

Mid-Project Review

Elliot Gardner

MSIA-423

Highlights

- Completed Data Pipeline for Eventbrite API
- Ingesting data roughly every 12 hours since mid-April
- Approximately 5% of events sell out prior to the start data in the data so far

Review Progress

- Data Gathering: API Integration completed
- Data Gathering: Data Capture completed
- Data Gathering: Data Formatting completed
- Web App Construction: Database Hosting partially completed

Demo

```
MINGW64:/c/Users/Elliott/Documents/Big Data/NW/MSiA/MSiA-423 Analytics Value Chain/sell_out_project
(sell_out_project)
Elliott@DESKTOP-SUT1VBG MINGW64 ~/Documents/Big Data/NW/MSiA/MSiA-423 Analytics Value Chain/sell_out_project (development)
$ python src/ingest_data.py --config config/config.yml --API_token 3IS3J3FEVFPVJ7D2AW6
2019-05-15 23:26:25,243 ingest_data_log INFO Setting API Request Headers
2019-05-15 23:26:25,243 ingest_data_log INFO Retrieving Page 1...
2019-05-15 23:26:28,188 root INFO Uploading 23_26_25_1.json to emg8426.msia423.project
2019-05-15 23:26:29,367 ingest_data_log INFO JSON uploaded as "b1107d78be59134b6f22bb50c6e1dd5a"
2019-05-15 23:26:29,367 root INFO Uploading 23_26_25_1.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:29,430 ingest_data_log INFO File 23_26_25_1.json written
2019-05-15 23:26:29,430 ingest_data_log INFO Retrieving Page 2...
2019-05-15 23:26:31,871 root INFO Uploading 23_26_25_2.json to emg8426.msia423.project
2019-05-15 23:26:32,716 ingest_data_log INFO JSON uploaded as "9880b7d47690362a32b2e4e297c48fad"
2019-05-15 23:26:32,716 root INFO Uploading 23_26_25_2.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:32,778 ingest_data_log INFO File 23_26_25_2.json written
2019-05-15 23:26:32,778 ingest_data_log INFO Retrieving Page 3...
2019-05-15 23:26:35,540 root INFO Uploading 23_26_25_3.json to emg8426.msia423.project
2019-05-15 23:26:36,307 ingest_data_log INFO JSON uploaded as "770f017b614559f33ede3b593f22dad5"
2019-05-15 23:26:36,307 root INFO Uploading 23_26_25_3.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:36,391 ingest_data_log INFO File 23_26_25_3.json written
2019-05-15 23:26:36,391 ingest_data_log INFO Retrieving Page 4...
2019-05-15 23:26:39,004 root INFO Uploading 23_26_25_4.json to emg8426.msia423.project
2019-05-15 23:26:39,806 ingest_data_log INFO JSON uploaded as "c7d69d5aae01c1ff708c911aeada3ba4"
2019-05-15 23:26:39,806 root INFO Uploading 23_26_25_4.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:39,886 ingest_data_log INFO File 23_26_25_4.json written
2019-05-15 23:26:39,889 ingest_data_log INFO Retrieving Page 5...
2019-05-15 23:26:42,375 root INFO Uploading 23_26_25_5.json to emg8426.msia423.project
2019-05-15 23:26:43,282 ingest_data_log INFO JSON uploaded as "e420378a5900c4b7a8c71225de11964a"
2019-05-15 23:26:43,282 root INFO Uploading 23_26_25_5.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:43,335 ingest_data_log INFO File 23_26_25_5.json written
2019-05-15 23:26:43,351 ingest_data_log INFO Retrieving Page 6...
2019-05-15 23:26:45,738 root INFO Uploading 23_26_25_6.json to emg8426.msia423.project
2019-05-15 23:26:46,428 ingest_data_log INFO JSON uploaded as "48e8bc58532c52032eab523daa1ff240"
2019-05-15 23:26:46,433 root INFO Uploading 23_26_25_6.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:46,502 ingest_data_log INFO File 23_26_25_6.json written
2019-05-15 23:26:46,504 ingest_data_log INFO Retrieving Page 7...
2019-05-15 23:26:48,712 root INFO Uploading 23_26_25_7.json to emg8426.msia423.project
2019-05-15 23:26:49,563 ingest_data_log INFO JSON uploaded as "b5a2bf84b4703d19987c72b27c6ed1c2"
2019-05-15 23:26:49,567 root INFO Uploading 23_26_25_7.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:49,635 ingest_data_log INFO File 23_26_25_7.json written
2019-05-15 23:26:49,638 ingest_data_log INFO Retrieving Page 8...
2019-05-15 23:26:52,149 root INFO Uploading 23_26_25_8.json to emg8426.msia423.project
2019-05-15 23:26:53,014 ingest_data_log INFO JSON uploaded as "69752e7835b0354071df3e3a20ed4940"
2019-05-15 23:26:53,014 root INFO Uploading 23_26_25_8.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:53,045 ingest_data_log INFO File 23_26_25_8.json written
2019-05-15 23:26:53,060 ingest_data_log INFO Retrieving Page 9...
2019-05-15 23:26:55,902 root INFO Uploading 23_26_25_9.json to emg8426.msia423.project
2019-05-15 23:26:56,645 ingest_data_log INFO JSON uploaded as "a782dfd11658e49e9b441bce173f4070"
2019-05-15 23:26:56,649 root INFO Uploading 23_26_25_9.json to C:\Users\Elliott\Documents\Big Data\NW\MSiA\MSiA-423 Analytics Value Chain\sell_out_project
2019-05-15 23:26:56,736 ingest_data_log INFO File 23_26_25_9.json written
```

Lessons learned

- It is better to keep all JSON files raw than to immediately ingest them into a database and lose some of the data
- Connecting to RDS can be challenging to set up
- Eventbrite doesn't provide an easy way to parse the artist that will be at a music event (Spotify-incorporation would be extremely challenging)

Recommendations

- Events Listings and Prediction - Sell-Out Prediction - Boosted Tree (1 pt)
- Events Listings and Prediction - Sell-Out Prediction - Random Forest (1 pt)
- Events Listings and Prediction - Sell-Out Prediction - Model Selection (2 pts)
- User Interaction - Web App Construction - Database Hosting (4 pts)
- User Interaction - Web App Construction - Model Hosting (4 pts)