

ナント中央学校

Projet d'Application

Sujet 6 : Conception d'un pipeline de manipulation de données au service des fans d'animation japonaise

Etudiants**Ahmad SAID****Rafael VAZQUEZ****Encadrants****Morgan MAGNIN****Tony RIBEIRO**

Soutenance le 09 - 01 - 2023



Table de matière

1. Contexte du Projet

Présentation de AniList

Présentation de l'API
AniList

Travail attendu

2. Méthodologie

Environnement de travail

Model conceptuel des
données

Récupération des
données

3. Analyse de données

Analyse via Power Bi

Analyse via Python
notebook

Analyse via Application
WEB



漫画

01

Contexte du Projet

Présentation de AniList

Présentation de l'API AniList

Travail attendu



The next-generation anime platform

Outil pour partager, suivre et découvrir des informations à propos des Mangas et animes

A screenshot of the Anilist homepage. It features a top navigation bar with links for Home, Profile, Anime List, Manga List, Browse, and Forum. Below the navigation is a search bar with a flame icon. The main content area is divided into several sections: 'Activity' (listing recent posts from users like dosau, onCrox, nadhdfdf, AlchimistaDacciaio, and LightKaworu), 'Following' (a global feed), 'In Progress' (manga and anime lists), 'Forum Activity' (a discussion thread about Fumetsu No Anata E Season 2), 'Activity feed' (another list of recent posts), 'Recent Reviews' (a grid of manga and anime covers), and a 'Write a status...' input field.

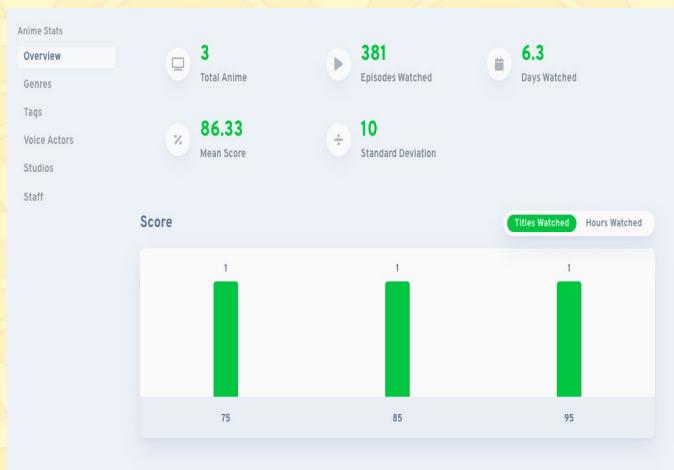
Home Page

A screenshot of the Anilist search results for 'SPY x FAMILY Cour 2'. The page shows the title 'SPY x FAMILY Cour 2' and a brief description 'The second half of SPYxFAMILY.' Below this are tabs for Overview, Characters, Staff, Stats, and Social. The Overview section includes a 'Relations' section with links to Source, Prequel, Sequel, and Side Story. It also lists characters and staff members with their roles and nationalities. The 'Overview' section also displays the title's rank (#70 Highest Rated All Time) and popularity (#194 Most Popular All Time).

Recherche d'Animes et de Mangas

A screenshot of a user profile page for 'rafaelvnm'. The profile picture is a cartoon character with black hair and blue eyes. The page includes sections for Overview, Anime List, Manga List, Favorites, Stats, Social, Reviews, and Submissions. The 'Activity History' section shows a timeline of recent activity, including pausing 'BLEACH: Thousand-Year Blood War', watching episode 25 of 'My Hero Academia Season 6', and watching episode 56 of 'ONE PIECE'. The 'Genre Overview' section shows the user's preferences across various genres. The 'Stats' section provides summary information: 3 Total Anime, 6.3 Days Watched, 86.3 Mean Score, 381 Episodes Watched, 6.3 Days Watched, 86.33 Mean Score, and 10 Standard Deviation.

Profil



Statistiques de profil



Exploration
D'Anilist



ANILIST



API Anilist GraphQL

Langage de requête pour accéder et manipuler des données tel que:

Les médias (anime ou manga)

Les studios

Les personnages

Les doubleurs

Les producteurs



The screenshot shows a GraphiQL interface with a query editor, a results panel, and a sidebar.

QUERY EDITOR (Left):

```

1 # Welcome to GraphiQL
2 #
3 # GraphiQL is an in-browser tool for writing, validating, and
4 # testing GraphQL queries.
5
6 query ($id: Int) { # Define which variables will be used in the query (id)
7   Media (id: $id, type: ANIME) { # Insert our variables into the query argument
8     id
9     title {
10       romaji
11       english
12       native
13     }
14     characters {
15       edges {
16         role
17         node {
18           name {
19             userPreferred
20           }
21         }
22       }
23     }
24   }
25 }
```

QUERY VARIABLES (Bottom Left):

```

1 {
2   "id": 231
3 }
```

RESULTS (Right):

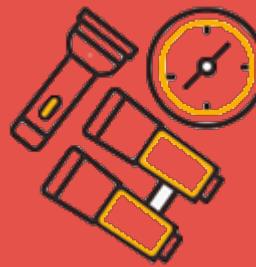
```

{
  "data": {
    "Media": {
      "id": 231,
      "title": {
        "romaji": "Asagiri no Miko",
        "english": "Shrine of the Morning Mist",
        "native": "朝霧の巫女"
      },
      "characters": {
        "edges": [
          {
            "role": "SUPPORTING",
            "node": {
              "name": {
                "userPreferred": "Kurako Hieda"
              }
            }
          },
          {
            "role": "SUPPORTING",
            "node": {
              "name": {
                "userPreferred": "Koma"
              }
            }
          },
          {
            "role": "SUPPORTING",
            "node": {
              "name": {
                "userPreferred": "Tama Hieda"
              }
            }
          },
          {
            "role": "MAIN",
            "node": {
              "name": {
                "userPreferred": "Yuzu Hieda"
              }
            }
          },
          {
            "role": "MAIN"
          }
        ]
      }
    }
  }
}
```

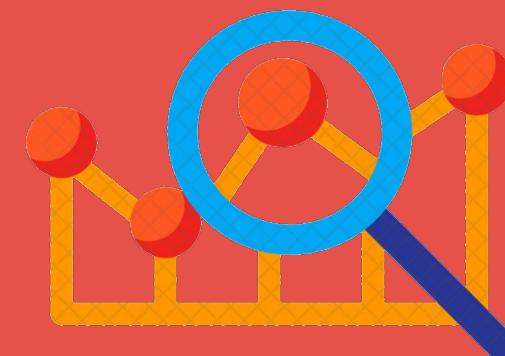


Travail Attendu

Exploration / Relations / Suivi Par AniList



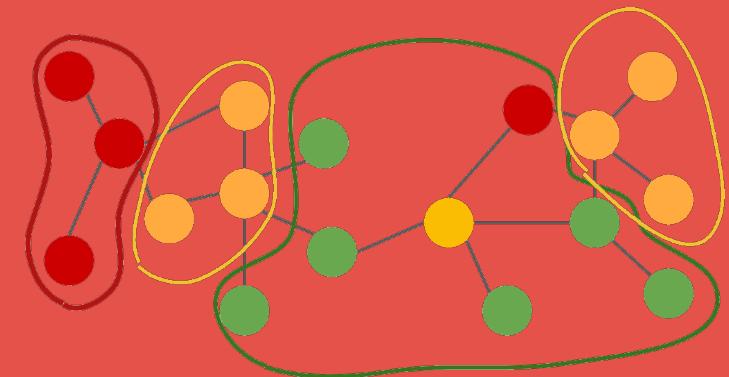
Pas d'Analyse
Ni Evolution



Évolution des genres d'anime
populaires au fil du temps ?



Sexes des personnages % à
leurs doubleurs ?



Graphe de connections
entre les doubleurs ?

