



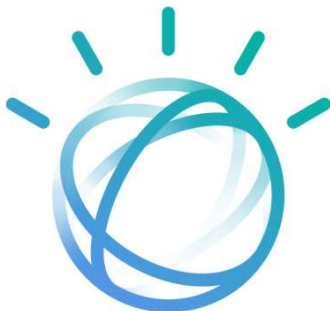
IBM Think 2018

Hands-on: Diventiamo Sommelier grazie al Machine Learning con Watson!

Daniela Zuppini, IBM, daniela_zuppini@it.ibm.com

Andrea Santi, IBM, andrea_santi@it.ibm.com

Alessio Murru, IBM, Alessio.Murru@ibm.com





Indice

Introduzione	3
Prerequisiti	4
Ingredienti	5
Registrazione a IBM Cloud	5
Set up di Watson Studio	6
Set up di un progetto in Watson Studio	9
Set up di Apache Spark in Watson Studio	12
Parte 1 - Lavoriamo con i Notebook	14
Creazione di un Notebook in Watson Studio	14
Import del dataset nel Notebook in Watson Studio	17
Esecuzione nel Notebook in Watson Studio	19
Parte 2 - Lavoriamo con Watson Modeler Flows	20
Set up di Watson Machine Learning in Watson Studio	20
Preparazione del modello con Modeler Flow	23
Deployment del modello con Watson Machine Learning	29
Reference	32



Introduzione

Vediamo da vicino cosa è il **Machine Learning** con questo semplice laboratorio. Scriveremo un programma il cui codice, però, non avrà lo scopo di definire in forma di **source code** le regole in base alle quali rispondere a degli input. Al contrario il codice che scriveremo avrà lo scopo di imparare dai dati, rappresentazione dell'esperienza passata, le regole in base alle quali rispondere agli input futuri. Il nostro codice sarà focalizzato sugli aspetti di lettura dati, sulla comprensione di quali parametri siano rilevanti per il problema e di quale sia il modello matematico migliore per effettuare apprendimento o meglio il training del programma.

Bordeaux e Machine Learning

Il caso che esaminiamo è reale e fa parte della letteratura.

Il valore di una bottiglia di vino di annata è un investimento difficile da decidere e la qualità del vino è qualcosa di misterioso affidato alla decisione dei Sommelier.

Nel 1990 un professore di economia si improvvisa Sommelier: guardando i dati delle aste del vino Bordeaux e i parametri ambientali legati alla vendemmia e alla fase di maturazione dei vigneti, il prof. Ashenfelter, dell'università di Princeton, mette a punto un modello predittivo in grado di predire la qualità del vino e il valore che potrà acquistare con l'invecchiamento. Questo il link alla pubblicazione originale:

http://media.terry.uga.edu/documents/economics/ashenfelter_predicting_quality.pdf

Il modello prende in considerazione parametri legati alle condizioni meteo nel periodo di maturazione dei vigneti e durante la vendemmia. Il risultato è un modello in grado di dire se da una vendemmia potrà nascere un vino d'annata capace di apprezzarsi con l'invecchiamento.

Il metodo che viene usato si chiama Regressione Lineare e predice i valori che può assumere una variabile dipendente a partire dai valori che assume un set di variabili indipendenti.

- una variabile del dataset viene considerata dipendente, o target, e sarà proprio questa variabile che il modello dovrà predire



- un set di variabili saranno indipendenti e saranno usate per determinare il valore della variabile dipendente

Obiettivi del laboratorio

In questo laboratorio proviamo a costruire un modello di predizione della qualità del vino partendo dal dataset originale delle aste di vino Bordeaux.

Utilizzeremo tecniche di Machine Learning che, esaminando i dati delle aste di vino passate arricchiti dei dati meteo, imparano ad associare il prezzo di vendita del vino, preso come valutazione di qualità, ai parametri meteo che hanno caratterizzato la maturazione e vendemmia del vino stesso.

Per mettere a punto il modello useremo due strade:

Parte 1) Scriveremo un Jupyter Notebook in linguaggio R. Il codice avrà il compito di leggere il dataset delle aste passate, capire quali sono i parametri rilevanti per il modello e costruire un modello di regressione lineare che collega andamento prezzi con anni di invecchiamento e dati meteo relative alla specifica bottiglia di vino.

