

Project Name

Application user acceptance and ranking project comparing Google and Apple

Project members

1. Trevor Laskey
2. Humberto Rojas
3. Orathailux (Jerica) Saengaram
4. Satoko Yamaguchi

Data Cleanup & Analysis

* The sources of data that you will extract from.

[Kaggle - Apple store](#)

[Kaggle - Google Play Store](#)

* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

Cleaning, joining, aggregating

* The type of final production database to load the data into (relational or non-relational).

MySQL, Python

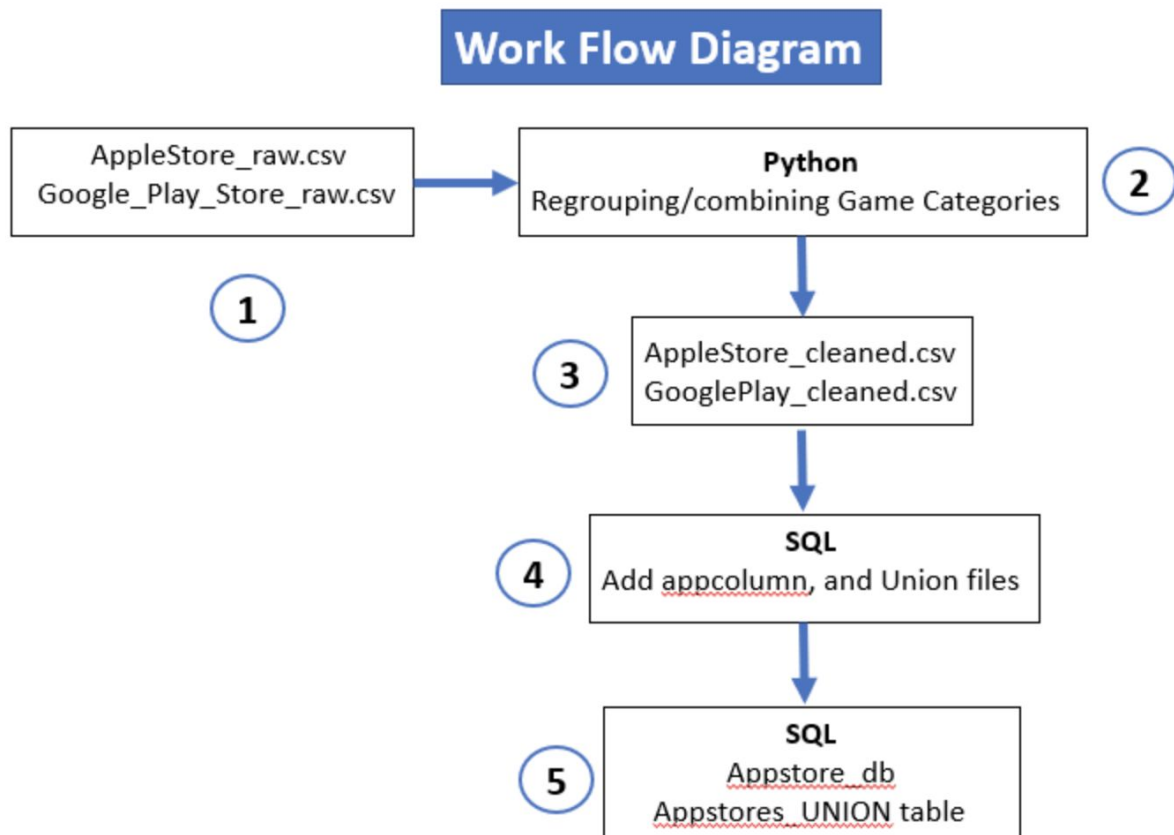
* The final tables or collections that will be used in the production database.

appstores_db(Tables: 1. applestore_cleaned 2. googleplay_cleaned 3. appstores_UNION)

We chose to store our data in SQL (relational database) because our data is related to each other and SQL is best fit to use in combining and keeping the database, and can easily be share as .csv file

Project Report

Work Flow Diagram



Extract**: your original data sources and how the data was formatted (CSV, JSON, MySQL, etc).

1) Two csv files

[Kaggle - Apple store](#)

[Kaggle - Google Play Store](#)

Transform**: what data cleaning or transformation was required.

2) We look into the raw data to see which category we can regroup or combine to make it make more sense and be able to combine the data

3) In Python: we regrouped and combined game category in both Apple and Google App Stores

4) In SQL: we deleted the columns that we don't use and added new column (app_source) to distinguish the data is from google play store or apple store

Load**: the final database, tables/collections, and why this was chosen as well as your final data schema.

