

PROJECTE DE TEORIA II

Base de Dades 24/25

Andres Rio, Laia Beni, Aleix Falgosa, Francesc Navarro

03/05/2025

Taula de continguts

1 Introducció.....	3
2 Objectius.....	3
3 Model ER.....	4
3.1 Creació.....	4
3.2 Diagrama.....	6
3.3 Justificació.....	7
4 Característiques de les Dades.....	7
5 Implementació.....	8
6 Anàlisi de les dades.....	10
7 Conclusions.....	19
1 Apèndix 1: Planificació pràctica.....	20
1.1 Tasques planificades.....	20

1 Introducció

El projecte consisteix a aprendre a treballar amb una base de dades d'un tema seleccionat per nosaltres, per generar noves conclusions a partir de les relacions entre les dades.

Primerament, hem reflexionat sobre el tema a escollir, ja que volem tractar una qüestió amb un impacte global, com el canvi climàtic per culpa del trànsit aeri, l'impacte de la intel·ligència artificial en el món laboral o els desastres naturals. Després de cercar informació i comparar diferents opcions, vam decidir centrar-nos en els desastres naturals i el seu impacte social, un tema d'actualitat i de gran rellevància, ja que afecta milions de persones arreu del món.

Aquesta elecció va estar influenciada per una notícia que ens va cridar l'atenció, *la previsió d'un possible gran terratrèmol al Japó que passa cada 30 anys, conegut com The Big One, que podria assolir una magnitud de 9 i causar greus conseqüències*. Aquest tipus d'esdeveniments posa en evidència la importància de comptar amb bases de dades exhaustives sobre desastres naturals, ja que ens permeten analitzar patrons històrics, avaluar els seus efectes econòmics/socials i millorar les estratègies de prevenció i resposta, per minimitzar aquestes conseqüències.

2 Objectius

Com ja hem mencionat prèviament, centrarem aquest projecte en l'anàlisi de desastres naturals i el seu impacte social i econòmic. Utilitzarem les dades de la plataforma EM-DAT (emdat.be), una base de dades internacional que recopila informació sobre l'ocurrència i els efectes de més de 26.000 desastres a tot el món des de 1900 fins a l'actualitat. Proporciona informació detallada sobre víctimes, danys materials i pèrdues econòmiques, facilitant l'anàlisi de l'impacte social i econòmic.

L'objectiu principal és identificar possibles patrons temporals i espacials d'aquests desastres, analitzar la freqüència i gravetat segons la seva tipologia i avaluar la relació entre vulnerabilitat i factors geogràfics.

Per estructurar les dades, desenvoluparem un model ER i definirem una estratègia de recollida de dades per garantir la seva qualitat. Finalment, posarem en pràctica la teoria

vista a classe sobre l'àlgebra relacional per formular consultes i extreure informació rellevant.

3 Model ER

3.1 Creació

En el procés de disseny del model entitat-relació, s'ha buscat estructurar la informació de manera clara i eficient per facilitar-ne l'emmagatzematge i consulta. Per aconseguir-ho, s'han identificat les entitats principals i s'han definit els seus atributs i relacions, assegurant que cada desastre natural i el seu impacte estiguin correctament modelats. A continuació, es detallen les decisions preses en la construcció del model.

L'entitat *Desastre* és l'element central del model i representa cadascun dels desastres naturals registrats. La seva clau primària és DisNo., que identifica de manera única cada desastre. Els seus atributs inclouen Start Year, Start Month, Start Day, End Year, End Month, End Day, que permeten registrar la seva durada, i Associated Types, que permet associar tipus de desastres relacionats. Aquesta entitat es relaciona amb Ubicació, ja que un desastre pot ocórrer en una o diverses ubicacions (relació 1:N). També es relaciona amb País, ja que un país pot patir múltiples desastres, però cada desastre ocorre en un sol país (relació N:1). A més, cada desastre té una classificació definida per l'entitat Classificació (relació N:1), pot tenir una magnitud registrada en l'entitat Magnitud (relació 1:0..1) i causa un impacte humà i econòmic registrat a les respectives entitats (relacions 1:1).

L'entitat *País* modelitza la informació geogràfica general i la seva clau primària és DisNo., que identifica de manera única cada país. Els seus atributs són ISO, Country, Subregion, Region, que permeten categoritzar els països en zones geogràfiques més àmplies. Aquesta entitat es relaciona amb Desastre en una relació 1:N, ja que cada desastre es registra en un país. També es relaciona amb Ubicació en una relació 1:N, atès que cada ubicació pertany a un únic país.

L'entitat *Classificació* permet categoritzar els desastres segons la seva naturalesa i té com a clau primària DisNo. Els seus atributs són Classification Key, Disaster Group, Disaster Subgroup, Disaster Type, Disaster Subtype, que permeten una classificació jeràrquica dels desastres. Aquesta entitat es relaciona amb Desastre (relació 1:N), atès que cada desastre

Projecte de Teoria II

ha de tenir una classificació determinada. A més, es relaciona amb Magnitud (relació 1:N), ja que determinats tipus de desastres tenen associada una mesura de magnitud específica.

