

Python Tutorial – Part II: Data

Mark A. Austin

University of Maryland

austin@umd.edu

ENCE 201, Fall Semester 2025

September 2, 2025

Overview

- ① Data-Driven Decision Making
- ② Tabular and Non-Tabular Data Models
 - Tabular and Non-Tabular Data Models
 - Homogeneous and Heterogeneous Data
- ③ Tabular Data and Dataset Transformation (Pandas)
 - Basic Operations (Data Series and Dataframes)
 - Intermediate Operations (Cleaning Data)
 - Advanced Operations (Data Filtering, Data Merge)
- ④ Spatial Data and Dataset Transformation (GeoPandas)
 - GeoPandas Data Model
 - Models of Geometric Objects (points, lines, polygons)
 - Applications (Urban and Global GeoDataModeling)
- ⑤ Appendix A: From Data Models to Data Structures

Applications

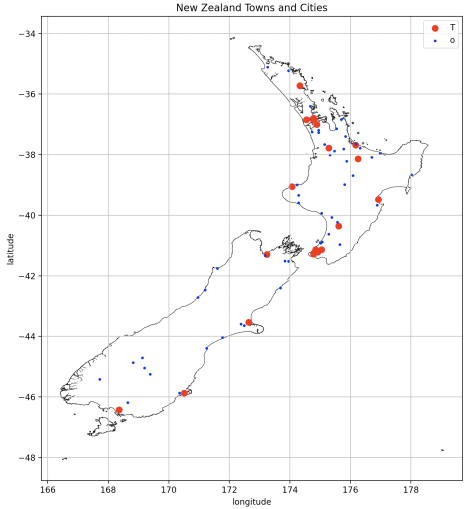
(Towns and Cities in New Zealand)

(The Worlds Megacities, Streets in Manhattan)

(Traffic Accidents in NYC, Flights to/from BWI)

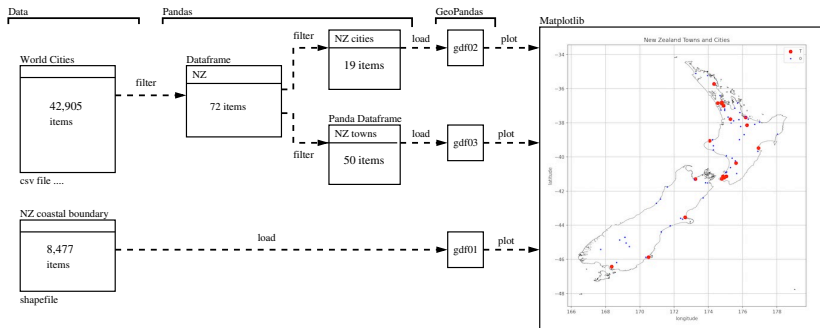
Example 2: Towns and Cities in New Zealand

Example 2: Towns and Cities in New Zealand.



Example 2: Towns and Cities in New Zealand

Part I: Data Processing Pipeline: Use sequence of filters to specialize views of data ...



Example 2: Towns and Cities in New Zealand

Part II: Program Source Code:

```
1  # =====
2  # TestNewZealandDataModel.py. Assemble data model for towns and cities in
3  # New Zealand.
4  #
5  # Written by: Mark Austin                                     February 2023
6  # =====
7
8  from pandas import DataFrame
9  from pandas import Series
10 from pandas import read_csv
11
12 import numpy as np
13 import pandas as pd
14 import geopandas
15
16 import matplotlib.pyplot as plt
17
18 # =====
19 # main method ...
20 # =====
21
22 def main():
23     print("--- Enter TestNewZealandDataModel.main()      ... ");
24     print("--- ===== ... ");
25
26     print("--- Part 01: Load world city dataset ... ");
```