漫画

02

Méthodologie Du travail

Environnement de travail

Modèle conceptuel des données

Récupération des données



Environnement du travail

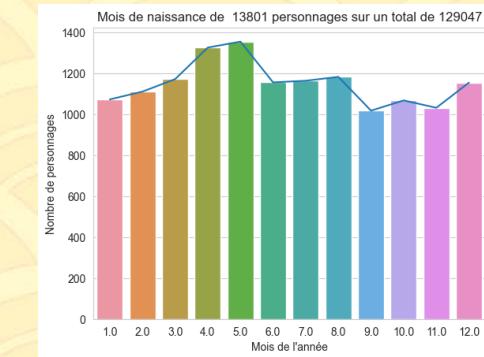


Power BI



Python

- Puissant
- Rapide
- Facile à maintenir



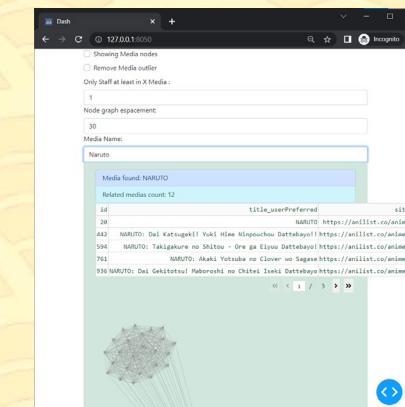
1. GraphQL request to Pandas dataframe

1.1 Extracting a single ressource

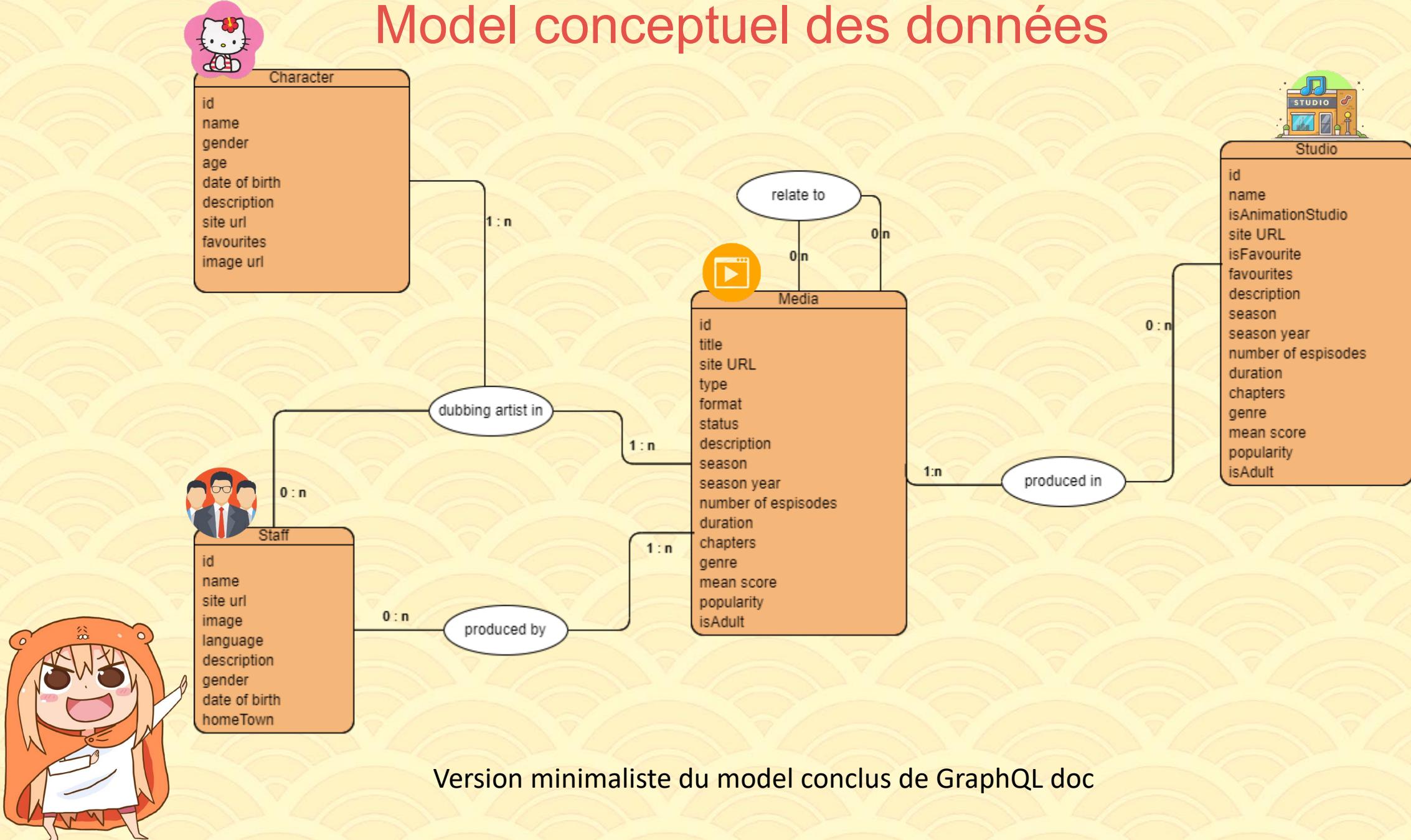
```
34 url = 'https://graphql.anilist.co'
35
36 response = requests.post(url, json={'query': query, 'variables': variables})
37
38 resultDict = response.json()['data']['Page']['media']
39 resultDataFrame = pandas.json_normalize(resultDict)
40 resultDataFrame
```

	id	season	seasonYear	source	episodes	popularity	title.romaji
0	98515	WINTER	2020	LIGHT_NOVEL	12	10507	Made in Abyss: Fukaki 1
1	99807	WINTER	2020	OTHER	12	14243	
2	100643	WINTER	2020	MANGA	1	92315	
3	101168	WINTER	2020	MANGA	24	75753	
4	101350	WINTER	2020	MANGA	13	24898	
5	101367	WINTER	2020	LIGHT_NOVEL	13	15042	Majutsushi O
6	101499	WINTER	2020	ORIGINAL	1	4518	High S
7	101574	WINTER	2020	ORIGINAL	1	13974	

10 rows x 10 columns [Open in new tab](#)



Model conceptuel des données



Récupération des données – Traitement via API

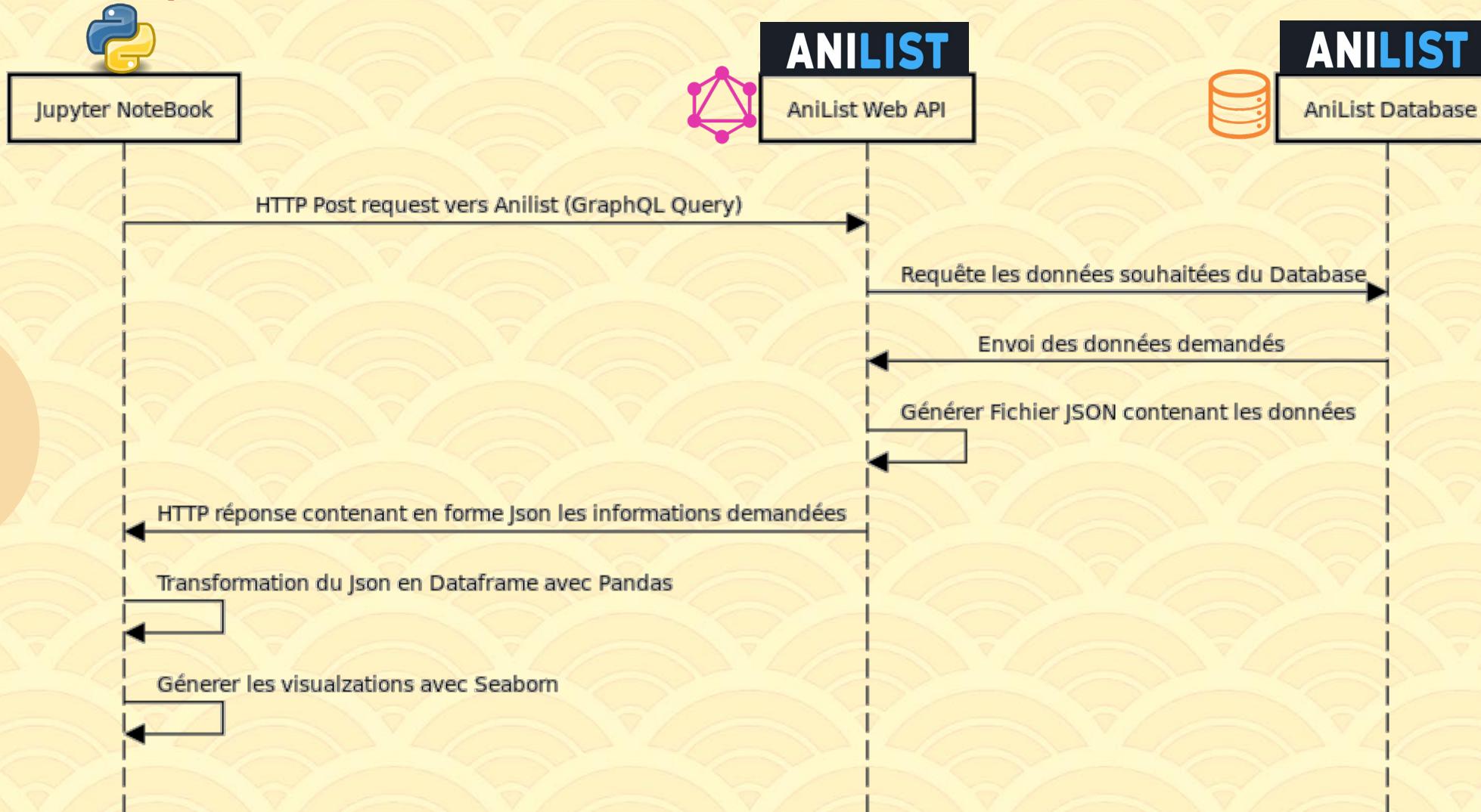


Diagramme de séquence : Accéder aux données d'AniList, les traiter, puis les visualiser

Récupération des données – Traitement local JSON

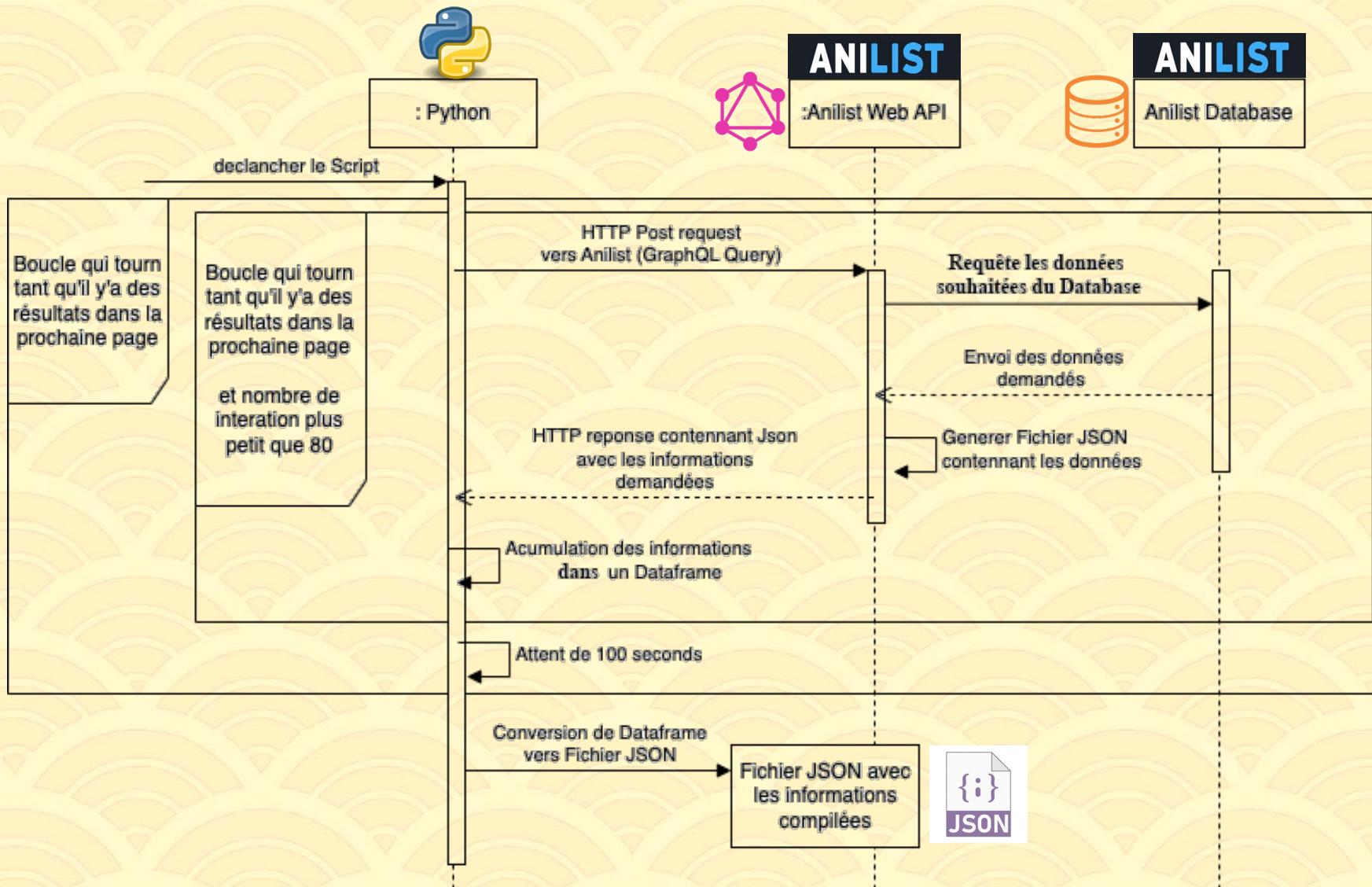


Diagramme de séquence : Sauvegarde de tous les données d'AniList en local format JSON

