## **IBM Cloud Gallery**

Estimated Time (45 min)

You will learn to:

IBM Cloud Resource hub is a growing collection of data sets, notebooks, and project templates. In this lab, you will use IBM Cloud Resource hub to explore different datasets. As you learned in the course, data can be more than just numbers. Data can be numeric, text, images, videos, audios and more. You will look at three samples. Sample 1 contains data with only numeric attributes.

Sample 2 contains data with numeric & text attributes.

**Sample 3** cantains a Jupyter Notebook, a tool which data scientists use to create models.

Let's take a look at how data scientists use different datasets. **Objectives:** 

• Examine a Jupyter Notebook **Exercise 1: Examine a numeric dataset** 

• Examine a dataset with non-numeric attributes

• Explore the IBM Cloud Resource hub

Examine a numeric dataset

1. Click on the link: https://dataplatform.cloud.ibm.com/gallery

2. Click the filter button in the top right of the window:

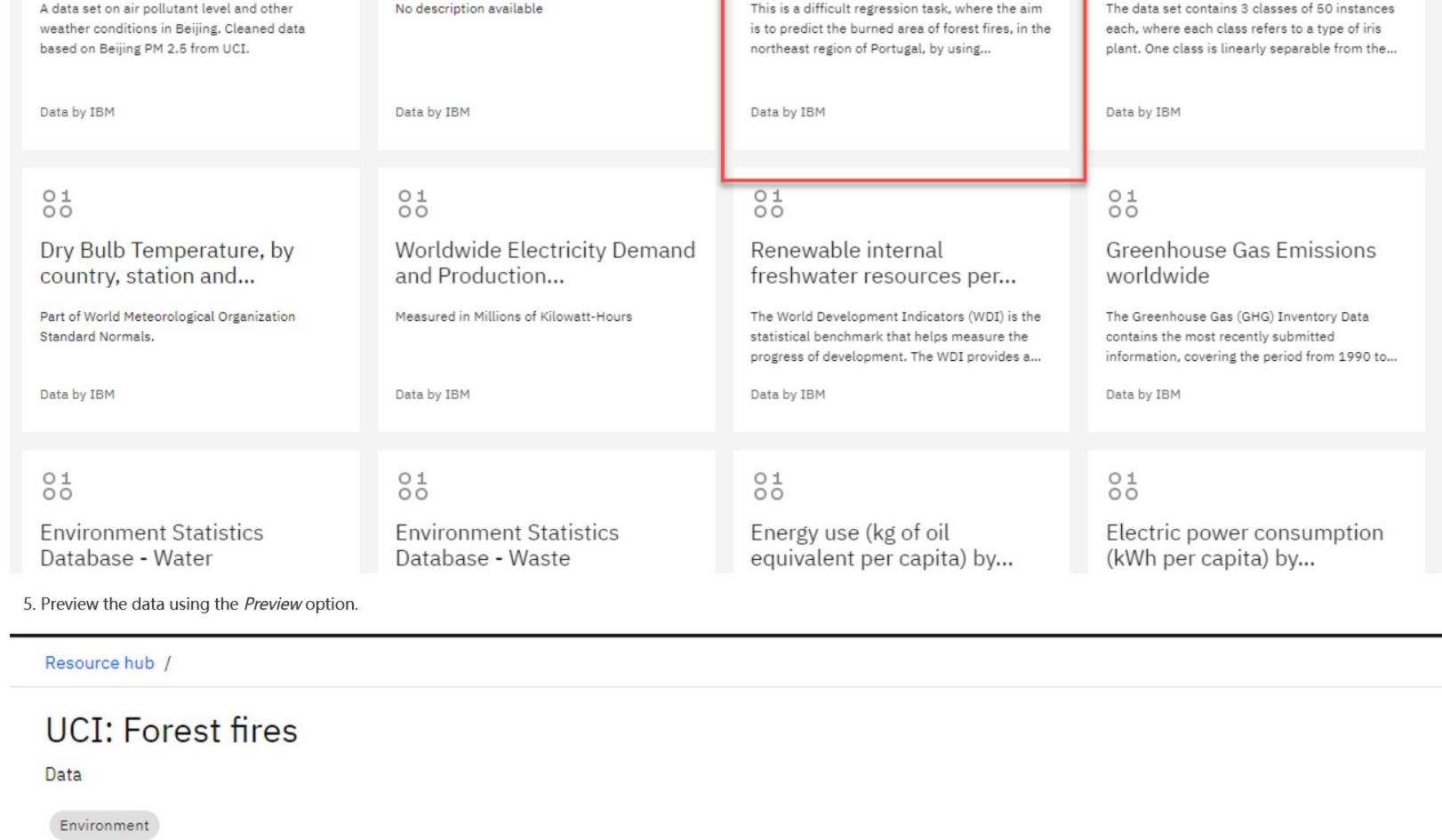
Resource hub ® Q Search samples 3. In the dropdown menu that appears, select the *Data* checkbox under *Sample type*. Then click on the *Tags* dropdown, and select the *Environment* checkbox.

