



IBM Cloud Watson Studio

Estimated Time (45 min)

IBM Watson Studio is a service from IBM, that provides a suite of tools and a collaborative environment for data scientists, developers and domain experts. In this lab, you will use Watson Studio and explore different datasets. As we have learnt in the course, the data is not only about numbers, it can be anything such as numeric data, text data, images, videos, audios etc. You will work on three samples.

Sample 1 in which you will learn about the dataset in which only numeric attributes are present.

Sample 2 in which you will learn about the dataset in which numeric & text attributes are present.

Sample 3 in which you will analyze how the Jupyter Notebooks look like. How a Data Scientist create the models?

Let's take a look that how different datasets are used by Data Scientist.

Objectives :

You will learn to:

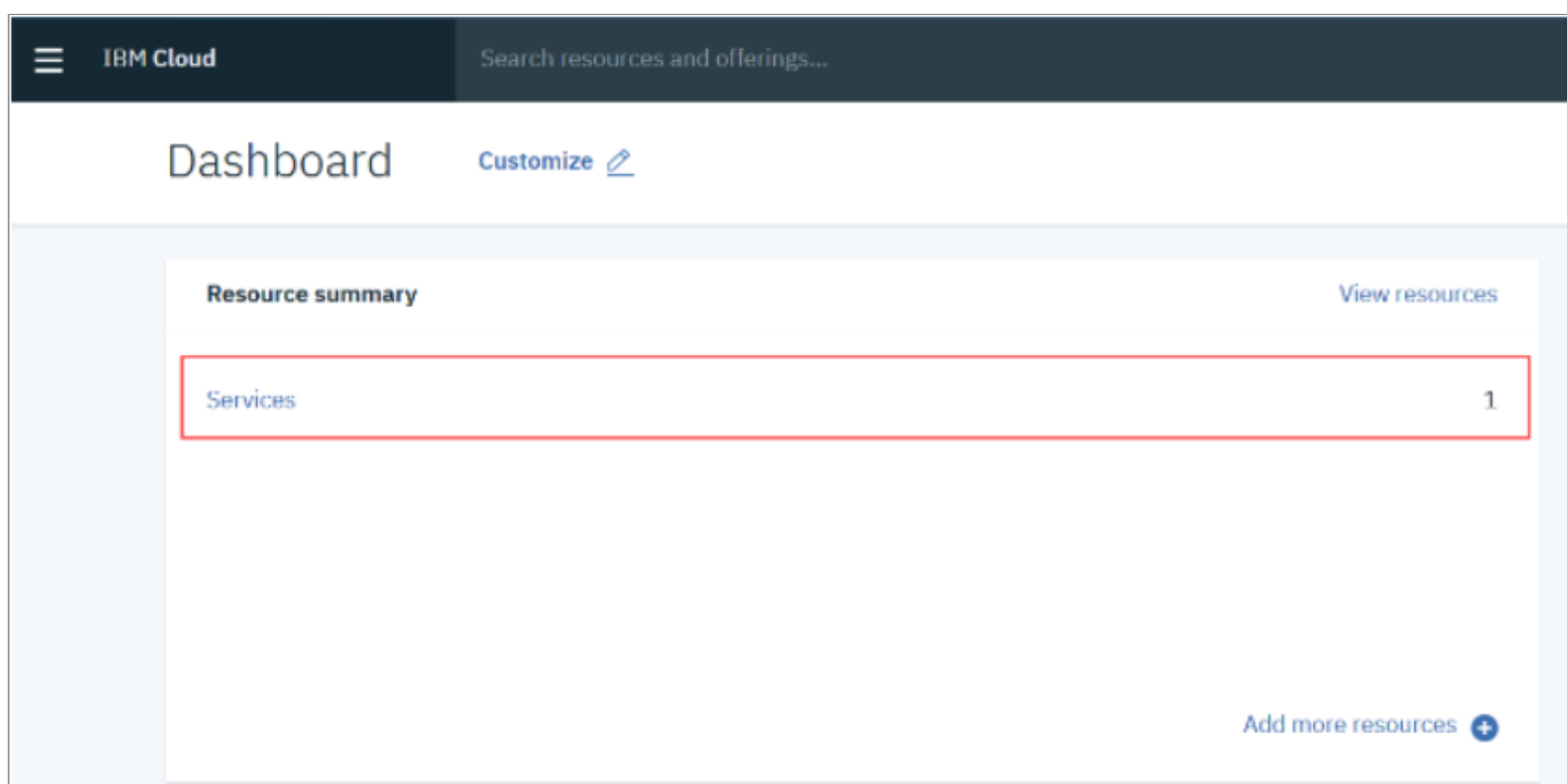
- Launch Watson Studio for accessing Data Science Problems
- Evaluate Numeric dataset
- Evaluate dataset with Non-Numeric attributes
- Evaluate Jupyter Notebook

