Content-based Recommender System

Utilizing User Data

(Applied Data Science Capstone)

by

Jay Mitchell

1/14/2020

Introduction

People often travel for work or pleasure. During that time, they may wish to find amenities similar to those that they normally utilize. Companies may also be looking for new patrons, and reaching out to the out-of-town crowd may increase sales.

A user may not desire to be reactive regarding venue choices. A user may desire that instead of inputting the desired venue, that a machine-coded program provide a venue that is desired without prompting.

Additionally, a company, such as Foursquare, which was utilized during this project, may desire to capitalize on its data. They may partner with local venues to drive business toward that venue. An example may be a coupon or similar incentive. The coupon may be presented by the user to the venue. The venue then may provide monetary reward to Foursquare. In addition, Mint data was utilized to generate a user profile. A company such as Mint may also benefit from such a scheme.

Data

Data was collected from two main sources: Mint and Foursquare. The Mint data was collected to create a user profile. This data may be collected from any such repository of user actions. Mint was selected as they categorize purchases, which simplifies the data cleaning. The data received was a text file. The text file was open and then read into a Pandas dataframe for further processing. Data processing included eliminating rows with "credit", as the aim was to determine the user's purchases. Also, many columns included unnecessary information and were eliminated. Simple dataframe manipulations suggested that the data comprised mainly object, and, thus, numerical analysis would not be useful. As a tool to determine the user profile a count was performed on the categories to determine those that were utilized significantly. Some categories were eliminated, such as "student loans" as a user would not likely wish to purchase more of those.

The Foursquare data was obtained using the categories derived from the Mint data. The closest five venues were pulled for each category. Foursquare was further utilized to get a user score for each venue. For those venues without a user score, one was generated at one standard deviation below the mean of those pulled in order to not lose much data and also to not over utilize Foursquare, which has data usage limits. The Foursquare and Mint data are then combined to create a recommendation engine.

Methodology

Functions include df.types and df.corr were utilized to gain familiarity with the data. See Figure 1.

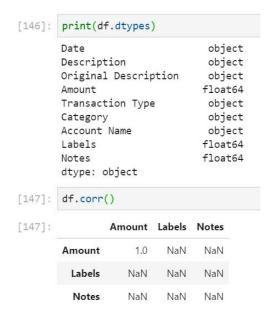


Figure 1. Initial analysis.

Based on this information, numerical analysis is unlikely to achieve results. A more categorical analysis would then be performed. To do so df.value_counts() was utilized to determine the count of the unique categories. This was visualized as a horizontal bar chart (see Figure 2). The bar chart aiding in eliminating certain categories from the data.

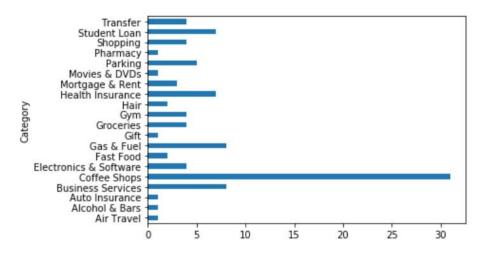


Figure 2. Categories.

Figure 2 also highlighted that there may be significant outliers. Such outliers may skew the recommendation too much. To determine the significance of any outlier, a box plot was generated. See Figure 3.

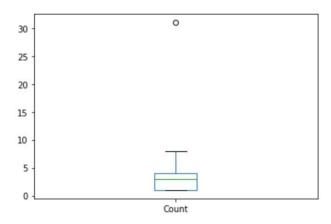


Figure 3. Initial box plot.

Figure 3 shows that one value is an extreme outlier. No other category was outside the top 75%. Thus, some reduction of weight was performed. Here, I determined to reduce the value of any outlier to 1.5X the next greatest value. It would still be weighted heavily, but not as extreme as depicted in Figure 3. The results are depicted in Figure 4.