× Filter Sample type  $\wedge$ 



Beijing PM 2.5 A data set on air pollutant level and other 01

Scoring for Beijing PM 2.5



01

UCI: Forest fires

01

UCI: Iris

Details

Publisher

Modified

Source

Sep 25, 2017

Terms of use

https://archive.ics.uci.edu/ml/datasets/forest+fires

IBM

X month

Description

## fri mar oct tue

oct

mar

Preview

day

sat

fri

FFMC

86.2

90.6

90.6

91.7

DMC

26.2

35.4

43.7

33.3

DC

94.3

669.1

686.9

77.5

ISI

5.1

6.7

6.7

9

temp

8.2

18

14.6

8.3

RH

51

33

33

97

wind

6.7

0.9

1.3

4

rain

0

0

0

0.2

area

0

0

0

0

6

8

https://archive.ics.uci.edu/ml/datasets/forest+fires 8 6 89.3 51.3 102.2 9.6 11.4 99 1.8 0 0 mar sun 22.2 0 85.3 488 14.7 29 5.4 0 8 aug 92.3 sun **Explore the data** The data is related to forest fires where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using meterological and other data. Attribute Information: 1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9 2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9 3. month - month of the year: 'jan' to 'dec' 4. day - day of the week: 'mon' to 'sun' 5. FFMC - FFMC index from the FWI system: 18.7 to 96.20 6. DMC - DMC index from the FWI system: 1.1 to 291.3 7. DC - DC index from the FWI system: 7.9 to 860.6 8. ISI - ISI index from the FWI system: 0.0 to 56.10 9. temp - temperature in Celsius degrees: 2.2 to 33.30 10. RH - relative humidity in %: 15.0 to 100

## The data doesn't have to be only based on numbers. Data can be text, images and other types as well. Let's look at a dataset which has text values. 1. At the top of the page, select the *Resource hub* option.

11. wind - wind speed in km/h: 0.40 to 9.40

12. rain - outside rain in mm/m2: 0.0 to 6.4

13. area - the burned area of the forest (in ha): 0.00 to 1090.84

**Exercise 2: Evaluate a non-numeric dataset** 

**UCI:** Forest fires

sense to model with the logarithm transform).

**IBM Cloud Pak for Data** 

Resource hub /

Data

2. Type Airbnb into the search bar.

(this output variable is very skewed towards 0.0, thus it may make

Environment

Airbnb listings for Washington, D.C., District of

from Inside Airbnb which aggregates and...

Data by IBM

01

Data by IBM

Columbia, United States. This dataset is sourced

aggregates and cleanses publicly available data...

01

Data by IBM

Data by IBM

01

Airbnb Data for Analytics:

Airbnb reviews for Vienna, Vienna, Austria. This

aggregates and cleanses publicly available data...

dataset is sourced from Inside Airbnb which

Airbnb Data for Analytics:

Airbnb listings for Trentino, Trentino-Alto

Adige/Südtirol, Italy. This dataset is sourced from

Inside Airbnb which aggregates and cleanses...

