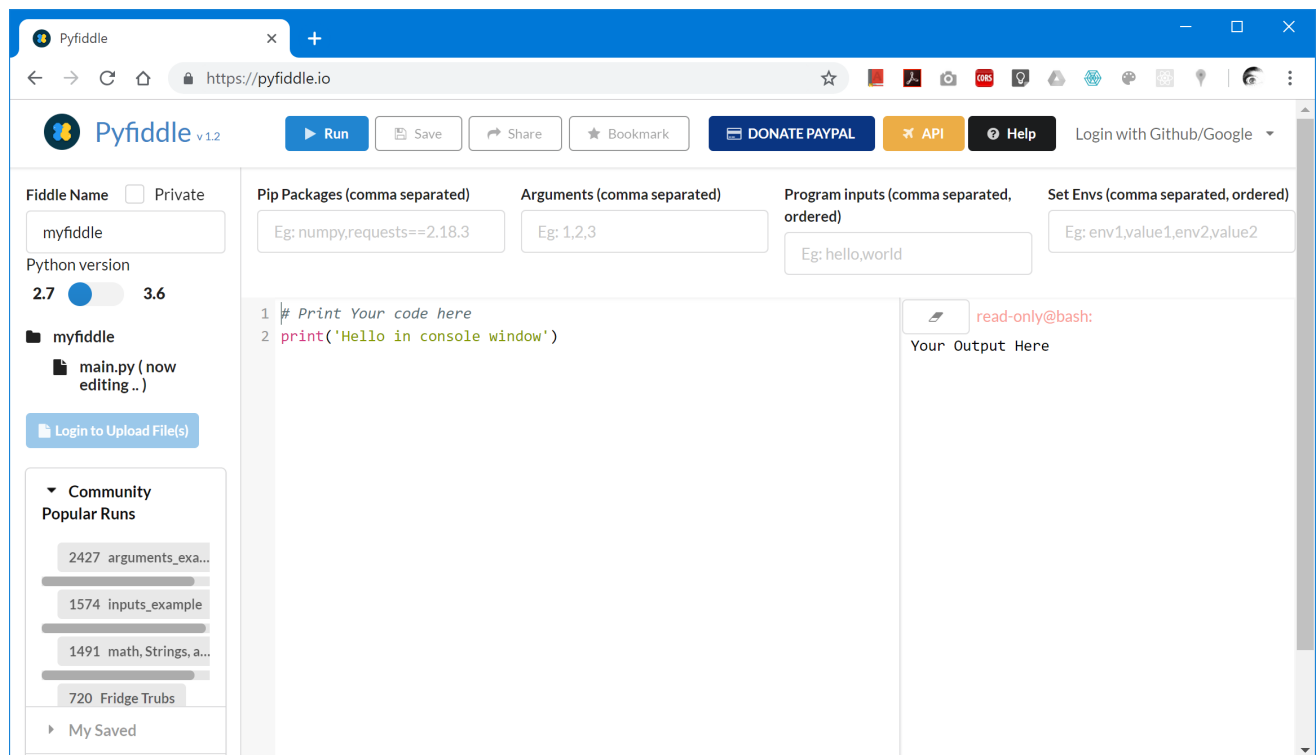


Operaciones básicas con conjuntos datos

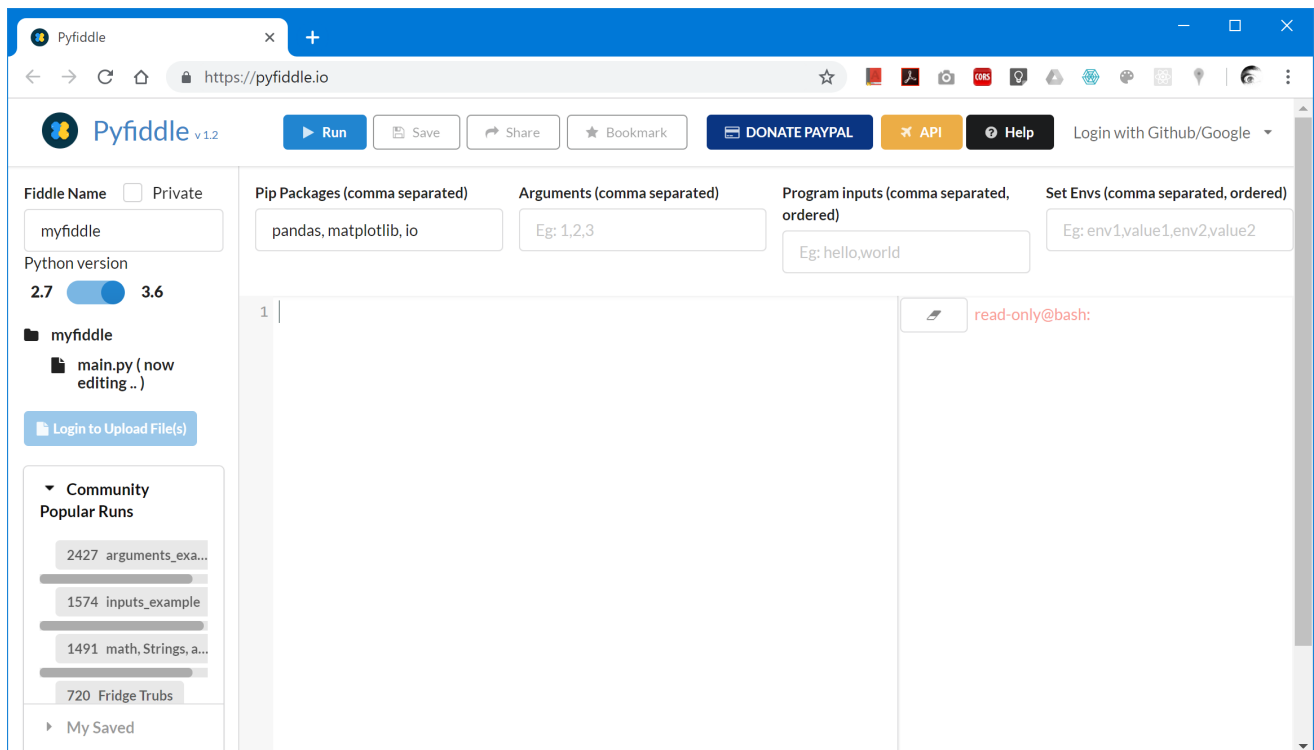
Preliminares

Esta actividad tiene como propósito mostrar algunos comandos útiles para describir tablas de datos. Para ello se empleará un editor en línea de código Python llamado [Pyfiddle](https://pyfiddle.io).

Al ingresar a encontrará una ventana como la siguiente:



Antes de iniciar, asegúrese de seleccionar la versión de Python 3.6 y de escribir en el cuadro **Pip Packages (comma separated)** el listado de librerías que se necesitarán, es decir `pandas` `matplotlib` e `io`. Su ventana ahora debería verse de la siguiente forma:



La base de datos a emplear es facilitada por <https://vincentarelbundock.github.io/Rdatasets/doc/boot/catsM.html> y consiste en una tabla que registra el sexo, el peso y el peso del corazón de 97 gatos. Para cargarla basta ejecutar los siguientes comandos:

```
import pandas as pd
import matplotlib.pyplot as plt
import io
from urllib.request import urlopen

link = "https://vincentarelbundock.github.io/Rdatasets/csv/boot/catsM.csv"
f = urlopen(link)
myfile = f.read()
datos = pd.read_csv(io.StringIO(myfile.decode('utf-8')), header = 0, sep=",")
```

Una vez ejecutado este código usted debería poder ejecutar métodos sobre la base de datos llamada `datos`.

Actividad 1

Ejecute las siguientes instrucciones y detalle su funcionamiento.

```
print(type(datos))
print(datos.head())
print(datos.tail())
print(datos.index)
print(datos.columns)
print(datos.dtypes)
print(datos.shape)
print(datos.values)
print(datos.info())
print(datos.describe())
```

Actividad 2

Calcule todas las medidas descriptivas de resumen que pueda obtener de este conjunto de datos.

Adicionalmente, compare las variables **Bwt** (peso total del gato) contra **Hwt** (el peso del corazón).
Describa el comportamiento de estas variables en conjunto.