Muhammed Göktepe
2017400162

Assignment 2

Describe the steps you have performed for data preprocessing.

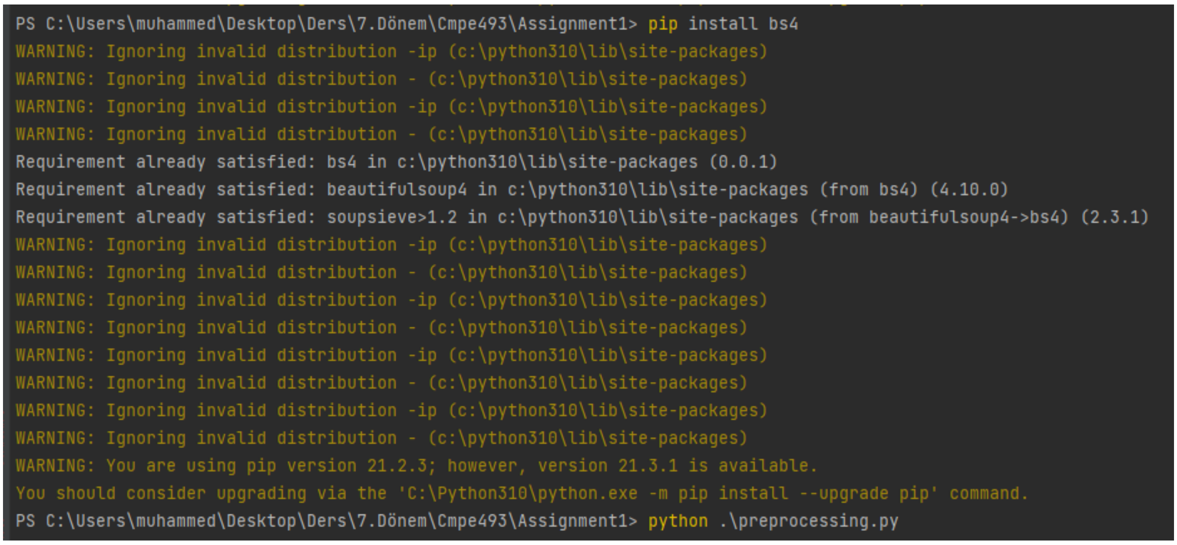
In this project we are assigned to implement a document retrieval system for simple boolean queries using the positional inverted indexing scheme. Reuters news is our dataset, and we have some stopwords to discard. First, I stored stopwords in a list for not to include these words while indexing texts. I have a dictionary of dictionary named index_dict. I store newId's of each word and their corresponding positions in documents in here. Then I started to traverse .sgm files in reuters21578 directory. I used BeautifulSoup library to ease .sgm file processing. Then, I store all articles in a file in news_list list. I iterate over each news and look for their titles and body's. If the article contains either of these, I combine the title and body and send it to the normalize function. In normalize function I made punctuation removal, case folding and finally stopwords removal. After these steps I added these words to dictionary and create a dictionary for each word. For these words I added their news ids and their positions in the document. And printed this dictionary to file.

Describe the data structures (hash, b-tree, linked list etc.) that you used for representing the inverted index (i.e., the dictionary and the postings lists).

I used dictionary of dictionary to store words, news ids and word positions in documents. Since .sgm files read in ascending order of news ids it stores new ids in ascending order. Also, positions are in ascending order in document. I used AND operator like algorithm while implementing phrase queries. In text free queries I used Lec03 as my reference.

Provide a screenshot of running the indexing module of your system.

Since I am using BeautifulSoup library it must be installed on working environment.



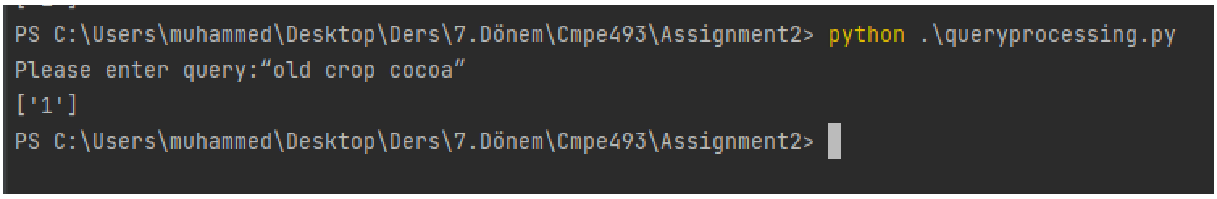
```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment1> pip install bs4
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
Requirement already satisfied: bs4 in c:\python310\lib\site-packages (0.0.1)
Requirement already satisfied: beautifulsoup4 in c:\python310\lib\site-packages (from bs4) (4.10.0)
Requirement already satisfied: soupsieve>1.2 in c:\python310\lib\site-packages (from beautifulsoup4->bs4) (2.3.1)
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -ip (c:\python310\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python310\lib\site-packages)
WARNING: You are using pip version 21.2.3; however, version 21.3.1 is available.
You should consider upgrading via the 'C:\Python310\python.exe -m pip install --upgrade pip' command.
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment1> python .\preprocessing.py
```

Muhammed Göktepe
2017400162

You can install it by typing “pip install bs4”. Then you can run my preprocessing code by typing “python preprocessing.py”.

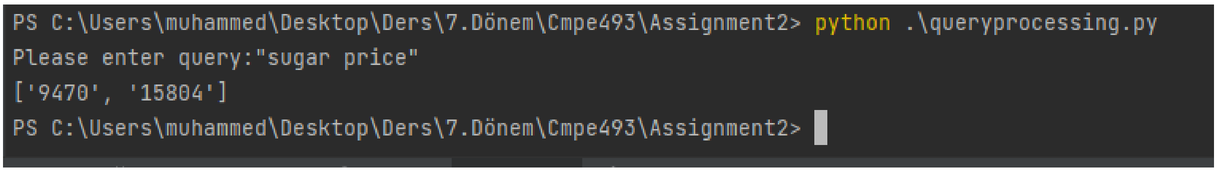
Provide four screenshots of running your system for each of the four types of queries.

