Advanced Database Assignment
OLAP operations and queries

Manoj S
2019506048

Aim

To analyse play store application data

Data Format

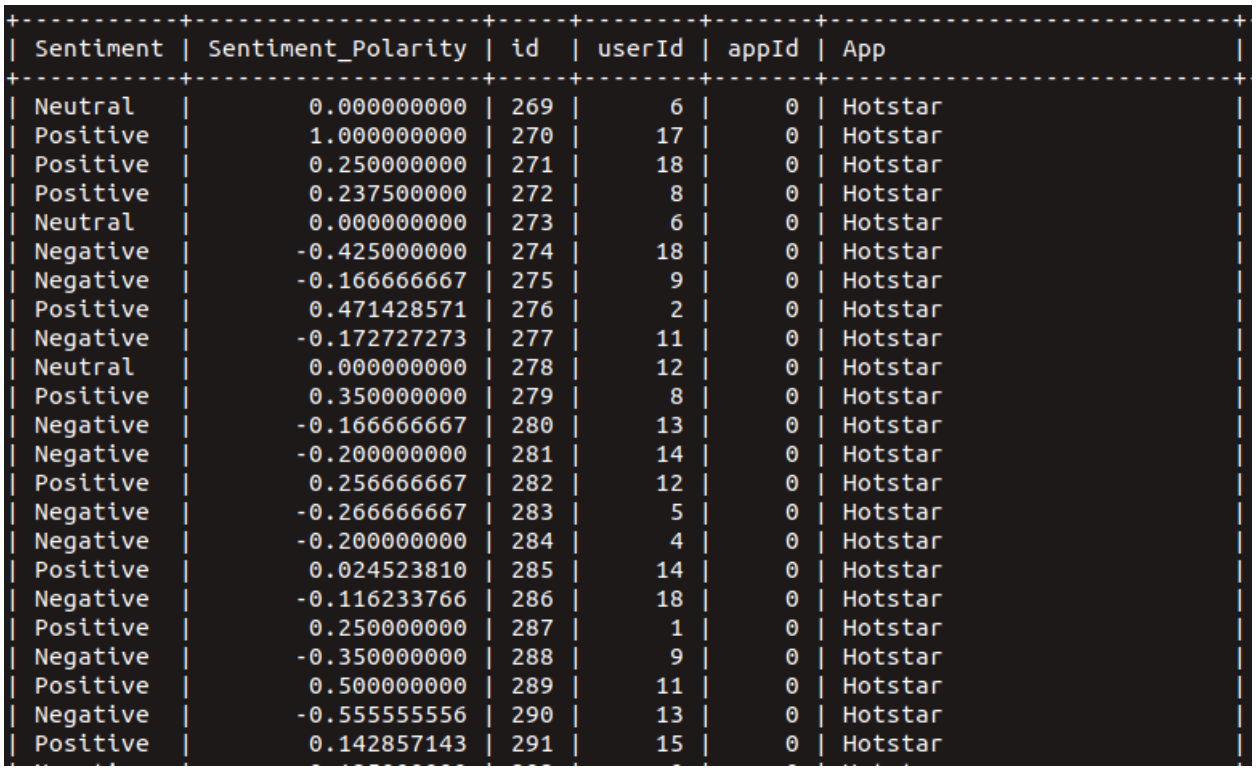
```
CREATE TABLE IF NOT EXISTS users (
    `state` VARCHAR(13) CHARACTER SET utf8,
    `location` VARCHAR(15) CHARACTER SET utf8,
    `version` INT,
    `id` INT PRIMARY KEY
);

CREATE TABLE IF NOT EXISTS apps (
    `App` VARCHAR(26) CHARACTER SET utf8,
    `Category` VARCHAR(15) CHARACTER SET utf8,
    `Rating` NUMERIC(2, 1),
    `Size` VARCHAR(4) CHARACTER SET utf8,
    `Installs` INT,
    `id` INT PRIMARY KEY
);


CREATE TABLE IF NOT EXISTS rating (
    `Sentiment` VARCHAR(8) CHARACTER SET utf8,
    `Sentiment_Polarity` NUMERIC(10, 9),
    `id` INT PRIMARY KEY,
    `userId` INT,
    `appId` INT,
    FOREIGN KEY (userId) REFERENCES users(id),
    FOREIGN KEY (appId) REFERENCES apps(id)
);
```

Here rating is the dimension table. apps , users are the fact table. The dimension table is related using the foreign keys userId and appId

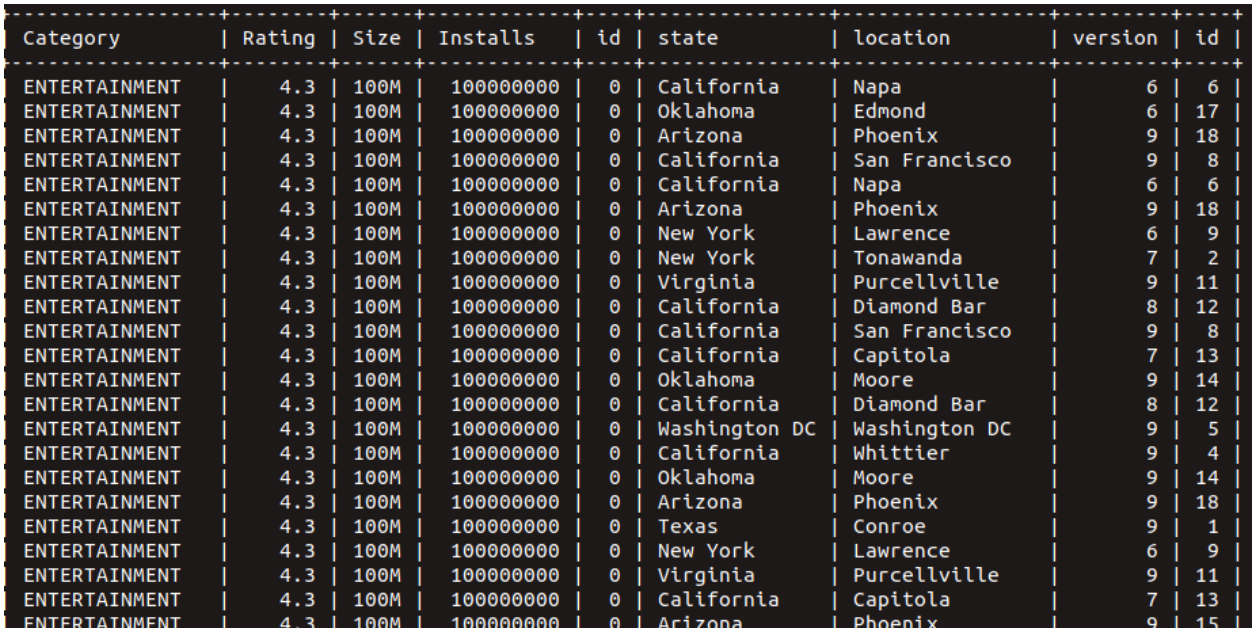
Data was cleansed and only selected 6 apps were kept from https://www.kaggle.com/lava18/google-play-store-apps?select=googleplaystore_user_reviews.csv