Parte 2) Useremo il servizio di Modeler Flows e Watson Machine Learning. Costruiremo un Modeler Flow che prende il dataset di input, specificheremo quali sono le variabili dipendenti/indipendenti e faremo il training di un modello di regressione lineare.

I servizi usati in questo tutorial sono: IBM Watson Machine Learning, Apache Spark, Jupyter notebooks in linguaggio R all'interno della piattaforma IBM Watson Studio e IBM Cloud Object Storage per salvare dati e contenuto dei notebook.

Prerequisiti

Skill Level: Principiante

Piattaforma: PC con un browser e una connessione internet



Ingredienti

1. **Accesso a IBM Cloud:** nel catalogo di IBM Cloud sono a disposizione servizi IaaS e PaaS. Useremo la componente PaaS e faremo attivazione di servizi in modalita' lite a costo zero.
2. **IBM Watson Studio:** questo servizio e' una piattaforma collaborativa per la preparazione di modelli analitici avanzati. E' possibile scrivere modelli usando Scala, Python, R ed e' possibile usare framework come ScikitLearn, Tensorflow e Keras. La piattaforma consente anche di utilizzare un Modeler Flows per scrivere modelli appoggiandosi a Watson Machine Learning Modeler e per costruire Neural Network usando un Neural Network Modeler.
3. **IBM Apache Spark:** questo servizio fornisce un ambiente Apache Spark semplicemente con un click. Con Spark e' possibile eseguire operazioni sui dati in modalita' parallela. Faremo attivazione di un servizio Apache Spark specificando il piano lite che prevede due Spark executor.
4. **IBM COS:** Cloud Object Storage e' un object storage basato sullo standard S3 che e' viene usato da Watson Studio per salvare dati e notebook. Nel piano lite sono compresi 25GB di IBM COS a costo zero.
5. **IBM Watson Machine Learning:** Con WML e' possibile scrivere modelli complessi, fare il loro training e il deploy. Lavorando in collaborazione con il Modeler Flow di Watson Studio e' possibile costruire i modelli usando una modalita' grafica.

Registrazione a IBM Cloud

Per registrarti ad IBM Cloud puoi utilizzare il promo code della Cognitive Class:

<https://cognitiveclass.ai/ibm-cloud-promotion/>

Usando questo link puoi iscriverti ad un portale di corsi on-line gratuiti e specializzati su Data Science, Big Data, Machine Learning, Deep Learning. Insieme ai corsi viene fornito un promo code che fornisce un credito di 1200\$ di risorse IBM Cloud.

Clicca qui (<https://courses.cognitiveclass.ai/login>) per registrarti e fare login alla Cognitive Class.

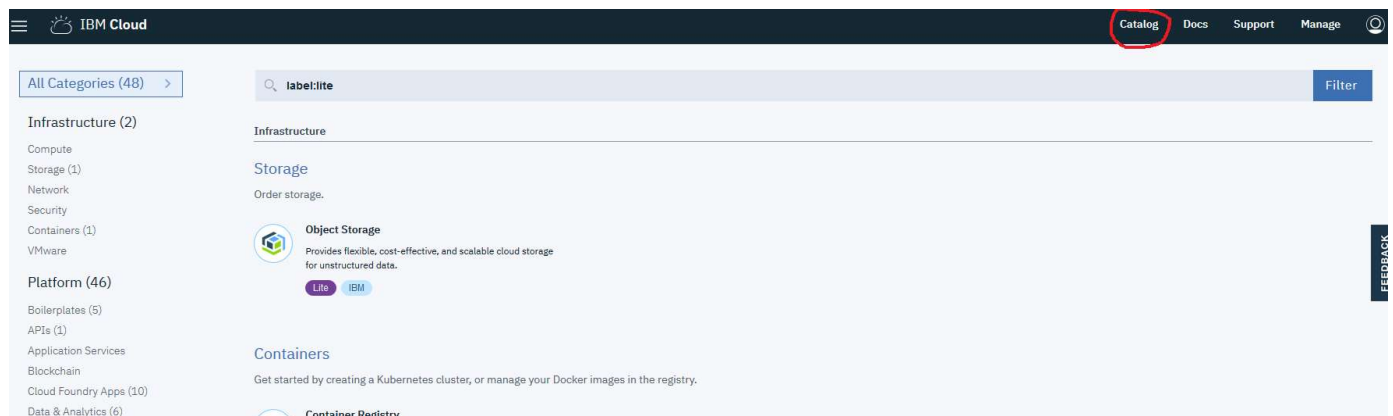
Ottieni il tuo promo code nella Welcome page (coloro che sono gia' iscritti alla CognitiveClass trovano il promo code nella loro [Dashboard](#))

[Registrati ad IBM Cloud](#) e applica il tuo promo code.

