

# Diventiamo sommelier grazie al Machine Learning con Watson!

## PARTE 1

### Skill Level: Basico

Il valore di una bottiglia di vino di annata e' un investimento difficile da decidere e la qualita' del vino e' 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'universita' di Princeton, mette a punto un modello predittivo in grado di predire la qualita' del vino e il valore che potra' acquistare con l'invecchiamento. Questo il link alla pubblicazione originale:

[http://media.terry.uga.edu/documents/economics/ashenfelter\\_predicting\\_quality.pdf](http://media.terry.uga.edu/documents/economics/ashenfelter_predicting_quality.pdf)

In questo laboratorio proviamo a costruire questo modello. Utilizzeremo tecniche di Machine Learning che, esaminando i dati delle aste passate, imparano ad associare il prezzo di vendita del vino, preso come valutazione di qualita', ai parametri meteo che hanno caratterizzato la maturazione e vendemmia del vino stesso. Per mettere a punto il modello useremo due strade:

1) scriveremo un Jupyter Notebook in linguaggio R. Il codice avra' 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.

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/Indipendente 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 notebook.

## Ingredienti

1. **Piattaforma:** PC con un browser e una connessione internet

2. **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.
3. **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.
4. **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.
5. **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.
6. **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.

## Step-by-step

### 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.

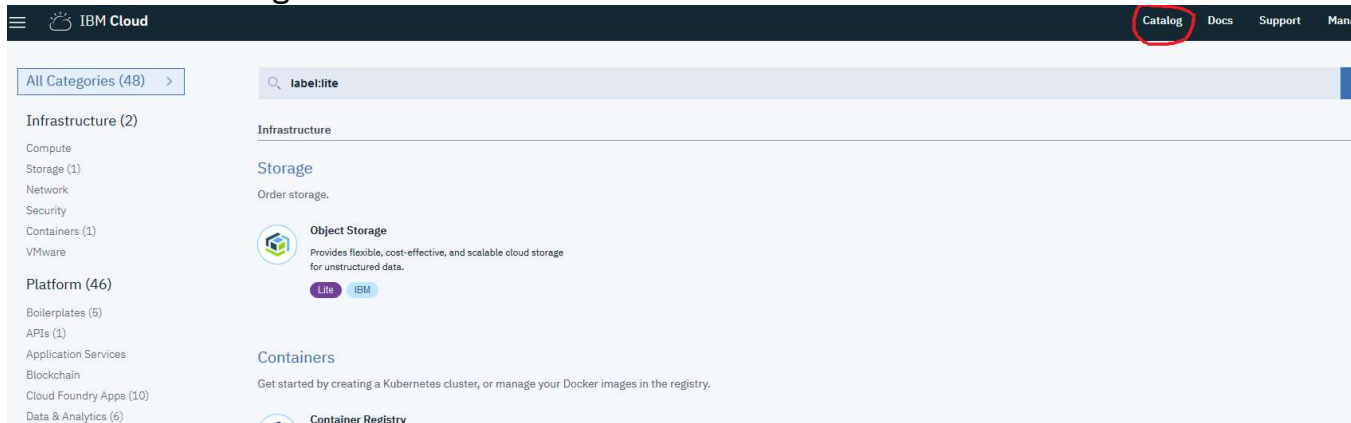
[Clicca qui \(https://courses.cognitiveclass.ai/login\)](https://courses.cognitiveclass.ai/login) per registrarti e fare login alla Cognitive Class.

