

DATA ACQUISITION PROJECT

Awarded Movies



Aymeric Behaegel
Alexandros Kouvatseas
Lucas Rousselet

INTRODUCTION

Focus on awarded movies

Data from famous ceremonies



PLAN

I - PRESENTATION OF THE DATA SOURCES

II - DIFFICULTIES

III - STRUCTURE OF THE FINAL BASE

I- PRESENTATION OF THE DATA SOURCES

- We selected various ceremonies, based on the variety of structures of the websites
- There are inherent variations between award titles, number of categories, number of nominees per categories, etc.
- Sources :
 - Oscars
 - BAFTA
 - BERLIN Golden Bear
 - Cannes Festival
 - Golden Globes

I- PRESENTATION OF THE DATA SOURCES

- Data we are interested in:
 - Ceremonies
 - Years
 - Award Titles
 - Winner
 - Associated Winner Item (Film/Person)
 - Nominees (optional)
 - Associated Nominees items (Film/Person) (optional)

II - DIFFICULTIES

II.1 - INHERENT DIFFICULTIES

- Variations and inconsistencies in awards titles among ceremonies: inconvenient for databases merging
- Inconsistencies within the ceremonies from years to years
- Awards bestowed to either a film or a person depending on its nature

I - PRESENTATION OF THE DATA SOURCES

The screenshot shows the official website for the Academy Awards (oscar.org). At the top, there is a navigation bar with links for AWARDS/INDUSTRY, MUSEUM/COLLECTION, SUPPORT AND SHOP, LEARN, and NEWS. Below this, a yellow button labeled "VIEW BY CATEGORY" is highlighted. A dropdown menu titled "SELECT A CATEGORY" is open, showing the option "ACTOR IN A LEADING ROLE". The main content area displays the nominees for the "Actor in a Leading Role" category. It includes a section for the WINNER (Cillian Murphy for Oppenheimer) and a list of NOMINEES: Bradley Cooper (Maestro), Colman Domingo (Rustin), Paul Giamatti (The Holdovers), and Jeffrey Wright (American Fiction).

ACTOR IN A LEADING ROLE

WINNER
CILLIAN MURPHY
Oppenheimer

NOMINEES
BRADLEY COOPER
Maestro

COLMAN DOMINGO
Rustin

PAUL GIAMATTI
The Holdovers

JEFFREY WRIGHT
American Fiction

ACTOR IN A SUPPORTING ROLE

WINNER
ROBERT DOWNEY JR.
Oppenheimer

NOMINEES
STERLING K. BROWN
American Fiction

ROBERT DE NIRO
Killers of the Flower Moon

RYAN GOSLING
Barbie

MARK RUFFALO
Poor Things

The screenshot shows the official website for the British Academy of Film and Television Arts (BAFTA). At the top, there is a navigation bar with links for BAFTA, Programmes, Membership, Awards, Events, and Stories. The main content area displays the "2024 RESULTS". It includes a section for the "ADAPTED SCREENPLAY" category, which has 5 nominees. The winner is "American Fiction" (Cord Jefferson), and the nominees are "All of Us Strangers" (Andrew Haigh) and "Oppenheimer" (Christopher Nolan). There are also sections for "ANIMATED FILM", "BEST FILM", and "BRITISH SHORT ANIMATION".

2024 RESULTS

ADAPTED SCREENPLAY (5) WINNER: AMERICAN FICTION

Winner
American Fiction
Cord Jefferson

Nominee
All of Us Strangers
Andrew Haigh

Nominee
Oppenheimer
Christopher Nolan

ANIMATED FILM (4) WINNER: THE BOY AND THE HERON

BEST FILM (5) WINNER: OPPENHEIMER

BRITISH SHORT ANIMATION (3) WINNER: CRAB DAY

Source : oscar.org & bafta.org

I - PRESENTATION OF THE DATA SOURCES

1966	<i>Un homme et une femme</i>		Claude Lelouch	France
	<i>Ces messieurs dames</i>	<i>Signore & signori</i>	Pietro Germi	Italie
1967	<i>Blow-Up</i>		Michelangelo Antonioni	Italie
1968	Arrêté en cours de déroulement à cause des événements de mai 68 ⁶¹			
1969	<i>If....</i>		Lindsay Anderson	Royaume-Uni
1970	<i>MASH</i>		Robert Altman	États-Unis
1971	<i>Le Messager</i>	<i>The Go-Between</i>	Joseph Losey	Royaume-Uni
1972	<i>La classe ouvrière va au paradis</i> ^{note 1,62}	<i>La classe operaia va in paradiso</i>	Elio Petri	Italie
	<i>L'Affaire Mattei</i> ^{note 1,62}	<i>Il caso Mattei</i>	Francesco Rosi	Italie
1973	<i>La Méprise</i>	<i>The Hireling</i>	Alan Bridges	Royaume-Uni
	<i>L'Épouvantail</i>	<i>Scarecrow</i>	Jerry Schatzberg	États-Unis
1974	<i>Conversation secrète</i>	<i>The Conversation</i>	Francis Ford Coppola	États-Unis