Trentino Listings

Vienna Reviews

01

Data by IBM

Data by IBM

01

Airbnb Data for Analytics:

Airbnb listings for Vienna, Vienna, Austria. This

aggregates and cleanses publicly available data...

dataset is sourced from Inside Airbnb which

Airbnb Data for Analytics:

Airbnb calendar for Venice, Veneto, Italy. This

aggregates and cleanses publicly available data...

dataset is sourced from Inside Airbnb which

Venice Calendar

Vienna Listings

01

Data by IBM

Data by IBM

01

Airbnb Data for Analytics:

Airbnb calendar for Vienna, Vienna, Austria, This

aggregates and cleanses publicly available data...

dataset is sourced from Inside Airbnb which

Airbnb Data for Analytics:

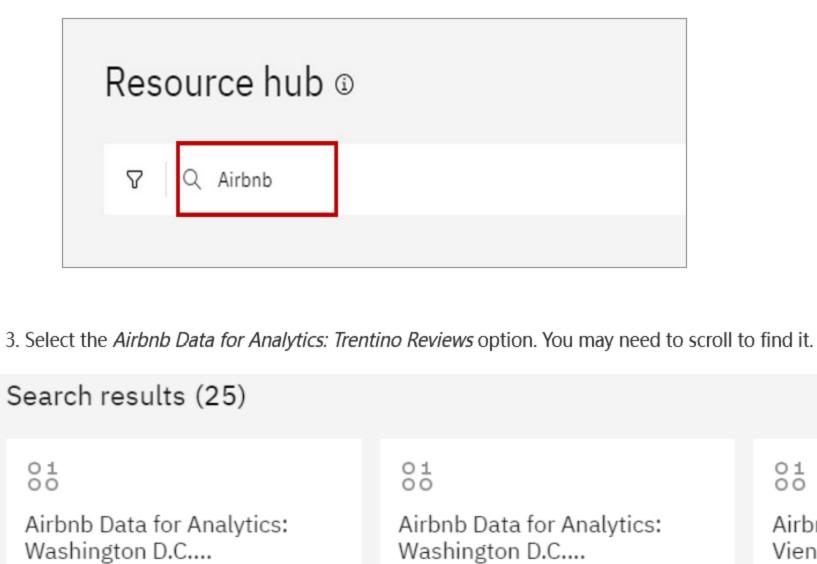
Airbnb reviews for Vancouver, British Columbia,

Airbnb which aggregates and cleanses publicly...

Canada. This dataset is sourced from Inside

Vancouver Reviews

Vienna Calendar



Airbnb Data for Analytics: Airbnb Data for Analytics: Venice Listings Venice Reviews Airbnb reviews for Venice, Veneto, Italy. This Airbnb listings for Venice, Veneto, Italy. This dataset is sourced from Inside Airbnb which dataset is sourced from Inside Airbnb which

Airbnb reviews for Washington, D.C., District of

from Inside Airbnb which aggregates and...

Data by IBM

Data by IBM

01

Columbia, United States. This dataset is sourced

aggregates and cleanses publicly available data...