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

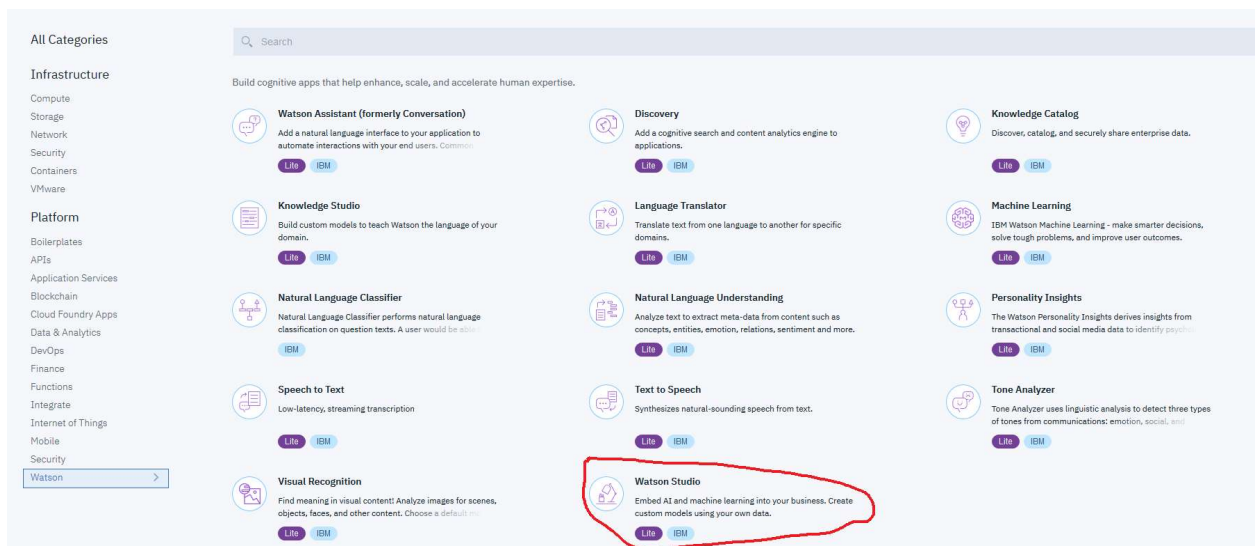
[Registrati ad IBM Cloud](#) e applica il tuo promo code.

# 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.

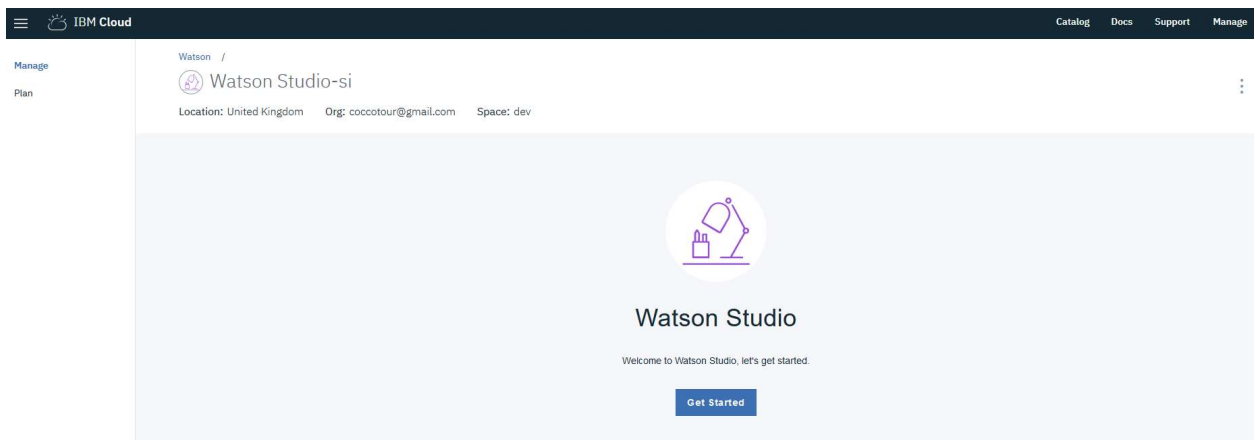


Quando si apre la pagina del servizio Watson Studio verifica che sia selezionato il piano lite e seleziona Create:

Pricing Plans			Monthly prices shown are for country or region: <a href="#">Italy</a>
PLAN	FEATURES	PRICING	
✓ Lite	<ul style="list-style-type: none"> <li>1 authorized user</li> <li>50 capacity unit-hours monthly limit</li> <li>1 free small compute environment with 1 vCPU and 4 GB RAM (does not require capacity unit-hours)</li> </ul> <p>The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment.</p> <p>Lite plan services are deleted after 30 days of inactivity.</p>	Free	
Standard v1	<ul style="list-style-type: none"> <li>1 authorized user + unlimited viewer collaborators</li> <li>50 capacity unit-hours included monthly (additional capacity available)</li> <li>Unlimited elastic compute environments</li> <li>Capacity Type: 1 vCPU and 4 GB RAM = 0.5 capacity units required per hour</li> <li>Capacity Type: 2 vCPU and 8 GB RAM = 1 capacity units required per hour</li> <li>Capacity Type: 4 vCPU and 16 GB RAM = 2 capacity units required per hour</li> <li>Capacity Type: 8 vCPU and 32 GB RAM = 4 capacity units required per hour</li> <li>Capacity Type: 16 vCPU and 64 GB RAM = 8 capacity units required per hour</li> <li>Data Flow, Sampling and Profiling jobs (2 concurrent jobs) = 6 capacity units required per hour</li> </ul>	€74.47 EUR/Instance €0.37 EUR/Capacity Unit-Hour €74.47 EUR/Authorized User	



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



Al primo accesso viene chiesto di confermare o modificare le informazioni di accesso allo spazio cloud.