L'entitat *Ubicació* permet registrar la informació geogràfica precisa dels desastres i té com a clau primària DisNo. Els seus atributs són Location, Latitude, Longitude, River Basin, que defineixen la posició exacta on s'ha produït el desastre. Aquesta entitat es relaciona amb Desastre (relació N:N), ja que un desastre pot afectar diverses ubicacions i una ubicació pot veure's afectada per diversos desastres al llarg del temps. També es relaciona amb País (relació N:1), ja que cada ubicació pertany a un sol país.

L'entitat *Impacte Humà* recull les conseqüències d'un desastre sobre la població i té com a clau primària DisNo. , ja que cada desastre té exactament un únic registre d'impacte humà. Això permet identificar de manera única cada registre sense la necessitat de crear una clau composta o artificial. Els seus atributs inclouen Total Deaths, que comptabilitza les morts i desapareguts, No. Injured, que representa els ferits i hospitalitzats, No. Affected, que quantifica el total de persones afectades pel desastre, No. Homeless, que indica aquelles que han perdut la seva llar, i Total Affected, que és la suma de totes les persones afectades de diferents maneres. Aquesta entitat es relaciona amb Desastre en una relació 1:1, garantint que cada desastre tingui un únic registre d'impacte humà.

L'entitat *Impacte Econòmic* recull la informació sobre les pèrdues econòmiques causades pels desastres i també té com a clau primària DisNo., perquè cada desastre té un únic registre d'impacte econòmic. Com en el cas de l'impacte humà, això assegura que cada desastre estigui associat a un únic conjunt de dades econòmiques, evitant la creació de claus artificials innecessàries. Els seus atributs inclouen Total Damage ('000 US\$), que indica els danys econòmics totals causats pel desastre, Total Damage Adjusted ('000 US\$), que ajusta aquest valor tenint en compte la inflació per a una millor comparació històrica, Reconstruction Costs ('000 US\$), que representa els costos estimats de reconstrucció, Reconstruction Costs Adjusted ('000 US\$), que n'inclou l'ajust per inflació, Insured Damage ('000 US\$), que quantifica els danys coberts per assegurances, i Insured Damage Adjusted ('000 US\$), que també inclou ajustos per inflació. Aquesta entitat es relaciona amb Desastre en una relació 1:1, assegurant que cada desastre tingui exactament un registre d'impacte econòmic.

Els dos impactes depenen d'un desastre per existir, per això al diagrama són entitats febles.

L'entitat *Magnitud* permet registrar la intensitat d'un desastre i té com a clau primària DisNo. Els seus atributs són Magnitude ,Magnitude Scale, que especifica l'escala emprada per mesurar el desastre. Aquesta entitat es relaciona amb Desastre en una relació 1:0..1, ja que no tots els desastres tenen una mesura de magnitud registrada.

