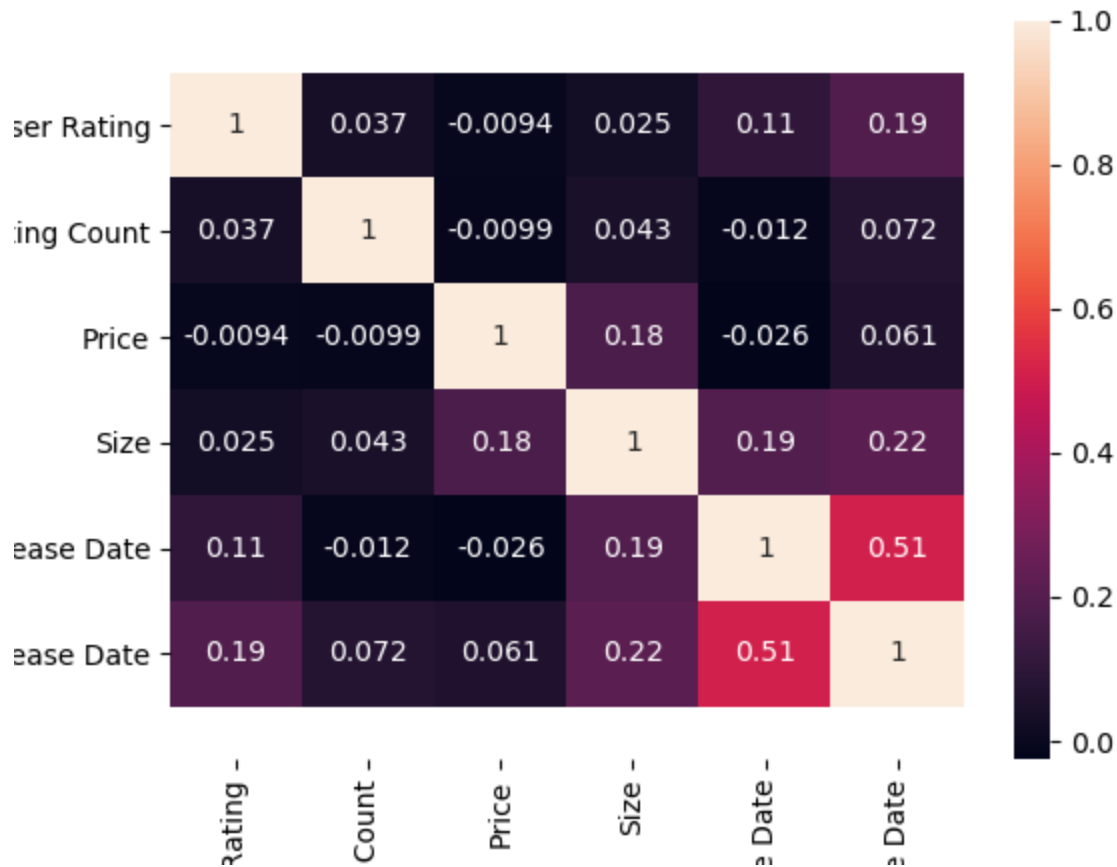| Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	4	4.01125	69.016502	15853568	4.036432	12.173	8	17
3.5	284	1.99	0	3.5	4.01125	4.0285739	12328960	4.036432	11.91386	8	18
3	8376	0	0	3.75	4.01125	4.0285739	674816	4.036432	11.91386	8	17
3.5	190394	0	0	4	4.01125	69.016502	21552128	4.036432	12.173	8	17
3.5	28	2.99	0	3.5	4.01125	61.11307	34689024	4.036432	15.92107	8	18
3	47	0	1.99	3.75	4.01125	4.0285739	48672768	4.036432	16.19213	8	19
2.5	35	0	0	2.5	4.01125	15.956244	6328320	4.036432	16.19213	8	13
2.5	125	0.99	0	2.5	4.01125	4.0285739	64333824	4.036432	11.91386	8	18
2.5	44	0	0	2.5	4.01125	3.7	2657280	4.036432	15.93299	8	17
2.5	184	0	0	2.5	4.01125	4.0285739	1466515	4.036432	12.22784	8	8
3.5	5072	0	0	3.5	4.01125	4.0285739	7086403	4.036432	12.173	8	8
3	526	0	0	4	4.01125	4.0285739	845008	4.036432	12.08678	8	9
3.5	989	0.99	0	3.5	4.01125	4.0285739	3643392	4.036432	12.173	8	19
2.5	2358	0	0	3	4.01125	4.0285739	731525	4.036432	16.19213	8	8
3	112	0	0.99	3	4.01125	20.323967	1.23E+08	4.036432	11.91386	8	15
3	112	0	0.99	3	4.01125	20.323967	1.23E+08	4.036432	11.91386	8	15
3	504	0	0	3	4.01125	4.0285739	444163	4.036432	11.91386	8	9
3.5	95971	0	2.132857	7.3510638	4.01125	4.0285739	26820692	4.036432	12.173	9	13
3.5	357	0	0.99	3.5	4.01125	4.0285739	52236288	4.036432	16.19213	8	19
3	7	0.99	0	3	4.01125	4.0285739	1530880	4.036432	15.93299	8	16
2	30	0.99	0	2	4.062102	4.0285739	4341760	3.821429	15.91121	8	19
4	637	0.99	0	4	4.01125	4.0285739	22873088	4.036432	16.19213	8	17
3.5	1563	0	0	3.5	4.01125	4.0285739	6550985	4.036432	12.173	8	8