5) appstores_UNION.sql

Our Work based on Work Flow Diagram

Work Flow Diagram #1: Download 2 csv data from Kaggle

AppleStore_raw.csv

Column1	id	track_name	size_bytes	currency	price	rating_count	rating_count	user_rating	user_rating	ver	cont_rating	prime_genre	sup_devic	ipadSc	ur	lang	num	vpp_lic
1	281656475	PAC-MAN Premium	100788224	USD	3.99	21292	26	4	4.5	6.3.5	4+	Games	38	5	10		1	
2	281796108	Evernote - stay orga	158578688	USD	0	161065	26	4	3.5	8.2.2	4+	Productivity	37	5	23		1	
3	281940292	WeatherBug - Local	100524032	USD	0	188583	2822	3.5	4.5	5.0.0	4+	Weather	37	5	3		1	
4	282614216	eBay: Best App to Bu	128512000	USD	0	262241	649	4	4.5	5.10.0	12+	Shopping	37	5	9		1	
5	282935706	Bible	92774400	USD	0	985920	5320	4.5	5	7.5.1	4+	Reference	37	5	45		1	
6	283619399	Shanghai Mahjong	10485713	USD	0.99	8253	5516	4	4	1.8	4+	Games	47	5	1		1	
7	283646709	PayPal - Send and re	227795968	USD	0	119487	879	4	4.5	6.12.0	4+	Finance	37	0	19		1	
8	284035177	Pandora - Music & R	130242560	USD	0	1126879	3594	4	4.5	8.4.1	12+	Music	37	4	1		1	
9	284666222	PiCalc - The Best Cal	49250304	USD	9.99	1117	4	4.5	5	3.6.6	4+	Utilities	37	5	1		1	
10	284736660	Ms. PAC-MAN	70023168	USD	3.99	7885	40	4	4	4.0.4	4+	Games	38	0	10		1	
11	284791396	Solitaire by Mobilit	49618944	USD	4.99	76720	4017	4.5	4.5	4.10.1	4+	Games	38	4	11		1	
12	284815117	SCRABBLE Premium	227547136	USD	7.99	105776	166	3.5	2.5	5.19.0	4+	Games	37	0	6		1	
13	284815942	Google " Search n	179979264	USD	0	479440	203	3.5	4	27	17+	Utilities	37	4	33		1	
14	284847138	Bank of America - M	160925696	USD	0	119773	2336	3.5	4.5	7.3.8	4+	Finance	37	0	2		1	
15	284862767	FreeCell	55153664	USD	4.99	6340	668	4.5	4.5	4.0.3	4+	Games	38	5	2		1	
16	284876795	TripAdvisor Hotels F	207907840	USD	0	56194	87	4	3.5	21.1	4+	Travel	37	1	26		1	
17	284882215	Facebook	389879808	USD	0	2974676	212	3.5	3.5	95	4+	Social Networki	37	1	29		1	
18	284910350	Yelp - Nearby Resta	167407616	USD	0	223885	3726	4	4.5	11.15.0	12+	Travel	37	5	18		1	
20	284993459	Shazam - Discover m	147093504	USD	0	402925	136	4	4.5	11.0.3	12+	Music	37	3	16		1	
21	285005463	Crash Bandicoot Nit	10735026	USD	2.99	31456	4178	4	3.5	1.0.0	4+	Games	47	0	1		1	
22	285946052	iQuran	70707916	USD	1.99	2929	966	4.5	4.5	3.3	4+	Reference	43	0	2		1	
23	285994151	:j) Sudoku +	6169600	USD	2.99	11447	781	5	5	5.2.6	4+	Games	40	5	1		1	
24	286058814	Yahoo Sports - Team	130583552	USD	0	137951	131	4	4.5	6.9	4+	Sports	37	2	6		1	
25	286070171	Microsoft Edge Leake	71203840	USD	5.00	8	0	4.5	0.0.0.5	4+	Business	37	5	3		1		

