DuckDB के साथ स्थानिक डेटा प्रबंधन

SQL की मूलभूत अवधारणाओं से उन्नत भू-स्थानिक विश्लेषण तक

Qiusheng Wu

DuckDB के साथ स्थानिक डेटा प्रबंधन

SQL की बुनियादी बातों से उन्नत भू-स्थानिक विश्लेषण तक

Qiusheng Wu

2026

Contents

प्रस्तावना **1**
- परिचय 3
- यह पुस्तक किसके लिए है 4
- यह पुस्तक क्या कवर करती है 5
- इस पुस्तक का अधिकतम लाभ उठाना 6
- इस पुस्तक में उपयोग किए गए सम्मेलन 7
- कोड उदाहरण डाउनलोड करना 8
- वीडियो ट्यूटोरियल और पूरक संसाधन 8
- समुदाय और प्रतिक्रिया 9
- आभार 9
- लेखक के बारे में 10
- लाइसेंसिंग और कॉपीराइट 10

I: DuckDB की बुनियादी बातें **11**

1. DuckDB के साथ शुरुआत ***13***
- 1.1. परिचय 13
- 1.2. सीखने के उद्देश्य 13
- 1.3. DuckDB को पारंपरिक डेटाबेस से क्या अलग बनाता है 13
- 1.4. स्थानिक कार्य के लिए DuckDB का कब उपयोग करें (और कब नहीं) 15
- 1.5. DuckDB CLI स्थापित करना और अपनी पहली क्वेरी चलाना 17
- 1.6. DuckDB Python क्लाइंट स्थापित करना 18
- 1.7. Visual Studio Code स्थापित करना 22
- 1.8. DuckDB UI का उपयोग करना 25
- 1.9. DBeaver SQL IDE स्थापित करना 26
- 1.10. मुख्य निष्कर्ष 30
- 1.11. अभ्यास 31

2. स्थानिक विश्लेषण के लिए आवश्यक SQL ***35***
- 2.1. परिचय 35
- 2.2. सीखने के उद्देश्य 35
- 2.3. नमूना डेटासेट 36
- 2.4. वातावरण की स्थापना 36
- 2.5. DuckDB से कनेक्ट करना 37
- 2.6. एक्सटेंशन स्थापित करें 38
- 2.7. URL से CSV फ़ाइलें पढ़ना 38
- 2.8. बेहतर प्रदर्शन के लिए तालिकाएँ बनाना 39
- 2.9. SQL SELECT स्टेटमेंट 41
- 2.10. WHERE क्लॉज के साथ डेटा फ़िल्टर करना 48
- 2.11. LIKE के साथ पैटर्न मैचिंग 50
- 2.12. IN ऑपरेटर 51
- 2.13. BETWEEN ऑपरेटर 52

2.14. SQL जॉइन के साथ डेटा का संयोजन ... 53
2.15. क्वेरी योजना और प्रदर्शन ... 59
2.16. सारांश आँकड़ों के लिए डेटा का एग्रीगेशन ... 60
2.17. सशर्त कथन ... 63
2.18. परिणाम सहेजना ... 64
2.19. स्थानिक विश्लेषण के लिए उन्नत SQL फ़ीचर ... 66
2.20. SQL टिप्पणियाँ और दस्तावेज़ीकरण ... 70
2.21. मुख्य निष्कर्ष ... 71
2.22. अभ्यास ... 72
***3. DuckDB* का *Python* के साथ एकीकरण ... 77**
3.1. परिचय ... 77
3.2. सीखने के उद्देश्य ... 77
3.3. नमूना डेटासेट ... 78
3.4. स्थापना और सेटअप ... 79
3.5. एक्सटेंशन स्थापित करना और लोड करना ... 80
3.6. कई स्रोतों से डेटा पढ़ना ... 81
3.7. Pandas DataFrames के साथ निर्बाध एकीकरण ... 84
3.8. Polars अंतरसंचालनीयता ... 87
3.9. परिणाम रूपांतरण और आउटपुट प्रारूप ... 88
3.10. डेटा को डिस्क पर लिखना ... 90
3.11. स्थायी स्टोरेज और डेटाबेस फ़ाइलें ... 91
3.12. तैयार स्टेटमेंट और पैरामीटर ... 93
3.13. मुख्य निष्कर्ष ... 94
3.14. अभ्यास ... 96

II: स्थानिक डेटा संचालन ... 101
***4.* स्थानिक डेटा फ़ॉर्मेट लोड करना ... 103**
4.1. परिचय ... 103
4.2. सीखने के उद्देश्य ... 104
4.3. नमूना डेटासेट ... 105
4.4. स्थापना और सेटअप ... 105
4.5. एक्सटेंशन स्थापित और लोड करना ... 106
4.6. नमूना डेटा डाउनलोड करना ... 107
4.7. निर्देशांक के साथ CSV फ़ाइलें लोड करना ... 107
4.8. JSON फ़ाइलें लोड करना ... 111
4.9. Pandas DataFrames को सीधे क्वेरी करना ... 113
4.10. प्रदर्शन के लिए Parquet फ़ाइलें लोड करना ... 114
4.11. स्थानिक ज्यामिति के साथ GeoJSON फ़ाइलें लोड करना ... 117
4.12. आधुनिक वर्कफ़्लो में Shapefile लोड करना ... 123
4.13. क्लाउड-नेटिव स्थानिक विश्लेषण के लिए GeoParquet लोड करना ... 125
4.14. डेटा लोडिंग प्रदर्शन रणनीतियाँ ... 127
4.15. सामान्य डेटा लोडिंग समस्याओं का समाधान ... 128

