

# How Many Videogames Will You Sell?

Sarah Floris




## Things to Note

1. Tweets are not differentiated by consoles.
2. The maximum number of tweets is 500 per week.
3. Weekly sales only provides the top 30 videogames.
4. Total sales includes all videogames provided by the VGChartz database.
5. Europe includes United Kingdom, Germany , and France
6. Global population is the population of United States, Japan, and Europe.

# Obtaining Data: VGChartz

Right: Online Database featuring, game, console, publisher, VGChartz score, critic score, user score, total sales, release date, and last update



Home

Charts


Tools

Game DB

Articles

Forum

Search Game DB



Weekly Chart Index

All Weekly Software Charts Available on VGChartz

28th Apr 2018	Global	USA	Europe	UK	Germany	France	Japan
21st Apr 2018	Global	USA	Europe	UK	Germany	France	Japan
14th Apr 2018	Global	USA	Europe	UK	Germany	France	Japan
07th Apr 2018	Global	USA	Europe	UK	Germany	France	Japan
31st Mar 2018	Global	USA	Europe	UK	Germany	France	Japan
24th Mar 2018	Global	USA	Europe	UK	Germany	France	Japan
17th Mar 2018	Global	USA	Europe	UK	Germany	France	Japan
10th Mar 2018	Global	USA	Europe	UK	Germany	France	Japan
03rd Mar 2018	Global	USA	Europe	UK	Germany	France	Japan
24th Feb 2018	Global	USA	Europe	UK	Germany	France	Japan
17th Feb 2018	Global	USA	Europe	UK	Germany	France	Japan
10th Feb 2018	Global	USA	Europe	UK	Germany	France	Japan
03rd Feb 2018	Global	USA	Europe	UK	Germany	France	Japan
27th Jan 2018	Global	USA	Europe	UK	Germany	France	Japan
20th Jan 2018	Global	USA	Europe	UK	Germany	France	Japan
13th Jan 2018	Global	USA	Europe	UK	Germany	France	Japan
06th Jan 2018	Global	USA	Europe	UK	Germany	France	Japan
30th Dec 2017	Global	USA	Europe	UK	Germany	France	Japan
23rd Dec 2017	Global	USA	Europe	UK	Germany	France	Japan
16th Dec 2017	Global	USA	Europe	UK	Germany	France	Japan
09th Dec 2017	Global	USA	Europe	UK	Germany	France	Japan
02nd Dec 2017	Global	USA	Europe	UK	Germany	France	Japan
25th Nov 2017	Global	USA	Europe	UK	Germany	France	Japan
18th Nov 2017	Global	USA	Europe	UK	Germany	France	Japan
11th Nov 2017	Global	USA	Europe	UK	Germany	France	Japan
04th Nov 2017	Global	USA	Europe	UK	Germany	France	Japan

