

Perform data engineering tasks in a notebook

Imagine that you work as a GIS analyst for a nonprofit community-planning organization in Portland, Oregon. Your firm has been awarded a contract from the state of Oregon to determine which metro areas are most prone to economic vulnerability. You have been provided with several datasets containing city- and block-level data for the Portland metro region.

In this exercise, you will perform several spatial data engineering tasks to prepare the datasets for analysis using an ArcGIS notebook.

Estimated completion time in minutes: 25

To complete exercises, you need the following:

- ArcGIS Pro 3.5 (Basic, Standard, or Advanced)

① Download the data

To complete the exercise, you must [download the data](#). If you have already downloaded and installed the data, continue to the next step.

Unable to find the data you downloaded?

After you have downloaded the data ZIP file, extract it to the C:\EsriTraining folder. The unzipped folder will have the same name as the ZIP file.

If you unzipped the data to a location other than C:\EsriTraining, browse to that location and locate the folder. You can also try searching for the folder within File Explorer.

② Start ArcGIS Pro and prepare a notebook

First, you will start ArcGIS Pro and open an ArcGIS notebook.

a Start ArcGIS Pro.

b If necessary, sign in to ArcGIS Pro using your ArcGIS Online organizational account.

- c Near Recent Projects, click Open Another Project .



Note: If you have configured ArcGIS Pro to start without a project template or with a default project, you will not see the Start page. On the Project tab, click Open, and then click Open Another Project.

- d On the left side of the Open Project dialog box, click This PC.
- e On the right side, browse to **C:\EsriTraining\NotebooksDSW\WorkingWithData**.
- f Select WorkingWithData.aprx and click OK.

The project opens to a map that displays areas of economic vulnerability in Portland, Oregon.

- g From the Catalog pane, expand Notebooks to view the notebooks that are included in the project.

- h Right-click the DataEngineering.ipynb notebook to view some of the file options.

From here, you can create and modify a notebook's metadata, rename or copy a notebook, or open a notebook.

- i Choose Open.

- j Right-click the DataEngineering tab and choose New Horizontal Tab Group.

Now that you have successfully opened the notebook, you will modify it to access data from a local file system.

③ Load the data

Using ArcGIS Notebooks functionality, you will code a notebook cell to load analysis-specific data into the notebook.

- a

In the notebook, confirm that the cell importing the libraries to be used in the notebook is selected.

[View result →](#)



```
[ ]: import arcpy
import os
```

Step 3a: Load the data.

b To import the libraries, click the Run button ►.

Which Python libraries are available to the notebook?

[Show answer](#)

arcpy, os

The second cell in the notebook is currently selected.

c

In the second cell, set the environment's workspace property to **r"C:\EsriTraining\NotebooksDSW\WorkingWithData\WorkingWithData.gdb"**.

[View code](#)

```
arcpy.env.workspace = r"C:\
EsriTraining\NotebooksDSW\WorkingWithData\WorkingWithData.gdb"
```

Setting the environment's workspace property to the file-based geodatabase makes the geodatabase's feature classes available for analysis in the notebook.

d

To set the environment's workspace, click the Run button ►.

The third cell in the notebook is currently selected.

The first two lines of code in the cell are used to access the WorkingWithData geodatabase's feature classes. The Vulnerability feature class represents areas of economic vulnerability in Portland using a score that ranges from 0 to 100, where a score of 60 or greater is considered vulnerable. The Socioeconomic Status feature class

represents socioeconomic status using an index that ranges from 0 to 100, where 100 represents a high socioeconomic status. First, you will clean each of these feature classes. Then, you will join these feature classes to better understand the economic vulnerability of residents that live in each census tract across the city.

e To access the feature classes in the specified geodatabase, click the Run button ►.

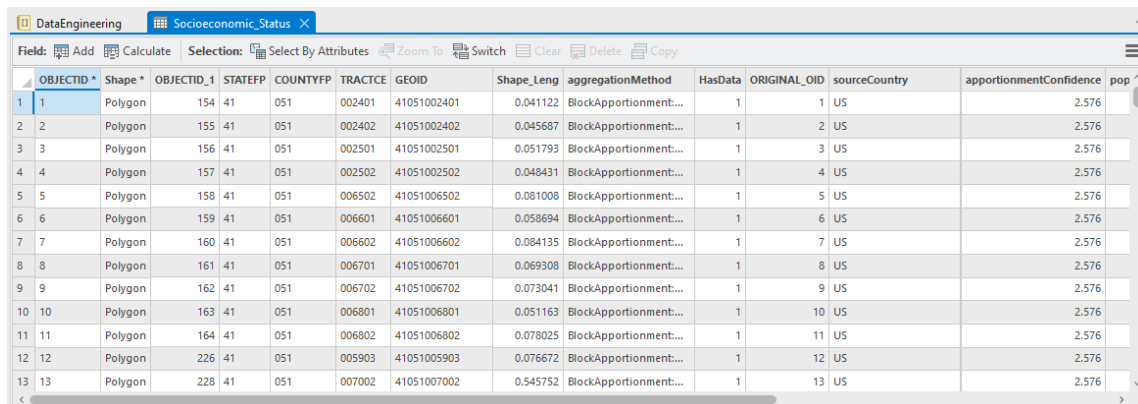
Now that you have successfully imported the libraries into the notebook and modified the notebook to access feature classes from the WorkingWithData geodatabase, you will modify the notebook to clean the data.