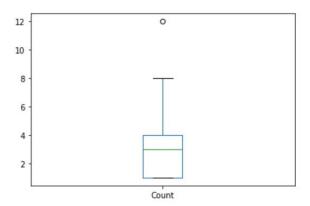


Figure 4. Final box plot.

The counts for each category was then converted into a weight. Here, I wanted weights of a positive values scaled to 10. Thus, I use simple feature scaling to get positive numbers between 0 and 10. I normalized to between 0 and 1 using Figure 5, the multiplied that score by 10). This was used instead of min-max scaling to avoid a zero value for the min. Also, this was used instead of Z-score as that would have given some negative values. The positive values will be utilized later as a weight. See Figure 6.

Figure 5. Simple Feature Scaling formula.

| | Category | Count | NormScore |
|----|------------------------|-------|-----------|
| 0 | Alcohol & Bars | 1 | 0.833333 |
| 1 | Coffee Shops | 12 | 10.000000 |
| 2 | Electronics & Software | 4 | 3.333333 |
| 3 | Fast Food | 2 | 1.666667 |
| 4 | Gas & Fuel | 8 | 6.666667 |
| 5 | Gift | 1 | 0.833333 |
| 6 | Groceries | 4 | 3.333333 |
| 7 | Gym | 4 | 3.333333 |
| 8 | Hair | 2 | 1.666667 |
| 9 | Movies & DVDs | 1 | 0.833333 |
| 10 | Pharmacy | 1 | 0.833333 |
| 11 | Shopping | 4 | 3.333333 |

Figure 6. Final Category Weights.

Foursquare data was then received based on the categories above. A Foursquare "search" was utilized with the search query set to the categories in Figure 6. The code utilizes Westminster as the location, but any may be utilized. The closest five (5) venues for each category were retrieved. In order process the data later, dummy rows were inserted in there are not 5 venues in the location. Figure 7 depicts the dataframe, though I encourage using the link to the GitHub code to read it.