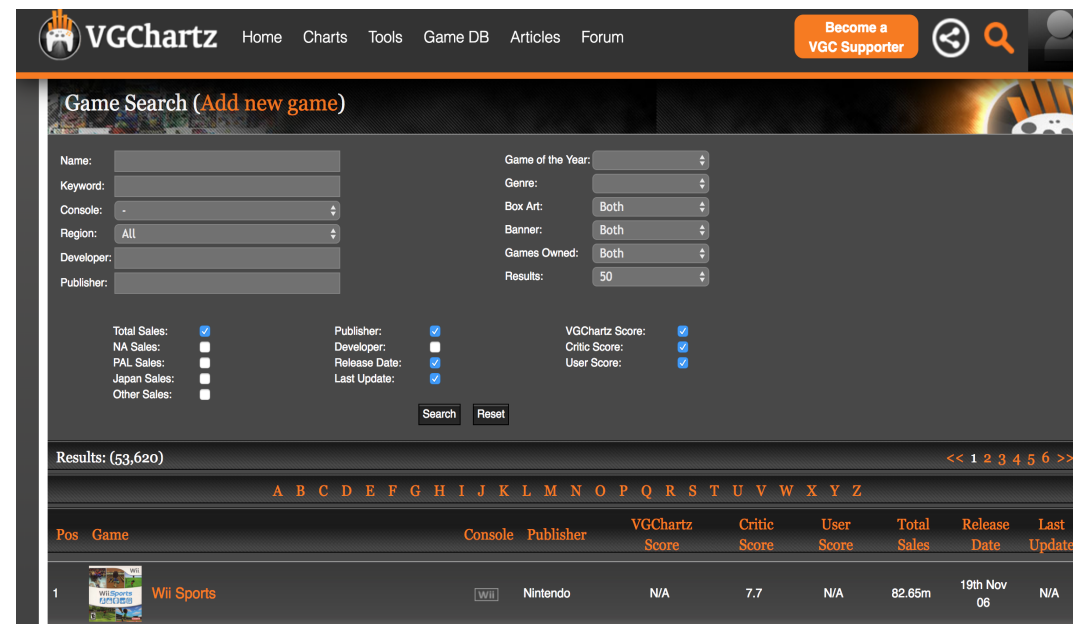
Global Top 10 28th April 2018

1	PS4	God of War (PS4)	780,388
2	PS4	Far Cry 5	144,754
3	PS4	Jikkyou Powerful Pro Baseball 2018	130,319
4	NS	Nintendo Labo: Toy-Con 01 Variety Kit	111,829
5	XONE	Far Cry 5	61,743
6	PSV	Jikkyou Powerful Pro Baseball 2018	57,196
7	NS	Mario Kart 8 Deluxe	56,860
8	NS	Splatoon 2	45,882
9	NS	The Legend of Zelda: Breath of the Wild	42,933
10	NS	Super Mario Odyssey	42,508

Top 10, ranked by number of units sold.

Most Popular Games

1	Hi Words - Crush Letters
2	Grand Theft Auto V
3	4 Pics 1 Word
4	Minecraft
5	Dynasty Warriors 8
6	Grand Theft Auto: San Andreas
7	Dark Souls
8	Woody Puzzle - Wood Block Jigsaw 10x10 Board Game
9	Bullet Hell Monday
10	Borderlands 2
11	Persona Q: Shadow of the Labyrinth
12	Smartphone Tycoon
13	Galaxy Attack: Alien Shooter
14	Ni no Kuni: Wrath of the White Witch
15	The Elder Scrolls V: Skyrim
16	Sniper 3D Assassin: Shoot to Kill
17	The Sims 3



The screenshot shows the VGChartz website's 'Game Search' page. It features a navigation bar with links to Home, Charts, Tools, Game DB, Articles, and Forum, along with a search bar and a 'Become a VGC Supporter' button. The main content area is a search form with various filters and a search button. Below the form, there is a 'Results: (53,620)' section with a pagination link '<< 1 2 3 4 5 6 >>'. The results are displayed in a table with columns: Pos, Game, Console, Publisher, VGChartz Score, Critic Score, User Score, Total Sales, Release Date, and Last Update. The first result is 'Wii Sports' by Nintendo, with a VGChartz Score of N/A, a Critic Score of 7.7, a User Score of N/A, and Total Sales of 82.65m.

Pos	Game	Console	Publisher	VGChartz Score	Critic Score	User Score	Total Sales	Release Date	Last Update
1	Wii Sports	Wii	Nintendo	N/A	7.7	N/A	82.65m	19th Nov 06	N/A

Left : Photo of the weekly sales database, listing weekly sales for global, United States, Europe, Germany , France, and Japan population.

# Obtaining Data: Twitter

Main problems with obtaining large set of tweets:

1. RESTful API only does historical tweets for the last 7 days
2. Only lets us query a limited amount (up to 1000) for free
3. Access to historical data if you want to pay for Enterprise

How do we solve these problems?