4.16. मुख्य निष्कर्ष 129
4.17. अभ्यास 131
5. स्थानिक डेटा का निर्यात और रूपांतरण 135
5.1. परिचय 135
5.2. सीखने के उद्देश्य 136
5.3. नमूना डेटासेट 137
5.4. स्थापना और सेटअप 137
5.5. एक्सटेंशन्स की स्थापना और लोडिंग 138
5.6. नमूना डेटा लोड करना 140
5.7. Pandas DataFrames में निर्यात करना 141
5.8. CSV फ़ाइलों में निर्यात करना 146
5.9. JSON फ़ाइलों में निर्यात करना 150
5.10. Excel फ़ाइलों में निर्यात करना 152
5.11. Parquet फ़ाइलों में निर्यात करना 155
5.12. GeoJSON फ़ॉर्मेट में निर्यात करना 157
5.13. Shapefile फ़ॉर्मेट में निर्यात करना 161
5.14. GeoPackage फ़ॉर्मेट में निर्यात करना 162
5.15. मुख्य निष्कर्ष 165
5.16. अभ्यास 166
6. ज्यामिति संचालन और फ़ंक्शन 173
6.1. परिचय 173
6.2. अधिगम उद्देश्य 173
6.3. नमूना डेटासेट 174
6.4. स्थापना और सेटअप 175
6.5. DuckDB से कनेक्ट करना और एक्सटेंशन लोड करना 176
6.6. ज्यामिति प्रकारों को समझना 178
6.7. बिंदुओं के साथ काम करना 181
6.8. Linestrings के साथ काम करना 184
6.9. बहुभुजों के साथ काम करना 186
6.10. संग्रहों के साथ काम करना 188
6.11. NYC स्थानिक डेटा का विज़ुअलाइज़ेशन 189
6.12. ज्यामिति प्रसंस्करण 192
6.13. ज्यामिति वैधता और मजबूती 196
6.14. मुख्य बातें 198
6.15. अभ्यास 199
7. स्थानिक क्वेरीज़ और संबंध 205
7.1. परिचय 205
7.2. सीखने के उद्देश्य 206
7.3. नमूना डेटासेट 206
7.4. इंस्टॉलेशन और सेटअप 207
7.5. DuckDB से जुड़ना और एक्सटेंशन लोड करना 208

7.6. स्थानिक संबंधों को समझना 209
7.7. ज्यामितीय पहचान का परीक्षण 209
7.8. टोपोलॉजिकल संबंध 212
7.9. दूरी-आधारित संबंध 220
7.10. निकटता सीमा द्वारा फ़िल्टरिंग 222
7.11. निकटतम पड़ोसी क्वेरी 223
7.12. मुख्य निष्कर्ष 223
7.13. अभ्यास 225
8. उन्नत स्थानिक जॉइन ***229***
8.1. परिचय 229
8.2. सीखने के उद्देश्य 229
8.3. नमूना डेटासेट 230
8.4. इंस्टॉलेशन 231
8.5. लाइब्रेरी आयात और कॉन्फ़िगरेशन 231
8.6. DuckDB से कनेक्ट करना 231
8.7. इंटरसेक्शन जॉइन 232
8.8. दूरी के भीतर जॉइन 238
8.9. उन्नत जॉइन 241
8.10. निर्देशांक प्रणाली परिवर्तन 243
8.11. स्थानिक संबंध फ़ंक्शन संदर्भ 246
8.12. मुख्य सीख 247
8.13. अभ्यास 248
9. इंटरैक्टिव डेटा विज़ुअलाइज़ेशन ***255***
9.1. परिचय 255
9.2. सीखने के उद्देश्य 256
9.3. नमूना डेटासेट 256
9.4. स्थापना और सेटअप 257
9.5. नमूना डेटा डाउनलोड करना 257
9.6. DuckDB से कनेक्ट करना और एक्सटेंशन लोड करना 258
9.7. बिंदु डेटा को विज़ुअलाइज़ करना 258
9.8. रेखा डेटा को विज़ुअलाइज़ करना 262
9.9. बहुभुज डेटा को विज़ुअलाइज़ करना 264
9.10. मुख्य निष्कर्ष 269
9.11. अभ्यास 271
10. वेक्टर टाइल्स और ***PMTiles*** के साथ कार्य ***275***
10.1. परिचय 275
10.2. अधिगम उद्देश्य 276
10.3. नमूना डेटासेट 277
10.4. इंस्टॉलेशन और सेटअप 278
10.5. फ़ाइलों से सीधे वेक्टर टाइल्स को विज़ुअलाइज़ करना 278
10.6. वेक्टर डेटा को PMTiles में परिवर्तित करना 283

10.7. PMTiles को विज़ुअलाइज़ करना ... 285
10.8. मुख्य निष्कर्ष ... 294
10.9. अभ्यास ... 296

III: वास्तविक दुनिया का भू-स्थानिक विश्लेषण ... 301

11. अमेरिकी राष्ट्रीय आर्द्रभूमि सूची का विश्लेषण ... ***303***
11.1. परिचय ... 303
11.2. सीखने के उद्देश्य ... 304
11.3. इस अध्याय में उपयोग किए गए डेटासेट ... 305
11.4. डेटासेट स्रोत को समझना ... 307
11.5. DuckDB के साथ आर्द्रभूमि डेटा तक पहुँचना ... 309
11.6. आर्द्रभूमि वितरण को विज़ुअलाइज़ करना ... 312
11.7. राष्ट्रीय-स्तरीय आर्द्रभूमि विश्लेषण ... 317
11.8. मुख्य निष्कर्ष ... 330
11.9. अभ्यास ... 332

12. वैश्विक भवन फुटप्रिंट का विश्लेषण ... ***335***
12.1. परिचय ... 335
12.2. सीखने के उद्देश्य ... 336
12.3. डेटासेट के बारे में ... 337
12.4. स्थापना और सेटअप ... 339
12.5. एक्सटेंशन स्थापित करना और लोड करना ... 340
12.6. उपलब्ध डेटा का अन्वेषण ... 340
12.7. बाउंडिंग बॉक्स फ़िल्टरिंग के साथ क्षेत्रीय विश्लेषण ... 344
12.8. बहु-क्षेत्रीय तुलना ... 354
12.9. डेटा स्रोत फ़िल्टरिंग और गुणवत्ता मूल्यांकन ... 359
12.10. ग्रिड-आधारित स्थानिक एकत्रीकरण ... 364
12.11. H3 हेक्सागोनल ग्रिड के साथ भवन डेटा का एकत्रीकरण ... 369
12.12. मुख्य बातें ... 380
12.13. अभ्यास ... 382