Full Join

```
select * from rating , apps , users where appId = apps.id and userId = users.id;
```



Sentiment	Sentiment_Polarity	id	userId	appId	App
Neutral	0.000000000	269	6	0	Hotstar
Positive	1.000000000	270	17	0	Hotstar
Positive	0.250000000	271	18	0	Hotstar
Positive	0.237500000	272	8	0	Hotstar
Neutral	0.000000000	273	6	0	Hotstar
Negative	-0.425000000	274	18	0	Hotstar
Negative	-0.166666667	275	9	0	Hotstar
Positive	0.471428571	276	2	0	Hotstar
Negative	-0.172727273	277	11	0	Hotstar
Neutral	0.000000000	278	12	0	Hotstar
Positive	0.350000000	279	8	0	Hotstar
Negative	-0.166666667	280	13	0	Hotstar
Negative	-0.200000000	281	14	0	Hotstar
Positive	0.256666667	282	12	0	Hotstar
Negative	-0.266666667	283	5	0	Hotstar
Negative	-0.200000000	284	4	0	Hotstar
Positive	0.024523810	285	14	0	Hotstar
Negative	-0.116233766	286	18	0	Hotstar
Positive	0.250000000	287	1	0	Hotstar
Negative	-0.350000000	288	9	0	Hotstar
Positive	0.500000000	289	11	0	Hotstar
Negative	-0.555555556	290	13	0	Hotstar
Positive	0.142857143	291	15	0	Hotstar

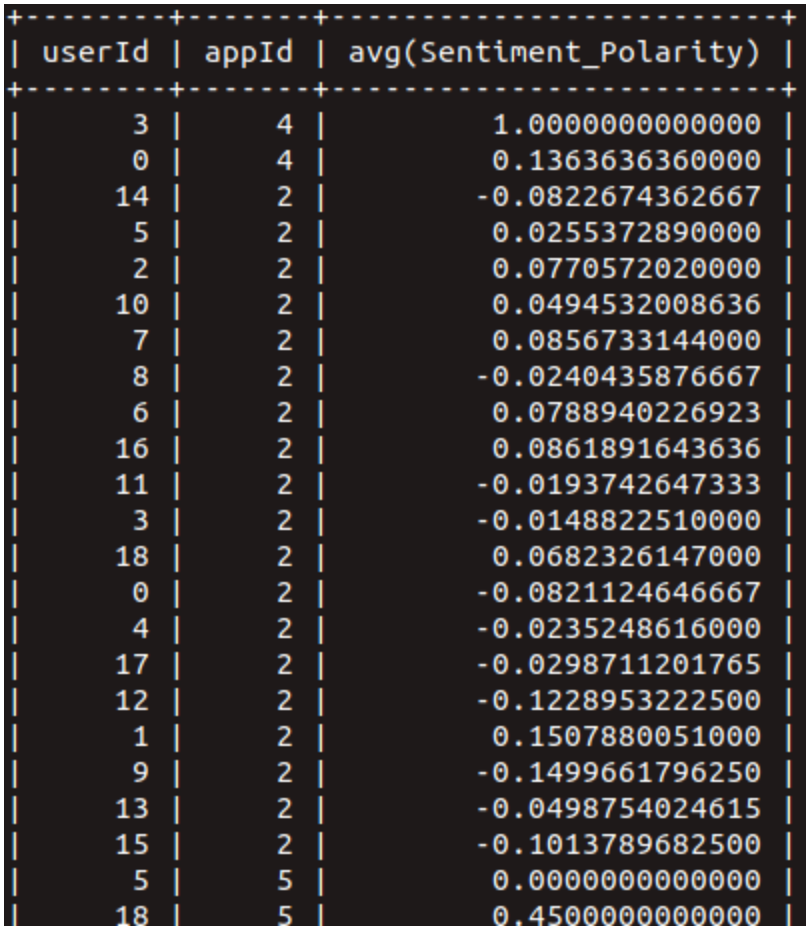


Category	Rating	Size	Installs	id	state	location	version	id
ENTERTAINMENT	4.3	100M	100000000	0	California	Napa	6	6
ENTERTAINMENT	4.3	100M	100000000	0	Oklahoma	Edmond	6	17
ENTERTAINMENT	4.3	100M	100000000	0	Arizona	Phoenix	9	18
ENTERTAINMENT	4.3	100M	100000000	0	California	San Francisco	9	8
ENTERTAINMENT	4.3	100M	100000000	0	California	Napa	6	6
ENTERTAINMENT	4.3	100M	100000000	0	Arizona	Phoenix	9	18
ENTERTAINMENT	4.3	100M	100000000	0	New York	Lawrence	6	9
ENTERTAINMENT	4.3	100M	100000000	0	New York	Tonawanda	7	2
ENTERTAINMENT	4.3	100M	100000000	0	Virginia	Purcellville	9	11
ENTERTAINMENT	4.3	100M	100000000	0	California	Diamond Bar	8	12
ENTERTAINMENT	4.3	100M	100000000	0	California	San Francisco	9	8
ENTERTAINMENT	4.3	100M	100000000	0	California	Capitola	7	13
ENTERTAINMENT	4.3	100M	100000000	0	Oklahoma	Moore	9	14
ENTERTAINMENT	4.3	100M	100000000	0	California	Diamond Bar	8	12
ENTERTAINMENT	4.3	100M	100000000	0	Washington DC	Washington DC	9	5
ENTERTAINMENT	4.3	100M	100000000	0	California	Whittier	9	4
ENTERTAINMENT	4.3	100M	100000000	0	Oklahoma	Moore	9	14
ENTERTAINMENT	4.3	100M	100000000	0	Arizona	Phoenix	9	18
ENTERTAINMENT	4.3	100M	100000000	0	Texas	Conroe	9	1
ENTERTAINMENT	4.3	100M	100000000	0	New York	Lawrence	6	9
ENTERTAINMENT	4.3	100M	100000000	0	Virginia	Purcellville	9	11
ENTERTAINMENT	4.3	100M	100000000	0	California	Capitola	7	13
ENTERTAINMENT	4.3	100M	100000000	0	Arizona	Phoenix	9	15