3.2 Diagrama

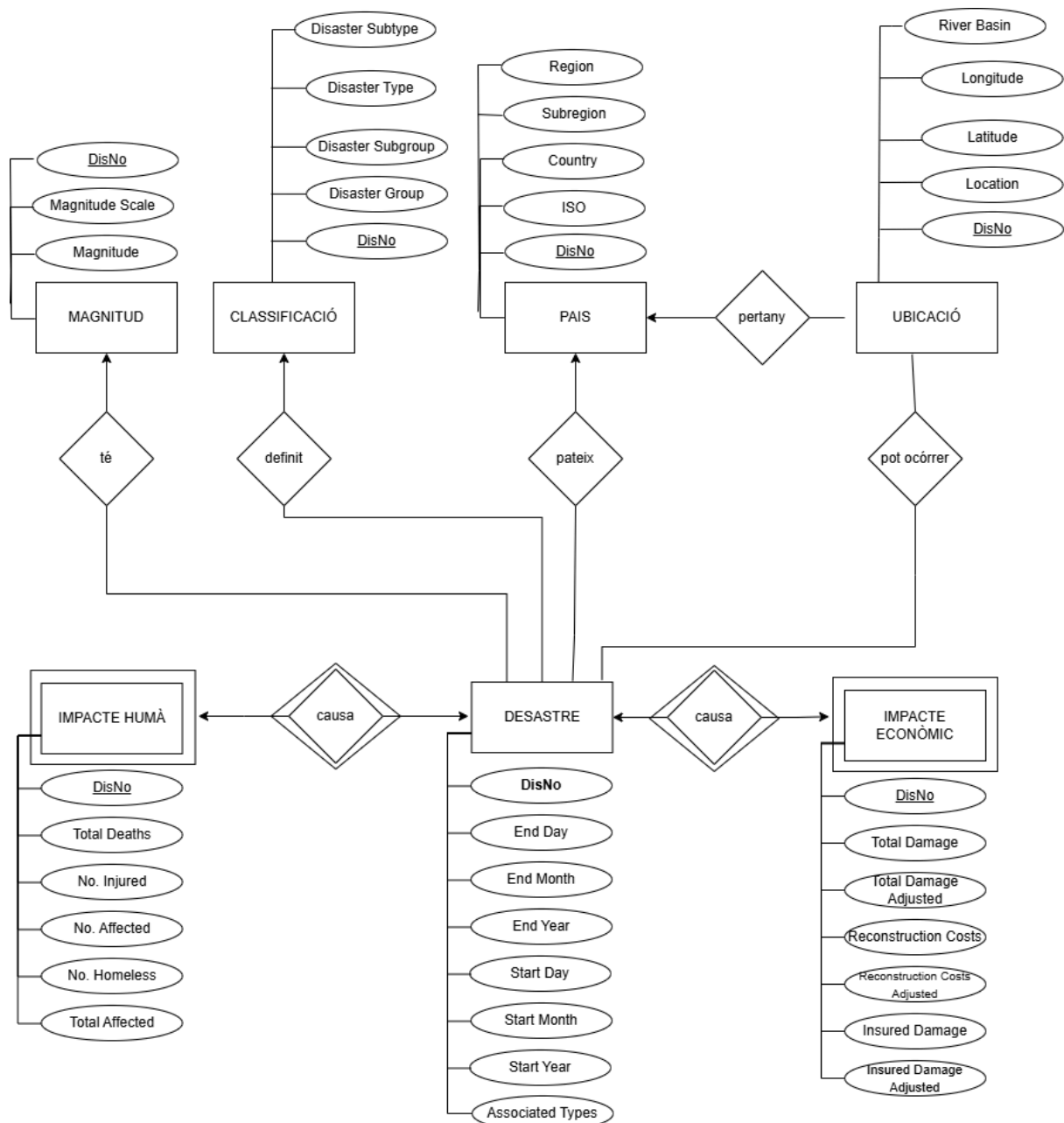


Fig 1. Model ER - Desastres naturals i el seu impacte

3.3 Justificació

Les entitats impacte (tant humana com econòmica) necessiten un desastre per tenir sentit, per això tenen la seva clau forana i són entitats dèbils. La cardinalitat seria de 1:1 perquè un desastre té un registre i cada registre pertany únicament a un desastre.

Desastre té 1 valor de magnitud, està classificat d'una manera, ocorre en un país (simplificació assumida: si el desastre afecta múltiples països, es consideraran com a desastres diferents per poder calcular l'impacte segons cada país) i pot ocórrer en diverses ubicacions.

Una magnitud pot pertànyer a diferents desastres, una classificació pot definir diversos desastres, un país pot patir diversos desastres igual que ubicació.

4 Característiques de les Dades

Les dades del projecte provenen de EM-DAT (Emergency Events Database), gestionada pel Centre de Recerca en Epidemiologia de Desastres (CRED) de la Universitat Catòlica de Lovaina. Aquesta base de dades recopila informació sobre desastres naturals i tecnològics des de 1900 fins a l'actualitat.

L'extracció de dades es va realitzar el 28 de març de 2025, amb una versió actualitzada del 25 de març de 2025, contenint 17.293 registres. Les dades s'obtenen de fonts fiables com les Nacions Unides (OCHA, OMS, OMM, FAO), governs (FEMA, NOAA, USGS), centres de recerca (DFO), organitzacions humanitàries (Creu Roja) i asseguradores (SwissRe, MunichRe, AON Benfield).

Cada registre inclou informació sobre el tipus de desastre, ubicació, dates, impacte humà i econòmic. La base de dades es pot exportar en Excel per facilitar l'anàlisi.

Tot i la seva fiabilitat, poden existir llacunes o biaixos en regions amb menys accés a dades. Malgrat això, EM-DAT és un referent mundial per a l'estudi dels desastres naturals i el seu impacte global.

5 Implementació

El primer pas en el nostre procés d'anàlisi va ser la creació de la base de dades, que vam anomenar *Desastres Naturals*. Aquesta tasca la vam dur a terme mitjançant un script SQL on, en primer lloc, vam definir l'esquema i vam crear les diferents taules que el componen. Un exemple de la creació d'una d'aquestes taules és la següent:

```
CREATE TABLE Desastre (  
    DisNo VARCHAR(20) PRIMARY KEY,  
    AssociatedTypes VARCHAR(255),  
    StartYear INT,  
    StartMonth INT,  
    StartDay INT,  
    EndYear INT,  
    EndMonth INT,  
    EndDay INT  
);
```

Les taules creades són: *Desastre*, *País*, *Classificació*, *Impacte Global*, *Impacte Econòmic*, *Magnitud* i *Ubicació*.