④ Clean the data

You will clean the data by adding code to the notebook to remove any null values and format some of the field values. The first step in cleaning the data is to open the feature class's attribute tables to inspect the data for irregularities and null values.

a In the Contents pane, right-click the Socioeconomic_Status feature class and choose Attribute Table to view the attribute table's values.

[View result ↗](#)



The screenshot shows a software interface with a tab titled 'Socioeconomic_Status'. Below the tab is a table with the following columns: OBJECTID, Shape, OBJECTID_1, STATEFP, COUNTYFP, TRACTCE, GEOID, Shape_Leng, aggregationMethod, HasData, ORIGINAL_OID, sourceCountry, apportionmentConfidence, and pop. The table contains 13 rows of data, each representing a census tract. The first row is highlighted in blue.

OBJECTID	Shape	OBJECTID_1	STATEFP	COUNTYFP	TRACTCE	GEOID	Shape_Leng	aggregationMethod	HasData	ORIGINAL_OID	sourceCountry	apportionmentConfidence	pop
1	Polygon	154	41	051	002401	41051002401	0.041122	BlockApportionment...	1	1	US	2.576	
2	Polygon	155	41	051	002402	41051002402	0.045687	BlockApportionment...	1	2	US	2.576	
3	Polygon	156	41	051	002501	41051002501	0.051793	BlockApportionment...	1	3	US	2.576	
4	Polygon	157	41	051	002502	41051002502	0.048431	BlockApportionment...	1	4	US	2.576	
5	Polygon	158	41	051	006502	41051006502	0.081008	BlockApportionment...	1	5	US	2.576	
6	Polygon	159	41	051	006601	41051006601	0.058694	BlockApportionment...	1	6	US	2.576	
7	Polygon	160	41	051	006602	41051006602	0.084135	BlockApportionment...	1	7	US	2.576	
8	Polygon	161	41	051	006701	41051006701	0.069308	BlockApportionment...	1	8	US	2.576	
9	Polygon	162	41	051	006702	41051006702	0.073041	BlockApportionment...	1	9	US	2.576	
10	Polygon	163	41	051	006801	41051006801	0.051163	BlockApportionment...	1	10	US	2.576	
11	Polygon	164	41	051	006802	41051006802	0.078025	BlockApportionment...	1	11	US	2.576	
12	Polygon	226	41	051	005903	41051005903	0.076672	BlockApportionment...	1	12	US	2.576	
13	Polygon	228	41	051	007002	41051007002	0.545752	BlockApportionment...	1	13	US	2.576	

Step 4a: Clean the data.



Note: You can also press Ctrl and double-click the layer to open its attribute table.

b

Browse each record to determine whether the feature class has any null feature values.

The Socioeconomic_Status feature class was previously cleaned and is ready for analysis. You will now verify that the Vulnerability feature class has been cleaned.

c

In the Contents pane, open the Vulnerability attribute table.

d

Browse the table to determine whether the Vulnerability feature class has any null feature values.

The Vulnerability feature class contains several rows that have <Null> values. Before you can begin cleaning the fields, you need to obtain a list of the feature class's fields.

e

In the notebook, in the cell that creates the `fieldList`, type the **`fc_vulnerability`** variable name to get a list of fields from the Vulnerability feature class.

[View code](#)

```
fieldList = [f.name for f in
arcpy.ListFields(fc_vulnerability) if not f.required]
print(fieldList)
```

f

To obtain a list of fields from the Vulnerability feature class, click the Run button ►.

```
fieldList = [f.name for f in arcpy.ListFields(fc_vulnerability) if not f.required]
print(fieldList)

['OBJECTID', 'Shape_Leng', 'STATEFP20', 'COUNTYFP20', 'TRACTCE20', 'GEOID20', 'NAME20', 'vulnerabil', 'flag_vulne', 'people_of',
_, 'black_nati', 'adults_wit', 'adjusted_h', 'housing_co', 'households', 'persons_wi', 'older_adul', 'youth_0_21', 'hi_people',
_, 'hi_black_n', 'hi_adults_', 'hi_adjuste', 'hi_housing', 'hi_househo', 'hi_persons', 'hi_retiree', 'hi_youth_0', 'pdx_peopl',
e', 'pdx_black_', 'PDX_adults', 'pdx_adjust', 'pdx_housin', 'pdx_househ', 'pdx_person', 'pdx_older_', 'pdx_youth_', 'Total_pop',
u', 'Total_hous', 'total_po_1', 'total_ho_1']
```

Step 4f: Clean the data.