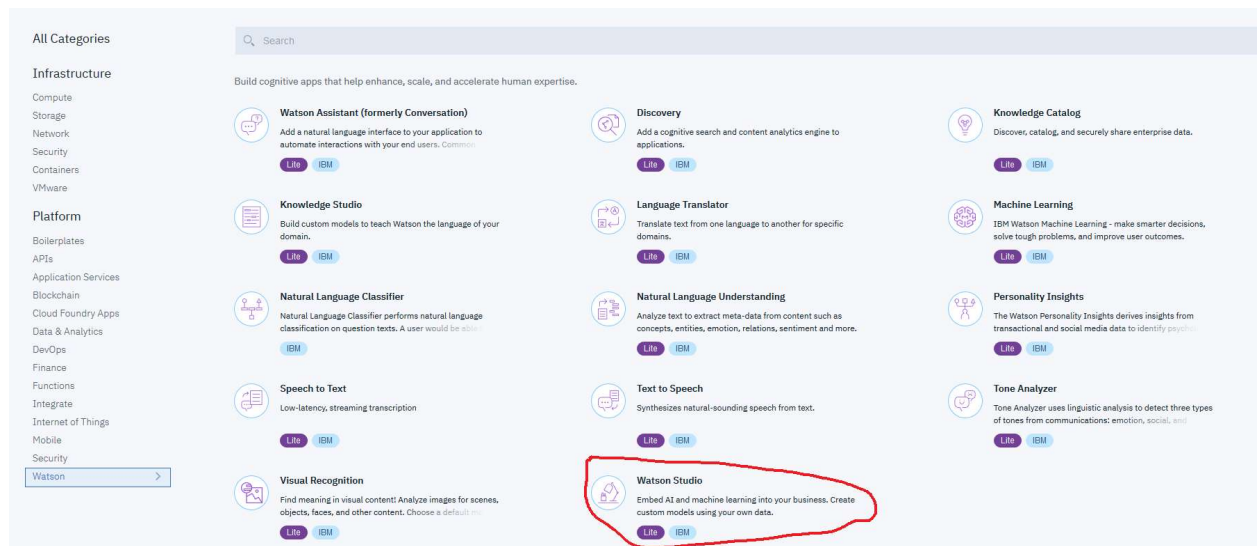


Set up di Watson Studio

1. Fai il Login ad [IBM Cloud](#).
2. Seleziona il Catalogo.



3. Nella sezione **Platform**, a sinistra dello schermo, seleziona la Platform Watson.
4. Scegli il servizio Watson Studio facendo doppio click.



Quando si apre la pagina del servizio Watson Studio verifica che sia selezionato il piano lite e seleziona Create:



Hands-on: Diventiamo Sommelier grazie al Machine Learning con Watson!

Watson Studio

Watson Studio democratizes machine learning and deep learning to accelerate infusion of AI in your business to drive innovation. Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts.

[Use IBM](#)

[View Docs](#) [Terms](#)

AUTHOR IBM
PUBLISHED 05/29/2018
TYPE Service
LOCATION United Kingdom, US South

Features

- Use what you know, learn what you don't**
Start from a tutorial, start from a sample, or start from scratch. Tap into the power of the best of open source (RStudio, Jupyter Notebooks) and Watson services for flexible model creation. Use Python, R, or Scala. Stop downloading and configuring analysis environments and start getting insights.
- Power on demand**
Enterprise-scale features on demand. From data exploration and preparation, to enterprise-scale performance. Manage your data, your analytical assets, and your projects in a secured cloud environment.
- Be a founding member**
Take advantage of shared data sets, notebooks, models, and tutorials. Share your work with your team and your peers across job roles. Join a vibrant community of data scientists, developers, and domain experts across industries, functions, and organization types.
- Collaborate for better outcomes**
Work with your peers on projects to find better solutions together. Share your knowledge and your work easily with visualizations and code – and help fuel the advancement of data science and AI for all.

Images

Click an image to enlarge and view screen captures, slides, or videos. Screen caps show the user interface for the service after it has been provisioned.