Pre-requisite:

Before you start, you need to have an IBM Cloud account. If not, follow the instructions given in the [link](#)

Exercise 1: Launch Watson Studio for accessing Data Science Problems

1. Login to IBM Cloud: <https://cloud.ibm.com/login>
2. Scroll down and click *Services* given in *Resource Summary*.



3. When you click on Services, all your existing services will be shown in the list. Click the Watson Studio service you created:

Resource list

Create resource

Collapse all | Expand all

Name ▲	Group	Location	Offering	Status	Tags
Filter by name or IP address...	Filter by group or org...	Filter...	Filter...	Filter...	Filter...
> Devices (0)					
> VPC Infrastructure (0)					
> Clusters (0)					
> Cloud Foundry Apps (0)					
> Cloud Foundry Services (0)					
> Services (1)					
Watson Studio-0o	Default	Dallas	Watson Studio	Provisioned	—
> Storage (1)					
> Network (0)					
> Cloud Foundry Enterprise Environments (0)					
> Functions Namespaces (0)					
> Apps (0)					
> Developer Tools (0)					

4. Click *Get Started*.

IBM Cloud

Search resources and offerings...

Catalog Docs Support Manage Account


Manage

Plan

Resource list /

Watson Studio-0o

Resource group: Default Location: Dallas Add Tags



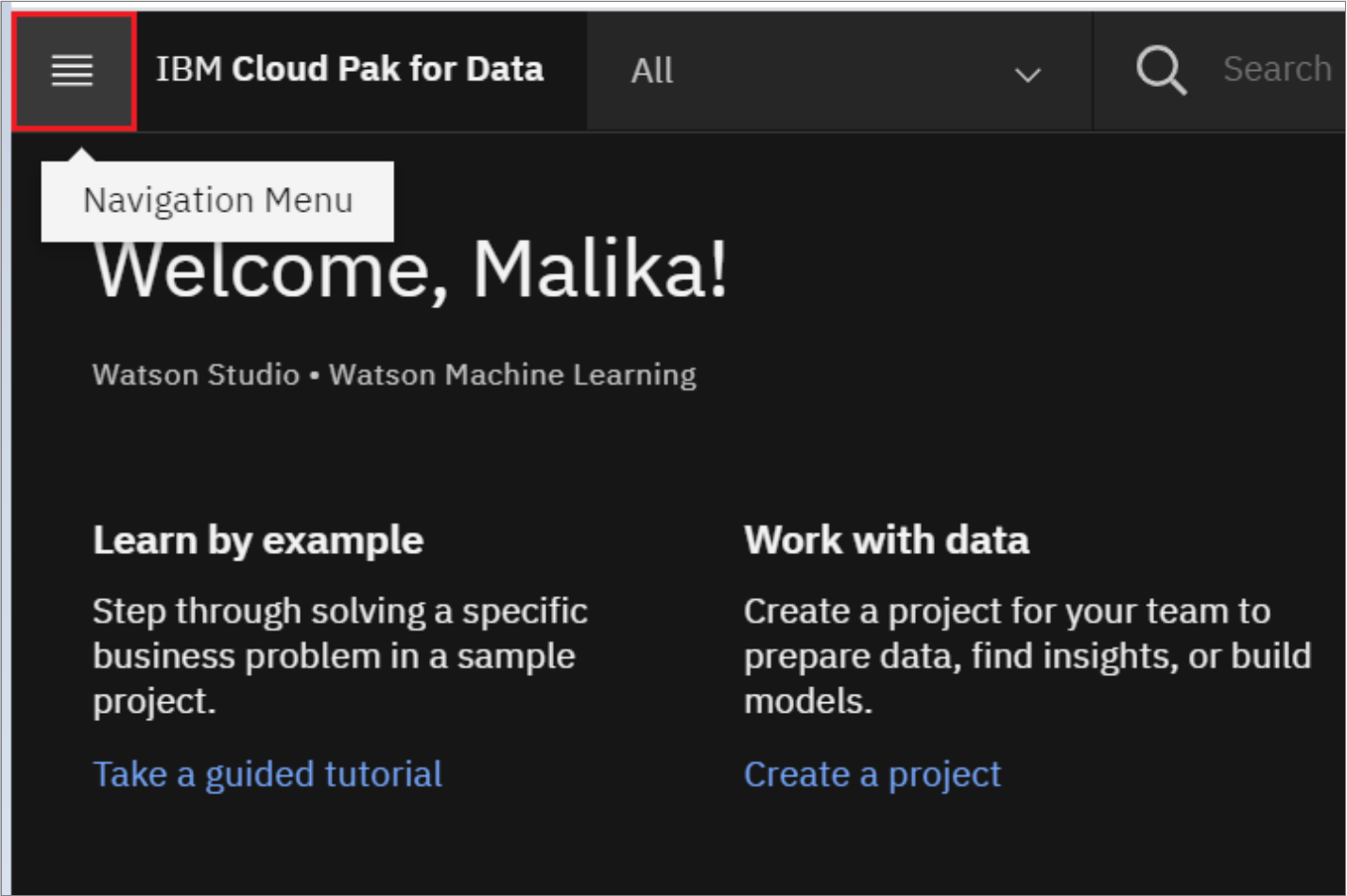
Watson Studio

Welcome to Watson Studio. Let's get started!

