

# Moscow neighborhoods vs. NYC neighborhoods

Guretskiy Evgeny

May 29, 2020

## 1. Introduction

### 1.1 Background

Moscow and NYC are among the biggest capitals in the world. Moscow has area of 2511 sq. km., while NYC has 784. They are both densely populated cosmopolitan metropolises (In 2012 Moscow population was almost 12 mln, NYC 8.4 mln), but they also have their differences. Assume that we have been asked to find the better city for launching a new branch of fitness clubs. We have conducted an experiment where we compare the two cities in terms of their public venues, including coffee shops, cafes, restaurants, and etc. to get a good measure of each city's residents towards their choices of venue.

### 1.2 Business

As mentioned above, sport has a certain degree of importance for nowadays people. Moreover, people globally more and more tend to follow healthy lifestyle and become more aware of their state. Therefore, we will model a situation in which we have decided to open a new fitness center full of areas for different kinds of sports, spa-facilities etc.

### 1.3 Target audience

A target audience are those of 12 mln people in Moscow or 8.4 mln in NYC who keen on sport.

## 2. Data

### 2.2 Moscow Data

Link: [https://en.wikipedia.org/wiki/Administrative\\_divisions\\_of\\_Moscow](https://en.wikipedia.org/wiki/Administrative_divisions_of_Moscow)

As there were no .csv file with ready data, it was necessary to scrape the data. For this purpose, we gathered data of Moscow administrative division from Wikipedia webpage. We got a list of neighborhoods names, but the problem was that the name consisted of Russian name and name in English in brackets, so we had to separate languages using "re" library.

```
11j: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 25
```

```
[17]:
```

	Name_R	Name_E
0	Арбат	Arbat
1	Басманный	Basmanny
2	Хамовники	Khamovniki
3	Красносельский	Krasnoselsky
4	Мещанский	Meshchansky
...	...	...
142	Поселение Новофёдоровское	Novofyodorovskoye Settlement
143	Поселение Первомайское	Pervomayskoye Settlement
144	Поселение Роговское	Rogovskoye Settlement
145	Поселение Троицк	Troitsky Settlement
146	Поселение Щаповское	Shchapovskoye Settlement

147 rows × 2 columns

### Data after separating languages

The next problem related to coordinates absence. Using “geocode” and “nominatim” libraries we easily fulfilled our table.

```

In [1]: #get coords by address in dfc
locator = Nominatim(user_agent="Moscow")
# 1 - convenient function to delay between geocoding calls
geocode = RateLimiter(locator.geocode, min_delay_seconds=1)
# 2 - create location column
df3['location'] = df3['address'].apply(geocode)
# 3 - create longitude, latitude and altitude from location column (returns tuple)
df3['point'] = df3['location'].apply(lambda loc: tuple(loc.point) if loc else None)
# 4 - split point column into latitude, longitude and altitude columns
df3[['latitude', 'longitude', 'altitude']] = pd.DataFrame(df3['point'].tolist(), index=df3.index)

In [2]:

In [3]: df4=df3.dropna()
df4

In [4]:
```

	Name_R	location	point	latitude	longitude	altitude
0	Арбат	(район Арбат, Центральный административный окр...	(55.751199, 37.5898715, 0.0)	55.751199	37.589872	0.0
1	Басманный	(Басманный район, Центральный административный...	(55.7672809, 37.6697732, 0.0)	55.767281	37.669773	0.0
2	Хамовники	(район Хамовники, Центральный административный...	(55.729229, 37.5729484, 0.0)	55.729229	37.572948	0.0
3	Красносельский	(Красносельский район, Центральный администрат...	(55.7774472, 37.6541598, 0.0)	55.777447	37.654160	0.0
4	Мещанский	(Мещанский район, Центральный административный...	(55.779172, 37.6279886, 0.0)	55.779172	37.627989	0.0
...	...	...	...	...	...	...
142	Поселение Новофёдоровское	(поселение Новофёдоровское, Троицкий администр...	(55.43792585, 36.99971817597381, 0.0)	55.437926	36.999718	0.0
143	Поселение Первомайское	(поселение Первомайское, Троицкий администрати...	(55.50190525, 37.21279504750347, 0.0)	55.501905	37.212795	0.0
144	Поселение Роговское	(поселение Роговское, Троицкий административны...	(55.2290933, 37.05519465465914, 0.0)	55.229093	37.055195	0.0
145	Поселение Троицк	(городской округ Троицк, Троицкий администрати...	(55.479412249999996, 37.29803052194669, 0.0)	55.479412	37.298031	0.0
146	Поселение Щаповское	(поселение Щаповское, Троицкий административны...	(55.384577699999994, 37.38527174994181, 0.0)	55.384578	37.385272	0.0

147 rows × 6 columns

### Code for cords gathering and data after

## 2.3 NYC data

Link: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

NYC data was structured in .csv file that was used in previous labs of course. So there were no any problems downloading it.

## 2.4 Foursquare API

We actively used this data provider during tutorials in previous labs. Therefore, this service is our powerful tool for getting the list of venues for each neighborhood in Moscow and NYC lists.

