

# IBM Cloud Gallery

Estimated Time (45 min)

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** contains a Jupyter Notebook, a tool which data scientists use to create models.

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

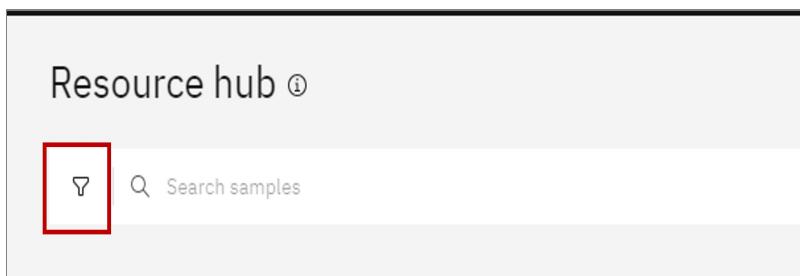
## Objectives :

You will learn to:

- Explore the IBM Cloud Resource hub
- Examine a numeric dataset
- Examine a dataset with non-numeric attributes
- Examine a Jupyter Notebook

## Exercise 1: 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:



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 X

Sample type ^

Data  

Governance Content

Notebook

Project

Tags ^

Communications

Cross Industry

Data fabric

Decision Optimization

Economy & Business

Energy & Utilities

Environment  

Financial Markets

Financial Services

Geography

Guided tutorial

Health

Industry Accelerator

Insurance

Knowledge Accelerators

Law & Government

4. In the search results, click on *UCI: Forest Fires*.

## Search results (35)

 Beijing PM 2.5  A data set on air pollutant level and other weather conditions in Beijing. Cleaned data based on Beijing PM 2.5 from UCI.  <a href="#">Data by IBM</a>	 Scoring for Beijing PM 2.5  No description available  <a href="#">Data by IBM</a>	 UCI: Forest fires  This is a difficult regression task, where the aim is to predict the burned area of forest fires, in the northeast region of Portugal, by using...  <a href="#">Data by IBM</a>	 UCI: Iris  The data set contains 3 classes each, where each class refers to a plant. One class is linearly separable.  <a href="#">Data by IBM</a>
 Dry Bulb Temperature, by country, station and...  Part of World Meteorological Organization Standard Normals.  <a href="#">Data by IBM</a>	 Worldwide Electricity Demand and Production...  Measured in Millions of Kilowatt-Hours  <a href="#">Data by IBM</a>	 Renewable internal freshwater resources per...  The World Development Indicators (WDI) is the statistical benchmark that helps measure the progress of development. The WDI provides a...  <a href="#">Data by IBM</a>	 Greenhouse Gas Emissions worldwide  The Greenhouse Gas (GHG) Inventory contains the most recently submitted information, covering the period 1990-2018.  <a href="#">Data by IBM</a>
 Environment Statistics Database - Water  <a href="#">Data by IBM</a>	 Environment Statistics Database - Waste  <a href="#">Data by IBM</a>	 Energy use (kg of oil equivalent per capita) by...  <a href="#">Data by IBM</a>	 Electric power consumption (kWh per capita) by...  <a href="#">Data by IBM</a>

5. Preview the data using the *Preview* option.

[Resource hub](#) /

# UCI: Forest fires

Data

Environment

Description

Preview

X

Y

month

day

FFMC

D

7

5

mar

fri

86.2

2

7

4

oct

tue

90.6

3

7

4

oct

sat

90.6

4

8

6

mar

fri

91.7

3

8

6

mar

sun

89.3

5

8

6

aug

sun

92.3

8

**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 meteorological 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
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m<sup>2</sup> : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84  
(this output variable is very skewed towards 0.0, thus it may make sense to model with the logarithm transform).

## Exercise 2: Evaluate a non-numeric dataset

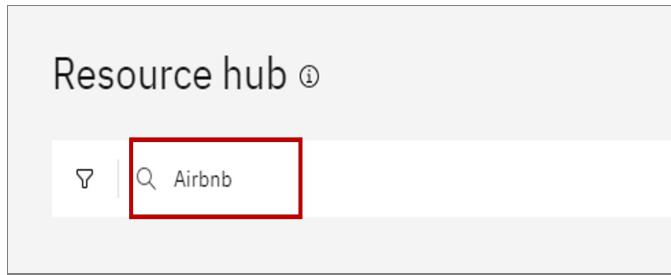
The data doesn't have to be only based on numbers. Data can be text, images and other types as well. Let's look at a dataset which has text values.