Summarized Joins

```
select userId , appId , avg(Sentiment_Polarity) FROM rating GROUP BY
userId , appId;
```

average of all reviews of a person to a particular app



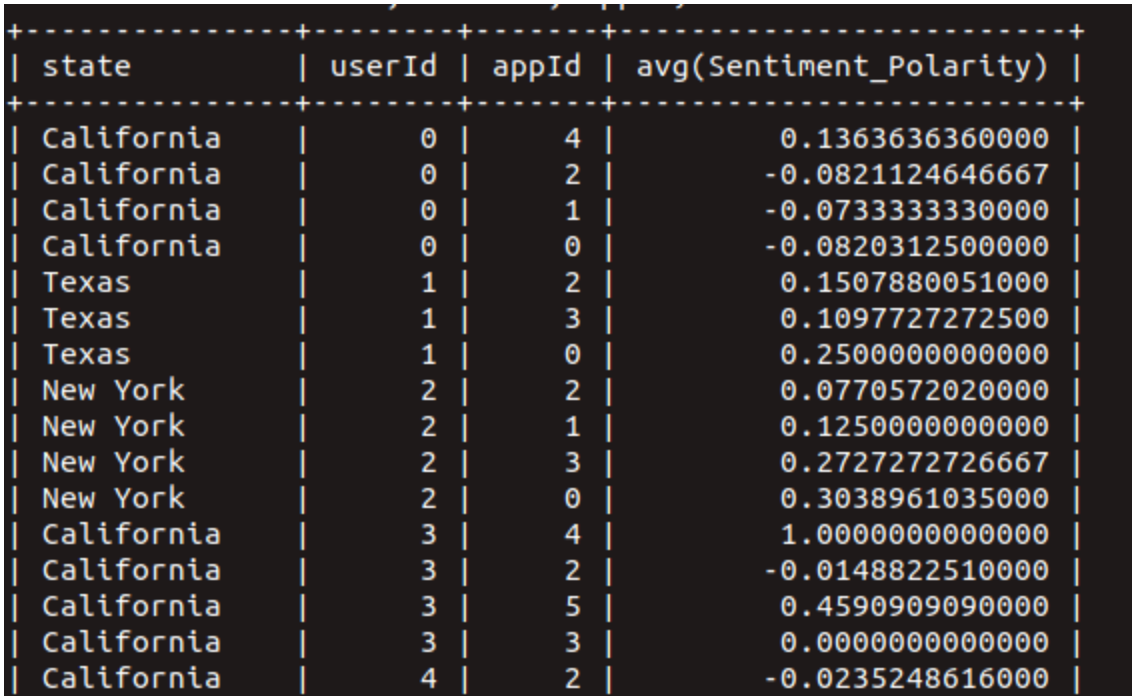
userId	appId	avg(Sentiment_Polarity)
3	4	1.0000000000000
0	4	0.1363636360000
14	2	-0.0822674362667
5	2	0.0255372890000
2	2	0.0770572020000
10	2	0.0494532008636
7	2	0.0856733144000
8	2	-0.0240435876667
6	2	0.0788940226923
16	2	0.0861891643636
11	2	-0.0193742647333
3	2	-0.0148822510000
18	2	0.0682326147000
0	2	-0.0821124646667
4	2	-0.0235248616000
17	2	-0.0298711201765
12	2	-0.1228953222500
1	2	0.1507880051000
9	2	-0.1499661796250
13	2	-0.0498754024615
15	2	-0.1013789682500
5	5	0.0000000000000
18	5	0.4500000000000

OLAP queries

Roll Up

```
select state, userId , appId , avg(Sentiment_Polarity) FROM rating ,
users WHERE users.id = rating.userId GROUP BY state, userId , appId;
```

Sentiment polarity per state per user per app
The dimension is reduced - compared with full join



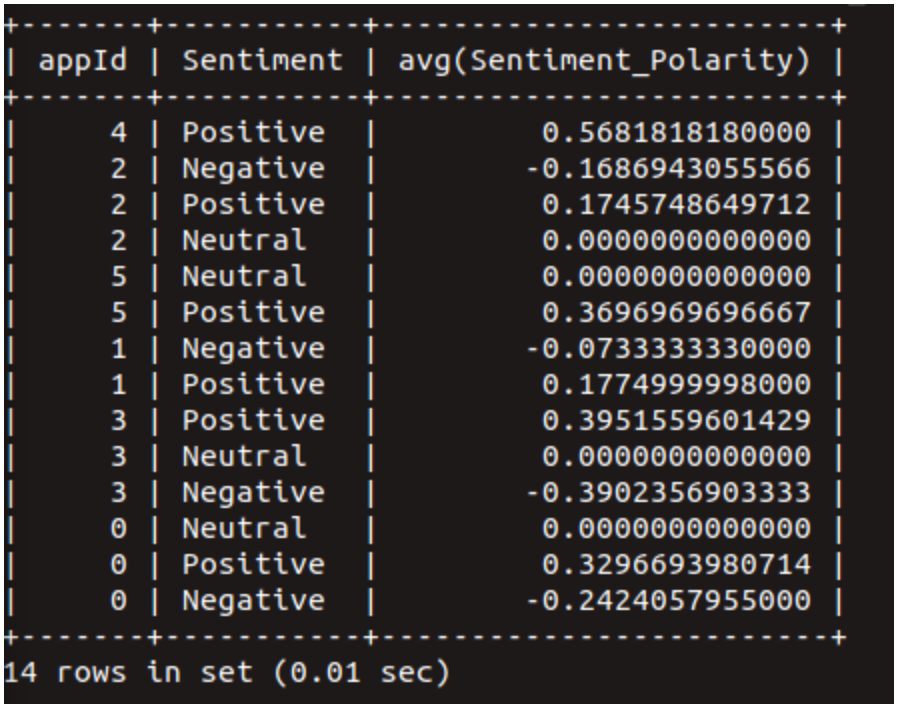
state	userId	appId	avg(Sentiment_Polarity)
California	0	4	0.1363636360000
California	0	2	-0.0821124646667
California	0	1	-0.0733333330000
California	0	0	-0.0820312500000
Texas	1	2	0.1507880051000
Texas	1	3	0.1097727272500
Texas	1	0	0.2500000000000
New York	2	2	0.0770572020000
New York	2	1	0.1250000000000
New York	2	3	0.2727272726667
New York	2	0	0.3038961035000
California	3	4	1.0000000000000
California	3	2	-0.0148822510000
California	3	5	0.4590909090000
California	3	3	0.0000000000000
California	4	2	-0.0235248616000

```
select appId , Sentiment , avg(Sentiment_Polarity) FROM rating GROUP BY appId , Sentiment;
```

| - no sentiment grouping

```
select appId , avg(Sentiment_Polarity) FROM rating GROUP BY appId;
```