Scegli di confermare i valori proposti.

**Select Organization and Space**

Confirm your IBM Cloud organization and space information below.  
[Or create new organization and space](#)

Select IBM Cloud account

IBM Cloud Organization

IBM Cloud Space  
dev

Continue

Il tuo account viene correttamente creato:

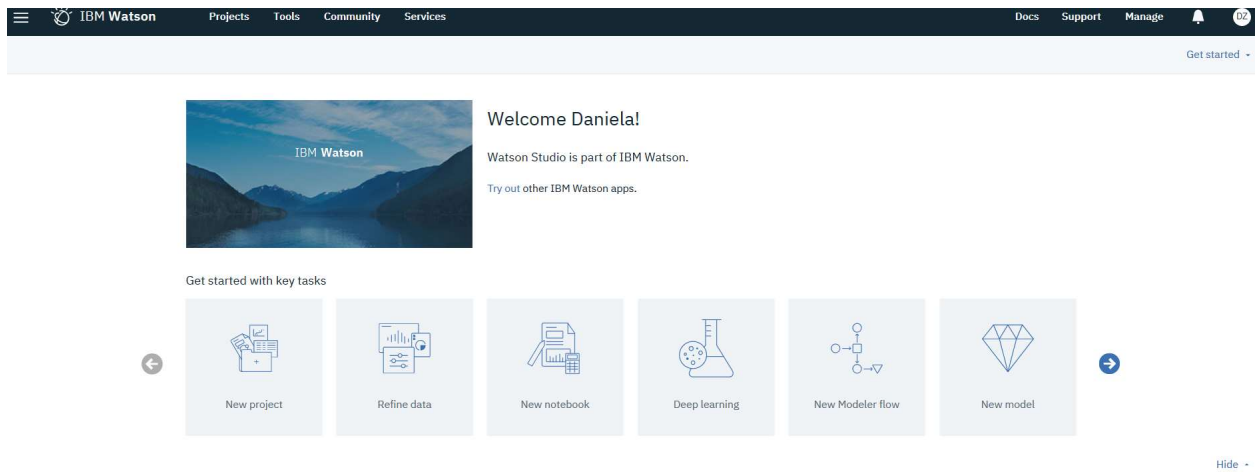
✓ Done!

Your account is created.

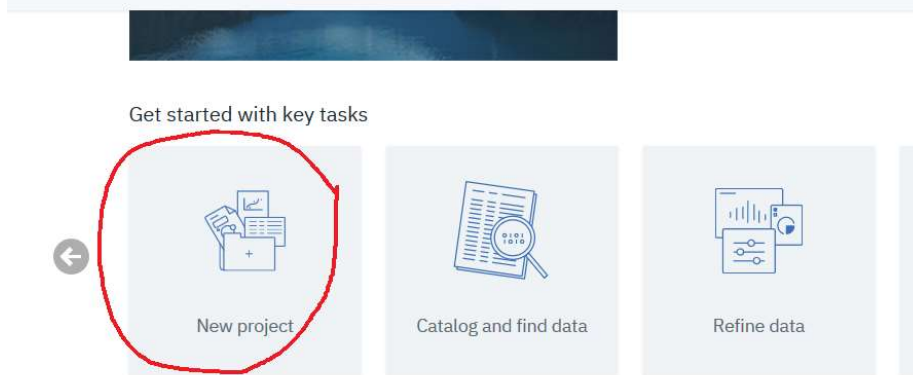
Get Started

## Set up di un progetto in Watson Studio

A questo punto si apre la 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 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	<p>1 COS Service Instance</p> <p>Storage up to 25 GB/mo.</p> <p>Up to 20,000 GET requests/mo.</p> <p>Up to 2,000 PUT requests/mo.</p> <p>Up to Data Retrieval 10 GB/mo.</p> <p>Up to 5GB Public Outbound</p> <p>Applies to aggregate total across all storage bucket classes</p>	Free

Conferma la creazione del servizio:

Choose for the IBM Cloud IAM Policies - Bucket level access Provisional endpoint Provisional endpoint

## Confirm Creation

Plan  
Lite

Resource group  
Default

Service name  
cloud-object-storage-nu

Cancel Confirm

Fai click su refresh del pannello creazione del progetto. Comparira' 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