1. At the top of the page, select the *Resource hub* option.



The screenshot shows the IBM Cloud Pak for Data interface. At the top, there is a dark header bar with the text "IBM Cloud Pak for Data". Below this, a navigation bar has a blue link labeled "Resource hub /" which is highlighted with a red rectangular box. The main content area displays the title "UCI: Forest fires" in large blue text. Underneath the title, there are two buttons: "Data" in orange and "Environment" in grey. The entire interface is set against a white background.

2. Type *Airbnb* into the search bar.



3. Select the *Airbnb Data for Analytics: Trentino Reviews* option. You may need to scroll to find it.

### Search results (25)

<p> Airbnb Data for Analytics: Washington D.C....</p> <p>Airbnb reviews for Washington, D.C., District of Columbia, United States. This dataset is sourced from Inside Airbnb which aggregates and...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Washington D.C....</p> <p>Airbnb listings for Washington, D.C., District of Columbia, United States. This dataset is sourced from Inside Airbnb which aggregates and...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Vienna Reviews</p> <p>Airbnb reviews for Vienna, Vienna, Austria. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly available data...</p> <p>Data by IBM</p>	<p> Airbnb Data for Vienna Listings</p> <p>Airbnb listings for Vienna, dataset is sourced from Inside Airbnb which aggregates and cleanses...</p> <p>Data by IBM</p>
<p> Airbnb Data for Analytics: Venice Reviews</p> <p>Airbnb reviews for Venice, Veneto, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly available data...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Venice Listings</p> <p>Airbnb listings for Venice, Veneto, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly available data...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Trentino Listings</p> <p>Airbnb listings for Trentino, Trentino-Alto Adige/Südtirol, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses...</p> <p>Data by IBM</p>	<p> Airbnb Data for Venice Calendar</p> <p>Airbnb calendar for Venice, dataset is sourced from Inside Airbnb which aggregates and cleanses...</p> <p>Data by IBM</p>
<p> Airbnb Data for Analytics: Vancouver Listings</p> <p>Airbnb listings for Vancouver, British Columbia, Canada. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Vancouver Calendar</p> <p>Airbnb calendar for Vancouver, British Columbia, Canada. This dataset is sourced from Inside Airbnb which aggregates and cleanses publicly...</p> <p>Data by IBM</p>	<p> Airbnb Data for Analytics: Trentino Reviews</p> <p>Airbnb reviews for Trentino, Trentino-Alto Adige/Südtirol, Italy. This dataset is sourced from Inside Airbnb which aggregates and cleanses...</p> <p>Data by IBM</p>	<p> Airbnb Data for Trentino Calendar</p> <p>Airbnb calendar for Trentino, dataset is sourced from Inside Airbnb which aggregates and...</p> <p>Data by IBM</p>

4. Preview the data using the *Preview* option.

## Airbnb Data for Analytics: Trentino Reviews

Economy & Business

Description		Preview					
listing_id	id	date	reviewer_id	reviewer_name	comments	listing_name	host_
listing_id	id	date	reviewer_id	reviewer_name	comments	listing_name	host_
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 living-room with a balcony that goes to the north. All in all a good flat to stay. Thanks!	apartment + Wi-FI + parking!	2845'
5064970	33481368	2015-05-28	20223641	Annika	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 tips for hiking, mountainbiking and restaurants. You have to ask her for the best Gelateria in Riva. The best ice cream I 've ever eaten! We will definitely come back! Thank you Marina for the awesome time we could spend in your flat. Annika & Joachim	apartment + Wi-FI + parking!	2845'

### Explore the data

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.

💡 What else might you use this data for?

The dataset comprises of three main tables:

- 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), is\_superhost (categorical), neighbourhood (categorical), ratings (continuous) among others.
- reviews - Detailed reviews given by the guests with 6 attributes. Key attributes include date (datetime), listing\_id (discrete), reviewer\_id (discrete) and comment (textual).
- 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).

