

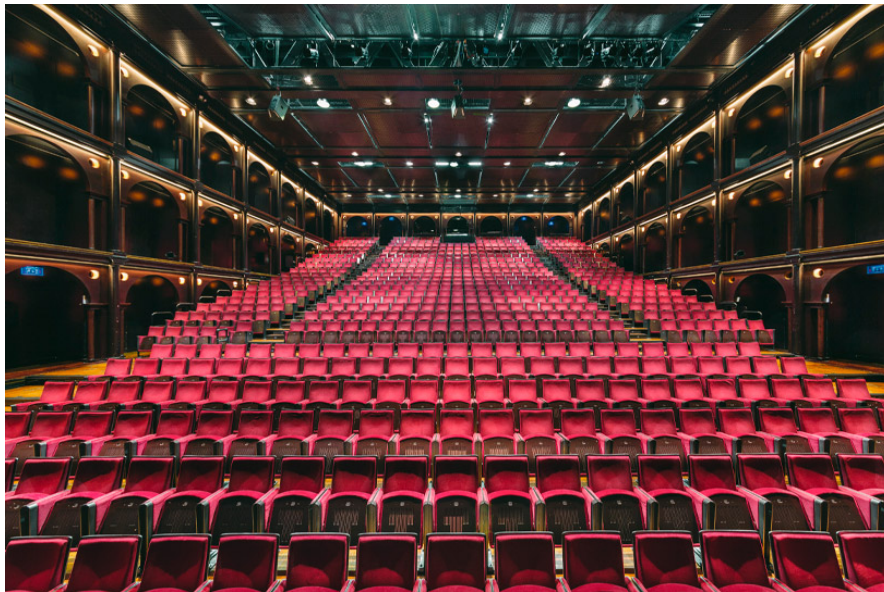
## **Dataset: “Obres de teatre a Barcelona, novembre 2023”**

Joana Llauradó i Maria Sopena

### **1. Context**

Aquesta pràctica està centrada en obtenir informació sobre els espectacles de teatre que tenen lloc a Barcelona aquest novembre de 2023. L'objectiu és generar un dataset amb informació rellevant de cada espectacle, com ara el lloc i el preu de l'entrada. La web seleccionada és Teatre Barcelona (<https://www.teatrebarcelona.com/>), una pàgina que conté tota la informació dels espectacles culturals a Barcelona i on es poden adquirir les entrades. L'objectiu d'aquest codi és que pugui ser reutilitzable i obtenir aquesta informació pels mesos següents.

### **2. Representació gràfica.**



**Figura 1. Sala Fabià Puigerver, Teatre Lliure Monjuïc, Barcelona** (Font: <https://www.teatrelliure.com/ca>)

### **3. Descripció del dataset**

El dataset conté informació sobre les obres de teatre que es realitzen a Barcelona aquest novembre del 2023. El contingut d'aquest dataset ens permetrà obtenir una descripció detallada de les característiques de les obres de teatre que ens interessin i així poder escollir un títol que s'adapti més als nostres interessos.

## 4. Contingut

Els camps que s'incoluen al dataset són els següents:

- **Nom de l'obra:** string que conté el nom de l'obra
- **Teatre:** string que conté la sala on es realitza l'obra
- **Preu de l'entrada:** cost de l'entrada per la web de Teatre Barcelona
- **Data d'inici:** dia que comença a estar en cartellera
- **Data de final:** dia que deixa d'estar en cartellera
- **Valoració:** percentage de valoració positiva de l'obra
- **Sinopsis:** Breu descripció de l'obra
- **Durada:** Llargada de l'obra en minuts
- **Idioma:** Si l'obra està en un únic idioma o en més d'un
- **Edat:** Quina és la recomanació d'edat del públic per l'obra de teatre

## 5. Propietari

Les dades han estat recopilades de la pàgina web Teatre Barcelona (<https://www.teatrebarcelona.com/>), propietat de ESCENES I PÚBLICS S.L. La propietat de les dels elements de la pàgina web i els drets d'autor pertanyen o han estat cedits a aquesta empresa, tal com s'indica a la web. L'objectiu fonamental d'aquesta plataforma és la venda d'entrades per a espectacles culturals a Barcelona, entre altres finalitats. No obstant això, el el propòsit d'aquesta pràctica és utilitzar la informació continguda en aquesta web per fer un anàlisi dels espectacles que es desenvoluparan a Barcelona.

Pel que fa als principis ètics i legals que hem seguit per aquest projecte, ens hem basat en informar-nos de la política de privacitat de la pàgina web. Per dur-ho a terme, hem consultat la secció d'avís legal del lloc (<https://www.teatrebarcelona.com/avis-legal>), on s'especifica la legalitat de l'ús de la informació continguda a la mateixa. El text rellevant diu el següent:

*“Es prohibeix la reproducció parcial o total del portal, i el seu tractament informàtic, així com la utilització comercial i explotació, en qualsevol de les seves formes del codi font, les imatges, els textos, la informació, els logotips o dissenys gràfics, les fotografies, les animacions, el software i qualsevol altre tipus de contingut escrit, audiovisual o sonor que es reculli a la pàgina Web. Sense perjudici de l'anterior, l'usuari pot accedir i visualitzar tots els elements, imprimir-los i emmagatzemar, sempre i quan sigui única i exclusivament, per al seu ús personal i privat.”*

Per tant, en el context d'aquesta pràctica, ja que no aspirem a realitzar un ús comercial o d'explotació, l'extracció d'informació d'aquesta web està permesa.