### 3. Methodology

#### 3.2 Moscow data preparation

As it was mentioned, first we parse the data from Wikipedia webpage, then clean it and separate names of neighborhoods. The next step was to get coordinates using “geocode” library (picture above).

#### 3.3 NYC data preparation

Downloading .json file.

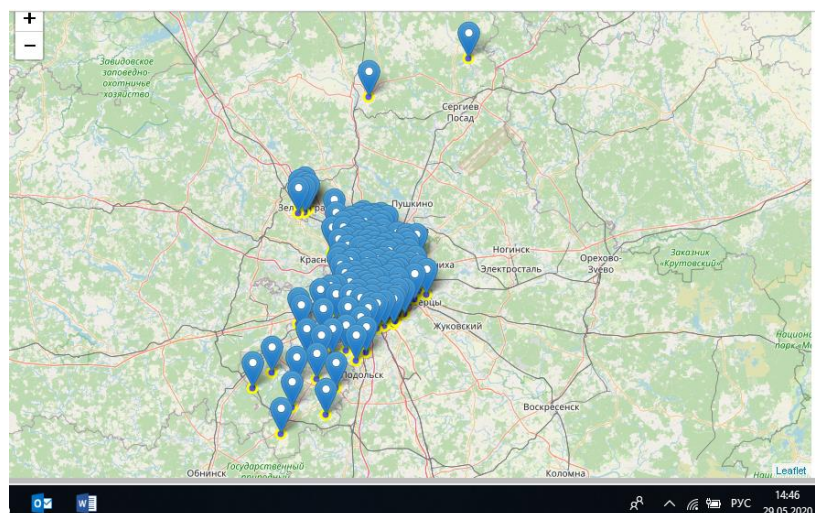
```
49]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

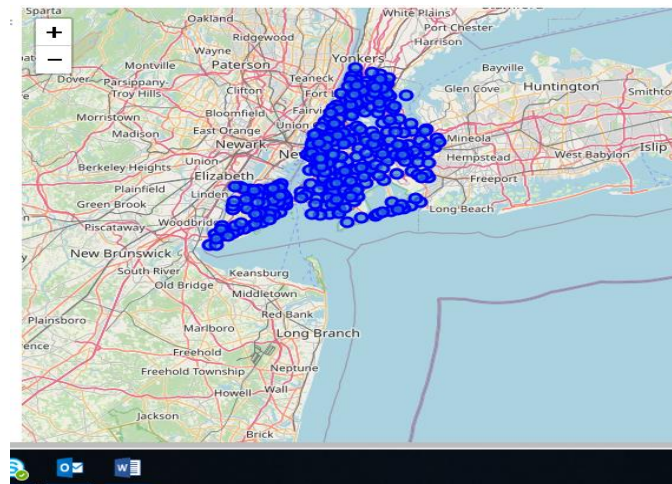
*NYC data after loading*

#### 3.4 Data visualization

Both prepared Moscow data and NYC data (latitude and longitude) were used to put point on the map using “folium” library, so we could better percept the data.



*Moscow map*



NYC map

### 3.5 API Calls to Foursquare

The Foursquare API is used to explore the neighborhoods and segment them. To access the API, 'CLIENT\_ID', 'CLIENT\_SECRET' and 'VERSION' are set. To overcome the redundancy of the process followed above, a function 'getNearbyVenues' is created. This function loops through all the neighborhoods of Moscow and New York City and creates an API request URL with radius = 500, LIMIT = 100. By limit, it is defined that the maximum of 100 nearby venues should be returned. Further, the GET request is made to Foursquare API and only relevant information for each nearby venue is extracted from it. The data is then appended to a python 'list'. Lastly, the python 'list' is unfolded or flattened to append it to the data frame being returned by the function.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list = []
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&lat={}&lng={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([
            name,
            lat,
            lng,
            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 = ['Name',
                             'Name_Latitude',
                             'Name_Longitude',
                             'Venue',
                             'Venue_Latitude',
                             'Venue_Longitude',
                             'Venue_Category']

    return(nearby_venues)
```

The returned data frames are as follows:

	Name	Name_Latitude	Name_Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Арбат	55.751199	37.589872	Шоколадница	55.751886	37.588327	Coffee Shop
1	Арбат	55.751199	37.589872	Corner Café & Kitchen	55.751496	37.586757	Japanese Restaurant
2	Арбат	55.751199	37.589872	Театр им. Вахтангова	55.749569	37.591638	Theater
3	Арбат	55.751199	37.589872	Спортивный комплекс «Арбат»	55.750928	37.587130	Pool
4	Арбат	55.751199	37.589872	Obedbufet (Обедбуфет)	55.752268	37.592275	Buffet
...	...	...	...	...	...	...	...
2748	Поселение Троицк	55.479412	37.298031	Т-клуб	55.475971	37.298819	Gym
2749	Поселение Троицк	55.479412	37.298031	Дворец спорта «Квант»	55.477341	37.298947	Stadium
2750	Поселение Троицк	55.479412	37.298031	Физическая Кунсткамера	55.475638	37.295264	Science Museum
2751	Поселение Троицк	55.479412	37.298031	Городской Рынок	55.476763	37.297657	Farmers Market
2752	Поселение Троицк	55.479412	37.298031	Центральный стадион	55.475311	37.297908	Soccer Field