Step (5) Data Scaling:

Scale all data to be ranged between (0, 1). Using MinMaxScalar

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	3553	2.99	0	4	4.01125	69.016502	15853568	4.036432	12.173	8	17
3.5	284	1.99	0	3.5	4.01125	4.0285739	12328960	4.036432	11.91386	8	18
3	8376	0	0	3.75	4.01125	4.0285739	674816	4.036432	11.91386	8	17
3.5	190394	0	0	4	4.01125	69.016502	21552128	4.036432	12.173	8	17
3.5	28	2.99	0	3.5	4.01125	61.11307	34689024	4.036432	15.92107	8	18
3	47	0	1.99	3.75	4.01125	4.0285739	48672768	4.036432	16.19213	8	19
2.5	35	0	0	2.5	4.01125	15.956244	6328320	4.036432	16.19213	8	13
2.5	125	0.99	0	2.5	4.01125	4.0285739	64333824	4.036432	11.91386	8	18
2.5	44	0	0	2.5	4.01125	3.7	2657280	4.036432	15.93299	8	17
2.5	184	0	0	2.5	4.01125	4.0285739	1466515	4.036432	12.22784	8	8
3.5	5072	0	0	3.5	4.01125	4.0285739	7086403	4.036432	12.173	8	8
3	526	0	0	4	4.01125	4.0285739	845008	4.036432	12.08678	8	9
3.5	989	0.99	0	3.5	4.01125	4.0285739	3643392	4.036432	12.173	8	19
2.5	2358	0	0	3	4.01125	4.0285739	731525	4.036432	16.19213	8	8
3	112	0	0.99	3	4.01125	20.323967	1.23E+08	4.036432	11.91386	8	15
3	112	0	0.99	3	4.01125	20.323967	1.23E+08	4.036432	11.91386	8	15
3	504	0	0	3	4.01125	4.0285739	444163	4.036432	11.91386	8	9
3.5	95971	0	2.132857	7.3510638	4.01125	4.0285739	26820692	4.036432	12.173	9	13
3.5	357	0	0.99	3.5	4.01125	4.0285739	52236288	4.036432	16.19213	8	19
3	7	0.99	0	3	4.01125	4.0285739	1530880	4.036432	15.93299	8	16
2	30	0.99	0	2	4.062102	4.0285739	4341760	3.821429	15.91121	8	19
4	637	0.99	0	4	4.01125	4.0285739	22873088	4.036432	16.19213	8	17
3.5	1563	0	0	3.5	4.01125	4.0285739	6550985	4.036432	12.173	8	8

Output:

Average User Rating	User Rating Count	Price	In-app Purchases	Developer	Age Rating	Languages	Size	Primary Genre	Genres	Original Release Date	Current Version Release Date
4	0.00117	0.021359	0	0.1764706	0.397489	0.1407515	0.003904	0.587042	0.263955	0	0.818182
3.5	9.20E-05	0.014215	0	0.1470588	0.397489	0.0011356	0.003024	0.587042	0.247293	0	0.909091
3	0.00276	0	0	0.1617647	0.397489	0.0011356	0.000115	0.587042	0.247293	0	0.818182
3.5	0.062778	0	0	0.1764706	0.397489	0.1407515	0.005327	0.587042	0.263955	0	0.818182
3.5	7.58E-06	0.021359	0	0.1470588	0.397489	0.1237723	0.008607	0.587042	0.504939	0	0.909091
3	1.38E-05	0	0.024982	0.1617647	0.397489	0.0011356	0.012098	0.587042	0.522367	0	1
2.5	9.89E-06	0	0	0.0882353	0.397489	0.0267602	0.001526	0.587042	0.522367	0	0.454545
2.5	3.96E-05	0.007072	0	0.0882353	0.397489	0.0011356	0.016008	0.587042	0.247293	0	0.909091
2.5	1.29E-05	0	0	0.0882353	0.397489	0.0004297	0.00061	0.587042	0.505705	0	0.818182
2.5	5.90E-05	0	0	0.0882353	0.397489	0.0011356	0.000312	0.587042	0.267481	0	0
3.5	0.001671	0	0	0.1470588	0.397489	0.0011356	0.001715	0.587042	0.263955	0	0
3	0.000172	0	0	0.1764706	0.397489	0.0011356	0.000157	0.587042	0.258412	0	0.090909
3.5	0.000324	0.007072	0	0.1470588	0.397489	0.0011356	0.000856	0.587042	0.263955	0	1
2.5	0.000776	0	0	0.1176471	0.397489	0.0011356	0.000129	0.587042	0.522367	0	0
3	3.53E-05	0	0.012428	0.1176471	0.397489	0.0361435	0.030612	0.587042	0.247293	0	0.636364
3	3.53E-05	0	0.012428	0.1176471	0.397489	0.0361435	0.030612	0.587042	0.247293	0	0.636364
3	0.000165	0	0	0.1176471	0.397489	0.0011356	5.70E-05	0.587042	0.247293	0	0.090909
3.5	0.031643	0	0.026776	0.373592	0.397489	0.0011356	0.006642	0.587042	0.263955	0.090909	0.454545
3.5	0.000116	0	0.012428	0.1470588	0.397489	0.0011356	0.012988	0.587042	0.522367	0	1
3	6.59E-07	0.007072	0	0.1176471	0.397489	0.0011356	0.000328	0.587042	0.505705	0	0.727273
2	8.24E-06	0.007072	0	0.0588235	0.842333	0.0011356	0.00103	0.494898	0.504305	0	1
4	0.000208	0.007072	0	0.1764706	0.397489	0.0011356	0.005657	0.587042	0.522367	0	0.818182
3.5	0.000514	0	0	0.1470588	0.397489	0.0011356	0.001582	0.587042	0.263955	0	0

