Sprint Review

Data File Formats & Conversion for N00bs

devleague

Sprint Objective

Introduce marketers to key file formats, show how they are applicable in the field, and demonstrate basic file conversions.

What will we cover today?

1.) Examine three file formats: CSV, XML, JSON
2.) Determine why each file type is important.
3.) How Python is used to convert files.
4.) Research scripts to convert CSV files to XML, and CSV files to JSON.

What are key distinguishing features of the different file formats?

Hierarchies

Size

Human Readability

File hierarchy comparison

CSV

Cersei, Arya, Tyrion

JSON

```
{"name":”Cersei","age":”36"},
{"name":”Arya","age":”14"},
{"name":”Tyrion","age":”34"}
```

XML

```
<person>
        <name> Cersei</name>
        <age> 36 </age>
</person>
<person>
        <name> Arya</name>
        <age> 14 </age>
 </person>
<person>
        <name> Tyrion</name>
        <age> 34 </age>
</person>
```

File size comparison

Public_Art.csv

93 kb

Public_Art.json

114 kb

Public_Art.xml

163kb

When would you use these types of files?

- CSV: Transferring large amounts of data.
- JSON: Web applications.
- XML: Document markups.

How does this apply to marketers?

- CSV: Pulling large data sets out of CRMs
- JSON: Extracting data from social media platforms
- XML: RSS feeds for Google alerts about the business

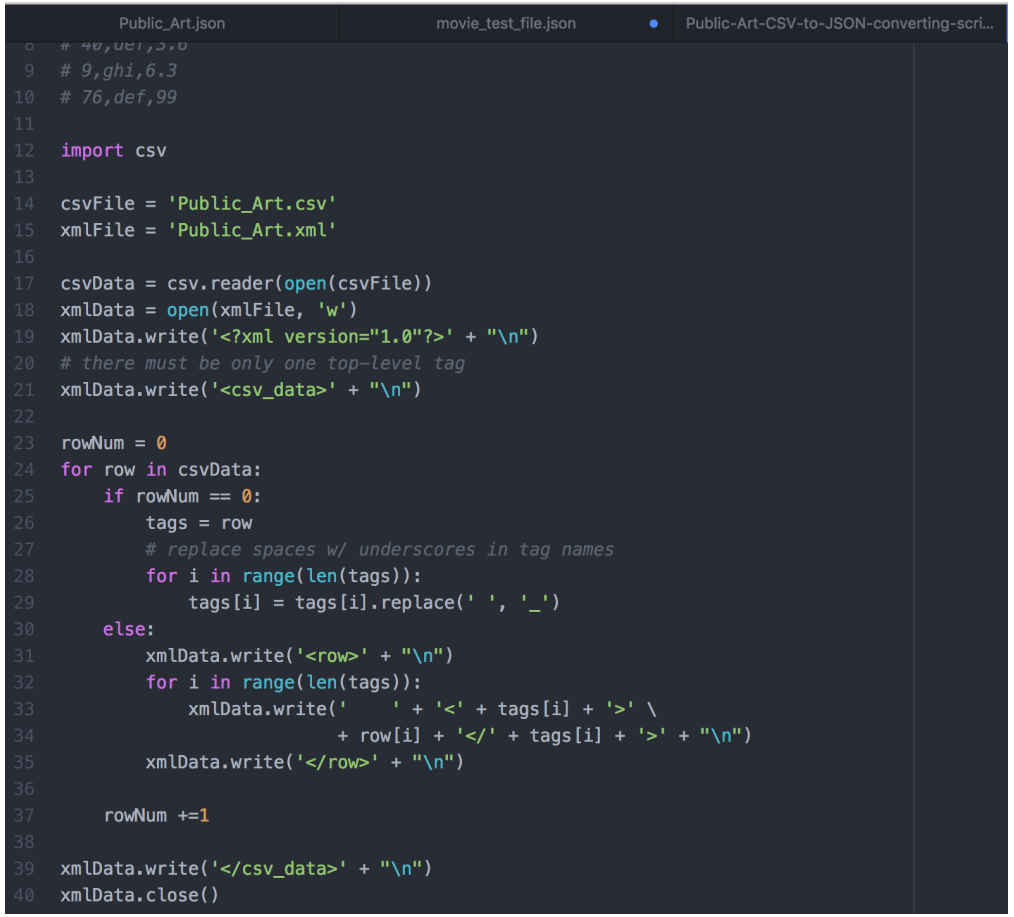
File Conversion Process

1.) Determine which file types are being converted.
2.) Research Python scripts online.
3.) Build scripts in Atom.
4.) Execute scripts in command line/terminal.

Python & Command Line Basics

1.) The command line is used in tandem with a text-editor like Atom.
2.) Scripts can be built in Atom, and executed in the command line.

CSV to XML Script



CSV to XML Script, Translated

1.) Read existing CSV file.
2.) Open a new writeable XML file.
3.) Translate CSV row headers into tags.
4.) Add brackets around the tags for each column of information to format as XML.

CSV to XML Output

```xml
<?xml version="1.0"?>
<csv_data>
<row>
    <creator>Robert Lee Eskridge</creator>
    <credit>Funded by the Works Progress Administration</credit>
    <date>1935</date>
    <description>One of a pair of murals at the Lester McCoy Pavilion at Ala Mona Regional Park. A Works Progre
    and fertility) being presented with ho'okupu (tribute). In the distance kahuna (priests) guard the kapa (ba
    <locationname>Lester McCoy Pavilion</locationname>
    <imagefile>http://hiculturearts.pastperfect-online.com/34250images/001/19300101.JPG</imagefile>
    <objectid>1930.01.01</objectid>
    <discipline>Mural</discipline>
    <title>The Makahiki Festival - The Makai Mural</title>
    <access>Limited</access>
    <latitude>21.290824</latitude>
    <longitude>-157.85131</longitude>
    <location>(21.290824, -157.85131)</location>
    <thumb></thumb>
</row>
```

CSV to JSON Script

```
1  import pandas as pd
2  df = pd.read_csv('Public_Art.csv')
3  # any operations on dataframe df
4  df.to_json('Public_Art.json')
5  print ("Conversion complete!")
6
```

CSV to JSON Script, Translated

1.) Import the pandas package.
2.) Read the CSV file.
3.) Convert to JSON.
4.) Print a “function complete!” line in the terminal once this script has been executed.

CSV to JSON Output

Bonus: Use the Atom-Beautify package to “prettify” JSON outputs!


```
{
  "creator": {
    "0": "Robert Lee Eskridge",
    "1": "Sean Browne",
    "2": "Robert Lee Eskridge",
    "3": "Charles Dickey",
    "4": "Kate Kelly",
    "5": "Marguerite Louis Blasingame",
    "6": "Juliette May Fraser",
    "7": "Charles Watson",
    "8": "Yoshinari Kochi",
    "9": "Marguerite Louis Blasingame",
    "10": "Giordano",
    "11": "Michael Weidenbach",
    "12": "Jodi Endicott",
    "13": "Ernest DeCoito",
    "14": null,
    "15": null,
```

Fin