Pricing Plans

Monthly prices shown are for country or region: [United States](#)

PLAN	FEATURES	PRICING
<input checked="" type="checkbox"/> Lite	1 authorized user 50 capacity unit-hours monthly limit 1 free email compute environment with 1 vCPU and 4 GB RAM (does not require capacity unit-hours) The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment. Lite plan services are deleted after 30 days of inactivity.	Free
Standard v1	1 authorized user + unlimited viewer collaborators	\$99.00 USD/instance

[Need Help? Contact IBM Cloud Sales](#)

Already have an account? [Log In](#) [Sign up to Create](#)

Quando il servizio Watson e' partito, nella pagina principale di Watson Studio clicca sul tasto **GET STARTED**

IBM Cloud

Catalog Docs Support Manage

Manage

Plan

Watson /

Watson Studio-si

Location: United Kingdom Org: coccotour@gmail.com Space: dev

Watson Studio

Welcome to Watson Studio, let's get started.

[Get Started](#)



Al primo accesso viene chiesto di confermare o modificare le informazioni di accesso allo spazio cloud. Scegli di confermare i valori proposti.

The screenshot shows a web form titled "Select Organization and Space". Below the title, it says "Confirm your IBM Cloud organization and space information below." and provides a link: "[Or create new organization and space](#)". There are three dropdown menus: "Select IBM Cloud account" (empty), "IBM Cloud Organization" (empty), and "IBM Cloud Space" (with "dev" selected). A "Continue" button is at the bottom right.

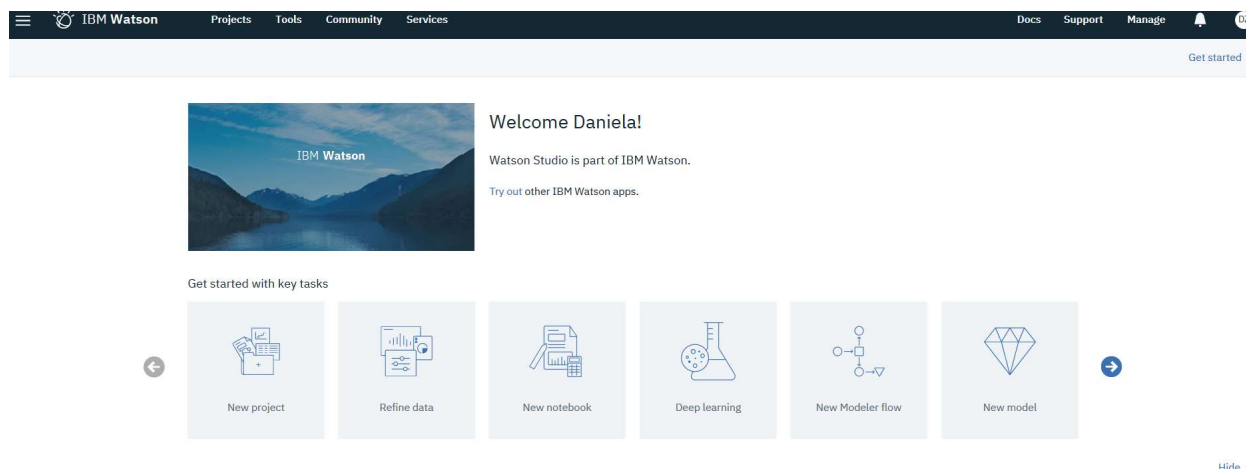
Il tuo account viene correttamente creato:

The screenshot shows a confirmation screen with a checkmark icon and the text "Done!". Below it, it says "Your account is created." and there is a "Get Started" button at the bottom right.

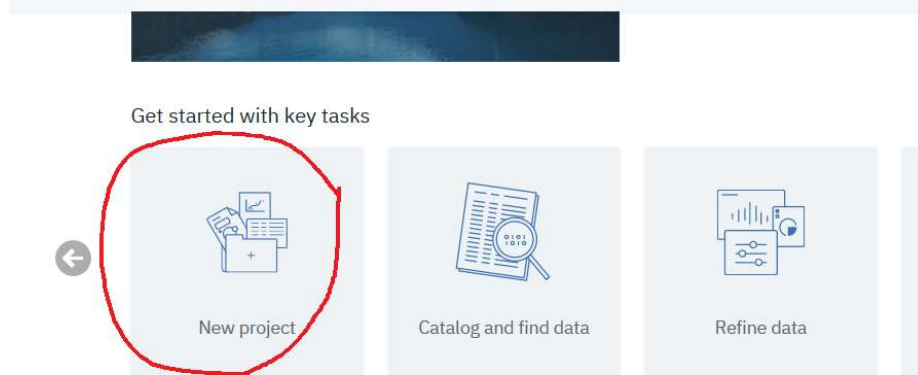


Set up di un progetto in Watson Studio

Nella pagina principale di Watson Studio:



Crea un nuovo progetto e accetta la proposta di attivare tutti i servizi.



