Google Play Store Analysis – PySpark on Google Colab

Problem Statement

Analyze the Google Play Store dataset using PySpark in Google Colab and present: 1) Top 10 Apps by Reviews 2) Top 10 Apps by Installs and Free vs Paid distribution 3) Category-wise distribution of installs 4) Top Paid Apps by Installs 5) Top Paid Apps by Rating (with review threshold)

Step 1: Load with Explicit Schema

We load the Excel as CSV and ingest into Spark with an explicit schema to avoid mixed-type inference errors.
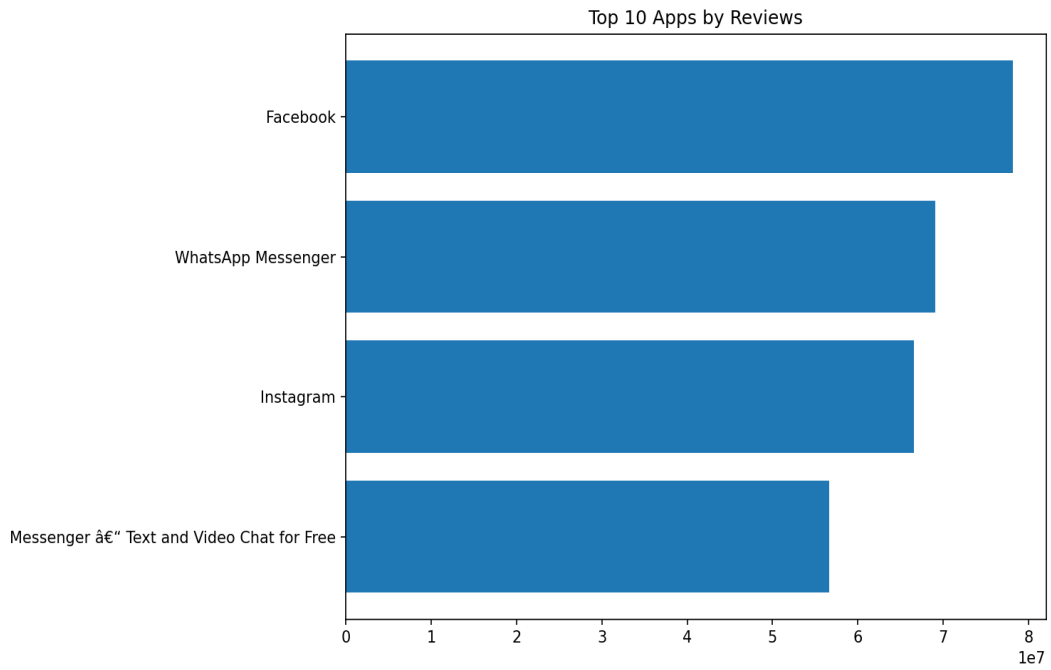
Step 2: Clean & Normalize

We normalize installs, price, reviews, rating, and tidy Type/Category. Ratings constrained to [0,5] when present.

Step 3: Compute Agenda Queries

We build the key aggregations and top-k selections required by the agenda.

Top 10 Apps by Reviews

Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	ins
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-08-03 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-08-03 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-03 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-03 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-03 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-07-31 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-07-31 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-07-31 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Teen	Social	2018-07-31 00:00:00	Varies with device	Varies with device	100
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-01 00:00:00	Varies with device	Varies with device	100

Top 10 Apps by Reviews

Top 10 Apps by Installs

| Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	i
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-07-21 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-01 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Teen	Books & Reference	2018-08-03 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-03 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-01 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-01 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-02 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-07-21 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-01 00:00:00	Varies with device	Varies with device	1
Varies with device	1,000,000,000+	Free	0	Everyone	Communication	2018-08-03 00:00:00	Varies with device	Varies with device	1