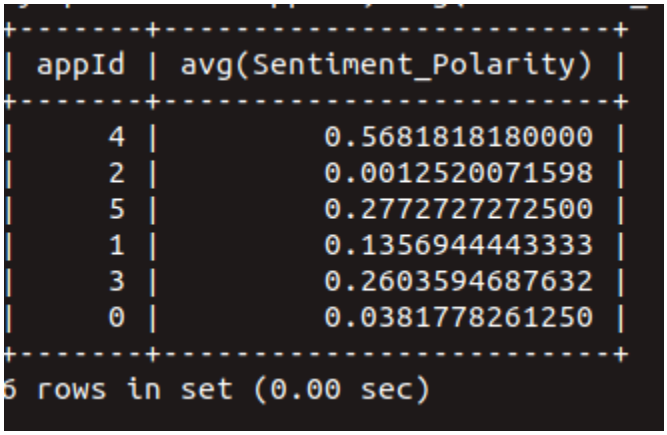
Particular apps positive , negative and neutral wise average polarity



appId	Sentiment	avg(Sentiment_Polarity)
4	Positive	0.5681818180000
2	Negative	-0.1686943055566
2	Positive	0.1745748649712
2	Neutral	0.0000000000000
5	Neutral	0.0000000000000
5	Positive	0.3696969696667
1	Negative	-0.0733333330000
1	Positive	0.1774999998000
3	Positive	0.3951559601429
3	Neutral	0.0000000000000
3	Negative	-0.3902356903333
0	Neutral	0.0000000000000
0	Positive	0.3296693980714
0	Negative	-0.2424057955000

14 rows in set (0.01 sec)

|- reduced dimensions



appId	avg(Sentiment_Polarity)
4	0.5681818180000
2	0.0012520071598
5	0.2772727272500
1	0.1356944443333
3	0.2603594687632
0	0.0381778261250

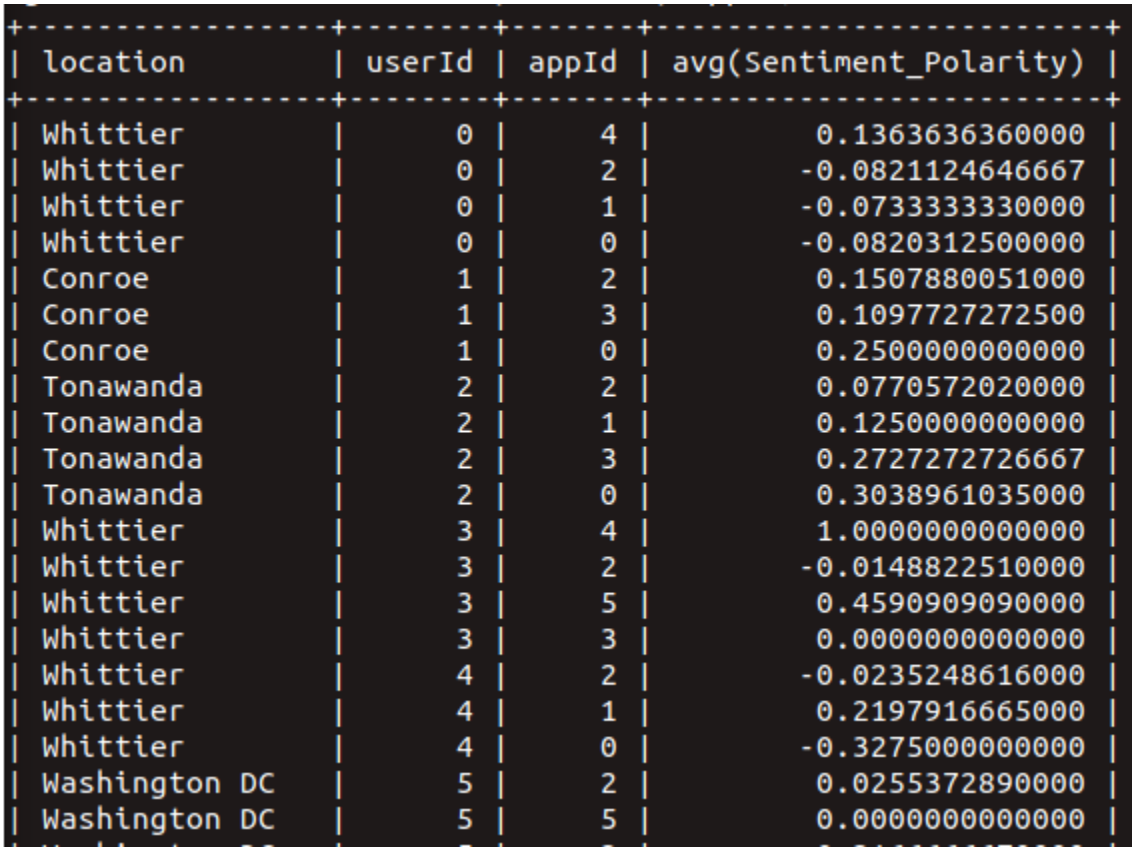
6 rows in set (0.00 sec)

Drill Down

```
select state, userId , appId , avg(Sentiment_Polarity) FROM rating ,
users WHERE users.id = rating.userId GROUP BY state, userId , appId;
```

| - state expanded to location

```
select location, userId , appId , avg(Sentiment_Polarity) FROM rating ,
users WHERE users.id = rating.userId GROUP BY location, userId , appId;
```



location	userId	appId	avg(Sentiment_Polarity)
Whittier	0	4	0.1363636360000
Whittier	0	2	-0.0821124646667
Whittier	0	1	-0.0733333330000
Whittier	0	0	-0.0820312500000
Conroe	1	2	0.1507880051000
Conroe	1	3	0.1097727272500
Conroe	1	0	0.2500000000000
Tonawanda	2	2	0.0770572020000
Tonawanda	2	1	0.1250000000000
Tonawanda	2	3	0.2727272726667
Tonawanda	2	0	0.3038961035000
Whittier	3	4	1.0000000000000
Whittier	3	2	-0.0148822510000
Whittier	3	5	0.4590909090000
Whittier	3	3	0.0000000000000
Whittier	4	2	-0.0235248616000
Whittier	4	1	0.2197916665000
Whittier	4	0	-0.3275000000000
Washington DC	5	2	0.0255372890000
Washington DC	5	5	0.0000000000000

```
select
avg(case when Rating > 4 and Rating <= 5 then Sentiment_Polarity else 0
end) as greater4,
avg(case when Rating > 3 and Rating <= 4 then Sentiment_Polarity else 0
end) as greater3
from rating , apps WHERE apps.id = rating.appId;
```