13. ***NYC*** टैक्सी डेटा का विश्लेषण ... ***385***
13.1. परिचय ... 385
13.2. सीखने के उद्देश्य ... 386
13.3. डेटासेट के बारे में ... 387
13.4. इंस्टॉलेशन ... 389
13.5. लाइब्रेरी आयात ... 389
13.6. एक्सटेंशन स्थापित करना और लोड करना ... 389
13.7. टैक्सी डेटा लोड करना ... 389
13.8. टेम्पोरल विश्लेषण ... 393
13.9. टैक्सी ज़ोन लुकअप डेटा लोड करना ... 401
13.10. स्थानिक विश्लेषण ... 403
13.11. यात्रा फ्लो विश्लेषण ... 411
13.12. भुगतान और आर्थिक विश्लेषण ... 416

13.13. यात्री व्यवहार विश्लेषण 419
13.14. बहु-माह विश्लेषण 421
13.15. दृश्यीकरण 423
13.16. प्रदर्शन अनुकूलन युक्तियाँ 429
13.17. मुख्य निष्कर्ष 430
13.18. अभ्यास 432
***14. Voilà* और *Solara* के साथ इंटरैक्टिव डैशबोर्ड विकसित करना 435**
14.1. परिचय 435
14.2. सीखने के उद्देश्य 437
14.3. Voilà और Solara इंस्टॉल करना 437
14.4. Hugging Face Spaces का परिचय 437
14.5. एक बुनियादी Voilà एप्लिकेशन बनाना 439
14.6. Solara के साथ एक उन्नत वेब एप्लिकेशन बनाना 449
14.7. मुख्य निष्कर्ष 458
14.8. अभ्यास 459

प्रस्तावना

परिचय

तेजी से डेटा-संचालित होती दुनिया में, स्थानिक जानकारी को प्रभावी ढंग से प्रबंधित करने और उसका विश्लेषण करने की क्षमता आवश्यक हो गई है। शहरी नियोजन और पर्यावरणीय निगरानी से लेकर लॉजिस्टिक्स और व्यक्तिगत स्थान-आधारित सेवाओं तक, भू-स्थानिक डेटा हमारे दैनिक जीवन को प्रभावित करने वाले अनेक अनुप्रयोगों की रीढ़ बनाता है। हालांकि, स्थानिक डेटा के साथ काम करना अक्सर एक विशिष्ट और जटिल क्षेत्र के रूप में देखा जाता रहा है, जिसके लिए पेचीदा उपकरणों और एक खड़ी सीखने की प्रक्रिया की आवश्यकता होती है।

DuckDB का प्रवेश।

DuckDB एक नवीन विश्लेषणात्मक डेटाबेस है जो दक्षता और उपयोगकर्ता-अनुकूलता के लिए डिज़ाइन किया गया है। एक इन-प्रोसेस OLAP (ऑनलाइन एनालिटिकल प्रोसेसिंग) डेटाबेस के रूप में, यह सीधे आपके एप्लिकेशन के भीतर संचालित होता है, जिससे अलग सर्वर परिनियोजन और बोझिल कॉन्फ़िगरेशन की आवश्यकता समाप्त हो जाती है। यह एम्बेडेड प्रकृति, इसके कॉलम-उन्मुख आर्किटेक्चर और वेक्टराइज़्ड निष्पादन इंजन के साथ मिलकर, DuckDB को विश्लेषणात्मक क्वेरी के लिए असाधारण रूप से तेज़ बनाती है, यहाँ तक कि बड़े डेटासेट के साथ भी। शुरुआत में अपनी सामान्य-उद्देश्य डेटा प्रसंस्करण क्षमताओं के लिए लोकप्रियता प्राप्त करते हुए, DuckDB का तेजी से विस्तार होता पारिस्थितिकी तंत्र (विशेष रूप से स्थानिक डेटा के लिए इसके एक्सटेंशन) एक परिवर्तनकारी बदलाव का प्रतिनिधित्व करता है।

यह पुस्तक, "***Spatial Data Management with DuckDB: From SQL Basics to Advanced Geospatial Analytics***," का उद्देश्य भू-स्थानिक डेटा को सरल बनाना और यह दिखाना है कि कैसे DuckDB सभी (डेटा विश्लेषकों और वैज्ञानिकों से लेकर डेवलपर्स और GIS पेशेवरों तक) को अभूतपूर्व सरलता और गति के साथ अपनी क्षमताओं का लाभ उठाने में सक्षम बनाता है। हमारा मानना है कि मजबूत स्थानिक विश्लेषण सुलभ होना चाहिए, न कि महंगे विशिष्ट सॉफ़्टवेयर या जटिल प्रोग्रामिंग भाषाओं तक सीमित। DuckDB के साथ, वह सुलभता वास्तविकता बन जाती है।

हमारी यात्रा स्थानिक डेटा की मूलभूत अवधारणाओं और इसके प्रतिनिधित्व से शुरू होती है, बिंदुओं, रेखाओं और बहुभुजों के साथ काम करने के लिए SQL में एक ठोस आधार स्थापित करती है। जैसे-जैसे हम आगे बढ़ते हैं, आप यह जानेंगे कि कैसे DuckDB की मूल भू-स्थानिक विशेषताएँ (इसके PostGIS-संगत एक्सटेंशन द्वारा बढ़ाई गई) सुरुचिपूर्ण SQL क्वेरी के माध्यम से स्थानिक जॉइन, बफ़रिंग और निकटतम पड़ोसी खोज जैसे परिष्कृत संचालन को सक्षम बनाती हैं। हम विभिन्न वास्तविक दुनिया के अनुप्रयोगों का पता लगाएंगे, यह प्रदर्शित करते हुए कि कैसे स्थानिक डेटासेट को लोड, रूपांतरित, विश्लेषण और विज़ुअलाइज़ किया जाए, जो आपको भौगोलिक जानकारी से सार्थक अंतर्दृष्टि निकालने में सक्षम बनाता है।

चाहे आप अपनी डेटा पाइपलाइन में स्थानिक विश्लेषण को एकीकृत करना चाहते हों, त्वरित तदर्थ भू-स्थानिक क्वेरी निष्पादित करना चाहते हों, या इंटरैक्टिव स्थान-जागरूक एप्लिकेशन विकसित करना चाहते हों, यह पुस्तक आपकी व्यापक मार्गदर्शिका के रूप में काम करेगी। हम आपके DuckDB वातावरण की स्थापना और विविध स्थानिक फ़ाइल प्रारूपों (जैसे Shapefiles, GeoJSON, और GeoParquet) को आयात करने से लेकर जटिल विश्लेषणात्मक कार्यों को निष्पादित करने और विज़ुअलाइज़ेशन उपकरणों के साथ एकीकृत करने तक के विषयों को कवर करेंगे।

