**Semana del 17 de mayo**

Detectar valores duplicados

Data[data.duplicated()]

Pasar a numérica una columna que, aunque tiene números aparecen como string

def try\_convert\_to int(x):

try:

return int(x)

except:

print(‘rare value:’, x)

return x

Admiración invertida

len[data[‘horsepower’] == ‘?’]

Hay que tomar una decisión sobre cómo actuar con esos valores raros

Una posibilidad es asignarles la media

data[data[‘horsepower’] ¡= ‘?’].horsepower.astype(int).mean()

data[‘horsepower’] .value\_counts (normalize = True)

asignar la media con Python a toda la columna

horspower\_mean = 0

count = 0

for pos, value in

import argpase

para ejecutar un archivo de Python pasándole los parámetros que queramos

librería que recoge cada uno de los valores que le vamos dando

copiamos la ruta absoluta que aparece.

Vamos a la terminal de Windows (pegamos con botón derecho). Se ejecuta el archivo pero por la terminal. Mientras se va ejecutando podemos seguir trabajando con visual studio code.

Ctrl+ C interrumpimos la ejecución

from my\_module import exponent

función exponent (x,y, v=None):

result = x\*\*y

return result \*v if v else result

import sys

aquí no hace falta especificar si es x, y, v. Siempre se recogen todos los valores como un string

**Martes 18 de mayo**

Instalar la librería streamlit y altair

Ploty.- librería para crear gráficas interactivas

Copiamos la ruta al fichero que aparece en el archivo app.py de la semana 6

St. Button lo coloco encima o debajo de la tabla según coloque el if button antes o después del if write

Importar la librería pandas\_profile

**Miércoles 19 de mayo**

APIs Flask

Endpoint consta de:

url + parámetro + valor

Flask crea un servidor que permite que tu ordenador sea una API

Las carpetas static y template han de estar siempre obligatoriamente. No pueden tener otro nombre. Si quiero utilizar algún archivo obtenido con flask en la html ha de ser guardado en la carpeta static

Ipconfig para ver mi IP privada

Cual es mi IP: para ver la IP del wifi en mi casa o donde esté conectada

Flask es un framework (paquete de librerías)

Fondo de pantalla

En el html hay primero un título, un body, form (formulario, para poder recoger los valores que el usuario selecciona)

host = ’0.0.0.0’ esto permite que nuestro ordenador sea público

port puede ir desde el 1000 hasta el 60.000 más o menos

API minimal.- Crea una página web en la que aparece un formulario con varias variables en las que se ha de responder accediendo a un desplegable de opciones de respuesta

Cuando lo rellenan cada usuario, devuelve un json al que podemos acceder con al api de nuestra página web creada.

All returned es la concatenación de todas las respuestas que se van dando al formulario

main flask.py.- hay una opción en el formulario en la que se puede escribir texto. Lo que se recoge en el submmit aparece luego para individualizar la interacción con el usuario.

Puerto público

Localhost6060/’nombre de la api a la que queremos entrar

[Greet form (aiconscience.ddns.net)](http://aiconscience.ddns.net:6060/)

Extensions de visual studio code.- poner los paréntesis en colorines indent-rainbow,

Api image.-

Se incluye un encabezado del cuestionario

Permite seleccionar un archivo .csv y se carga el número de filas y columnas en html

Y seleccionar una imagen y que se suba como jpg o png

Get, post, put y delete son las acciones que se pueden utilizar con una API (http)

Con streamlit realizamos dashboard

En la carpeta others hay una carpeta EDA. Mira el de archivo 2 dentro de esta carpeta

Fecha para entregar el proyecto: 02-06-2021

Presentación del proyecto los días 3 y 4 de junio de 2021. No más de 10 minutos

Se entregará en un .zip

Carpetas:

* Carpeta src que contenga todo el código del proyecto
* Carpeta documentación (documentos relacionados)
* Recursos: imágenes, …
* Reports.- todo lo creado: imágenes, html, pdf
* Data.- contiene los datos del proyecto
* Notebooks.-
* Src/Utils.- librerías creadas:
  + Folders.py
  + Visualización.py
  + Mining data.py
  + Dashboard.py
  + Api.py
  + otros
* Main
* Dashboard.-
* Src/ crear un token id

Trello. Metodología ágil. Se utiliza para gestionar tareas de un proyecto

En trello. com crearse un usuario

Poner tareas del c, b, a ponerles etiquetas e ir planificando y

To do, doing, done. Backlog y review son opcionales

Kanbam template

Tenemos que agregar a los profesores

**webscrapping**

permite extraer información de páginas webs y guardar la información como un csv

librería de Python beautifulsoup4

Entro en la página, botón derecho/inspeccionar

Aparece el código

Pantalla superior del código en la parte de arriba, el cuadradito con la flecha pequeña en una esquina nos pasa a la página y seleccionar aquella parte de la que queremos ver el código

El parse: ‘lxml’ permite leer html de forma más fácil

Hay que instalar la librería lxml con pip 3 pero luego no hace falta utilizar import porque la librería BeautifulSoup lo utiliza por dentro

Podemos obtener todo el html. Nos interesan partes especificas de la web

Seleccionamos producto-image. Obtenemos el código de cada cerveza

Cada elemento del producto se abre y cierra con o < y se cierra con </>

Selenium

Botón derecho sobre la página a la que se accede desde Python e inspeccionar. Aparece el código.

Mejor que copiar el código que aparece en la página es más seguro y estable en el tiempo recoger el class o el tag porque si el que crea la página cambia algo en ella es más fácil que cambie parte del código que el class o el tag

Se puede acceder a elementos específicos con sus etiquetas

Cuando accedo a imágenes me las descarga en el directorio en el que tengo selenium