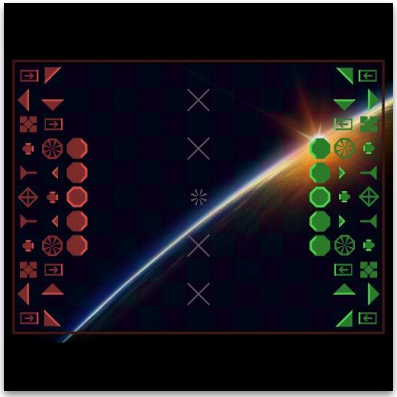

# Apple App Store Game Icons linked to User Rating

Team 7
Leona Yiu & Xiaoxu Zhang

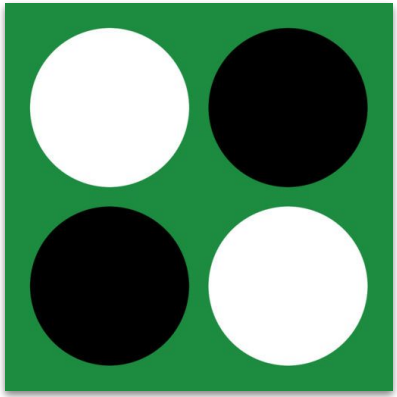
# Objective

Can we predict the user rating of a game from the Apple App Store based on the appearance of the game's icon (512 x 512 pixels)?



**Laser Chess**
Average User Rating: 1.5
User Rating Count: 200


**Tank Ace 1944 HD Lite**
Average User Rating: 2
User Rating Count: 697

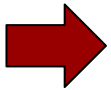


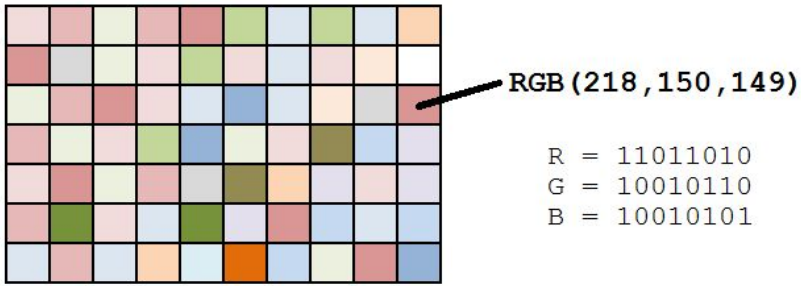
**Morocco**
Average User Rating: 3
User Rating Count: 8,376


**Kingdom Rush HD**
Average User Rating: 5
User Rating Count: 36,212

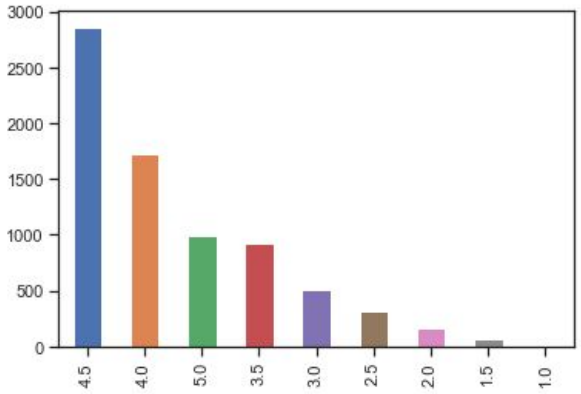
# Data Preparation

- Data set from Kaggle consists of 17,007 strategy games from Apple App Store collected on August 3rd, 2019
- Each game icon jpg (taken from Icon URL) converted into 512 x 512 strings of color code (RGB format)
- Based on color codes in icon, 3 features created:
  - hpercent_c_cnt: count of colors taking >10% of full image
  - lpercent_c_cnt: count of colors taking <0.1% of the full image
  - max_c_cnt: count of pixels of the color that appears the most in the icon image