[View result](#)

A list of the fields is displayed in the cell's output.

The fifth cell in the notebook is currently selected. This cell is used to remove features

based on a field value. You will now modify this cell to remove all features that have a null value.

g

In the fifth cell, locate the line of code in the `for` loop where the value of the field is being tested using an `if` conditional statement, as specified in the following graphic.

```
for field in fieldList:

    with arcpy.da.UpdateCursor(fc_vulnerability, field) as cursor:

        for row in cursor:
            value = row[0]
            if value == :
                print("Deleted NULL value in the " + field + " field")
                cursor.deleteRow()

print("Deleting NULL values completed")
```

h

Type **None** as the value to be tested.

[View code](#)

```
for field in fieldList:

    with arcpy.da.UpdateCursor(fc_vulnerability, field) as
cursor:

        for row in cursor:
            value = row[0]
            if value == None:
                print("Deleted NULL value in the " + field
+ " field")
                cursor.deleteRow()

print("Deleting NULL values completed")
```

i

To delete all features with <Null> attribute values, click the Run button ►.

What is the name of the field that contained NULL values?

[Show answer](#)

vulnerabil

You successfully removed all features with null values. There are three fields in the feature class that are not required for future analysis. The `drop_fields` variable represents a list of fields by name to remove from the feature class. You will now use the **DeleteField_management** geoprocessing tool to drop the specified field from the feature class.

What are the names of the fields to be removed from the Vulnerability feature class?

[Show answer](#)

STATEFP20, COUNTYFP20, and GEOID20

j

In the sixth cell, to drop the specified fields from the Vulnerability feature class, click the Run button ►.

k

In the Vulnerability attribute table, browse to verify that the null feature values have been removed from the feature class and that the specified fields have been removed.

Now that you have successfully cleaned the data, you will enrich the data by joining the Vulnerability feature class to another feature class.

5 Join the data

Before you can discover areas that are most prone to economic vulnerability, you will join the Vulnerability feature class with the Socioeconomic_Status feature class. In this step, you will add code to the notebook that uses the Add Join tool to join the two feature classes.

a

Browse the fields in the Vulnerability and Socioeconomic_Status attribute tables to determine which fields are common to both feature classes and could be used to join the tables.

Which fields could be used to join the two feature classes and why?

[Show answer](#)

The TRACTCE20 and TRACTCE fields are specific to each feature and would be appropriate to use.

b

In the notebook, in the cell used to create the join, type **TRACTCE20** as the `in_field`.

c

For the `join_field`, type **TRACTCE**.

[View code](#)

```
arcpy.env.overwriteOutput = False
joined_features =
arcpy.AddJoin_management(in_layer_or_view=fc_vulnerability,
in_field="TRACTCE20", join_table=fc_socioeconomic,
join_field="TRACTCE", join_type="KEEP_COMMON")
```

d

To join the Socioeconomic_Status feature class to the Vulnerability feature class, click the Run button ►.

e

In the Vulnerability attribute table, scroll horizontally through the table to view the fields inserted into the Vulnerability feature class.

What are some of the fields added to the Vulnerability feature class?

[Show answer](#)

The new fields include aggregationMethod, apportionmentConfidence, and 2023 Socioeconomic Status Index.

- f** Close any open attribute tables.
- g** If necessary, activate the DataEngineering notebook.
- h** On the Notebook tab, click Save.
- i** Close the DataEngineering notebook.

You will now select census tracts that have high economic vulnerability (with a vulnerability score greater than or equal to 60 on a scale of 0 to 100) and a lower socioeconomic status (with an index value less than 50 on a scale of 0 to 100).

- j** On the Map tab, in the Selection group, click Select By Attributes.
- k** For Input Rows, choose Vulnerability, if necessary.
- l** Under Expression, for Select A Field, choose Vulnerabil.
- m** For Is Equal To, click the down arrow and choose Is Greater Than Or Equal To.
- n** For the value field, type **60**.
- o** Click Add Clause.

- p**
Build the following expression: And 2023 Socioeconomic Status Index Is Less Than 50.

