

Lab3 - CIS*6180/DATA*6300

Student ID – 1238693

Kusumitha Korrapati

1. Import data into Postgres SQL [1 point]

```
SQL Shell (psql)
Machine_Learning_IS=# SELECT * FROM "AdClicks";
time_stamp | tx_Id | user_session_Id | team_Id | user_Id | ad_Id | ad_Category
-----+-----+-----+-----+-----+-----+-----
(0 rows)

Machine_Learning_IS=# \COPY "AdClicks" ("time_stamp","tx_Id","user_session_Id","team_Id","user_Id","ad_Id","ad_Category") FROM 'C:\lab03_data\ad-clicks.csv' C
SV HEADER DELIMITER ',';
ERROR: column "team_id" of relation "AdClicks" does not exist
Machine_Learning_IS=# \COPY "AdClicks" ("time_stamp","tx_Id","user_session_Id","team_Id","user_Id","ad_Id","ad_Category") FROM 'C:\lab03_data\ad-clicks.csv'
CSV HEADER DELIMITER ',';
COPY 16323
Machine_Learning_IS=# SELECT * FROM "AdClicks";
time_stamp | tx_Id | user_session_Id | team_Id | user_Id | ad_Id | ad_Category
-----+-----+-----+-----+-----+-----+-----
2016-05-26 | 5974 | 5809 | 27 | 611 | 2 | electronics
2016-05-26 | 5976 | 5705 | 18 | 1874 | 21 | movies
2016-05-26 | 5978 | 5791 | 53 | 2139 | 25 | computers
2016-05-26 | 5973 | 5756 | 63 | 212 | 10 | fashion
2016-05-26 | 5980 | 5920 | 9 | 1027 | 20 | clothing
2016-05-26 | 5977 | 5954 | 77 | 595 | 4 | games
2016-05-26 | 5981 | 5674 | 54 | 770 | 6 | movies
2016-05-26 | 5975 | 5919 | 59 | 2133 | 3 | electronics
2016-05-26 | 5979 | 5945 | 75 | 253 | 3 | electronics
2016-05-26 | 5972 | 5914 | 78 | 1821 | 12 | computers
2016-05-26 | 6000 | 5955 | 64 | 2009 | 17 | games
2016-05-26 | 6001 | 5892 | 53 | 1538 | 21 | movies
2016-05-26 | 5999 | 5705 | 18 | 1874 | 3 | electronics
2016-05-26 | 6002 | 5702 | 20 | 355 | 29 | automotive
2016-05-26 | 6003 | 5920 | 9 | 1027 | 0 | sports
2016-05-26 | 5998 | 5892 | 53 | 1538 | 25 | computers
2016-05-26 | 6031 | 5938 | 11 | 1155 | 2 | electronics
2016-05-26 | 6037 | 5711 | 18 | 128 | 23 | fashion
2016-05-26 | 6040 | 5945 | 75 | 253 | 11 | games
2016-05-26 | 6035 | 5886 | 64 | 68 | 2 | electronics
2016-05-26 | 6038 | 5775 | 35 | 868 | 22 | computers
2016-05-26 | 6039 | 5917 | 51 | 2030 | 10 | fashion
2016-05-26 | 6034 | 5655 | 9 | 1277 | 1 | sports
2016-05-26 | 6032 | 5676 | 44 | 1926 | 14 | fashion
2016-05-26 | 6028 | 5886 | 64 | 68 | 21 | movies
```

```
SQL Shell (psql)
Machine_Learning_IS=# SELECT * FROM "BuyClicks";
timeStamp | tx_Id | user_session_Id | team | user_Id | buy_Id | price
-----+-----+-----+-----+-----+-----+-----
(0 rows)

Machine_Learning_IS=# \COPY "BuyClicks" ("timeStamp","tx_Id","user_session_I
d","team","user_Id","buy_Id","price") FROM 'C:\lab03_data\buy-clicks.csv' CS
V HEADER DELIMITER ',';
COPY 2947
Machine_Learning_IS=# SELECT * FROM "BuyClicks";
timeStamp | tx_Id | user_session_Id | team | user_Id | buy_Id | price
-----+-----+-----+-----+-----+-----+-----
2016-05-26 | 6004 | 5820 | 9 | 1300 | 2 | 3.0
2016-05-26 | 6005 | 5775 | 35 | 868 | 4 | 10.0
2016-05-26 | 6006 | 5679 | 97 | 819 | 5 | 20.0
