

SPATIAL DATA SCIENCE & ANALYTICS

Trainer 1: Lebogang Mashishi

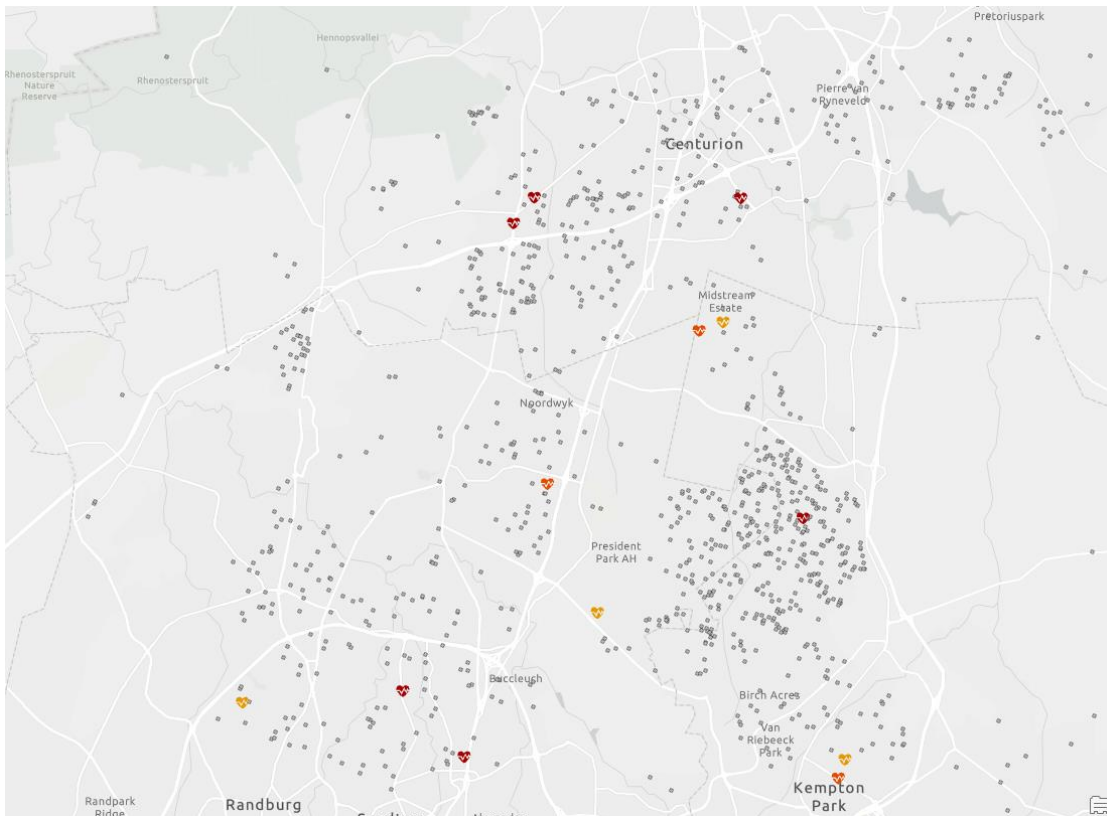
Trainer 2: Liezel Botha

DATA ENGINEERING

What do we need?

Ask | Clients:

1. Data from Dr's – Location
2. Data from patients – Location, patient info, disease



Explore the data?

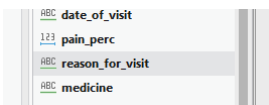
3. Map < Data Engineering
4. Patient | Visit < Data Engineering < Add All Fields & Calculate

Field Name	Alias	Field Type	Nulls	Chart Preview	Minimum	Maximum	Mean	Standard Deviation	Medi
first_name	first_name	Text (255)	0 (0%)	[Bar Chart]					
last_name	last_name	Text (255)	0 (0%)	[Bar Chart]					
address	address	Text (8000)	4 (0.4%)	[Bar Chart]					
gender	gender	Text (255)	12 (1.2%)	[Bar Chart]					
birth_year	birth_year	Long	36 (3.6%)	[Histogram]	1,921	2,022	1,979.7	29.8	1
age_2023	age_2023	Long	13 (1.3%)	[Histogram]	-10	115	43.3	29.9	
weight	weight	Long	6 (0.6%)	[Histogram]	1	120	57.3	33.7	
health_status	health_status	Text (255)	24 (2.4%)	[Bar Chart]					
date_of_visit	date_of_visit	Text (255)	22 (2.2%)	[Bar Chart]					
pain_perc	pain_perc	Long	8 (0.8%)	[Histogram]	1	125	49.8	29.5	