Example 2: Towns and Cities in New Zealand

Part II: Program Source Code: (Continued) ...

```

27
28     df = pd.read_csv("../data/cities/world-cities.csv")
29
30     print("--- Part 02: Print dataframe info and contents ... ");
31
32     print(df)
33     print(df.info() )
34
35     print("--- Part 03: Filter dataframe to keep only cities from New Zealand ... ")
36
37     options = ['New Zealand']
38     dfNZ      = df [ df['country'].isin(options) ].copy()
39
40     print("--- Part 04: Filter data to find NZ cities and towns ... ")
41
42     dfNZcities = dfNZ [ (dfNZ['population'] > 40000) ].sort_values( by=['population'] )
43
44     dfNZtowns  = dfNZ [ (dfNZ['population'] > 1000) & (dfNZ['population'] < 40000) ]
45     dfNZtowns  = dfNZtowns.sort_values( by=['population'] )
46
47     print('--- New Zealand Cities:\n', dfNZcities )
48     print('--- New Zealand Towns:\n', dfNZtowns )
49
50     print("--- Part 05: Read NZ coastline shp file into geopandas ... ")
51
52     nzboundarydata = geopandas.read_file("../data/geography/nz/Coastline02.shp")
53     print(nzboundarydata)

```

Example 2: Towns and Cities in New Zealand

Part II: Program Source Code: (Continued) ...

```
55     print("--- Part 06: Define geopandas dataframes ... ")
56
57     gdf01 = geopandas.GeoDataFrame(nzboundarydata)
58     gdf02 = geopandas.GeoDataFrame( dfNZcities ,
59         geometry=geopandas.points_from_xy(dfNZcities.lng, dfNZcities.lat))
60     gdf03 = geopandas.GeoDataFrame( dfNZtowns ,
61         geometry=geopandas.points_from_xy( dfNZtowns.lng, dfNZtowns.lat))
62
63     print(gdf01.head())
64
65     print("--- Part 07: Create boundary map for New Zealand ... ")
66
67     # We can now plot our ‘GeoDataFrame’.
68
69     ax = gdf01.plot( color='white', edgecolor='black')
70     ax.set_aspect('equal')
71     ax.set_title("New Zealand Towns and Cities")
72
73     gdf01.plot(ax=ax, color='white')
74
75     gdf02.plot(ax=ax, color = 'red', markersize = 50, label= 'Cities')
76     gdf03.plot(ax=ax, color = 'blue', markersize = 5, label= 'Towns' )
77
78     plt.legend('Towns/Cities:')
79     plt.xlabel('longitude')
80     plt.ylabel('latitude')
```

Example 2: Towns and Cities in New Zealand

Part II: Program Source Code: (Continued) ...

```
81     plt.grid(True)
82     plt.show()
83
84     print("--- ===== ... ");
85     print("--- Leave TestNewZealandDataModel.main() ... ");
86
87     # =====
88     # call the main method ...
89     # =====
90
91     main()
```

Source Code: See: python-code.d/geopandas/

Example 2: Towns and Cities in New Zealand

Part III: Abbreviated Output:

```
--- Enter TestNewZealandDataModel.main() ...
--- ===== ...
--- Part 01: Load world city dataset ...
--- Part 02: Print dataframe info and contents ...
      city city_ascii   lat ... capital population      id
0      Tokyo      Tokyo  35.6839 ... primary 39105000.0 1392685764
1      Jakarta  Jakarta -6.2146 ... primary 35362000.0 1360771077
...      ...      ...      ...      ...      ...      ...
42903 Timmiarmiut Timmiarmiut 62.5333 ... NaN      10.0 1304206491
42904 Nordvik      Nordvik  74.0165 ... NaN      0.0 1643587468
[42905 rows x 11 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42905 entries, 0 to 42904
Data columns (total 11 columns):
```

#	Column	Dtype	#	Column	Dtype
0	city	object	6	iso3	object
1	city_ascii	object	7	admin_name	object
2	lat	float64	8	capital	object
3	lng	float64	9	population	float64
4	country	object	10	id	int64
5	iso2	object			

```
dtypes: float64(3), int64(1), object(7)
```

```
memory usage: 3.6+ MB
```

Example 2: Towns and Cities in New Zealand

Part III: Abbreviated Output (Continued) ...

```
--- Part 03: Filter dataframe to keep only cities from New Zealand ...
```

```
--- Part 04: Filter data to find NZ cities and towns ...
```

```
--- New Zealand Cities:
```

	city	city_ascii	...	population	id
14169	Upper Hutt	Upper Hutt	...	41000.0	1554000042
6159	Invercargill	Invercargill	...	47625.0	1554148942
.....					
741	Wellington	Wellington	...	418500.0	1554772152
516	Auckland	Auckland	...	1346091.0	1554435911

```
[19 rows x 11 columns]
```

```
--- New Zealand Towns:
```

	city	city_ascii	...	population	id
42142	Kaikoura	Kaikoura	...	2210.0	1554578431
.....					
14309	Whanganui	Whanganui	...	39400.0	1554827998

```
[50 rows x 11 columns]
```

```
--- Part 05: Read NZ coastline shp file into geopandas ...
```

```
0 POLYGON ((174.00369 -40.66489, 174.00372 -40.6...
.....
8476 POLYGON ((173.01384 -34.39348, 173.01395 -34.3...
[8477 rows x 1 columns]
```