Un cop establerta l'estructura de la base de dades, vam procedir a carregar-hi les dades. Aquestes dades estaven inicialment recollides en un fitxer Excel amb tota la informació necessària. Per tal de facilitar el procés d'importació, vam separar el contingut en diverses fulles d'Excel, una per cada taula, incloent-hi únicament les columnes necessàries. Després, cada fulla va ser exportada individualment en format CSV.

Inicialment vam intentar fer la importació de les dades mitjançant l'eina *Table Data Import Wizard* de MySQL Workbench. Tot i això, ens vam trobar amb un obstacle important: els valors numèrics en format decimal als fitxers CSV utilitzaven la coma com a separador decimal en lloc del punt, cosa que dificultava la correcta interpretació dels floats.

Com a solució, vam intentar utilitzar *Visual Studio* per modificar els fitxers CSV mitjançant expressions regulars, substituint totes les comes per punts. Tot i això, aquesta estratègia no va donar els resultats esperats i vam continuar buscant alternatives amb l'ajut d'eines d'IA.

Projecte de Teoria II

Finalment, vam trobar una solució utilitzant la instrucció *LOAD DATA INFILE*, que ens permet controlar de manera més precisa el procés d'importació. Per a utilitzar-la vam passar tots els fitxers CSV a la carpeta *Uploads* de MySQL, des d'on vam extreure les rutes completes per a poder fer les importacions. Un exemple del codi implementat és el següent:

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Classificacio.csv'
INTO TABLE Classificacio
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\n'
IGNORE 1 LINES
(@DisNo, ClassificationKey, DisasterGroup, DisasterSubgroup, DisasterType, DisasterSubtype)
SET DisNo = NULLIF(TRIM(@DisNo), '');
```

Aquest fragment de codi carrega les dades des del fitxer *Classificacio.csv*, que es troba dins la carpeta de pujada de MySQL. S'especifica que els camps estan separats per punt i coma i que s'ha d'ignorar la primera línia. Les dades es llegeixen primer en variables temporals (@) per poder fer neteges, com eliminar espais innecessaris i tractar valors buits.

En el cas de la taula *Magnitud*, també vam necessitar substituir les comes per punts en els decimals. Per fer-ho, vam utilitzar el següent codi:

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Magnitud.csv'
INTO TABLE Magnitud
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\n'
IGNORE 1 LINES
(@DisNo, @Magnitude, MagnitudeScale)
SET
  DisNo = NULLIF(TRIM(@DisNo), ''),
  Magnitude = CASE
    WHEN TRIM(@Magnitude) = '' THEN NULL
    ELSE REPLACE(@Magnitude, ',', '.') + 0
  END;
```

Aquest codi aplica una transformació als valors del camp *Magnitude*, convertint les comes en punts i assegurant que, en cas de cel·les buides, el valor sigui interpretat com a *NULL*. Això ens va permetre importar correctament dades numèriques sense errors de format.

Un cop finalitzat aquest procés per a cada taula, vam tenir la base de dades completament funcional i preparada per iniciar les tasques d'anàlisi i consulta.

6 Anàlisi de les dades

Per fer l'anàlisi de les dades, s'ha decidit analitzar cada pregunta per separat. En cada pregunta, hem utilitzat una comanda SQL, per tal d'obtenir una taula amb els resultats desitjats per facilitar l'estudi de les dades. Posteriorment, hem exportat les taules resultants en fitxers.csv, per poder realitzar gràfics, de manera que fos més fàcil la seva anàlisi.

1. Quin ha estat l'augment en la freqüència dels desastres naturals en els últims 50 anys?

Per tal de donar resposta a la pregunta plantejada, hem formulat una consulta SQL específica que ens permet extreure la informació rellevant de la nostra base de dades de desastres naturals.