5. Reason for Visit < Missing Values (15%)

pain_perc	pain_perc	Long	8 (0.8%)	[Histogram]
reason_for_visit	reason_for_visit	Text (255)	15 (1.5%)	[Bar Chart]
medicine	medicine	Text (255)	16 (1.6%)	[Bar Chart]

6. Reason for Visit < Update Symbology & Create Chart (colours)



7. Chart Preview < Encode Field

c	reason_for_visit	medicine	type_cover	cost_of_visit	date_x	abdominal pain (reason_for_visit_One-hot)	allergic reaction (reason_for_visit_One-hot)	anxiety (re
1	7	short of breath	item 5	medical aid	3905.24	2022/09/06	0	0
2	5	allergic reaction	item 3	cash	6651.74	2023/07/28	0	1
3	4	cough	item 3	cash	3927.18	2023/03/26	0	0
4	4	tonsillitis	item 3	medical aid	9104.46	2023/07/06	0	0
5	3	sore throat	item 5	cash	8674.85	2022/11/22	0	0
6	3	sore throat	item 1	medical aid	9180.13	2022/07/19	0	0
7	3	sore throat	item 5	cash	4575.51	2022/12/18	0	0
8	7	tonsillitis	item 1	medical aid	4547.1	2023/07/03	0	0
9	5	fever	item 3	cash	4668.42	2022/12/28	0	0
10	9	cough	item 3	medical aid	7657.8	2022/07/29	0	0
11	7	abdominal pain	item 5	medical aid	1052.32	2023/08/03	1	0
12	3	headache	item 2	medical aid	3843.94	2022/04/08	0	0

- 8. Attribute Table < Add 2 Fields < Date X (date) & Day of Week (text)
- 9. Calculate Field < Python < Date X < Day of Visit & Day
- 10. Calculate Field < Arcade < Day of Week < ISOWeekday(\$feature.date_x)

SPATIAL ANALYSIS

11. Identify times & areas of high & low costs:
12. Map < Health Analysis
13. Patient | Visit < Data Clock < Date X < Rings = Years < Wedges = Months < Mean < Cost of Visit

Variables

Date
date_x

Rings
Years

Wedges
Months

Aggregation
Mean

Number
cost_of_visit

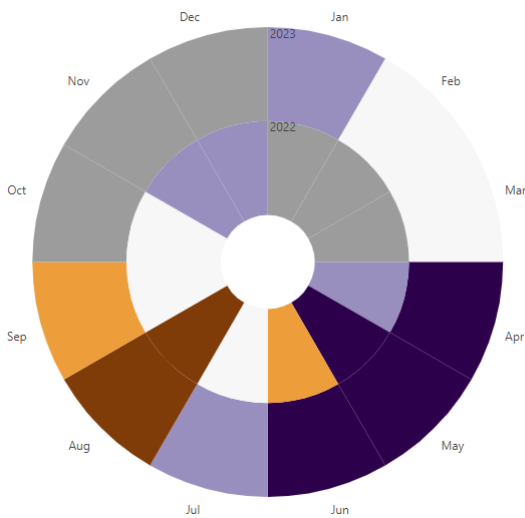
Method
Natural Breaks (Jenks)

Classes
5

Color scheme

Null

Cost of Visit | By Month | Over Years



1. Cluster & Outlier Analysis < Patient | Visit < Cost of Visit < Dr Cost Moran I < K Nearest Neighbours < 8



Geoprocessing ⌵ ⌵ ✕

⬅ Cluster and Outlier Analysis (Anselin Local Moran's I) ➕

Parameters Environments ?