हमारा उद्देश्य केवल आपको सिंटैक्स सिखाना नहीं है, बल्कि यह समझ विकसित करना है कि ये उपकरण और तकनीकें क्यों शक्तिशाली हैं। इस पुस्तक के अंत तक, आप स्थानिक डेटा प्रबंधन और विश्लेषण के लिए अपने पसंदीदा इंजन के रूप में DuckDB का उपयोग करने में निपुण हो जाएंगे, अपने प्रोजेक्ट के लिए नई संभावनाओं को खोलेंगे और सूचित, स्थानिक रूप से जागरूक निर्णय लेने में सक्षम होंगे।

जैसे ही हम DuckDB की विश्लेषणात्मक क्षमताओं और भू-स्थानिक डेटा की समृद्ध दुनिया के रोमांचक चौराहे में उतरते हैं, हमारे साथ जुड़ें। सुलभ स्थानिक विश्लेषण का भविष्य यहाँ है, और यह DuckDB पर चलता है।

यह पुस्तक किसके लिए है

यह पुस्तक उन सभी के लिए डिज़ाइन की गई है जो आधुनिक स्थानिक डेटा विश्लेषण की जटिलताओं से जूझ रहे हैं। यदि आपने कभी एक स्थानिक जॉइन के समाप्त होने की प्रतीक्षा में घंटों बिताए हैं, बड़े भौगोलिक डेटासेट को मेमोरी में लोड करने में संघर्ष किया है, या SQL की शक्ति को स्थानिक संचालन के साथ संयोजित करने का अधिक सीधा तरीका चाहा है, तो यह पुस्तक आपके लिए है।

आपको सबसे अधिक मूल्य मिलेगा यदि आप हैं

एक **GIS** पेशेवर जो बड़े डेटासेट को संभालते समय डेस्कटॉप सॉफ़्टवेयर की सीमाओं से निराश हैं। आप QGIS या ArcGIS से परिचित हैं, लेकिन आपको लाखों फ़ीचर का विश्लेषण करने, व्यापक GPS ट्रैक को संसाधित करने, या स्वचालित वर्कफ़्लो में स्थानिक विश्लेषण को एकीकृत करने की आवश्यकता है।

एक डेटा वैज्ञानिक या विश्लेषक जो अक्सर स्थान डेटा का सामना करते हैं। आप Python और pandas के साथ सहज हैं, लेकिन स्थानिक डेटा अक्सर एक रहस्य की तरह लगता है। आप जटिल GIS सॉफ़्टवेयर में उतरे बिना अपने विश्लेषण में भौगोलिक आयाम शामिल करना चाहते हैं।

एक सॉफ़्टवेयर डेवलपर जो स्थानिक विशेषताओं को शामिल करने वाले एप्लिकेशन का निर्माण कर रहे हैं। आपको तेज़ स्थानिक क्वेरी की आवश्यकता है, भारी डेटाबेस इन्फ्रास्त्रक्चर से बचना चाहते हैं, और विशेष स्थानिक लाइब्रेरी के बजाय परिचित SQL के साथ काम करना पसंद करते हैं।

एक शोधकर्ता या शिक्षाविद भूगोल, पर्यावरण विज्ञान, या शहरी नियोजन जैसे क्षेत्रों में। आपके शोध में बड़े स्थानिक डेटासेट शामिल हैं, और आपको पुनरुत्पादनीय, स्केलेबल विश्लेषण विधियों की आवश्यकता है जो बढ़ते डेटा वॉल्यूम के अनुकूल हो सकें।

एक बिजनेस इंटेलिजेंस पेशेवर स्थान-आधारित व्यावसायिक डेटा से निपट रहे हैं। चाहे वह स्टोर स्थान हों, डिलीवरी रूट हों, ग्राहक क्षेत्र हों, या रियल एस्टेट पोर्टफोलियो हों, आपको व्यावसायिक मेट्रिक्स को स्थानिक अंतर्दृष्टि के साथ मर्ज करने की आवश्यकता है।

आवश्यक पूर्वापेक्षाएँ

आपको इनके साथ सहज होना चाहिए:

- **Python** प्रोग्रामिंग: वेरिएबल्स, फ़ंक्शन, और लाइब्रेरी आयात करने का तरीका समझना (विशेषज्ञता आवश्यक नहीं है)
- डेटा विश्लेषण अवधारणाएँ: रिकॉर्ड फ़िल्टर करना, डेटा एकत्र करना, और तालिकाओं को जॉइन करना
- **SQL** मूल बातें: SELECT, WHERE, और GROUP BY क्लॉज से परिचय (हम स्थानिक पहलुओं को कवर करेंगे)
- स्थानिक डेटा की मूल बातें: यह समझना कि डेटा का एक स्थान होता है (अक्षांश/देशांतर, प्रक्षेपण)

सहायक पृष्ठभूमि (लेकिन आवश्यक नहीं)

- pandas, GeoPandas, या Jupyter नोटबुक के साथ अनुभव
- डेटाबेस या डेटा वेयरहाउस के साथ पूर्व परिचय
- GIS सॉफ़्टवेयर (QGIS, ArcGIS, PostGIS) से परिचय
- स्थानिक फ़ाइल प्रारूपों (GeoJSON, Shapefiles, Parquet) का ज्ञान

यदि आप Python प्रोग्रामिंग में नए हैं

यदि आप भू-स्थानिक Python प्रोग्रामिंग में नए हैं, तो निम्नलिखित पुस्तक मूलभूत GIS अवधारणाओं और Python प्रोग्रामिंग दोनों का एक उत्कृष्ट परिचय प्रदान करती है:

Wu, Q. (2025). *Introduction to GIS Programming: A Practical Python Guide to Open Source Geospatial Tools.* Independently published. ISBN 979-8286979455. https://www.amazon.com/dp/B0FFW34LL3