漫画

03

Analyse des Données

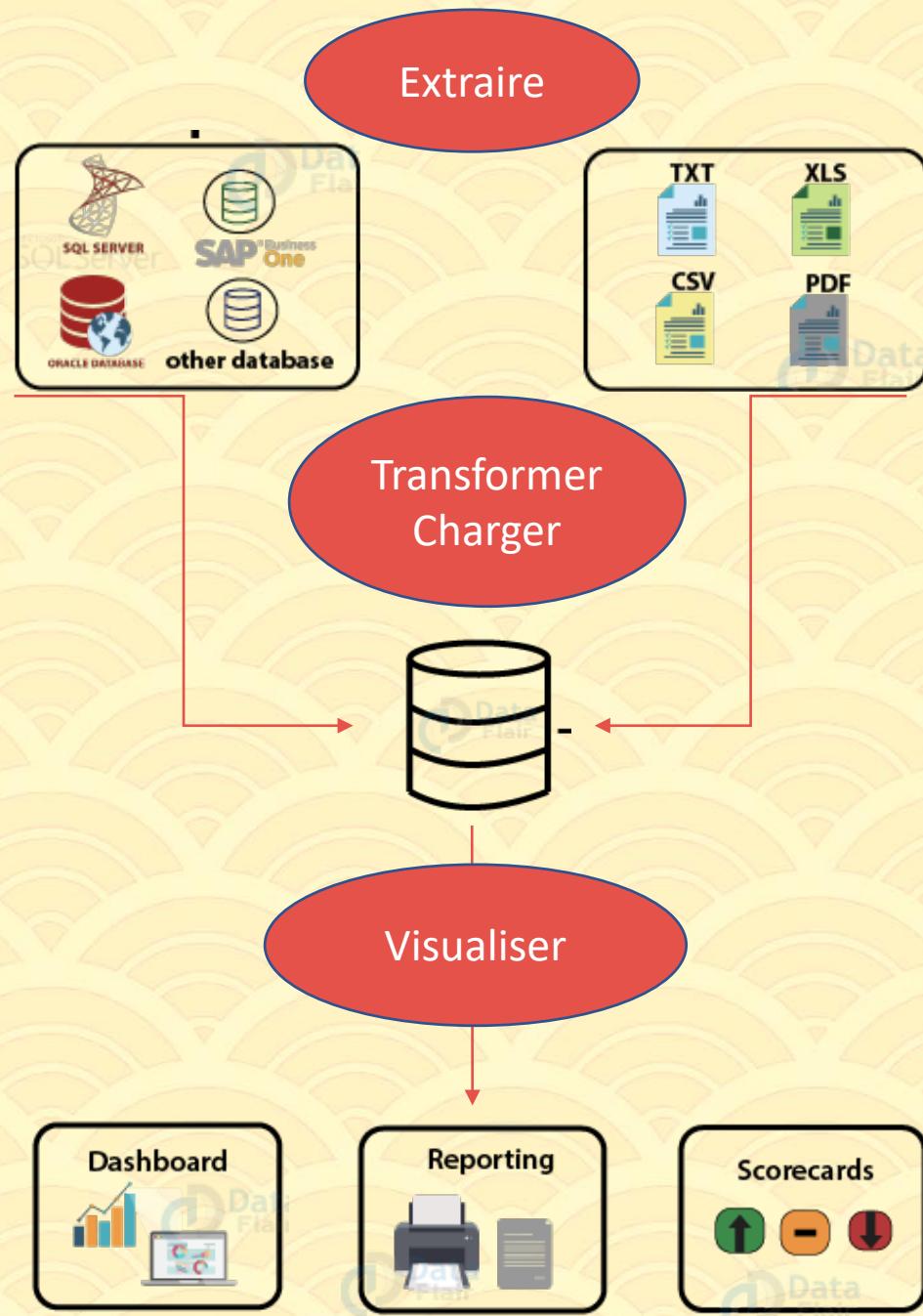
Analyse via Power Bi

Analyse via Python notebook

Analyse via Application WEB



Microsoft Power BI





Disponibilité des données



Characters	126.31K Count of id_charac	115.72K Count of image.medium	79.82K Count of description	39.41K Count of gender	11.27K Count of age
Media	102.27K Count of id_media	12.30K Count of seasonYear	84.53K Count of meanScore	12.30K Count of season	
	102.27K Count of isAdult	95.61K Count of id_media	102.22K Count of status_media	102.21K Count of format_media	
Staff	119.13K Count of id_staff	11.35K Count of homeTown	555 Count of Height	37.51K Count of gender	7386 Count of Years_Active
		13.18K Count of Birthday.date	1227 Count of DeathDay.date	13.05K Count of Age	3855 Count of bloodType

- **Characters** : Données problématiques
 - 1/3 sur leur sexe et 10% sur leur âge.
- **Media** : Données mieux remplis.
 - Problème liée à la saison et l'année de sortie; que 10%
- **Staff** : Données problématiques.
 - Analyse probablement sous-représentée. Maximum de 30% des renseignement.



Profil des tous les médias





Profil des médias avec une score Maximale



- 802 médias
- Type Manga
- Plus de personnages principaux (54 % contre 27%)
- Plus de contenus Adultes (45% contre 23%)
- Hentai monte de quatrième à premier categorie plus présent dans les médias



Analyse Spécifique des Données

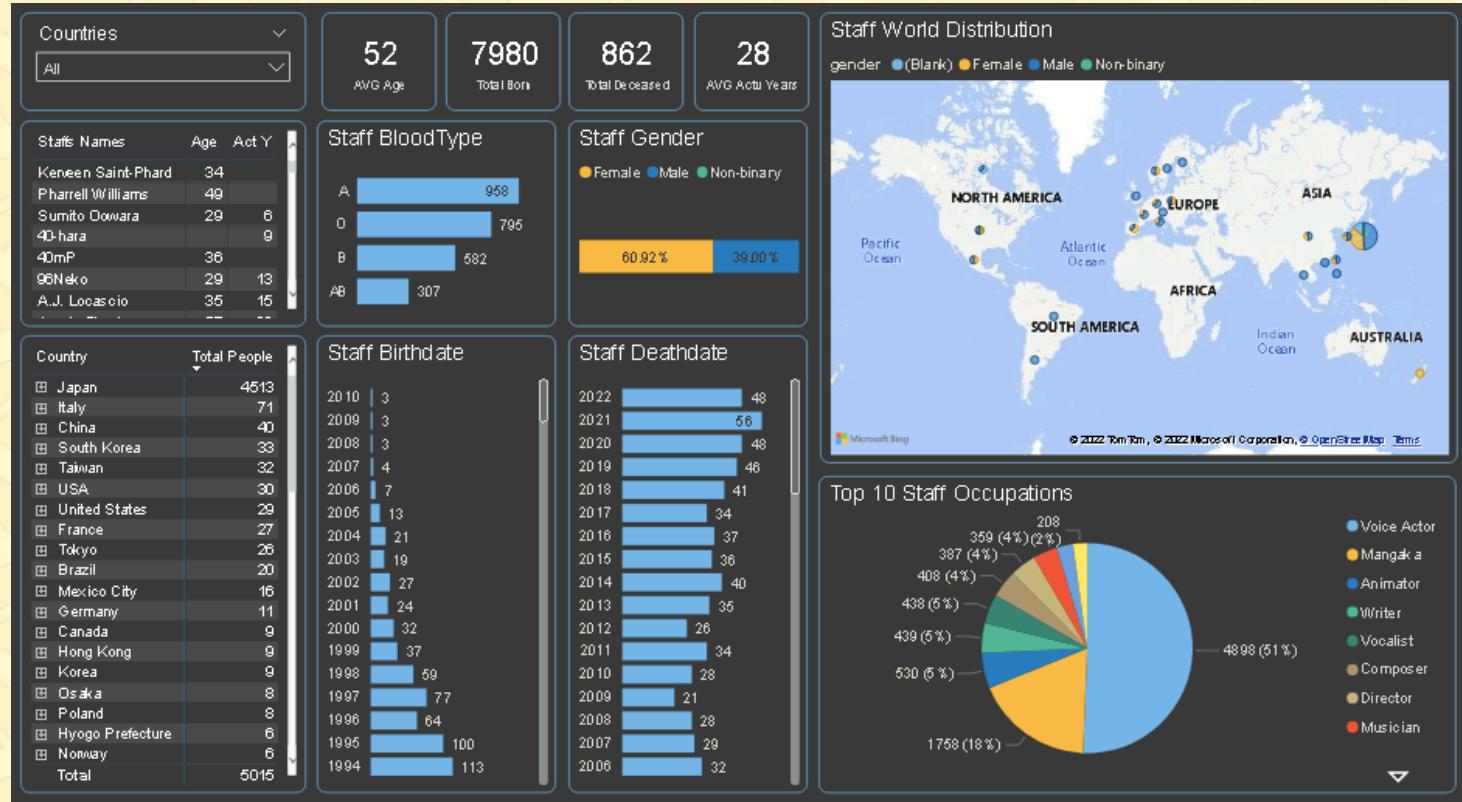


Title	Score	Season	Type	Status	Adult	Dur.	Format	Country	Charac	Genres	Staff	Studio	Year
Saezuru Toriwa Habatakanai: The clouds gather	72	WINTER	ANIME	FINISHED	1	85	MOMEM	JP	8	2	13	1	2020
Love Slave	80		ANIME	FINISHED	1	88	OVA	KR		2	2		
Kanojo × Kanojo × Kanojo (2013)	71	SPRING	ANIME	FINISHED	1	90	OVA	JP		1			2013
Kanashimi no Belladonna	71	SUMMER	ANIME	FINISHED	1	89	MOMEM	JP	8	2	24	1	1973

Exploration des médias d'une durée supérieure à 60 minutes avec une note supérieure à 70 sur la plateforme

- 100% Contenu Adulte
- Distribution uniforme sur la saison et l'année de sortie
- Durée comprise entre 80 et 90 minutes
- Plupart produit au Japon

Répartition du Personnel



- Âge moyen de 52 ans
- Moyenne de 28 ans d'activité
- Distribution des groupes sanguins selon le modèle de distribution au Japon
- Le japon, l'italie et la chine avec la plus grande concentration de personnel
- plus de 50% classés comme doubleur

漫画

03

Analyse des Données

Analyse via
Python notebook





Analyse via Python notebook

1. Groupe des images de caractères nés en même journée

Method A
Requête filtrage par GraphQL

Method B
Fichier JSON et filtrage par pandas DataFrame

Extraction du URL "image.medium"