Input Feature Class
Patient | Visit

Input Field
cost_of_visit

Output Feature Class
Dr_Cost_Moran_I

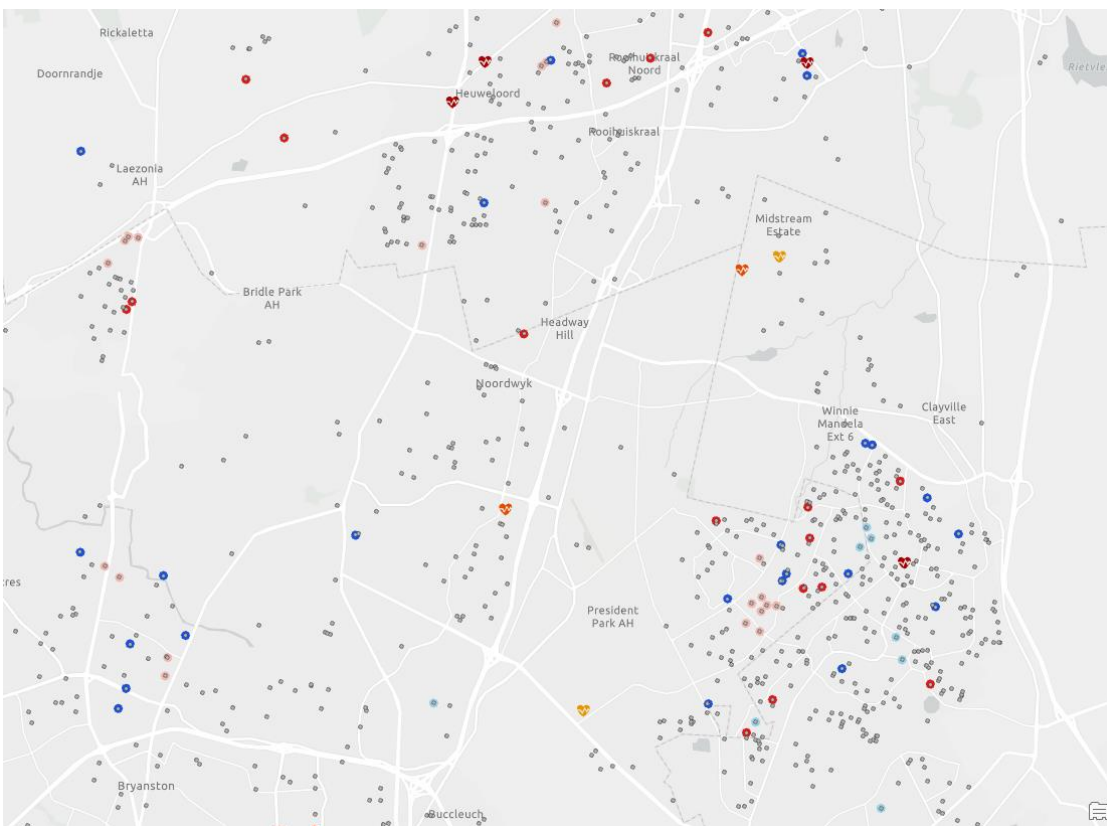
Conceptualization of Spatial Relationships
K nearest neighbors

Standardization
Row

Number of Neighbors

Apply False Discovery Rate (FDR) Correction

Number of Permutations



What factors explain higher cost?

1. Calculate Composite Index < Patient | Visit < Patient Visit Index < Age 2023, Weight, Pain Perc < Mean



Geoprocessing ⌵ ⌵ ✕

← Calculate Composite Index ⊕

Parameters Environments ?