| di cutegories | | hasPerk id | | location, address 1 | location.cc | location city | location.country | location.crossStreet location.dista | ce location.formattedAdd | was location.lebeledLetL | ngs locatio | n.lat location.lng | location.neighborhood loc | ation.postalCode location.state | name refer |
|-------------------------|----------------------------|------------|---|---|-------------|---------------------|------------------|-------------------------------------|--|--|-------------|--------------------|---------------------------|---------------------------------|--|
| 0 The Helibertosans | 554125941735 (name) W. | 0.0 | 5645c#StareS4e0008ec02dr | Lihit IS, Old School PI | 68 | Drevelon | United Kenndom | Nehi 4 | Ell Strid St. Old School Pt. Crowdon, Greater Lo | ol. Elabel display Tel: 51.5000.007/8816 | | 2111 -0.130645 | NeN | CRD 4GA Greater London | Seer Gor I Alcohol Delivery = 15/9054 |
| 1 THE MEDICALISMS | 988419941735 name 1 | 0.0 | 5/DARMINI-1-1332596/021 | 11 faces litted | GR. | Limiten | Child Kingdon | Note to | 2.0 [10 Fases Street, Loredon, Greater Loredon, WC | B. Clater Visiter, No. 51 Spinsocration | 01. 5150 | unio di trianni | Nett | WCZESAA Greater London | Allought Alcohol Delivery 1578034 |
| 2 | NaN. | took | . hats | Nehi | TON | Nati | Nahi | | | | | total Nets | NaN | NeN Nets | Pieli 7 |
| 4 | ton | Black | No. | Nell | Blabi | Net | Net | | | | | Tools tools | Nett | Nati Nati | Toda 7 |
| | NAN | Tools | Tuni. | Turki . | Tiers | NaN | Tork | | | | | Tools Note: | NaN | Nation Nation | Nahi 7 |
| | 88d1e0931735; neme: C | 0.0 | | Westmooder SuberStation | - 68 | | United Kingdom | | 42 (Westminster Fulse Station, London, Greater). | | | | Net | SWIA Greater London | Code Coffee (=15700)4 |
| | BBd1eD31735, name: C | | 4cm/3r4d3bletdrbSbetdb/3 | Particular House | 68 | | United Kingdom | | | | | | NaN | SWIA Greate London | Departs for Coffee Stop - 15/1004 |
| | | | | | | | | | | | | | | | |
| | 8881+0931735; name: C | 0.0 | | St. Thomas Hospital | GE | | | | | in [[label] display, lat: \$1,49907,009003 | | | NeN | SET 75H. Greater London | ANT Coffee +157034 |
| | 88d143941335; name: 16 | | 4654942895445209542745 | 49 York Rd. | GB | | United Keepform | | Ell. 149 York Rd., Waterlass, Greater London, SE1 | | | | NaN | SET 7NF Greater Lambon | Coffee Cultura: +1579054 |
| a (Unit 40129909455) | 886140931735; henr: C | | 48:60120=49-0x5+32/119658 | Cine Great George St. | GE | | United Kingdom | | 1.0 Cine Great George St, Condon, Greater Londo | | | | NeN | Naft Greater London | Costa Coffee v-1579034 |
| 10 | NeN | flah | Trans. | Neti | NeN | Nett | Nati | | | | | NeN NeN | NaN | Nehi Nehi | feels 7 |
| 0 | tiets | NaN | Nation 1 | Nell | Nati | Nets | NeN | | | | | reats reats | Nett | Nati Nati | Teef4 F |
| 12 | NeN | flaN | Turki . | Nels | NeN | NeN | Nahi | 1,117 | | | | Neti Neti | NaN | Nehi Nehi | Plaft 7 |
| 3 | tiats | NaN | No. | 1661 | NaN | Nati | NeN | | | | iaNi | rouns rouns | Net : | Neti Neti | Nell 1 |
