

## Prova Data Science Hackató Saló Ocupació

### Descripció del projecte

En aquesta prova tècnica disposaràs de tres conjunts de dades:

1. Detall de les subvencions atorgades per l'Ajuntament de Barcelona (exceptuant les subvencions gestionades per Barcelona Activa), l'Institut de Cultura de Barcelona, l'Institut Municipal Barcelona Esports, l'Institut Municipal d'Educació, l'Institut Municipal de Mercats, l'Institut Municipal de Persones amb Discapacitat, l'Institut Municipal de Serveis Socials, l'Institut Municipal del Paisatge Urbà i la Qualitat de Vida i l'Institut Municipal d'Hisenda, per a les quals s'han realitzat pagaments de l'any 2019 i següents: [Subvencions atorgades per l'Ajuntament de Barcelona](#)
2. Enquesta sobre la bretxa digital a la ciutat de Barcelona: [Enquesta sobre la bretxa digital a la ciutat de Barcelona - Conjunts de dades - Open Data Barcelona](#)

Per a més informació, es recomana accedir al següent enllaç:

- Qüestionari: [20201008 Annex I Questionari-def.pdf \(barcelona.cat\)](#)
- Variables i valors: [Informacio de variables i valors.pdf \(barcelona.cat\)](#)

3. Accidents segons la causa gestionats per la Guàrdia Urbana a la ciutat de Barcelona: [Accidents segons causa conductor gestionats per la Guàrdia Urbana a la ciutat de Barcelona](#)

La teva missió en aquesta prova és:

1. Seleccionar un conjunt de dades, realitzar una neteja exhaustiva i duu a terme un EDA (Anàlisi Exploratòria de Dades). Es requereix que realitzis almenys dues visualitzacions que descriuen la informació continguda en les teves dades, destacant insights clau i tendències observades.
2. Implementa un pipeline que integri tant el preprocés de les dades com el modelatge. S'haurà d'avaluar almenys dos models diferents per determinar quin ofereix el millor rendiment en funció de l'estudi de les seves mètriques. Recorda avaluar si és pertinent realitzar un ajust fi dels hiperparàmetres dels models. Hauràs de justificar les teves decisions.

## Avaluació

- Correcta depuració de les dades. Hauràs de fer l'estandardització i/o normalització de les variables que ho requereixin.
- Correcta documentació del procés. S'avaluarà la llegibilitat, organització i bones pràctiques de programació.
- Correcta interpretació dels resultats.

## Lliurament

Has d'entregar la prova tècnica a través d'un **enllaç al teu repositori de GitHub** (com a resposta a aquest mail). En el repositori, has d'incloure:

- Un arxiu Jupyter Notebook amb totes les línies de codi executades i les interpretacions que consideris necessàries.
- Un arxiu README amb la següent informació:
  1. Introducció: Presentació del conjunt de dades i de les variables seleccionades.
  2. Depuració de dades: Descripció detallada de les tècniques de preprocessat aplicades i els criteris d'avaluació utilitzats.
  3. Resultats: Presentació dels resultats obtinguts.
  4. Conclusions: Principals inferències derivades dels resultats aconseguits.

Molta sort! Esperem veure el teu projecte!