Analysis:

Regression Techniques:

1. Multivariable Linear Regression.
   According to our dataset we have more than one factor that influences the 'Average User Rate' output. The model describes how a single response variable 'Average User Rate' depends linearly on several predictor variables

2. Polynomial Linear Regression.
   Polynomial regression is about improving the model's closeness to the data by increasing the order.

Testing:

- K-fold Cross Validation technique.
- K is equal to 10
- Testing Size is equal to 1/10 of the dataset size.
- Training Size is equal to 9/10 of the dataset size.

Model 1: Multivariable Linear Regression

```
Running for fold # 1 ...
Train Error for fold # 1 : 0.4255015999624942
Test Error for fold # 1 : 0.8881235766495046

Running for fold # 2 ...
Train Error for fold # 2 : 0.427304712145356
Test Error for fold # 2 : 0.8531520202172945

Running for fold # 3 ...
Train Error for fold # 3 : 0.4258056364686977
Test Error for fold # 3 : 0.9051681282292576

Running for fold # 4 ...
Train Error for fold # 4 : 0.4251707935418063
Test Error for fold # 4 : 0.8957181251130909

Running for fold # 5 ...
Train Error for fold # 5 : 0.42901410533426143
Test Error for fold # 5 : 0.8864650576699296

Running for fold # 6 ...
Train Error for fold # 6 : 0.424765827225875
Test Error for fold # 6 : 0.9196371471883962


Running for fold # 7 ...
Train Error for fold # 7 : 0.4289023708846424
Test Error for fold # 7 : 0.8755365681829681

Running for fold # 8 ...
Train Error for fold # 8 : 0.4276969700610949
Test Error for fold # 8 : 0.9487503415300051

Running for fold # 9 ...
Train Error for fold # 9 : 0.4300235298803108
Test Error for fold # 9 : 0.8476379122305379

Running for fold # 10 ...
Train Error for fold # 10 : 0.42137388948280335
Test Error for fold # 10 : 0.9897970023286472

_____Finish_____
Average Train Error:  0.42655594349873416
Average Test  Error:  0.900998587933963

Process finished with exit code 0
```

Model 2: Polynomial Linear Regression

```
Running for fold # 1 ...
Train Error for fold # 1 : 0.25844558776799426
Test Error for fold # 1 : 8.165576209246842

Running for fold # 2 ...
Train Error for fold # 2 : 0.2653713697100027
Test Error for fold # 2 : 7.268575970427257

Running for fold # 3 ...
Train Error for fold # 3 : 0.2597346035181071
Test Error for fold # 3 : 8.0694510797589

Running for fold # 4 ...
Train Error for fold # 4 : 0.23300436144127656
Test Error for fold # 4 : 10.757178633044157

Running for fold # 5 ...
Train Error for fold # 5 : 0.265558117162353
Test Error for fold # 5 : 7.669147537657147

Running for fold # 6 ...
Train Error for fold # 6 : 0.2682649393751873
Test Error for fold # 6 : 7.520705349135002


Running for fold # 7 ...
Train Error for fold # 7 : 0.2625408789426293
Test Error for fold # 7 : 7.390679938907383

Running for fold # 8 ...
Train Error for fold # 8 : 0.2608781611491316
Test Error for fold # 8 : 7.207854902008725

Running for fold # 9 ...
Train Error for fold # 9 : 0.26773617281009154
Test Error for fold # 9 : 6.7752799577896345

Running for fold # 10 ...
Train Error for fold # 10 : 0.2632214838025626
Test Error for fold # 10 : 7.642462365887023

_____Finish_____
Average Train Error:  0.2604755675679336
Average Test  Error:  7.846691194386207
```

Conclusion:

Our intuition was that Polynomial linear regression would better, but it was disproved by the error of the output.

Multivariable linear regression has proven to be better than Polynomial linear regression according to the features that we have selected.