[illegible]

Now, data allows us to count top 5 venues both for Moscow and NYC, and conduct high-level description.

```

]:

```

	count	mean	std	min	25%	50%	75%	max
Supermarket	137.0	0.861314	1.237778	0.0	0.0	1.0	1.0	9.0
Coffee Shop	137.0	0.795620	1.591054	0.0	0.0	0.0	1.0	9.0
Clothing Store	137.0	0.350365	0.895998	0.0	0.0	0.0	0.0	7.0
Cosmetics Shop	137.0	0.569343	1.069604	0.0	0.0	0.0	1.0	6.0
Hotel	137.0	0.262774	0.788428	0.0	0.0	0.0	0.0	6.0
Theater	137.0	0.145985	0.612854	0.0	0.0	0.0	0.0	6.0
Gym / Fitness Center	137.0	0.474453	0.832080	0.0	0.0	0.0	1.0	5.0
Arcade	137.0	0.153285	0.554268	0.0	0.0	0.0	0.0	5.0
Café	137.0	0.532847	0.970530	0.0	0.0	0.0	1.0	5.0
Bar	137.0	0.175182	0.554365	0.0	0.0	0.0	0.0	4.0

Moscow data description

```

]:

```

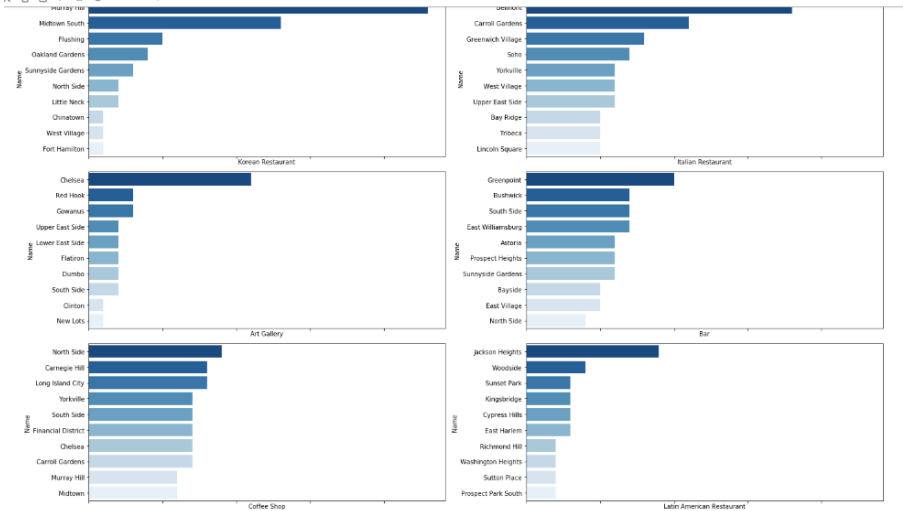
	count	mean	std	min	25%	50%	75%	max
Korean Restaurant	301.0	0.232558	1.589257	0.0	0.0	0.0	0.0	23.0
Italian Restaurant	301.0	0.916944	1.810086	0.0	0.0	0.0	1.0	18.0
Art Gallery	301.0	0.169435	0.762362	0.0	0.0	0.0	0.0	11.0
Bar	301.0	0.710963	1.402205	0.0	0.0	0.0	1.0	10.0
Coffee Shop	301.0	1.003322	1.774821	0.0	0.0	0.0	1.0	9.0
Latin American Restaurant	301.0	0.275748	0.783396	0.0	0.0	0.0	0.0	9.0
Beach	301.0	0.146179	0.786486	0.0	0.0	0.0	0.0	9.0
Deli / Bodega	301.0	0.930233	1.210420	0.0	0.0	1.0	1.0	8.0
Park	301.0	0.596007	1.163842	0.0	0.0	0.0	1.0	8.0
Gym / Fitness Center	301.0	0.385382	0.947445	0.0	0.0	0.0	0.0	8.0