| 4 | NeN | TON | Tues: | Nets | 7647 | Nett | Nets | Nels 1 | eN 1 | iaN 7 | SelV. | Nets Nets | NeN | NeN Nets | Nebi 7 |
| 1 [[vf: 46/588/88489 | 88211851735; hwnw: G., | 0.0 | 486424560393603653 | 244-246 Westmirsder Bridge Rd | GB | Lightlen | United Kingdom | 160 4 | 6.8 244-245 Westminiter Bridge Rd, London, Gree | n. [[label] display], lat: \$1,920/28823638 | 85. 5150 | 3729 -0.117016 | Net | Natt. Greater London. | A. P. Food Express -+1579034 |
| 6 IDH YHSRARAYIN | 884118551735", henri G. | 0.0 | 4:58:0925418x\930ee4760e | St. Thomas: Hospital | GB | Lendon | United Keeplern | Westminist Bridge Rd 2 | 2.0 (Si Thomas' Hospital (Westminster Bridge Ro | [[label: display, lat: \$1,49995081,8105 | 11. 51.49 | 2051 -0.110922 | NaN | SE1 7831. Greater Lambon | MISS Simply Food = 15/9054 |
| 7 Did Witsenberg | 884128941735, hwnw: G., | 0.0 | 49545e4e(964e525725e27e3 | 17 Smith Sq | GB | Lienten | Crited Kingdom | NeN 6 | S.D. (17 Smith Sig. Condon; Greater London; SW1.F | 2,- (flater) display, fact \$1,495615014550 | 76. 51.49 | 9616 -0.12589E | NeN | SWTP2 Greater London | Department for Emissionwest, Food & Bural Affair 1579004 |
| B IDM HATBARDS | 884118951715', henry 'G. | 0.0 | 5cH79u8863cd000x4d2dsd | Sirvit To Sesufficients Places, York Road | 68 | London | United Kingdom | NeN 5 | LD SUNT In Southbank Place, York Road, London | G. (Field: Migray, Set: \$1.500021, Se | 97- 5130 | 0.119984 | NaN | N11 /ND Greater Lambon | NSS Food Hell = 1579034 |
| 9 [[uf: HartstadesHess | 8821185173V, harrie G. | 0.0 | Addb-Ne1415206a543791371a | 12-16 Imbankment Place | GB | Lignation | United Kingdom | Nell 6 | 2.0 (17-18 Embankment Place, London, Greater Lon | ts (flatel) 'display', 'lat: 51.500917905052 | 87. 55.50 | 1918 -0.122930 | Neti | Natl Greater London | A.P. Food express =-1570034 |
| P (Feb 45/53666489 | 884198941735; harrer 14 | 20 | 105de457didede5/16b4b0e7 | Lambeth Palace Road | 68 | Limiten | United Keepdom | Nels 4 | 40 Sambelli Pelace Road, London, Greeler Londo | | g1 51.400 | 9811 -0.116327 | Lambeth | 517 71W Greater Lambon | Sessot House v-15/9054 |
| 1 | 754% | NaN | Nation 1 | 200 | NaN | Net | NeN | 200 7 | en a | iaN 7 | (aN | rown rowns | NeN | Neti Neti | 766 7 |
| 2 | Net | Studi | Name (| Tuels: | 2045 | Nati | Tools | total t | eli i | uN r | Self. | ton ton | NeN | NeN Net | total 1 |
| 3 | Tub. | ton | No. | Total | NeN | NeN | NeN | 1961 7 | di i | iah i | (aN | TOURS TOURS | Nett | Neti Neti | Toda 7 |
| | tish | took | t total | Neti | NaN | Nati | Nett | | eN . | | | tuni tuni | NaN | Nehi Nehi | Rehi P |
| | 88412851735, harrer G. | | 3#Oh (2014### bob (2008)205500 | Neti | 68 | | Cristed Kingdom | | | on) (Tabel Stepley, Sel. 55.500909, Se | | | Net | SET Greater London | GR Shep = 157004 |
| | S/HOMBORICE THEW S. | | 575/etra6490e13aalle00901d | Nation (| - 08 | | United Kreadow | | | on) [Table" shapley, let: \$1,500001, by | | | Nati | Nahi Nahi | Excel Gift & Internet Cale TS P3054 |