यह पुस्तक क्या कवर करती है

यह पुस्तक SQL मूल बातों से लेकर उन्नत भू-स्थानिक विश्लेषण तक एक संरचित यात्रा प्रदान करती है, जो आपको वास्तविक दुनिया के उदाहरणों के माध्यम से व्यावहारिक कौशल से लैस करती है। प्रत्येक अध्याय सरल क्वेरी से जटिल स्थानिक विश्लेषण तक आगे बढ़ता है, आधुनिक भू-स्थानिक डेटा प्रबंधन में आपकी विशेषज्ञता का निर्माण करता है।

भाग I: DuckDB नींव *(अध्याय 1-3)*

उन आवश्यक अवधारणाओं में महारत हासिल करें जो बाद की सभी सामग्री को रेखांकित करती हैं:

- अध्याय **1: DuckDB** के साथ शुरुआत: इंस्टॉलेशन, प्रारंभिक क्वेरी, और इस बारे में अंतर्दृष्टि कि DuckDB स्थानिक विश्लेषण में कैसे क्रांति लाता है।
- अध्याय **2:** स्थानिक विश्लेषण के लिए आवश्यक **SQL**: स्थानिक डेटा के लिए तैयार किए गए फ़िल्टरिंग, एकत्रीकरण, जॉइनिंग, और क्वेरी अनुकूलन के लिए मुख्य SQL पैटर्न।
- अध्याय **3: DuckDB Python** एकीकरण: एक सहज स्थानिक विश्लेषण वर्कफ़्लो तैयार करने के लिए SQL की शक्ति को pandas की लचीलेपन के साथ संयोजित करें।

भाग *I* के अंत तक, आप आत्मविश्वास से स्थानिक डेटासेट को क्वेरी करेंगे और किसी भी Python-आधारित विश्लेषण पाइपलाइन में DuckDB को एकीकृत करेंगे।

भाग II: स्थानिक डेटा संचालन *(अध्याय 4-10)*

मुख्य स्थानिक टूलकिट में गोता लगाएँ, जो डेटा लोडिंग से लेकर उन्नत विश्लेषण तक सब कुछ कवर करता है:

- अध्याय **4:** स्थानिक डेटा प्रारूप लोड करना: CSV निर्देशांक, API से GeoJSON, विशाल Shapefiles, और क्लाउड-होस्टेड GeoParquet सहित विभिन्न प्रारूपों को आयात करें।
- अध्याय **5:** स्थानिक डेटा निर्यात और रूपांतरण: अपने परिणामों को किसी भी प्रारूप में रूपांतरित करें जिसकी आपके हितधारकों को आवश्यकता हो।
- अध्याय **6:** ज्यामिति संचालन और फ़ंक्शन: SQL फ़ंक्शन का उपयोग करके स्थानिक फ़ीचर बनाएँ, मापें, और रूपांतरित करें।
- अध्याय **7:** स्थानिक क्वेरी और संबंध: तालिका जॉइन के स्थानिक समतुल्य में महारत हासिल करें: समावेशन, प्रतिच्छेदन, और निकटता।
- अध्याय **8:** उन्नत स्थानिक जॉइन: ID के बजाय स्थान के अनुसार डेटासेट को संयोजित करें, जो स्थानिक विश्लेषण का सार है।
- अध्याय **9:** इंटरैक्टिव डेटा विज़ुअलाइज़ेशन: आकर्षक मानचित्र और चार्ट उत्पन्न करें जो आपके डेटा की स्थानिक कथाओं को प्रभावी ढंग से संप्रेषित करते हैं।
- अध्याय **10:** वेक्टर टाइलें और **PMTiles** के साथ काम करना: इंटरैक्टिव मानचित्र परिनियोजित करें जो लाखों फ़ीचर को सहजता से संभालने में सक्षम हैं।

भाग *II* के अंत तक, आप किसी भी स्थानिक डेटा प्रारूप का प्रबंधन करने, जटिल संचालन निष्पादित करने, और पेशेवर विज़ुअलाइज़ेशन बनाने में निपुण हो जाएंगे।

भाग III: वास्तविक दुनिया का भू-स्थानिक विश्लेषण *(अध्याय 11-14)*

बड़े पैमाने पर, वास्तविक डेटासेट का उपयोग करके चार व्यापक केस स्टडी का अन्वेषण करें:

- अध्याय **11:** यूएस राष्ट्रीय आर्द्रभूमि सूची का विश्लेषण: सभी 50 राज्यों में पर्यावरणीय विश्लेषण करें, लाखों आर्द्रभूमि बहुभुजों को संसाधित करें।
- अध्याय **12:** वैश्विक भवन फ़ुटप्रिंट का विश्लेषण: Overture Maps के वैश्विक भवन फ़ुटप्रिंट का उपयोग करके शहरी डेटा का विश्लेषण करें।
- अध्याय **13: NYC** टैक्सी ट्रिप डेटा का विश्लेषण: करोड़ों टैक्सी यात्राओं से स्थानिक-कालिक पैटर्न का पता लगाएँ, शहरी गतिशीलता में अंतर्दृष्टि प्रकट करें।
- अध्याय **14: Voilà** और **Solara** के साथ इंटरैक्टिव डैशबोर्ड विकसित करना: वेब एप्लिकेशन बनाएँ और परिनियोजित करें जो आपके विश्लेषण को हितधारकों के लिए सुलभ बनाते हैं।

भाग *III* के अंत तक, आपके पास पोर्टफोलियो-योग्य प्रोजेक्ट होंगे जो आपकी उन्नत स्थानिक विश्लेषण क्षमताओं को प्रदर्शित करते हैं।

पूरे पुस्तक में क्रॉस-कटिंग विषय

- पैमाने पर प्रदर्शन: तकनीकें जो हजारों या लाखों स्थानिक फ़ीचर को संभालने में प्रभावी हैं।
- क्लाउड-नेटिव वर्कफ़्लो: सीधे S3 से डेटा संसाधित करें और आधुनिक डेटा स्टैक के साथ निर्बाध रूप से एकीकृत करें।
- पुनरुत्पादनीय विश्लेषण: साझा करने योग्य कोड और विधियाँ जिन्हें संस्करण-नियंत्रित किया जा सकता है और उत्पादन में परिनियोजित किया जा सकता है।
- वास्तविक दुनिया की डेटा चुनौतियाँ: प्रक्षेपण समस्याओं, अनुपलब्ध मानों, और डेटा गुणवत्ता संबंधी चिंताओं से निपटें।
- एकीकरण पैटर्न: बढ़ी हुई कार्यक्षमता के लिए DuckDB को व्यापक Python भू-स्थानिक पारिस्थितिकी तंत्र के साथ संयोजित करें।