Input Table
 Patient | Visit 📁

Append Fields to Input Table

Output Features or Table
 Patient_Visit_Index 📁

Input Variables ⌵ ⚙️

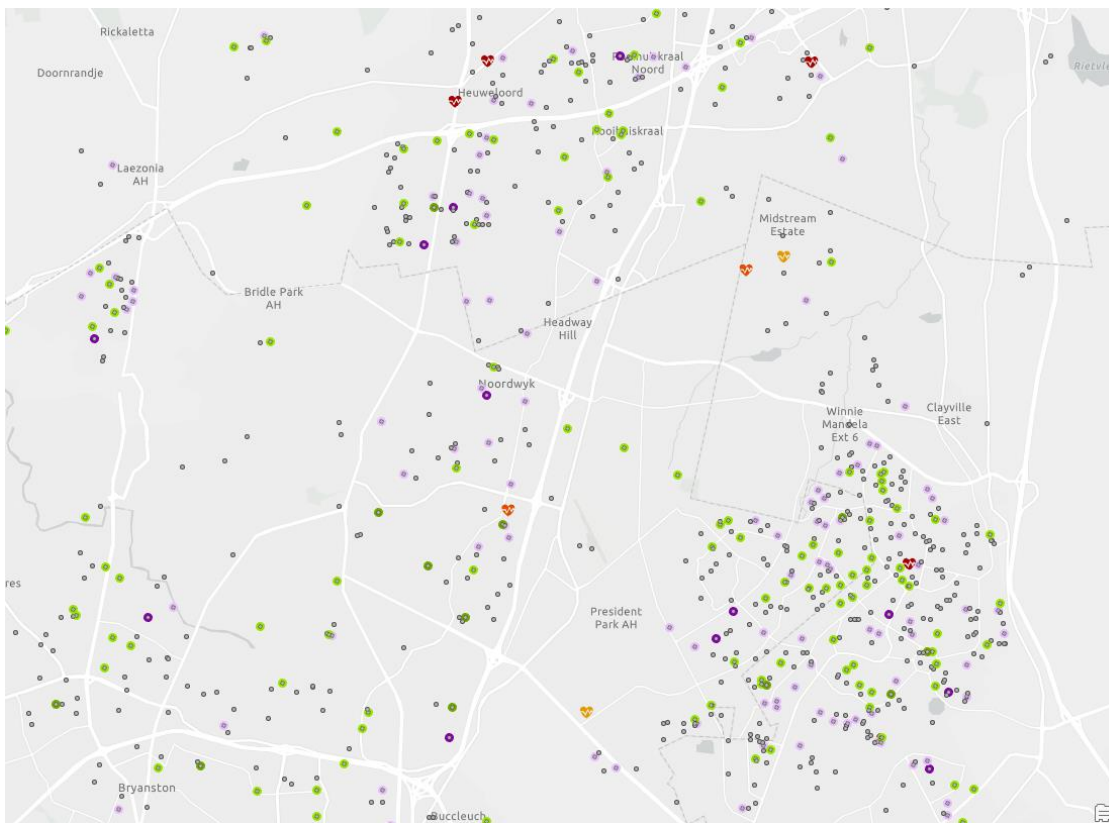
Field	Reverse Direction
age_2023	<input type="checkbox"/>
weight	<input type="checkbox"/>
pain_perc	<input type="checkbox"/>

⊕ Add another

Preset Method to Scale and Combine Variables
 Combine values (Mean of scaled values) ⌵

Method to Scale Input Variables
 Minimum-maximum ⌵

Method to Combine Scaled Variables
 Mean ⌵



14. Turn On < Patient Visit Index < Standard Deviation
15. Open < Counts per Patient | Index – Mean < Select 2 & 3 (Higher Index)
16. Style < Patient | Visit < Graduated Symbols < Cost of Visit < 5
17. Closest vs popular dr:
18. Generate Desire Lines < Dr | Locations < Patient Visit < Dr 2 Locations < Name < Address < Straight Line < Meters < 7,500