| | BBST285173V Name G. | | Artis10/41405acts0ct0078bar | 12 Bridge M | GE | | United Knodom | | | er. [fabel display, fat: \$1,500/356/4786 | | | Net | SWIA ZIX Greater London | Places of Pallament Shop in 157(0)4 |
| | 958219941735, hame: U. | | 51/0/26/01/00/00/00/00/00/00/00/00/00/00/00/00/ | From Waterloo Station | 68 | South Bank | United Krepdom | | | | | | NaN | NeN Greater London | Social Cirtis, Top-ups, Mobile Limbolang in 15/10094 |
| | Notice Transport | TO/N | 517 DZSOKOBROZODIA-GZOZ | From Waterson states | NaN. | South Barrie | | | | | | NAME OF TAKE | NAM NAM | | |
| 9 | | | | | | | NeN | | | | | | | Nahi Nahi | Toda 7 |
| 0 | NeN | TON | Tuels . | Nets | 7647 | NeN | Net | | | | | Nets Nets | NaN | fosti fosti | Plats 7 |
| 1 | tots | NaN | Total Sealer | Net | NaN | Net | NeN | | | | | rosti rosti | Net | Neti Neti | Tuft 1 |
| 2 | NeN | fusN | Nets | Neti | NeN | NeN | Nett | | | | | Neti Neti | NaN | Nehi Nehi | NeN 1 |
| 1 | fishs | NaN | Nation 1 | Nell | Nati | Nati | NeN | Nell 1 | eli I | iahi r | ishi. | nun nun | NeN: | NeN NeN | Neft f |
| 14 | NeN | flaN | Turk . | Neti | NeN | NeN | Nets | Nels 1 | | | | tiets tiets | NaN | Nahi Nahi | NeN 7 |
| S (Sid Approximate) | 00x176941735, hwnw: G- | 0.0 | SIDAteObe-InObdZecS157affb | 10 Whitehall 25 | GE | Lightlen | United Kingdom | Nell 9 | 7.0 (10 Whitehall P), London, Greeter London, U | nt. [[label] display, lat: 51.526318434150 | EL 5110 | 1118 -0.124648 | Net | NeN Greater London | ESPA Life Cym. (=1578034 |
| 6 ((nf) 45/58/d8/488 | 88417941795 henv: G. | 0.0 | 55118w6408x10x0x61xdix4 | 17-21 Northumberland Ave. | GB | City of London | United Kingdom | NeN 5 | 5.0 [17-21 Northumberland Ave, City of London, C | re [[label: display, lat: 5150/018/23184 | 41 5130 | 019 -0.124760 | NeN | WCINSEA Greater London | The Gym London Charing Cross = 15/9054 |
| 17 [[id: 46/586/86/480] | 882176941735; hwnw: G., | 0.0 | 4ccs/2xf566ad203f582dfd | County Hell | GE | Lipsdon | United Kingdom | Betredene Road 3 | 2.0 (County Hall (Between Road), London, Great | er [[label] display], far: \$1.503078935344 | 76 5150 | 2019 -0.119739 | Net | SET 776 Greater London | The Gym Ni Club et County Hell ->-1579034 |
| 18 IDH 16586486488 | 8841 (9941 (15), hwnw (G., | 20 | 4:slefe19400leF40057165b2 | Derby Galle, 1 Carron Ross | 68 | City of Westminster | United Kingdom | Nels 1 | DD | er, [[label: display, lat. 51.501859293004 | 93. 5130 | 1859 -0.124989 | NeN | SW14 ZIN. Greater Lorelon. | Westmanske Gym. 9-75/9054 |
| 9 [Cut 45/Southeatt | 888176941735; hwnw: G. | 0.0 | 525/40364384164661354007 | 76/0 | GB | Nett | United Kingdom | Nell 4 | S.D. (United Kingd | orn) [[fabel: Mopley, Set : \$1,500/4, Tex | 5150 | 2742 -0.117479 | NeN | NeN NeN | Helel Com 1578034 |
| O per debassions | 8841105105, new S. | 0.0 | tationidaleHispopoleetial | 635 High Road Leytonstone | 68 | Limiter | United Keeplern | NeN S | S.D. 1835 Prigh Blood Leytronitone, London, Greater | in. [label: dopley, lat: \$1,504148054725 | 15. 5130 | 1148 -0.529160 | NaN | 511 400 Greater Lambon | Huby's Her & Beauty Selon : v-15/9054 |