Airbnb Data for Analytics: Ancouver Listings  bnb listings for Vancouver, British Columbia, nada. This dataset is sourced from Inside bnb which aggregates and cleanses publicly  ta by IBM  Airbnb Data for Analytics: Vancouver Calendar  Airbnb calendar for Vancouver, British Columbia, Canada. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly  Data by IBM  Peview the data using the Preview option.				Airbnb Data for Analytics: Trentino Reviews  Airbnb reviews for Trentino, Trentino-Alto Adige/Südtirol, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses  Data by IBM		Airbnb Data for Analytics: Trentino Calendar  Airbnb calendar for Trentino, Trentino-Alto Adige/Südtirol, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses  Data by IBM		Airbnb Data for Analytics: Toronto Reviews  Airbnb reviews for Toronto, Ontario, Canada. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly available data  Data by IBM		
Airbnb Data for Analytics: Tr			Trentino Reviews		Economy & Business		Dec 20, 2016		Add to project	
Description	on I	Preview								
listing_id	id	date	reviewer_id	reviewer_name	comments	listing_name	e host_id	listing_latitude	listing_longitude	host_name
listing_id	id	date	reviewer_id	reviewer_name	comments	listing_name	host_id	listing_latitude	listing_longitude	host_name
5064970 29436648		2015- 04-07	11582326	Stephan	Marina is very kind and friendly. We enjoyed her apartment, that was very modern and clean with two rooms, a bathroom and the kitchen inside the livingroom with a balcony that goes to the north. All in all a good flat to stay. Thanks!	apartment + Wi-FI + parking!	2845951	45.88512254895795	5 10.859054481189382 Marina	
					Marinas flat was a dream! Spotlessly clean, very cute decorated and the balcony was the biggest plus! Marina welcomed us in					

her flat and gave us many

restaurants. You have to ask

her for the best Gelateria in Riva. The best ice cream I 've ever eaten! We will

definitly come back! Thank you Marina for the awesome time we could spend in your

flat. Annika & Joachim

apartment +

2845951

45.88512254895795

skills

Notebook by IBM

House Building with worker

This tutorial includes everything you need to set

constraint programming models to help solve th...

up decision optimization engines and build

10.859054481189382

Marina

Wi-FI +

parking!

tips for hiking,

mountainbiking and

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb guests may leave a review after their stay, and these can be used as an indicator of airbnb activity. The minimum stay, price and number of reviews have been used to estimate the occupancy rate, the number of nights per year and the income per month for each listing. You could use this data in multitude of ways - to analyze the star ratings of places, to analyze the location preferences of the customers, to analyze the tone and sentiment of customer reviews and many more. Airbnb uses location data to improve guest satisfaction.

The dataset comprises of three main tables:

What else might you use this data for?

**Exercise 3: Evaluate Jupyter Notebook** 

5064970

**Explore the data** 

new stores using...

Data

✓ Notebook

Project

Tags

Governance Content

33481368

20223641

is\_superhost (categorical), neighbourhood (categorical), ratings (continuous) among others.

Annika

Resource hub ®  $\nabla$ Finding optimal locations × × Sample type: Notebook X Reset filters Filter Sample type

Finding optimal locations of

Optimization can help to prescribe decisions for a

complex constrained problem using CPLEX...

Terms of Use

new stores using...

Notebook by IBM

Privacy

This notebook shows you how Decision

Return to the Resource hub. Select Notebook from the Sample type menu that appears after clicking on the filter button. In the search bar type Finding optimal locations Select the card that says Finding optimal locations of

• listings - Detailed listings data showing 96 attributes for each of the listings. Some of the attributes used in the analysis are price(continuous), longitude (continuous), latitude (continuous), listing\_type (categorical),

• reviews - Detailed reviews given by the guests with 6 attributes. Key attributes include date (datetime), listing\_id (discrete), reviewer\_id (discrete) and comment (textual).

Search results (3)

spatial data

Notebook by IBM

This Jupyter notebook uses *Decision Optimization* with Python to help determine the optimal location of a new store.