Télécharger et enregistrer l'image

Uniformiser le ratio de forme de l'image

Regrouper en une seule image



Groupe des personnage nées dans la même journée 29-12



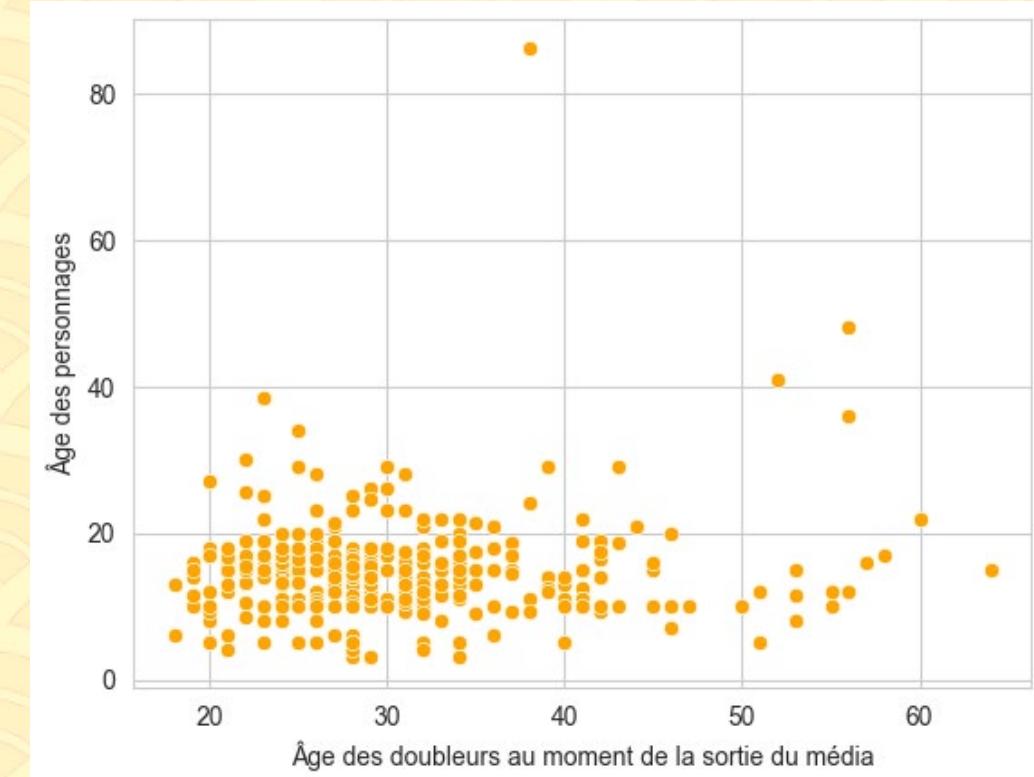
Analyse via Python notebook

2. Corrélation du sexe des personnages avec leurs doubleurs

Sexe des personnages	Féminin	Masculin	Total
Sexe des doubleurs			
Féminin	13504	1332	14836
Masculin	76	14518	14594
Total	13580	15850	29430

Contingence entre les sexes des personnages et leurs doubleurs

Les femmes peuvent doubler des enfants masculins!

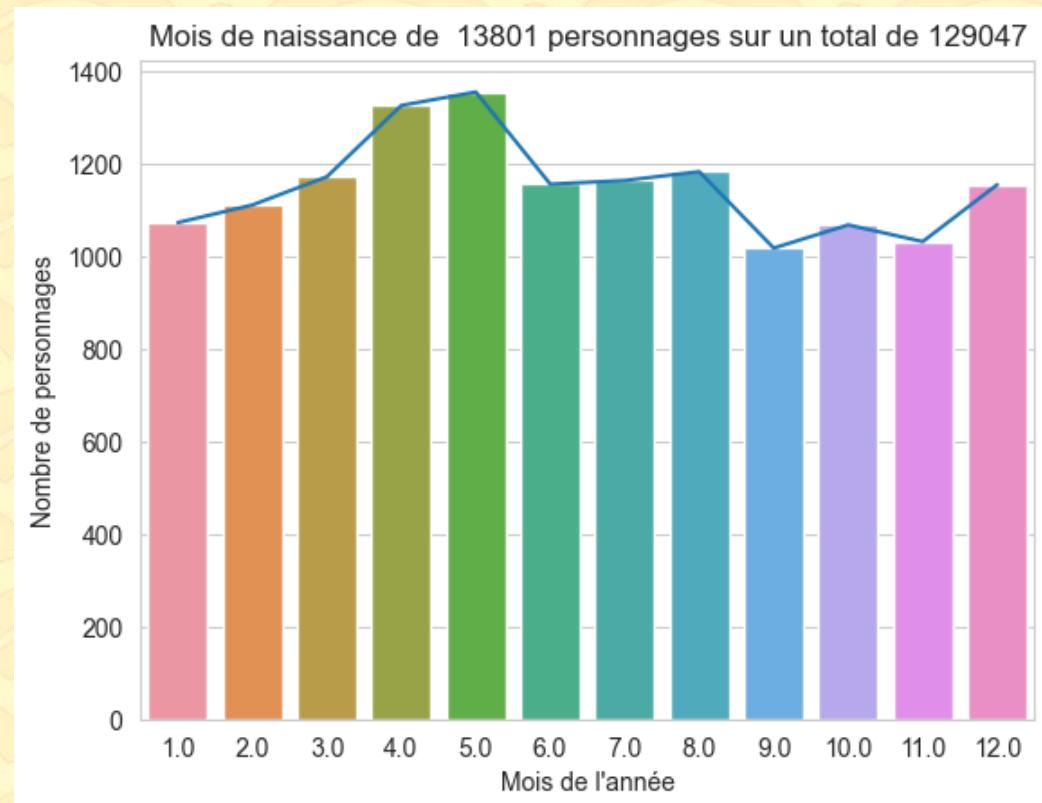


NUAGE DE POINTS DES PERSONNAGES MASQUELS DOUBLÉS PAR DES DOUBLEURS FÉMININS
MOYENNE D'ÂGE DES PERSONNAGES = 15,048065 ANS

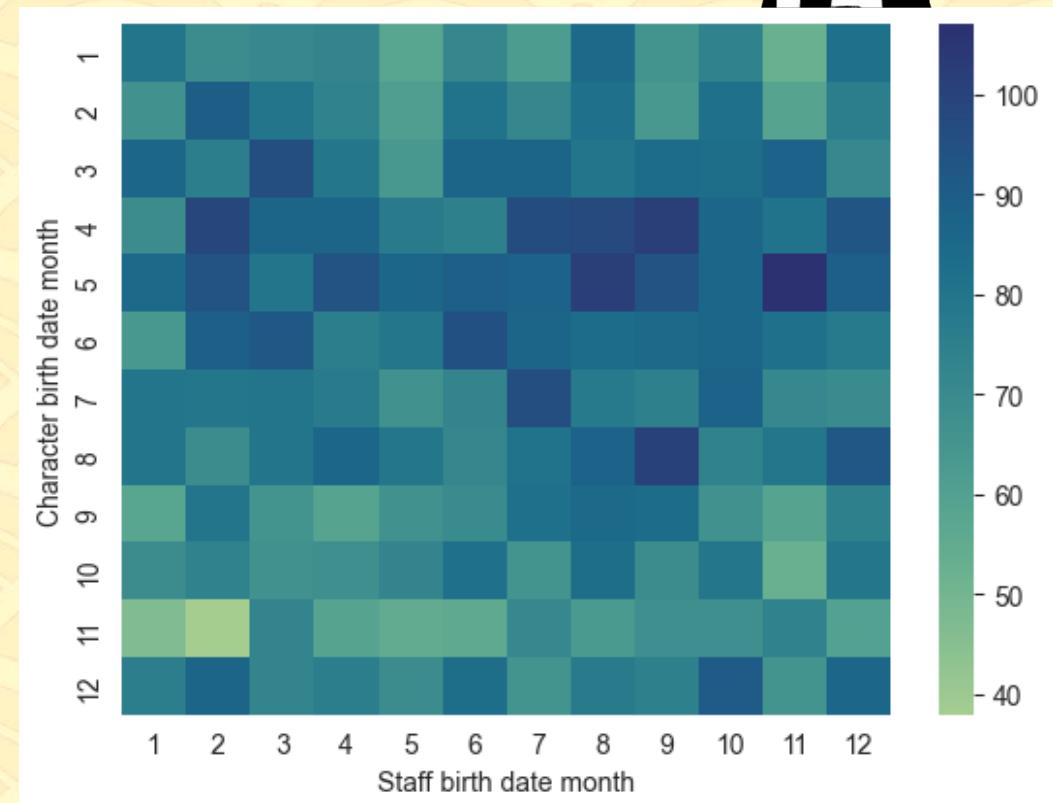


Analyse via Python notebook

3.1 Corrélation des dates de naissance des personnages avec leurs doubleurs



Répartition des personnages sur leurs mois de naissance



Heatmap du tableau de contingence entre les mois de naissance des personnages et ceux des doubleurs





