

# **VISUALITZACIÓ DE DADES**

## **ENGINYERIA DE DADES**

### **PROJECTE DE VISUALITZACIÓ DE DADES**

#### **Funcionament**

Judit Chamorro Servent, Guillermo Marin Getino, Enric Martí Gòdia  
Bellaterra, Abril 2024

## 1. INTRODUCCIÓ

Aquesta activitat pretén ser un projecte el més real possible de Visualització de Dades. L'objectiu principal és treballar amb un o varis datasets reals, formular hipòtesis, fer els anàlisis necessaris que ajudin a verificar-les o refutar-les, i presentar els resultats mitjançant un sistema coherent de visualitzacions de dades.

Es tracta d'un treball en **grups d'entre 5 i 6 persones**, que treballareu i us coordinareu en equip, realitzant les tasques per a verificar o refutar les hipòtesis. El grup triarà el o els datasets amb els que vol treballar en la primera sessió de projecte i formularà les hipòtesis i les tasques a realitzar. Al final, el grup presentarà una visualització composta de les gràfiques necessàries basada en el/els data set/s escollits.

Pretenem que en el projecte treballem amb els conceptes i tècniques apreses a classe, entre les que volem destacar:

- *Data Massage & Transformation*
- *Data Reduction* (PCA, LDA, t-SNE)
- Ús de gràfiques exploratòries per a comprendre les dades y per explorar possibles hipòtesis (scatterplots, diagrames de línia, diagrames de barres, histogrames, violins, boxplots, regressions, etc.)
- Ús de gràfiques avançades, com gràfiques per múltiples variables i múltiples dimensions (multi-panels-facets, SPLOM, etc), mapes, xarxes, 3D, etc.
- Interactivitat i/o animació
- Tècniques de personalització de gràfiques

Apart d'aquests, tindreu que aplicar tècniques d'anàlisi y transformació de dades apreses a altres assignatures, com pas complementari a la presentació gràfica de la informació. Es valorarà el nombre d'aquestes tècniques que utilitzeu i l'ús coherent que en feu per a validar o refutar les hipòtesis.

Referències d'on podeu treure datasets:

- **Open Data BCN:** Servei de dades obertes de l'Ajuntament de Barcelona (<https://opendata-ajuntament.barcelona.cat/es>)
- **Our World in Data:** Banc de dades amb activitats humanes (<https://ourworldindata.org/>)
- **Kaggle:** Banc de dades obert per usuaris (<https://www.kaggle.com/datasets>)

A continuació us expliquem més concretament el funcionament, horaris i avaluació.

## 2. FUNCIONAMENT

Com hem dit, el projecte consisteix en:

- Verificar o refutar una o varies hipòtesis sobre un tema a partir d'un o varis data sets d'elecció lliure entre les referències que us donem o, previ acord amb els professors, elegit per vosaltres mateixos d'altres fonts.
- Presentar els resultats mitjançant un sistema coherent de visualitzacions de dades. El format final (web, Tableau workbook, pòster estàtic, article) serà d'elecció lliure segons les característiques de cada projecte i es consensuarà amb els professors durant les sessions de projecte.

Per a realitzar el projecte cal formar grups de treball de 5 o 6 persones que us heu d'apuntar en una activitat d'inscripció de grups de projecte en el Campus Virtual (<http://cv.uab.cat>). Els grups han d'estar formats quan arribeu a la primera sessió de projecte

En la primera sessió amb el professor, aquest us explicarà breument en què consisteix aquesta activitat i el grup començarà a treballar en:

- Cercar el o els datasets amb els que voleu treballar, a partir de les referències donades o d'altres que pugueu trobar. En aquest darrer cas, cal consultar al professor.
- Examinar el data set i pensar en hipòtesis a verificar o refutar.
- Al final d'aquesta sessió s'hauria d'haver fet una feina de 'Brain Storming' i 'Conceptualització': Establir un punt d'inici per contextualitzar la informació que es voldrà presentar i tenir una idea del punt final /objectiu/conclusió a la que es vol arribar.
- Definir un conjunt de tasques a realitzar i assignar-les als membres del grup per a que reportin la feina feta a la següent sessió de projecte.

Al final de cada sessió cal pujar a l'aula virtual una acta de la reunió que reporti els temes i discussions que fa el grup durant la sessió i els acords a que s'arriba per a la següent sessió.

Les tasques i gràfiques es poden fer en R o Tableau. Els alumnes també poden decidir fer-ho en un altre llenguatge però tenint en compte que en aquest cas els professors no podran ajudar-los amb problemes de tipus tècnic.

En les sessions de projecte la feina del professor és:

- Fer un seguiment del treball que va realitzant cada grup per evitar que es desviï excessivament dels objectius inicialment plantejats en el problema o projecte,
- Atendre dubtes que el grup es faci procurant no limitar la iniciativa del grup.
- Observar el comportament del grup, tant a nivell individual com col·lectiu.

## 2.1 Horaris de les sessions de projecte.

Hi ha un únic horari que es mostra a la *Taula 1*.