Top 10 Apps by Installs
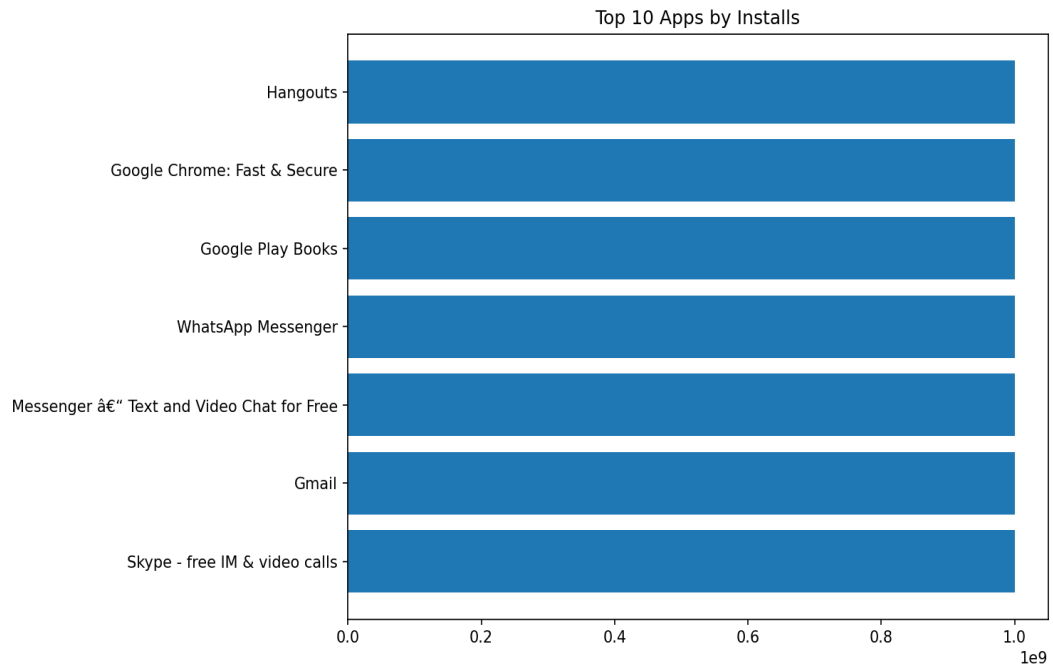
Free vs Paid (Count)

type_norm	app_count
102248	1
	1
Free	10037
Paid	800
2509	1

Free vs Paid (Count)
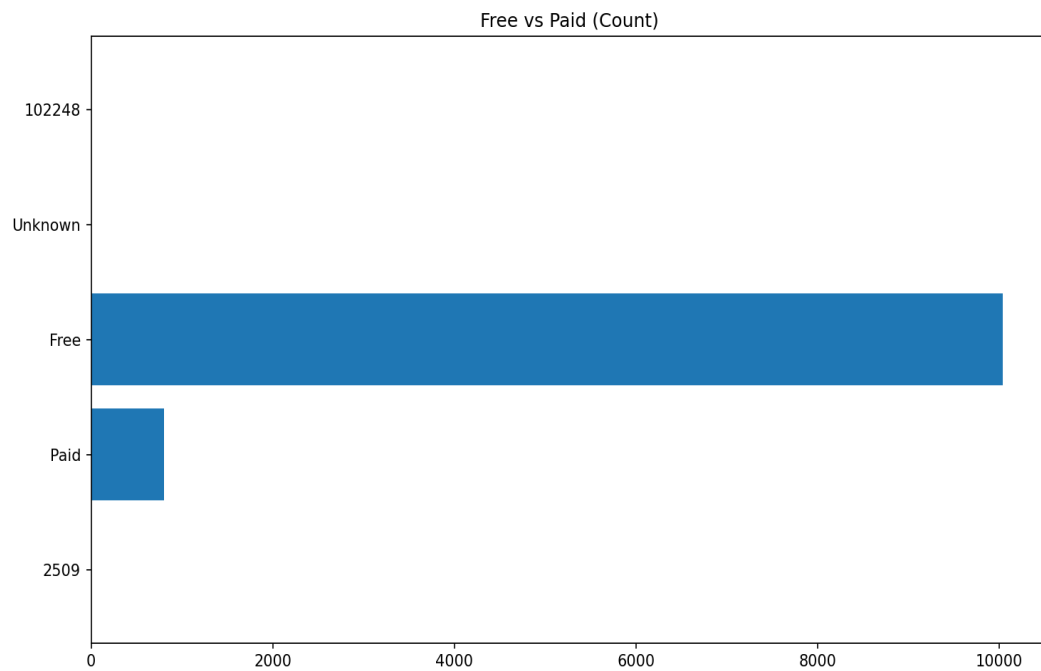
Free vs Paid (Installs)

type_norm	total_installs
102248	4.0
	0.0
Free	167554477406.0
Paid	72956081.0
2509	4.0

