capstone (3)

July 30, 2020

0.1 # Capstone Project - The Battle of Neighborhoods

0.2 Table of contents

0.3 Introduction

Background:

- The café and coffee shop industry is an important industry in terms of employment and contribution to the UK economy, according to "Investigating the Success of Independent Coffee Shops and Cafes in the UK: Findings from a Pilot Study" (by Jacqueline Douglas, Alexander Douglas, Michele Cano, David Moyes).
- However, sustainability in terms of longevity is an issue. Despite the low barriers to entry into the industry, cafés are a very high risk business and most start-ups fail. According to the *Office for National Statistics*, only 34.6% of accommodation and food services survive longer than five years (*Office for National Statistics, 2016*).
- It's clear that to survive coffee shops owners need to have some feedback from their clients. The feedback should be analyzed (in this research with the help of Machine Learning) and interpreted, as with this information, entrepreneurs can react and change their coffee shops according to their customers opinions and prosper in this business area.

Business Problem:

In this research, the Foursquare Rating has been chosen as the measure of the customers' loyalty for the venues, which have placed their menus on the Foursquare Website. The aim of this research project is to determine if the connection between Foursquare Rating (customers loyalty measure) and availability of the most popular items in a coffee shop menu exists (also the prices for these most popular items are being taken into consideration).

0.4 Data Sources

- Foursquare Rating is based on a number of signals that from social data mines (likes and dislikes, and positive versus negative tips). It's going to be used as the main measure of

the project, as the goal is to find the connection between this metric and availability in the published menus the most popular of the coffee shop items.
- Published on the Foursquare, coffee shop Menus with the items and their prices.
- Foursquare Geospatial Data - latitude and longitude of the venues.
- Some additional information: number of the coffee shop photos, number of guest tips, etc.
- Wikipedia page with the list of the districts of London ordered by population density, based on population estimates for 2019. https://en.wikipedia.org/wiki/List_of_English_districts_by_population_density

0.5 Methodology

- Data Collection:

  1. Import of the Python libraries.
  2. Getting the list of the London districts from the Wikipedia.
  3. Foursquare API, getting the venues for the London districts from the previous step.
  4. Foursquare API, getting the data of every venue (venue's rating, URL, number of photos and tips, etc.).
  5. Web crawling, getting the menu (items and prices) for every venue.
  6. Data cleaning.
  7. Identification of the most popular items in the obtained menus.
  8. Representation on the map analyzed coffee shops.

- Data Analysis:

  1. Data preparation.
  2. Linear Regression construction.
  3. Random Forest Regression construction.

0.6 ## Data Collection

1. Import of the necessary for the research Python libraries.

[1]:
```
import lxml
import time
import folium
import random
import requests
import warnings
import numpy as np
import pandas as pd
import matplotlib.cm as cm
import matplotlib.colors as colors

from bs4 import BeautifulSoup
from matplotlib import pyplot
from sklearn import linear_model
from sklearn.cluster import KMeans
from sklearn.metrics import r2_score
from pandas.io.json import json_normalize
```

```
from sklearn.ensemble import RandomForestRegressor
```

Setting of some visualization parameters.

[2]:
```
%matplotlib notebook
warnings.filterwarnings('ignore')
pd.set_option('display.max_rows',None)
pd.set_option('display.max_columns',None)
```

2. Getting the list of the London districts from the Wikipedia page with density more than 10,000 per km²: https://en.wikipedia.org/wiki/List_of_English_districts_by_population_density.

[3]:
```
url = 'https://en.wikipedia.org/wiki/
↪List_of_English_districts_by_population_density'
soup = BeautifulSoup(requests.get(url).text,features = 'lxml')
table = soup.find_all(class_ = 'wikitable')
t = pd.read_html(str(table))[0]
neighborhoods = t.copy()
neighborhoods['neighborhood'] = t['District'] + ', London, Greater London,␣
↪United Kingdom'
neighborhoods = neighborhoods[['neighborhood']]
neighborhoods
```

[3]:

	neighborhood
0	Islington, London, Greater London, United Kingdom
1	Tower Hamlets, London, Greater London, United …
2	Hackney, London, Greater London, United Kingdom
3	Kensington and Chelsea, London, Greater London…
4	Lambeth, London, Greater London, United Kingdom
5	Camden, London, Greater London, United Kingdom
6	Westminster, London, Greater London, United Ki…
7	Hammersmith and Fulham, London, Greater London…
8	Southwark, London, Greater London, United Kingdom

[4]:
```
CLIENT_ID = 'M5QE21FVOFLBC5SHM4GPU2ETFQ4DXIYPWFAFX1S1BJUGHOAP'
CLIENT_SECRET = 'HLPNQSI5ICRNSLXTPZ4PEARAGFCKI1AYNYTP5AFV1WY5S2HH'
VERSION = '20190425'
```

3. We will get from Foursquare API the list of the venues, located in the obtained from Wikipedia London districts.

For that, we will execute the API-call function, which will retrieve the venues in 900 meters distance from the neighborhood center.

[5]:
```
def getNearbyVenues(neighborhoods,category,radius = 900,limit = 3):
    venues_list = []
    for neighborhood in zip(neighborhoods):
```

```
        url = 'https://api.foursquare.com/v2/venues/explore?
↪&client_id={}&client_secret={}&v={}&near={}&categoryId={}&radius={}&limit={}'.
↪format(
            CLIENT_ID,CLIENT_SECRET,VERSION,neighborhood,category,radius,limit)
        results = requests.get(url).json()['response']['groups'][0]['items']
        venues_list.append([(neighborhood[0],
                             v['venue']['id'],
                             v['venue']['name'],
                             v['venue']['location']['lat'],
                             v['venue']['location']['lng'],
                             v['venue']['categories'][0]['name'])
                            for v in results])
    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item␣
↪in venue_list])
    nearby_venues.columns = ['neighborhood',
                             'venue_id',
                             'venue_name',
                             'venue_latitude',
                             'venue_longitude',
                             'venue_category']
    return(nearby_venues)
```

[6]:
```
cat = '4bf58dd8d48988d1e0931735' # the coffee shops category
neighborhood_venue = getNearbyVenues(neighborhoods =␣
↪neighborhoods['neighborhood'],category = cat,limit = 495)
neighborhood_venue.head()
```

[6]:
```
                                    neighborhood  \
0  Islington, London, Greater London, United Kingdom
1  Islington, London, Greater London, United Kingdom
2  Islington, London, Greater London, United Kingdom
3  Islington, London, Greater London, United Kingdom
4  Islington, London, Greater London, United Kingdom

                   venue_id               venue_name  venue_latitude  \
0  58a576b8f8572431aec041b8                Kobo Cafe       51.534988
1  584bfe4544587f042f5ff30c              Katsute 100       51.534286
2  4fc9ff8ce4b087d229f75ce4  The Coffeeworks Project       51.534254
3  50338ecbe4b0c160f73b46d0                 Giddy Up       51.536374
4  5a9ea6c2d03360028695111e              Six Degrees       51.535228

   venue_longitude venue_category
0        -0.104149           Café
1        -0.104540       Tea Room
2        -0.104684    Coffee Shop
3        -0.102930    Coffee Shop
4        -0.103486    Coffee Shop
```

So, in the given dataframe we have the neighborhood, venue id, venue name, venue's latitude, longitude, and category.