Geoprocessing Generate Desire Lines

Parameters Environments

Store Layer: Dr | Locations

Customer Layer: Patient | Visit

Output Feature Class: Dr_2_Locations

Store ID Field: name

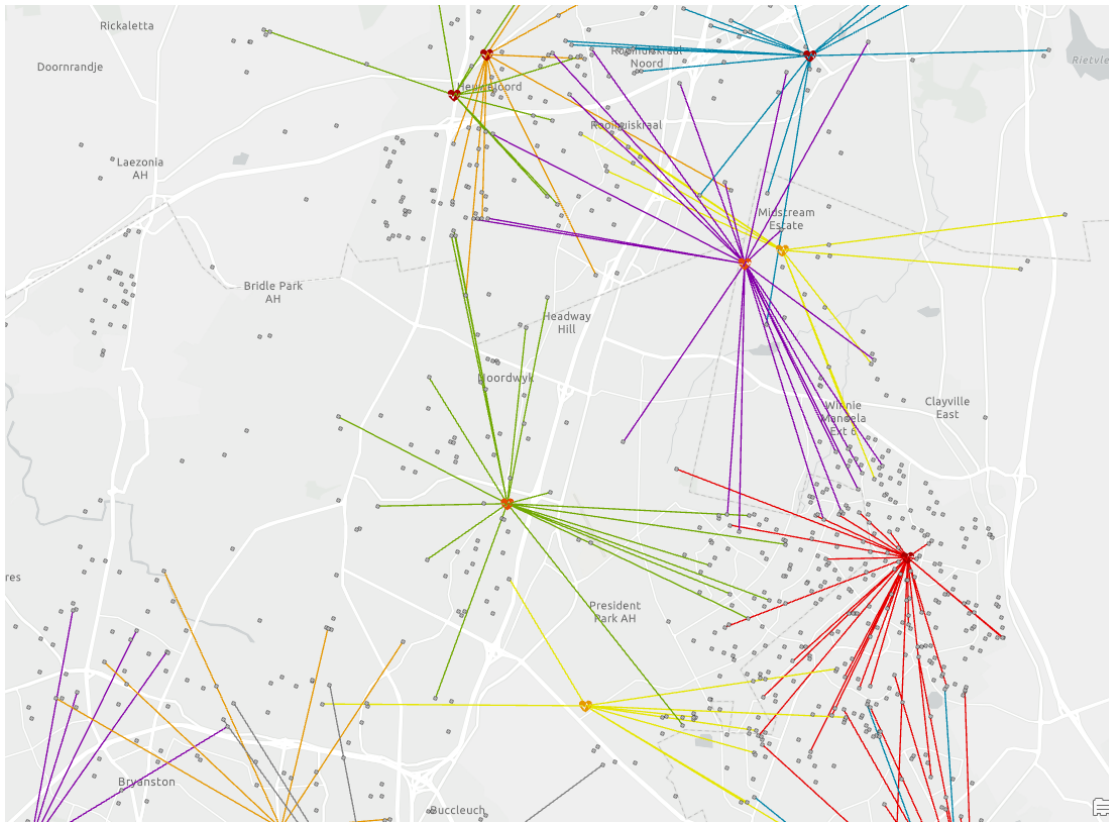
Associated Store ID Field: address

Distance Type: Straight Line

Measure Units: Meters

Cutoff: 7500

Create Report



Is medicine accessible to everyone?

- 19. Generate Trade Area Rings < Medicine | Pick Up < Medicine Locations Trade Area < 2,500 5,000 7,500 < Meters < Mall Name < Overlap