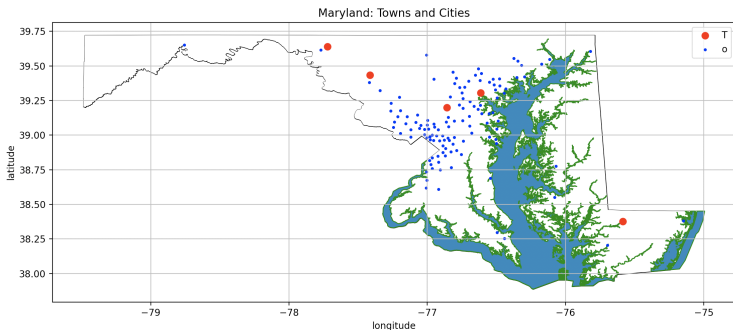
```
--- Part 07: Create boundary map for New Zealand ...
```

```
--- ===== ...
```

```
--- Leave TestNewZealandDataModel.main() ...
```

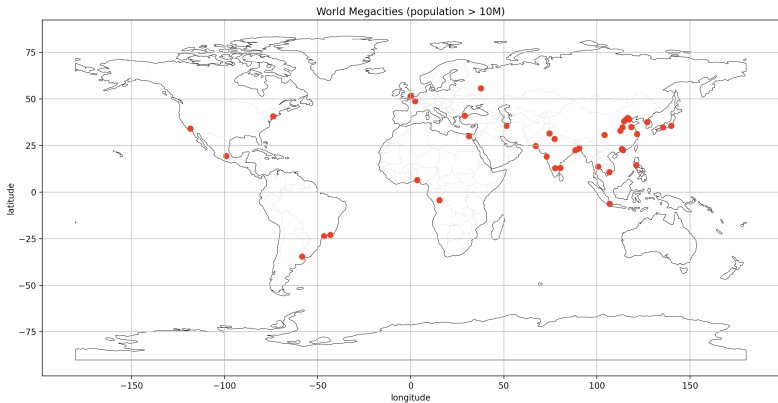
Example 3: Towns and Cities in Maryland

Example 3: Towns and Cities in Maryland.



Cities: Columbia (pop. 103991), Salisbury (pop. 106447), Frederick (pop. 156787), Hagerstown (pop. 184755), Baltimore (pop. 2106068).

Example 5: The World's Megacities



Example 5: The World's Megacities

```
--- Part 02: Filter to keep only large cities (pop. > 10M) ...
```

	city	city_ascii	...	population	id
0	Tokyo	Tokyo	...	39105000.0	1392685764
1	Jakarta	Jakarta	...	35362000.0	1360771077
2	Delhi	Delhi	...	31870000.0	1356872604
3	Manila	Manila	...	23971000.0	1608618140
4	São Paulo	Sao Paulo	...	22495000.0	1076532519
5	Seoul	Seoul	...	22394000.0	1410836482
6	Mumbai	Mumbai	...	22186000.0	1356226629
7	Shanghai	Shanghai	...	22118000.0	1156073548
8	Mexico City	Mexico City	...	21505000.0	1484247881
9	Guangzhou	Guangzhou	...	21489000.0	1156237133
10	Cairo	Cairo	...	19787000.0	1818253931
11	Beijing	Beijing	...	19437000.0	1156228865
12	New York	New York	...	18713220.0	1840034016
13	Kolkāta	Kolkata	...	18698000.0	1356060520
14	Moscow	Moscow	...	17693000.0	1643318494
15	Bangkok	Bangkok	...	17573000.0	1764068610

```
... details removed ...
```

33	London	London	...	11120000.0	1826645935
34	Paris	Paris	...	11027000.0	1250015082
35	Tianjin	Tianjin	...	10932000.0	1156174046
36	Linyi	Linyi	...	10820000.0	1156086320
37	Shijiazhuang	Shijiazhuang	...	10784600.0	1156217541
38	Zhengzhou	Zhengzhou	...	10136000.0	1156183137
39	Nanyang	Nanyang	...	10013600.0	1156192287

Example 6: Main Streets in Lower Manhattan



Example 6: Main Streets in Lower Manhattan

Abbreviated Datafile: 720 lines of data ...