4. Now, let's get the Foursquare Rating and other valuable information (url, number of tips, photos, etc.) for every venue.

[8]:
```
def get_venue_data(venues):
    venues_list = []
    for vn in venues:
        url = 'https://api.foursquare.com/v2/venues/{}?
↪&client_id={}&client_secret={}&v={}'.
↪format(vn,CLIENT_ID,CLIENT_SECRET,VERSION)
        global results
        try:
            results = requests.get(url).json()['response']['venue']
        except:
            print(vn,'Something went wrong with the Foursquare response')
            pass
        try:
            venues_list.append([(results['id'],
                                 results['name'],
                                 results['location']['formattedAddress'],
                                 results['canonicalUrl'],
                                 results['categories'][0]['name'],
                                 results['verified'],
                                 results['stats']['tipCount'],
                                 results['price']['tier'],
                                 results['price']['message'],
                                 results['price']['currency'],
                                 results['rating'],
                                 results['photos']['count'],
                                 results['tips']['count']
                                )])
        except:
            print(vn,'Something went wrong, critical data is missing')
            pass
    df_venues = pd.DataFrame([item for venue_list in venues_list for item in␣
↪venue_list])
    df_venues.columns = ['v_id',
                         'name',
                         'formattedAddress',
                         'canonicalUrl',
                         'categories',
                         'verified',
                         'tipCount',
                         'tier',
                         'message',
                         'currency',
```

```
                             'rating',
                             'photos',
                             'tips'
                            ]
    return(df_venues)
```

[9]:
```
df = pd.
 ↪concat([neighborhood_venue,get_venue_data(neighborhood_venue['venue_id'])],axis␣
 ↪= 1,join = 'inner')
df =␣
 ↪df[['neighborhood','venue_id','venue_name','venue_latitude','venue_longitude','venue_catego
        ,'categories','tier','message','rating','photos','tips']]
#df.head()
```

```
5cb32d7bad910e002cdabc0f Something went wrong, critical data is missing
5c979e1e86bc490039527670 Something went wrong, critical data is missing
4cb44eb2562d224b81203388 Something went wrong, critical data is missing
4d5fcc4f865a224b17f7a285 Something went wrong, critical data is missing
4f671617e4b0e78c02e9c203 Something went wrong, critical data is missing
4eb10e889a52c49f4cd2e953 Something went wrong, critical data is missing
552057e4498e1d3a865aeeb4 Something went wrong, critical data is missing
530f34f911d20da2504590af Something went wrong, critical data is missing
56da524ecd108d4b2ff7ed86 Something went wrong, critical data is missing
4ce24b4cf8a4a14321a2efbc Something went wrong, critical data is missing
50408782e4b06c0c2b2d9654 Something went wrong, critical data is missing
4d5a55bb1d6cf04d08a140fe Something went wrong, critical data is missing
5eb534a96c4752000849cd61 Something went wrong, critical data is missing
4b0730e7f964a5203ef922e3 Something went wrong, critical data is missing
4d441ba1c3e5f04d46c79620 Something went wrong, critical data is missing
4dbd309acda109aa6cad9616 Something went wrong, critical data is missing
5a16a5a098fbfc0aa529ba70 Something went wrong, critical data is missing
5c4c7c2335811b002c3bac56 Something went wrong, critical data is missing
4c5fc3f490b2c9b6df663b22 Something went wrong, critical data is missing
4d661f7ff75c3704343dd49a Something went wrong, critical data is missing
5811f99a38faba8934b1a504 Something went wrong, critical data is missing
5355007c498e29540c91f51d Something went wrong, critical data is missing
4dea272ed164ef597ce26c8d Something went wrong, critical data is missing
4b1ff5c2f964a520612b24e3 Something went wrong, critical data is missing
4def81757d8bb02cbefb44ad Something went wrong, critical data is missing
57ab50e0498e038d1a8e69ad Something went wrong, critical data is missing
5324a700498e5ce42695e4e4 Something went wrong, critical data is missing
5794ed31cd1010a4f772c0e7 Something went wrong, critical data is missing
4eb3e9546c2590eb875e4b1d Something went wrong, critical data is missing
4c8a4e3752a98cfa323e2be9 Something went wrong, critical data is missing
4ba5fe75f964a5204e2d39e3 Something went wrong, critical data is missing
4c3ed5f00e0d0f476b11167f Something went wrong, critical data is missing
50cc7653e4b0e6c83bc62e02 Something went wrong, critical data is missing
```

4bf789704a67c928abdf23cf Something went wrong, critical data is missing
4ddd1aaab0fba481fc908415 Something went wrong, critical data is missing
5cb1c5a5acb00b002c3deb92 Something went wrong, critical data is missing
4cde9bed7e2e236a7e39801b Something went wrong, critical data is missing
4c0c507b009a0f474987ecbf Something went wrong, critical data is missing
4f0c0583e4b0baf830708a2f Something went wrong, critical data is missing
4d9495265241a1cd6ea0624b Something went wrong, critical data is missing
4ff85f4fe4b0484033aae831 Something went wrong, critical data is missing
4f5b3fcde4b0695cb9f23d52 Something went wrong, critical data is missing
54ccc56f498e4249447bc5bf Something went wrong, critical data is missing
4d5b8292b815b60c5a174c16 Something went wrong, critical data is missing
4ff96529e4b0b8fdaafdacf5 Something went wrong, critical data is missing
5a58d588f193c07e2ad86312 Something went wrong, critical data is missing
58c6f9de109dfe5fc57d9192 Something went wrong, critical data is missing
5e263227d7da7600085822d3 Something went wrong, critical data is missing
5065b0bbe4b0d3feb1401e8b Something went wrong, critical data is missing
501bd7b6e4b08e76e648ccdc Something went wrong, critical data is missing
4d934737d176a1cd55a035f0 Something went wrong, critical data is missing
5c55a3ed464d65002cf55aea Something went wrong, critical data is missing
51e2a265498e6f1748fa0794 Something went wrong, critical data is missing
5d90614c7223320007ad49a0 Something went wrong, critical data is missing
56192801498e53358f49f687 Something went wrong, critical data is missing
59cf9b920c9f3155d652f56c Something went wrong, critical data is missing
5965e239135b391718065382 Something went wrong, critical data is missing
56f29bc0498ef1b351d85581 Something went wrong, critical data is missing
4bf536676a31d13a6693962e Something went wrong, critical data is missing
4fae71f2e4b0f4503a3ecc6b Something went wrong, critical data is missing
4be14bc4d816c92854a7efd9 Something went wrong, critical data is missing
50aa3212e4b00340979caebf Something went wrong, critical data is missing
59f47607f62f2b383336762b Something went wrong, critical data is missing
4bb08ffef964a5201f4c3ce3 Something went wrong, critical data is missing
5815e34638fac9dab7d64ff5 Something went wrong, critical data is missing
57ffa75e38fa774bca04d69e Something went wrong, critical data is missing
4f6de4d7e4b08a0bba90116d Something went wrong, critical data is missing
4ae97969f964a520e6b421e3 Something went wrong, critical data is missing
50e55d04e4b01244a9f7617f Something went wrong, critical data is missing
4b264571f964a520b37924e3 Something went wrong, critical data is missing
4bf24a58324cc9b6c4f5cc92 Something went wrong, critical data is missing
4d4170c3bd53f04d815d4d15 Something went wrong, critical data is missing
580de55c38fa89037956ab51 Something went wrong, critical data is missing
4b8fc239f964a520b56033e3 Something went wrong, critical data is missing
4e48dbb1fa76a07fde78c109 Something went wrong, critical data is missing
56d1687acd10a99383caa3de Something went wrong, critical data is missing
4b977f6ff964a520150635e3 Something went wrong, critical data is missing
4c35eed3452620a1747a270f Something went wrong, critical data is missing
59c38413bb8d3664aa5beef2 Something went wrong, critical data is missing
5d442a460372ce00070e1829 Something went wrong, critical data is missing
50ab4f9be4b0e77721143877 Something went wrong, critical data is missing