Analyse via Python notebook

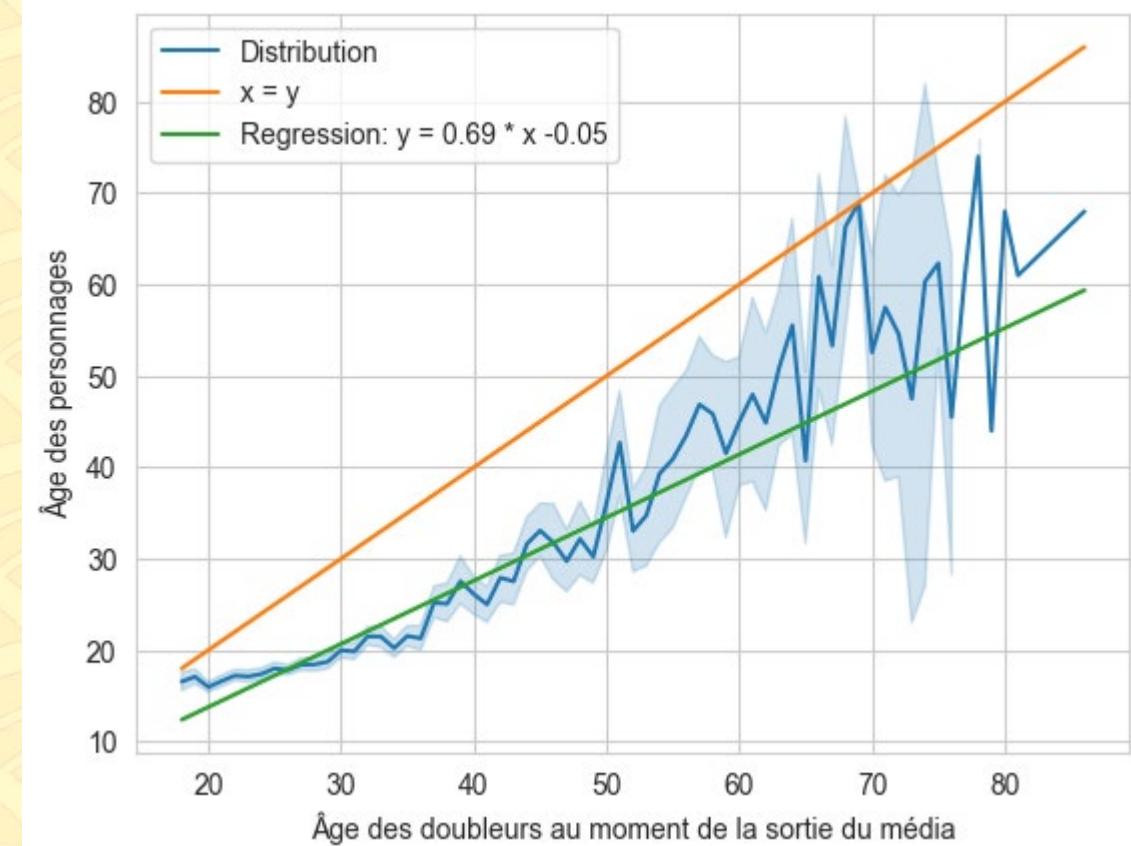
3.2 Corrélation des âges des personnages avec leurs doubleurs

Echantillon	Ages de doubleurs	Ages de personnages
Nombre d'échantillon	7199	7199
moyenne	31.638551	21.807430
écart-type	9.768778	11.759488

Corrélation : 0,574728

Moyennes et écarts-types des âges des doubleurs et des personnages

Tendance à doubler des personnages plus jeunes d'environ 10 ans





Analyse via Python notebook

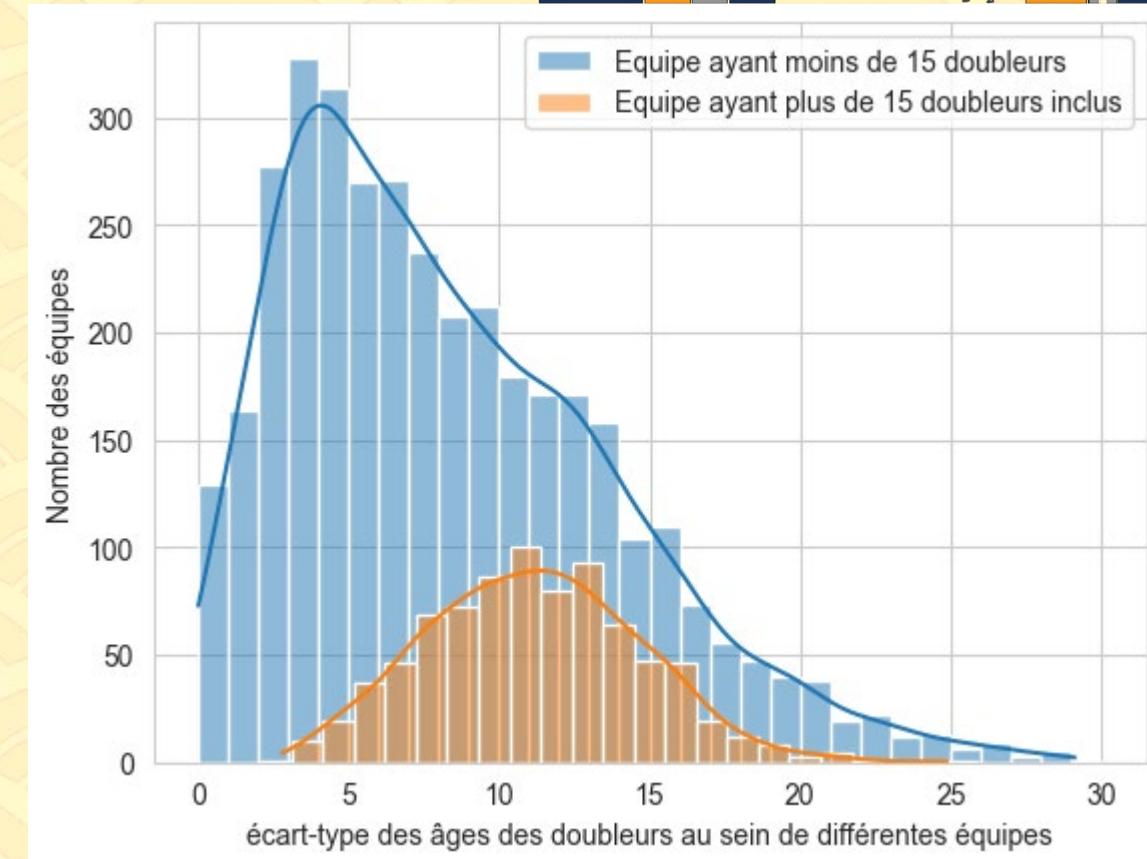
4. L'âge de l'équipe de travailleurs au moment du lancement d'un média

Echantillon	Petit Equipe < 15	Grand Equipe >= 15
Nombre d'échantillon	3648/4736	818/4736
Moyenne d'écart-type	8,35 ans	11,13 ans

Comparaison entre les équipes de petite taille et de grande taille

Âge similaire pour les médias de petite taille

Age plus disperse pour les médias de grande taille



Écart-type des âges des doubleurs au sein de différentes équipes

漫画

03

Analyse des Données

Analyse via
Application WEB





Analyse via Application WEB

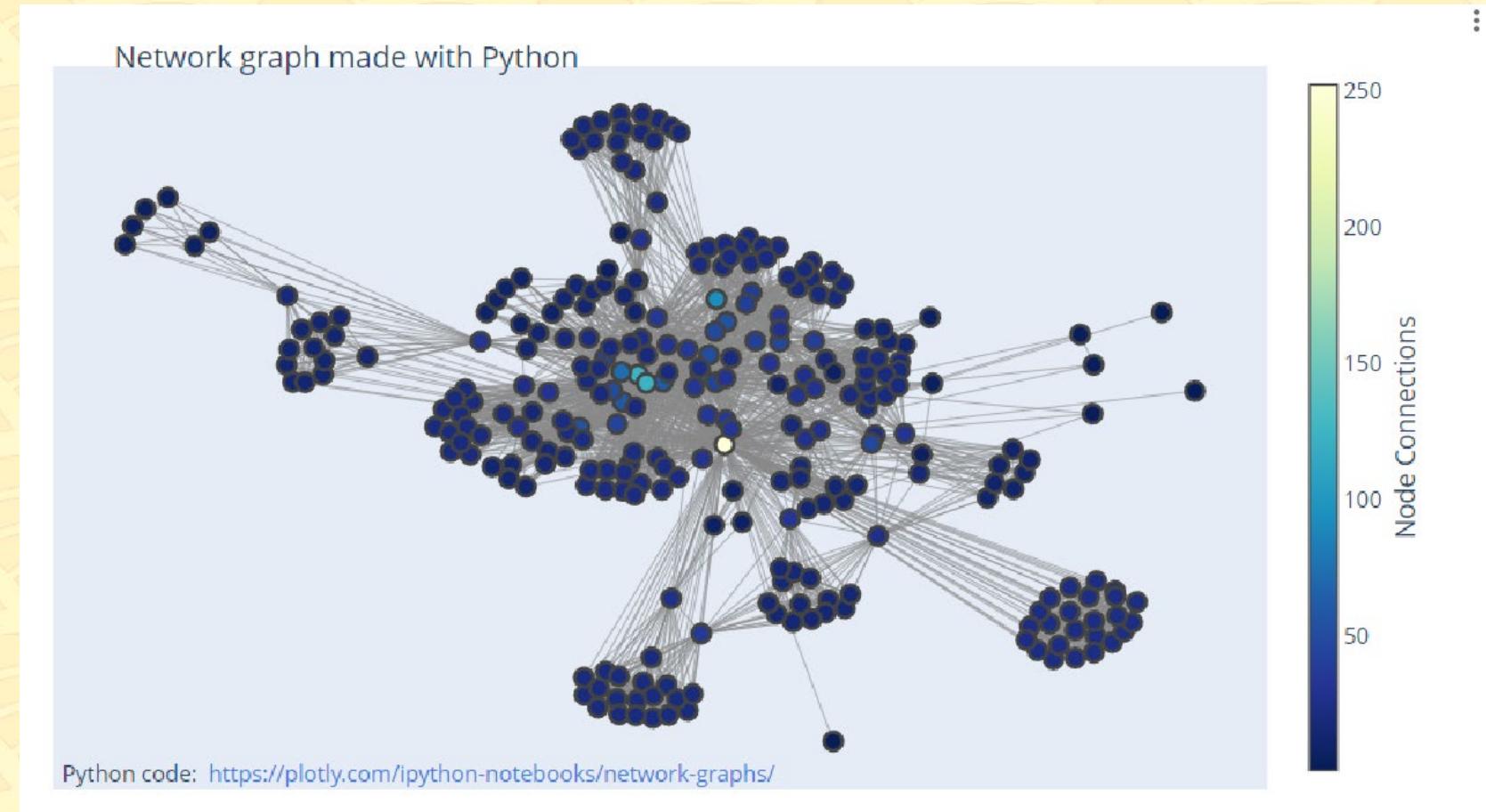
Graphe des équipes de travailles

Manque
d'interactivité

Hard Code le filtre
dans le script

Graphe condensée et
difficile d'interpreter

Limitation de Jupiter Notebook



Graphe complexe et indéchiffrable des travailleurs d'une série japonaise dans NoteBook



漫画

04

Live demo

Live vidéo démo

YouTube Playlist: AniList Analysis

- <https://www.youtube.com/playlist?list=PLEhbrzQmnpg3TizSdQpvAj8bIKYqB1Bf>

PowerBi Analysis

- <https://www.loom.com/share/acb80df936234b31805ea6650ca9a13d>



Analyse via Application WEB – Alternative du live

Dash

127.0.0.1:8050

Showing Media nodes

Remove Media outlier

Only Staff at least in X Media :

1

Node graph espacement:

30

Media Name:

[redacted]
[redacted]





Analyse via Application WEB – Alternative du live

Updating... x +

127.0.0.1:8050

Showing Media nodes
 Remove Media outlier
Only Staff at least in X Media :
1

Node graph espacement:
30

Media Name:
ayakashi

Media found: Pia Carrot e Youkoso!!: Sayaka no Koi Monogatari

Related medias count: 2

id	title_userPreferred	siteUrl
215	Pia Carrot e Youkoso!! 2 DX	https://anilist.co/anime/215
216	Pia Carrot e Youkoso!!: Sayaka no Koi Monogatari	https://anilist.co/anime/216