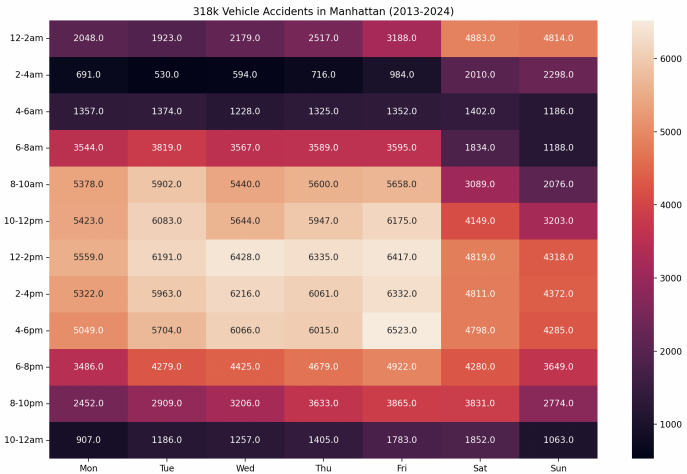
the_geom,Borough,Boro_Code,Route_Name,Route_Type,Route_Sub,Route_Stat

```
"MULTILINESTRING ( (-73.9910241043406 40.72364943251929,  
-73.99102457806407 40.72373532313096,  
-73.99099489093685 40.72400907472583,  
-73.99081902422084 40.72427429772359, ... ))",  
Manhattan,1,2 Avenue,Major Streets,  
Major street to be improved,Existing
```

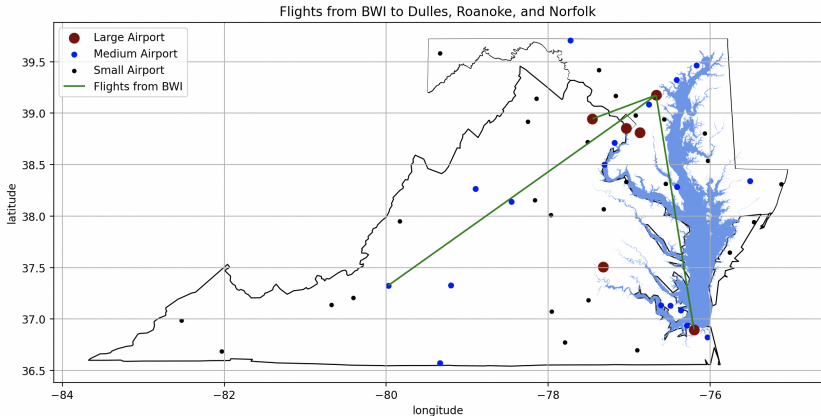
....

```
"MULTILINESTRING ((-74.01166334821232 40.71279805585768,  
-74.0109993711713 40.71332952150041,  
-74.01056054567228 40.71384627551561,  
-74.01027954718754 40.71419488854311,  
-74.01012810711585 40.71438276644835,  
-74.00968038665206 40.71494543483944,  
...  
-74.00923907482891 40.71552016776998, ... ))",  
Manhattan,1,West Broadway,Major Streets,  
Major Street,Existing
```