2016-05-26 | 6067 | 5665 | 18 | 121 | 2 | 3.0
2016-05-26 | 6093 | 5709 | 11 | 2222 | 5 | 20.0
2016-05-26 | 6094 | 5798 | 77 | 1304 | 5 | 20.0
2016-05-26 | 6155 | 5920 | 9 | 1027 | 5 | 20.0
2016-05-26 | 6156 | 5697 | 35 | 2199 | 2 | 3.0
2016-05-26 | 6183 | 5893 | 64 | 1544 | 5 | 20.0
2016-05-26 | 6184 | 5697 | 35 | 2199 | 1 | 2.0
2016-05-26 | 6243 | 5659 | 13 | 1623 | 4 | 10.0
2016-05-26 | 6244 | 5920 | 9 | 1027 | 3 | 5.0
2016-05-26 | 6269 | 5785 | 27 | 1065 | 5 | 20.0
2016-05-26 | 6270 | 5661 | 63 | 83 | 3 | 5.0
2016-05-26 | 6271 | 5706 | 9 | 1652 | 0 | 1.0
2016-05-26 | 6292 | 5921 | 2 | 518 | 0 | 1.0
2016-05-26 | 6327 | 5860 | 57 | 2221 | 5 | 20.0
2016-05-26 | 6394 | 6088 | 32 | 1815 | 2 | 3.0
2016-05-26 | 6395 | 5880 | 35 | 2146 | 1 | 2.0
2016-05-26 | 6411 | 6230 | 77 | 1457 | 0 | 1.0
2016-05-26 | 6412 | 5870 | 9 | 371 | 2 | 3.0
2016-05-26 | 6458 | 5698 | 9 | 1143 | 1 | 2.0
2016-05-27 | 6473 | 5910 | 69 | 1162 | 1 | 2.0
2016-05-27 | 6474 | 5777 | 64 | 1567 | 3 | 5.0
2016-05-27 | 6549 | 5777 | 64 | 1567 | 2 | 3.0
2016-05-27 | 6550 | 5938 | 11 | 1155 | 5 | 20.0
```

```
SQL Shell (psql)
Machine_Learning_IS=# SELECT * FROM "GameClicks";
timestamp | click_Id | user_Id | user_session_Id | isHit | team_Id | team_Level
-----+-----+-----+-----+-----+-----+-----
(0 rows)

Machine_Learning_IS=# \COPY "GameClicks" ("timestamp","click_Id","user_Id","user_session_Id","isHit","team_Id","team_Level") FROM 'C:\lab03_data\game-clicks.csv' CSV HEADER DELIMITER ',';
COPY 755806
Machine_Learning_IS=# SELECT * FROM "GameClicks";
timestamp | click_Id | user_Id | user_session_Id | isHit | team_Id | team_Level
-----+-----+-----+-----+-----+-----+-----
2016-05-26 | 105 | 1038 | 5916 | 0 | 25 | 1
2016-05-26 | 154 | 1099 | 5898 | 0 | 44 | 1
2016-05-26 | 229 | 899 | 5757 | 0 | 71 | 1
2016-05-26 | 322 | 2197 | 5854 | 0 | 99 | 1
2016-05-26 | 22 | 1362 | 5739 | 0 | 13 | 1
2016-05-26 | 107 | 1071 | 5939 | 0 | 27 | 1
2016-05-26 | 289 | 2359 | 5764 | 0 | 85 | 1
2016-05-26 | 301 | 1243 | 5900 | 0 | 86 | 1
2016-05-26 | 274 | 1628 | 5896 | 0 | 82 | 1
2016-05-26 | 66 | 453 | 5662 | 0 | 20 | 1
2016-05-26 | 124 | 2336 | 5761 | 0 | 32 | 1
2016-05-26 | 204 | 1664 | 5831 | 0 | 69 | 1
2016-05-26 | 162 | 243 | 5943 | 0 | 54 | 1
2016-05-26 | 308 | 2333 | 5701 | 0 | 90 | 1
2016-05-26 | 160 | 181 | 5787 | 0 | 53 | 1
2016-05-26 | 183 | 212 | 5756 | 0 | 63 | 1
2016-05-26 | 314 | 1142 | 5871 | 0 | 93 | 1
2016-05-26 | 171 | 1085 | 5882 | 0 | 61 | 1
2016-05-26 | 67 | 1215 | 5905 | 0 | 20 | 1
2016-05-26 | 332 | 2306 | 5783 | 0 | 100 | 1
2016-05-26 | 239 | 489 | 5692 | 0 | 72 | 1
2016-05-26 | 205 | 1162 | 5910 | 0 | 69 | 1
2016-05-26 | 23 | 1072 | 5746 | 0 | 13 | 1
2016-05-26 | 0 | 624 | 5833 | 0 | 2 | 1
2016-05-26 | 52 | 217 | 5726 | 0 | 18 | 1
2016-05-26 | 261 | 1191 | 5719 | 0 | 77 | 1
2016-05-26 | 92 | 1349 | 5755 | 0 | 22 | 1
2016-05-26 | 227 | 1665 | 5953 | 0 | 70 | 1
```