The screenshot shows the Berlinale Archive website's 'Awards & Juries' section. At the top, there are navigation links: Archive Overview, Yearbooks, Programme, Awards & Juries (which is underlined in red), Photos & Videos, and News & Topics. Below this, a large heading says 'Awards'. There is a search bar with a magnifying glass icon and a dropdown menu for 'Festival edition(s)'. To the right of the search bar, there are pagination controls showing '1 ... 38 >'. Below the search area, there are filters for 'Award', 'Section', 'Country', and a sorting option 'Sort: by festival edition [descending]'. The result count is '1488 results'. A specific award entry is highlighted: 'Golden Bear for Best Film (awarded to the film's producers) by Mati Diop'. The Berlinale bear logo is visible on the left side of the header.

Source : wikipedia.org & berlinale.de

II - DIFFICULTIES

II.1 - INHERENT DIFFICULTIES

List of discontinued Awards of Merit categories by year introduced,
sortable by category

Year introduced	Year discontinued	Category
1927/28	1927/28	Best Director, Comedy Picture
1927/28	1927/28	Best Director, Dramatic Picture
1927/28	1927/28	Best Engineering Effects
1927/28	1927/28	Best Title Writing
1927/28	1927/28	Best Unique and Artistic Production
1927/28	1956	Best Original Story
1931/32	1935	Best Short Subject – Comedy
1931/32	1935	Best Short Subject – Novelty
1932/33	1937	Best Assistant Director
1935	1937	Best Dance Direction
1936	1956	Best Short Subject – 1 Reel
1936	1956	Best Short Subject – 2 Reel
1936	1937	Best Short Subject – Color
1963	2019	Best Sound Editing
1995	1998	Best Original Musical or Comedy Score

List of current Awards of Merit categories by
year introduced, sortable by category

Year introduced	Category
1927/28	Best Picture
1927/28	Best Director
1927/28	Best Actor
1927/28	Best Actress
1927/28	Best Cinematography
1927/28	Best Production Design
1927/28	Best Adapted Screenplay
1929/30	Best Sound
1931/32	Best Animated Short Film
1931/32	Best Live Action Short Film
1934	Best Film Editing
1934	Best Original Score
1934	Best Original Song
1936	Best Supporting Actor
1936	Best Supporting Actress
1939	Best Visual Effects
1940	Best Original Screenplay
1941	Best Documentary Short Film
1943	Best Documentary Feature Film
1947	Best International Feature Film
1948	Best Costume Design
1981	Best Makeup and Hairstyling
2001	Best Animated Feature Film
2025	Best Casting ^[104]

Inconsistent Oscar Awards through time (Source : Wikipedia)

II - DIFFICULTIES

II.2 - HTML STRUCTURES DIFFICULTIES

- Persons and films treated without distinction

```
▼<div class="nomination__body"> [flex]
  ▼<div class="nomination__body__top">
    <div class="nomination__status color-gold font-bold">Winner</div>
    <h3 class="tile__heading nomination__heading text-reg font-bold">American Fiction</h3>
  ▼<div class="nomination__nominees text-sm o-50">
    Cord Jefferson
  </div>
```

```
▼<div class="nomination__body__top">
  <div class="nomination__status color-gold font-bold">Winner</div>
  <h3 class="tile__heading nomination__heading text-reg font-bold">Cillian Murphy</h3>
  ▼<div class="nomination__nominees text-sm o-50">
    Oppenheimer
  </div>
</div>
```

II - DIFFICULTIES

II.2 - HTML STRUCTURES DIFFICULTIES

Golden Bear for Best Film (awarded to the
film's producers)