50645842e4b03558bb3552bd Something went wrong, critical data is missing
5bde4abb6e69af00257a60c0 Something went wrong, critical data is missing
59281f57780eee608ab3450e Something went wrong, critical data is missing
4bf536676a31d13a6693962e Something went wrong, critical data is missing
56f29bc0498ef1b351d85581 Something went wrong, critical data is missing
4fae71f2e4b0f4503a3ecc6b Something went wrong, critical data is missing
56192801498e53358f49f687 Something went wrong, critical data is missing
59cf9b920c9f3155d652f56c Something went wrong, critical data is missing
504494d4e4b0aa9bdce06d77 Something went wrong, critical data is missing
5965e239135b391718065382 Something went wrong, critical data is missing
4bf24a58324cc9b6c4f5cc92 Something went wrong, critical data is missing
4b264571f964a520b37924e3 Something went wrong, critical data is missing
59f47607f62f2b383336762b Something went wrong, critical data is missing
4cac1b4d36fa6dcb80e6db78 Something went wrong, critical data is missing
4cac512344a8224bbe1a3240 Something went wrong, critical data is missing
5066c3d2e4b007c93710590a Something went wrong, critical data is missing
4d91d34a033ba35d5dbd770d Something went wrong, critical data is missing
4e48dbb1fa76a07fde78c109 Something went wrong, critical data is missing
580de55c38fa89037956ab51 Something went wrong, critical data is missing
500846f2e4b0117eceecdce1 Something went wrong, critical data is missing
4bb08ffef964a5201f4c3ce3 Something went wrong, critical data is missing
4d4170c3bd53f04d815d4d15 Something went wrong, critical data is missing
4ae97969f964a520e6b421e3 Something went wrong, critical data is missing
56d1687acd10a99383caa3de Something went wrong, critical data is missing
5cdef4efe55d8b002cbfa441 Something went wrong, critical data is missing
56e96acd498ea60d05b6fcff Something went wrong, critical data is missing
4c6269e886b6be9a40048c34 Something went wrong, critical data is missing
4ba22d52f964a52081e137e3 Something went wrong, critical data is missing
4bb9dbda53649c74534e48fb Something went wrong, critical data is missing
4bc5694941cb76b0055f3e6f Something went wrong, critical data is missing
5c7ac51c64c8e1002c698544 Something went wrong, critical data is missing
4b8ea85af964a520ff2e33e3 Something went wrong, critical data is missing
4cb045f4562d224ba7e01488 Something went wrong, critical data is missing
4bcf2fc99854d13ac58cf54d Something went wrong, critical data is missing
4bb0a164f964a520a5503ce3 Something went wrong, critical data is missing
5c77ad413149b9002cdfb69c Something went wrong, critical data is missing
4bc816e66501c9b6489b3f29 Something went wrong, critical data is missing
4b115b12f964a520d17a23e3 Something went wrong, critical data is missing
4f59d8bee4b0a5edd03198db Something went wrong, critical data is missing
4ff54bf0e4b009b18e5387a0 Something went wrong, critical data is missing
51b433282fc6c7927e38f886 Something went wrong, critical data is missing
4e79e12c2271920e473bcfda Something went wrong, critical data is missing
59088e50c876c8497c9b2886 Something went wrong, critical data is missing
4af560b1f964a520f9f821e3 Something went wrong, critical data is missing
5cb200be3fcee8002c14275a Something went wrong, critical data is missing
53f31a8d498ec7b047200ae3 Something went wrong, critical data is missing
561ba93f498e228017f39041 Something went wrong, critical data is missing
4c3638df3849c92816a9bbb1 Something went wrong, critical data is missing

```
5af1e191cbcdee002c36f193 Something went wrong, critical data is missing
4bd97c8c5cf276b072aa9d00 Something went wrong, critical data is missing
5a16d2f2a4ba7c160bcb8c60 Something went wrong, critical data is missing
593bdcd167af3a7da0951229 Something went wrong, critical data is missing
5b5f10633af988002c559704 Something went wrong, critical data is missing
4c0768680ed3c9289b7f797d Something went wrong, critical data is missing
5a88264195d986447ea4e22e Something went wrong, critical data is missing
4c651051ae719521c276967b Something went wrong, critical data is missing
```

[9]:

```
                                      neighborhood  \
0  Islington, London, Greater London, United Kingdom
1  Islington, London, Greater London, United Kingdom
2  Islington, London, Greater London, United Kingdom
3  Islington, London, Greater London, United Kingdom
4  Islington, London, Greater London, United Kingdom

                   venue_id               venue_name  venue_latitude  \
0  58a576b8f8572431aec041b8                Kobo Cafe       51.534988
1  584bfe4544587f042f5ff30c              Katsute 100       51.534286
2  4fc9ff8ce4b087d229f75ce4  The Coffeeworks Project       51.534254
3  50338ecbe4b0c160f73b46d0                 Giddy Up       51.536374
4  5a9ea6c2d03360028695111e              Six Degrees       51.535228

   venue_longitude venue_category  \
0        -0.104149           Café
1        -0.104540       Tea Room
2        -0.104684    Coffee Shop
3        -0.102930    Coffee Shop
4        -0.103486    Coffee Shop

                                        canonicalUrl   categories  tier  \
0  https://foursquare.com/v/kobo-cafe/58a576b8f85…         Café     1
1  https://foursquare.com/v/katsute-100/584bfe454…     Tea Room     2
2  https://foursquare.com/v/the-coffeeworks-proje…  Coffee Shop     2
3  https://foursquare.com/v/giddy-up/50338ecbe4b0…  Coffee Shop     1
4  https://foursquare.com/v/six-degrees/5a9ea6c2d…  Coffee Shop     1

    message  rating  photos  tips
0     Cheap     8.5      28    15
1  Moderate     8.1      64    17
2  Moderate     8.1     360   140
3     Cheap     7.8      26     2
4     Cheap     7.6       6     2
```

As you can see above, we have added to the dataframe URL of the venue, it's rating, categories, number of tiers and photos.