Google_Play_Store_raw.csv

App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
Photo Editor & Candy Camera & Grid & Scrap	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Prete	15-Jan-18	2.0.0	4.0.3 and up
U Launcher Lite " "FREE Live Cool Themes, H	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	8-Jun-18	Varies with dev	4.2 and up
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creat	20-Jun-18		1.1 4.4 and up
Paper flowers instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17		1 2.3 and up
Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18		1.1 4.0.3 and up
Infinite Painter	ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
Garden Coloring Book	ART_AND_DESIGN	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;Creat	3-Jul-18		2.8 4.0.3 and up
Text on Photo - Fontee	ART_AND_DESIGN	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
Name Art Photo Editor - Focus n Filters	ART_AND_DESIGN	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
Tattoo Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	2-Apr-18		3.8 4.1 and up
Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	26-Jun-18	1.0.4	4.4 and up
3D Color Pixel by Number - Sandbox Art Colc	ART_AND_DESIGN	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
Learn To Draw Kawaii Characters	ART_AND_DESIGN	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
Photo Designer - Write your name with shap	ART_AND_DESIGN	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	31-Jul-18		3.1 4.1 and up
350 Diy Room Decor Ideas	ART_AND_DESIGN	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	7-Nov-17		1 2.3 and up
FlipaClip - Cartoon animation	ART_AND_DESIGN	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
ibis Paint X	ART_AND_DESIGN	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up
Logo Maker - Small Business	ART_AND_DESIGN	4	450	14M	100,000+	Free	0	Everyone	Art & Design	20-Apr-18		4 4.1 and up
Boys Photo Editor - Six Pack & Men's Suit	ART_AND_DESIGN	4.1	654	12M	100,000+	Free	0	Everyone	Art & Design	20-Mar-18		1.1 4.0.3 and up
Superheroes Wallpapers 4K Backgrounds	ART_AND_DESIGN	4.7	7699	4.2M	500,000+	Free	0	Everyone 10+	Art & Design	12-Jul-18	2.2.6.2	4.0.3 and up

Work Flow Diagram #2: We regrouped and combined game category in both Apple and Google App Stores. For example, for Apple, we put Catalog together with Shopping. For Google, I changed the Business, Art_and_Design, and Events categories to Business so it can be compared with Apple Store's Business category.

Apple Store	Google					
Book	BOOKS_AND_REFERENCE					
Business	BUSINESS	ART_AND_DESIGN	EVENTS			
Catalogs --> Shopping						
Education	EDUCATION					
Entertainment	ENTERTAINMENT	COMICS				
Finance	FINANCE					
Food & Drink	FOOD_AND_DRINK					
Games	GAME	FAMILY				
Health & Fitness	HEALTH_AND_FITNESS					
Lifestyle	LIFESTYLE	BEAUTY	DATING	HOUSE_AND_HOME	PERSONALIZATION	PARENTING
Medical	MEDICAL					
Music --> Entertainment						
Navigation	MAPS_AND_NAVIGATION					
News	NEWS_AND_MAGAZINES					
Photo & Video	PHOTOGRAPHY	VIDEO_PLAYERS				
Productivity	PRODUCTIVITY					
Reference --> Book						
Shopping	SHOPPING					
Social Networking	SOCIAL	COMMUNICATION				
Sports	SPORTS					
Travel	TRAVEL_AND_LOCAL	AUTO_AND_VEHICLES				
Utilities	LIBRARIES_AND_DEMO					
Weather	WEATHER					

Work Flow Diagram #2: Python Codes

```
In [1]: 1 import pandas as pd
2
3 # File to Load
4 AppleStore_data_to_load = "Resources/AppleStore_raw.csv"
5 GooglePlay_data_to_load = "Resources/Google_Play_Store_raw.csv"
6
7 # Read into Pandas Data Frames
8 AppleStore_data = pd.read_csv(AppleStore_data_to_load)
9 GooglePlay_data = pd.read_csv(GooglePlay_data_to_load)
```

```
In [2]: 1 print(AppleStore_data.columns)
2 print(len(AppleStore_data.columns))

Index(['Column1', 'id', 'track_name', 'size_bytes', 'currency', 'price',
       'rating_count_tot', 'rating_count_ver', 'user_rating',
       'user_rating_ver', 'ver', 'cont_rating', 'prime_genre',
       'sup_devices.num', 'ipadSc_urls.num', 'lang.num', 'vpp_lic'],
      dtype='object')
17
```

```
In [3]: 1 print(GooglePlay_data.columns)
2 print(len(GooglePlay_data.columns))

Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
13
```

```
In [4]: 1 Apple_CatGenresList = AppleStore_data["prime_genre"].unique()
2 print(Apple_CatGenresList)
3 print(len(Apple_CatGenresList))

['Games' 'Productivity' 'Weather' 'Shopping' 'Reference' 'Finance' 'Music'
 'Utilities' 'Travel' 'Social Networking' 'Sports' 'Business'
 'Health & Fitness' 'Entertainment' 'Photo & Video' 'Navigation'
 'Education' 'Lifestyle' 'Food & Drink' 'News' 'Book' 'Medical' 'Catalogs']
23
```



```
In [5]: 1 Google_CatGenresList = GooglePlay_data["Category"].unique()
2 print(Google_CatGenresList)
3 print(len(Google_CatGenresList))
4
5
```

```
['ART_AND_DESIGN' 'AUTO_AND_VEHICLES' 'BEAUTY' 'BOOKS_AND_REFERENCE'
'BUSINESS' 'COMICS' 'COMMUNICATION' 'DATING' 'EDUCATION' 'ENTERTAINMENT'
'EVENTS' 'FINANCE' 'FOOD_AND_DRINK' 'HEALTH_AND_FITNESS' 'HOUSE_AND_HOME'
'LIBRARIES_AND_DEMO' 'LIFESTYLE' 'GAME' 'FAMILY' 'MEDICAL' 'SOCIAL'
'SHOPPING' 'PHOTOGRAPHY' 'SPORTS' 'TRAVEL_AND_LOCAL' 'TOOLS'
'PERSONALIZATION' 'PRODUCTIVITY' 'PARENTING' 'WEATHER' 'VIDEO_PLAYERS'
'NEWS_AND_MAGAZINES' 'MAPS_AND_NAVIGATION' '1.9']
34
```