Geoprocessing Generate Trade Area Rings

Parameters Environments

Input Features
Medicine | Pick Up

Output Feature Class
Medicine_Locations_Trade_Area

Input Method
Values

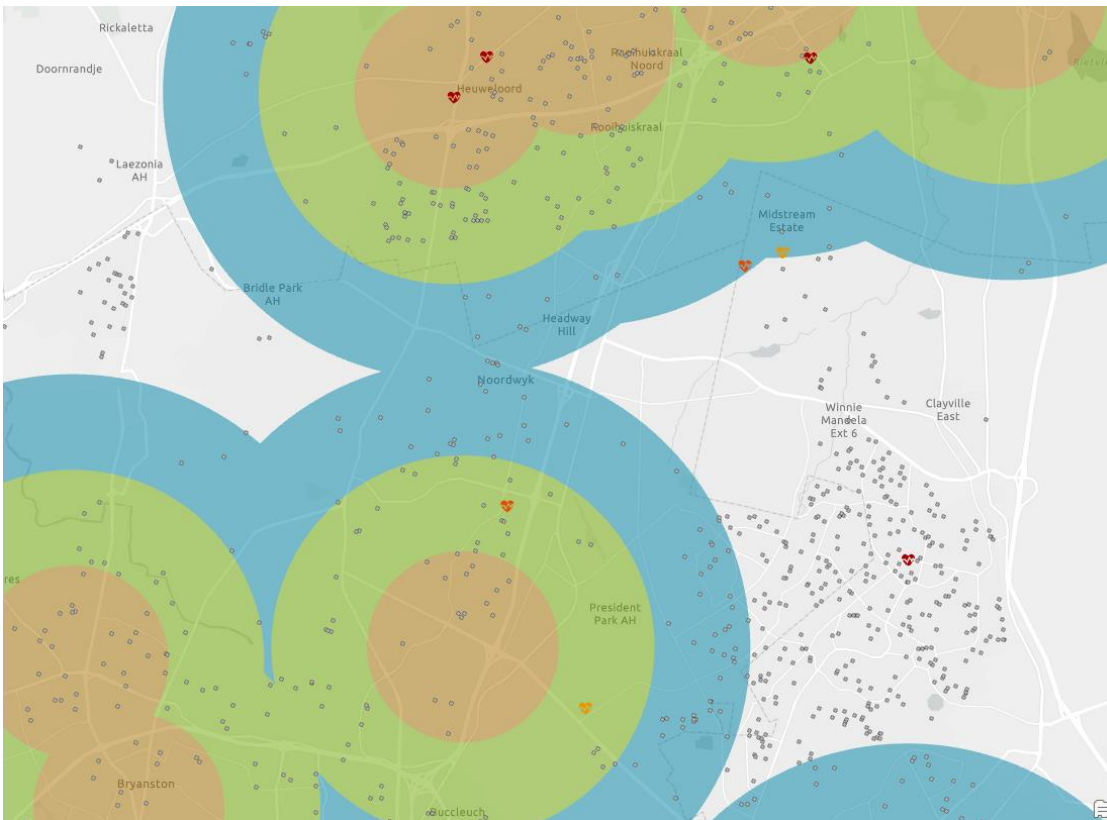
Distances
2500
5000
7500

Distance Units
Meters

ID Field
Mall Name

Remove Overlap

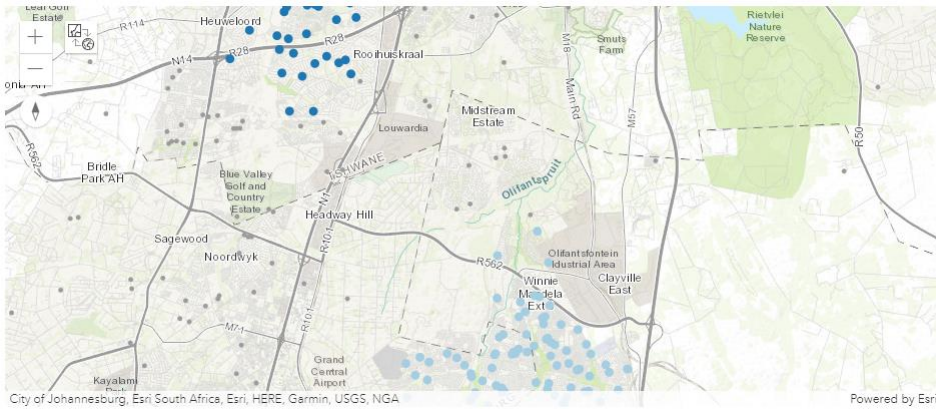
Dissolve Option
Overlap



1. Turn on < Dwelling Frame (bins) < higher density (customer growth)

Identify areas of patients without medical aid?

- 20. Cluster Analysis Notebook < Run
- 21. Note : Update output name (cluster)



GEO AI

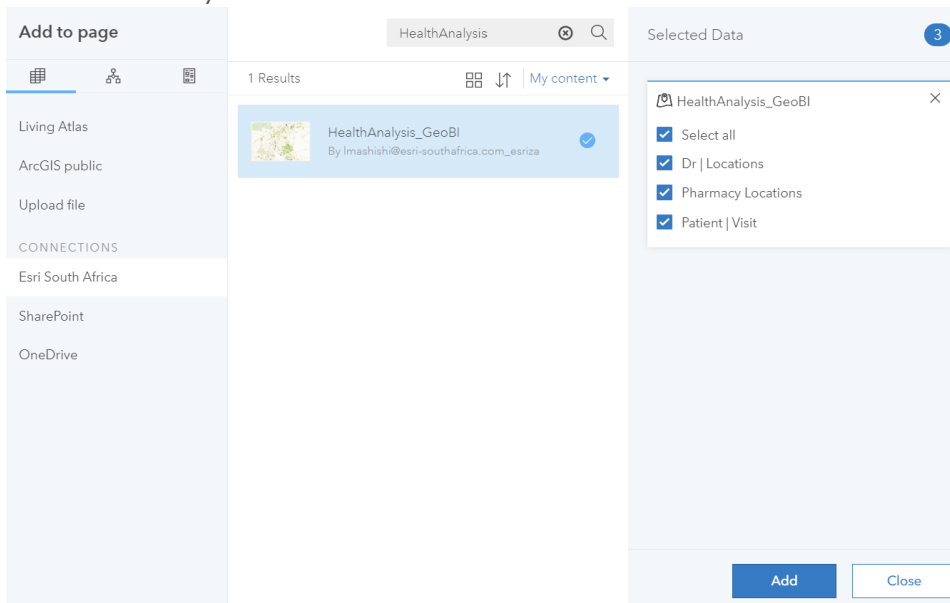
From DOS to spatial (Named Entity Recognition)?