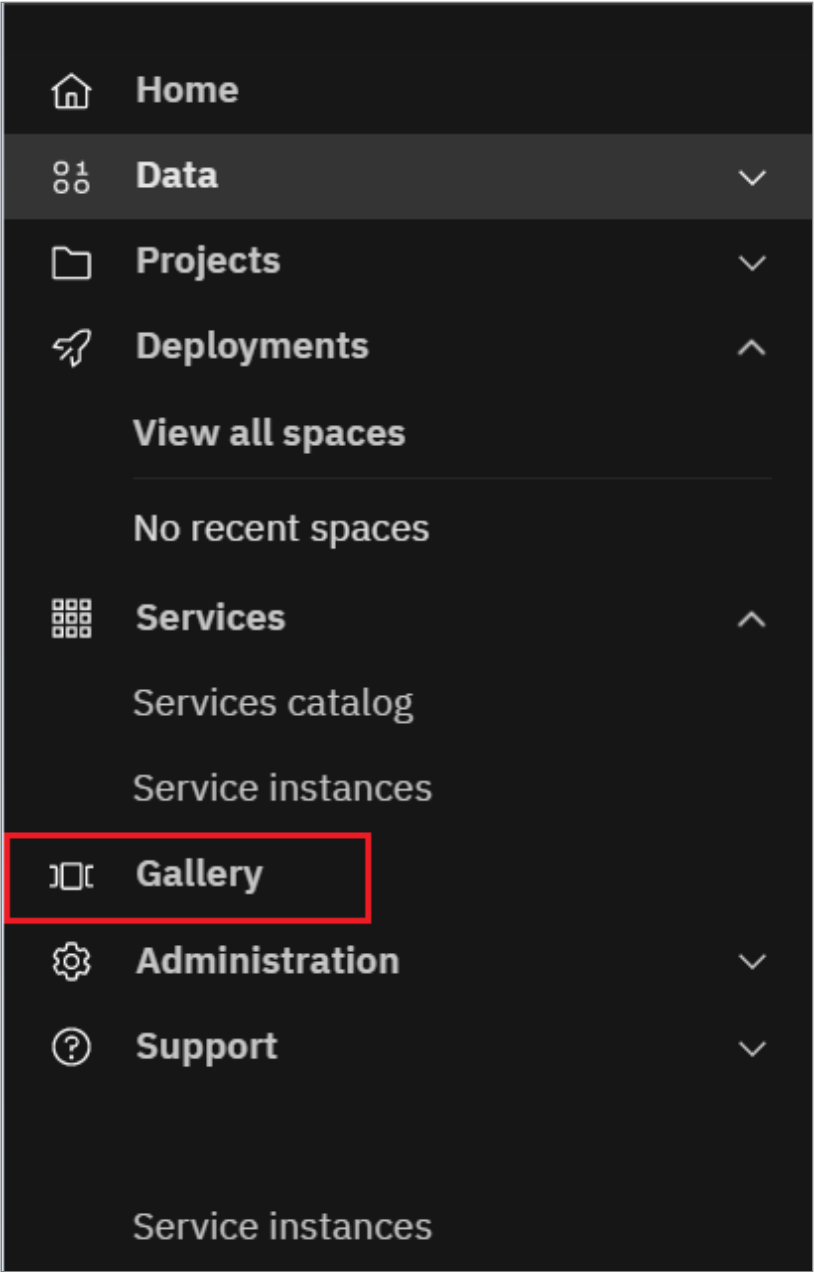
Get Started

Exercise 2: Evaluate Numeric dataset

1. Click on Navigation Menu.



2. Click on *Gallery*.



3. Select *All Filters*. From *Format* select *Data* and from *Topic* select *Energy & Utilities*, *Environment and Industry Accelerator*

All filters

Reset filters

Format

☒ Data

☐ Governance Content

☐ Notebook

☐ Sample Project

Topic

☐ Filter

☐ Art

☐ Communications

☐ Decision Optimization

☐ Economy & Business

☒ Energy & Utilities

☒ Environment

☐ Financial Markets

☐ Geography

☐ Health

☒ Industry Accelerator

☐ Insurance

☐ Law & Government

☐ Leisure

4. Click on *UCI: Forest Fires*.

IBM Watson Studio

All

Search

All filters

Reset filters

Format

☒ Data

☐ Governance Content

☐ Notebook

☐ Sample Project

Topic

☐ Filter

☐ Art

☐ Communications

☐ Decision Optimization

☐ Economy & Business

Gallery

Search results (25)

DATA

UCI: Forest fires

AUTHOR

IBM

MODIFIED

Sep 25, 2017

Environment

DATA

UCI: Iris

AUTHOR

IBM

MODIFIED

Sep 20, 2017

Environment

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

IBM Watson Studio

All

Search

Upgrade

Msingla Singla's Account

Gallery / UCI: Forest fires /

← Back

UCI: Forest fires

Tags

Environment

Modified

Sep 25, 2017

Add to project

Description		Preview										
X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
7	4	oct	tue	90.6	35.4	669.1	6.7	18	33	0.9	0	0
7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0	0
8	6	mar	fri	91.7	33.3	77.5	9	8.3	97	4	0.2	0

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
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

(this output variable is very skewed towards 0.0, thus it may make sense to model with the logarithm transform).

Exercise 2: Evaluate 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 into data having text values.