SESSIÓ	Dia	Hora	Aula
1	19 Abril	Divendres, 13:00 a 15:00	Q5/1012
2	10 Maig	Divendres, 13:00 a 15:00	Q5/1012
3	21 Maig	Dimarts, 15:00 a 17:00	Q5/1012
4 (Examen Individual)	24 Maig	Divendres, 13:00 a 15:00	Q5/1012
5	28 Maig	Dimarts, 15:00 a 17:00	Q5/1012
6 (Presentació Projecte)	31 Maig	Divendres, 17:00 a 19:00	Sala de Graus

*Taula 1. Horaris de les sessions de projecte.*

**L'assistència a les sessions de projecte tutelades per professor és OBLIGATÒRIA per a poder fer un bon seguiment del treball que feu per part del professorat i que aquest us pugui respondre dubtes sobre el projecte. S'estudiaran casos d'absència degudament justificats.**

## 2.2. Entregues.

Hi hauran dues entregues de documentació mitjançant el Campus Virtual per a fer un seguiment i avaluació del projecte:

1. **Informe Inicial (3 Maig):** Plantejament (motivació), objectius o hipòtesis del projecte, explicació del datasets que treballareu, tasques a realitzar amb una previsió de durada i feina realitzada fins el moment. **Es recomana fer aquesta entrega per a tenir feedback dels professors i per tant, millorar el resultat final del projecte. L'informe inicial tindrà un màxim de 4 pàgines més els annexes.**

## 2. Entrega Final. Documentació final del projecte en el Campus Virtual:

- **ARTICLE (28 Maig) :** Resum del treball realitzat i firmat per tot el grup. L'article contindrà: a) Una introducció explicant les hipòtesis plantejades i els objectius del treball, b) Explicació dels datasets utilitzats, c) Resum de la feina de neteja, formatat, massatge i anàlisi de les dades, més una selecció de les gràfiques exploratòries utilitzades abans d'arribar a les gràfiques finals, d) Justificació de la visualització final. E) Conclusions i possibles millores del vostre treball, així com bibliografia i referències utilitzades. Es pot afegir el codi utilitzat per a generar les gràfiques en annexes. **L'article tindrà un màxim de 10 pàgines, mes els annexes.**
- **PRESENTACIÓ (31 Maig):** Breu presentació (10 minuts) en Power Point amb transparències on s'explica el treball realitzat, amb seccions semblants a la memòria.

En el Campus Virtual s'ofereixen documents patró en format Word (actes, informe inicial i memòria) i Power Point (presentació) per a guiar en els continguts dels documents i unificar estils de presentació. **El seu ús és optatiu.**

La darrera sessió es farà una classe presencial a **TOTS** els membres del grup (sempre que sigui possible per les restriccions sanitàries) on cada grup farà una presentació del seu treball en un màxim de 10 minuts davant d'un tribunal format per tres professors del grau d'Enginyeria de Dades. Després de cada presentació hi haurà un torn de preguntes per part del tribunal de com a màxim 5 minuts.

## 3. AVALUACIÓ

L'avaluació es basa en dues notes: Nota Grupal i Nota Individual.

La nota grupal té dos indicadors:

- Nota de l'Informe Inicial (**Prj\_1**) a partir del feedback donat pels professors.
- Nota de l'Entrega final (**Prj\_2**). Es valora la feina realitzada (el grau de dificultat de les hipòtesis proposades així com la forma i la qualitat de la solució), la documentació presentada (actes, informe inicial, article, transparències) i l'exposició oral. La nota és grupal i serà la mateixa a tots els membres. En funció de les observacions en les sessions de projectes (positives o negatives) es podrà modificar individualment la nota grupal d'algun membre del grup. La nota és sobre 10.

La nota individual (**Prj\_Exa**) serà d'un examen que es realitzarà la sessió anterior a la presentació oral. Serà un examen pràctic per a valorar els coneixements i expertesa adquirides en el projecte. La nota és sobre 10.

La nota de l'assignatura es definirà segons el grau de dificultat de les hipòtesis que el grup s'hagi proposat així com la forma i la qualitat de la solució. La nota es definirà de la següent forma:

$$\text{Nota Projecte} = \text{Prj\_2} * 0,8 + \text{Prj\_Exa} * 0,2 \quad [\text{Prj\_1}, \text{Prj\_2} \geq 5]$$

Per a no perjudicar en excés l'informe inicial, es valorarà el màxim entre la ponderació dels dos indicadors grupals (0,3 i 0,7) i la nota de l'entrega final.

## 4. DEDICACIÓ DE L'ALUMNE

Per a la realització del projecte cal tenir en compte les hores de sessió i les de feina entre sessions.

En la taula 2 mostrem les hores de les sessions de projecte (8 hores tret de l'examen individual) en les que el grup treballa cooperativament per a discutir el treball fet i el treball a fer, arribant a acords i feines per a la següent sessió. Per a cada hora de treball, comptem una hora de treball per part de cada alumne. En cada sessió cal distribuir la feina a fer per a cada membre, pel que no tothom ha de fer el mateix. Això fa que el nombre d'hores dedicades al projecte augmenti considerablement.

# alumnes	Total hores treball alumnes en projecte
5	8h sessions + 8h x 5 persones + (5h article i presentació) = 53 hores
6	8h sessions + 8h x 6 persones + (6h article i presentació) = 62 hores

*Taula 2. Esforç en hores per un grup de 5 o 6 persones en el projecte.*

Tal com es mostra en la taula 2, un grup de 5 persones disposa de 41 hores de feina per a realitzar el projecte. Un grup de 6 de 53 hores i un de 7 71 hores. Aquest temps inclou l'assistència a les sessions de projecte (4 sessions efectives de treball de 2h.), la feina pròpia (8 h per persona, una per a hora de sessió) i la preparació de la memòria i presentació (1 h per persona). Tots els valors són aproximats. El que queda clar és que el fet que hi hagi una persona més en un grup ha de repercutir en la quantitat i qualitat de feina que realitzi el grup.

**Bellaterra, Abril 2024**