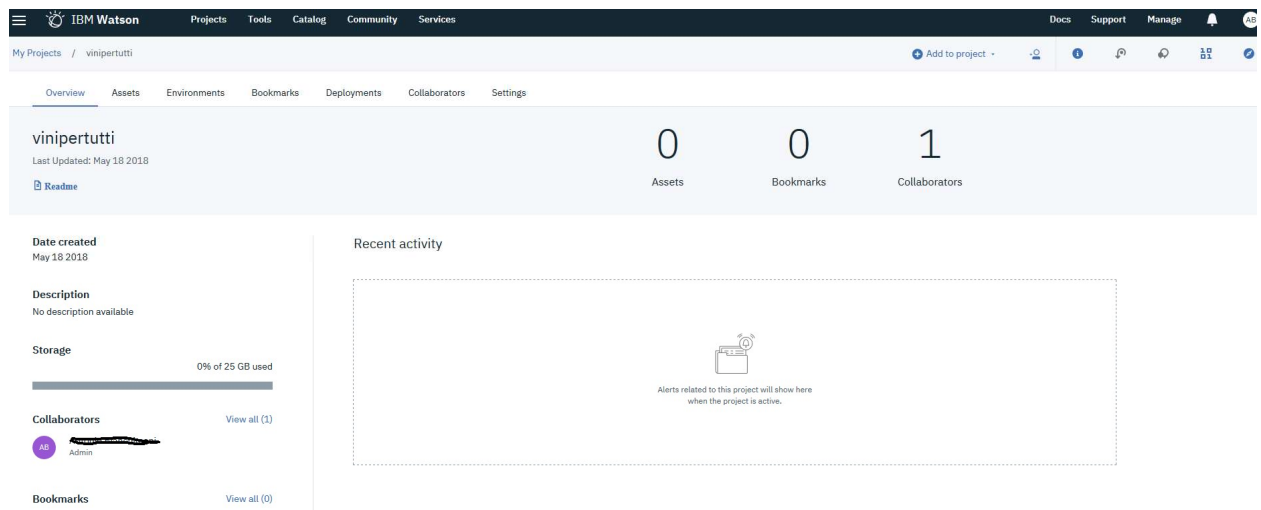
### Define project details

Name  
vinipertutti

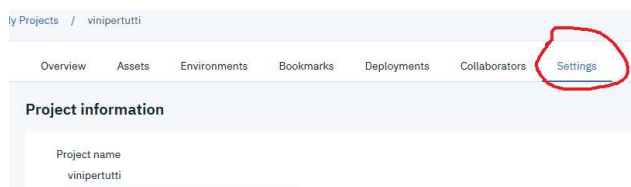
88

## Set up di Apache Spark in Watson Studio

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



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:



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



### Associated services

NAME	ACTIONS

### Access tokens

NAME	ACTIONS

[+ Add service](#)

- Amazon EMR Spark
- IBM Analytics Engine
- Spark**
- Machine Learning
- Streaming Analytics
- Analytics Dashboard
- Watson

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 servizi. Il piano lite attiva due Spark executor.

IBM Watson

[Projects](#)
[Tools](#)
[Catalog](#)
[Community](#)
[Services](#)

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 an the service is 100% compatible Spark, developers can build the them against the IBM managed benefit from operational, maint hardware excellence.

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

PLAN	FEATURES	PRICING
<div> <div></div> <div>Lite</div> </div> <div>An entry level plan to run programs using up to 2 Spark executors</div>	2 Spark Executors	Free

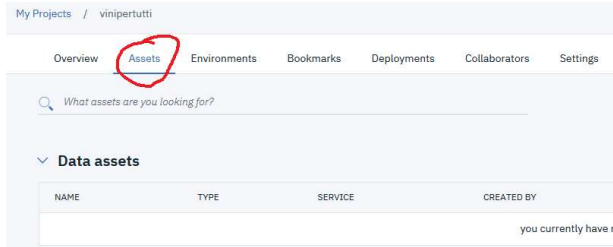
A questo punto il servizio Spark compare tra i servizi associati al progetto.