- 22. New Map
- 23. Explorer < Dr Reports < report_x.txt
- 24. Extract Entities Using Deep Learning < Dr Reports < Dr Report < Dr Locator.loc

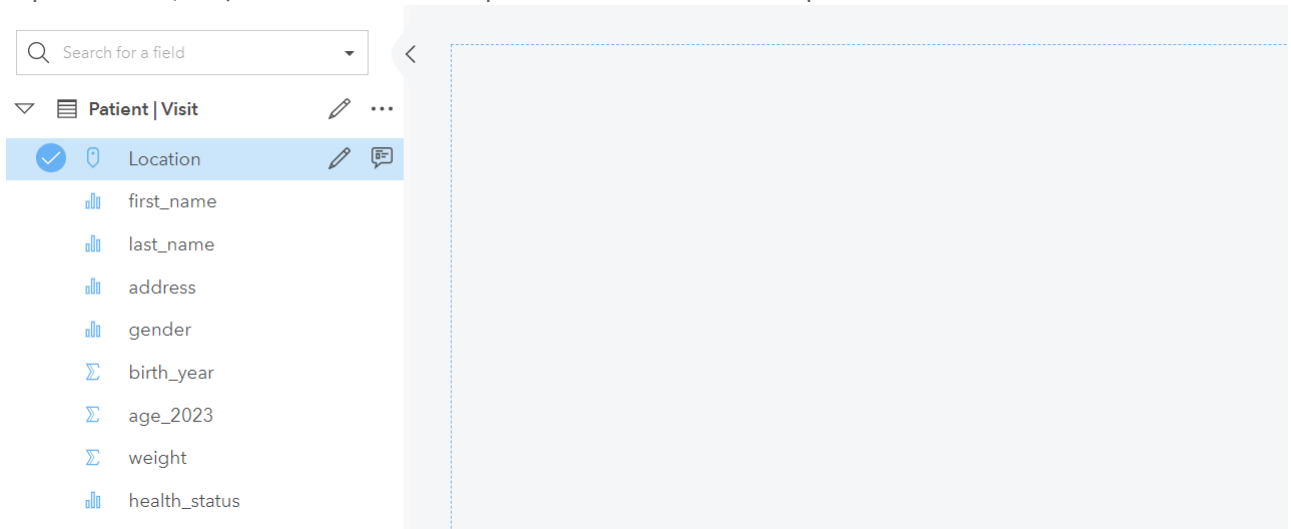
ip	money	geopolitical_area	person	cardinal_number	location	ordinal_number	percent	event
1	R450		Sheldon Cooper				90%	
2	320		Liezel Botha				70%	
3	R250		Tracy Skinner				30%	
4	750		Renchia Olivier				90%	
5	R500		Nico Roets				100%	
6	550		Carley Olivier				20%	
7	250		Cayden Olivier				70%	

GEO BI

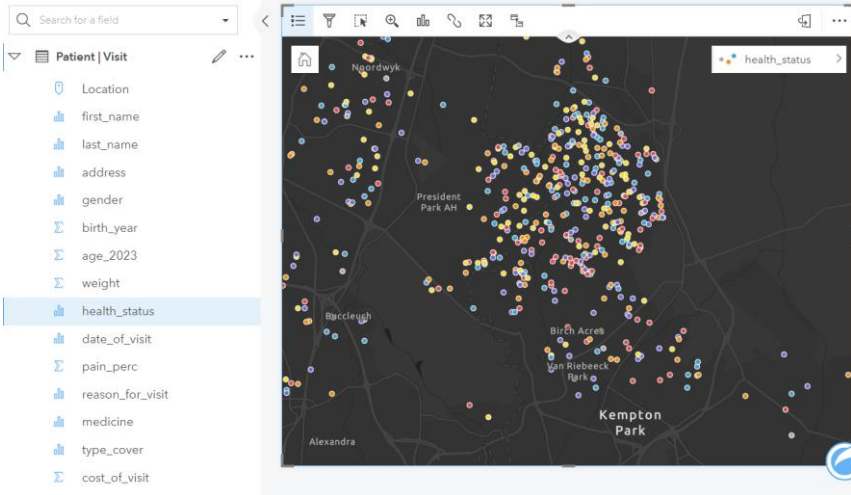
25. On a web browser, open www.arcgis.com
26. Sign in with the ArcGIS credentials assigned to you
27. From the app launch pad, choose insights
28. Click on workbook, then choose New workbook
29. Search HealthAnalysis and select it and click add