Example 7: Traffic Accidents in Lower Manhattan

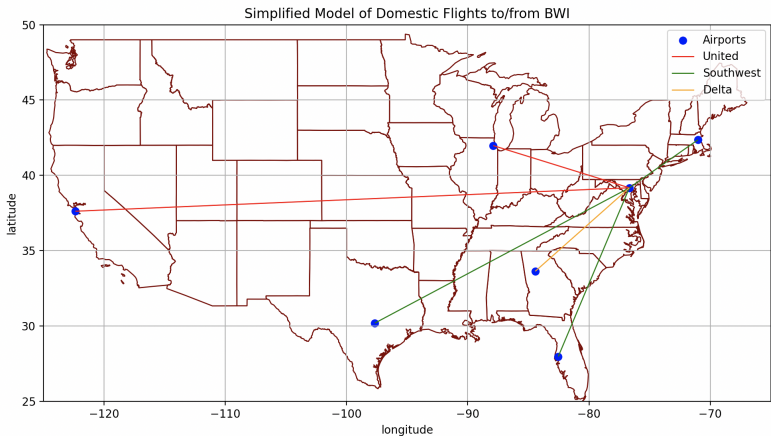


Example 8: Flights from BWI



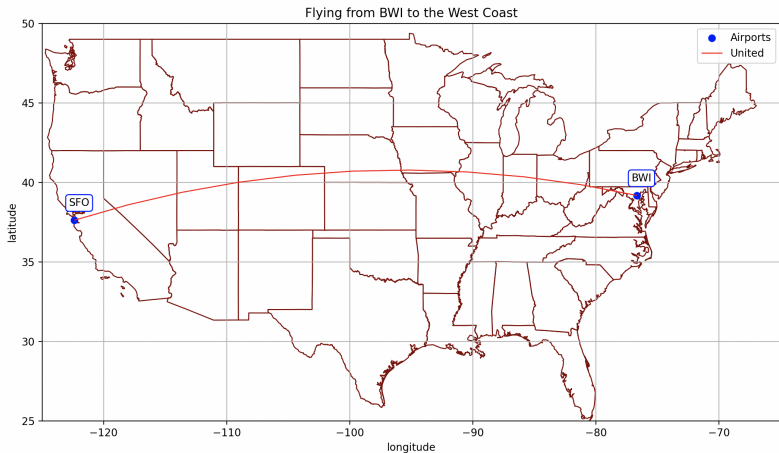
BWI Airport: GeoLocation (long, lat) = (-76.668297, 39.175400) ...

Example 8: Flights from BWI



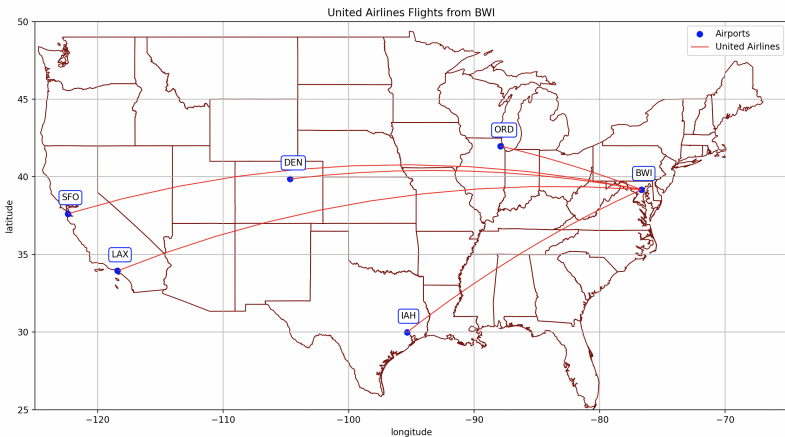
Source Code: See [python-code.d/applications/transportation/air/TestAirTransportationUSA01.py](#)

Example 8: Flights from BWI

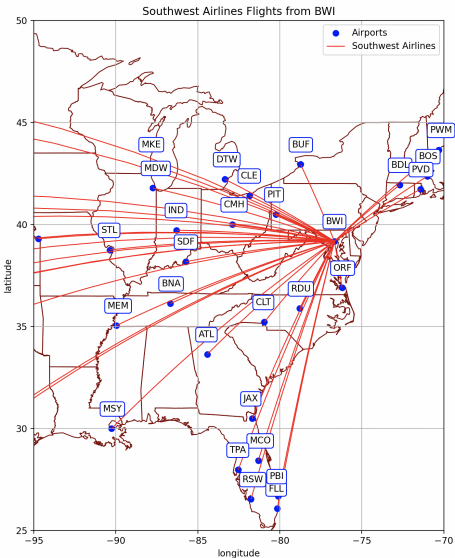


Source Code: See python-code.d/applications/transportation/air/TestAirTransportationUSA02.py

Example 8: United Flights from BWI



Example 8: Southwest Flights from BWI



New for Fall Semester 2025