“old crop cocoa”



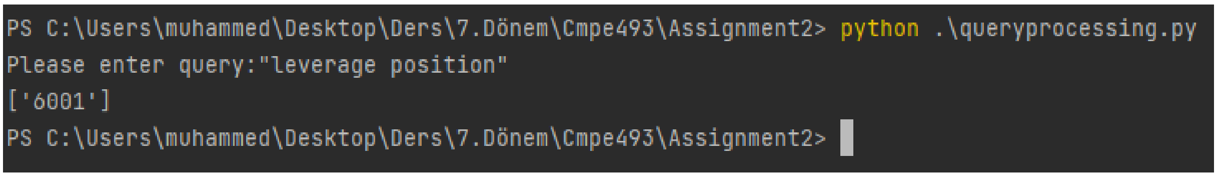
```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:“old crop cocoa”
['1']
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2>
```

"sugar price"



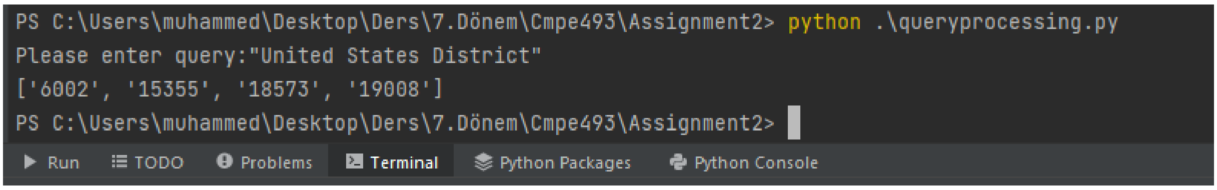
```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:"sugar price"
['9470', '15804']
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2>
```

"leverage position"



```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:"leverage position"
['6001']
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2>
```

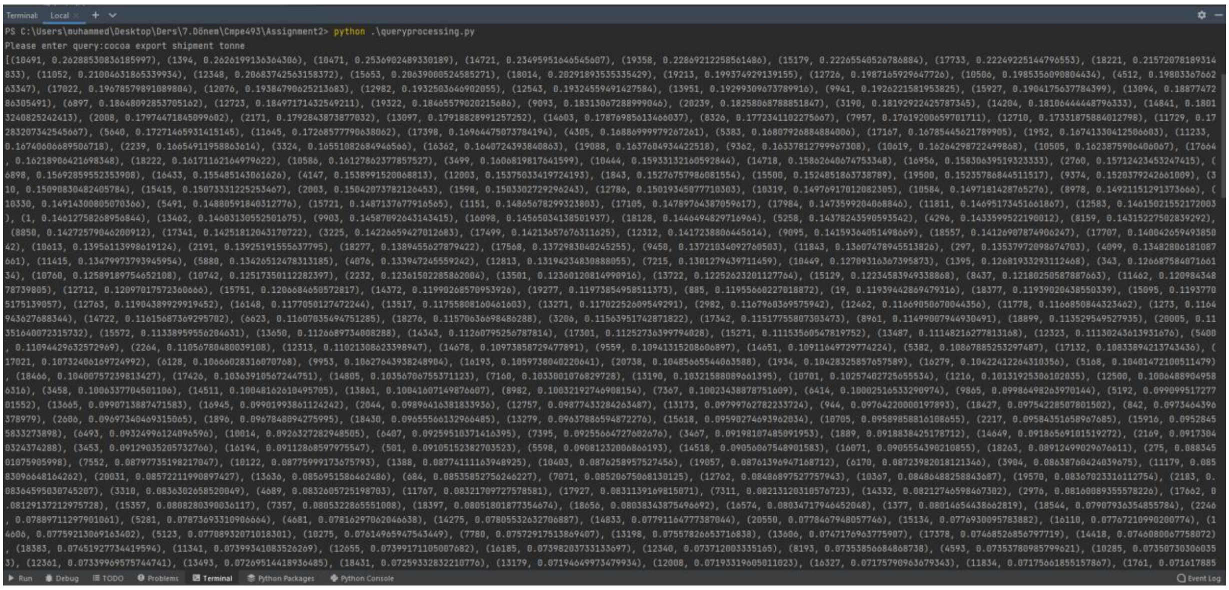
"United States District"



```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:"United States District"
['6002', '15355', '18573', '19008']
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2>
```