क्या इस पुस्तक को अलग बनाता है

सैद्धांतिक चर्चाओं या उपकरण-विशिष्ट ट्यूटोरियल के विपरीत, यह पुस्तक वास्तविक समस्याओं को हल करने पर जोर देती है। प्रत्येक तकनीक वास्तविक विश्लेषणात्मक चुनौतियों में निहित है, वास्तविक डेटासेट के साथ प्रदर्शित की गई है, और स्पष्ट शब्दों में समझाया गया है कि कब और क्यों इसका उपयोग किया जाए।

इस पुस्तक का अधिकतम लाभ उठाना

इस पुस्तक के साथ अपने सीखने के अनुभव को अधिकतम करने के लिए, निम्नलिखित अनुशंसाओं पर विचार करें:

एक उचित विकास वातावरण स्थापित करें: अध्याय 1 में वर्णित अनुसार Python और आवश्यक लाइब्रेरी स्थापित करें। एक अच्छी तरह से कॉन्फ़िगर किया गया वातावरण आपके सीखने की यात्रा के दौरान आपका समय और निराशा बचाएगा। अपने Python पैकेज प्रबंधित करने के लिए conda या uv का उपयोग करने पर विचार करें, क्योंकि यह भू-स्थानिक लाइब्रेरी की स्थापना को सरल बनाता है।

कोड उदाहरणों के साथ अनुसरण करें: यह पुस्तक इंटरैक्टिव होने के लिए डिज़ाइन की गई है। केवल कोड न पढ़ें; इसे टाइप करें, चलाएँ, और संशोधनों के साथ प्रयोग करें। समझ अभ्यास के माध्यम से आती है, और प्रत्येक उदाहरण ऐसे कौशल का निर्माण करता है जिनकी आपको बाद में आवश्यकता होगी।

अभ्यासों के माध्यम से काम करें: प्रत्येक अध्याय में आपके द्वारा सीखी गई अवधारणाओं को सुदृढ़ करने के लिए डिज़ाइन किए गए अभ्यास शामिल हैं। ये वैकल्पिक अतिरिक्त नहीं हैं; वे सीखने की प्रक्रिया का अभिन्न अंग हैं। निर्देशित अभ्यासों से शुरू करें, फिर अपनी परियोजनाओं के साथ खुद को चुनौती दें।

वास्तविक डेटा का उपयोग करें: जबकि पुस्तक उदाहरणों और अभ्यासों के लिए डेटासेट प्रदान करती है, अपने स्वयं के क्षेत्र या रुचियों से डेटा पर तकनीकों को लागू करने का प्रयास करें। इससे आपको यह समझने में मदद मिलेगी कि अवधारणाएँ वास्तविक दुनिया के परिदृश्यों पर कैसे लागू होती हैं और आपकी क्षमताओं में आत्मविश्वास का निर्माण होगा।

परियोजनाएँ बनाएँ: जैसे-जैसे आप पुस्तक के माध्यम से प्रगति करते हैं, एक व्यक्तिगत परियोजना पर काम करने पर विचार करें जो आपको रुचिकर लगती है। यह आपके शोध से डेटा का विश्लेषण करना, अपने समुदाय के लिए मानचित्र बनाना, या आपके काम में आपके सामने आने वाली समस्या को हल करना हो सकता है।

अपने आप के साथ धैर्य रखें: प्रोग्रामिंग निराशाजनक हो सकती है, खासकर जब आप सीख रहे हों। त्रुटियों का सामना करने, डिबगिंग में समय बिताने, और कभी-कभी अटकने की उम्मीद करें। यह सामान्य है और सीखने की प्रक्रिया का हिस्सा है। आवश्यकता पड़ने पर ब्रेक लें, और याद रखें कि विशेषज्ञता निरंतर अभ्यास के माध्यम से धीरे-धीरे विकसित होती है। यदि आप अटक जाते हैं, तो पुस्तक के GitHub रिपॉजिटरी पर मदद माँगने में संकोच न करें।

अभ्यास करते रहें: इस पुस्तक के कौशलों को बनाए रखने और विकसित करने के लिए नियमित अभ्यास की आवश्यकता है। भू-स्थानिक प्रोग्रामिंग परियोजनाओं पर काम करने के लिए नियमित रूप से समय निकालें, भले ही वे छोटी हों।

इस पुस्तक में उपयोग किए गए सम्मेलन

यह पुस्तक आपको सामग्री में नेविगेट करने और कोड उदाहरणों को समझने में मदद करने के लिए कई सम्मेलनों का उपयोग करती है:

कोड स्वरूपण: सभी Python कोड कोड ब्लॉक के भीतर मोनोस्पेस्ड फ़ॉन्ट में दिखाई देते हैं। जब कोड नियमित पाठ के भीतर दिखाई देता है, तो इसे `like this` स्वरूपित किया जाता है। फ़ाइल और निर्देशिका नाम भी मोनोस्पेस्ड फ़ॉन्ट में स्वरूपित होते हैं।

कोड उदाहरण: अधिकांश कोड उदाहरण पूर्ण और चलने योग्य हैं। उनमें प्रदर्शित की जा रही प्रमुख अवधारणाओं और तकनीकों को समझाने वाली टिप्पणियाँ शामिल हैं। साथ देने वाले पाठ में संदर्भ के लिए लाइन नंबर शामिल किए जा सकते हैं।

```
# This is an example of a code block
import leafmap
m = leafmap.Map()
m.add_basemap("OpenTopoMap") # add a basemap to the map
m
```

SQL स्टाइल गाइड: निरंतरता और पठनीयता के लिए, SQL उदाहरण इन पैटर्नों का अनुसरण करते हैं:

- **UPPERCASE** में कीवर्ड: `SELECT`, `FROM`, `WHERE`, `JOIN`
- फ़ंक्शन नाम केस संरक्षित करते हैं: `ST_Area()`, `read_csv_auto()`
- लोअरकेस में तालिका और कॉलम नाम: `cities`, `population`
- पठनीयता के लिए इंडेंटेशन: मल्टी-लाइन क्वेरी स्पष्टता के लिए स्वरूपित होती हैं

```
SELECT name, ST_Area(geometry) as area
FROM neighborhoods
WHERE borough = 'Manhattan'
ORDER BY area DESC;
```