5. Now let's get the items and prices from the Foursquare's menu for each venue.

[13]:

```
headers = {"User-Agent":"Mozilla/5.0 (Macintosh; Intel Mac OS X 10.14; rv:66.0)␣
↪Gecko/20100101 Firefox/66.0",
           "Accept":"text/html,application/xhtml+xml,application/xml;q=0.9,*/*;
↪q=0.8","Accept-Language":"en-US,en;q=0.5",
           "Accept-Encoding":"gzip, deflate","DNT":"1","Connection":
↪"close","Upgrade-Insecure-Requests":"1"}
df_wm = pd.DataFrame([])
for i in df.index:
    r = requests.get(df['canonicalUrl'][i] + '/menu',headers = headers)
    if r:
        #print(i,r.status_code,'Venue''s menu exists on the Foursquare Website.
↪')
        df_wm = pd.concat([df_wm,df.iloc[[i]]],ignore_index = True)
    #else:
        #print(i,r.status_code,'Venue''s menu does not exist on the Foursquare␣
↪Website.')
    time.sleep(random.randint(5,10))

df_menu = pd.DataFrame([])
for i in df_wm.index:
    url = df_wm.loc[i,'canonicalUrl']+'/menu'
    time.sleep(random.randint(5,10))
    r = requests.get(url,headers = headers)
    soup = BeautifulSoup(r.text,features = 'lxml')
    menu_items = [i.get_text() for i in soup.find_all(class_ = 'title')]
    menu_prices = [i.get_text() for i in soup.find_all(class_ = 'entryPrice')]
    df_menu = pd.concat([df_menu,pd.DataFrame({'venue_id':df_wm.
↪loc[i,'venue_id'],'venue_name':df_wm.loc[i,'venue_name']
                                               ,'item':menu_items,'price':
↪menu_prices})],ignore_index = True)
    #print(i,r.status_code,df_wm.loc[i,'venue_name'])
```

```
0 404 Venues menu does not exist on the Foursquare Website.
1 404 Venues menu does not exist on the Foursquare Website.
2 404 Venues menu does not exist on the Foursquare Website.
3 404 Venues menu does not exist on the Foursquare Website.
4 404 Venues menu does not exist on the Foursquare Website.
5 404 Venues menu does not exist on the Foursquare Website.
6 404 Venues menu does not exist on the Foursquare Website.
7 404 Venues menu does not exist on the Foursquare Website.
8 404 Venues menu does not exist on the Foursquare Website.
9 404 Venues menu does not exist on the Foursquare Website.
10 404 Venues menu does not exist on the Foursquare Website.
11 404 Venues menu does not exist on the Foursquare Website.
12 404 Venues menu does not exist on the Foursquare Website.
13 404 Venues menu does not exist on the Foursquare Website.
14 404 Venues menu does not exist on the Foursquare Website.
```

```
15 404 Venues menu does not exist on the Foursquare Website.
16 404 Venues menu does not exist on the Foursquare Website.
17 404 Venues menu does not exist on the Foursquare Website.
18 404 Venues menu does not exist on the Foursquare Website.
19 404 Venues menu does not exist on the Foursquare Website.
20 404 Venues menu does not exist on the Foursquare Website.
21 200 Venues menu exists on the Foursquare Website.
22 404 Venues menu does not exist on the Foursquare Website.
23 404 Venues menu does not exist on the Foursquare Website.
24 404 Venues menu does not exist on the Foursquare Website.
25 404 Venues menu does not exist on the Foursquare Website.
26 404 Venues menu does not exist on the Foursquare Website.
27 404 Venues menu does not exist on the Foursquare Website.
28 404 Venues menu does not exist on the Foursquare Website.
29 404 Venues menu does not exist on the Foursquare Website.
30 404 Venues menu does not exist on the Foursquare Website.
31 404 Venues menu does not exist on the Foursquare Website.
32 200 Venues menu exists on the Foursquare Website.
33 404 Venues menu does not exist on the Foursquare Website.
34 404 Venues menu does not exist on the Foursquare Website.
35 404 Venues menu does not exist on the Foursquare Website.
36 404 Venues menu does not exist on the Foursquare Website.
37 404 Venues menu does not exist on the Foursquare Website.
38 404 Venues menu does not exist on the Foursquare Website.
39 404 Venues menu does not exist on the Foursquare Website.
40 404 Venues menu does not exist on the Foursquare Website.
41 404 Venues menu does not exist on the Foursquare Website.
42 404 Venues menu does not exist on the Foursquare Website.
43 404 Venues menu does not exist on the Foursquare Website.
44 404 Venues menu does not exist on the Foursquare Website.
45 404 Venues menu does not exist on the Foursquare Website.
46 404 Venues menu does not exist on the Foursquare Website.
47 404 Venues menu does not exist on the Foursquare Website.
48 404 Venues menu does not exist on the Foursquare Website.
49 404 Venues menu does not exist on the Foursquare Website.
50 404 Venues menu does not exist on the Foursquare Website.
51 404 Venues menu does not exist on the Foursquare Website.
52 404 Venues menu does not exist on the Foursquare Website.
53 404 Venues menu does not exist on the Foursquare Website.
54 404 Venues menu does not exist on the Foursquare Website.
55 404 Venues menu does not exist on the Foursquare Website.
56 404 Venues menu does not exist on the Foursquare Website.
57 404 Venues menu does not exist on the Foursquare Website.
58 404 Venues menu does not exist on the Foursquare Website.
59 404 Venues menu does not exist on the Foursquare Website.
60 404 Venues menu does not exist on the Foursquare Website.
61 404 Venues menu does not exist on the Foursquare Website.
62 404 Venues menu does not exist on the Foursquare Website.
```