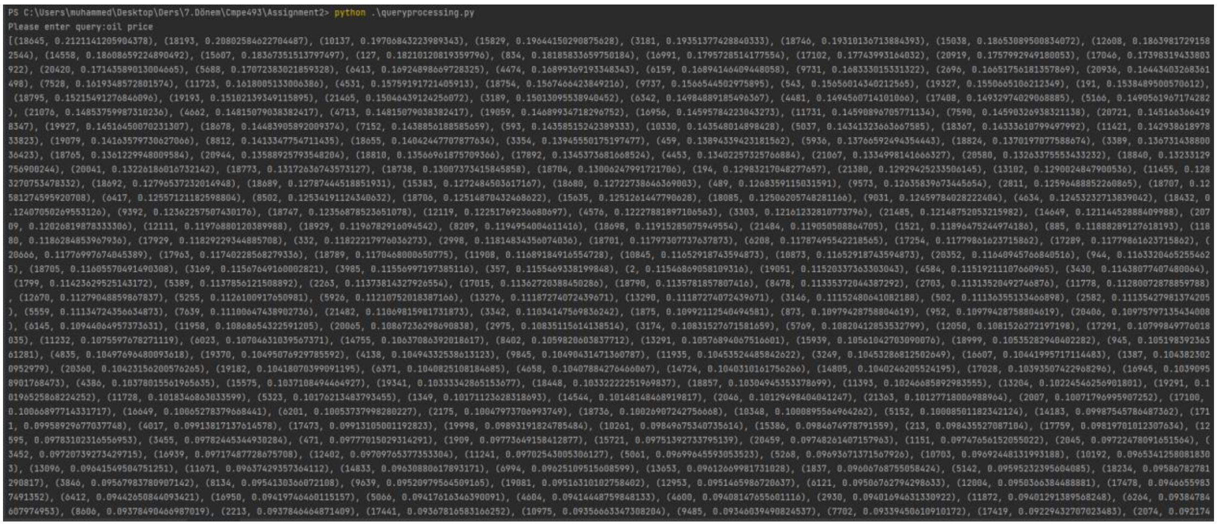
Muhammed Göktepe
2017400162

cocoa export shipment tonne



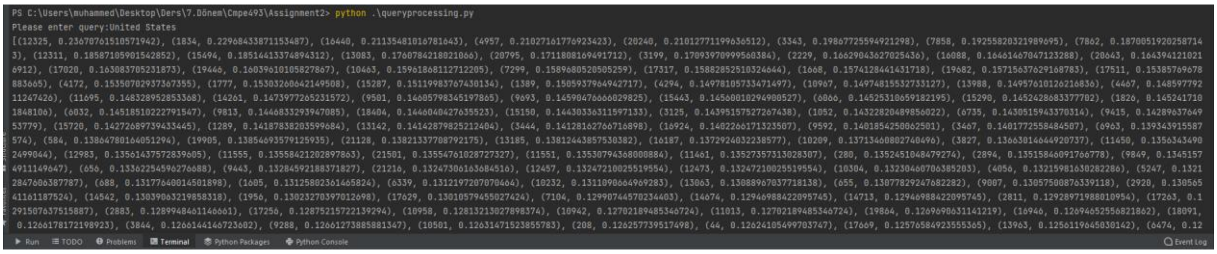
```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:cocoa export shipment tonne
[(18491, 0.26288530836185997), (1394, 0.2626199136364306), (18471, 0.2536902489330189), (14721, 0.23495951646545607), (19358, 0.22869212258561486), (15179, 0.22265540526786884), (17733, 0.22249225144796553), (18221, 0.21572078189314
833), (11052, 0.21004631865339934), (12348, 0.20683742563158372), (15653, 0.20639080524585271), (18014, 0.20291893535335429), (19213, 0.199374929139155), (12726, 0.1987165929647726), (10506, 0.1985356090804434), (4512, 0.19803367662
63347), (17022, 0.19678579891089804), (12076, 0.19384790625213683), (12982, 0.1932503646902055), (12543, 0.19324559491427584), (13951, 0.19299309673789916), (9941, 0.1926221581953825), (15927, 0.1904175637784399), (13094, 0.18877472
86305491), (6897, 0.18648092853705162), (12723, 0.18497171432549211), (19322, 0.18465579020215686), (9093, 0.18313067288999046), (20239, 0.18258068788851847), (3190, 0.18192922425787345), (14204, 0.18106444448796333), (14841, 0.1801
3240825242413), (2008, 0.17974471845099602), (2171, 0.1792843873877032), (13097, 0.17918828991257252), (14603, 0.17876985613466037), (8326, 0.1772341102275667), (7957, 0.17619200659701711), (12710, 0.17331875884012798), (11729, 0.17
283207342545667), (5640, 0.17271465931415145), (11645, 0.17268577790638062), (17398, 0.16964475073784194), (4305, 0.16886999979267261), (5383, 0.16807926884884006), (17167, 0.16785445621789905), (1952, 0.16741330412506603), (11233,
0.16740606689506718), (2239, 0.16654911958863614), (3324, 0.16551082684946566), (16362, 0.1640724393840863), (19088, 0.1637604934422518), (9362, 0.16337812799967308), (10619, 0.16264298722499868), (10505, 0.1623875906406067), (17664
, 0.16218906421698348), (18222, 0.16171162164979622), (10586, 0.16127862377857527), (3499, 0.1606819817641599), (10444, 0.15933132160592844), (14718, 0.15862640674753348), (16956, 0.15830639519323333), (2760, 0.15712423453247415), (
6898, 0.15692859552353908), (16433, 0.155485143061626), (4147, 0.1538991520068813), (12003, 0.15375033419724193), (1843, 0.15276757986081554), (15500, 0.1524851863738789), (19500, 0.15235786844511517), (9374, 0.15203792426610009), (3
10, 0.15090830482405784), (15415, 0.15073331225253467), (2003, 0.15042073782126453), (1598, 0.1503302729296243), (12786, 0.15019345077710303), (10319, 0.14976917012082305), (10584, 0.1497181428765276), (8978, 0.14921151291373666), (
10330, 0.14914300805070366), (5491, 0.14880591840312776), (15721, 0.14871376779165565), (1151, 0.14865678299323803), (17105, 0.14789764387059617), (17984, 0.14735992040068846), (11811, 0.14695173451661867), (12583, 0.14615021552172003
), (1, 0.14612758268956844), (13462, 0.14603130552501675), (9903, 0.14587092643143415), (16098, 0.14565034138501937), (18128, 0.1446494829716964), (5258, 0.14378243590593542), (4296, 0.1433599522190012), (8159, 0.14315227502839292),
 (8850, 0.14272579046200912), (17341, 0.14251812043170722), (3225, 0.14226659427012683), (17499, 0.14213657676311625), (12312, 0.1417238806445614), (9095, 0.14159364051498669), (18557, 0.14126907874906247), (17707, 0.140042659493850
42), (10613, 0.13956113998619124), (2191, 0.13925191555637795), (18277, 0.1389455627879422), (17568, 0.13729830402452555), (9450, 0.13721034092760503), (11843, 0.13607478945513826), (297, 0.13537972098674703), (4099, 0.13482806181087
661), (11415, 0.13479973793945954), (5880, 0.13426512478313185), (4076, 0.133947245559242), (12813, 0.13194234830888055), (7215, 0.1301279439711459), (10449, 0.12709316367395873), (1395, 0.12681933293112468), (343, 0.126687584071661
34), (10760, 0.12589189754652108), (10742, 0.12517350112282397), (2232, 0.12361502285862004), (13501, 0.12360120814990916), (13722, 0.12252623201127764), (15129, 0.12234583949338868), (8437, 0.12180250587887663), (11462, 0.120984348
78739805), (12712, 0.12097017572360666), (15751, 0.12066846505728817), (14372, 0.11990268570953926), (19277, 0.11973854958511373), (885, 0.11955660227018872), (19, 0.11939442869479316), (18377, 0.11939020438550339), (15095, 0.1193770
5175139057), (12763, 0.11904389929919452), (16148, 0.1177050127472244), (13517, 0.11755808160461603), (13271, 0.11702252609549291), (2982, 0.1167960369575942), (12462, 0.11669050670044356), (11778, 0.11668508443234462), (1273, 0.1164