कमांड लाइन निर्देश: कमांड लाइन या टर्मिनल पर दर्ज किए जाने वाले कमांड `$` प्रॉम्प्ट के साथ दिखाए गए हैं (`$` प्रतीक स्वयं टाइप न करें):

```
$ pip install leafmap
$ python script.py
```

कोड उदाहरण डाउनलोड करना

इस पुस्तक के लिए सभी कोड उदाहरण, डेटासेट, और पूरक सामग्री GitHub पर स्वतंत्र रूप से उपलब्ध हैं:

https://github.com/giswqs/duckdb-spatial

सामग्री डाउनलोड करने के लिए, आप निम्नलिखित विधियों में से एक का उपयोग कर सकते हैं:

- रिपॉजिटरी क्लोन करें (यदि आपके पास Git स्थापित है):

```
$ git clone https://github.com/giswqs/duckdb-spatial.git
```

- **ZIP** के रूप में डाउनलोड करें (यदि आप Git का उपयोग करना पसंद नहीं करते हैं):
  - GitHub रिपॉजिटरी पृष्ठ पर जाएँ
  - हरे **Code** बटन पर क्लिक करें
  - **Download ZIP** चुनें
  - फ़ाइलों को अपने पसंदीदा स्थान पर निकालें
- यदि आपको केवल विशिष्ट उदाहरणों की आवश्यकता है तो GitHub इंटरफ़ेस के माध्यम से ऑनलाइन व्यक्तिगत फ़ाइलें ब्राउज़ करें

रिपॉजिटरी को सुधार, सुधार, और अतिरिक्त उदाहरणों के साथ नियमित रूप से अद्यतन किया जाता है। अद्यतन के लिए समय-समय पर वापस आएँ, या परिवर्तनों की सूचना प्राप्त करने के लिए GitHub पर रिपॉजिटरी देखें।

यदि आपको कोड में त्रुटियाँ मिलती हैं या सुधार के लिए सुझाव हैं, तो कृपया GitHub पर एक मुद्दा खोलें या एक पुल अनुरोध सबमिट करें। सामुदायिक योगदान इस संसाधन को सभी के लिए बेहतर बनाने में मदद करते हैं।

वीडियो ट्यूटोरियल और पूरक संसाधन

लिखित सामग्री को पूरक करते हुए, यह पुस्तक वीडियो ट्यूटोरियल की एक व्यापक श्रृंखला द्वारा समर्थित है जो प्रमुख अवधारणाओं के माध्यम से चलते हैं और अतिरिक्त उदाहरण प्रदान करते हैं:

https://tinyurl.com/duckdb-spatial-videos

वीडियो लिखित सामग्री के पूरक के लिए डिज़ाइन किए गए हैं, उसे प्रतिस्थापित करने के लिए नहीं। वे विशेष रूप से इनके लिए सहायक हैं:

- विज़ुअल शिक्षार्थी जो कोड लिखे और निष्पादित होते देखने से लाभान्वित होते हैं
- कई स्पष्टीकरणों के माध्यम से जटिल अवधारणाओं को समझना
- विकास वर्कफ़्लो और सर्वोत्तम प्रथाओं के बारे में सीखना
- समस्याओं से कैसे निपटें और मुद्दों को कैसे डिबग करें यह देखना

प्लेलिस्ट को पुस्तक की संरचना का अनुसरण करने के लिए व्यवस्थित किया गया है। आप उन्हें क्रम में देख सकते हैं जैसे-जैसे आप पुस्तक के माध्यम से प्रगति करते हैं, या आवश्यकतानुसार विशिष्ट विषयों पर जा सकते हैं।

वीडियो 2023 की शरद ऋतु में बनाए गए थे जब मैं टेनेसी विश्वविद्यालय में **Spatial Data Management**[1] पाठ्यक्रम पढ़ा रहा था। हालाँकि पाठ्यक्रम समाप्त हो गया है, वीडियो प्रासंगिक रहते हैं और इन्हें पुस्तक के संदर्भ के रूप में उपयोग किया जा सकता है। भविष्य में अतिरिक्त वीडियो जोड़े जाएंगे।

[1] https://geog-414.gishub.org

समुदाय और प्रतिक्रिया

मैं पाठकों से प्रतिक्रिया, प्रश्नों, और सुझावों का स्वागत करता हूँ। आपके इनपुट से पुस्तक को बेहतर बनाने में मदद मिलती है और यह भू-स्थानिक प्रोग्रामिंग समुदाय के लिए अधिक उपयोगी बनती है।

पुस्तक से संबंधित प्रश्नों और चर्चाओं के लिए:

- GitHub Issues: https://github.com/giswqs/duckdb-spatial/issues
- GitHub Discussions: https://github.com/giswqs/duckdb-spatial/discussions

विशेष रूप से सहायक प्रकार की प्रतिक्रिया:

- पाठ या कोड में त्रुटियाँ या अस्पष्ट स्पष्टीकरण
- अतिरिक्त उदाहरणों या उपयोग के मामलों के लिए सुझाव
- नए विषयों या अध्यायों के लिए विचार
- विभिन्न ऑपरेटिंग सिस्टम या लाइब्रेरी संस्करणों के साथ संगतता समस्याओं की रिपोर्ट
- पुस्तक से तकनीकों को कैसे लागू किया, इसकी सफलता की कहानियाँ

आभार

यह पुस्तक कई व्यक्तियों के योगदान और व्यापक ओपन-सोर्स भू-स्थानिक समुदाय के कारण मौजूद है।

सबसे पहले, मैं ऐसी असाधारण डेटाबेस प्रणाली बनाने के लिए **DuckDB** विकास टीम को धन्यवाद देना चाहता हूँ। प्रदर्शन, सरलता, और ओपन-सोर्स सिद्धांतों के प्रति उनकी प्रतिबद्धता ने स्थानिक विश्लेषण को एक बहुत व्यापक दर्शकों के लिए सुलभ बना दिया है। DuckDB की स्थानिक क्षमताओं में योगदान देने वाले टीम के सदस्यों को विशेष पहचान दी जाती है।