1. Set a tweet's class attributes to formulate an url (such as video game title, date ranges, etc.)
2. Use this url to web scrape the tweets that contain the videogame title both in the body and in the hashtag
3. Keep searching until max number of tweets has been reached



# Obtaining Data: Twitter

Web scraped the tweets that contain the videogame hashtag from week to week after searching

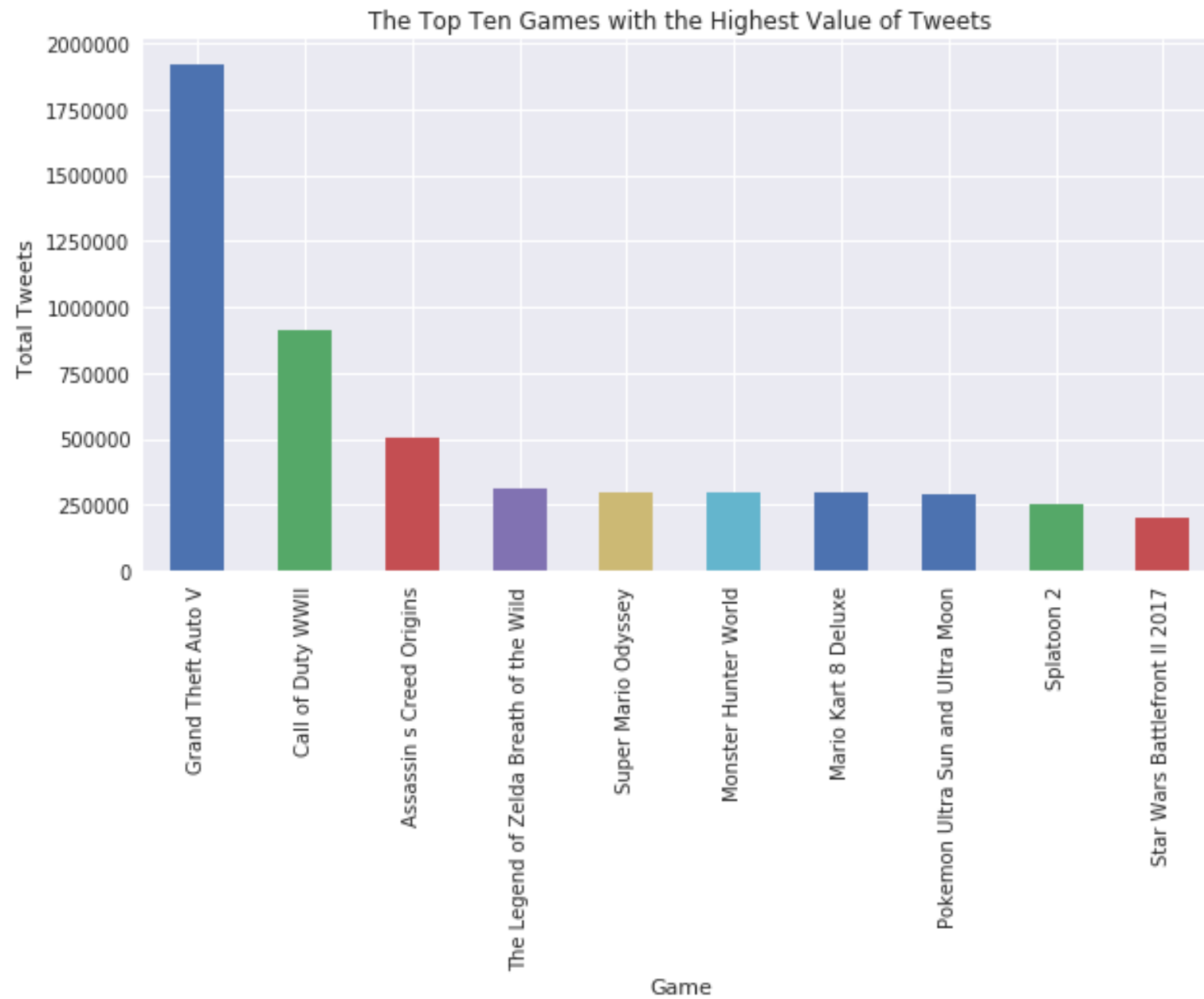
## Example

```
def finding_tweets(videogame):  
    """Description"""  
    name = []  
    start_date = []  
    tweet_date = []  
    tweet_text = []  
    assert isinstance(videogame[0],str)  
    assert isinstance(videogame[1],str)  
    assert isinstance(videogame[2],str)  
    try:  
        tweetCriteria = TweetCriteria().setQuerySearch(videogame[0]) \  
.setSince(videogame[1]).setUntil(videogame[2]).setMaxTweets(500)  
        tweets = TweetManager.getTweets(tweetCriteria)  
        for tweet in tweets:  
            tweet_date.append(tweet.date)  
            tweet_text.append(tweet.text)  
    except:  
        print(videogame[0],videogame[1],videogame[2], 'empty')  
    df = pd.DataFrame(np.column_stack((tweet_date,tweet_text)))  
    df['Name']= videogame[0]  
    df['start_date'] = videogame[1]  
    df['end_date'] = videogame[2]  
    return df
```