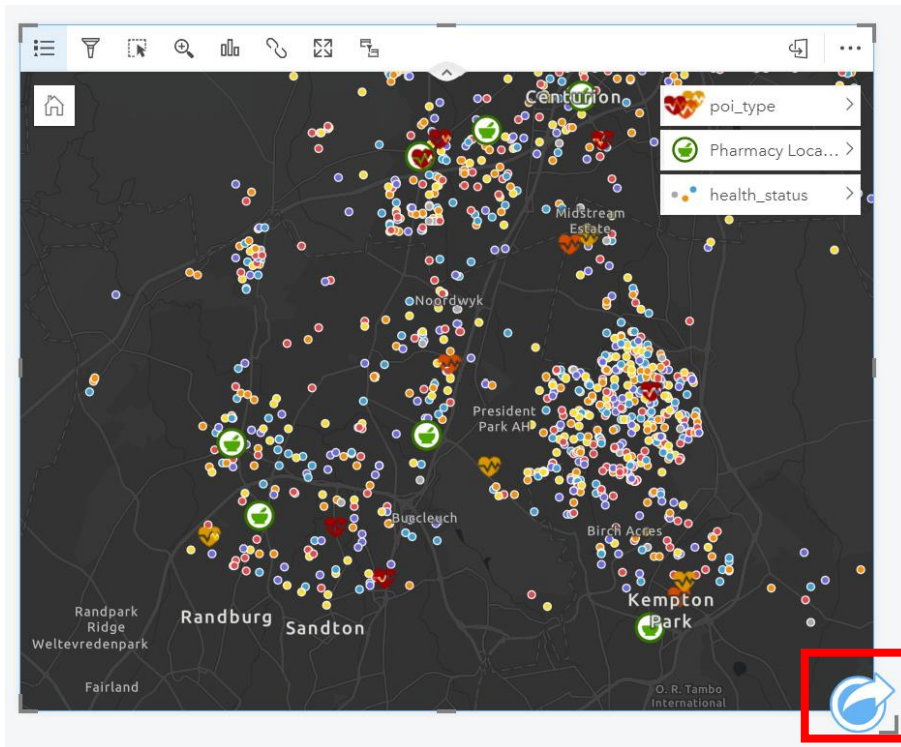
30. Delete the card on the screen
31. Expand Patient | Visit, select Location and drop it on the screen. Choose Map



32. Under Patient | Visit, Select, drag & drop health status to the map



- 33. On your own, add the locations for the other two layers: Pharmacy Locations and Dr Locations onto the map
- 34. Click on the action button at the bottom right of your card



- 35. Choose find answers, then pick how is it distributed
- 36. Choose Calculate Density and then enter the following parameters



← Calculate Density ×

1 Choose a point layer

Patient | Visit

Additional options

Weight

Σ age_2023

Search radius (bandwidth)

Default km

Cell size

Default km²

▶ Run

37. Click Run
38. Once again, click on the action button
39. Click find answers, then choose How is it related?
40. Choose regression and fill in the following properties

← Regression Model ×

1 Choose a layer

Patient | Visit

2 Choose a dependent variable

Σ cost_of_visit

3 Choose explanatory variables

Select explanatory variables

- Select all
- Σ birth_year
- Σ age_2023
- Σ weight
- Σ pain_perc

Select Cancel

41. Click run and then click visualize
42. Once again, click the action button
43. Click on What is nearby and choose

Find Nearest

1 features to search near
Patient | Visit

2 Choose a layer with features you wish to find
Dr | Locations

3 For each location in the input layer

Limit the number of nearest locations
3

Limit the search range
5 km

Run

44. Click Run
45. Once again, click on the action button
46. Click on How has it changed?
47. Choose time series

Time Series

1 Choose a layer
Patient | Visit

2 Choose number fields
cost_of_visit

3 Choose date/time fields
date_x

Run

48. Click Run
49. Click and Explore the analysis view