Se hai appena creato il servizio di Watson Studio, e questo e' il tuo primo progetto, devi attivare anche un servizio di IBM Cloud Object Storage (COS).

Scegli Add per attivare il servizio di IBM COS.



New project

Define project details

Name

Project name

100

Description

Project description

Define storage

① Select storage service

Add

Add an object storage instance and then ret
Refresh.

② Refresh

A questo punto si apre il servizio di Cloud Object Storage. Verifica che sia stato selezionato il piano lite e clicca su **create** per creare il servizio.

Cloud Object Storage

Existing **New**

Cloud Object Storage

IBM Cloud Object Storage is a highly scalable cloud storage service, designed for high durability, resiliency and security. Store, manage and access your data via our self-service portal and RESTful APIs. Connect applications directly to Cloud Object Storage use other IBM Cloud Services with your data.

Features

Storage for the IBM Cloud

IBM Cloud Object Storage provides unstructured data storage for cloud applications. Libraries and SDKs support a common set of S3 API functions for connecting new applications to scalable cloud storage and integrating your data into other services on the IBM Watson and Cloud Platform.

IAM Policies - Bucket level access management

IBM Identity and Access Management (IAM) integration allows for granular access control the bucket level using role-based policies.

Lite and pay-as-you-go plans

Pricing Plan: Monthly Process shown above reflect the: [United States](#)

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	1 COS Service Instance Storage up to 25 GB/mo. Up to 20,000 GET requests/mo. Up to 2,000 PUT requests/mo. Up to Data Retrieval 10 GB/mo. Up to 5GB Public Outbound Applies to aggregate total across all storage bucket classes	Free



Conferma la creazione del servizio:

Confirm Creation

Plan
Lite

Resource group
Default

Service name
cloud-object-storage-nu

Cancel Confirm

Fai click su refresh del pannello creazione del progetto. Comparirà il nome dell'object storage appena creato nel cloud:

Storage

cloud-object-storage-nu

Diamo un nome al nuovo progetto e scegliamo **create**:

New project

Define project details

Name
vinipertutti



Set up di Apache Spark in Watson Studio

Quando il progetto viene creato, si apre la finestra di lavoro del progetto:

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'My Projects / vinipertutti'. Below it, a sub-navigation bar includes 'Overview', 'Assets', 'Environments', 'Bookmarks', 'Deployments', 'Collaborators', and 'Settings'. The main content area displays project details for 'vinipertutti', including 'Last Updated: May 18 2018' and a 'Readme' link. On the right, there are three large numbers: 0 Assets, 0 Bookmarks, and 1 Collaborators. Below this, there's a 'Recent activity' section with a placeholder for alerts. On the left, there's a 'Date created' section (May 18 2018), a 'Description' section (No description available), a 'Storage' section (0% of 25 GB used), a 'Collaborators' section (showing 'Admin' with a redacted name), and a 'Bookmarks' section.

Abbiamo bisogno di attivare un servizio Apache Spark su cui fare lavorare il notebook. Dal menu di progetto scegli di aprire la pagina dei settings:

The screenshot shows the 'Settings' page for the 'vinipertutti' project. The 'Settings' tab is highlighted in the sub-navigation bar. The main content area is titled 'Project information' and shows the 'Project name' as 'vinipertutti'.

Nella pagina settings cerca la sezione Associated services, apri il menu a tendina e scegli il servizio Spark:

The screenshot shows the 'Associated services' section in the settings page. There's a table with columns 'NAME' and 'ACTIONS'. To the right of the table, there's a dropdown menu with the following options: 'Amazon EMR Spark', 'IBM Analytics Engine', 'Spark' (which is highlighted with a red underline), 'Machine Learning', 'Streaming Analytics', 'Analytics Dashboard', and 'Watson'. Above the dropdown, there's a button labeled 'Add service'.