Dahomey
by Mati Diop
produced by: Eve Robin, Judith Lou Lévy, Mati Diop

Silver Bear Grand Jury Prize

Yeohaengjaui pilyo
by Hong Sangsoo

Silver Bear for Best Director

Nelson Carlos De Los Santos Arias
Pepe

II - DIFFICULTIES

II.2 - HTML STRUCTURES DIFFICULTIES

- Unclean nodes

The screenshot shows the Academy Award website's category page for 'ANIMATED SHORT FILM'. At the top, there are sections for 'OPPENHEIMER' (Production Design: Arthur Max; Set Decoration: Elli Griff) and 'AWARDS/INDUSTRY' (Production Design: Ruth De Jong; Set Decoration: Claire Kaufman). Below these are three main navigation links: 'AWARDS/INDUSTRY', 'MUSEUM/COLLECTION', and 'SUPPORT AND SHOP'. A large, stylized Oscar statuette logo is centered above the category title. The category title 'ANIMATED SHORT FILM' is prominently displayed in large, gold-colored letters. Below it, the word 'WINNER' is highlighted in gold. The winning film is listed as 'WAR IS OVER! INSPIRED BY THE MUSIC OF JOHN & YOKO' by 'Dave Mullins and Brad Booker'.

```
<div class="field field--name-field-award-category-oscars field--type-entity-reference field--label-hidden field__item">Animated Short  
▼<div class="field field--name-field-award-honorees field--type-entity-reference-revisions field--label-hidden field__items" data-once=""  
▼<div class="field__item">  
    ▼<div class="paragraph paragraph--type--award-honoree paragraph--view-mode--oscars">  
        <div class="field field--name-field-honoree-type winner">Winner</div>  
        ▼<div class="field">  
            <div class="field field--name-field-award-film field--type-entity-reference field--label-hidden field__item">  
                WAR IS OVER! Inspired by the Music of John & Yoko</div>  
            </div>  
        ▶<div class="field">...</div>  
    </div>
```