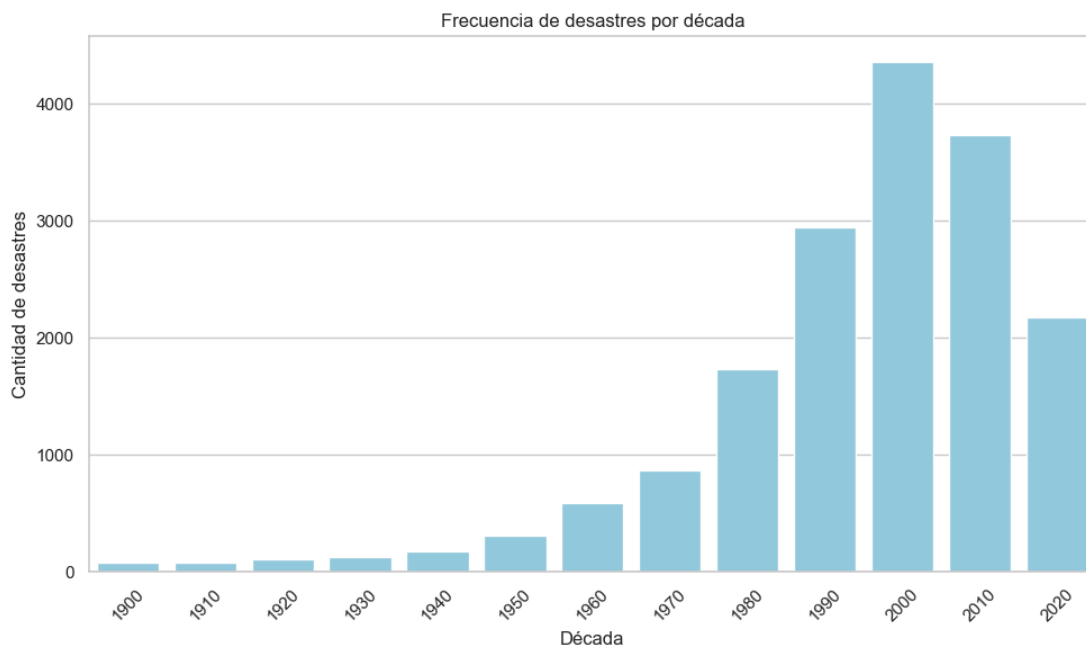
La consulta utilitzada és la següent:

```
5 • SELECT
6     FLOOR(StartYear / 10) * 10 AS Decade,
7     COUNT(DisNo) AS DisasterCount
8 FROM
9     Desastre
10 GROUP BY
11     Decade;
```

Figura 1: Consulta SQL

Projecte de Teoria II

Un cop executada la consulta, els resultats obtinguts s'han representat mitjançant un gràfic de barres per facilitar-ne la interpretació visual:



Com podem observar en el gràfic, es veu que hi ha un ascens en la freqüència dels desastres naturals al llarg de les últimes 5 dècades. Aquest és un increment constant, que podria estar relacionat amb diversos factors com el canvi climàtic, la urbanització accelerada de les ciutats o el fet que l'avanç tecnològic, ha fet que hi hagi hagut una millor capacitat de detecció i registre. El fet de veure aquesta evolució, ens demostra la necessitat d'incrementar les polítiques de prevenció i resposta, ja que el nombre de fenòmens va augmentant amb el temps.

2. Quins són els desastres més recurrents en zones de gran densitat de població?

Per tal de donar resposta a la pregunta plantejada, hem formulat una consulta SQL específica que ens permet extreure la informació rellevant de la nostra base de dades de desastres naturals.

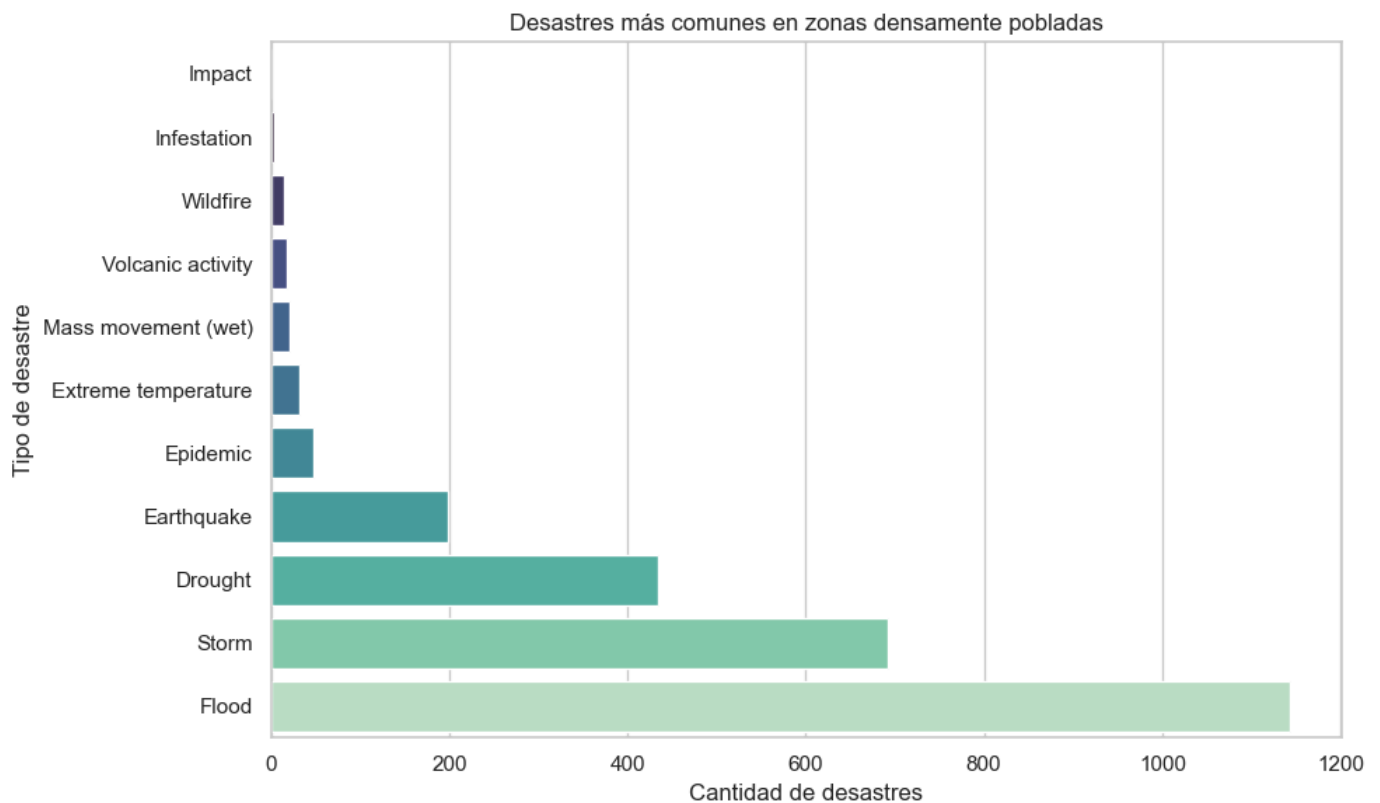
La consulta utilitzada és la següent:

```
15 • SELECT
16     c.DisasterType,
17     COUNT(d.DisNo) AS DisasterCount
18 FROM
19     ImpacteHuma ih
20 JOIN Desastre d ON ih.DisNo = d.DisNo
21 JOIN Classificacio c ON d.DisNo = c.DisNo
22 WHERE
23     ih.TotalAffected > 100000 -- ajusta aquest llindar segons el criteri
24 GROUP BY
25     c.DisasterType;
```