| - expanding the Rating column

```
select
avg(case when Rating > 4.5 and Rating <= 5 then Sentiment_Polarity else
0 end) as greater45,
avg(case when Rating > 4 and Rating <= 4.5 then Sentiment_Polarity else
0 end) as greater4,
avg(case when Rating > 3.5 and Rating <= 4 then Sentiment_Polarity else
0 end) as greater35,
avg(case when Rating > 3 and Rating <= 3.5 then Sentiment_Polarity else
0 end) as greater3
from rating , apps WHERE apps.id = rating.appId;
```

```
+------------------+------------------+
| greater4         | greater3         |
+------------------+------------------+
| 0.0443191698306 | 0.0036846874053 |
+------------------+------------------+
1 row in set (0.01 sec)
```

```
    -> from rating , apps WHERE apps.id = rating.appId;
+------------------+------------------+------------------+------------------+
| greater45        | greater4         | greater35        | greater3         |
+------------------+------------------+------------------+------------------+
| 0.0000000000000 | 0.0443191698306 | 0.0036846874053 | 0.0000000000000 |
+------------------+------------------+------------------+------------------+
1 row in set (0.00 sec)
```

```
select appId , avg(Sentiment_Polarity) FROM rating GROUP BY appId;
```

| - Including version of use by user

```
select version , appId , avg(Sentiment_Polarity) FROM rating, users
WHERE rating.userId = users.id GROUP BY version , appId;
```

appId	avg(Sentiment_Polarity)
0	0.0381778261250
1	0.1356944443333
2	0.0012520071598
3	0.2603594687632
4	0.5681818180000
5	0.2772727272500

6 rows in set (0.00 sec)

version	appId	avg(Sentiment_Polarity)
8	4	0.5681818180000
8	2	-0.0203181487234
8	1	-0.0733333330000
8	0	0.0899494048000
9	2	-0.0042704228721
9	3	0.1770454546154
9	0	0.0262033279375
7	2	0.0474648981250
7	1	0.0697916665000
7	3	0.4071969697000
7	0	-0.0286075040000
8	5	0.3295454545000
8	3	0.3212500000000
9	1	0.2493055553333
9	5	0.2250000000000
6	2	-0.0179451627632
6	3	0.1357284580000
6	0	0.0667314544286

18 rows in set (0.00 sec)

Dice

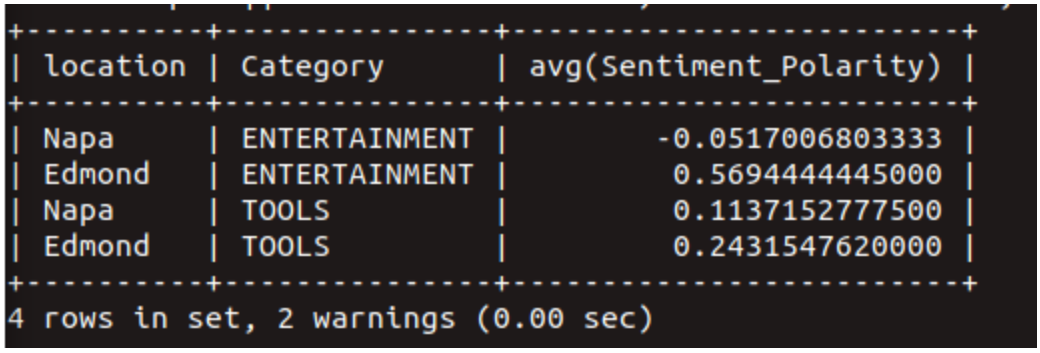
```
select location , Category , avg(Sentiment_Polarity) FROM rating ,
users, apps WHERE users.id = rating.userId AND rating.appId = apps.id
GROUP BY location , Category;
```

| - selecting subcube

```
select location , Category , avg(Sentiment_Polarity) FROM rating ,
users, apps WHERE users.id = rating.userId AND rating.appId = apps.id
AND ( Category = "ENTERTAINMENT" || Category = "TOOLS" ) AND ( location
= "Napa" || location= "Edmond" ) GROUP BY location , Category;
```

```
| Capitola          | ENTERTAINMENT |        -0.2358796300000 |
| Moore             | ENTERTAINMENT |        -0.0877380950000 |
| Washington DC     | ENTERTAINMENT |        -0.2666666670000 |
| Whittier          | ENTERTAINMENT |        -0.0646875000000 |
| Conroe            | ENTERTAINMENT |         0.2500000000000 |
| Moore             | GAME          |        -0.0822674362667 |
| Washington DC     | GAME          |         0.0255372890000 |
| Tonawanda         | GAME          |         0.0824422722500 |
| Purcellville      | GAME          |         0.0215501742703 |
| San Francisco     | GAME          |        -0.0240435876667 |
| Napa              | GAME          |         0.0788940226923 |
| Vadnais Heights   | GAME          |         0.0861891643636 |
| Whittier          | GAME          |        -0.0479633128519 |
| Phoenix           | GAME          |        -0.0071503110556 |
| Edmond            | GAME          |        -0.0298711201765 |
| Diamond Bar       | GAME          |        -0.1228953222500 |
| Conroe            | GAME          |         0.1507880051000 |
| Lawrence          | GAME          |        -0.1499661796250 |
| Capitola          | GAME          |        -0.0498754024615 |
| Vadnais Heights   | TOOLS         |         0.6000000000000 |
| Moore             | TOOLS         |         0.2000000000000 |
| Capitola          | TOOLS         |         0.4845959596667 |
| Diamond Bar       | TOOLS         |         0.4375000000000 |
| Tonawanda         | TOOLS         |         0.2045454545000 |
| Whittier          | TOOLS         |         0.0000000000000 |
| Napa              | TOOLS         |         0.1137152777500 |
```

|



location	Category	avg(Sentiment_Polarity)
Napa	ENTERTAINMENT	-0.0517006803333
Edmond	ENTERTAINMENT	0.5694444445000
Napa	TOOLS	0.1137152777500
Edmond	TOOLS	0.2431547620000

4 rows in set, 2 warnings (0.00 sec)

Slicing

```
select Category, version , avg(Sentiment_Polarity) FROM rating , apps , users WHERE apps.id = rating.appId AND users.id = rating.userId GROUP BY Category , version ;
```

| - getting details of version 9

```
select Category , avg(Sentiment_Polarity) FROM rating , apps , users WHERE apps.id = rating.appId AND users.id = rating.userId AND version = 9 GROUP BY Category , version ;
```

Category	version	avg(Sentiment_Polarity)
ENTERTAINMENT	6	0.0667314544286
ENTERTAINMENT	9	0.0614299954211
ENTERTAINMENT	7	0.0041922195000
ENTERTAINMENT	8	0.0627356151667
GAME	9	-0.0042704228721
GAME	7	0.0474648981250
GAME	8	-0.0203181487234
GAME	6	-0.0179451627632
TOOLS	7	0.4071969697000
TOOLS	9	0.1770454546154
TOOLS	8	0.3212500000000
TOOLS	6	0.1357284580000
PERSONALIZATION	8	0.5681818180000
FAMILY	9	0.2250000000000
FAMILY	8	0.3295454545000