2. For each table write a query to return first five rows of timestamp and userid.

```
Machine_Learning_IS=# SELECT "time_stamp","user_Id" FROM "AdClicks" LIMIT 5;
time_stamp | user_Id
-----+-----
2016-05-26 | 611
2016-05-26 | 1874
2016-05-26 | 2139
2016-05-26 | 212
2016-05-26 | 1027
(5 rows)

Machine_Learning_IS=# SELECT "timeStamp","user_Id" FROM "BuyClicks" LIMIT 5;
timeStamp | user_Id
-----+-----
2016-05-26 | 1300
2016-05-26 | 868
2016-05-26 | 819
2016-05-26 | 121
2016-05-26 | 2222
(5 rows)

Machine_Learning_IS=# SELECT "timeStamp","user_Id" FROM "GameClicks" LIMIT 5;
timeStamp | user_Id
-----+-----
2016-05-26 | 1038
2016-05-26 | 1099
2016-05-26 | 899
2016-05-26 | 2197
2016-05-26 | 1362
(5 rows)

Machine_Learning_IS=#
```

3. What is the maximum, minimum and average level a team has reached in game clicks (1 point)?

```
Machine_Learning_IS=# select max("team_Level") as "MaximumLevel",min("team_Level") as "MinimumLevel",avg("team_Level") as "AvergaeLevel" from "GameClicks"
Machine_Learning_IS=# select max("team_Level") as "MaximumLevel",min("team_Level") as "MinimumLevel",avg("team_Level") as "AvergaeLevel" from "GameClicks";
ERROR: syntax error at or near "select"
LINE 2: select max("team_Level") as "MaximumLevel",min("team_Level")...
^

Machine_Learning_IS=# SELECT max("team_Level") as "MaximumLevel",min("team_Level") as "MinimumLevel",avg("team_Level") as "AvergaeLevel" FROM "GameClicks";
MaximumLevel | MinimumLevel | AvergaeLevel
-----+-----+-----
            8 |             1 | 4.5617327726956388
(1 row)

Machine_Learning_IS=#
```