Figura 2: Consulta SQL

Projecte de Teoria II

Un cop executada la consulta, els resultats obtinguts s'han representat mitjançant un gràfic de barres per facilitar-ne la interpretació visual:



El gràfic resultant indica que els desastres més habituals en zones densament poblades són les inundacions i les tempestes. Això és coherent amb la realitat de moltes ciutats que es troben en zones costaneres o riberenques, on les pluges intenses i els ciclons tropicals tenen un gran potencial destructiu. Aquest gràfic ens mostra la importància de planificar el creixement urbà tenint en compte les diferents vulnerabilitats que hi ha davant dels riscos naturals, així com s'ha de millorar les infraestructures de drenatge i protecció civil.

3. Com ha canviat la distribució geogràfica dels desastres en el darrer segle?

Per tal de donar resposta a la pregunta plantejada, hem formulat una consulta SQL específica que ens permet extreure la informació rellevant de la nostra base de dades de desastres naturals.

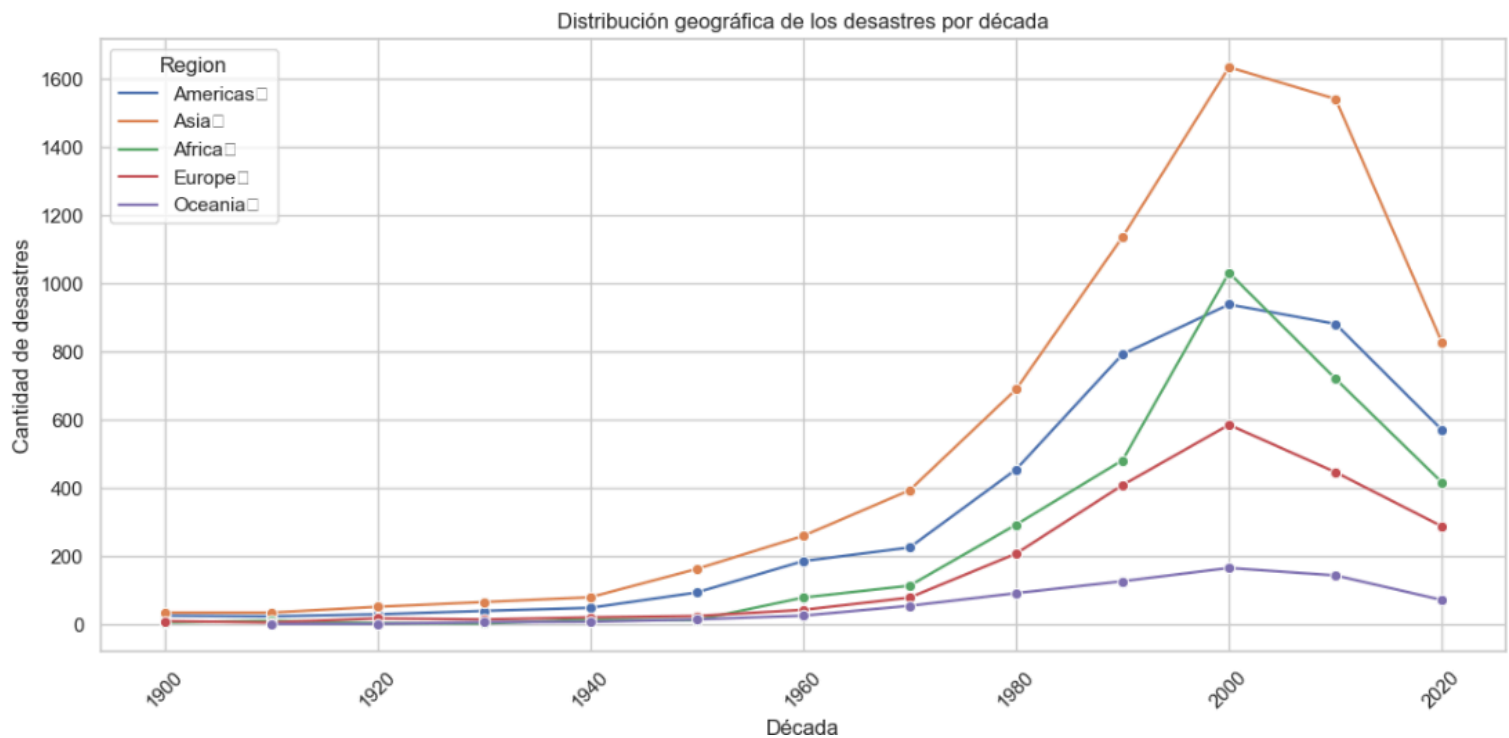
La consulta utilitzada és la següent:

```
29 • SELECT
30     FLOOR(d.StartYear / 10) * 10 AS Decade,
31     p.Region,
32     COUNT(d.DisNo) AS DisasterCount
33 FROM
34     desastre d
35     JOIN pais p ON d.DisNo = p.DisNo
36 GROUP BY
37     Decade, p.Region;
38
```