| n pic wisawawa | 004110051735, name: %. | 0.0 | 5c5273b4c0cm/80000x863776 | Si Thomas Hospital | GB | Lignation | Crebed Kingdom | Nell 3 | 5.0 Di Thomas Hospital, Lorsfon, Greater Lorsdon | 5. (Tabel: Vispley, Sel: 51-499247, Sr. | 91. 5149 | 2247 -0.1192/00 | NeN | SET Greater London | creative hair studio. in 1579034 |
| 2 | NeN | Tools | Nation 1 | NeN | 70470 | Nett | Nets | Nels 1 | eN 1 | iati i | Self. | tests trent | NaN | Nels Nels | Tests 7 |
| 13 | Tials | NaN | Tests: | Nelli | Nati | Tinti | NeN | Nell 1 | di i | iaN 1 | iahi | rears rears | Neti | Neti Neti | Toda 7 |
| | NeN. | TON | Turk . | Neti | 70475 | Nett | Net | | | | | ton ton | NaN | Nels Nels | Tubi 7 |
| | 760 | ton | No. | Nation (| NaN. | Name . | Nan | | | | | NATE THAT | Nation 1 | Nation Nation | Table 1 |
| | Net | took | Tools . | Tests | Nati | Nett | Nett | | | | | tools tools | NeN | Nels Nels | Nell 7 |
| * | Tiefs. | NaN | Today . | Netic Netic | 1000 | Net | NeN | | | | | room room | Net | Nation Nation | Nell 1 |
| | NeN. | Tours | Teatre . | Neti | NaN . | Nati | Nan Nah | | | | | Nation Nation | Nati | room room | Nati I |
| | NaNi NaNi | TOWN. | | Nation Nation | Plant | Nation Nation | Nati | | | | | reers reers | Nati | Nation Nation | Nati |
| | | 7000 | Control of the second | | 1000 | 1465 | | | | | | | | | |
| | le5684a65b475, 'name' 'C | | \$2,003f111d26c041a139d14 | St Thomas Yospital | GB | | United Keepforn | | (3) Thomas Hospital (London St), London, Gre | | | | NaN | SET 7531. Greater London | Sambuy's Phermacy = 157905 |
| | 884196941735; harrer 31. | | 50007050s4b01s0btcs3c800 | Nell | GE | | United Kingdom | | | om): [flabel1:display]; fart.51.499025895840 | | | NeN | Nati Nati | Pharmacy PR, St Thomac = 157903 |
| | 884198941735; harrer: 71 | | 50753927+4509356475Hec4 | Nets | GB | | United Kingdom | | | rm) [[label: display, lat: 51.499/131/4950 | | | NaN | Yeah) Neft | St Humas Pharmacy Sea Room (+15/905 |
| TH 465866400 | 884196941735; hwret: H | 0.0 | 4x:5188954x525xx620x3 | Westminster Bridge Rd | GE | Limiten | United Kingdom | Nell 3 | ED: Westminder Bridge Rd, London, Greeker Lond | os. [[label] display], lat: 51.499710477209 | 51.49 | | Net | SET 7511 Greater Landon | 38 Herman Hospital +-157803 |
| | NeN | flaN | Turki . | Neti | 70470 | Nett | Net | | | | WN. | held held | NeN | Nehi Nehi | NeN |
| | fishs | NaN | No. | Nell | NeN | Nets. | NeN | Nell 1 | di I | iati i | Self- | reals reals | Nett | Neti Neti | York |
| | NeN | forN | tori | Neti | NeN | Nett | Net | Nels 1 | eN 1 | ian r | Selv. | tieti tieti | NeN | Nehi Nehi | turki |
| | tions | NaN | No. | Nell | NaN | Nati | NeN | Nell I | ali I | iaN f | ishi . | reals reals | Nati | Nati Nati | NeN 1 |
| | NeN | TorN | Turk . | Page 1 | 70470 | Nett | Nets | Nels 1 | eN 1 | ian r | Self. | turk Nets | NeN | Nels Nels | Nebi 7 |
| | tots | NaN | No. | Nation 1 | NaNi | Net | NeN | 166 7 | di i | iaN 7 | iani | rown rowns | Net | Net Net | Toda 7 |
| | | | | 74810 | | | | | | | | | | | |