Si apre la finestra del servizio Apache Spark disponibile nel catalogo di IBM Cloud. Seleziona il piano Lite e procedi con **create** per istanziare il servizio. Il piano lite attiva due Spark executor.

Apache Spark

Existing **New**

Apache Spark

Apache Spark is an open source cluster computing framework optimized for extremely fast and large scale data processing, which you can access via the newly integrated notebook interface IBM Analytics for Apache Spark. You can connect to your existing data sources or take advantage of the on-demand big data optimization of Object Storage. Spark plans are based on the maximum number of executors available to process your analytic jobs. Executors exist only as long as they're needed for processing, so you're charged only for processing done.

Features

Incredibly Fast
Apache Spark delivers 100x the performance of Apache Hadoop for certain workloads because of its advanced in-memory computing engine.

Easy to Use and Powerful
Apache Spark's Streaming and programming models backed by GraphX make it incredibly easy and data scientists to build app machine learning and graph analytics. The service is 100% compatible Spark, developers can build the them against the IBM managed benefit from operational, maintain hardware excellence.

Pricing Plan: Monthly Process shown above reflect the: United States

PLAN	FEATURES	PRICING
Lite	2 Spark Executors	Free

An entry level plan to run programs using up to 2 Spark executors

A questo punto il servizio Spark compare tra i servizi associati al progetto.

Associated services	
NAME	SERVICE TYPE
spark-yc	Spark



Parte 1 - Lavoriamo con i Notebook

Creazione di un Notebook in Watson Studio

Un Notebook e' un ambiente web-based di tipo interattivo. L'ambiente e' organizzato in celle che possono contenere codice sorgente oppure testo descrittivo on formato markdown. Il codice contenuto nella singola cella puo' essere eseguito e l'output viene mostrato a video immediatamente sotto alla cella stessa.

Quindi e' possibile eseguire celle il cui risultato e' la produzione di plot che vengono poi immediatamente visualizzati.

All'interno di Watson Studio e' possibile scrivere Notebook nei linguaggi Scala, Python e R ed e' possibile utilizzare framework come SparkMLlib, Scikit-Learn, TensorFlow e Keras. Qui trovi la lista completa:

https://datascience.ibm.com/docs/content/analyze-data/pm_service_supported_frameworks.html?context=analytics

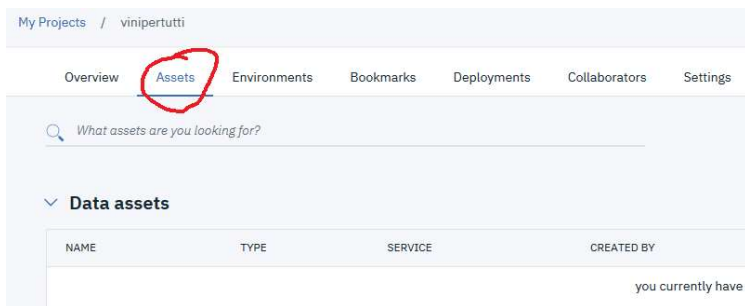
E' possibile avere piu' Notebook aperti contemporaneamente lavorando su finestre browser distinte. E' possibile costruire un ambiente collaborativo a livello di progetto: i membri del team sono invitati a fare parte del progetto stesso e a condividerne non solo il codice sorgente ma anche gli output prodotti.

Un nuovo Notebook puo' essere creato partendo da zero e creando tutte le sue celle, oppure caricando il Notebook da file o da URL.

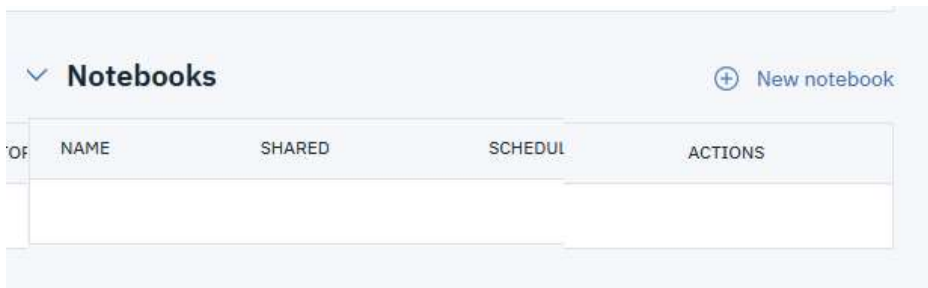
Watson Studio include una grande vastita' di librerie e package. Per installare una nuova libreria e' possibile eseguirne l'installazione da cella di tipo Code usando le istruzioni standard del linguaggio in uso.