943627688344), (14722, 0.11615687369295702), (6623, 0.11607035494751285), (18276, 0.11570636698486288), (3206, 0.11563951742871822), (17342, 0.11517755807303473), (8961, 0.11499007944930491), (18899, 0.113529549527935), (20005, 0.11
351640072315732), (15572, 0.11338959556204631), (13650, 0.1126689734008288), (14343, 0.11260795256787814), (17301, 0.11252736399794028), (15271, 0.11153560547819752), (13487, 0.11148216277813168), (12323, 0.11130243613931676), (5400
, 0.11094429632572969), (2264, 0.11056780480039108), (12313, 0.11021308623398947), (14678, 0.10973858729477891), (9559, 0.10941315208606897), (14651, 0.10911649729774224), (5382, 0.10867885253297487), (17132, 0.10833894213743436), (
17021, 0.10732406169724992), (6128, 0.10666028316070768), (9953, 0.10627643938248904), (16193, 0.1059738040220641), (20738, 0.10485665544063588), (1934, 0.10428325857657589), (16279, 0.10422412264310356), (5168, 0.10401472100511479)
, (18466, 0.10400757239813427), (17426, 0.10363910567244751), (14805, 0.10356706755371123), (7160, 0.1033001076829728), (13190, 0.10321588089661395), (10701, 0.10257402725655534), (1216, 0.10131925306102035), (12500, 0.1006488904958
6316), (3458, 0.1006337704501106), (14511, 0.10048162610495705), (13861, 0.10041607149876607), (8982, 0.10032192746908154), (7367, 0.10023438878751609), (6414, 0.10002516533290974), (9865, 0.09986498263970144), (5192, 0.099099517277
01552), (13665, 0.0990713887471583), (16945, 0.09901993861124242), (2044, 0.09896416381833936), (12757, 0.09877433284263487), (13173, 0.09799762782233724), (944, 0.09764220000197893), (18427, 0.09754228507801502), (842, 0.0973464396
378979), (2686, 0.09697340469315065), (1896, 0.0967848094275995), (18430, 0.09655566132966485), (13279, 0.09637886594872276), (15618, 0.09590274693962034), (10705, 0.09589858816108655), (2217, 0.09584351658967685), (15916, 0.0952845
5833273898), (6493, 0.09324996124096596), (10014, 0.09263272829485505), (6407, 0.09259510371416395), (7395, 0.09255664727602676), (3467, 0.09198107485091953), (1889, 0.0918838425178712), (14649, 0.09186569101519272), (2169, 0.0917304
0324374288), (3453, 0.09129035205732766), (16194, 0.09112868597975547), (501, 0.09105152382703523), (5598, 0.09081232006866193), (14518, 0.09056067548901583), (16071, 0.0905554390210855), (18263, 0.08912499029676611), (275, 0.088345
01075905998), (7552, 0.08797735198217047), (10122, 0.08775999173675793), (1388, 0.08774111163948925), (10403, 0.0876258957527456), (19057, 0.08761396947168712), (6170, 0.08723982018121346), (3904, 0.08638760424039675), (11179, 0.085
83096648164262), (20031, 0.08572211990897427), (13636, 0.0856951586462486), (684, 0.08535852756246227), (7071, 0.08520675068130125), (12762, 0.08486897527757943), (10367, 0.08486488258843687), (19570, 0.08367023316112754), (2183, 0.
08364595030745207), (3310, 0.0836302658520049), (4689, 0.0832605725198703), (11767, 0.08321709727578581), (17927, 0.0831139169815071), (7311, 0.08213120310576723), (14332, 0.08212746598467302), (2976, 0.08160089355578226), (17662, 0
.08129137212975728), (15357, 0.0808280390036117), (7357, 0.0805322865551008), (18397, 0.08051801877354674), (18656, 0.08038343875496692), (16574, 0.08034717946452048), (1377, 0.08014654438662819), (18544, 0.07907936354855784), (2246
, 0.07889711297901061), (5281, 0.07873693310906664), (4681, 0.07816297062046638), (14275, 0.07805532632706887), (14833, 0.07791164777387044), (20550, 0.0778467948057746), (15134, 0.0776930095783882), (16110, 0.07767210990200774), (1
4606, 0.07759213069163402), (5123, 0.07708932071018301), (10275, 0.07614965947543449), (7780, 0.07572917513869407), (13198, 0.07557826653716838), (13606, 0.0747176963775907), (17378, 0.07468526856797719), (14418, 0.0746080067758072)
, (18383, 0.07451927734419594), (11341, 0.07399341083526269), (12655, 0.07399171105007682), (16185, 0.07398203733133697), (12340, 0.073712003335165), (8193, 0.07353856684868738), (4593, 0.07353780985799621), (10285, 0.07350730306035
3), (12361, 0.07339969575744741), (13493, 0.07269514418936485), (18431, 0.07259332832210776), (13179, 0.07194649973479934), (12008, 0.07193319605011023), (16327, 0.07175790963679343), (11834, 0.07175661855157867), (1761, 0.071617885
```