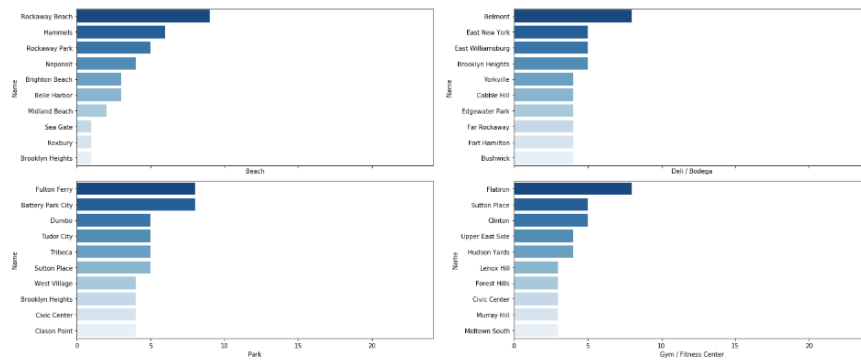
NYC data description

4. Data Visualization

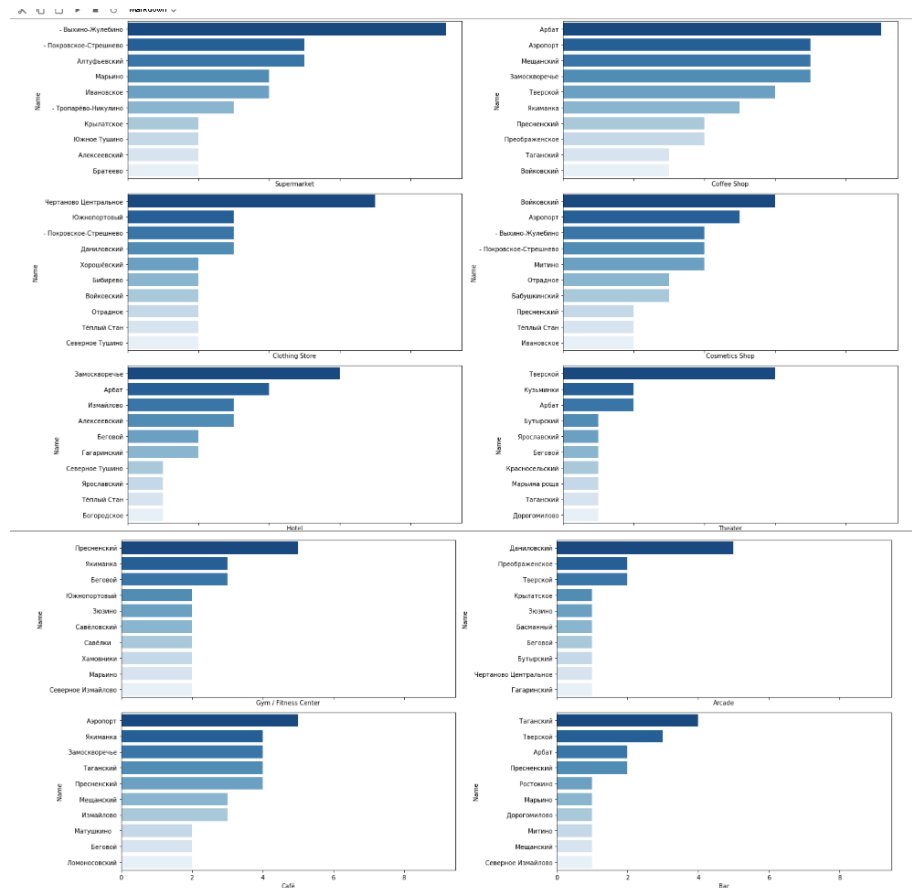
These top 10 categories are further plotted individually on bar graph using python ‘seaborn’ library. A code block is created which loops and plots the graph of top 10 neighborhoods for a category.

NYC graphs:





## Moscow graphs:



Next, the rows of the neighborhoods are grouped together and the frequency of occurrence of each category is calculated by taking the mean. As the limit is set to be 100, there will be many venues being returned by the Foursquare API. The created data frame is then fed with the top 5 most common venues categories in the respective neighborhoods.

```

num_top_venues = 5

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighbourhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
nyc_venues_sorted = pd.DataFrame(columns=columns)
nyc_venues_sorted['Neighbourhood'] = nyc_grouped['Neighbourhood']

for ind in np.arange(nyc_grouped.shape[0]):
    nyc_venues_sorted.iloc[ind, 1:] = return_most_common_venues(nyc_grouped.iloc[ind, :], num_top_venues)

nyc_venues_sorted.head()

```

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allerton	Pizza Place	Bakery	Chinese Restaurant	Deli / Bodega	Department Store
1	Annadale	Pizza Place	American Restaurant	Dance Studio	Sports Bar	Train Station
2	Arden Heights	Pizza Place	Deli / Bodega	Pharmacy	Bus Stop	Business Service
3	Arlington	Grocery Store	Deli / Bodega	American Restaurant	Intersection	Bus Stop
4	Arrochar	Italian Restaurant	Bus Stop	Deli / Bodega	Pizza Place	Cosmetics Shop

NYC

## 5. ML

We used The Silhouette Method to determine number of cluster, as the elbow method was inefficient.

```

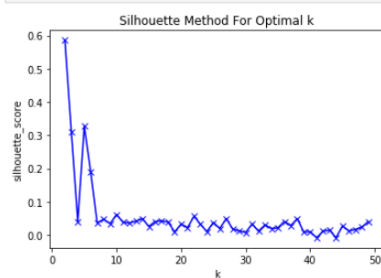
[72]: from sklearn.metrics import silhouette_score

sil = []
K_sil = range(2,50)
# minimum 2 clusters required, to define dissimilarity
for k in K_sil:
    print(k, end=" ")
    kmeans = KMeans(n_clusters = k).fit(nyc_grouped_clustering)
    labels = kmeans.labels_
    sil.append(silhouette_score(nyc_grouped_clustering, labels, metric = 'euclidean'))

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49

[73]: plt.plot(K_sil, sil, 'bx-')
plt.xlabel('k')
plt.ylabel('silhouette_score')
plt.title('Silhouette Method For Optimal k')
plt.show()