1. Use the *All Filters*. From *Format* select *Data* and from *Topic* select *Economy and Business*.

You will get mutiple datasets given. Scroll down and select *Airbnb Data for Analytics: Trentino Reviews* (If you will not get the data using **Load More** option)

DATA

Airbnb Data for Analytics: Trentino Listings

AUTHOR
IBM

MODIFIED
Dec 20, 2016

Economy & Business

DATA

Airbnb Data for Analytics: Venice Calendar

Airbnb calendar for Venice, Veneto, Italy. This dataset is sourced from [Inside Airbnb] (<http://insideairbnb.com/about.html>) which aggregates and cleanses publicly available data from Airbnb for the purpose of supporting

DATA

Airbnb Data for Analytics: Vancouver Reviews

AUTHOR
IBM

MODIFIED
Dec 20, 2016

Economy & Business

DATA

Airbnb Data for Analytics: Vancouver Listings

AUTHOR
IBM

MODIFIED
Dec 20, 2016

Economy & Business

DATA

Airbnb Data for Analytics: Vancouver Calendar

AUTHOR
IBM

MODIFIED
Dec 20, 2016

Economy & Business

DATA

Airbnb Data for Analytics: Trentino Reviews

AUTHOR
IBM

MODIFIED
Dec 20, 2016

Economy & Business

DATA

2. Preview the data using the *Preview* option.

Airbnb Data for Analytics: Trentino Reviews

Economy & Business

Dec 20, 2016

Add to project

Description		Preview								
listing_id	id	date	reviewer_id	reviewer_name	comments	listing_name	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 living-room with a balcony that goes to the north. All in all a good flat to stay. Thanks!	apartment + Wi-Fi + parking!	2845951	45.88512254895795	10.859054481189382	Marina
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!	2845951	45.88512254895795	10.859054481189382	Marina

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.

This data can be used in various 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.