II - DIFFICULTIES

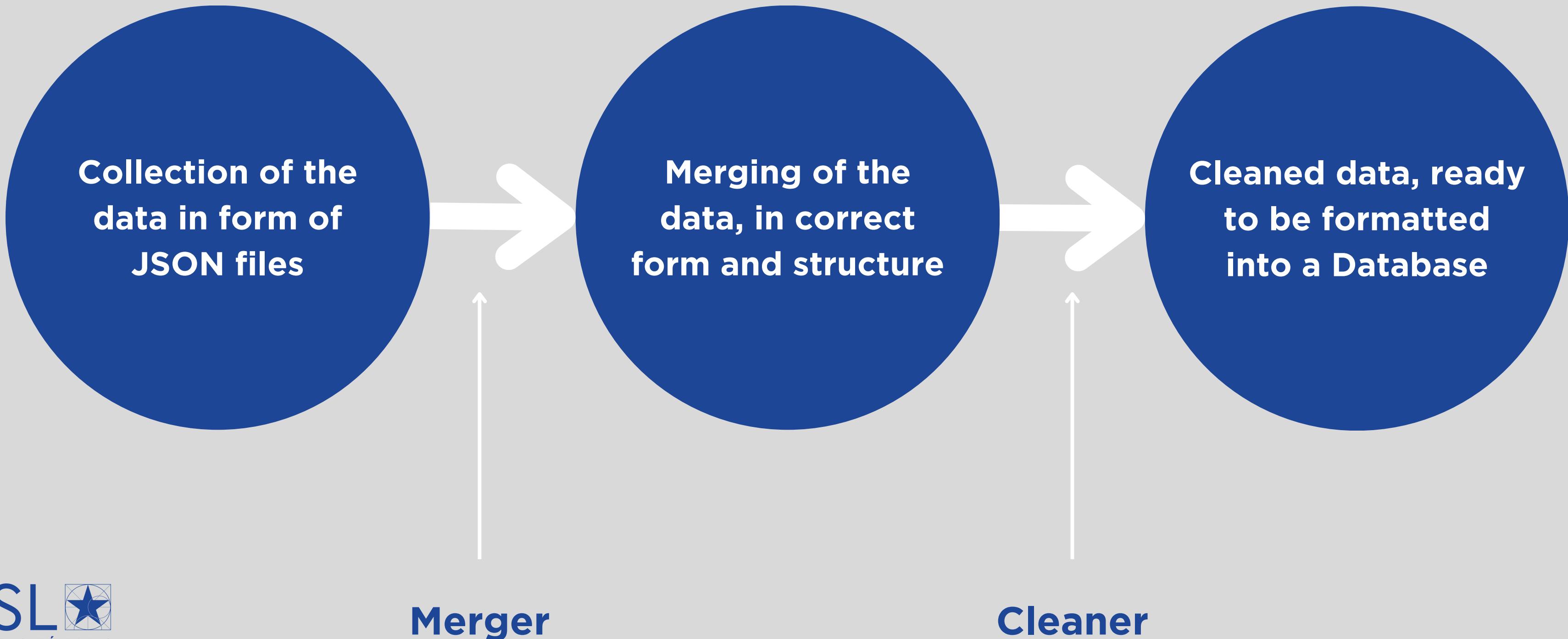
II.2 - HTML STRUCTURES DIFFICULTIES

- Missing entries
- **rowspan**
- different tables on the page

1966	<i>Un homme et une femme</i>		Claude Lelouch	 France
	<i>Ces messieurs dames</i>	<i>Signore & signori</i>	Pietro Germi	
1967	<i>Blow-Up</i>		Michelangelo Antonioni	 Italie
1968	Arrêté en cours de déroulement à cause des événements de mai 68 ⁶¹			
1969	<i>If....</i>		Lindsay Anderson	 Royaume-Uni
1970	<i>MASH</i>		Robert Altman	 États-Unis
1971	<i>Le Messager</i>	<i>The Go-Between</i>	Joseph Losey	 Royaume-Uni
1972	<i>La classe ouvrière va au paradis</i> ^{note 1,62}	<i>La classe operaia va in paradiso</i>	Elio Petri	
	<i>L'Affaire Mattei</i> ^{note 1,62}	<i>Il caso Mattei</i>	Francesco Rosi	 Italie
1973	<i>La Méprise</i>	<i>The Hireling</i>	Alan Bridges	 Royaume-Uni
	<i>L'Épouvantail</i>	<i>Scarecrow</i>	Jerry Schatzberg	
1974	<i>Conversation secrète</i>	<i>The Conversation</i>	Francis Ford Coppola	 États-Unis

III. STRUCTURE OF THE FINAL DATABASE

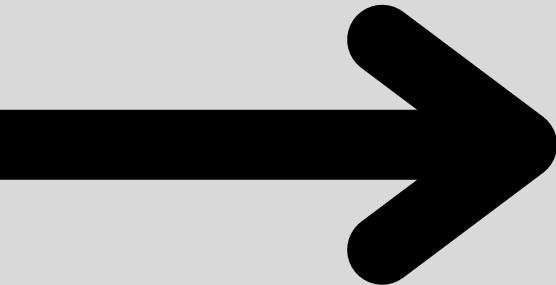
III.1 - MERGING AND CLEANING OF DATA



III. STRUCTURE OF THE FINAL DATABASE

III.1 - MERGING AND CLEANING OF DATA

Merging each award
gained by each
movie



```
{  
  "Year": "2013",  
  "Film": "Life Of Pi",  
  "Details": [  
    {"Award": "Adapted Screenplay",  
     "Nominee(s)": [  
       "David Magee"  
     ]  
   }  
 ],  
},
```

```
"Year": "2013",  
"Film": "Life Of Pi",  
"Details": [  
  {  
    "Award": "Adapted Screenplay",  
    "Nominee(s)": [  
      "David Magee"  
    ]  
  },  
  {  
    "Award": "Cinematography",  
    "Nominee(s)": [  
      "Claudio Miranda"  
    ]  
  },  
  {  
    "Award": "Director",  
    "Nominee(s)": [  
      "Ang Lee"  
    ]  
  },  
  {  
    "Award": "Editing",  
    "Nominee(s)": [  
      "Tim Squyres"  
    ]  
  },  
],
```

III. STRUCTURE OF THE FINAL DATABASE

III.2 - DIFFICULTIES

- **Inconsistencies in movie titles**
- **Character confusion, and necessary format change**
- **During the Merge, handling of some inconsistencies on the variables that represent the movie titles**
- **Handling of small differences in titles**

III. STRUCTURE OF THE FINAL DATABASE

III.3 - CREATING THE DATABASE

By continuing the format of the Jsons, we create the table .

First we contain the **Year** and the **Film**.