### Exercise 3: Evaluate Jupyter Notebook

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 new stores using...*

The screenshot shows the Microsoft Power BI Resource hub interface. At the top, the title "Resource hub" is displayed with an information icon. Below it, there is a search bar containing the text "Finding optimal locations", which is highlighted with a red border. To the left of the search bar is a filter icon (a funnel symbol) and a magnifying glass icon. Below the search bar, the word "Filter" is visible, followed by a close button (an 'X'). Further down, under the heading "Sample type", there is a list of items. The first item is "Data", preceded by a checkbox icon, and the second item is partially visible as "...".

## Governance Content



Notebook



Project

---

## Tags

---



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

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.

IBM Cloud Pak for Data All Search

[Back](#)

**Finding optimal locations of new stores using Decision Optimization**

Tags: Decision Optimization

This notebook shows you how Decision Optimization can help to prescribe decisions for a complex constrained problem using CPLEX Modeling for Python to help determine the optimal location for a new store. This notebook requires the Commercial Edition of CPLEX engines, which is included in the latest Python XS + DO environment in Watson Studio.

## Finding Optimal Locations for New Stores

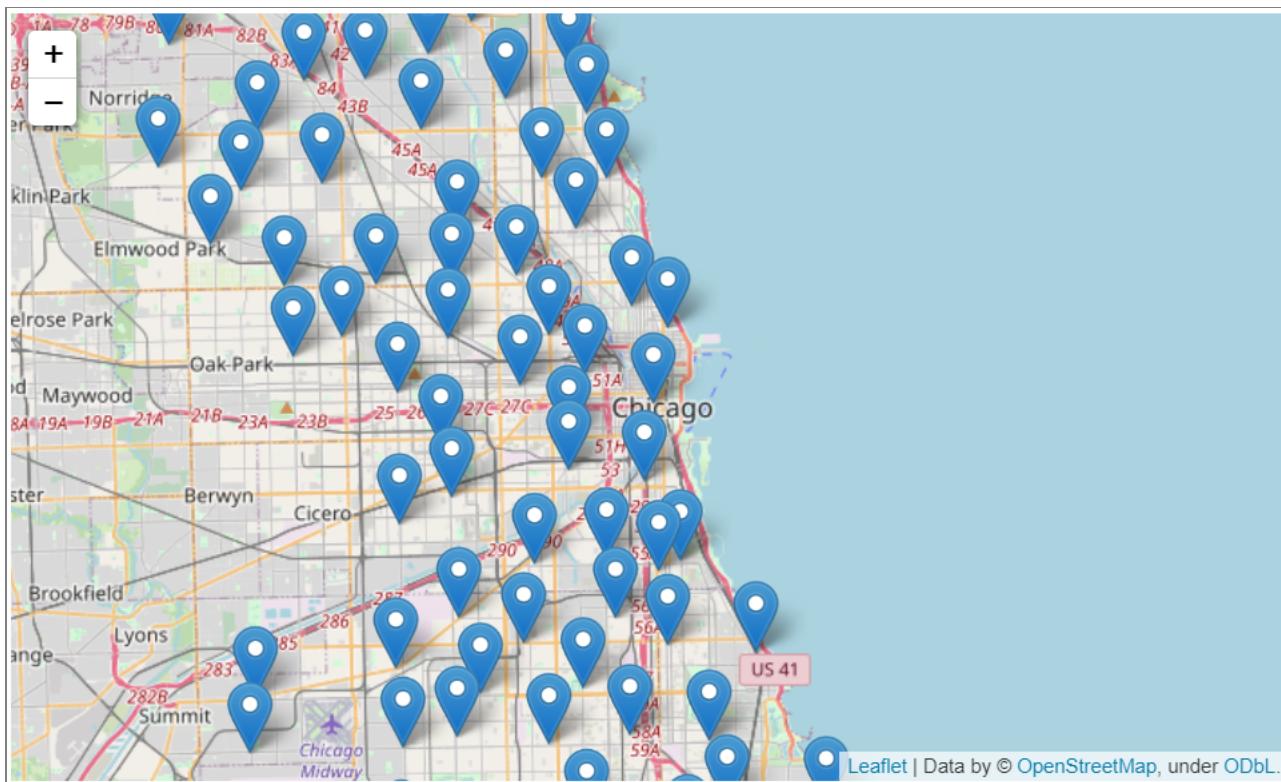
This notebook is an example of how **Decision Optimization** can help to prescribe decisions for a constrained problem. When you finish this notebook, you'll have a foundational knowledge of *Prescriptive Analytics*.