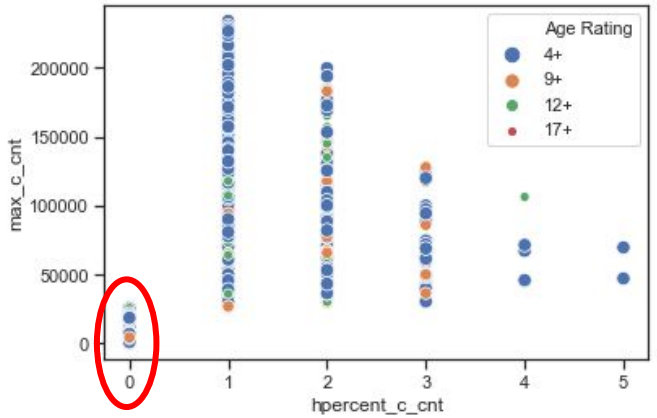


# Data Exploration




Prelim observation - data set consists heavily of highly rated games (potentially impacting model)

| hpercent_ c_cnt | Average User Rating | Count |
|---|---|---|
| **0** | 4.0 | 676 |
| | 4.5 | 1,435 |
| | 5.0 | 206 |
| **1** | 4.0 | 153 |
| | 4.5 | 277 |
| | 5.0 | 55 |
| **2** | 4.0 | 37 |
| | 4.5 | 91 |
| | 5.0 | 41 |




Majority of high rated games have icon images with 0 color taking >10% of full image (hpercent_c_cnt) and <27K pixels of color that appears most in icon (max_c_cnt)

# Data Modeling - Random Forest Classifier

```
X = df[['hpercent_c_cnt','max_c_cnt']] # feature variables tried: hpercent_c_cnt & max_c_cnt ; or aggre
y = df['Great_App']

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)

rf = RandomForestClassifier(n_estimators=100)

rf.fit(X_train, y_train)

# Predictions: Apply classifier to test data
rf_predict = rf.predict(X_test)
```

Accuracy: 0.63199647

```
X = df[['hpercent_is_0','max_c_cnt_is_less_27000']] # feature variables tried: hpercent_c_cnt & max_c_c
y = df['Great_App']

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)

rf2 = RandomForestClassifier(n_estimators=100)

rf2.fit(X_train, y_train)

# Predictions: Apply classifier to test data
rf2_predict = rf2.predict(X_test)
```

Accuracy (after fine tuning): 0.73336271

# Data Modeling - Random Forest Classifier

```
=== Confusion Matrix ===
[[   0  605]
 [   0 1664]]


=== Classification Report ===
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       605
           1       0.73      1.00      0.85      1664

   micro avg       0.73      0.73      0.73      2269
   macro avg       0.37      0.50      0.42      2269
weighted avg       0.54      0.73      0.62      2269



=== All AUC Scores ===
[0.436497   0.45195962 0.4696196  0.45510767 0.50147055 0.50494844
 0.51503478 0.46703445 0.53500898 0.48154344]


=== Mean AUC Score ===
Mean AUC Score - Random Forest:  0.4818224538568163
```

Model accurately predicted 73% cases that game is a Great App (rating 4+), 27% incorrectly predicted

73% precision classifying game is a Great App based on feature variables hpercent_is_0 & max_c_cnt_is_less70000; 100% recall

Mean AUC Score of 0.48, however, indicating mediocre classifier to distinguish between highly rated games vs. lower rated games

# Results - True Positives : Game icons correctly classified as highly rated and actual rating is 4+

**Logic Puzzles+**
Average User Rating: 5
User Rating Count: 50
Hpercent_c_cnt: 0
max_c_cnt: 17,388

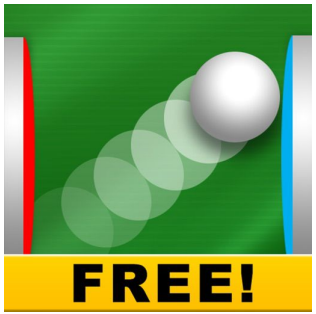

**Land Air Sea Warfare HD RTS**
Average User Rating: 4
User Rating Count: 660
Hpercent_c_cnt: 0
max_c_cnt: 518

**Solitaire 2017 HD**
Average User Rating: 4
User Rating Count: 241
Hpercent_c_cnt: 0
max_c_cnt: 9,790

**Age of Z**
Average User Rating: 4.5
User Rating Count: 3,553
Hpercent_c_cnt: 0
max_c_cnt: 814

# Results - False Positives : Game icons incorrectly classified as highly rated but actual rating is <4

**Sport Car Driving : City Adventure**
Average User Rating: 3
User Rating Count: 97
Hpercent_c_cnt: 0
max_c_cnt: 1,707




**Free Ping Pong Table Tennis**
Average User Rating: 2.5
User Rating Count: 660
Hpercent_c_cnt: 0
max_c_cnt: 2,678

**Tools for MTG**
Average User Rating: 3.5
User Rating Count: 13
Hpercent_c_cnt: 0
max_c_cnt: 24,240

**Stick Castles War**
Average User Rating: 3.5
User Rating Count: 71
Hpercent_c_cnt: 0
max_c_cnt: 40,396