Figura 3: Consulta SQL

Un cop executada la consulta, els resultats obtinguts s'han representat mitjançant un gràfic de punts per facilitar-ne la interpretació visual:

Projecte de Teoria II



Segons el gràfic obtingut, es veu com els desastres naturals han anat variant la seva localització al llarg del segle XX i inicis del XXI. Regions com Àsia i Amèrica presenten un augment notable del nombre d'esdeveniments registrats, mentre que altres àrees, com Oceania o Europa, presenten una variabilitat menor. Aquest canvi pot estar marcat per la combinació de factors ambientals, socioeconòmics i demogràfics, així com per una millora en els sistemes de recopilació de dades en determinades zones del planeta.

4. Com ha augmentat la magnitud dels desastres en els últims anys?

Per tal de donar resposta a la pregunta plantejada, hem formulat una consulta SQL específica que ens permet extreure la informació rellevant de la nostra base de dades de desastres naturals.

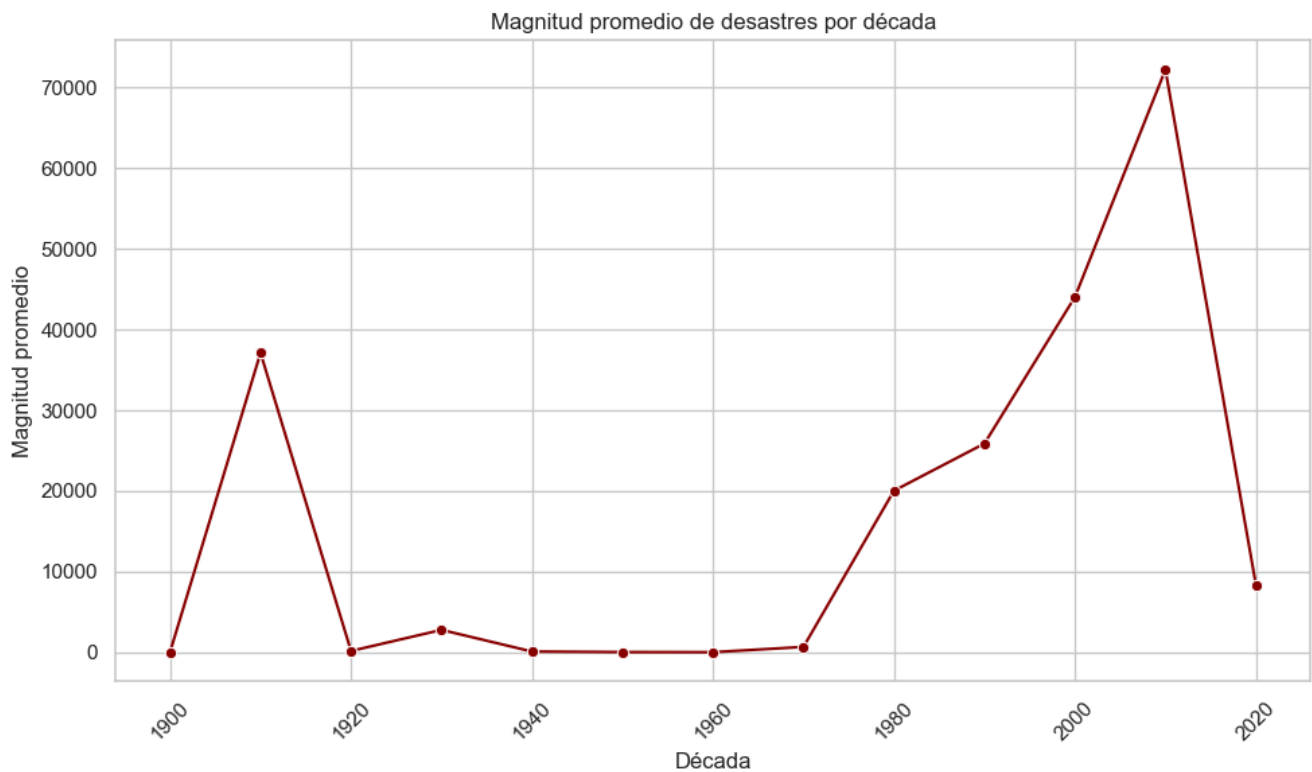
La consulta utilitzada és la següent:

```
42 ● SELECT
43     FLOOR(d.StartYear / 10) * 10 AS Decade,
44     AVG(m.Magnitude) AS AvgMagnitude
45 FROM
46     Desastre d
47     JOIN Magnitud m ON d.DisNo = m.DisNo
48 GROUP BY
49     Decade;
--
```