15 rows in set (0.01 sec)

Category	avg(Sentiment_Polarity)
GAME	-0.0042704228721
TOOLS	0.1770454546154
ENTERTAINMENT	0.0614299954211
FAMILY	0.2250000000000

4 rows in set (0.00 sec)

```
select appId , location , avg(Sentiment_Polarity) FROM rating , users
WHERE users.id = rating.userId GROUP BY location , appId;
```

| - App 0 : location wise details

```
select location , avg(Sentiment_Polarity) FROM rating , users WHERE
users.id = rating.userId AND rating.appId = 0 GROUP BY location ;
```

appId	location	avg(Sentiment_Polarity)
4	Whittier	0.5681818180000
2	Whittier	-0.0479633128519
1	Whittier	0.1220833333333
0	Whittier	-0.2047656250000
2	Conroe	0.1507880051000
3	Conroe	0.1097727272500
0	Conroe	0.2500000000000
2	Tonawanda	0.0824422722500
1	Tonawanda	0.1250000000000
3	Tonawanda	0.2045454545000
0	Tonawanda	0.3038961035000
5	Whittier	0.4590909090000
3	Whittier	0.0000000000000
2	Washington DC	0.0255372890000
5	Washington DC	0.0000000000000
3	Washington DC	0.3166666670000
0	Washington DC	-0.2666666670000
2	Napa	0.0788940226923
3	Napa	0.1137152777500

location	avg(Sentiment_Polarity)
Napa	-0.0517006803333
Edmond	0.5694444445000
Phoenix	-0.0370941557500
San Francisco	0.1958333333333
Lawrence	-0.2583333335000
Tonawanda	0.3038961035000
Purcellville	0.2757575756667
Diamond Bar	0.2046031746667
Capitola	-0.3611111115000
Moore	-0.0877380950000
Washington DC	-0.2666666670000
Whittier	-0.2047656250000
Conroe	0.2500000000000

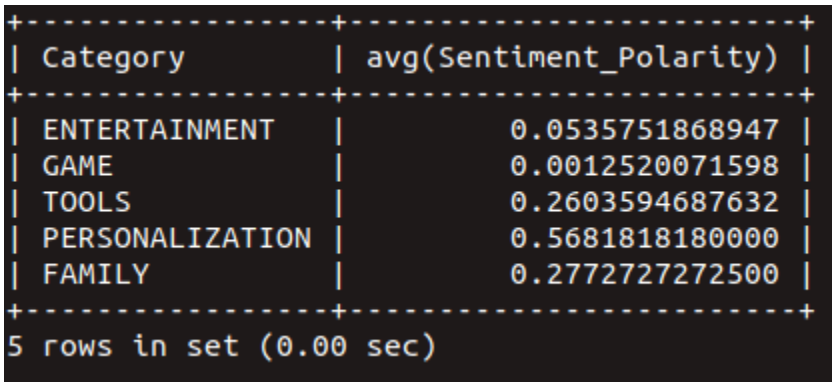
13 rows in set (0.00 sec)

Pivoting

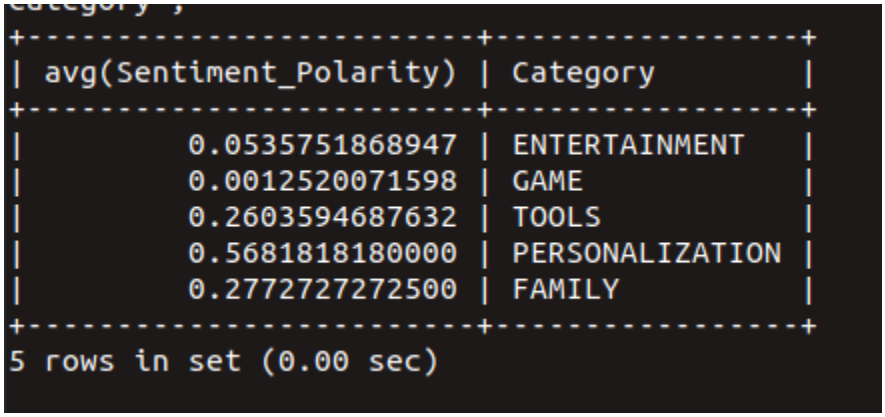
```
select Category , avg(Sentiment_Polarity) FROM rating , apps WHERE
apps.id = rating.appId GROUP BY Category ;
```

|

```
select avg(Sentiment_Polarity) , Category FROM rating , apps WHERE
apps.id = rating.appId GROUP BY Category ;
```



```
+------------------+--------------------------+
| Category         | avg(Sentiment_Polarity)  |
+------------------+--------------------------+
| ENTERTAINMENT    |          0.0535751868947 |
| GAME             |          0.0012520071598 |
| TOOLS            |          0.2603594687632 |
| PERSONALIZATION  |          0.5681818180000 |
| FAMILY           |          0.2772727272500 |
+------------------+--------------------------+
5 rows in set (0.00 sec)
```



```
+--------------------------+------------------+
| avg(Sentiment_Polarity)  | Category         |
+--------------------------+------------------+
|          0.0535751868947 | ENTERTAINMENT    |
|          0.0012520071598 | GAME             |
|          0.2603594687632 | TOOLS            |
|          0.5681818180000 | PERSONALIZATION  |
|          0.2772727272500 | FAMILY           |
+--------------------------+------------------+
5 rows in set (0.00 sec)
```

Other queries

```
select Category , avg(rating) FROM rating , apps , users WHERE apps.id =
rating.appId AND users.id = rating.userId GROUP BY Category;
```

Category	avg(rating)
ENTERTAINMENT	4.30000
GAME	4.50000
TOOLS	4.40000
PERSONALIZATION	4.50000
FAMILY	3.70000

5 rows in set (0.00 sec)

```
select * from users;
```

state	location	version	id
California	Whittier	8	0
Texas	Conroe	9	1
New York	Tonawanda	7	2
California	Whittier	8	3
California	Whittier	9	4
Washington DC	Washington DC	9	5
California	Napa	6	6
New York	Tonawanda	7	7
California	San Francisco	9	8
New York	Lawrence	6	9
Virginia	Purcellville	8	10
Virginia	Purcellville	9	11
California	Diamond Bar	8	12
California	Capitola	7	13
Oklahoma	Moore	9	14
Arizona	Phoenix	9	15
Minnesota	Vadnais Heights	7	16
Oklahoma	Edmond	6	17

```
select * from rating;
```