Oil price



```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:oil price
[(18645, 0.2121141205904378), (18193, 0.20802584622704487), (10137, 0.19706843223989343), (15829, 0.19644150290875628), (3181, 0.19351377428840333), (18746, 0.19310136713884393), (15038, 0.18653089500834072), (12608, 0.1863981729158
2544), (14558, 0.18608659224890492), (15607, 0.18367351513797497), (127, 0.18210120819359796), (834, 0.18185833659750184), (16991, 0.1795728514177554), (17102, 0.17743993164032), (20919, 0.1757992949180053), (17046, 0.17398319433803
922), (20420, 0.17143589013004665), (5688, 0.17072383021859328), (6413, 0.16924898669728325), (4474, 0.16899369193348343), (6159, 0.16894146409448058), (9731, 0.168333015331322), (2696, 0.16651756181357869), (20936, 0.16443403268361
498), (7528, 0.1619348572801574), (11723, 0.1618005133006386), (4531, 0.15759191721405913), (18754, 0.1567466423849216), (9737, 0.1566544502975895), (543, 0.15656014340212565), (19327, 0.1550665106212349), (191, 0.1538489500570612),
 (18795, 0.15215491276846096), (19193, 0.15102139349115895), (21465, 0.15046439124256072), (3189, 0.15013095538940452), (6342, 0.14984889185496367), (4481, 0.1494560714101066), (17408, 0.14932974029068885), (5166, 0.1490561967174282
), (21076, 0.14853759987310236), (4662, 0.14815079038382417), (4713, 0.14815079038382417), (19059, 0.14689934718296752), (16956, 0.14595784223043273), (11731, 0.14590896705771134), (7590, 0.14590326938321138), (20721, 0.145166366419
8347), (19927, 0.14516450070231307), (18678, 0.14483905892009374), (7152, 0.1438856188585659), (593, 0.14358515242389333), (10330, 0.143548014898428), (5037, 0.14341323663667585), (18367, 0.14333610799497992), (11421, 0.142938618978
33823), (19079, 0.14163579730627066), (8812, 0.1413347754711435), (18655, 0.14042447707877634), (3354, 0.13945550175197477), (459, 0.13894339423181562), (5936, 0.13766592494354443), (18824, 0.1370197077588674), (3389, 0.136731438800
36423), (18765, 0.1361229948009584), (20944, 0.13588925793548204), (18810, 0.13566961875709366), (17892, 0.13453736816685 24), (4453, 0.13402257325766884), (21067, 0.1334998141666327), (20580, 0.13263375553433232), (18840, 0.13233129
756900244), (20041, 0.13226186016732142), (18773, 0.13172636743573127), (18738, 0.13007373415845858), (18704, 0.13006247991721706), (194, 0.12983217048277657), (21380, 0.12929425233506145), (13102, 0.1290024847900536), (11455, 0.128
3270753478332), (18692, 0.12796537232014948), (18689, 0.12787444518851931), (15383, 0.1272484503617167), (18680, 0.12722738646369003), (489, 0.1268359115031591), (9573, 0.12635839673445654), (2811, 0.12596488852260865), (18707, 0.12
581274595920708), (6417, 0.12557121182598804), (8502, 0.12534191124340632), (18706, 0.12514870432468622), (15635, 0.1251261447790628), (18085, 0.12506205748281166), (9031, 0.12459784028222404), (4634, 0.12453232713839042), (18432, 0
.12407050269553126), (9392, 0.12362257507430176), (18747, 0.12356878523651078), (12119, 0.12251769236680697), (4576, 0.12227881897106563), (3303, 0.12161232810773796), (21485, 0.12148752053215982), (14649, 0.12114452888409988), (207
09, 0.12026819878333306), (12111, 0.11976880120389988), (18929, 0.1196782916094542), (8209, 0.1194954004611416), (18698, 0.11915285075949554), (21484, 0.119050508864705), (1521, 0.11896475244974186), (885, 0.11888289127618193), (118
80, 0.1186284853967936), (17929, 0.11829229344885708), (332, 0.11822217976036273), (2998, 0.11814834356074036), (18701, 0.11797307737637873), (6208, 0.11787495542218565), (17254, 0.11779861623715862), (17289, 0.11779861623715862), (
20666, 0.11776997674045389), (17963, 0.11740228568279336), (18789, 0.1170468000650775), (11908, 0.11689184916554728), (10845, 0.11652918743594873), (10873, 0.11652918743594873), (20352, 0.11640945766840516), (944, 0.1163320465255462
5), (18705, 0.11605570491490308), (3169, 0.11567649160002821), (3985, 0.11556997197385116), (357, 0.1155469338199848), (2, 0.11546869058109316), (19051, 0.11520337363303043), (4584, 0.11519211107660965), (3430, 0.11438077407480064),