Free vs Paid (Installs)
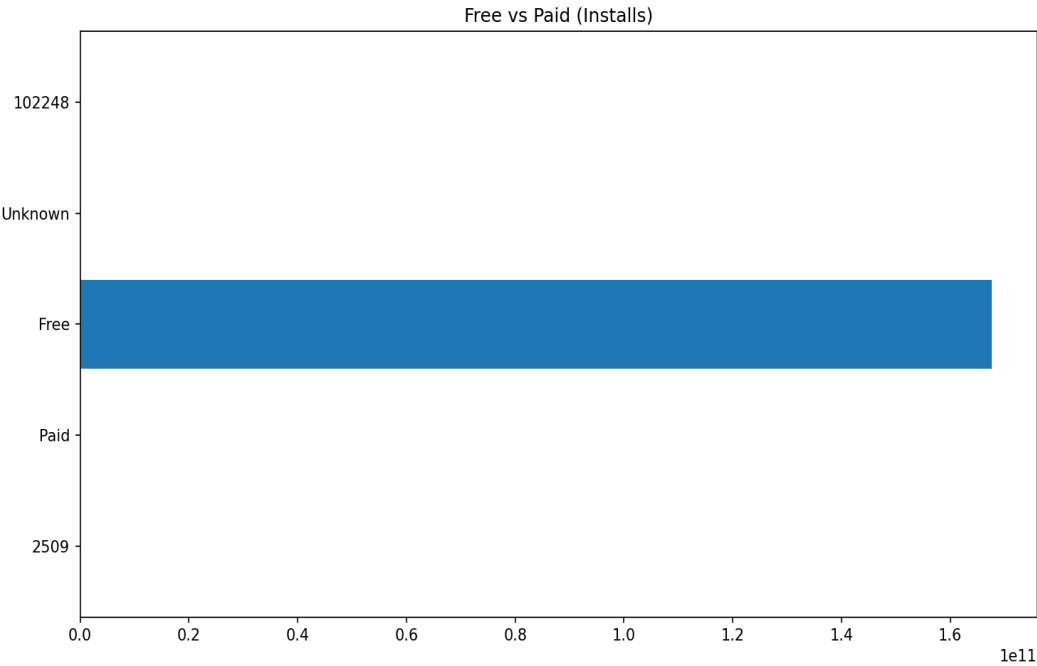
Top Categories by Installs

| category_norm	num_apps	total_installs	avg_rating
GAME	1144	35086024415.0	4.2863263445761195
COMMUNICATION	387	32647276251.0	4.158536585365852
PRODUCTIVITY	424	14176091369.0	4.211396011396012
SOCIAL	295	14069867902.0	4.255598455598457
TOOLS	843	11452771915.0	4.047411444141691
FAMILY	1972	10258263505.0	4.192272467086437
PHOTOGRAPHY	335	10088247655.0	4.192113564668767
NEWS_AND_MAGAZINES	283	7496317760.0	4.1321888412017165
TRAVEL_AND_LOCAL	258	6868887146.0	4.10929203539823
VIDEO_PLAYERS	175	6222002720.0	4.063750000000001

Top Categories by Installs

Top Categories by Installs
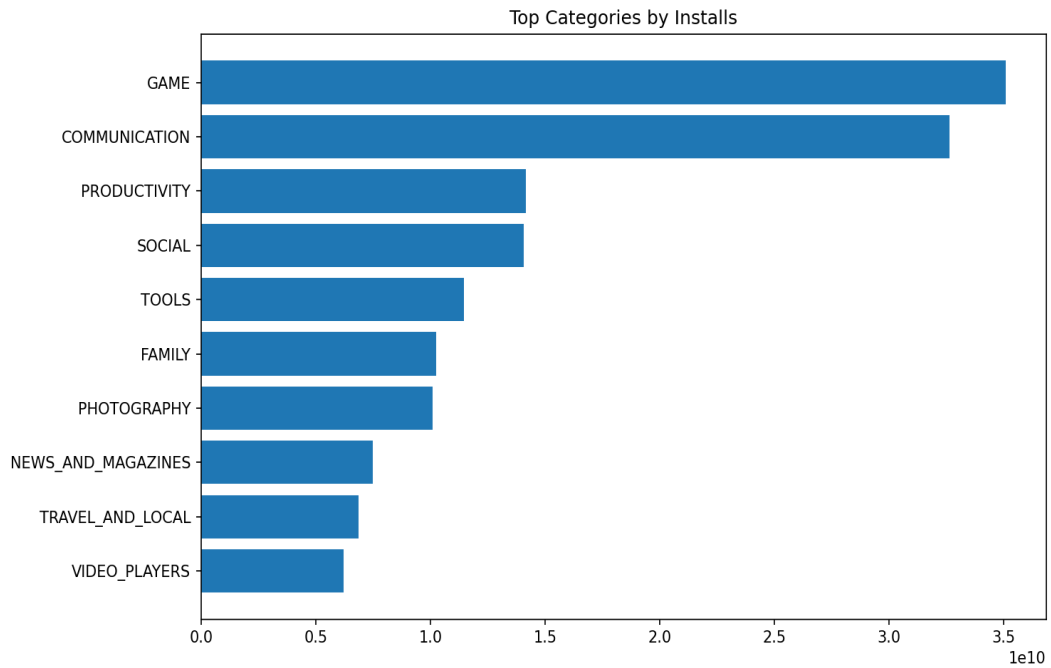
Top Paid Apps by Installs

Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	ins
Varies with device	10,000,000+	Paid	6.99	Everyone 10+	Arcade;Action & Adventure	2018-07-24 00:00:00	1.5.2.1	Varies with device	10
29M	10,000,000+	Paid	0.99	Mature 17+	Action	2018-07-12 00:00:00	1.7.110758	4.1 and up	10
Varies with device	10,000,000+	Paid	6.99	Everyone 10+	Arcade;Action & Adventure	2018-07-24 00:00:00	1.5.2.1	Varies with device	10
48M	1,000,000+	Paid	5.99	Everyone	Photography	2018-07-25 00:00:00	1.3.1	4.1 and up	10
96M	1,000,000+	Paid	0.99	Teen	Action	2018-06-26 00:00:00	2.4.4	2.3 and up	10
43M	1,000,000+	Paid	0.99	Everyone	Puzzle	2018-06-19 00:00:00	3.7.0	4.1 and up	10
73M	1,000,000+	Paid	1.99	Everyone	Sports	2018-08-04 00:00:00	1.5.1	4.0.3 and up	10
14M	1,000,000+	Paid	2.49	Everyone	Personalization	2016-08-24 00:00:00	5.7.8	2.3 and up	10
872k	1,000,000+	Paid	5.99	Everyone	Lifestyle	2018-06-27 00:00:00	20180608	4.0 and up	10
23M	1,000,000+	Paid	2.99	Everyone 10+	Card;Action & Adventure	2016-02-11 00:00:00	1.11.0	2.3.3 and up	10

Top Paid Apps by Installs
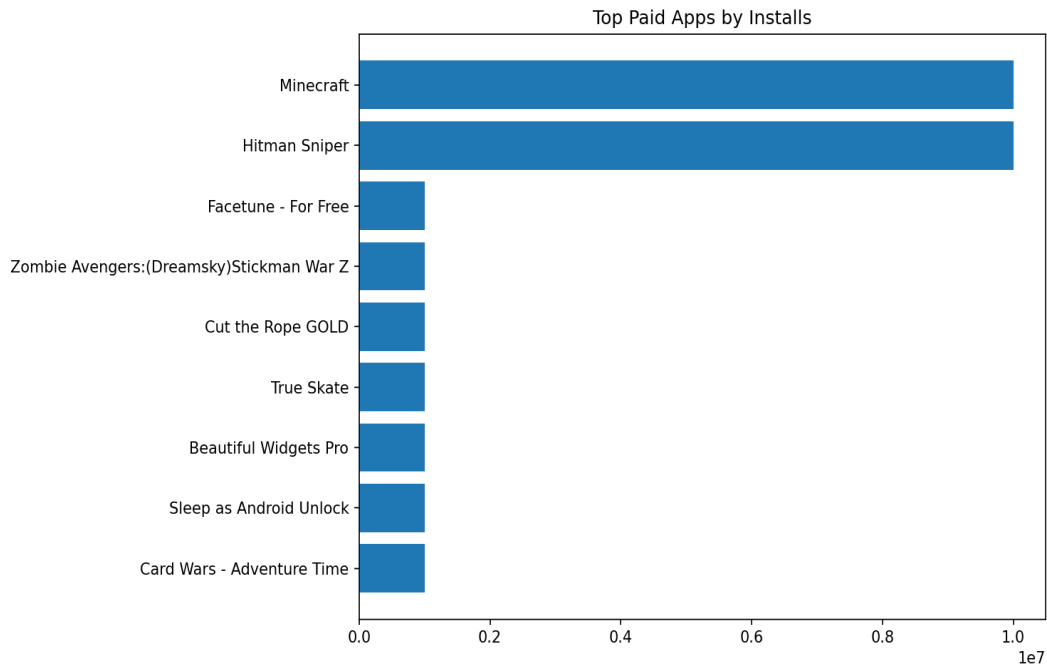
Top Paid Apps by Rating

Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	installs_n
48M	100,000+	Paid	4.99	Everyone	Puzzle	2018-04-18 00:00:00	1.0.1	4.4 and up	100000
57M	100,000+	Paid	4.99	Teen	Role Playing	2018-06-21 00:00:00	1.574	4.0.3 and up	100000
Varies with device	100,000+	Paid	2.99	Everyone	Health & Fitness	2018-01-19 00:00:00	Varies with device	Varies with device	100000
11M	100,000+	Paid	4.49	Everyone	Weather	2018-04-20 00:00:00	1.3	4.4 and up	100000
10M	100,000+	Paid	3.49	Everyone	Tools	2018-08-03 00:00:00	3.2.5	5.0 and up	100000
Varies with device	100,000+	Paid	4.99	Everyone	Puzzle	2018-06-14 00:00:00	1.0.4	4.1 and up	100000
39M	100,000+	Paid	4.99	Everyone	Business	2017-04-11 00:00:00	3.4.6	3.0 and up	100000
39M	100,000+	Paid	4.99	Everyone	Business	2017-04-11 00:00:00	3.4.6	3.0 and up	100000
Varies with device	10,000+	Paid	2.49	Everyone	Lifestyle	2018-07-30 00:00:00	Varies with device	Varies with device	10000
15M	100,000+	Paid	4.49	Everyone	Productivity	2018-07-18 00:00:00	1	4.1 and up	100000

Top Paid Apps by Rating
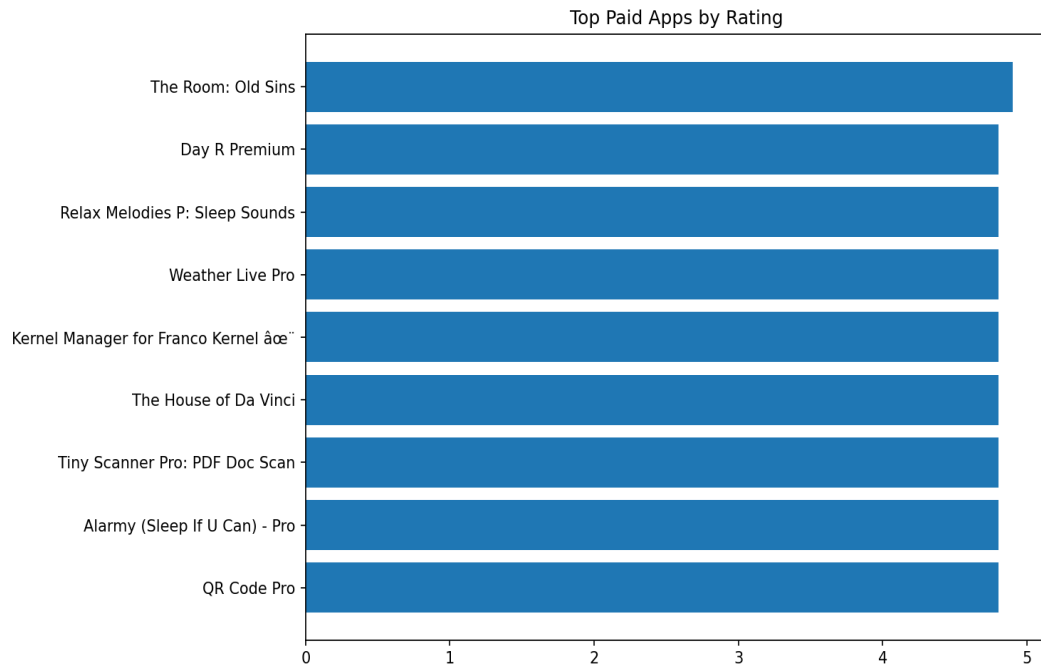
Conclusions & Insights

Total apps analyzed: **10840**.

Total installs (sum): **167,627,433,495**.

Free apps share (by app count): **92.6%**; Free apps share (by installs): **100.0%**.

Top category by total installs: **GAME** with **35,086,024,415** installs.

Average rating — Free: **4.19**, Paid: **4.27**.

Best paid app by rating (≥ 1000 reviews): **The Room: Old Sins** (Rating **4.90**, Reviews **21,119**, Price **$4.99** if listed, Category **GAME**).

Correlation (paid apps) between price and rating: **-0.11** (near 0 implies weak linear relationship).