Finalment, volem esmentar que hi ha altres pàgines webs amb seccions culturals, com ara Time Out (<https://www.timeout.cat/>), que realitzen un anàlisi similar sobre les millors obres de teatre del mes. No obstant això, no donen més informació més ellà del nom de l'obra, el lloc i la sinopsi. En el nostre cas, volem anar més enllà i proporcionar informació addicional sobre cada obra, com ara el preu, la durada i el gènere.

## 6. Inspiració

La motivació d'escollir la pàgina web de Teatre Barcelona per dur a terme aquesta pràctica es fonamenta en la nostra passió compartida pel teatre. El que succeeix, però, és que l'oferta cultural a Barcelona és sumament àmplia, i sovint costa triar quina funció volem anar a veure. Amb el dataset generat mitjançant el web scraping, volem tenir la informació necessària per seleccionar una obra de teatre de manera fàcil, considerant diversos paràmetres com el preu o la durada. Algunes de les preguntes que ens plantejem a l'hora de fer aquesta elecció, i que, per tant, s'han de poder respondre el dataset, són les següents:

- Quin espectacle ha rebut les millors puntuacions?
- De què va l'obra de teatre?
- Quines obres estan destinades a tots els públics?
- Quina és la duració de cada obra?
- En quin idioma es representa l'obra de teatre?

Aquest conjunt de preguntes pretén guiar-nos a l'hora de seleccionar una obra que s'ajusti a les nostres preferències, proporcionant una visió detallada dels aspectes rellevants com la qualitat, la temàtica, l'accessibilitat per a tots els públics o la durada

## 7. Llicència

El nostre dataset estarà sota la llicència **"CC BY-NC-SA 4.0"**. Aquesta llicència permet l'ús i l'adaptació de les dades sempre que es reconegui l'autoria original. No obstant això, s'estableix la restricció de no permetre'n un ús comercial i de requerir que les obres derivades es distribueixin amb la mateixa llicència. La decisió de seleccionar aquesta llicència es fonamenta en què les dades ja son d'accés públic a la web de Teatre Barcelona, però aquesta plataforma s'explicita clarament que no es pot fer ús comercial de les dades. Per mantenir la coherència amb aquesta política i, al mateix temps, facilitar l'ús i la col·laboració amb les dades, la llicència "CC BY-NC-SA 4.0" s'ha considerat la més adequada per als nostres propòsits.

## 8. Dataset

El dataset `teatre_novembre_2023.csv` es troba al gihub dins la carpeta de data i també al Zenodo amb DOI: 10.5281/zenodo.10123665

## 9. Codi

El codi es troba partit en tres fitxer `.py` per tal de que l'usuari no toqui el codi i sigui més fàcil d'utilitzar. També ens és útil per a que el codi principal sigui mes net i fàcil de llegir.

- En el fitxer `paramters.py` trobem els parameters que l'usuari pot escollir per tal de realitar una recerca personalitzada. En primera instància es va provar de fer usant chorme drive ja que la pagina utilitza java per fer aquesta busca avançada, per incompatibilitat de versions s'ha descartat aquesa opció. Per tal de personificar la recerca es construeix la url cada slash implica un paràmetre, i es pot reesciure.
- En el fitxer `functions_scrpaing.py` obtenim la url i parsejem l'html. Per accedir a les url fem servir la llibreria `request`. Un cop obtenim l'html es tractat d'anar mirant on es troben els elements que ens interessin. Hem fet servir web tools developers de Google chrome, per trobar el nom de les classes en que es trobava el text que ens

interessa. Per cada informació trobem una capsula/caixa i, per tant, una funció per accedir a cada classe de l'html i poder-ne parsejar la informació. En la majoria dels casos hi han petites variacions pel que tenir-ho separat en funcions també ho facilita. Per exemple, en el cas del preu el nom de les classes varia una mica en funció de si hi ha o no descompte aplicat.

- En el fitxer main.py cridem les funcions i anem cosntruint el dataset final. Per utiltzar-lo escrivim la següent comanda a la terminal: `python main.py`

El document source/requirements.txt del github conté totes les llibreries i versions emprades en aquesta pràctica.

## 10. Vídeo

El vídeo de la pràctica és troba al següent link:

[https://drive.google.com/drive/folders/1QEgzVbrwCcX2uf\\_j7vpl5f2R96bMzIkh?usp=drive\\_link](https://drive.google.com/drive/folders/1QEgzVbrwCcX2uf_j7vpl5f2R96bMzIkh?usp=drive_link)