[View result →](#)

Input Rows
Vulnerability

Selection Type
New selection

Expression
Load Save Remove

SQL ☐

Where vulnerabil is greater 60

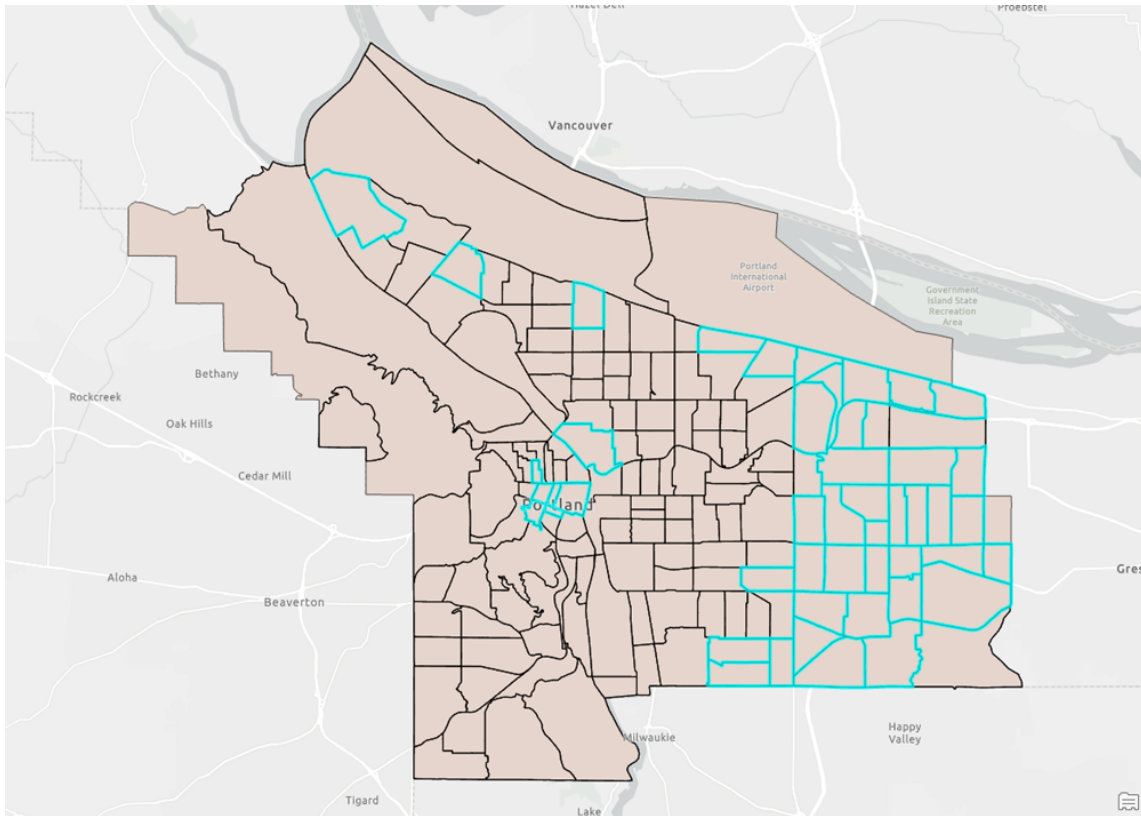
And 2023 Socioecon is less than 50

+ Add Clause

☐ Invert Where Clause

Step 5p: Join the data.

q Click OK.



Step 5q: Join the data.

[View result](#) 

The selected features in the map represent areas that have higher economic vulnerability and a lower socioeconomic status.

r On the Map tab, in the Selection group, click Clear .

s Close the Data Engineering map.

t Save the project.

u

If you are continuing to the next exercise, leave ArcGIS Pro open; otherwise, exit ArcGIS Pro.

In this exercise, you used an ArcGIS notebook in ArcGIS Pro to create a mechanism for cleaning the Vulnerability feature class and joining the feature class to the Socioeconomic_Status attribute table for future analysis.

Copyright © Esri
All rights reserved.

Published in the United States of America.

The information contained in this document is the exclusive property of Esri. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by Esri. All requests should be sent to Attention: Director, Contracts and Legal, Esri, 380 New York Street, Redlands, CA 92373-8100, USA.

Export Notice: Use of these Materials is subject to U.S. export control laws and regulations including the U.S. Department of Commerce Export Administration Regulations (EAR). Diversion of these Materials contrary to U.S. law is prohibited.

The information contained in this document is subject to change without notice.

Commercial Training Course Agreement Terms: The Training Course and any software, documentation, course materials or data delivered with the Training Course is subject to the terms of the Master Agreement for Products and Services found at the following website: esri.com/TrainingTerms. The license rights in the Master Agreement strictly govern Licensee's use, reproduction, or disclosure of the software, documentation, course materials and data. Training Course students may use the course materials for their personal use and may not copy or redistribute for any purpose. Contractor/Manufacturer is Esri, 380 New York Street, Redlands, CA 92373-8100, USA.

Esri Marks: Esri marks and product names mentioned herein are subject to the terms of use found at the following website: esri.com/EsriMarks.

Other companies and products or services mentioned herein may be trademarks, service marks, or registered marks of their respective mark owners.