 (1799, 0.11423629525143172), (5389, 0.1137856121508892), (2263, 0.11373814327926554), (17015, 0.11362720388450286), (18790, 0.1135781857807416), (8478, 0.11335372044387292), (2703, 0.11313520492746876), (11778, 0.11280072878859788)
, (12670, 0.11279048859867837), (5255, 0.1126100917650981), (5926, 0.11210752018387166), (13276, 0.11187274072439671), (13290, 0.11187274072439671), (3146, 0.11152480641082188), (502, 0.11136355133466898), (2582, 0.11135427981374205
), (5559, 0.11134724356634873), (7639, 0.11100647438902736), (21482, 0.11069815981731873), (3342, 0.11034147569836242), (1875, 0.10992112540494581), (873, 0.10979428758804619), (952, 0.10979428758804619), (20406, 0.10975797135434008
), (6145, 0.10944064957373631), (11958, 0.10868654322591205), (20065, 0.10867236298690838), (2975, 0.10835115614138514), (3174, 0.10831527671581659), (5769, 0.10820412853532799), (12050, 0.1081526272197198), (17291, 0.10799849776018
035), (11232, 0.1075597678271119), (6023, 0.10704631039567371), (14755, 0.10637086392018617), (8402, 0.10598206038377 12), (13291, 0.10576894067516601), (15939, 0.10561042703090076), (18999, 0.10535282940402282), (945, 0.105198392363
61281), (4835, 0.10497696480093618), (19370, 0.10495076929785592), (4138, 0.10494332538613123), (9845, 0.10490431471360787), (11935, 0.10453524485842622), (3249, 0.10453286812502649), (16607, 0.10441995717114483), (1387, 0.104382302
0952979), (20360, 0.10423156200576265), (19182, 0.10418070399091195), (6371, 0.1040825108184685), (4658, 0.10407884276466067), (14724, 0.1040310161756266), (14805, 0.1040246205524195), (17028, 0.10393507422968296), (16945, 0.1039095
8901768473), (4386, 0.10378015561965635), (15575, 0.1037108494464927), (19341, 0.10333342865153677), (18448, 0.10332222251969837), (18857, 0.10304945353378699), (11393, 0.10246685892983555), (13204, 0.10224546256901801), (19291, 0.1
0196525868224252), (11728, 0.1018346863033599), (5323, 0.10176213483793455), (1349, 0.10171123628318693), (14544, 0.10148148468919817), (2046, 0.10129498404041247), (21363, 0.10127718006988964), (2007, 0.10071796995907252), (17100,
0.10066897714331717), (16649, 0.10065278379668441), (6201, 0.10053737998280227), (2175, 0.10047973706993749), (18736, 0.10026907242756668), (10348, 0.1000895564964262), (5152, 0.10008501182342124), (14183, 0.09987545786487362), (171
1, 0.09958929677037748), (4017, 0.09913817137614578), (17473, 0.09913105001192823), (19998, 0.09893191824785484), (10261, 0.09849675340735614), (15386, 0.0984674978791559), (213, 0.098435527087104), (17759, 0.09819781012307634), (12
595, 0.09783102316556953), (3455, 0.09782445344930284), (471, 0.09777015029314291), (1909, 0.09773649158412877), (15721, 0.09751392733795139), (20459, 0.09748261407157963), (1151, 0.09747656152055022), (2045, 0.09722478091651564), (
3452, 0.09720739273429715), (16939, 0.09717487728675708), (12402, 0.09709765377353304), (11241, 0.09702543005306127), (5061, 0.09699645593053523), (5268, 0.09693671371567926), (10703, 0.09692448131993188), (10192, 0.0965341258081830
3), (13096, 0.09641549504751251), (11671, 0.09637429357364112), (14833, 0.09630880617893171), (6994, 0.09625109515608599), (13653, 0.09612669981731028), (1837, 0.09606768755058424), (5142, 0.09595232395604085), (18234, 0.09586782781
290817), (3846, 0.09567983780907142), (8134, 0.0954130366072108), (9639, 0.09520979564509165), (19081, 0.09516310102758402), (12953, 0.0951465986720637), (6121, 0.09506762794298633), (12004, 0.0950366384488881), (17478, 0.0946655983
7491352), (6412, 0.09442650844093421), (16950, 0.09419746460115157), (5066, 0.09417616346390091), (4604, 0.09414448759848133), (4600, 0.09408147655601116), (2930, 0.09401694631330922), (11872, 0.09401291389568248), (6264, 0.09384784
607974953), (8606, 0.09378490466987019), (2213, 0.0937846464871409), (17441, 0.09367816583166252), (10975, 0.09356663347308204), (9485, 0.09346039490824537), (7702, 0.09339450610910172), (17419, 0.09229432707023483), (2074, 0.092174
```