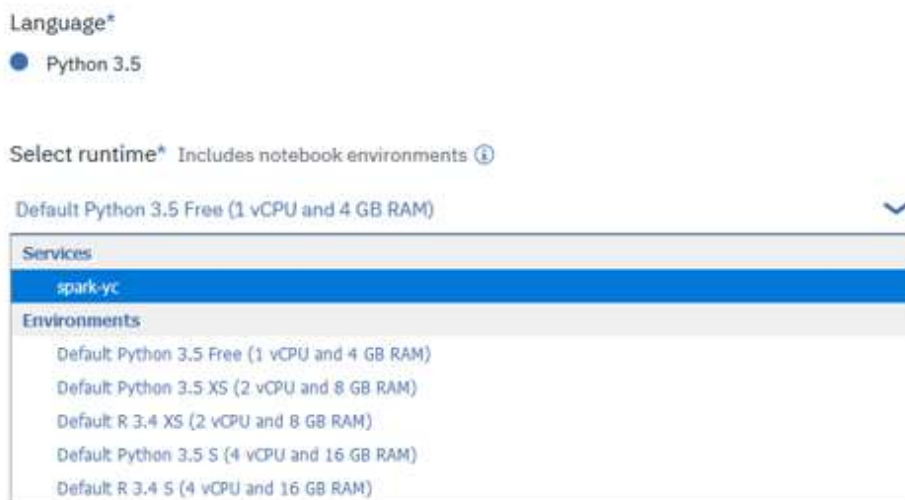
Dal menu di progetto scegli di aprire la pagina degli asset:



Vai nella sezione notebook e scegli di creare un new notebook:



Vai nella sezione runtime e cambia il runtime che viene presentato scegliendo il servizio Spark appena creato:





A questo punto cambia il linguaggio di programmazione e scegli R:

Language*

☐ Python 3.5 ☐ Python 2 ☒ R ☐ Scala

Spark version*

☒ 2.1

Select runtime* Includes notebook environments ⓘ

spark-yc

Associate this notebook with the runtime service of your choice.

Il nuovo Notebook va creato a partire da URL. Quindi scegli **From URL** e nella box a fianco Notebook URL specifica questo URL

<https://github.com/Danizu/siamosommelier/blob/master/Notebookvini.ipynb>

Notebook URL*

<https://github.com/Danizu/siamosommelier/blob/master/Notebookvini.ipynb>

Select runtime* Includes notebook environments ⓘ

spark-yc

Associate this notebook with the runtime service of your choice.

Crea il nuovo notebook dando un nome:

New notebook

Blank From file From URL

Name*

Notebookvini

38 Characters
Remaining

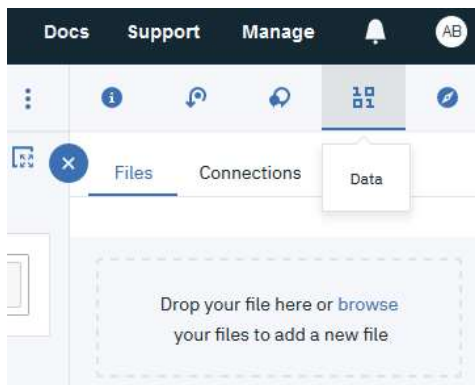
Procedi con **create notebook**.



Nel Notebook leggi con attenzione le celle di testo con le spiegazioni.

Import del dataset nel Notebook in Watson Studio

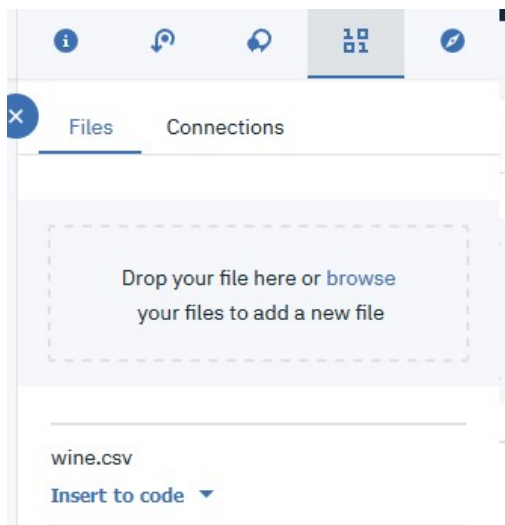
Si apre il notebook e compare la prima cella del notebook. La cella e' vuota al momento. Carichiamo il dataset wine.csv. Nel menu in alto a destra scegliete di aprire la sezione dati cliccando sull'icona data:



Seguendo le istruzioni, porta il file sull'object storage con drag and drop oppure facendo browse del disco. Il file puo essere scaricato da questo link:

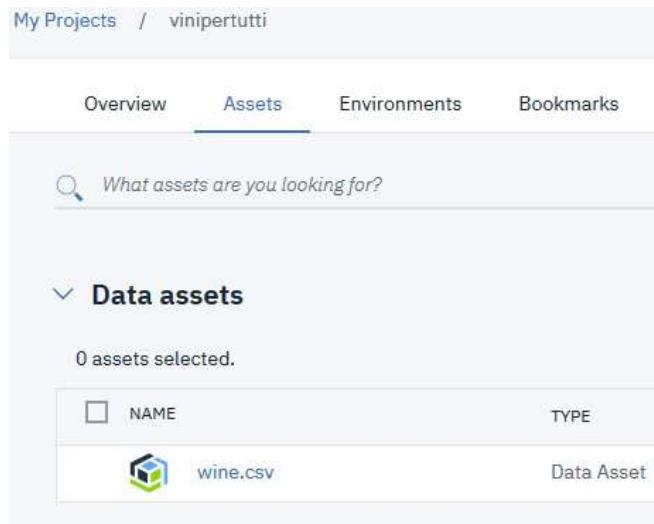
<https://github.com/Danizu/siamosommelier>

Per caricarlo nel progetto vai alla sezione dati e fai Load del file:





Quando viene completato il caricamento del file nel progetto, il file compare anche nella sezione dei **data asset**:



Posizionati con il cursore all'interno della prima cella di codice che contiene il commento : **# The code was removed by DSX for sharing.**

Apri il menu **insert to code** e scegli **Insert R DataFrame.**

Watson Studio genera automaticamente una cella di codice in linguaggio R che contiene le variabili ambiente per accedere e leggere l'object storage, effettua la lettura del file wine.csv che abbiamo caricato sull'object storage e carica il contenuto del file in un dataset. Fai attenzione alle istruzioni contenute nel testo del Notebook. In particolare **rinomina il nome del dataset da df.data.1 in wine.**

Con la combinazione di tasti CTRL-ENTER, la cella viene eseguita:

```
In [1]: library("aws.s3")

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share your notebook.
Sys.setenv("AWS_ACCESS_KEY_ID" = "al42ee6423149e3b40da6ac254fb0bc", "AWS_SECRET_ACCESS_KEY" = "ef51f0d23fafb363939a730e16279f93ceale2e7e0da29f7")
url <- "s3.eu-geo-objectstorage.service.networklayer.com"
bucket <- "vinipertutti-donotdelete-pr-yla9xhyfhhc2z"
headers <- list("x-amz-content-sha256"="e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855")

obj <- s3HTTP(
  verb = "GET",
  bucket = bucket,
  headers = headers,
  path = "wine.csv",
  key = Sys.getenv("AWS_ACCESS_KEY_ID"),
  secret = Sys.getenv("AWS_SECRET_ACCESS_KEY"),
  check_region = FALSE,
  base_url = url)

df.data.1 <- read.csv(text = rawToChar(obj$content))
head(df.data.1)
```

Year	Price	WinterRain	AGST	HarvestRain	Age	FrancePop
1952	7.4950	600	17.1167	160	31	43183.57
1953	8.0393	690	16.7333	80	30	43495.03
1955	7.6858	502	17.1500	130	28	44217.86
1957	6.9845	420	16.1333	110	26	45152.25
1958	6.7772	582	16.4167	187	25	45653.81



Esecuzione nel Notebook in Watson Studio

Leggi con attenzione il contenuto del Notebook ed esegui, in modo interattivo, il codice contenuto nelle celle.

Per eseguire una cella di codice, posizionali con il mouse sulla cella ed esegui CTRL-ENTER.

Il notebook contiene delle celle incomplete da usare come esercizio. Esegui le celle completando con il codice mancante.

Una volta completato passa alla seconda parte del Laboratorio.