4. How many user-id's (both distinct and with repetition) were on the lowest level as found in the previous question? (1 point)

```
Machine_Learning_IS=# select count("user_Id") as Cost_Low_Level_Users from "GameClicks" where "team_Level"=(select min("team_Level") from "GameClicks");
cost_low_level_users
-----
        67271
(1 row)

Machine_Learning_IS=#
```

```
Machine_Learning_IS=# SELECT COUNT(DISTINCT("user_Id")) as "distinct_Users",COUNT("user_Id") as "Repetition_Count" FROM "GameClicks" WHERE "team_Level" = (
SELECT MIN("team_Level") FROM "GameClicks");
distinct_Users | Repetition_Count
-----+-----
          467 |          67271
(1 row)

Machine_Learning_IS=#
```

5. How many user-id's (both distinct and with repetition) were on the lowest level in game-clicks and clicked the highest costing price in buy-clicks? (1 point)

6. How many user-id's (both distinct and with repetition) were on the highest level in game-clicks and clicked the highest costing price in buy-clicks? (1 point)

```
Machine_Learning_IS=# SELECT COUNT(*) as "total_Users",COUNT(DISTINCT("GameClicks"."user_Id")) as "distinct_Users" FROM "GameClicks" INNER JOIN "BuyClicks"
ON "GameClicks"."user_Id" = "BuyClicks"."user_Id" WHERE "GameClicks"."team_Level"=(SELECT MIN("team_Level") FROM "GameClicks") AND "BuyClicks"."price" = (SE
LECT MAX("price") FROM "BuyClicks");
total_Users | distinct_Users
-----+-----
        36591 |          112
(1 row)

Machine_Learning_IS=# SELECT COUNT(*) as "total_Users",COUNT(DISTINCT("GameClicks"."user_Id")) as "distinct_Users" FROM "GameClicks" INNER JOIN "BuyClicks"
ON "GameClicks"."user_Id" = "BuyClicks"."user_Id" WHERE "GameClicks"."team_Level"=(SELECT MAX("team_Level") FROM "GameClicks") AND "BuyClicks"."price" = (SE
LECT MAX("price") FROM "BuyClicks");
total_Users | distinct_Users
-----+-----
        32747 |          190
(1 row)

Machine_Learning_IS=# s
```

7. Import data into MongoDB (1 point)

```
Command Prompt
Microsoft Windows [Version 10.0.22621.1413]
(c) Microsoft Corporation. All rights reserved.

C:\Users\16478>cd C:\Program Files\MongoDB\Server\5.0\bin

C:\Program Files\MongoDB\Server\5.0\bin>show dbs
'show' is not recognized as an internal or external command,
operable program or batch file.

C:\Program Files\MongoDB\Server\5.0\bin>mongoimport --jsonArray --db twitter
db --collection twitter --file "C:\lab03_data\twitter_users.json"
2023-03-16T15:19:28.425-0400    connected to: mongodb://localhost/
2023-03-16T15:19:28.731-0400    11188 document(s) imported successfully. 0 d
ocument(s) failed to import.

C:\Program Files\MongoDB\Server\5.0\bin>
```

8. How many distinct users are in this dataset? (1 point)

```
ObjectId("64136bc0d4d6d01c4fe3f381"),
ObjectId("64136bc0d4d6d01c4fe3f382"),
ObjectId("64136bc0d4d6d01c4fe3f383"),
ObjectId("64136bc0d4d6d01c4fe3f384"),
ObjectId("64136bc0d4d6d01c4fe3f385"),
ObjectId("64136bc0d4d6d01c4fe3f386"),
ObjectId("64136bc0d4d6d01c4fe3f387"),
ObjectId("64136bc0d4d6d01c4fe3f388")
]
> db.twitter.distinct("_id").length
11188
>
```