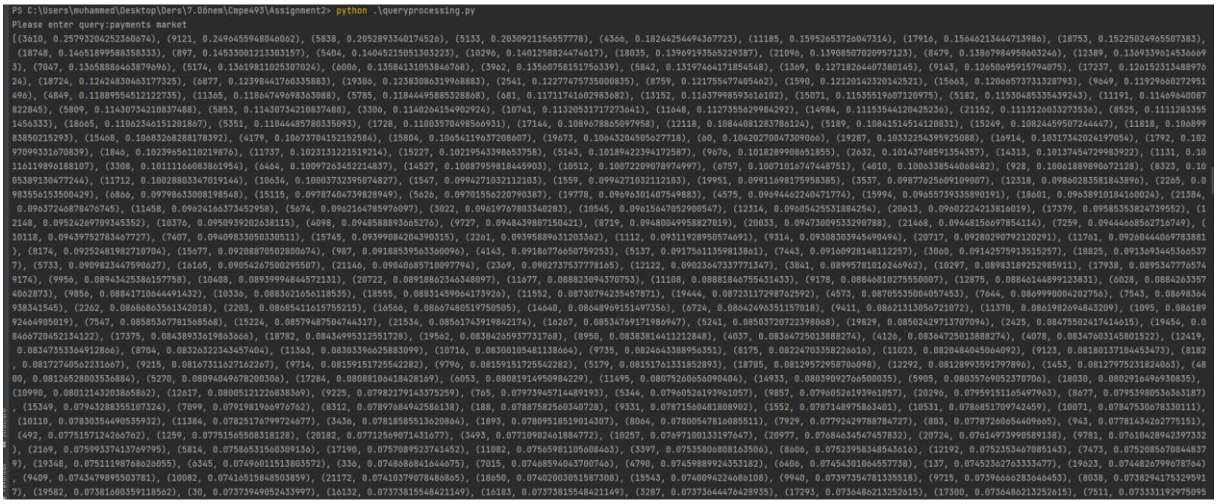
United States



```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:United States
[(12325, 0.23670761510571942), (1834, 0.22968433871153487), (16440, 0.21135481016781643), (4957, 0.21027161776923423), (20240, 0.21012771199636512), (3343, 0.19867725594921298), (7858, 0.19255820321989695), (7862, 0.1870051920258714
3), (12311, 0.18587105901542852), (15494, 0.18514413374894312), (13083, 0.1760784218021066), (20795, 0.1711808169491712), (3199, 0.17093970999560384), (2229, 0.16629043627025436), (16088, 0.16461467047123288), (20643, 0.164394121021
6912), (17020, 0.163083705231873), (19446, 0.16039610105827867), (10463, 0.15961868112712205), (7299, 0.1589680520505259), (17317, 0.15882852510324644), (1668, 0.1574128441431718), (19682, 0.15715637629168783), (17511, 0.15385769678
883665), (4172, 0.15350702937367355), (1777, 0.15303260642149508), (15287, 0.15119983767430134), (1389, 0.1505937964942717), (4294, 0.14978105733471497), (10967, 0.14974815532733127), (13988, 0.14957610126216836), (4467, 0.148597792
11247426), (11695, 0.148328952853368), (14261, 0.1473977265231572), (9501, 0.14605798345197865), (9693, 0.14590476666029825), (15443, 0.14560010294900527), (6066, 0.14525310659182195), (15290, 0.1452428683377702), (1826, 0.145241710
1848106), (6032, 0.14518510222791547), (9813, 0.1446833293947085), (18404, 0.1446040427635523), (15150, 0.14430336311597133), (3125, 0.14395157527267438), (1052, 0.14322820489856022), (6735, 0.1430515943370314), (9415, 0.14289637649
53779), (15720, 0.14272689739433445), (1289, 0.14187838203599684), (13142, 0.14142879825212404), (3444, 0.14128162766716898), (16924, 0.1402266171323507), (9592, 0.1401854250062501), (3467, 0.1401772558484507), (6963, 0.139343915587
574), (584, 0.13864780164051294), (19905, 0.13854693579125935), (21128, 0.13821337708792175), (13185, 0.13812443857530382), (16187, 0.1372924032238577), (10209, 0.13713460802740496), (3827, 0.13663014644920737), (11450, 0.1356343490
2499044), (12983, 0.13561437572839605), (11555, 0.13558421202897863), (21501, 0.13554761028727327), (11551, 0.13530794368000884), (11461, 0.13527357313028307), (280, 0.1352451048479274), (2894, 0.13515846091766778), (9849, 0.1345157
4911149647), (656, 0.13362254596276688), (9443, 0.13284592188371827), (21216, 0.13247306163684516), (12457, 0.13247210025519554), (12473, 0.13247210025519554), (10304, 0.13230460706385203), (4056, 0.13215981630282286), (5247, 0.1321
2847606387787), (688, 0.13177640014501898), (1605, 0.13125802361465824), (6339, 0.1312197207070464), (10232, 0.1311090664969283), (13063, 0.13088967037718138), (655, 0.13077829247682282), (9007, 0.13057500876339118), (2920, 0.130565
41161187524), (14542, 0.13039063219858318), (1956, 0.13023270397012698), (17629, 0.13010579455027424), (7104, 0.12990744570234403), (14674, 0.12946988422095745), (14713, 0.12946988422095745), (2811, 0.12928971988010954), (17263, 0.1
291507637515887), (2883, 0.1289948461146661), (17256, 0.12875215722139294), (10958, 0.12813213027898374), (10942, 0.12702189485346724), (11013, 0.12702189485346724), (19864, 0.1269690631141219), (16946, 0.12694652556821862), (18091,
 0.1266178172198923), (3844, 0.1266144146723602), (9288, 0.12661273885881347), (10501, 0.12631471523855783), (208, 0.126257739517498), (44, 0.12624105499703747), (17669, 0.12576584923555365), (13963, 0.1256119645030142), (6474, 0.12
```