Use spatial indexing to query

Learn how to accelerate time-critical searches by

using spatial indexing to query spatial data. This

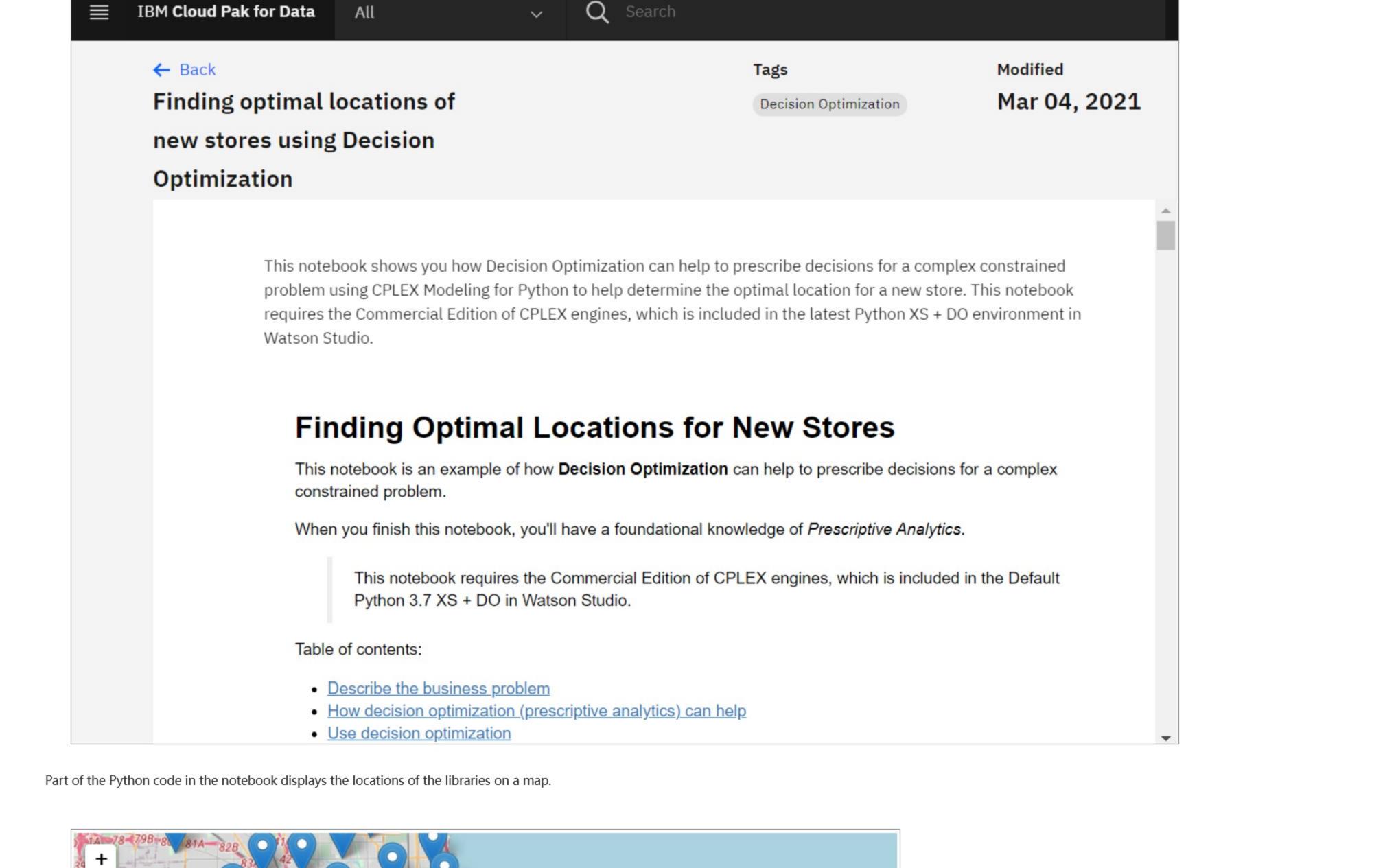
will help you find locations points within a certai...