Figura 4: Consulta SQL

Projecte de Teoria II

Un cop executada la consulta, els resultats obtinguts s'han representat mitjançant un gràfic de punts per facilitar-ne la interpretació visual:



L'evolució de la magnitud mitjana dels desastres naturals per dècada mostra una tendència d'augment en els darrers anys. Aquest increment pot ser interpretat com una evidència del creixent impacte dels fenòmens extrems, possiblement agreujats pel canvi climàtic, com ens hem trobat casos de Danes extremes en els darrers anys. Per altra banda, cal considerar que la magnitud registrada pot estar condicionada per la disponibilitat de dades i les metodologies de mesura emprades en cada període. De manera que aquests dos factors, poden explicar aquest augment en la magnitud dels desastres naturals.

El gràfic obtingut demostra que els terratrèmols i les inundacions són els tipus de desastre que han generat més pèrdues econòmiques a escala global. Aquesta afectació econòmica varia segons el país, on podem observar valors especialment alts en estats amb alta densitat urbana i infraestructura vulnerable. Tot això ens fa veure que cal invertir en mesures que ens ajudin a fer front als desastres, i també que és important tenir bons sistemes d'assegurança i reconstrucció, per poder fer front a les conseqüències d'aquests desastres naturals.

7 Conclusions

En aquesta primera fase del projecte hem establert una base sòlida per analitzar l'impacte dels desastres naturals. Hem definit els objectius, seleccionat una base de dades fiable (EM-DAT) i creat un model entitat-relació que estructura la informació de manera clara i coherent.

També hem començat a formular consultes per identificar patrons i relacions rellevants, i hem valorat les característiques de les dades, tenint en compte tant el seu valor com les seves possibles limitacions. Tot plegat ens permet continuar amb l'anàlisi amb una estructura ben definida.

A partir de l'anàlisi de les dades, s'ha pogut observar que el nombre de desastres naturals ha augmentat de manera notable en els darrers cinquanta anys. Aquest increment sembla estar relacionat amb diversos factors, com ara el creixement de la urbanització, els efectes del canvi climàtic i la millora de les tecnologies de detecció i registre, que han permès identificar amb més precisió aquests esdeveniments.

A més, no només han augmentat en nombre, sinó també en magnitud, fet que pot indicar que els desastres són cada vegada més intensos. Davant d'aquesta realitat, és important aplicar polítiques més decidides per frenar el canvi climàtic i, al mateix temps, invertir en infraestructures més segures i preparades per afrontar situacions extremes. També cal tenir en compte la distribució geogràfica dels desastres i el seu impacte en països amb més densitat de població, per tal de prioritzar actuacions preventives i de resposta eficient.

1 Apèndix 1: Planificació pràctica

1.1 Tasques planificades

Describeix les tasques i quin/s integrants del grup són els encarregats.

- Tasca Búsqueda del tema i fonts d'informació – Noms: Aleix Falgosa, Francesc Navarro.
- Tasca Objectius – Noms: Laia Beni .
- Tasca Model ER – Noms: Andres Rio, Aleix Falgosa.
- Tasca Característiques de les dades – Noms: Andres Rio, Laia Beni.
- Tasca Avanç de consultes mitjançant algebra relacional – Noms: Aleix Falgosa, Francesc Navarro.
- Tasca Implementació de la Base de Dades - Noms: Aleix Falgosa, Andres Rio.
- Tasca Anàlisi de Dades - Noms: Francesc Navarro, Laia Beni.
- Conclusions - Noms: Aleix Falgosa, Francesc Navarro.