### Associated services

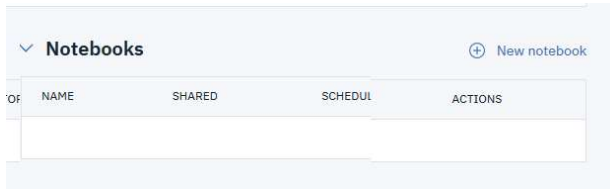
NAME	SERVICE TYPE
spark-yc	Spark

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

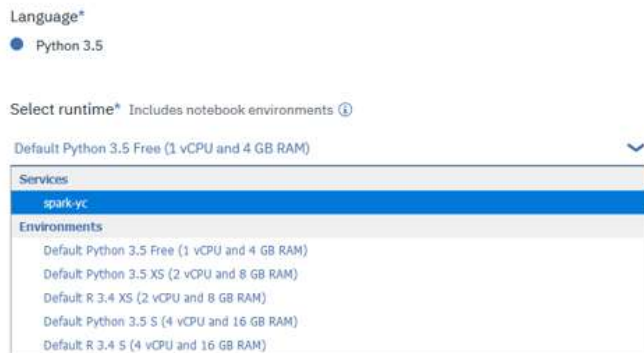
9



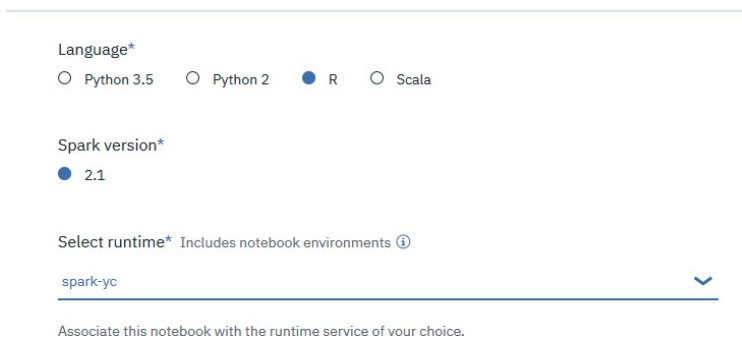
Vai nella sezione notebook e scegli di creare un new notebook:



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 scegliendo R:



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\*

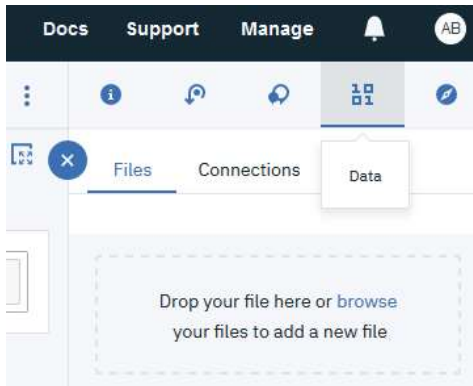
Notebookvini|

38 Characters  
Remaining

Procedi con **create notebook**.

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

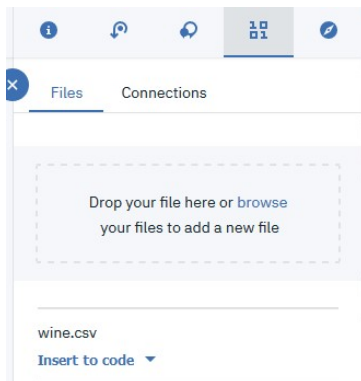
Si apre il notebook e compare la prima cella del notebook. La cella e' vuota al momento. Carichiamo il dataset wine.csv. Nel menu in alto a destra scegliete di aprire la sezione dati cliccando sull'icona data:



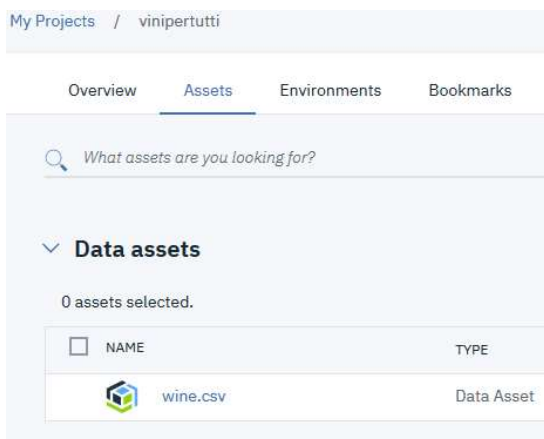
Seguendo le istruzioni, porta il file sull'object storage con drag and drop oppure facendo browse del disco. Il file puo essere scaricato da questo link:

<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 these credentials before you share your notebook.
Sys.setenv("AWS_ACCESS_KEY_ID" = "a142ee6e423149e3b40da6ac254fb0bc", "AWS_SECRET_ACCESS_KEY" = "ef51f0d23fafb383939a730e16279f93cea1e2e7e0da29f7")
url <- "s3.eu-gw-objectstorage.service.networklayer.com"
bucket <- "vinipertutti-donotdelete-pr-yla9khyfbc2z"
headers <- list('x-amz-content-sha256'="e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855")

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

Esegui il codice contenuto nel Notebook eseguendo le celle.

Esegui le celle completando con il codice mancante.

Una volta completato passa alla seconda parte del Laboratorio.