Work on Google Pay Store

```
In [34]: 1 # reassigned category names for Google Play Store apps
2
3 for x in range(len(GooglePlay_data["Category"])):
4
5     if GooglePlay_data["Category"][x] == "BOOKS_AND_REFERENCE":
6         GooglePlay_data["Category"][x] = "Book"
7
8     elif GooglePlay_data["Category"][x] == "BUSINESS" \
9         or GooglePlay_data["Category"][x] == "ART_AND_DESIGN" \
10        or GooglePlay_data["Category"][x] == "EVENTS":
11         GooglePlay_data["Category"][x] = "Business"
12
13     elif GooglePlay_data["Category"][x] == "EDUCATION":
14         GooglePlay_data["Category"][x] = "Education"
15
16     elif GooglePlay_data["Category"][x] == "ENTERTAINMENT" \
17         or GooglePlay_data["Category"][x] == "COMICS":
18         GooglePlay_data["Category"][x] = "Entertainment"
19
20     elif GooglePlay_data["Category"][x] == "FINANCE":
21         GooglePlay_data["Category"][x] = "Finance"
22
23     elif GooglePlay_data["Category"][x] == "FOOD_AND_DRINK":
24         GooglePlay_data["Category"][x] = "Food & Drink"
25
26     elif GooglePlay_data["Category"][x] == "GAME" \
27         or GooglePlay_data["Category"][x] == "FAMILY":
28         GooglePlay_data["Category"][x] = "Games"
29
30     elif GooglePlay_data["Category"][x] == "HEALTH_AND_FITNESS":
31         GooglePlay_data["Category"][x] = "Health & Fitness"
32
33     elif GooglePlay_data["Category"][x] == "LIFESTYLE" \
34         or GooglePlay_data["Category"][x] == "BEAUTY" \
35         or GooglePlay_data["Category"][x] == "DATING" \
36         or GooglePlay_data["Category"][x] == "HOUSE_AND_HOME" \
37         or GooglePlay_data["Category"][x] == "PERSONALIZATION" \
38         or GooglePlay_data["Category"][x] == "PARENTING":
39         GooglePlay_data["Category"][x] = "Lifestyle"
40
41     elif GooglePlay_data["Category"][x] == "MEDICAL":
42         GooglePlay_data["Category"][x] = "Medical"
43
```

```

44 elif GooglePlay_data["Category"][x] == "MAPS_AND_NAVIGATION":
45     GooglePlay_data["Category"][x] = "Navigation"
46
47 elif GooglePlay_data["Category"][x] == "NEWS_AND_MAGAZINES":
48     GooglePlay_data["Category"][x] = "News"
49
50 elif GooglePlay_data["Category"][x] == "PHOTOGRAPHY":
51     or GooglePlay_data["Category"][x] == "VIDEO_PLAYERS":
52         GooglePlay_data["Category"][x] = "Photo & Video"
53
54 elif GooglePlay_data["Category"][x] == "PRODUCTIVITY":
55     GooglePlay_data["Category"][x] = "Productivity"
56
57 elif GooglePlay_data["Category"][x] == "SHOPPING":
58     GooglePlay_data["Category"][x] = "Shopping"
59
60 elif GooglePlay_data["Category"][x] == "SOCIAL":
61     or GooglePlay_data["Category"][x] == "COMMUNICATION":
62         GooglePlay_data["Category"][x] = "Social Networking"
63
64 elif GooglePlay_data["Category"][x] == "SPORTS":
65     GooglePlay_data["Category"][x] = "Sports"
66
67 elif GooglePlay_data["Category"][x] == "TRAVEL_AND_LOCAL":
68     or GooglePlay_data["Category"][x] == "AUTO_AND_VEHICLES":
69         GooglePlay_data["Category"][x] = "Travel"
70
71 elif GooglePlay_data["Category"][x] == "LIBRARIES_AND_DEMO":
72     GooglePlay_data["Category"][x] = "Utilities"
73
74 elif GooglePlay_data["Category"][x] == "WEATHER":
75     GooglePlay_data["Category"][x] = "Weather"
76
77