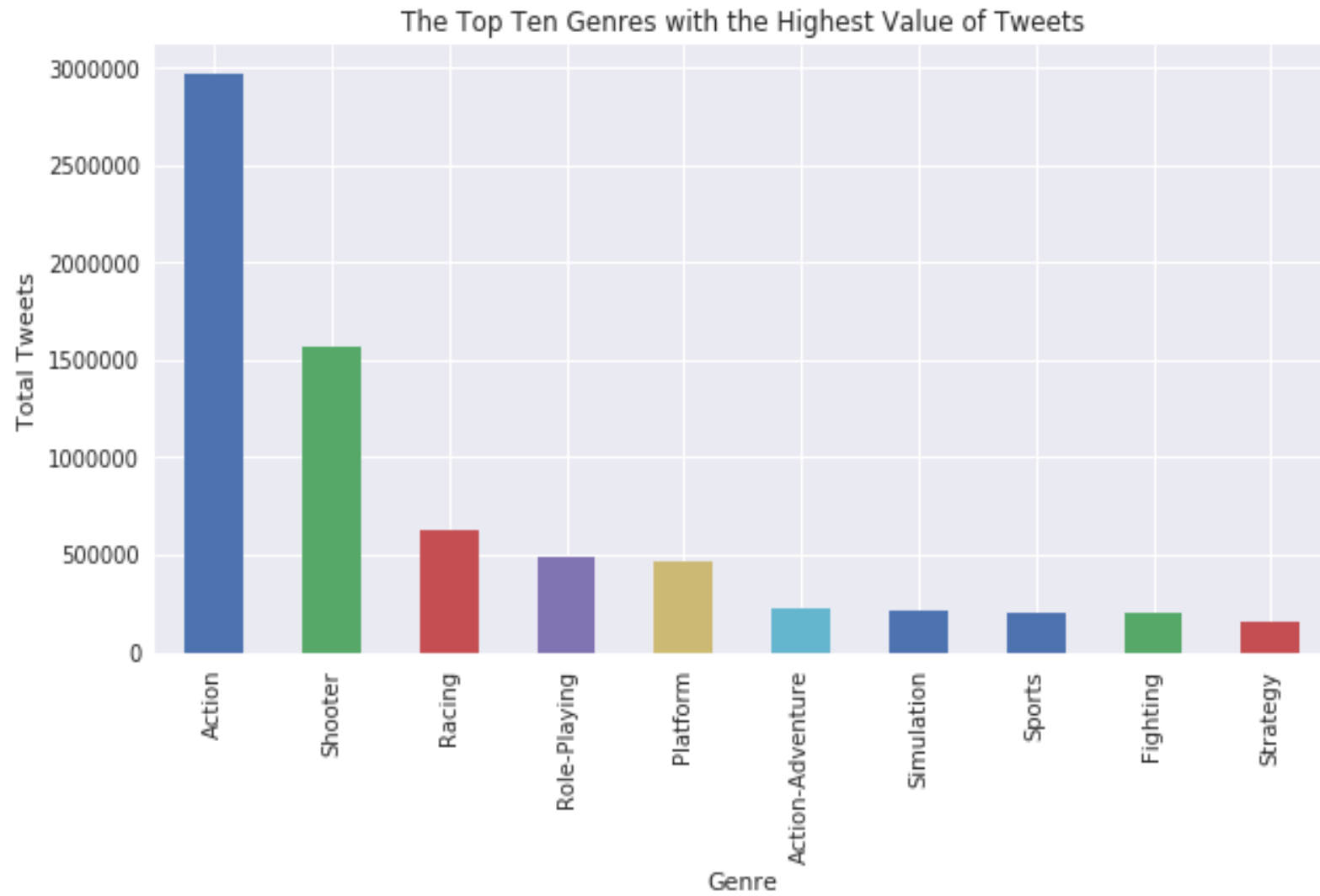
Interested in seeing all of the code? Look at my github!