30 <>



Analyse via Application WEB – Alternative du live

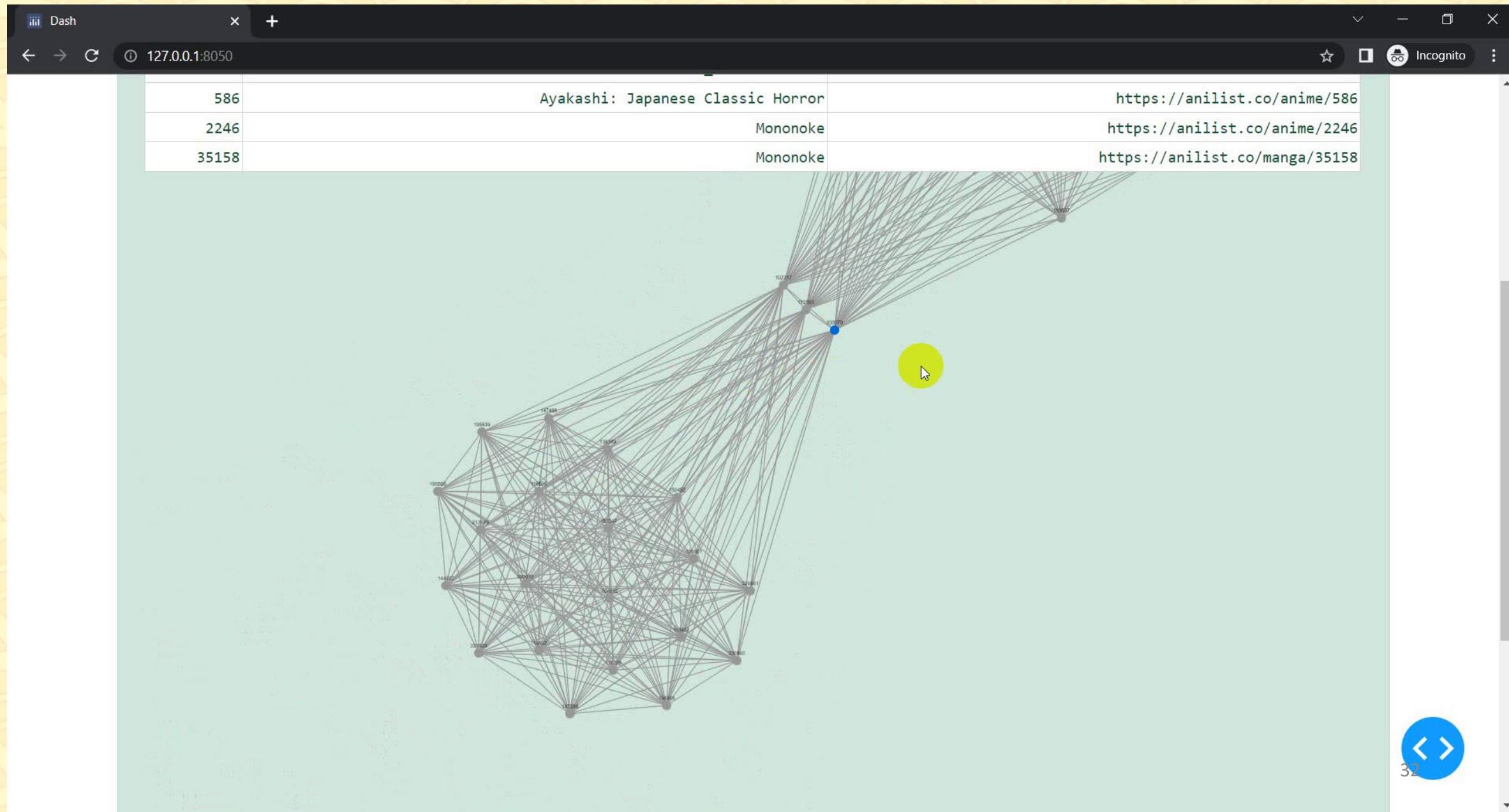
Dash x + 127.0.0.1:8050 Incognito :

Related media count: 3

id	title_userPreferred	siteUrl
586	Ayakashi: Japanese Classic Horror	https://anilist.co/anime/586
2246	Mononoke	https://anilist.co/anime/2246
35158	Mononoke	https://anilist.co/manga/35158

31 <>

Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live

Dash × + 127.0.0.1:8050 ☆ Incognito :

Staff participated in 2 medias of total 3

id	name_userPreferred	siteUrl
100070	Yasuharu Takanashi	https://anilist.co/staff/100070

id	title_userPreferred	siteUrl	Role
35158	Mononoke	https://anilist.co/manga/35158	Did not participate
586	Ayakashi: Japanese Classic Horror	https://anilist.co/anime/586	Music
2246	Mononoke	https://anilist.co/anime/2246	Music

33 < >



Analyse via Application WEB – Alternative du live

Dash x + 127.0.0.1:8050 ☆ Incognito :

Showing Media nodes
 Remove Media outlier

Only Staff at least in X Media : 1

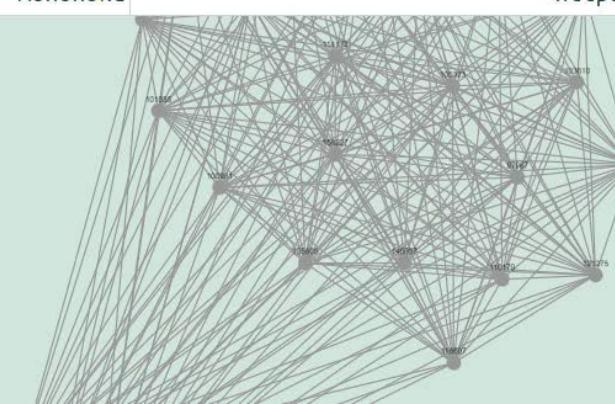
Node graph espacement: 30

Media Name: ayakashi

Media found: Ayakashi: Japanese Classic Horror

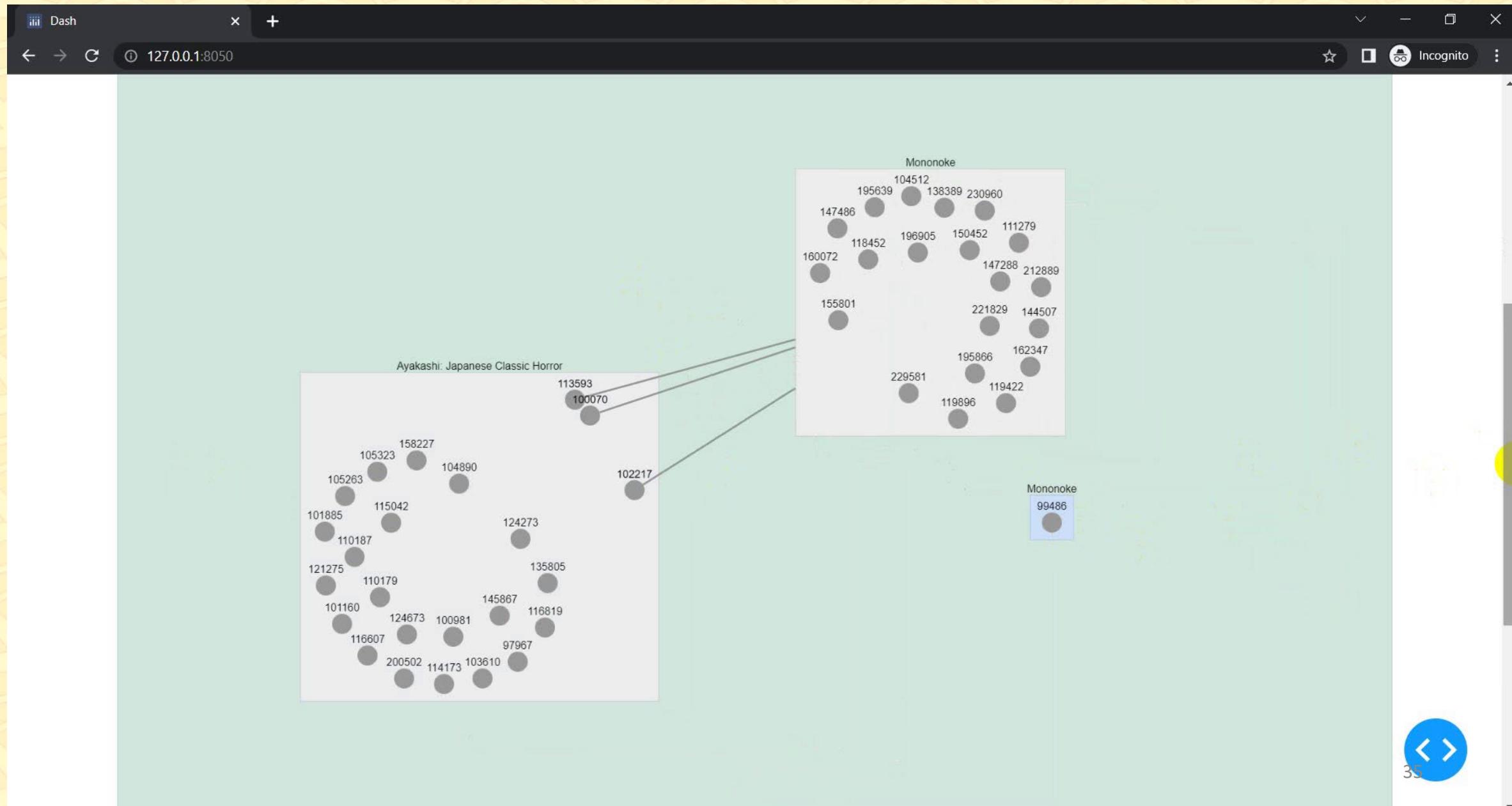
Related medias count: 3

id	title_userPreferred	siteUrl
586	Ayakashi: Japanese Classic Horror	https://anilist.co/anime/586
2246	Mononoke	https://anilist.co/anime/2246
35158	Mononoke	https://anilist.co/manga/35158



< >

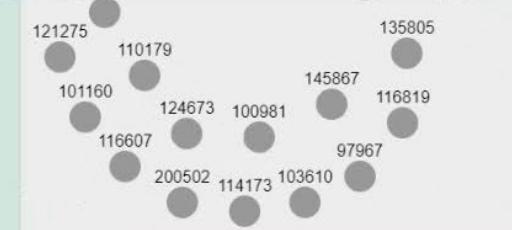
Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live

Dash x + 127.0.0.1:8050 Incognito :