Blog

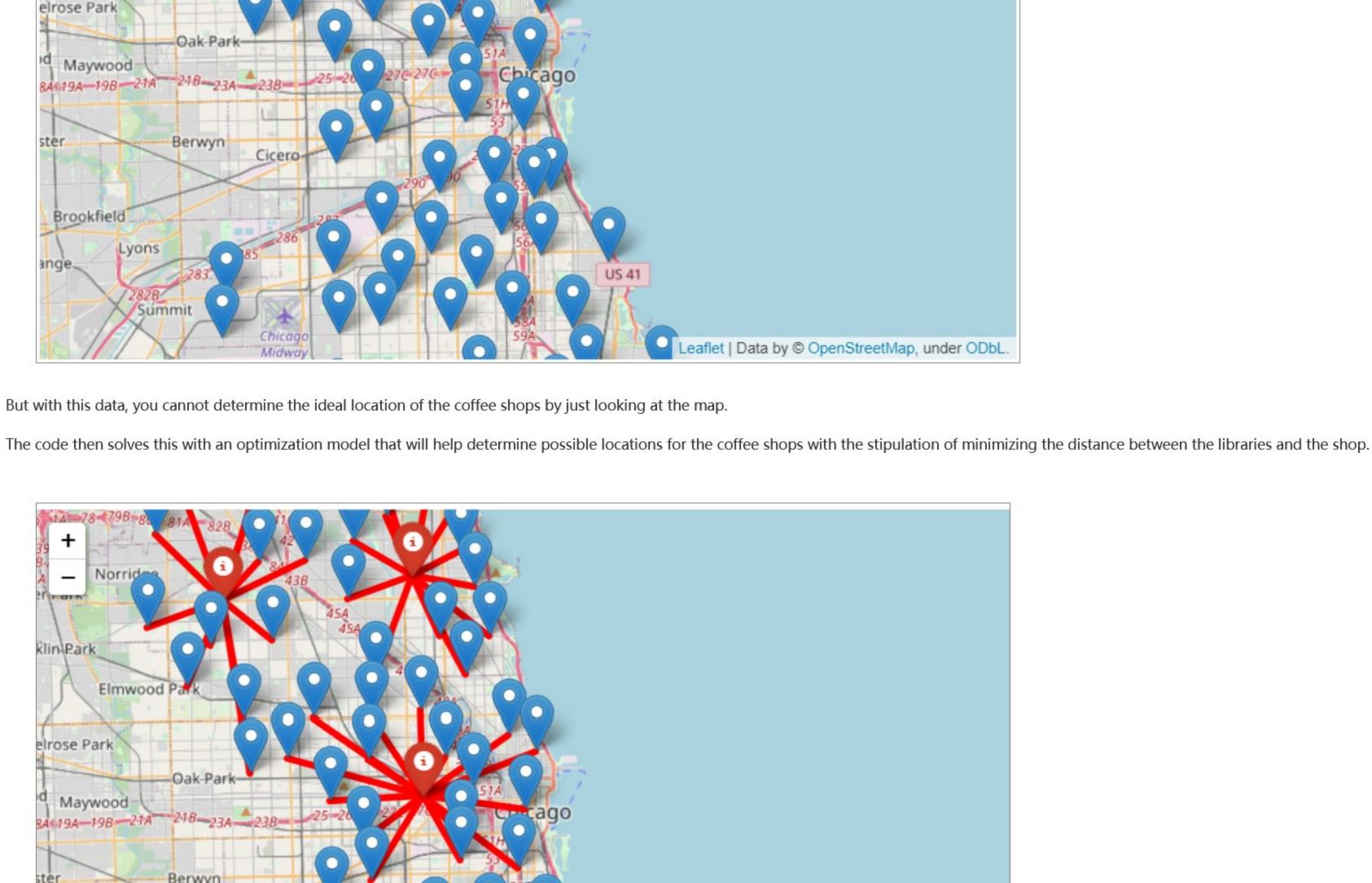
This Notebook aims to identify where to place a coffee shop that minimizes the total distance from libraries in the area to the shop so that a book reader can get to the shop easily.