```



NYC silhouette score

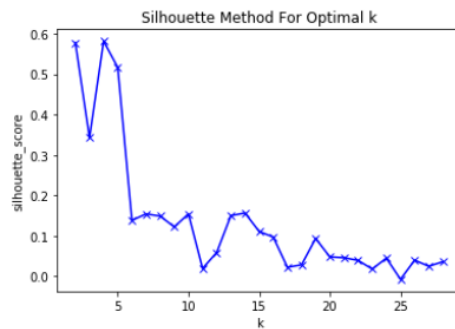


```
[102]: from sklearn.metrics import silhouette_score

sil = []
K_sil = range(2,29)
# minimum 2 clusters required, to define dissimilarity
for k in K_sil:
    print(k, end=' ')
    kmeans = KMeans(n_clusters = k).fit(moscow_grouped_clustering)
    labels = kmeans.labels_
    sil.append(silhouette_score(moscow_grouped_clustering, labels, metric = 'euclidean'))

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

[103]: plt.plot(K_sil, sil, 'bx-')
plt.xlabel('k')
plt.ylabel('silhouette_score')
plt.title('Silhouette Method For Optimal k')
plt.show()
```



## Moscow silhouette score

For NYC There is a peak at  $k = 2$ ,  $k = 5$  and  $k = 6$ . Two clusters will give a very broad classification of the venues. Therefore, the number of New York clusters (i.e. 'k') is chosen to be 5.

For Moscow there is a peak at  $k=4$ .

As we are going to find diversities of clusters then it will be more reasonable to use  $k=4$  for both NYC and Moscow.

```
.]: #Run k-means to cluster the neighborhood into 4 clusters.
# set number of clusters
kclusters = 4

moscow_grouped_clustering = moscow_grouped.drop('Name', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(moscow_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
# to change use .astype()
```

Further, the cluster labels curated are added to the data frame to get the desired results of segmenting the neighborhood based upon the most common venues in its vicinity. Moscow and New York City's neighborhoods are visualized utilizing the python 'folium' library as we did it before.

## 6. Results

### 6.1 NYC

## Cluster 1: Top 1 – Carrebian restaurant, Top 2 – cosmetic shops

```
nyc_cluster_1=nyc_merged.loc[nyc_merged['Cluster Labels1'] == 0, nyc_merged.columns[[0]+[1]+ list(range(5, nyc_merged.shape[1]))]]
nyc_cluster_1
```

	Borough	Neighborhood	Cluster Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Bronx	Williamsbridge	0.0	Caribbean Restaurant	Bar	Nightclub	Convenience Store	Soup Place
41	Bronx	Olinville	0.0	Caribbean Restaurant	Fried Chicken Joint	Laundromat	Basketball Court	Supermarket
165	Queens	St. Albans	0.0	Caribbean Restaurant	Dance Studio	Southern / Soul Food Restaurant	Shopping Mall	Grocery Store
168	Queens	Cambria Heights	0.0	Caribbean Restaurant	Cosmetics Shop	Restaurant	Liquor Store	Nightclub
188	Queens	Laurelton	0.0	Caribbean Restaurant	Cosmetics Shop	Sculpture Garden	Train Station	Yoga Studio
257	Staten Island	Howland Hook	0.0	0	0	0	0	0
300	Brooklyn	Erasmus	0.0	Caribbean Restaurant	Grocery Store	Yoga Studio	Convenience Store	Bank

```
: for col in nyc_required_column:
    print(nyc_cluster_1[col].value_counts(ascending=False))
    print('-----')
```

```
Queens      3
Bronx       2
Brooklyn    1
Staten Island 1
Name: Borough, dtype: int64
-----
Caribbean Restaurant    6
0                        1
Name: 1st Most Common Venue, dtype: int64
-----
Cosmetics Shop    2
Grocery Store     1
Fried Chicken Joint 1
0                 1
Bar               1
Dance Studio      1
Name: 2nd Most Common Venue, dtype: int64
-----
```