Let's take a look at the data

# Top Ten Games



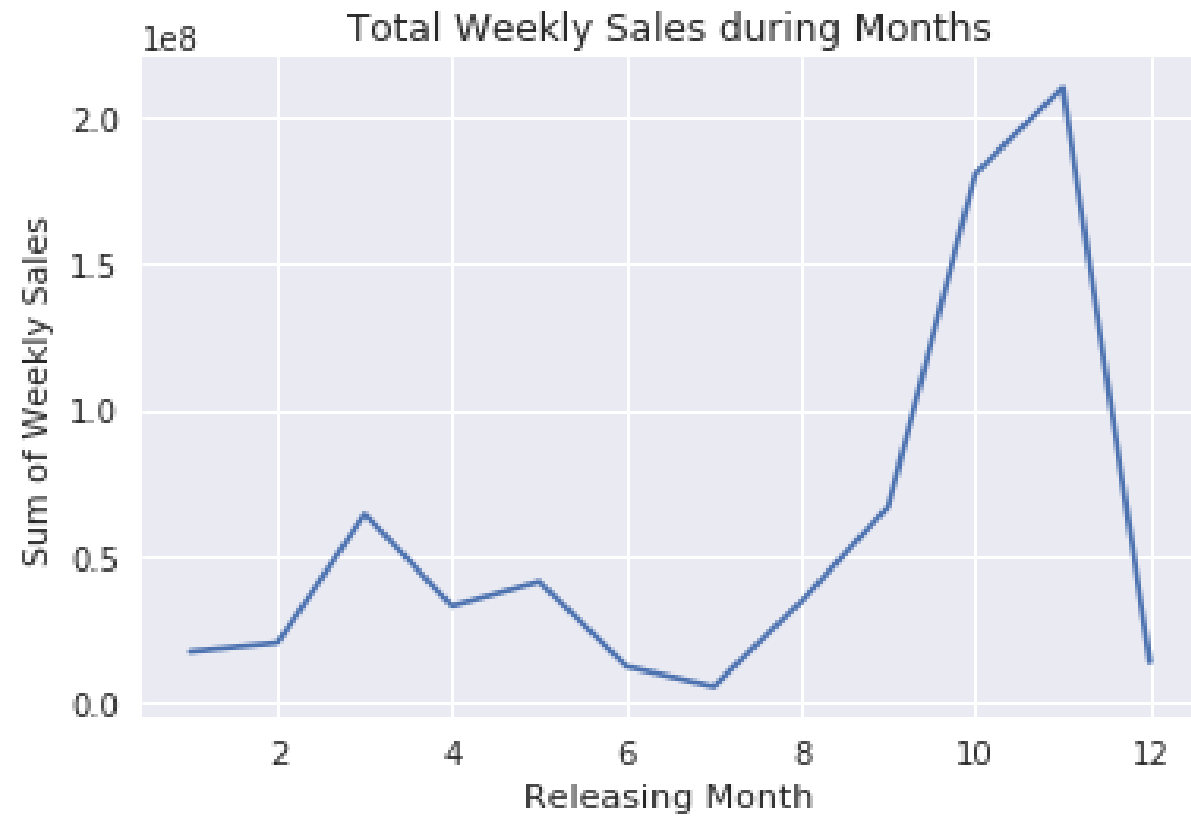
## Top Ten Genres



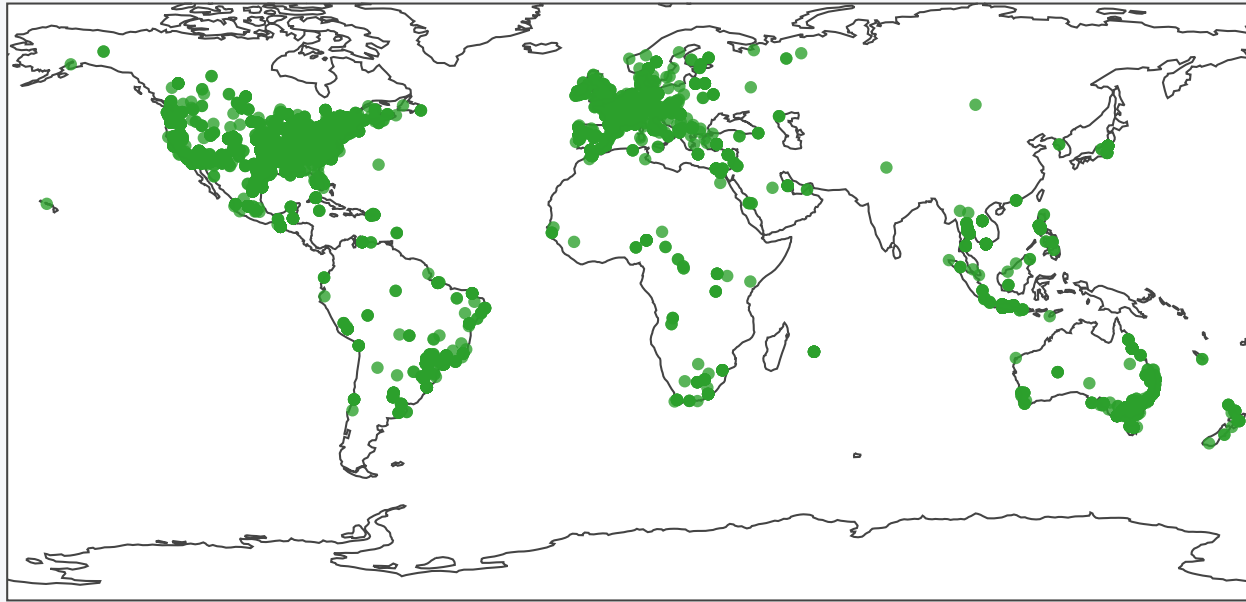


## Monthly Game Unit Sales and Total Sales

Month	Number of Games
1	519
2	1138
3	2525
4	1491
5	1554
6	730
7	343
8	1134
9	1596
10	4748
11	5223
12	892



## Videogame Publisher's Location



Predicting the next weeks sales

# Machine Learning Methods

- Linear Regression - 0.198
- Elastic Net - 0.160
- Random Forest Regression - 0.035
- GradientBoostingregressor - 0.187
- AdaBoostRegressor with RandomForest Regression as initial estimator - 0.384

The best machine learning method was AdaBoostRegressor with a RandomForestRegressor as a base estimator.  
Test set accuracy score for best params: 0.375

```
AdaBoostRegressor(base_estimator=RandomForestRegressor(bootstrap=True, criterion='mse',  
max_depth=None, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0,  
min_impurity_split=None, min_samples_leaf=1, min_samples_split=2,  
min_weight_fraction_leaf=0.0, n_estimators=9, n_jobs=-1, oob_score=False,  
random_state=None, verbose=0, warm_start=False), learning_rate=1.0,  
loss='linear', n_estimators=50, random_state=None)
```

begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction.

# Predictions for the Top 30 Games Sold