Then we insert inside:

**{(Award1, Nominee1),
(Award2,Nominee2),...}**

```
CREATE TABLE IF NOT EXISTS Movies (
    id INTEGER PRIMARY KEY AUTOINCREMENT,
    year TEXT,
    film TEXT UNIQUE
);
```

```
CREATE TABLE IF NOT EXISTS Awards (
    id INTEGER PRIMARY KEY AUTOINCREMENT,
    movie_id INTEGER,
    award TEXT,
    nominees TEXT,
    FOREIGN KEY (movie_id) REFERENCES Movies(id)
);
```

Year, Film, Awards and Nominees

2022,A Quiet Place Part II,"Sound (Ethan Van der Ryn, Erik Aadahl, Malte Bieler, Brandon Proctor, Michael Barosky)"
2022,Affairs of the Art,"British Short Animation (Joanna Quinn, Les Mills)"
2022,After Love,"Director (Aleem Khan),Leading Actress (Joanna Scanlan),Outstanding British Film (Aleem Khan, Matthew Warchus)"
2022,Ali & Ava,"Leading Actor (Adeel Akhtar),Outstanding British Film (Clio Barnard, Tracy O'Riordan)"
2022,An Irish Goodbye,"British Short Film (Ross White, Tom Berkeley),SHORT FILM (Live Action) (),SHORT FILM (Live Action) (No nominees listed)"
2022,Ariana DeBose,"EE Rising Star Award (),EE Rising Star Award (No nominees listed)"
2022,Avatar: The Way of Water,"Golden Globe Award for Best Director (),Golden Globe Award for Best Director (No nominees listed)"
2022,Becoming Cousteau,"Documentary (Mridu Chandra, Liz Garbus, Dan Cogan, Evan Hayes)"
2022,Being The Ricardos,"Original Score (Daniel Pemberton),Original Screenplay (Aaron Sorkin)"
2022,Black Panther: Wakanda Forever,"COSTUME DESIGN (),COSTUME DESIGN (No nominees listed),COSTUME DESIGN (Ruth Carter)"
2022,Boiling Point,"Casting (Carolyn McLeod),Leading Actor (Stephen Graham),Outstanding British Film (James Cumming)"
2022,C'mon C'mon,Supporting Actor (Woody Norman)
2022,Cow,"Documentary (Andrea Arnold, Kat Mansoor)"
2022,Cyrano,"Costume Design (Massimo Cantini Parrini),Make Up & Hair (Siân Miller, Alessandro Bertolazzi),Outstanding Production Design (No nominees listed)"
2022,Do Not Feed The Pigeons,"British Short Animation (Antonin Niclass, Vladimir Krasilnikov, Jordi Morera)"
2022,Don't Look Up,"Best Film (Kevin Messick, Adam McKay),Leading Actor (Leonardo DiCaprio),Original Score (Nicholas Heidke)"
2022,Everybody's Talking About Jamie,"Outstanding British Film (Tom MacRae, Dan Gillespie Sells, Mark Herbert, Jonathan Teplitzky)"
2022,Everything Everywhere All at Once,"ACTOR IN A SUPPORTING ROLE (),ACTOR IN A SUPPORTING ROLE (Ke Huy Quan),ACTOR IN A SUPPORTING ROLE (No nominees listed)"
2022,Femme,"British Short Film (Ng Choon Ping, Hayley Williams, Rienkje Attoh, Sam H. Freeman, Sam Ritzenberg)"
2022,Flee,"Animated Film (Monica Hellström, Jonas Poher Rasmussen, Signe Byrge Sørensen, Charlotte De La Gournerie)"
2022,Free Guy,"Special Visual Effects (Daniel Sudick, Bryan Grill, Swen Gillberg, Nikos Kalaitzidis)"
2022,Ghostbusters: Afterlife,"Special Visual Effects (Pier Lefebvre, Alessandro Ongaro, Sheena Duggal, Aharon Bourla)"
2022,Guillermo del Toro's Pinocchio,"ANIMATED FEATURE FILM (),ANIMATED FEATURE FILM (Guillermo del Toro, Mark Gustafson)"
2022,Happening,Director (Audrey Diwan)

[Year](#),[Film](#),[Awards](#) and [Nominees](#)

2009,Avatar,"ART DIRECTION (),ART DIRECTION (No nominees listed),ART DIRECTION (Production Design: Rick Cart

2022,Avatar: The Way of Water,"Golden Globe Award for Best Director (),Golden Globe Award for Best Director

Films to contain the word “Avatar” in the title



III. STRUCTURE OF THE FINAL DATABASE

III.4-SQL DIFFICULTIES

- **SQL conversion was leading to repetitions of movies because of wrong format**
- **Some data don't contain nominees, thus the SQL conversion pipeline needed to address null spaces**
- **Joining the awards and nominees to the movie proved quite tricky sometimes, leading to wrong results**

Thank you.



Aymeric Behaegel
Alex Kouvatseas
Lucas Rousselet