## Cluster 2: Top 1- Bus stops, Top – 2 parks

0]:	Borough	Neighborhood	Cluster Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
27	Bronx	Clason Point	1.0	Park	American Restaurant	Grocery Store	Boat or Ferry	Pool
77	Brooklyn	Manhattan Beach	1.0	Bus Stop	Café	Ice Cream Shop	Harbor / Marina	Playground
172	Queens	Breezy Point	1.0	Trail	Beach	Monument / Landmark	Yoga Studio	Filipino Restaurant
179	Queens	Neponsit	1.0	Beach	Yoga Studio	Ethiopian Restaurant	Event Space	Exhibit
198	Staten Island	New Brighton	1.0	Bus Stop	Park	Deli / Bodega	Bowling Alley	Discount Store
202	Staten Island	Grymes Hill	1.0	Dog Run	Bus Stop	Yoga Studio	Event Space	Exhibit
212	Staten Island	Oakwood	1.0	Bar	Bus Stop	Lawyer	Financial or Legal Service	Event Space
224	Staten Island	Park Hill	1.0	Bus Stop	Athletics & Sports	Hotel	Coffee Shop	Gym / Fitness Center
228	Staten Island	Arrochar	1.0	Bus Stop	Deli / Bodega	Italian Restaurant	Food Truck	Supermarket
229	Staten Island	Grasmere	1.0	Bus Stop	Bagel Shop	Bank	Grocery Store	Deli / Bodega
238	Staten Island	Butler Manor	1.0	Baseball Field	Pool	Bus Stop	Convenience Store	Yoga Studio
241	Staten Island	Arden Heights	1.0	Pizza Place	Pharmacy	Coffee Shop	Bus Stop	Lawyer
245	Staten Island	Bloomfield	1.0	Recreation Center	Park	Theme Park	Bus Stop	Discount Store
256	Staten Island	Randall Manor	1.0	Bus Stop	Pizza Place	Bagel Shop	Carpet Store	Playground
285	Staten Island	Willowbrook	1.0	Bus Stop	Chinese Restaurant	Pizza Place	Yoga Studio	Filipino Restaurant
286	Staten Island	Sandy Ground	1.0	Bus Stop	Food Truck	Art Gallery	Intersection	Racetrack
302	Queens	Hammels	1.0	Beach	Neighborhood	Diner	Building	Gym / Fitness Center
303	Queens	Bayswater	1.0	Playground	Park	Yoga Studio	Event Service	Event Space
305	Staten Island	Fox Hills	1.0	Bus Stop	Print Shop	Sandwich Place	Yoga Studio	Field

```
: for col in nyc_required_column:
    print(nyc_cluster_2[col].value_counts(ascending=False))
    print('-----')
```

```
Staten Island    13
Queens           4
Brooklyn         1
Bronx            1
Name: Borough, dtype: int64
-----
Bus Stop         9
Beach            2
Playground       1
Park             1
Baseball Field   1
Trail            1
Bar              1
Dog Run          1
Pizza Place      1
Recreation Center 1
Name: 1st Most Common Venue, dtype: int64
-----
Park             3
Bus Stop         2
Bagel Shop       1
Beach            1
```

### Cluster 3: Top 1 – Pizza place, Top 2 – coffee shop

cluster 2:Bus stops (9) and parks (3)

```
112]: nyc_cluster_3=nyc_merged.loc[nyc_merged['Cluster Labels1'] == 2, nyc_merged.columns[[0]*[1] + list(range(5, nyc_merged.shape[1]))]]
nyc_cluster_3
```

	Borough	Neighborhood	Cluster Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronx	Wakefield	2.0	Pharmacy	Ice Cream Shop	Dessert Shop	Sandwich Place	Gas Station
1	Bronx	Co-op City	2.0	Bus Station	Baseball Field	Park	Gift Shop	Pharmacy
2	Bronx	Eastchester	2.0	Bus Station	Caribbean Restaurant	Deli / Bodega	Bus Stop	Diner
3	Bronx	Fieldston	2.0	Plaza	River	Bus Station	Filipino Restaurant	Event Service
4	Bronx	Riverdale	2.0	Park	Bus Station	Home Service	Food Truck	Bank
...	...	...	...	...	...	...	...	...
297	Bronx	Bronxdale	2.0	Performing Arts Venue	Pizza Place	Paper / Office Supplies Store	Italian Restaurant	Gym
298	Bronx	Allerton	2.0	Pizza Place	Supermarket	Deli / Bodega	Bus Station	Chinese Restaurant
299	Bronx	Kingsbridge Heights	2.0	Pizza Place	Coffee Shop	Chinese Restaurant	Grocery Store	Mexican Restaurant
301	Manhattan	Hudson Yards	2.0	Gym / Fitness Center	Hotel	Italian Restaurant	American Restaurant	Dog Run
304	Queens	Queensbridge	2.0	Hotel	Sandwich Place	Gym / Fitness Center	Scenic Lookout	Performing Arts Venue

278 rows × 8 columns

Queens	73
Brooklyn	68
Bronx	49
Staten Island	48
Manhattan	40
Name: Borough, dtype: int64	
-----	
Pizza Place	27
Deli / Bodega	23
Italian Restaurant	19
Coffee Shop	17
Chinese Restaurant	15
..	
Plaza	1
Wine Bar	1
Liquor Store	1
Ice Cream Shop	1
Vegetarian / Vegan Restaurant	1
Name: 1st Most Common Venue, Length: 78, dtype: int64	
-----	
Pizza Place	28
Coffee Shop	19
Deli / Bodega	17
Italian Restaurant	14
Chinese Restaurant	11

### Cluster 4: Top 1 – parks, Top 2- yoga

Cluster 3: definitely pizza and coffee

```
]: nyc_cluster_4=nyc_merged.loc[nyc_merged['Cluster Labels1'] ==3, nyc_merged.columns[[0]*[1] + list(range(5, nyc_merged.shape[1]))]]
nyc_cluster_4
```

	Borough	Neighborhood	Cluster Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
192	Queens	Somerville	3.0	Park	Yoga Studio	Ethiopian Restaurant	Event Space	Exhibit
203	Staten Island	Todt Hill	3.0	Park	Yoga Studio	Ethiopian Restaurant	Event Space	Exhibit

```
]: for col in nyc_required_column:
    print(nyc_cluster_4[col].value_counts(ascending=False))
    print('-----')

Queens      1
Staten Island  1
Name: Borough, dtype: int64
-----
Park        2
Name: 1st Most Common Venue, dtype: int64
-----
Yoga Studio    2
Name: 2nd Most Common Venue, dtype: int64
-----
```