This media have 1 staffs

id	title_userPreferred	siteUrl
35158	Mononoke	https://anilist.co/manga/35158

id	name_userPreferred	siteUrl	Role
99486	Yaeko Ninagawa	https://anilist.co/staff/99486	Story & Art

36



Analyse via Application WEB – Alternative du live

Dash x + ⌂ 127.0.0.1:8050 ☆ Incognito :

Showing Media nodes
 Remove Media outlier

Only Staff at least in X Media : 1

Node graph espacement: 30

Media Name: conan

Media found: Meitantei Conan

Related medias count: 63

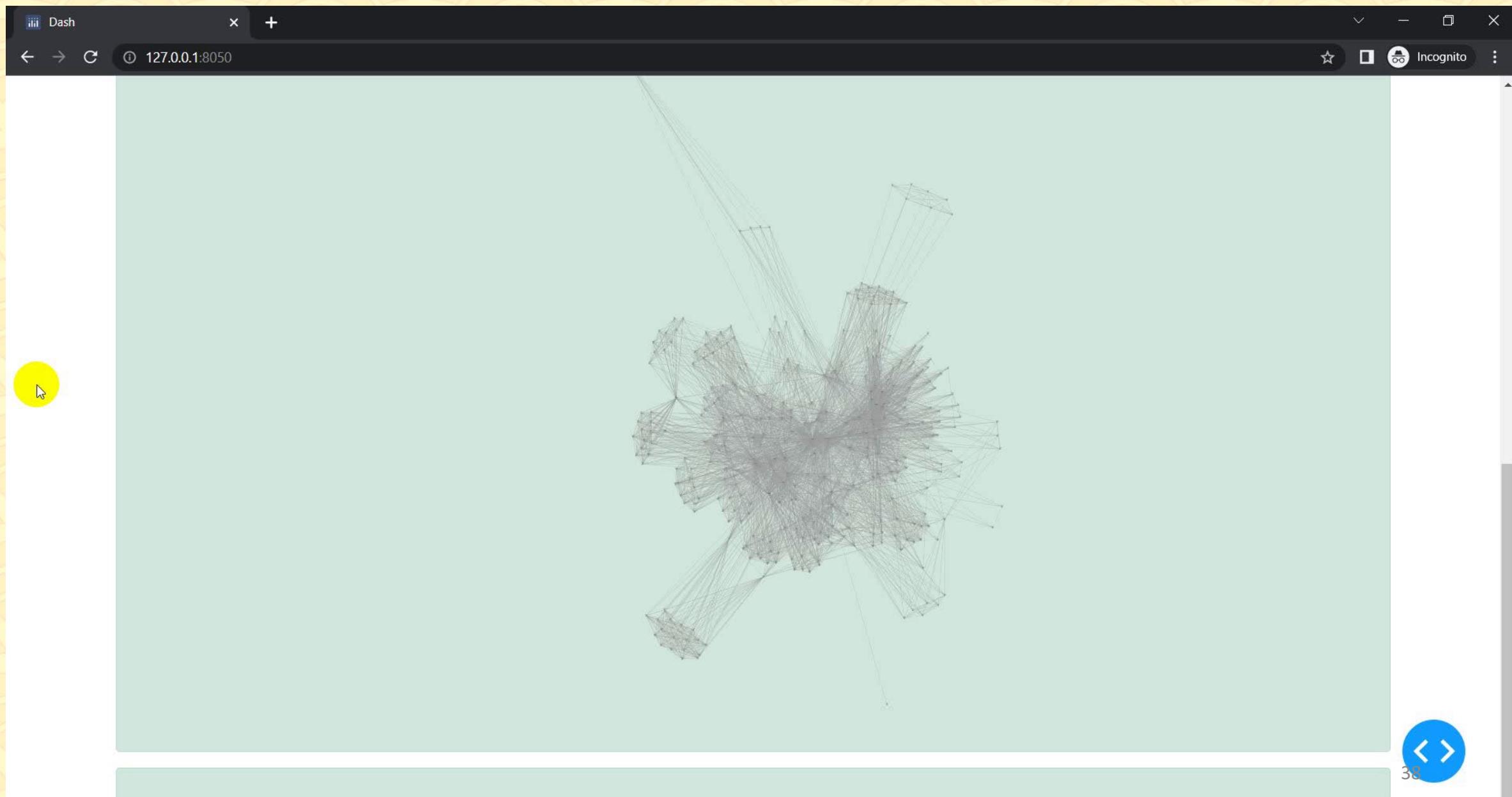
id	title_userPreferred	siteUrl
235	Meitantei Conan	https://anilist.co/anime/235
779	Meitantei Conan: Tokei Jikake no Matenrou	https://anilist.co/anime/779
780	Meitantei Conan: Juuyonbanme no Target	https://anilist.co/anime/780
781	Meitantei Conan: Seikimatsu no Majutsushi	https://anilist.co/anime/781
1293	Urusei Yatsura	https://anilist.co/anime/1293

<< < 1 / 13 > >>



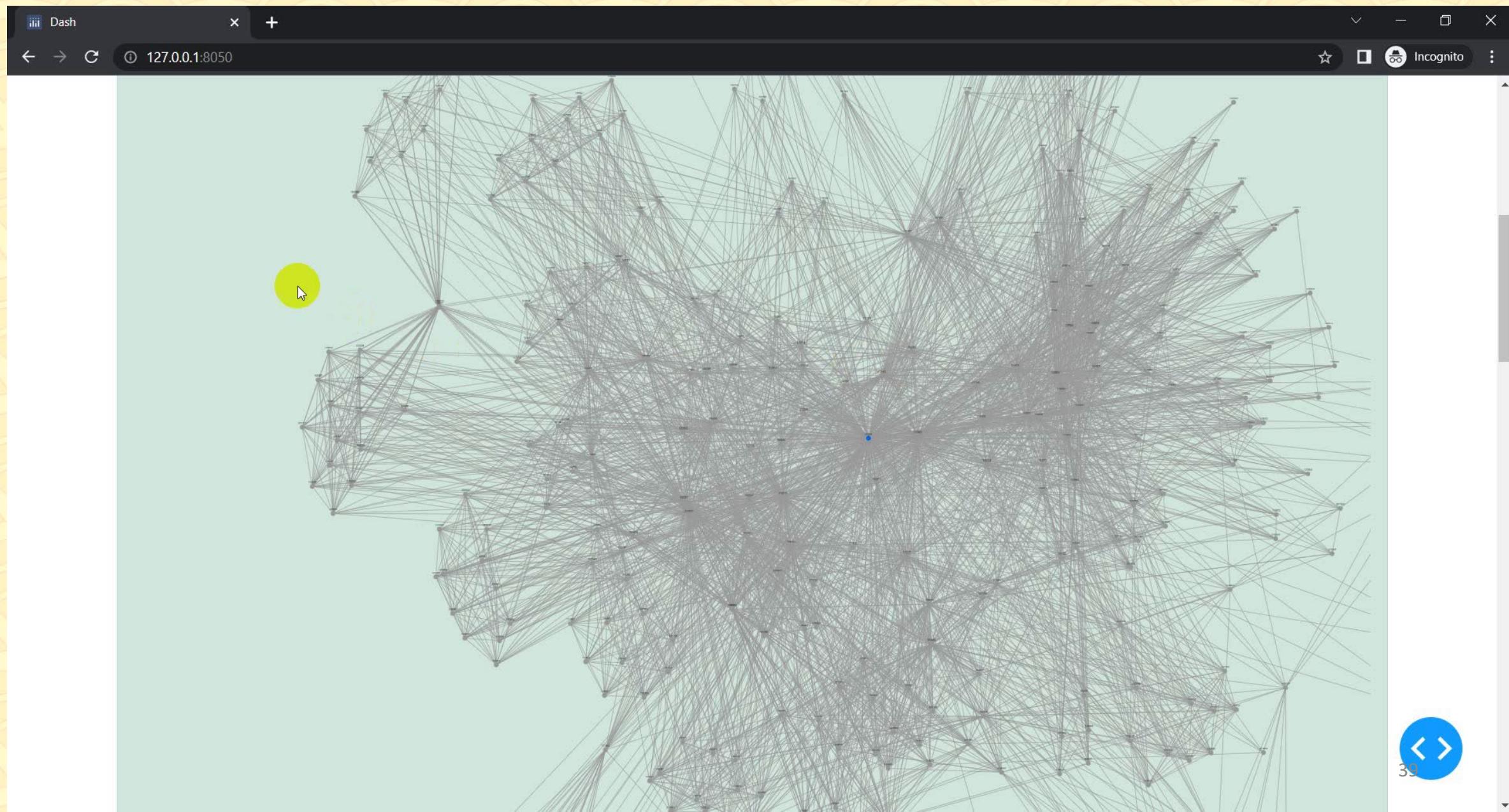


Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live

Staff participated in 61 medias of total 63

id	name_userPreferred	siteUrl
97804	Gosho Aoyama	https://anilist.co/staff/97804

id	title_userPreferred	siteUrl	Role
139179	Meitantei Conan: Tokubetsu-hen - "Kinga Shinnen Mouri Kogorou"	https://anilist.co/anime/139179	Did not participate
1293	Urusei Yatsura	https://anilist.co/anime/1293	Did not participate
235	Meitantei Conan	https://anilist.co/anime/235	Original Creator
9785	Meitantei Conan: Conan vs. Kid - SHARK & JEWEL	https://anilist.co/anime/9785	Original Creator
9786	Meitantei Conan: Conan vs. Kid - Shikkoku no Sniper	https://anilist.co/anime/9786	Original Creator
9963	Meitantei Conan: Chinmoku no Quarter	https://anilist.co/anime/9963	Original Creator
10531	Meitantei Conan MAGIC FILE 2011: Niigata - Tokyo Omiyage Capriccio	https://anilist.co/anime/10531	Original Creator
10703	Meitantei Conan: London kara no Maru Hi Shirei	https://anilist.co/anime/10703	Original Creator
10766	Meitantei Conan vs. Wooo	https://anilist.co/anime/10766	Original Creator
12117	Meitantei Conan: 11 Nin-me no Striker	https://anilist.co/anime/12117	Original Creator
13837	Meitantei Conan BONUS FILE: Fantasista no Hana	https://anilist.co/anime/13837	Original Creator
13839	Meitantei Conan: Excalibur no Kiseki	https://anilist.co/anime/13839	Original Creator
14735	Meitantei Conan: Sekkai no Private Eye	https://anilist.co/anime/14735	Original Creator
20546	Meitantei Conan: Ijigen no Sniper	https://anilist.co/anime/20546	Original Creator
20790	Magic Kaito 1412	https://anilist.co/anime/20790	Original Creator
20878	Meitantei Conan: Edogawa Conan Shissou Jiken - Shijou Saiaku no Futsukakan	https://anilist.co/anime/20878	Original Creator
21470	Meitantei Conan: Junkoku no Nightmare	https://anilist.co/anime/21470	Original Creator
21646	Meitantei Conan: Gouka no Himawari	https://anilist.co/anime/21646	Original Creator
98222	Meitantei Conan: Kara Kurenai Love Letter	https://anilist.co/anime/98222	Original Creator
22624	Meitantei Conan: Final "ONE" Shijou no Heiwa	https://anilist.co/anime/22624	Original Creator