Sentiment	Sentiment_Polarity	id	userId	appId
Positive	1.000000000	0	3	4
Positive	0.136363636	1	0	4
Negative	-0.013333333	2	14	2
Negative	-0.140000000	3	5	2
Negative	-0.303571429	4	2	2
Negative	-0.066666667	5	10	2
Positive	0.013636364	6	7	2
Negative	-0.572916667	7	8	2
Positive	0.243333333	8	6	2
Negative	-0.083333333	9	16	2
Positive	0.140625000	10	11	2
Positive	0.154761905	11	10	2
Negative	-0.132142857	12	3	2
Positive	0.137500000	13	10	2

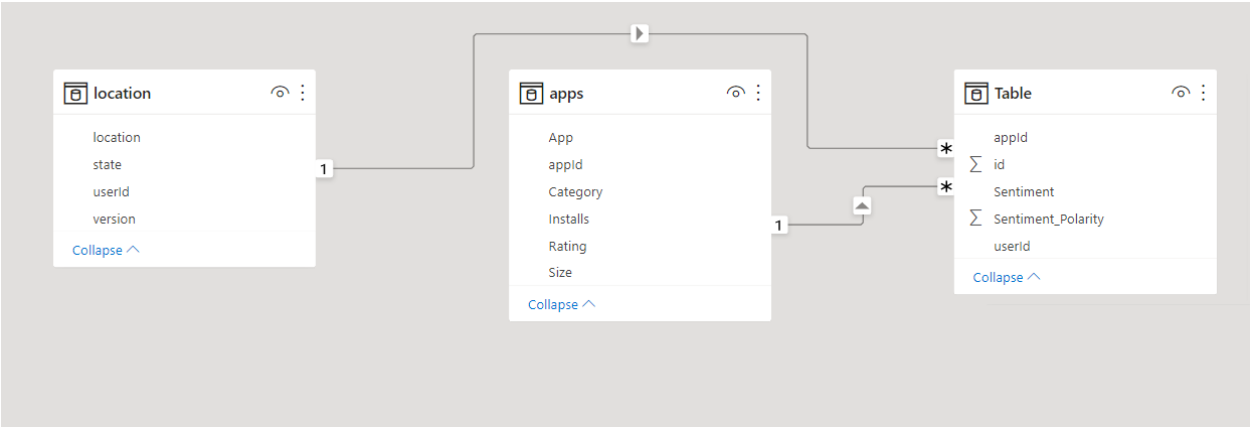
```
select * from apps;
```

App	Category	Rating	Size	Installs	id
Hotstar	ENTERTAINMENT	4.3	100M	100000000	0
AMC Theatres	ENTERTAINMENT	4.3	72M	1000000	1
8 Ball Pool	GAME	4.5	52M	100000000	2
Google	TOOLS	4.4	50M	1000000000	3
3D Live Neon Weed Launcher	PERSONALIZATION	4.5	21M	100000	4
A Word A Day	FAMILY	3.7	2.9M	500000	5

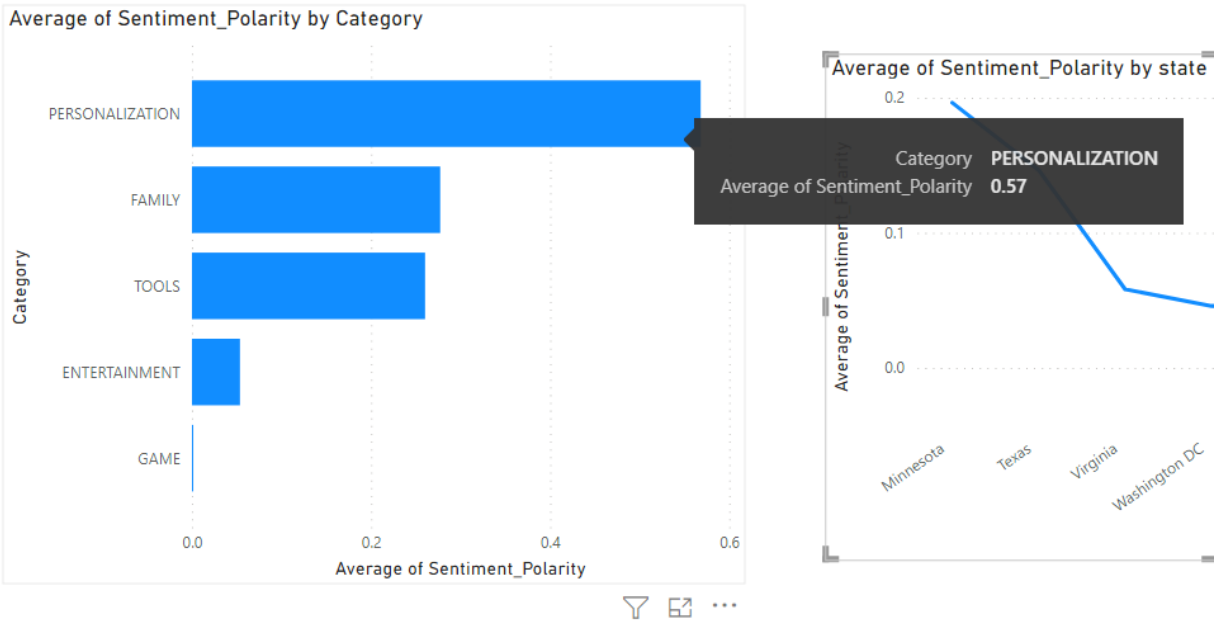
6 rows in set (0.00 sec)