9. What is the username of the twitter account who has a tweet_followers_count of exactly 1241159? (1 point)

```
> db.twitter.find({"tweet_followers_count":1241159});
{ "_id" : ObjectId("64136bc0d4d6d01c4fe3ca9b"), "user_name" : "midoahm", "retweet_count" : 0, "tweet_followers_count" : 1241159, "source" : "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\">Twitter for iPhone</a>", "coordinates" : null, "tweet_mentioned_count" : 1, "tweet_ID" : "757637082596700161", "tweet_text" : "RT @TSBible: Throwback to FIFA 94 when you could run away from the ref when he tried to book you. 🤖 https://t.co/Bxyq3vrLNH", "user" : { "CreatedAt" : NumberLong("1278165896000"), "FavouritesCount" : 181, "FollowersCount" : 1241159, "FriendsCount" : 500, "UserId" : 162385414, "Location" : "Doha, Qatar" } }
> db.twitter.find({"tweet_followers_count":1241159,{"user_name":1,"_id":0});
{ "user_name" : "midoahm" }
>
```

10. How many tweets contain the word "FIFA"? (1 point)

```
" : 5084, "FollowersCount" : 341, "FriendsCount" : 212, "UserId" : 363162955, "Location" : "Beserah" } }
{ "_id" : ObjectId("64136bc0d4d6d01c4fe3c7ee"), "user_name" : "jjswest", "retweet_count" : 0, "tweet_followers_count" : 237, "source" : "<a href='\"http://twitter.com/download/android\"' rel='\"nofollow\"'>Twitter for Android</a>", "coordinates" : null, "tweet_mentioned_count" : 1, "tweet_ID" : "755892045311569920", "tweet_text" : "RT @RobHarris: IOC withdrawing support for sporting events in Russia sharpens scrutiny on FIFA staging 2018 World Cup in Russia https://t... ", "user" : { "CreatedAt" : NumberLong("1377531962000"), "FavouritesCount" : 2149, "FollowersCount" : 237, "FriendsCount" : 210, "UserId" : 1702284013, "Location" : "Cape Town, SOUTH AFRICA" } }
Type "it" for more
> db.twitter.find({"tweet_text" : {$regex : "FIFA"}}).length
function() {
  return this.toArray().length;
}
> db.twitter.find({"tweet_text" : {$regex : "FIFA"}}).length()
3697
>
```

11. How many tweets mention the word FIFA and have the tweet_mentioned_count greater than 5? (1 point)

```
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":5}}).count()
0
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":1}}).count()
402
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":2}}).count()
76
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":3}}).count()
6
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":4}}).count()
1
twitterdb> db.twitter.find({"tweet_text":{"$regex":"FIFA"},"tweet_mentioned_count":{"$gt":5}}).count()
0
twitterdb>
```

12. What location has the most distinct users talking about FIFA in this dataset? (1 point)

```
twitterdb> db.twitter.aggregate([{$match: {"tweet_text" : {$regex : "FIFA"}}},{$match: {"user.Location": { $ne: null } }},{$group: {_id: "$user.Location", count: { $sum: 1 } }},{$sort: { count: -1 } }])
[
  { _id: 'Nigeria', count: 81 },
  { _id: 'United States', count: 37 },
  { _id: 'Lagos, Nigeria', count: 36 },
  { _id: 'Lagos', count: 25 },
  { _id: 'UK', count: 23 },
  { _id: 'London', count: 23 },
  { _id: 'London, England', count: 22 },
  { _id: 'Abuja, Nigeria', count: 15 },
  { _id: 'England', count: 15 },
  { _id: 'Earth', count: 13 },
  { _id: 'Manchester', count: 10 },
  { _id: 'France', count: 9 },
  { _id: 'Manchester, England', count: 9 },
  { _id: 'Ireland', count: 8 },
  { _id: 'United Kingdom', count: 8 },
  { _id: 'Abuja', count: 8 },
  { _id: 'Bournemouth', count: 8 },
  { _id: 'Liverpool', count: 8 },
  { _id: 'Dublin', count: 7 },
  { _id: 'England, United Kingdom', count: 7 }
]
Type "it" for more
twitterdb> db.twitter.aggregate([{$match: {"tweet_text" : {$regex : "FIFA"}}},{$match: {"user.Location": { $ne: null } }},{$group: {_id: "$user.Location", count: { $sum: 1 } }},{$sort: { count: -1 } },{$limit:1}])
[ { _id: 'Nigeria', count: 81 } ]
twitterdb> s
```