Analyse via Application WEB – Alternative du live

Updating... x +

127.0.0.1:8050

Showing Media nodes

Remove Media outlier

Only Staff at least in X Media :

15

Node graph espacement:

1500

Media Name:

conan

Media found: Meitantei Conan

Related medias count: 63

id	title_userPreferred	siteUrl
235	Meitantei Conan	https://anilist.co/anime/235
779	Meitantei Conan: Tokei Jikake no Matenrou	https://anilist.co/anime/779
780	Meitantei Conan: Juuyonbanme no Target	https://anilist.co/anime/780
781	Meitantei Conan: Seikimatsu no Majutsushi	https://anilist.co/anime/781
1293	Urusei Yatsura	https://anilist.co/anime/1293

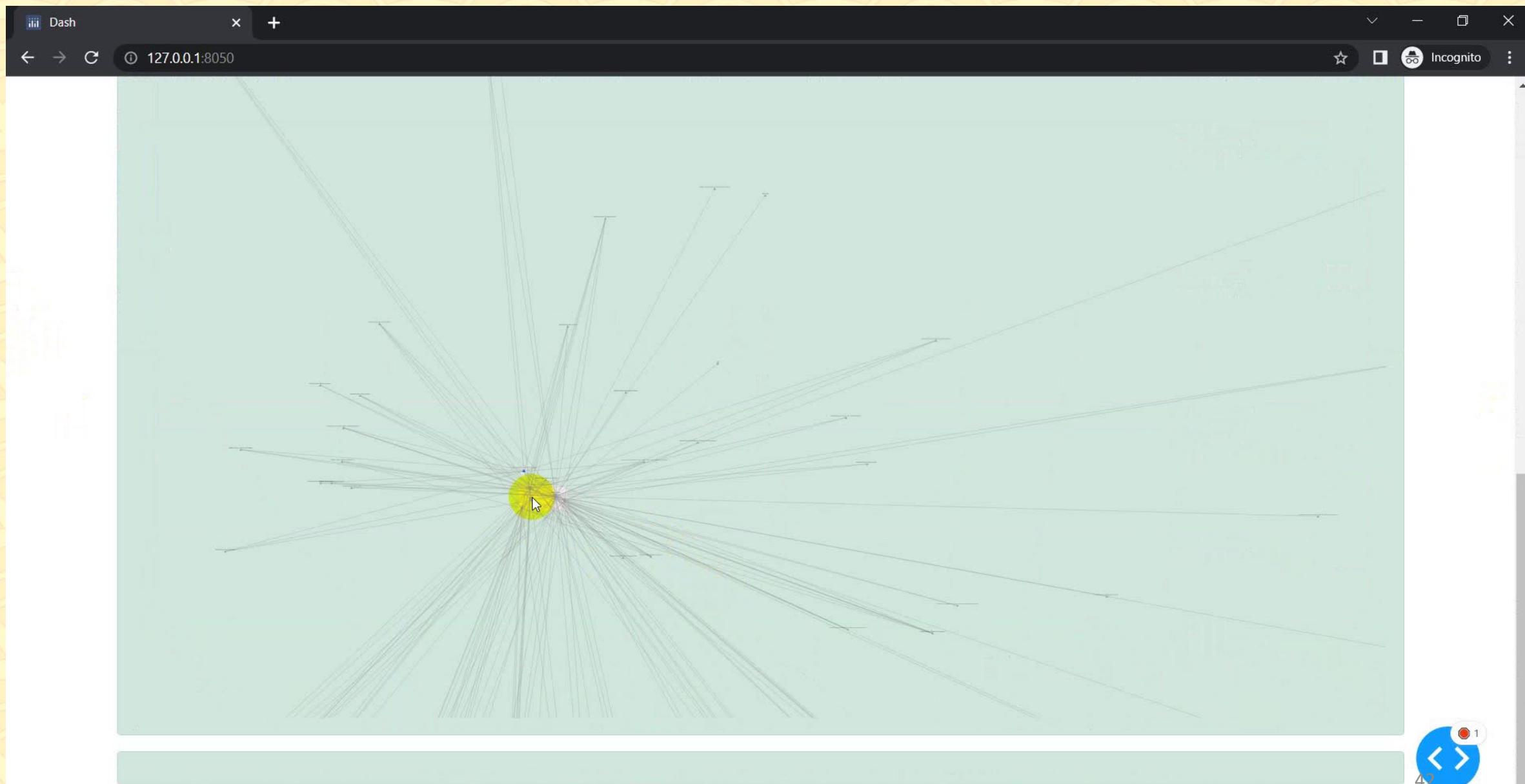
<< < 1 / 13 > >>

Magic Kaito

41

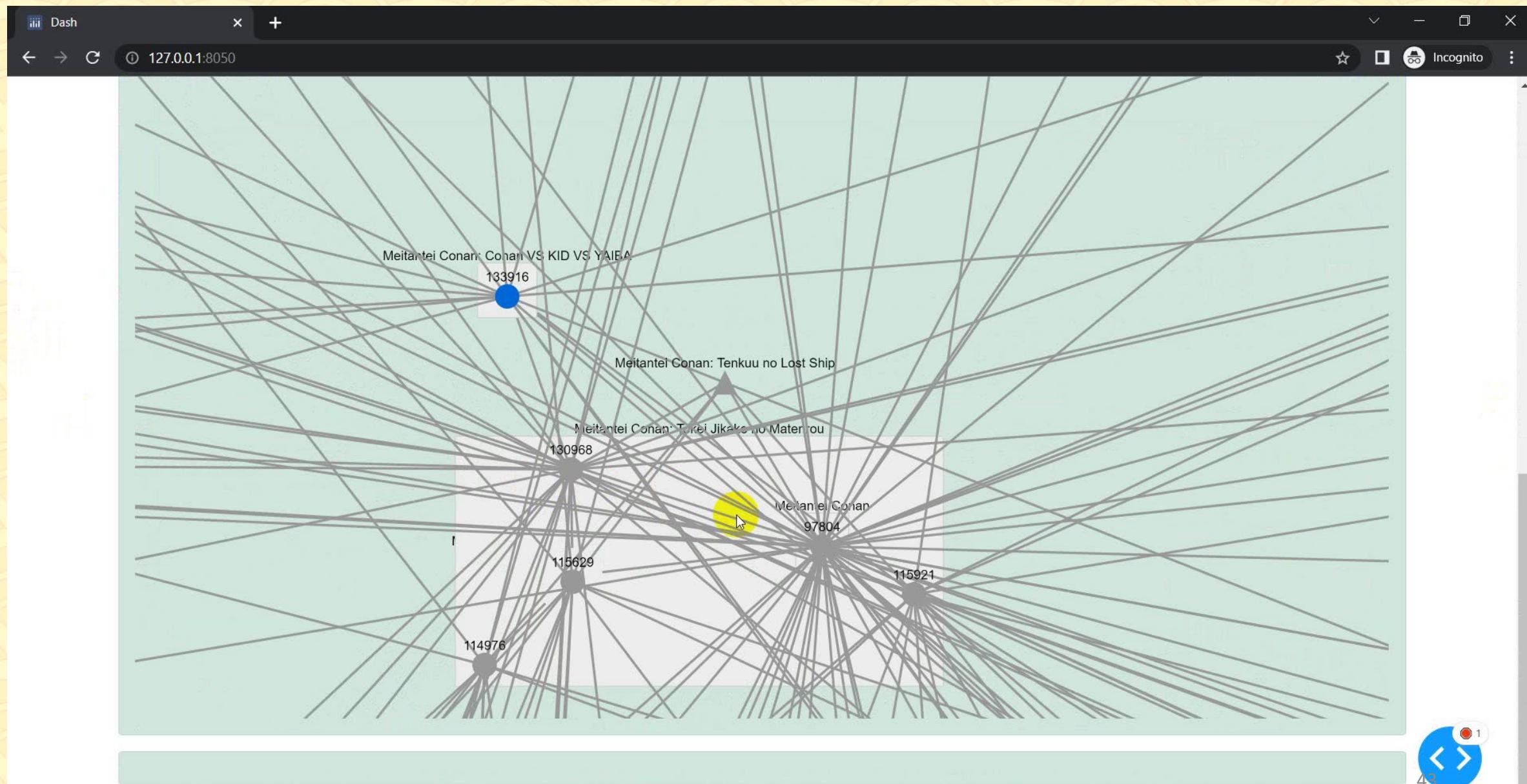


Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live





Analyse via Application WEB – Alternative du live

Nom	Participation / 49 médias	Rôle
Eiichirou Oda	46	Créateur d'origine
Kouhei Tanaka	19	Music
Inez Günther	14	Directeur ADR (allemand)
Shiro Hamaguchi	11	Music

Résultat des travailleurs restants "One Pièce" après le filtrage

漫画

05

Conclusion



Conclusion

Extraire des insights pertinents à partir de données AniList nettoyées

Apprendre de nouvelles technologies (Python, GraphQL, Power BI)

GitHub Repos

<https://github.com/Ahmad-Said/AniList-Analysis>

<https://github.com/RafaelVMoraes/Anilist-PowerBi-Analysis>

Screenshot of a GitHub repository page for "Ahmad-Said / AniList-Analysis". The repository is public and has 1 branch and 0 tags. It contains 11 commits from last week, all by Ahmad-Said, adding a Power BI report repository. The commits are listed under the main branch. The README.md file is present. On the right, there is an illustration of a character in a dynamic pose.

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

Ahmad-Said Add Power Bi report repository 68e530d last week 11 commits

- 1.Getting Started Initial commit
- 2.Mass Data Collection Initial commit
- 3.DataFrame Manipulations Initial commit
- 4.PowerBI Reports Add Power Bi report repository
- 5.Graph Reports Initial commit
- 6.DashApp Initial commit
- data Initial commit
- doc remove space in header doc
- x.ProjetDatabase TP Initial commit
- .gitattributes Initial commit
- .gitignore Initial commit
- LICENSE Initial commit
- README.md Add Power Bi report repository

README.md

AniList Analysis

Technologies Used

Technologie	Comments
Python	pandas, seaborn, plotly...
GraphQL	API Anilist

Introduction

The project aim for general analysis present in anime world. Data was collected from Anilist, a website for anime and manga enthusiasts, and analyzed using various statistical techniques.

漫画

END

Thank you!