63 404 Venues menu does not exist on the Foursquare Website.
64 404 Venues menu does not exist on the Foursquare Website.
65 404 Venues menu does not exist on the Foursquare Website.
66 200 Venues menu exists on the Foursquare Website.
67 404 Venues menu does not exist on the Foursquare Website.
68 404 Venues menu does not exist on the Foursquare Website.
69 404 Venues menu does not exist on the Foursquare Website.
70 404 Venues menu does not exist on the Foursquare Website.
71 404 Venues menu does not exist on the Foursquare Website.
72 404 Venues menu does not exist on the Foursquare Website.
73 404 Venues menu does not exist on the Foursquare Website.
74 404 Venues menu does not exist on the Foursquare Website.
75 404 Venues menu does not exist on the Foursquare Website.
76 200 Venues menu exists on the Foursquare Website.
77 404 Venues menu does not exist on the Foursquare Website.
78 404 Venues menu does not exist on the Foursquare Website.
79 404 Venues menu does not exist on the Foursquare Website.
80 404 Venues menu does not exist on the Foursquare Website.
81 404 Venues menu does not exist on the Foursquare Website.
82 404 Venues menu does not exist on the Foursquare Website.
83 404 Venues menu does not exist on the Foursquare Website.
84 404 Venues menu does not exist on the Foursquare Website.
85 404 Venues menu does not exist on the Foursquare Website.
86 404 Venues menu does not exist on the Foursquare Website.
87 404 Venues menu does not exist on the Foursquare Website.
88 404 Venues menu does not exist on the Foursquare Website.
89 404 Venues menu does not exist on the Foursquare Website.
90 404 Venues menu does not exist on the Foursquare Website.
91 404 Venues menu does not exist on the Foursquare Website.
92 404 Venues menu does not exist on the Foursquare Website.
93 404 Venues menu does not exist on the Foursquare Website.
94 404 Venues menu does not exist on the Foursquare Website.
95 404 Venues menu does not exist on the Foursquare Website.
96 404 Venues menu does not exist on the Foursquare Website.
97 200 Venues menu exists on the Foursquare Website.
98 200 Venues menu exists on the Foursquare Website.
99 404 Venues menu does not exist on the Foursquare Website.
100 404 Venues menu does not exist on the Foursquare Website.
101 404 Venues menu does not exist on the Foursquare Website.
102 404 Venues menu does not exist on the Foursquare Website.
103 404 Venues menu does not exist on the Foursquare Website.
104 404 Venues menu does not exist on the Foursquare Website.
105 404 Venues menu does not exist on the Foursquare Website.
106 404 Venues menu does not exist on the Foursquare Website.
107 404 Venues menu does not exist on the Foursquare Website.
108 404 Venues menu does not exist on the Foursquare Website.
109 404 Venues menu does not exist on the Foursquare Website.
110 404 Venues menu does not exist on the Foursquare Website.

```
111 404 Venues menu does not exist on the Foursquare Website.
112 404 Venues menu does not exist on the Foursquare Website.
113 404 Venues menu does not exist on the Foursquare Website.
114 200 Venues menu exists on the Foursquare Website.
115 404 Venues menu does not exist on the Foursquare Website.
116 404 Venues menu does not exist on the Foursquare Website.
117 404 Venues menu does not exist on the Foursquare Website.
118 404 Venues menu does not exist on the Foursquare Website.
119 404 Venues menu does not exist on the Foursquare Website.
120 404 Venues menu does not exist on the Foursquare Website.
121 404 Venues menu does not exist on the Foursquare Website.
122 404 Venues menu does not exist on the Foursquare Website.
123 404 Venues menu does not exist on the Foursquare Website.
124 404 Venues menu does not exist on the Foursquare Website.
125 404 Venues menu does not exist on the Foursquare Website.
126 404 Venues menu does not exist on the Foursquare Website.
127 404 Venues menu does not exist on the Foursquare Website.
128 404 Venues menu does not exist on the Foursquare Website.
129 404 Venues menu does not exist on the Foursquare Website.
130 404 Venues menu does not exist on the Foursquare Website.
131 404 Venues menu does not exist on the Foursquare Website.
132 404 Venues menu does not exist on the Foursquare Website.
133 404 Venues menu does not exist on the Foursquare Website.
134 404 Venues menu does not exist on the Foursquare Website.
135 404 Venues menu does not exist on the Foursquare Website.
136 404 Venues menu does not exist on the Foursquare Website.
137 404 Venues menu does not exist on the Foursquare Website.
138 404 Venues menu does not exist on the Foursquare Website.
139 404 Venues menu does not exist on the Foursquare Website.
140 404 Venues menu does not exist on the Foursquare Website.
141 200 Venues menu exists on the Foursquare Website.
142 200 Venues menu exists on the Foursquare Website.
143 404 Venues menu does not exist on the Foursquare Website.
144 404 Venues menu does not exist on the Foursquare Website.
145 404 Venues menu does not exist on the Foursquare Website.
146 404 Venues menu does not exist on the Foursquare Website.
147 404 Venues menu does not exist on the Foursquare Website.
148 404 Venues menu does not exist on the Foursquare Website.
149 404 Venues menu does not exist on the Foursquare Website.
150 404 Venues menu does not exist on the Foursquare Website.
151 404 Venues menu does not exist on the Foursquare Website.
152 404 Venues menu does not exist on the Foursquare Website.
153 404 Venues menu does not exist on the Foursquare Website.
154 404 Venues menu does not exist on the Foursquare Website.
155 404 Venues menu does not exist on the Foursquare Website.
156 404 Venues menu does not exist on the Foursquare Website.
157 404 Venues menu does not exist on the Foursquare Website.
158 404 Venues menu does not exist on the Foursquare Website.
```

```
159 404 Venues menu does not exist on the Foursquare Website.
160 404 Venues menu does not exist on the Foursquare Website.
161 404 Venues menu does not exist on the Foursquare Website.
162 404 Venues menu does not exist on the Foursquare Website.
163 404 Venues menu does not exist on the Foursquare Website.
164 404 Venues menu does not exist on the Foursquare Website.
165 404 Venues menu does not exist on the Foursquare Website.
166 404 Venues menu does not exist on the Foursquare Website.
167 404 Venues menu does not exist on the Foursquare Website.
168 404 Venues menu does not exist on the Foursquare Website.
169 404 Venues menu does not exist on the Foursquare Website.
170 404 Venues menu does not exist on the Foursquare Website.
171 404 Venues menu does not exist on the Foursquare Website.
172 404 Venues menu does not exist on the Foursquare Website.
173 404 Venues menu does not exist on the Foursquare Website.
174 404 Venues menu does not exist on the Foursquare Website.
175 404 Venues menu does not exist on the Foursquare Website.
176 404 Venues menu does not exist on the Foursquare Website.
177 404 Venues menu does not exist on the Foursquare Website.
178 404 Venues menu does not exist on the Foursquare Website.
179 404 Venues menu does not exist on the Foursquare Website.
180 404 Venues menu does not exist on the Foursquare Website.
181 404 Venues menu does not exist on the Foursquare Website.
182 404 Venues menu does not exist on the Foursquare Website.
183 404 Venues menu does not exist on the Foursquare Website.
184 404 Venues menu does not exist on the Foursquare Website.
185 404 Venues menu does not exist on the Foursquare Website.
186 404 Venues menu does not exist on the Foursquare Website.
187 404 Venues menu does not exist on the Foursquare Website.
188 404 Venues menu does not exist on the Foursquare Website.
189 200 Venues menu exists on the Foursquare Website.
190 404 Venues menu does not exist on the Foursquare Website.
191 404 Venues menu does not exist on the Foursquare Website.
192 200 Venues menu exists on the Foursquare Website.
193 404 Venues menu does not exist on the Foursquare Website.
194 404 Venues menu does not exist on the Foursquare Website.
195 404 Venues menu does not exist on the Foursquare Website.
196 404 Venues menu does not exist on the Foursquare Website.
197 404 Venues menu does not exist on the Foursquare Website.
198 404 Venues menu does not exist on the Foursquare Website.