This notebook requires the Commercial Edition of CPLEX engines, which is included in the latest Python 3.7 XS + DO environment in Watson Studio.

Table of contents:

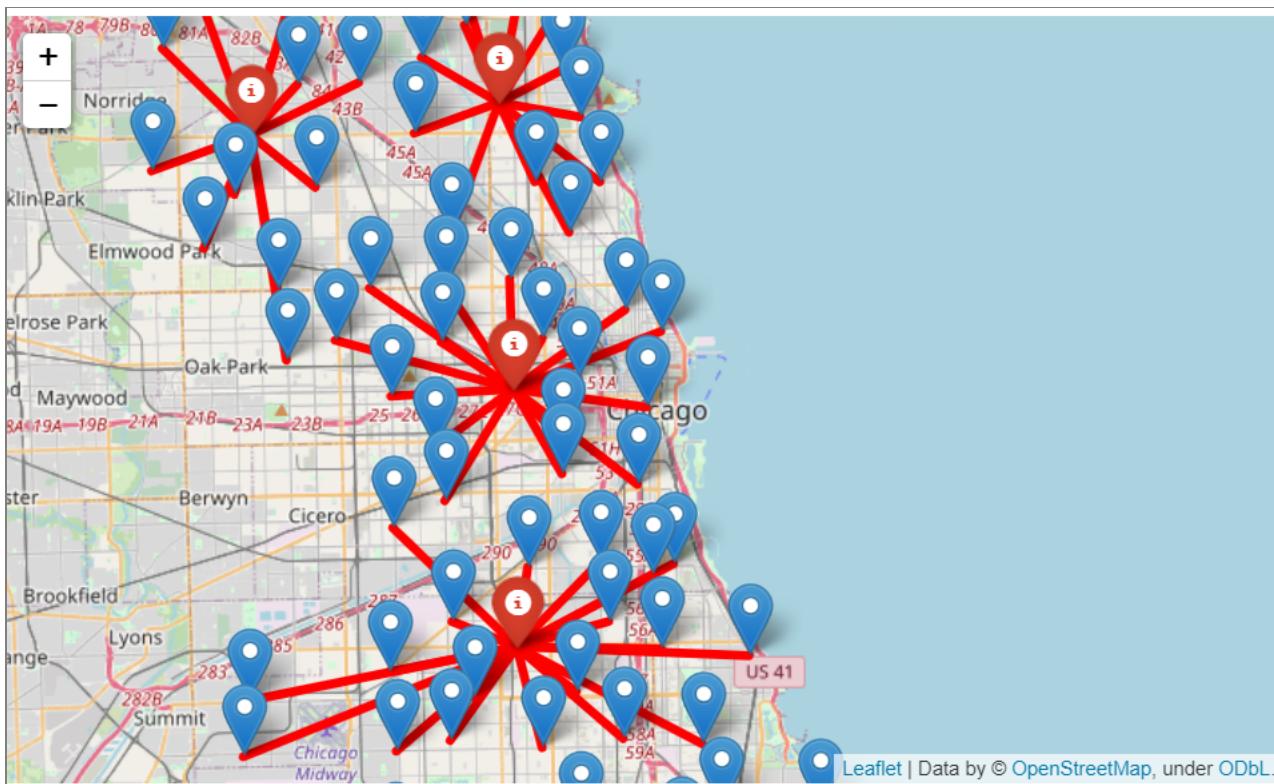
- [Describe the business problem](#)
- [How decision optimization \(prescriptive analytics\) can help](#)
- [Use decision optimization](#)

Part of the Python code in the notebook displays the locations of the libraries on a map.



But with this data, you cannot determine the ideal location of the coffee shops by just looking at the map.

The code then solves this with an optimization model that will help determine possible locations for the coffee shops with the stipulation of minimizing the distance between the libraries and the shop.



## Summary

In this lab, you have learnt about to explore datasets and notebooks in IBM cloud Resource hub.

## Author(s)

[Malika Singla](#)

## Other Contributor(s)

[Lavanya](#)



**Skills Network**