

# MEMORIA PROYECTO ESTEBAN ANDUEZA

2 de abril de 2024

## EXPOSICIONES TEMPORALES EN BARCELONA

### INTRODUCCIÓN

Hasta la fecha es difícil que en el sector de la cultura se presenten datos cuantitativos sobre su desempeño. Por múltiples razones intrínsecas al propio sector (variables difíciles de cuantificar, maneras de estructuración, agrupación y catalogación de las diferentes formas de producción cultural, gran intervención de instituciones politizadas y poco preparadas, falta de conciencia de "industria", etc.), conseguir una base de datos estructurada y completa es prácticamente imposible. Solo los sectores como el cinematográfico y el editorial, que están claramente catalogados como "industria", son capaces de llevar un resumen de su actividad traducida en datos. En el resto de sectores, en el caso de que se recoja algún tipo de información de la actividad, es mayoritariamente en base a datos categóricos, y raramente numéricos.

### OBJETIVOS

El propósito general de este trabajo es valorar en qué medida la Ciencia de Datos y los sistemas de Machine Learning son válidos y efectivos para ser aplicados en el sector de la cultura, la creación artística y la producción cultural.

Los objetivos específicos del trabajo son muchos y variados, pero se podrían resumir en:

- Experimentar diferentes formas de trabajo en bases de datos cuyas variables sean mayoritariamente categóricas.
- Aplicar modelos de Machine Learning en bases de datos procedentes del sector cultural.
- Detectar deficiencias y carencias en la forma de recopilar los datos en el ámbito de la cultura con el fin de proponer mejoras y sistemas adecuados a las instituciones responsables de cuantificar la producción cultural en este país.

### CONJUNTO DE DATOS

Para la realización de este trabajo se ha contado con los datos que presenta el Ayuntamiento de Barcelona sobre la programación de Exposiciones Temporales en la ciudad. El Ayuntamiento presenta un informe anual en el que recoge las visitas realizadas a cada exposición temporal, y de cada registro aporta información complementaria como la Sede o Centro que acoge la exposición, la temática, el régimen de producción, etc.

Los datos han sido obtenidos de la web del Ayuntamiento de Barcelona:  
<https://opendata-ajuntament.barcelona.cat/es>

Con el fin de tener una muestra representativa del número de exposiciones y significativa en el tiempo se han seleccionado los datos de 13 años consecutivos, desde el año 2010 hasta 2022 inclusive.

El data dataframe desde el que se inicia el estudio consta de 2.180 registros en 21 variables, que son las siguientes:

#	Column	Dtype	Significado y contenido
---	-----	-----	-----
0	Any	int64	Año de la toma del dato
1	Nom_Equipament	object	Supraentidad organizadora
2	Seu	object	Sede que acoge la exposición
3	TitolExposicio	object	Título identificativo de la expo
4	Visites	float64	Número de visitas
5	DataInici	object	Fecha de inicio de la exposición
6	DataFi	object	Fecha de clausura de la exposición
7	TipusProduccio	object	Si es producida o no por la sede
8	Codi_Districte	int64	Código Distrito de la sede
9	Nom_Districte	object	Nombre Distrito de la sede
10	Codi_Barri	int64	Código Barrio de la sede
11	Nom_Barri	object	Nombre Barrio de la sede
12	Titularitat	object	Tipo de propiedad de la sede
13	Entitat	object	Entidad propietaria de la sede
14	Tipus_Equipament	object	Catalogación de la sede
15	Tipus_Us	object	Uso al que se destina la sede
16	Ambit	object	Tipología de la exposición
17	Latitud	float64	Latitud geográfica de la sede
18	Longitud	float64.	Longitud geográfica de la sede
19	Web	object	Web de la sede o de la exposición
20	Notes_Equipament	object	Notas de interés

## METODOLOGÍA

Para el desarrollo del trabajo se seguirá la siguiente metodología:

- 1.- Obtención y preparación del conjunto de datos a partir de los datos anuales publicados.
- 2.- Limpieza de datos: Tratamiento de valores Nan y corrección de extravagancias y errores que puedan aparecer en el conjunto de datos.
- 3.- EDA (exploratory data analysis): Visualizaciones y conclusiones sobre los datos a estudiar.
- 4.- Preprocesado de las variables en función de que sean numéricas o categóricas y de la distribución de datos y valores atípicos que presenten, con el fin de prepararlas para su modelización.

5.- Modelización. Ensayo con modelos de Machine Learning no supervisado: Agrupación mediante Clústering con Kmeans y Clústering Jerárquico.

6.- Comparación de los modelos de agrupación.

## **CONCLUSIONES QUE SE ESPERA OBTENER**

- Con los métodos aplicados se pretende llegar a concluir si la aplicación de métodos de Machine Learning no supervisados sobre los datos que brindan las instituciones culturales pueden llegar a aportar catalogaciones ocultas al sentido común que sirvan para emprender acciones específicas sobre el sector de la cultura.
- En última medida se pretende detectar carencias en la obtención y presentación de los datos, con el fin de corregir las deficiencias y aprender a crear conjuntos de datos que Sí permitan evaluaciones con los métodos y sistemas que nos brinda el Data Science.