199 404 Venues menu does not exist on the Foursquare Website.
200 404 Venues menu does not exist on the Foursquare Website.
201 404 Venues menu does not exist on the Foursquare Website.
202 404 Venues menu does not exist on the Foursquare Website.
203 404 Venues menu does not exist on the Foursquare Website.
204 200 Venues menu exists on the Foursquare Website.
205 404 Venues menu does not exist on the Foursquare Website.
206 404 Venues menu does not exist on the Foursquare Website.
```

```
207 200 Venues menu exists on the Foursquare Website.
208 404 Venues menu does not exist on the Foursquare Website.
209 404 Venues menu does not exist on the Foursquare Website.
210 200 Venues menu exists on the Foursquare Website.
211 404 Venues menu does not exist on the Foursquare Website.
212 404 Venues menu does not exist on the Foursquare Website.
213 404 Venues menu does not exist on the Foursquare Website.
214 404 Venues menu does not exist on the Foursquare Website.
215 404 Venues menu does not exist on the Foursquare Website.
216 404 Venues menu does not exist on the Foursquare Website.
217 404 Venues menu does not exist on the Foursquare Website.
218 404 Venues menu does not exist on the Foursquare Website.
219 404 Venues menu does not exist on the Foursquare Website.
220 404 Venues menu does not exist on the Foursquare Website.
221 404 Venues menu does not exist on the Foursquare Website.
222 404 Venues menu does not exist on the Foursquare Website.
223 404 Venues menu does not exist on the Foursquare Website.
224 200 Venues menu exists on the Foursquare Website.
225 404 Venues menu does not exist on the Foursquare Website.
226 404 Venues menu does not exist on the Foursquare Website.
227 404 Venues menu does not exist on the Foursquare Website.
228 404 Venues menu does not exist on the Foursquare Website.
229 404 Venues menu does not exist on the Foursquare Website.
230 404 Venues menu does not exist on the Foursquare Website.
231 404 Venues menu does not exist on the Foursquare Website.
232 404 Venues menu does not exist on the Foursquare Website.
233 404 Venues menu does not exist on the Foursquare Website.
234 404 Venues menu does not exist on the Foursquare Website.
235 404 Venues menu does not exist on the Foursquare Website.
236 404 Venues menu does not exist on the Foursquare Website.
237 404 Venues menu does not exist on the Foursquare Website.
238 404 Venues menu does not exist on the Foursquare Website.
239 200 Venues menu exists on the Foursquare Website.
240 404 Venues menu does not exist on the Foursquare Website.
241 404 Venues menu does not exist on the Foursquare Website.
242 404 Venues menu does not exist on the Foursquare Website.
243 404 Venues menu does not exist on the Foursquare Website.
244 404 Venues menu does not exist on the Foursquare Website.
245 404 Venues menu does not exist on the Foursquare Website.
246 404 Venues menu does not exist on the Foursquare Website.
247 404 Venues menu does not exist on the Foursquare Website.
248 200 Venues menu exists on the Foursquare Website.
249 404 Venues menu does not exist on the Foursquare Website.
250 404 Venues menu does not exist on the Foursquare Website.
251 200 Venues menu exists on the Foursquare Website.
252 404 Venues menu does not exist on the Foursquare Website.
253 404 Venues menu does not exist on the Foursquare Website.
254 404 Venues menu does not exist on the Foursquare Website.
```

```
255 404 Venues menu does not exist on the Foursquare Website.
256 404 Venues menu does not exist on the Foursquare Website.
257 404 Venues menu does not exist on the Foursquare Website.
258 404 Venues menu does not exist on the Foursquare Website.
259 200 Venues menu exists on the Foursquare Website.
260 404 Venues menu does not exist on the Foursquare Website.
261 404 Venues menu does not exist on the Foursquare Website.
262 404 Venues menu does not exist on the Foursquare Website.
263 404 Venues menu does not exist on the Foursquare Website.
264 404 Venues menu does not exist on the Foursquare Website.
265 404 Venues menu does not exist on the Foursquare Website.
266 404 Venues menu does not exist on the Foursquare Website.
267 200 Venues menu exists on the Foursquare Website.
268 404 Venues menu does not exist on the Foursquare Website.
269 404 Venues menu does not exist on the Foursquare Website.
270 404 Venues menu does not exist on the Foursquare Website.
271 404 Venues menu does not exist on the Foursquare Website.
272 404 Venues menu does not exist on the Foursquare Website.
273 404 Venues menu does not exist on the Foursquare Website.
274 404 Venues menu does not exist on the Foursquare Website.
275 404 Venues menu does not exist on the Foursquare Website.
276 404 Venues menu does not exist on the Foursquare Website.
277 404 Venues menu does not exist on the Foursquare Website.
278 404 Venues menu does not exist on the Foursquare Website.
279 404 Venues menu does not exist on the Foursquare Website.
280 404 Venues menu does not exist on the Foursquare Website.
281 404 Venues menu does not exist on the Foursquare Website.
282 404 Venues menu does not exist on the Foursquare Website.
283 404 Venues menu does not exist on the Foursquare Website.
284 404 Venues menu does not exist on the Foursquare Website.
285 404 Venues menu does not exist on the Foursquare Website.
286 404 Venues menu does not exist on the Foursquare Website.
287 200 Venues menu exists on the Foursquare Website.
288 404 Venues menu does not exist on the Foursquare Website.
289 404 Venues menu does not exist on the Foursquare Website.
290 404 Venues menu does not exist on the Foursquare Website.
291 404 Venues menu does not exist on the Foursquare Website.
292 404 Venues menu does not exist on the Foursquare Website.
293 404 Venues menu does not exist on the Foursquare Website.
294 404 Venues menu does not exist on the Foursquare Website.
295 404 Venues menu does not exist on the Foursquare Website.
296 404 Venues menu does not exist on the Foursquare Website.
297 200 Venues menu exists on the Foursquare Website.
298 404 Venues menu does not exist on the Foursquare Website.
299 404 Venues menu does not exist on the Foursquare Website.
300 404 Venues menu does not exist on the Foursquare Website.
301 404 Venues menu does not exist on the Foursquare Website.
302 200 Venues menu exists on the Foursquare Website.
```

```
303 404 Venues menu does not exist on the Foursquare Website.
304 404 Venues menu does not exist on the Foursquare Website.
305 404 Venues menu does not exist on the Foursquare Website.
306 200 Venues menu exists on the Foursquare Website.
307 404 Venues menu does not exist on the Foursquare Website.
308 404 Venues menu does not exist on the Foursquare Website.
309 404 Venues menu does not exist on the Foursquare Website.
310 200 Venues menu exists on the Foursquare Website.
311 404 Venues menu does not exist on the Foursquare Website.
312 404 Venues menu does not exist on the Foursquare Website.
313 404 Venues menu does not exist on the Foursquare Website.
314 404 Venues menu does not exist on the Foursquare Website.
315 404 Venues menu does not exist on the Foursquare Website.
316 404 Venues menu does not exist on the Foursquare Website.
317 404 Venues menu does not exist on the Foursquare Website.
318 404 Venues menu does not exist on the Foursquare Website.
319 404 Venues menu does not exist on the Foursquare Website.
320 404 Venues menu does not exist on the Foursquare Website.
321 404 Venues menu does not exist on the Foursquare Website.
322 404 Venues menu does not exist on the Foursquare Website.
323 404 Venues menu does not exist on the Foursquare Website.
324 404 Venues menu does not exist on the Foursquare Website.
0 200 Canonbury Kitchen
1 200 Nibbles
2 200 Costa Pronto
3 200 Jim's Cafe
4 200 Amhurst Cafe
5 200 Dilara's
6 200 Starbucks
7 200 WFM Coffee Bar
8 200 Cafe Bar
9 200 IWM The Tea Room
10 200 espressamente illy
11 200 Urban Baristas
12 200 Bouquets and Beans
13 200 Flamingo
14 200 Enough To Feed An Elephant
15 200 Coffee Circus
16 200 Chez Nous
17 200 The Cakehouse by Foodilicious
18 200 Curators Coffee
19 200 Sista Barista @ the smokehouse
20 200 Espressamente Illy
21 200 The Borough Produce Cafe
22 200 Costa Coffee
23 200 The Kennington Coffee Shop
24 200 Black Sheep Coffee
```