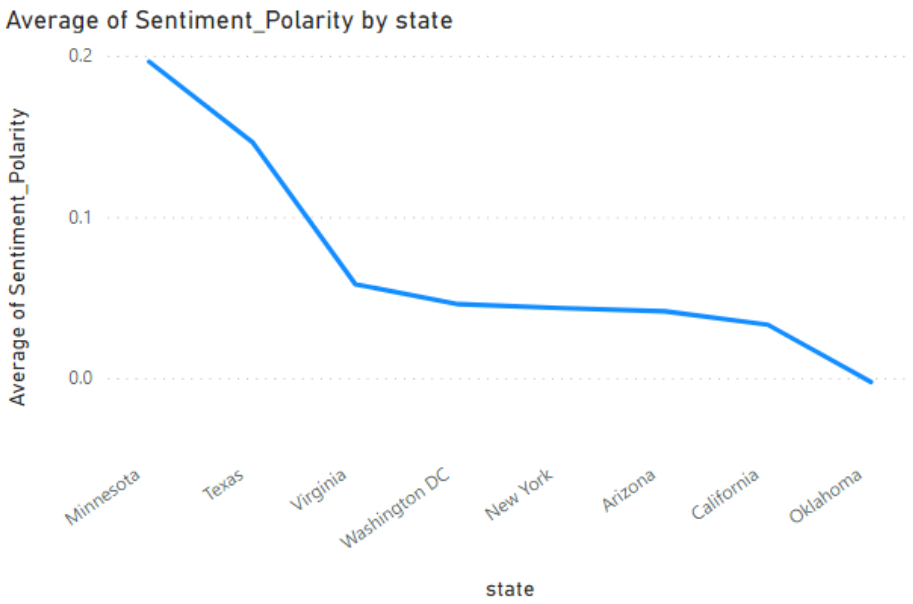
OLAP in power BI

Importing data and relational model

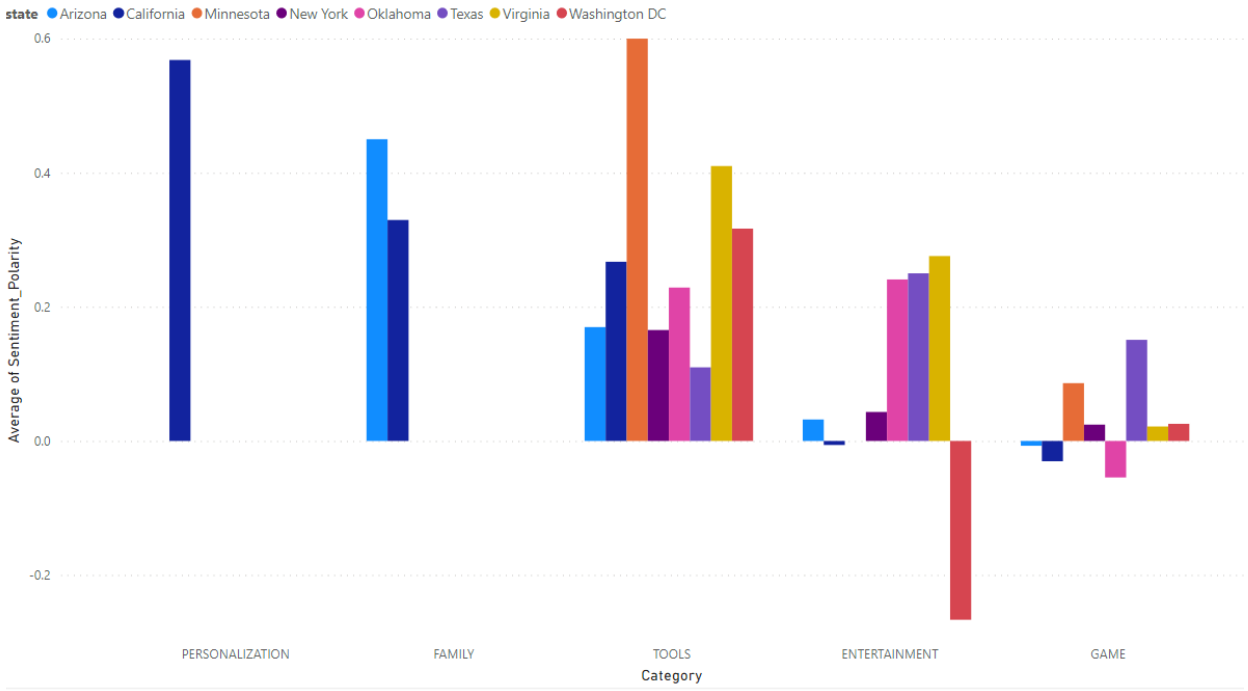


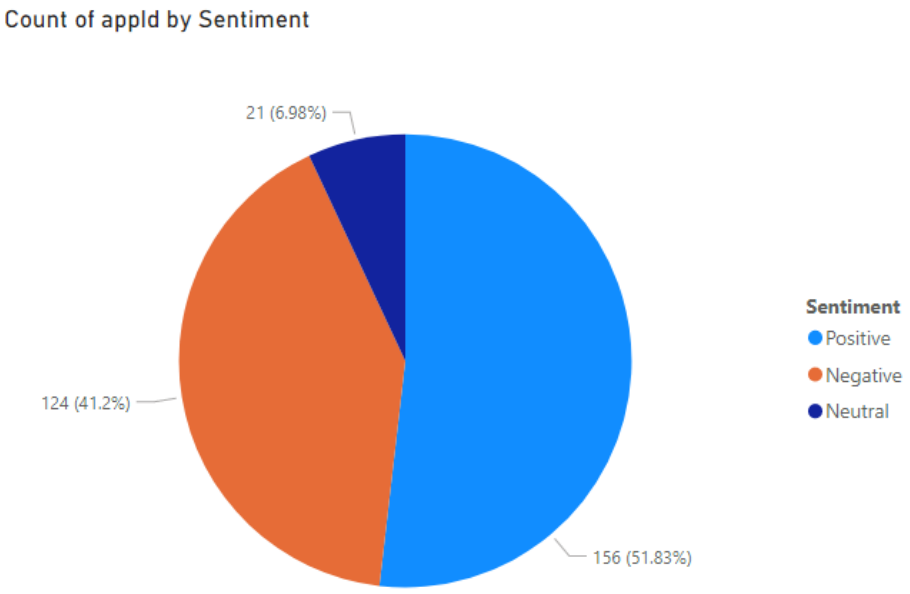
Creating plots



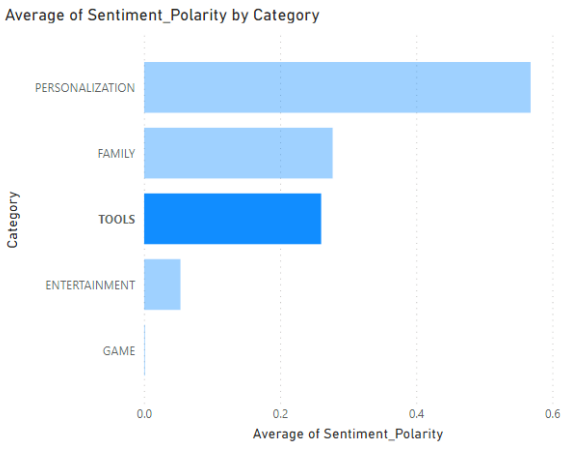


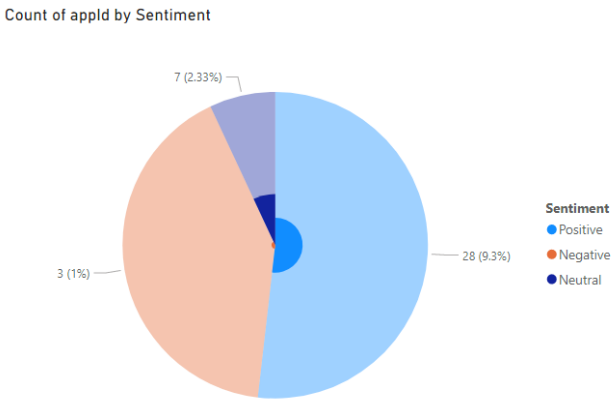
Average of sentiment polarity by category and state





Sliced out Tools





Slicing only negative sentiments