मैं टेनेसी विश्वविद्यालय में अपने सहयोगियों और छात्रों का आभारी हूँ जिन्होंने प्रतिक्रिया प्रदान की है, उदाहरणों का परीक्षण किया है, और **Spatial Data Management** पाठ्यक्रम के माध्यम से सामग्री को परिष्कृत करने में मदद की है। उनके प्रश्नों और अंतर्दृष्टि ने इस पुस्तक को बहुत मजबूत बनाया है।

ओपन-सोर्स भू-स्थानिक समुदाय विशेष पहचान के योग्य है। GDAL, GeoPandas, Shapely जैसी परियोजनाएँ, और अनगिनत अन्य उस आधार का निर्माण करती हैं जो आधुनिक स्थानिक विश्लेषण को संभव बनाता है। इस समुदाय की सहयोगी भावना मेरे काम को प्रेरित करती रहती है।

प्रारंभिक पाठकों का धन्यवाद जिन्होंने मसौदा अध्यायों पर प्रतिक्रिया प्रदान की और उन क्षेत्रों की पहचान करने में मदद की जिन्हें स्पष्टीकरण या सुधार की आवश्यकता थी। उनके विविध दृष्टिकोणों (अनुभवी GIS पेशेवरों से लेकर डेटा विज्ञान नवागंतुकों तक) ने यह सुनिश्चित करने में मदद की कि यह पुस्तक अपने इच्छित दर्शकों की सेवा करती है।

मैं लेखन में बिताई गई कई शामों और सप्ताहांतों के दौरान उनके धैर्य और समर्थन के लिए अपने परिवार को स्वीकार करना चाहता हूँ। उनकी समझ और प्रोत्साहन ने इस परियोजना को संभव बनाया।

अंत में, स्थानिक डेटा विश्लेषण और ओपन-सोर्स उपकरणों में आपकी रुचि के लिए आपको, पाठक को धन्यवाद। आप जैसे अभ्यासकर्ता ही नवाचार को आगे बढ़ाते हैं और भू-स्थानिक क्षेत्र को इतना रोमांचक बनाते हैं जहाँ काम किया जा सके।

यदि यह पुस्तक आपको स्थानिक समस्याओं को अधिक प्रभावी ढंग से हल करने में मदद करती है, तो सारा प्रयास सार्थक रहा है।

लेखक के बारे में

डॉ. Qiusheng Wu टेनेसी विश्वविद्यालय, नॉक्सविले में भूगोल और स्थिरता विभाग में एसोसिएट प्रोफेसर हैं। वे एक Amazon विद्वान भी हैं। डॉ. Wu का शोध क्लाउड कंप्यूटिंग और GeoAI के माध्यम से ओपन-सोर्स भू-स्थानिक विश्लेषण को आगे बढ़ाने पर केंद्रित है। वे कई व्यापक रूप से उपयोग किए जाने वाले ओपन-सोर्स Python पैकेजों के निर्माता और अनुरक्षक हैं, जिनमें Geemap[2], Leafmap[3], SAMGeo[4], और GeoAI[5] शामिल हैं, जो क्लाउड-आधारित भू-स्थानिक प्लेटफ़ॉर्म को AI-संचालित विश्लेषण और विज़ुअलाइज़ेशन के साथ एकीकृत करते हैं। डॉ. Wu का काम रिमोट सेंसिंग, पृथ्वी अवलोकन, और कृत्रिम बुद्धिमत्ता को जोड़ता है ताकि बड़े पैमाने पर भू-स्थानिक डेटा को दुनिया भर के शोधकर्ताओं, शिक्षकों, और अभ्यासकर्ताओं के लिए अधिक सुलभ, पुनरुत्पादनीय, और बुद्धिमान बनाया जा सके। उनकी ओपन-सोर्स परियोजनाएँ GitHub पर https://github.com/opengeos पर पाई जा सकती हैं।

लाइसेंसिंग और कॉपीराइट

यह पुस्तक खुले विज्ञान और खुली शिक्षा के सिद्धांतों को अपनाती है। पारदर्शिता, सीखने, और पुन: उपयोग का समर्थन करने के लिए, इस पुस्तक के कोड उदाहरण एक Creative Commons Attribution 4.0 International (CC BY 4.0) लाइसेंस के तहत जारी किए गए हैं। इसका मतलब है कि आप कोड को कॉपी, संशोधित, और वितरित करने के लिए स्वतंत्र हैं, यहाँ तक कि व्यावसायिक उद्देश्यों के लिए भी, जब तक उचित श्रेय दिया जाता है।

कृपया पुस्तक का हवाला देकर या GitHub रिपॉजिटरी से लिंक करके कोड उपयोग को श्रेय दें:

> Wu, Q. (2025). *Spatial Data Management with DuckDB: From SQL Basics to Advanced Geospatial Analytics*. Independently published. PDF edition ISBN 979-8993859705; Print edition ISBN 979-8274710572. https://duckdb.gishub.org

जबकि कोड स्वतंत्र रूप से उपलब्ध है, इस पुस्तक के पाठ, आंकड़े, और चित्र लेखक द्वारा कॉपीराइट हैं और स्पष्ट अनुमति के बिना पुन: प्रस्तुत, पुनर्वितरित, या संशोधित नहीं किए जा सकते हैं। इसमें सभी लिखित सामग्री, कस्टम आरेख, और एम्बेडेड विज़ुअलाइज़ेशन शामिल हैं जब तक अन्यथा नोट न किया गया हो।

यदि आप पुस्तक से किसी भी गैर-कोड सामग्री का पुन: उपयोग या अनुकूलन करना चाहते हैं (उदाहरण के लिए, शिक्षण, प्रस्तुतियों, या प्रकाशनों के लिए), तो कृपया अनुमति का अनुरोध करने के लिए लेखक से संपर्क करें।

यह दोहरा लाइसेंसिंग दृष्टिकोण मूल रचनात्मक कार्य की सुरक्षा के साथ सीखने की सामग्री तक खुली पहुँच को संतुलित करने में मदद करता है। इन शर्तों का सम्मान करने और ओपन-सोर्स भू-स्थानिक समुदाय का समर्थन करने के लिए धन्यवाद।

[2] https://geemap.org
[3] https://leafmap.org
[4] https://samgeo.gishub.org
[5] https://opengeoai.org