6. Data cleaning. Let's remove the rows without prices, convert prices in decimal numbers, etc.

[14]:
```
df_menu = df_menu.loc[(df_menu['price'] != "")]
df_menu.reset_index(inplace = True)
df_menu.drop('index',axis = 1,inplace = True)

df_venue_menu = pd.merge(df_wm,df_menu.drop('venue_name',axis = 1),how =␣
↪'inner',on = 'venue_id',sort = False)
for i in df_venue_menu.index:
    if '.' in df_venue_menu.loc[i,'price']:
        df_venue_menu.loc[i,'price'] = df_venue_menu.loc[i,'price'][:
↪df_venue_menu.loc[i,'price'].find('.') + 3]
df_venue_menu.loc[:,'price'] = df_venue_menu.loc[:,'price'].apply(pd.
↪to_numeric,errors = 'coerce')
df_venue_menu.dropna(inplace = True)
df_venue_menu = df_venue_menu.loc[(df_menu['item'] != "House")]
df_venue_menu = df_venue_menu.loc[(df_menu['item'] != "Soup of the Day")]
df_venue_menu.reset_index(drop = True,inplace = True)
df_venue_menu.head()
```

[14]:
```
                                       neighborhood  \
0  Islington, London, Greater London, United Kingdom
1  Islington, London, Greater London, United Kingdom
2  Islington, London, Greater London, United Kingdom
3  Islington, London, Greater London, United Kingdom
4  Islington, London, Greater London, United Kingdom

                   venue_id         venue_name  venue_latitude  \
0  4d2dfe6494013704c6b0e1da  Canonbury Kitchen       51.543211
1  4d2dfe6494013704c6b0e1da  Canonbury Kitchen       51.543211
2  4d2dfe6494013704c6b0e1da  Canonbury Kitchen       51.543211
3  4d2dfe6494013704c6b0e1da  Canonbury Kitchen       51.543211
4  4d2dfe6494013704c6b0e1da  Canonbury Kitchen       51.543211

   venue_longitude venue_category  \
0        -0.102633           Café
1        -0.102633           Café
2        -0.102633           Café
3        -0.102633           Café
4        -0.102633           Café

                                        canonicalUrl categories  tier message  \
0  https://foursquare.com/v/canonbury-kitchen/4d2…       Café     1   Cheap
1  https://foursquare.com/v/canonbury-kitchen/4d2…       Café     1   Cheap
2  https://foursquare.com/v/canonbury-kitchen/4d2…       Café     1   Cheap
3  https://foursquare.com/v/canonbury-kitchen/4d2…       Café     1   Cheap
4  https://foursquare.com/v/canonbury-kitchen/4d2…       Café     1   Cheap
```

```
   rating  photos  tips                                              item  \
0     7.1      10     7                              Fresh Soup of the Day
1     7.1      10     7                 Tiger Prawns W/ White Wine & Chili
2     7.1      10     7               Grilled Whole Calamari W/ Salsa Verde
3     7.1      10     7  Beef Carpaccio W/ Shaved Parmesan & White Truf…
4     7.1      10     7   Warm Goats Cheese, Speck, Rocket & Mint Crostini

   price
0    6.0
1    8.5
2    8.5
3    9.5
4    9.5
```

7. Let's find top 10 of the most popular items across the analyzed venues.

[15]:
```python
mpi = df_venue_menu.copy()
mpi = mpi[['item','price','venue_id']].groupby(['item']).agg({'price':
 ↪['mean'],'venue_id':[pd.Series.nunique]}) # most popular items
mpi.columns = mpi.columns.map('-'.join).str.strip('-')
mpi.sort_values(by = 'venue_id-nunique',ascending = False,inplace = True)
mpi.reset_index(inplace = True)
mpi = mpi.head(10)
mpi
```

[15]:

	item	price-mean	venue_id-nunique
0	Americano	2.267143	11
1	Cappuccino	2.345000	10
2	Espresso	1.686667	10
3	Macchiato	1.856250	7
4	Mocha	2.617500	6
5	Hot Chocolate	2.492500	6
6	Latte	2.411250	6
7	Cheese	3.050000	4
8	Lasagne	6.225000	4
9	Mixed Salad	3.383333	3

All right, so we see, as the feasibility check, Americano, Cappuccino, and Espresso are the most popular items.

8. Let's look at the map of the analyzed coffee shops.

Blue markers represent all the venues which have been found around neighborhood center.

Black icons represent venues with published menus, which will be analyzed in the next section.

[16]:
```python
all_venues_map = folium.Map(location = [neighborhood_venue['venue_latitude'].
 ↪mean(),
```

```
                                                    neighborhood_venue['venue_longitude'].
 ↪mean()],zoom_start = 12)
for venue,lat,lng,v_id in␣
 ↪zip(neighborhood_venue['venue_name'],neighborhood_venue['venue_latitude'],
                                          ␣
 ↪neighborhood_venue['venue_longitude'],neighborhood_venue['venue_id']):
    label = folium.Popup('{}'.format(venue),parse_html = True)
    if v_id in df_venue_menu['venue_id'].unique():
        folium.Marker([lat,lng],icon = folium.Icon(color = 'black',icon =␣
 ↪'coffee',prefix = 'fa'),popup = label).add_to(all_venues_map)
    else:
        folium.CircleMarker([lat,lng],radius = 5,popup = label,color =␣
 ↪'blue',fill = True,fill_color = '#3186cc',fill_opacity = 0.7,
                            parse_html = False).add_to(all_venues_map)
all_venues_map
```

[16]:
```
<folium.folium.Map at 0x7fac416a0fd0>
```

0.7 ## Data Analysis

In this section we will analyze data. As was stated in the beginning, the aim of this research project is to determine if the connection between Foursquare Rating and availability of the most popular items in a coffee shop menu exists.

1. Let's prepare the data for the analysis.

As we want to analyze *popular* items, let's remove items which available only in three or less venues. Also, as we need to focus only on the availaibility of popular items, let's remove all columns, excluding popular items columns.