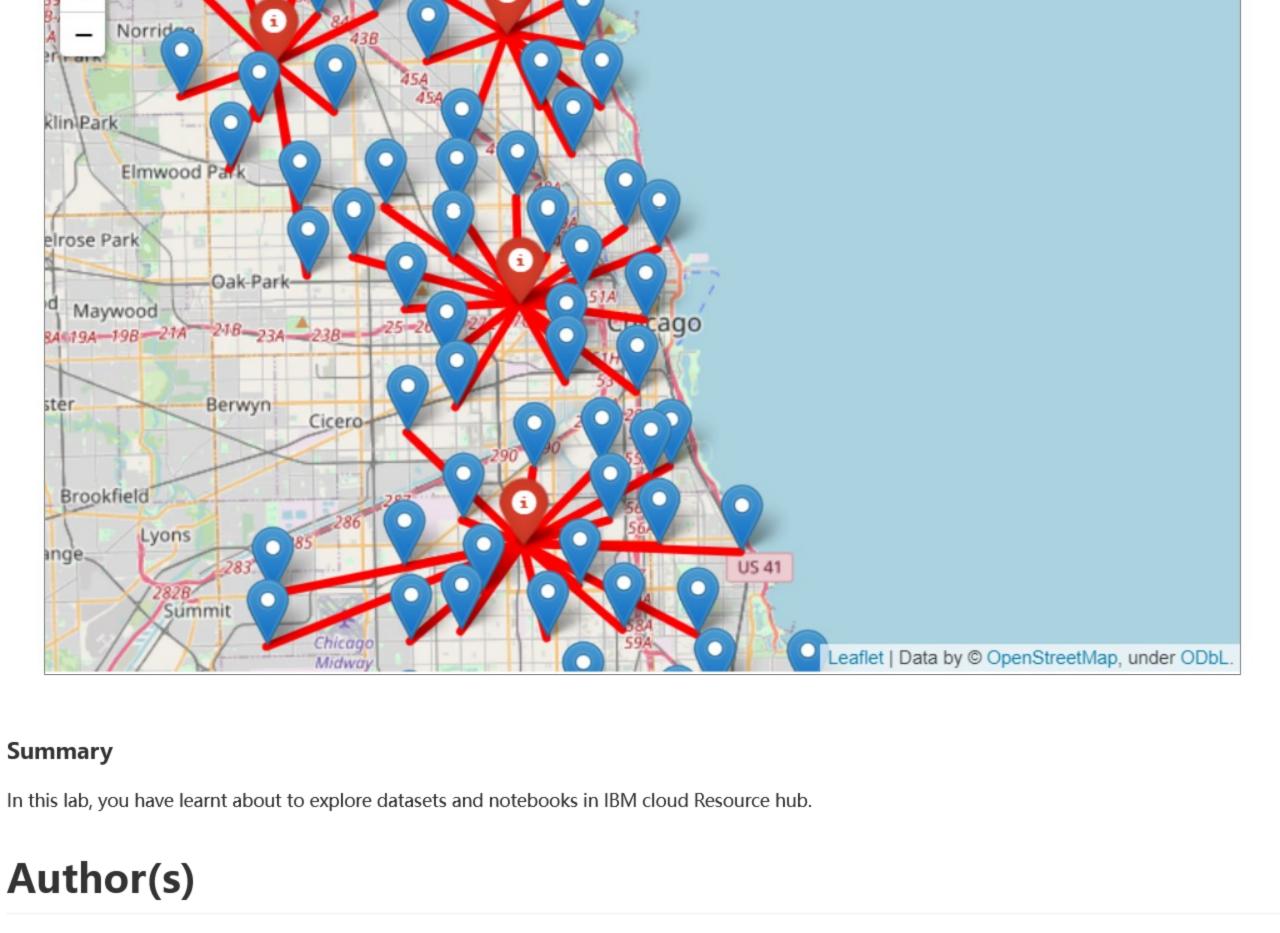
Contact

• calendar - Provides details about booking for the next year by listing. Four attributes in total including listing\_id (discrete), date(datetime), available (categorical) and price (continuous).



klin Park Elmwood Park





## **Malika Singla** Other Contributor(s)