```

Work on Apple Store

```

In [48]: 1 # reassigned category names for Apple Store apps
2
3 for x in range(len(AppleStore_data["prime_genre"])):
4
5     if AppleStore_data["prime_genre"][x] == "Catalogs":
6         AppleStore_data["prime_genre"][x] = "Shopping"
7
8     elif AppleStore_data["prime_genre"][x] == "Music":
9         AppleStore_data["prime_genre"][x] = "Entertainment"
10
11    elif AppleStore_data["prime_genre"][x] == "Reference":
12        AppleStore_data["prime_genre"][x] = "Book"
13

```

```

In [49]: 1 # write the new Apple Store dataframe to a new csv file
2
3 df = pd.DataFrame(AppleStore_data)
4 df.to_csv("Resources/AppleStore_cleaned.csv")

```

Work Flow Diagram #3: upload the Apple_cleaned.csv and GooglePlay_cleaned.csv into SQL database

Work Flow Diagram #4: MySQL code

```

CREATE DATABASE appstores_db;

USE appstores_db;

#SHOW CREATE TABLE googleplaystore;
#SHOW CREATE TABLE applestore;

#SELECT * FROM googleplaystore;
#SELECT * FROM applestore;

DROP TABLE IF EXISTS appstores_UNION;

CREATE TABLE appstores_UNION as
SELECT App as app_name, Category as category, Rating as rating, Reviews as rating_count, Price as price, 'google' as app_source FROM googleplay_cleaned
UNION
SELECT track_name as app_name, prime_genre as category, user_rating as rating, rating_count_tot as rating_count, price, 'apple' as app_source FROM applestore_cleaned
;

SELECT * FROM appstores_UNION;

```

appstore_UNION table

	A	B	C	D	E	F
1	app_name	category	rating	rating_count	price	app_source
2	Photo Editor	Business	4.1	159	0	google
3	Coloring boo	Business	3.9	967	0	google
4	U Launcher L	Education	4.7	87510	0	google
5	Sketch - Dra	Business	4.5	215644	0	google
6	Pixel Draw -	Business	4.3	967	0	google
7	Paper flower	Business	4.4	167	0	google
8	Smoke Effect	Business	3.8	178	0	google
9	Infinite Paint	Business	4.1	36815	0	google
10	Garden Color	Business	4.4	13791	0	google
11	Kids Paint Fr	Business	4.7	121	0	google
12	Text on Phot	Business	4.4	13880	0	google
13	Name Art Ph	Business	4.4	8788	0	google
14	Tattoo Name	Business	4.2	44829	0	google
15	Mandala Coli	Business	4.6	4326	0	google
16	3D Color Pixe	Business	4.4	1518	0	google
17	Learn To Dra	Business	3.2	55	0	google
18	Photo Desigr	Business	4.7	3632	0	google
19	350 Diy Roor	Business	4.5	27	0	google
20	FlipaClip - C	Business	4.3	194216	0	google
21	ibis Paint X	Business	4.6	224399	0	google
22	Logo Maker -	Business	4	450	0	google
23	Boys Photo E	Business	4.1	654	0	google
24	Superheroes	Business	4.7	7699	0	google
25	HD Mickey M	Business	4.7	118	0	google
26	Harley Quinn	Business	4.8	192	0	google
27	Colorfit - Dra	Business	4.7	20260	0	google
28	Animated Ph	Business	4.1	203	0	google
29	Pencil Sketch	Business	3.9	136	0	google
30	Easy Realisti	Business	4.1	223	0	google
31	Pink Silver Br	Business	4.2	1120	0	google
32	Art Drawing	Business	4.1	227	0	google
33	Anime Mang	Business	4.5	5035	0	google
34	Easy Origam	Business	4.2	1015	0	google
35	I Creative Ide	Business	4.7	353	0	google
36	How to draw	Business	3.8	564	0	google
37	UNICORN - C	Business	4.7	8145	0	google

Visualization of Schema