[17]:
```
df_vm_proc = df_venue_menu.copy()

df_vm_proc.drop_duplicates(keep = 'first',inplace = True)
df_vm_proc.reset_index(drop = True,inplace = True)

df_vm_proc = df_vm_proc.groupby(df_vm_proc.drop(['price'],axis = 1).columns.
 ↪values.tolist(),as_index = False).agg({'price':['min']})
df_vm_proc.columns = df_vm_proc.columns.map('-'.join).str.strip('-')
df_vm_proc.rename(columns = {'price-min':'price'},inplace = True)
df_vm_proc.reset_index(drop = True,inplace = True)

df_rep = df_vm_proc.copy()
df_rep = df_rep.pivot_table(index = ['item'],aggfunc = 'size')
df_rep = df_rep.to_frame()
df_rep.reset_index(level = df_rep.index.names,inplace = True)
df_rep.columns = ['item','count']

df_vm_proc = pd.merge(df_vm_proc,df_rep,how = 'inner',on = 'item',sort = False)
```

```
df_vm_proc = df_vm_proc.loc[df_vm_proc['count'] >= 4]
df_vm_proc.reset_index(drop = True,inplace = True)
```

[18]:
```
df_for_model = df_vm_proc.copy()
df_for_model =␣
 ↪df_for_model[['rating','neighborhood','venue_name','tips','message','photos','item','price'
df_for_model = pd.pivot_table(df_for_model,values = 'price',index =␣
 ↪['rating','neighborhood','venue_name','tips','message','photos']
                              ,columns = ['item'],aggfunc = np.sum,fill_value =␣
 ↪0)
df_for_model.reset_index(drop = False,inplace = True)
df_for_model.columns.name = None
```

In the ‘x’ dataframe we put all columns of the given dataframe except Foursquare Rating. The ‘y’ will reresent Foursquare Rating column.

Our goal is to find the connection between Foursquare Rating and available popular items in coffee shops (‘x’ dataframe).

[19]:
```
x = df_for_model.copy()
y = x.loc[:,'rating']
x.drop(['neighborhood','venue_name','message','rating','tips','photos'],axis =␣
 ↪1,inplace = True)
x
```

[19]:

	Americano	Cappuccino	Cheese	Espresso	Hot Chocolate	Lasagne	Latte	\
0	2.00	2.10	0.0	1.60	2.10	0.0	2.10	
1	0.00	2.55	0.6	1.75	2.55	0.0	0.00	
2	2.50	2.50	0.0	2.00	2.70	0.0	2.50	
3	2.00	2.20	0.0	1.50	0.00	0.0	2.20	
4	2.99	2.99	0.0	1.99	2.99	8.5	2.99	
5	2.00	2.50	0.0	1.50	0.00	0.0	2.50	
6	2.00	2.60	0.0	2.20	0.00	0.0	0.00	
7	2.20	2.30	0.0	1.30	0.00	0.0	2.30	
8	1.90	1.90	4.0	1.20	1.90	4.4	0.00	
9	1.95	0.00	3.8	0.00	0.00	6.0	0.00	
10	1.95	0.00	3.8	0.00	0.00	6.0	0.00	
11	2.50	2.00	0.0	0.00	2.50	0.0	0.00	
12	0.00	0.00	0.0	1.70	0.00	0.0	0.00	

	Macchiato	Mocha
0	1.80	0.00
1	1.75	0.00
2	2.50	2.50
3	1.50	2.75
4	2.60	2.99
5	1.70	2.80

```
6          0.00    0.00
7          1.50    2.40
8          0.00    0.00
9          0.00    0.00
10         0.00    0.00
11         0.00    2.50
12         0.00    0.00
```

2. Linear regression.

Let's construct multiple linear regression and check if we can predict venue rating, based on the available in the venue items.

[23]:
```
lr = linear_model.LinearRegression()
lr_fit = lr.fit(x,y)
rating_hat = lr.predict(x)

print ('intercept: %40.5f' % lr.intercept_)
for i,v in enumerate(lr.coef_):
        print('feature: %30s,   coefficient: %.5f' % (x.columns[i],v))

print("\n\nRating predictions:")
print(pd.concat([y,pd.DataFrame(rating_hat,columns = {'predicted␣
 ↪rating'})],axis = 1,sort = False))

print("\nmean absolute error: %.2f" % np.mean(np.absolute(rating_hat - y)))
print("residual sum of squares (MSE): %.2f" % np.mean((rating_hat - y) ** 2))
```

```
intercept:                                    8.59221
feature:                       Americano,   coefficient: -0.00129
feature:                      Cappuccino,   coefficient: -0.70410
feature:                          Cheese,   coefficient: -0.20854
feature:                        Espresso,   coefficient: 0.14962
feature:                   Hot Chocolate,   coefficient: 0.03346
feature:                         Lasagne,   coefficient: 0.10019
feature:                           Latte,   coefficient: -0.32538
feature:                       Macchiato,   coefficient: -0.69140
feature:                           Mocha,   coefficient: 0.61519


Rating predictions:
    rating  predicted rating
0      5.4          5.492871
1      5.8          5.808861
2      6.3          6.214366
3      6.7          7.203893
4      6.8          6.801386
5      7.0          6.787530
```

```
6       7.1           7.088145
7       7.1           6.855445
8       7.1           7.101769
9       8.4           8.398370
10      8.4           8.398370
11      8.8           8.802433
12      8.9           8.846562

mean absolute error: 0.09
residual sum of squares (MSE): 0.03
```

As we see above, our model quite well predict the rating at Foursquare, based on the popular items availability and their prices.

3. Let's construct Random Forest Regression. This analysis will help us to identify the most influenceable products for the coffee shop ratings.

For this we will use impurity-based feature importances. The higher, the more important the feature. The importance of a feature is computed as the (normalized) total reduction of the criterion brought by that feature. It is also known as the Gini importance. Gini Importance or Mean Decrease in Impurity (MDI) calculates each feature importance as the sum over the number of splits (across all tress) that include the feature, proportionally to the number of samples it splits.

[22]:
```
rf = RandomForestRegressor(random_state = 5)
rf.fit(x,y)
importance = rf.feature_importances_

l = []
for i,v in enumerate(importance):
    l.append([x.columns[i],v])

df_l = pd.DataFrame(l,columns = ['feature','importance'])
df_l.sort_values(by = 'importance',ascending = False,inplace = True)
df_l.reset_index(drop = True,inplace = True)
df_l

print("R2-score: %.5f" % rf.score(x,y))
pyplot.rcParams["figure.figsize"] = (12,6)
pyplot.bar(df_l['feature'],df_l['importance'])
pyplot.xticks(rotation = 'vertical')
pyplot.tight_layout()

pyplot.show()
df_l
```

```
R2-score: 0.87677

<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

[22]:

	feature	importance
0	Cappuccino	0.429172
1	Espresso	0.185238
2	Macchiato	0.128230
3	Cheese	0.059913
4	Hot Chocolate	0.058257
5	Americano	0.049117
6	Latte	0.046562
7	Mocha	0.032183
8	Lasagne	0.011328

0.8 ## Results

- Cappuccino has the highest influence power on the venue's rating and guest loyalty.
- Despite the fact, that Americano is the most popular item among the coffee shops, it is only on the fifth place of the products, ranked by importance.

0.9 Discussion

Clearly, before the research was obvious, that the coffee shops owners have to have available in café most popular coffee drinks, such as Cappuccino, Macchiato, Espresso, Hot Chocolate, and Americano. But which items are more important for their guests was undetermined.

Now, entrepreneurs precisely know that the mentioned above 5 items are most important and their prices have an influence on the overall customer's rating. So, the recommendation for business owners is to keep the mentioned items on stock and carefully track their prices.

0.10 Conclusion

To sum up, in this research, we have determined the connection between venue's rating and availability of the most popular items (incl. their prices). We have identified the most important for coffee shop's rating items and numerically described their influence on the overall venue's rating.