## 6.2 Moscow

### Cluster 1: Top 1-lakes, Top 2 - zoo

3]:

	location	altitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
111	(Внуково, Дмитровский городской округ, Московс...	0.0	0	Lake	Zoo Exhibit	Fabric Shop	Farmers Market	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop	Food
124	(район Силино, Зеленоградский административный...	0.0	0	Lake	Zoo Exhibit	Fabric Shop	Farmers Market	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop	Food

1]:

```
for col in moscow_required_column:
    print(moscow_cluster_1[col].value_counts(ascending=False))
    print('-----')
```

(Внуково, Дмитровский городской округ, Московская область, Центральный федеральный округ, Россия, (56.3651669, 37.58425)) 1  
(район Силино, Зеленоградский административный округ, Москва, Центральный федеральный округ, Россия, (56.0017567, 37.1688978)) 1  
Name: location, dtype: int64  
-----  
Lake 2  
Name: 1st Most Common Venue, dtype: int64  
-----  
Zoo Exhibit 2  
Name: 2nd Most Common Venue, dtype: int64  
-----

### Cluster 2: Top 1- coffee shops, top 2 - parks

```
1]: moscow_cluster_2=moscow_merged.loc[moscow_merged['Cluster_Labels'] == 1, moscow_merged.columns[[1] + list(range(5, moscow_merged.shape[1]))]]
2: moscow_cluster_2
```

1]:

	location	altitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	(район Арбат, Центральный административный окр...	0.0	1	Coffee Shop	Hotel	Yoga Studio	Hostel	Museum	Caucasian Restaurant	Burger Joint	Italian Restaurant	Concert Hall	Bar
1	(Басманный район, Центральный административный...	0.0	1	History Museum	Gun Range	Vegetarian / Vegan Restaurant	Flower Shop	Dance Studio	Supermarket	Park	Gift Shop	Massage Studio	Pet Café
2	(район Хамовники, Центральный административный...	0.0	1	Gym / Fitness Center	Park	Bath House	Coffee Shop	Bakery	Sushi Restaurant	BBQ Joint	Seafood Restaurant	Supermarket	Farmers Market
3	(Красносельский район, Центральный администрат...	0.0	1	Gym / Fitness Center	Hotel Bar	Coffee Shop	Varenyky restaurant	Gaming Cafe	Bus Stop	Flower Shop	Mobile Phone Shop	Café	Fast Food Restaurant
4	(Мещанский район, Центральный административный...	0.0	1	Coffee Shop	Café	Middle Eastern Restaurant	Bakery	Restaurant	Cosmetics Shop	Spa	Caucasian Restaurant	Museum	Health Food Store
...	...	...	...	...	...	...	...	...	...	...	...	...	...
129	(поселение Кокоскино, Новоомосковский администрат...	0.0	1	Farmers Market	Zoo Exhibit	Food & Drink Shop	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop	Food	Food Court
131	(поселение Московский, Новоомосковский администрат...	0.0	1	Café	Resort	Wine Bar	Middle Eastern Restaurant	Bus Stop	Big Box Store	Zoo Exhibit	Food	Film Studio	Fish Market

Mode: Command Ln 1. Col 1 Capstone.iovnb

```
Python
1: (Зеленоград, Зеленоградский административный округ, Москва, Центральный федеральный округ, Россия, (55.9954233, 37.198534))
2: (Измайлово, Окружной проезд, Соколиная Гора, район Соколиная Гора, Восточный административный округ, Москва, Центральный федеральный округ, 105613, Р
   оссия, (55.7897928, 37.7431771))
3: (район Южное Бутово, Юго-Западный административный округ, Москва, Центральный федеральный округ, Россия, (55.5412721, 37.5325392))
4: (Матушкино, район Матушкино, Зеленоградский административный округ, Москва, Центральный федеральный округ, 103489, Россия, (56.0095351, 37.1973782))
5: (Северное Бутово, район Северное Бутово, Юго-Западный административный округ, Москва, Центральный федеральный округ, 117628, Россия, (55.5714094, 37.
   5776237))
6: Name: location, Length: 110, dtype: int64
7: -----
8: Coffee Shop 11
9: Park 8
10: Gym / Fitness Center 6
11: Café 6
12: Cosmetics Shop 5
13: Pharmacy 5
14: Fast Food Restaurant 4
15: Supermarket 4
16: Mobile Phone Shop 4
17: Bus Stop 4
18: Soccer Field 3
19: Auto Workshop 3
20: Pizza Place 3
21: Restaurant 2
22: Sushi Restaurant 2
23: Music Venue 2
24: Electronics Store 2
25: Clothing Store 2
26: Flower Shop 2
27: Sporting Goods Shop 2
28: ..
29: ..
30: ..
```