Figure 7. Venues.

Once the venues were obtained, the categories from the Mint data were added using df.concat. Once this was done, the dummy rows were then deleted. The dummy rows may the concatenation of the dataframes much easier.

Next, each venue included a venue id in Figure 7 above. These venue ids were utilized along with the Foursquare "venues" option to return the values for each venue. Some venues, however, did not have a value and were initially recorded as 'NaN'. In order to provide some, but not a lot of weight to unrated venues, the mean and standard deviation of the rated venues were determined and the unrated venues were given a rating of (mean – standard deviation). This puts them on the low end, but not as outliers. As these ratings are already on scale to ten (10), they were not further manipulated. See Figure 8.

```
0
     6.240819
1
     6.240819
2
    7.900000
     6.240819
3
4
    6.600000
5
    6.200000
6
     6.240819
7
     6.240819
8
    6.800000
9
    6.240819
10 6.240819
11
     6.240819
12 6.240819
13
     6.240819
    6.240819
14
15
    6.240819
16
     6.240819
17
    7.400000
18
    6.100000
19
    7.600000
20 6.240819
21
    6.240819
22 6.240819
23 6.240819
24 6.240819
25 6.240819
26
    6.240819
27
     6.240819
```

Figure 8. Venue Ratings.

Results

Now that there is a rating from the venue, each venue has a category, and there is a weight for each category, a final WeightScore (Venue Rating x NormScore) is determined for each venue. The dataframe is sorted in descending order by WeightScore and the top 10 venues are selected. Further dataframe manipulation is performed to ensure that the name of the venue and its latitude and longitude are in the dataframe for plotting. The dataframe is depicted in Figure 9. Again, dataframe viewing is best done via GitHub.

| | name | categories | address | cc | city | country | crossStreet | distance | formattedAddress | labeledLatLngs | lat | Ing | neighborhood | postalCode | state | id | WeightScore | Category |
|----|----------------------------------|---|-----------------------------|----|------------------------|-------------------|------------------------|----------|---|---|-----------|-----------|--------------|------------|-------------------|-----|-------------|-----------------|
| 2 | Costa Coffee | [['id': '4bf58dd8d48988d1e0931735', 'name': 'C | Westminster Tube Station | G8 | London | United Kingdom | NaN | 74.0 | [Westminster Tube Station, London, Greater Lon | [('label': 'display', 'lat': 51.50093364186963 | 51.500934 | -0.124805 | NaN | S W1A | Greater London | NaN | 79.000000 | Coffee Shops |
| 4 | AMT Coffee | [('id': '4bf58dd8d48988d1e0931735', 'name': 'C | St. Thomas Hospital | GB | London | United Kingdom | Lambeth Palace Road | 348.0 | [St. Thomas Hospital (Lambeth Palace Road), Lo | [('label': 'display', 'lat': 51.49997320903263 | 51.499973 | -0.118974 | NaN | SE1 7EH | Greater London | NaN | 66.000000 | Coffee Shops |
| 3 | Despatch Box Coffee Shop | [('id': '4bf58dd8d48988d1e0931735', 'name': 'C | Portcullis House | GB | London | United Kingdom | NaN | 71.0 | [Portcullis House, London, Greater London, SW1 | [{'label': 'display', 'lat': 51.50111406437572 | 51.501114 | -0.124743 | NaN | SW1 A 2 | Greater London | NaN | 62.408193 | Coffee Shops |
| 6 | Costa Coffee | [('id': '4bf58dd8d48988d1e0931735', 'name': 'C | One Great George St | GB | London | United Kingdom | NaN | 373.0 | [One Great George St, London, Greater London, | [{'label': 'display', 'lat': 51.50121966342386 | 51.501220 | -0.129109 | NaN | NaN | Greater London | NaN | 62.408193 | Coffee Shops |
| 5 | Coffee Culture | name: b | | | Waterloo | United Kingdom | NaN | 584.0 | [49 York Rd., Waterloo, Greater London, SE1 7N | [{'label': 'display', 'lat': 51.50304865891913 | 51.503049 | -0.115980 | NaN | SE1 7NJ | Greater London | NaN | 62.000000 | Coffee Shops |
| 12 | Gassiot House | [['id': '4bf58dd8d48988d196941735', 'name': 'H | Lambeth Palace Road | GB | London | United Kingdom | NaN | 444.0 | [Lambeth Palace Road, London, Greater London, | [{'label': 'display', 'lat': 51.498833, 'lng': | 51.498833 | -0.118327 | Lambeth | SE1 7EW | Greater London | NaN | 41.605462 | Gas & Fuel |
| 19 | The Gym & Club at County Hall | [['id': '4bf58dd8d48988d176941735', 'name': 'G | County Hall | | London | United Kingdom | Belvedere Road | 302.0 | [County Hall (Belvedere Road), London, Greater | [{'label': 'display', 'lat': 51.50207893534426 | 51.502079 | -0.119739 | NaN | SE1 7PB | Greater London | NaN | 25.333333 | Gym |
| 17 | ESPA Life Gym | [['id': '4bf58dd8d48988d176941735', 'name': 'G | 10 Whitehall Pl | GB | London | United Kingdom | NaN | 597.0 | [10 Whitehall Pl, London, Greater London, Unit | [{'label': 'display', 'lat': 51.50631843415083 | 51.506318 | -0.124648 | NaN | NaN | Greater London | NaN | 24.666667 | Gym |
| 21 | Hotel Gym | [['id': '4bf58dd8d48988d176941735', 'name': 'G | NaN | GB | NaN | United Kingdom | NaN | 435.0 | [United Kingdom] | [['label': 'display', 'lat': 51.50074, 'lng': | 51.500740 | -0.117470 | NaN | NaN | NaN | NaN | 20.802731 | Gym |
| 20 | Westminster Gym | [['id': '4bf58dd8d48988d176941735', 'name': 'G | Derby Gate, 1 Canon Row | GB | City of Westminster | United Kingdom | NaN | 130.0 | [Derby Gate, 1 Canon Row, City of Westminster | [('label': 'display', 'lat': 51.50185929300493 | 51.501859 | -0.124989 | NaN | SW1A 2JN | Greater London | NaN | 20.802731 | Gym |