💡

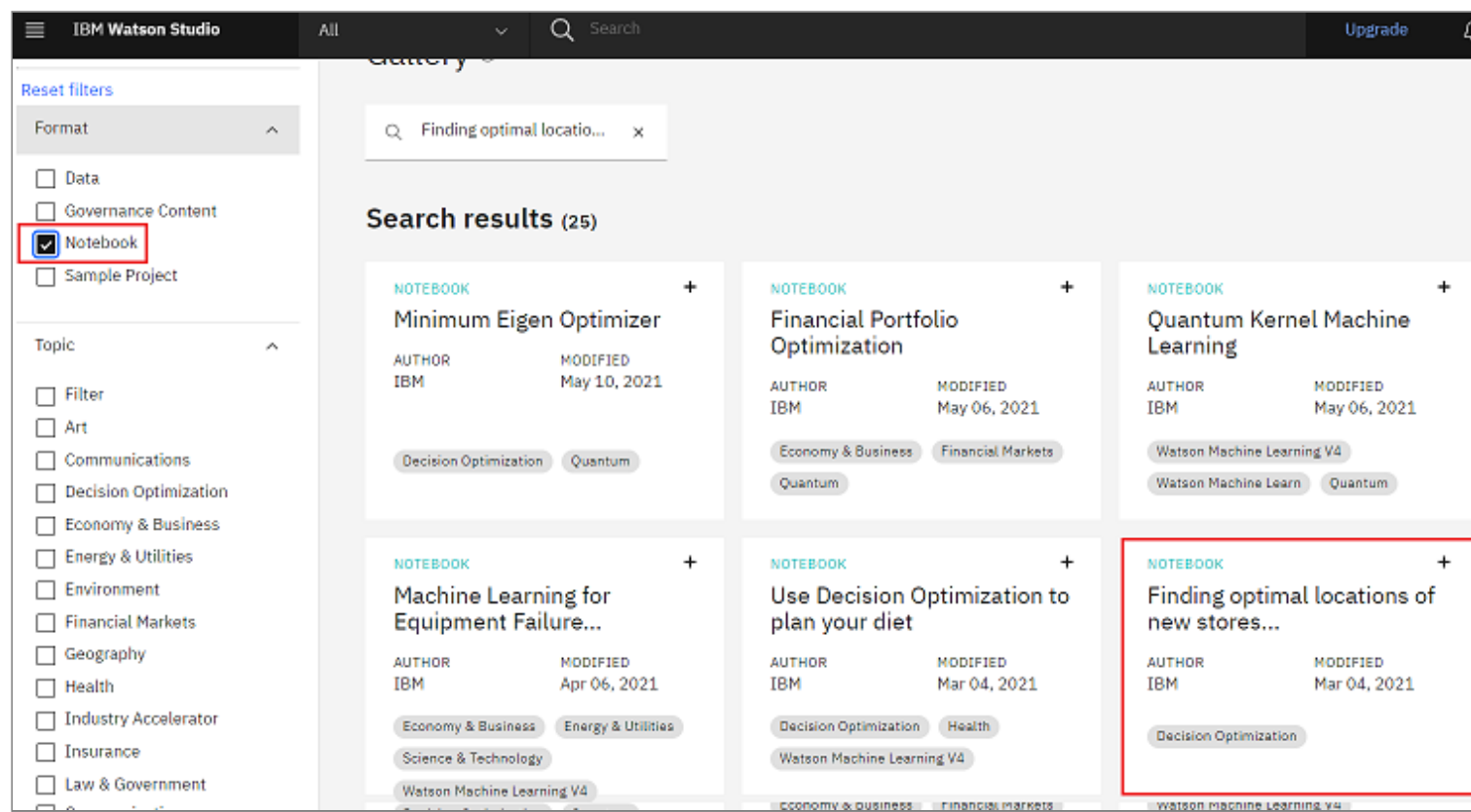
Can you think of what you can use this data for?

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).