## Cluster 3: Top 1 – supermarkets, top 2 – gyms. That is what we have been looking for.

```
4]: moscow_cluster_3=moscow_merged.loc[moscow_merged['Cluster_Labels'] == 2, moscow_merged.columns[[1] + list(range(5, moscow_merged.shape[1]))]]
moscow_cluster_3
```

	location	altitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	(Дмитровский, Москва, Северный административны...	0.0	2	Supermarket	Gym	Pizza Place	Shipping Store	Eastern European Restaurant	Bus Stop	Modern European Restaurant	Bath House	Park	Restaurant
16	(Коврино, улица Дыбенко, район Ховрино, Северн...	0.0	2	Supermarket	Auto Workshop	Park	Campground	French Restaurant	Fountain	Forest	Food Truck	Food Court	Food & Drink Shop
23	(Восточное Дегунино, район Восточное Дегунино, ...	0.0	2	Supermarket	Bus Stop	Park	Pedestrian Plaza	Toy / Game Store	Business Service	Pet Service	Soccer Field	Health Food Store	Film Studio
27	(Алтуфьевский, Москва, Северо-Восточный админи...	0.0	2	Supermarket	Dry Cleaner	Gym Pool	Grocery Store	Café	Health Food Store	Pharmacy	Salon / Barbershop	Bus Stop	Food & Drink Shop
42	(район Южное Медведково, Северо-Восточный адми...	0.0	2	Supermarket	Flower Shop	Pharmacy	Gym / Fitness Center	Cafeteria	Bakery	BBQ Joint	Bus Stop	Gym	Sushi Restaurant
45	(район Ивановское, Восточный административный ...	0.0	2	Supermarket	Cosmetics Shop	Shopping Mall	Bus Stop	Event Space	Pizza Place	Fast Food Restaurant	Electronics Store	Pet Store	Auto Workshop

```
1
(Алтуфьевский, Москва, Северо-Восточный административный округ, Москва, Центральн
1
(поселение Сосенское, Новомосковский административный округ, Москва, Центральн
1
(Орехово-Борисово Южное, Москва, Южный административный округ, Москва, Центральн
1
Name: location, dtype: int64
-----
Supermarket      18
Pizza Place       1
Bus Stop          1
Name: 1st Most Common Venue, dtype: int64
-----
Gym                4
Cosmetics Shop     2
Bus Stop           2
Convenience Store  2
Auto Workshop      1
Korean Restaurant  1
Sushi Restaurant   1
Dry Cleaner        1
Supermarket        1
Flower Shop        1
Grocery Store      1
Bakery             1
Park               1
Gym / Fitness Center 1
Name: 2nd Most Common Venue, dtype: int64
-----
```

## Cluster 4: Top 1 – conveniences store, top 2 – bus stop.

```
Python
]: moscow_cluster_4=moscow_merged.loc[moscow_merged['Cluster_Labels'] == 3, moscow_merged.columns[[1] + list(range(5, moscow_merged.shape[1]))]]
moscow_cluster_4
```

	location	altitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
39	(Северный, Талдомский городской округ, Московс...	0.0	3	Convenience Store	Motel	Zoo Exhibit	Food Court	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop	Food
83	(Орехово-Борисово Северное, Москва, Южный адми...	0.0	3	Convenience Store	Bus Stop	Dairy Store	Grocery Store	Pharmacy	Playground	Dance Studio	Discount Store	Fast Food Restaurant	Fried Chicken Joint
105	(Ново-Переделкино, район Ново-Переделкино, зап...	0.0	3	Convenience Store	Bus Stop	Tennis Court	Pool	Hockey Rink	Zoo Exhibit	Food	Fast Food Restaurant	Film Studio	Fish Market
137	(поселение Вороновское, Троицкий административ...	0.0	3	Convenience Store	Zoo Exhibit	Food & Drink Shop	Farmers Market	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop	Food
143	(поселение Первомайское, Троицкий администрати...	0.0	3	Recreation Center	Beach	Convenience Store	Zoo Exhibit	Food & Drink Shop	Fast Food Restaurant	Film Studio	Fish Market	Flea Market	Flower Shop

```

9): for col in moscow_required_column:
    print(moscow_cluster_4[col].value_counts(ascending=False))
    print('-----')

(Ново-Переделкино, район Ново-Переделкино, Западный административный округ, Москва, Центральный федеральный округ, 119634, Россия, (55.6461043, 37.35
2684)) 1
(Орехово-Борисово Северное, Москва, Южный административный округ, Москва, Центральный федеральный округ, 115569, Россия, (55.619223, 37.7011025))
1
(Северный, Талдомский городской округ, Московская область, Центральный федеральный округ, 141912, Россия, (56.725784, 37.6219781))
1
(поселение Вороновское, Троицкий административный округ, Москва, Центральный федеральный округ, 108830, Россия, (55.321132399999996, 37.1219867578463
4)) 1
(поселение Первомайское, Троицкий административный округ, Москва, Центральный федеральный округ, Россия, (55.50190525, 37.21279504750347))
1
Name: location, dtype: int64
-----
Convenience Store 4
Recreation Center 1
Name: 1st Most Common Venue, dtype: int64
-----
Bus Stop 2
Beach 1
Zoo Exhibit 1
Motel 1
Name: 2nd Most Common Venue, dtype: int64
-----

```

## Conclusion

We have done a lot: loading data, formatting and wrangling it, getting coordinates, putting points on the map, preparation of data for ML, building graphs, conducting k-mean algorithm in order to define better place for launching a new branch of sport gyms. Therefore, the winner is Moscow, as the only cluster that had “gym” in top 2 attendance places was cluster 3 in Moscow.