Muhammed Göktepe
2017400162

payments market



```
PS C:\Users\muhammed\Desktop\Ders\7.Dönem\Cmpe493\Assignment2> python .\queryprocessing.py
Please enter query:payments market
[(3610, 0.25793204252360674), (9121, 0.2496455948046062), (5838, 0.2052893340174526), (5133, 0.2030921156557778), (4366, 0.18244254494367723), (11185, 0.15952653726047314), (17916, 0.15646213444713986), (18753, 0.15225024965507383),
 (18748, 0.14651899588358333), (897, 0.14533001213303157), (5404, 0.14045215051303223), (10296, 0.1401258824474617), (18035, 0.13969193565229387), (21096, 0.13908507020957123), (8479, 0.13867984950603246), (12389, 0.1369339614536669
3), (7047, 0.13658886463879696), (5174, 0.13619811025307024), (6006, 0.13584131053846768), (3962, 0.13560758151756339), (5842, 0.13197464171854548), (1369, 0.12718264407380145), (9143, 0.12650695915794075), (17237, 0.126152313488976
24), (18724, 0.12424830463177325), (6877, 0.12398441760335883), (19306, 0.12383086319968883), (2541, 0.12277475735000835), (8759, 0.121755477405462), (1590, 0.12120142320142521), (15663, 0.12066573731328793), (9649, 0.11929660272951
496), (4849, 0.11889554512122735), (11365, 0.11864749698363088), (5785, 0.11844495885328868), (681, 0.11711741602983682), (13152, 0.11637998593616102), (15071, 0.11535519607120975), (5182, 0.11530485335439243), (11191, 0.11469640087
822845), (5809, 0.11430734210837488), (5853, 0.11430734210837488), (3306, 0.1140264154902924), (10741, 0.11320531717273641), (11648, 0.1127355629984292), (14984, 0.11153544120425236), (21152, 0.1113126033273536), (8525, 0.1111283355
1456333), (18665, 0.11062346151201867), (5351, 0.11044485780335093), (1728, 0.11003570498566931), (17144, 0.1089678865097958), (12118, 0.10844081283786124), (5189, 0.10841514514120831), (15249, 0.10824459507244447), (11818, 0.106899
83850215293), (15468, 0.10683268288178392), (4179, 0.10673704152152584), (15804, 0.10654119637208607), (19673, 0.10643204505627718), (60, 0.10420270047309066), (19287, 0.10332254395925088), (16914, 0.10317342024197054), (1792, 0.102
97099331670839), (1846, 0.10239656110219876), (11737, 0.1023131221519214), (15227, 0.10219543398653758), (5143, 0.10189422394172587), (9676, 0.1018289908651855), (2632, 0.10143768591354357), (14313, 0.10137454729983922), (1131, 0.10
116119896188107), (3308, 0.10111166083861954), (6464, 0.10097263452214837), (14527, 0.10087959818445903), (10512, 0.10072209878974997), (6757, 0.10071016747448751), (4010, 0.1006338544068482), (928, 0.10061889890672128), (8323, 0.10
05389130477244), (11712, 0.10028803347019144), (10634, 0.10003732395074827), (1547, 0.0994271032112103), (1559, 0.0994271032112103), (19951, 0.09911698175958385), (3537, 0.09877625609109007), (12318, 0.09860283581843896), (2265, 0.0
983556153500429), (6866, 0.09798633008198548), (15115, 0.09787404739828949), (5626, 0.09701556220790387), (19778, 0.09696301407549883), (4575, 0.09694462240471774), (15994, 0.09655739335890191), (18601, 0.09638910184160024), (21384,
 0.09637246878476745), (11458, 0.09624166373452958), (5674, 0.0962164785976097), (3022, 0.0961976780334 0283), (10545, 0.09615647052900547), (12314, 0.09605425531884254), (20613, 0.09602224213816019), (17379, 0.09585353824739552), (1
2148, 0.09524269709345352), (10376, 0.0950939202638115), (4098, 0.0948588893665276), (9727, 0.0948439807150421), (8719, 0.09480049958827019), (20033, 0.09473009533290788), (21468, 0.09448156697854114), (7259, 0.09444668562716749), (
10118, 0.09439752783467727), (7407, 0.09409833050330511), (15745, 0.09399084204390315), (2261, 0.09395889631203362), (1112, 0.09311928950574691), (9314, 0.09308303945490494), (20717, 0.09280290792120291), (11761, 0.09260444069783881
), (8174, 0.09252481982710704), (15677, 0.09208870502800674), (987, 0.09188539563360096), (4143, 0.09186776650759253), (5137, 0.09175611359813861), (7443, 0.09160928148112257), (3060, 0.09142575913515257), (18825, 0.0913693445366537
7), (5733, 0.0909823447590627), (16165, 0.09054267500295507), (21146, 0.09040685710097794), (2369, 0.0902737537778165), (12122, 0.09023647337771347), (3841, 0.08995781016246962), (10297, 0.08983189252985911), (17938, 0.0895347776574
9174), (9956, 0.08943425386157758), (10408, 0.08939994844572131), (20722, 0.08918862346348097), (11677, 0.088823094370753), (11108, 0.08881846755431433), (9178, 0.08846810275550007), (12875, 0.08846144899123831), (6028, 0.0884263357
4062873), (9856, 0.08841710644491432), (10336, 0.0883621656118535), (18555, 0.08831459064173926), (11552, 0.08730794235457871), (19444, 0.08723117298762592), (4573, 0.08705535004057453), (7644, 0.0869990004202756), (7543, 0.08698364
938341545), (2262, 0.08686863561342018), (2203, 0.08685411615755215), (16566, 0.08667480519750505), (14640, 0.08648969151497356), (6724, 0.08642496351157018), (9411, 0.08621313056721072), (11370, 0.0861982694843209), (1095, 0.086189
92464905019), (7547, 0.08585367781568568), (15224, 0.08579487504744317), (21534, 0.08561743919842174), (16267, 0.08534769171986947), (5241, 0.08503720722398068), (19829, 0.08502429713707094), (2425, 0.08475502417414615), (19454, 0.0
8466720452134122), (17375, 0.08438933619863666), (18782, 0.08434995312551728), (19562, 0.08384265937731768), (8950, 0.08383814411212848), (4037, 0.08364725013888274), (4126, 0.08364725013888274), (4078, 0.08347603145801522), (12419,
 0.08347353364912866), (8704, 0.08326322343457404), (11363, 0.08303396625883099), (10716, 0.08300105481138664), (9735, 0.0824643388956351), (8175, 0.08224703358226616), (11023, 0.0820484045064092), (9123, 0.08180137104453473), (8182
, 0.08172740562231667), (9215, 0.08167311627162267), (9714, 0.08159151725542282), (9796, 0.08159151725542282), (5179, 0.08151761331852893), (18785, 0.08129572958706098), (12292, 0.08128993591797896), (1453, 0.08127975231824063), (48
00, 0.08126528003536884), (5270, 0.08094049678200306), (17284, 0.08088106418428169), (6053, 0.0808191495098422 9), (11495, 0.08075260656090404), (14933, 0.08039092766500035), (5905, 0.08035769052370706), (18030, 0.0802916496930835),
(10990, 0.08012143203865862), (12617, 0.08005121226838369), (9225, 0.07982179143375259), (765, 0.07973945714489193), (5344, 0.07960526193961057), (9857, 0.07960526193961057), (20296, 0.07959151165497963), (8677, 0.07953980536363187)
, (15349, 0.07943288355107324), (7099, 0.0791981966976762), (8312, 0.0789768494258613 8), (188, 0.07887582560340728), (9331, 0.07871560481808902), (1552, 0.0787148975863401), (10531, 0.0786851709742459), (10071, 0.0784753067833 0111),
 (10110, 0.07830354490535932), (11384, 0.07825176799724677), (3436, 0.07818585513620864), (1893, 0.07809518519014307), (8064, 0.07800547816085511), (7929, 0.07792429788784727), (803, 0.07787260654409665), (943, 0.07781434262775151),
 (492, 0.0775157124266762), (1259, 0.07751565508318128), (20182, 0.07712569071431677), (3493, 0.07710902461884772), (10257, 0.07697100133197647), (20977, 0.07684634547457832), (20724, 0.07614973990589138), (9781, 0.07610428942397332
), (2169, 0.07599337413769795), (5814, 0.07586531560309136), (17190, 0.0757089523741452), (11082, 0.07565981105608463), (3397, 0.07535806808163506), (8606, 0.07523958348543616), (12192, 0.0752353467085143), (7473, 0.0752085670844837
9), (19348, 0.07511198768626055), (6345, 0.07496011513803572), (336, 0.0748686841644675), (7015, 0.07468594043700746), (4790, 0.07459889924353182), (6406, 0.07454301064557738), (137, 0.07452362763333477), (19623, 0.0744826799678764)
, (9409, 0.0743479895503781), (10082, 0.07416515848503859), (21172, 0.07410379078486865), (18650, 0.07402003051587308), (15543, 0.0740094224686108), (9940, 0.07397354781335518), (9715, 0.07396666283646453), (8038, 0.0738294175329591
7), (19582, 0.07381600359118562), (30, 0.07373949052433997), (16132, 0.07373815548421149), (16183, 0.07373815548421149), (3287, 0.07373644476428935), (17293, 0.0736486213252615), (17300, 0.0736486213252615), (7511, 0.073331192975095
```