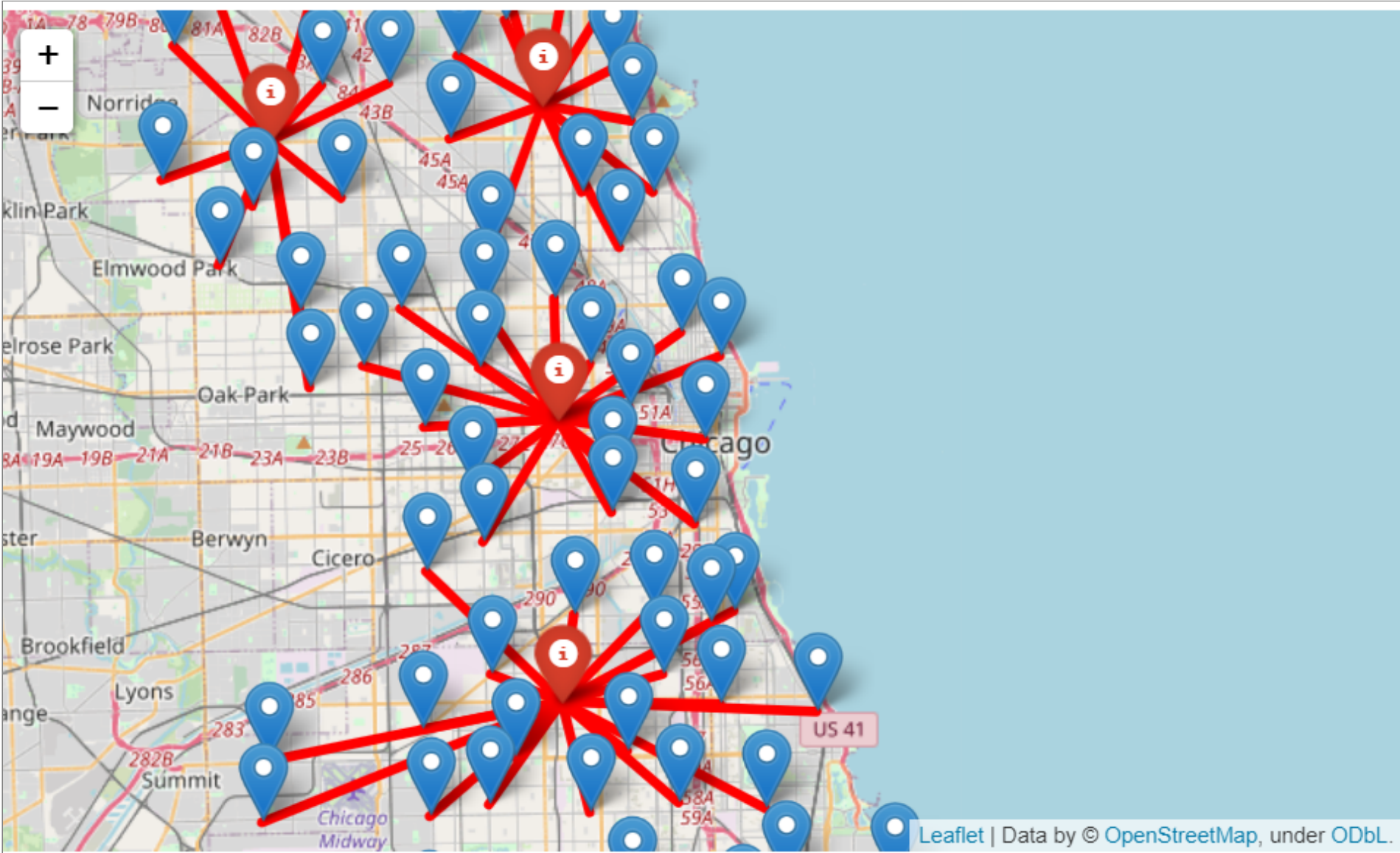
Exercise 3: Evaluate Jupyter Notebook

Use the *All Filters*. From *Format* select *Notebook* and select *Finding optimal locations of new stores using Decision Optimization*



This notebook shows you how Decision Optimization can help to prescribe decisions for a complex constrained problem using Python to help determine the optimal location for a new store.

The objective is to minimize the total distance from libraries to coffee shops so that a book reader always gets to our coffee shop easily. It can be done by analyzing and displaying the location of the coffee shops on a map.



Summary

In this lab, you have learnt about how different datasets are available and how a data scientist create and predict the models using the model building in IBM Watson Jupyter Notebook.

Author(s)

[Malika Singla](#)

Other Contributor(s)

[Lavanya](#)

Change log

Date	Version	Changed by	Change Description
2021-06-010	1.0	Malika Singla	Initial Version

© IBM Corporation 2021. All rights reserved.