Figure 9. Mapped Venues.

The Folium mapping tool was then used to depict the venues in Figure 9. To determine the current location, a large red dot was plotted. The blue dots then represent each of the venues. The radius of the blue dots plotted was based on the WeightScore, such that the relative size is proportional to the WeightScore. Each dot may be interacted with to show the venue name and the category to which it belongs. A sample map is depicted in Figure 10, with one venue interacted with. Note that the map figure does not fit into my browser.

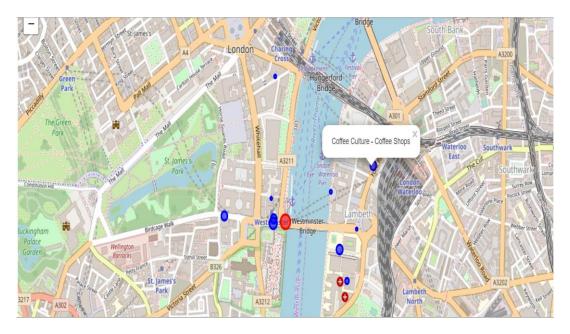


Figure 10. Local Points of Interest.

Discussion

The above recommendation engine did produce local venues of interest based on my interests. Further implementation may focus on improving data, processing, and profit.

Data

The above recommendation engine did produce local venues of interest based on my interests. Further implementations my utilize other data sources based on the particular user. Not everyone utilizes Mint and Foursquare. Also, the data from Mint may be updated from time to time to ensure relevancy. More generalized categories may be used as well. With further data, a k means clustering algorithm may be used to determine which categories are similar.

Processing

The final recommendation was very much skewed toward those categories with many counts. Counts may not be the best way to determine a user profile. Other methods may reduce the impact of the outlier. Binning may be utilized to get items that are 'high', 'medium', and 'low', or some other scheme. Each bin may then be associated with a weight. This may result than more than just coffee shops. Though, I not complaining.

Profit

This implementation also needs more profit. The popup label may also include a hyperlink to a coupon or other incentive. Such an incentive should include a way for the data sources (and the app developer) to make a cut of the earnings from the directed customer.

Conclusion

This recommendation engine profiled a user based on historical data. That data was then used to determine nearby venues, such as when traveling, in a different part of town, or just looking for hidden gems. Previous users had rated those venues. Those ratings, along with the user profile, determined the venues that were displayed to the user. The user then could interact with a Folium map to determine their present location and the locations and weight score of venues of interest.

Code: https://github.com/jmitchell4390/Coursera Capstone/blob/master/Capstone Project.ipynb

Map:

https://nbviewer.jupyter.org/github/jmitchell4390/Coursera_Capstone/blob/master/Capstone_Project.ipynb

Text file: https://github.com/jmitchell4390/Coursera Capstone/blob/master/transactions.txt