

# PROJET KAFKA

EXPLORATION DU JEU DE DONNÉES ORBITAAL POUR  
L'ANALYSE DES TRANSACTIONS BITCOIN

SAIBI MOHAMED  
MARZOUKI AHMED  
HAMIDOUNI EL-KAIM

ENCADRE PAR : COQUIDE CELESTIN



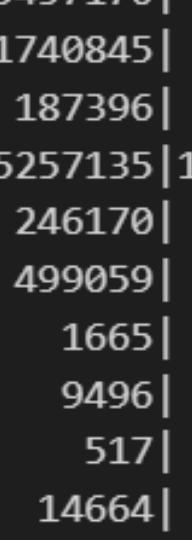
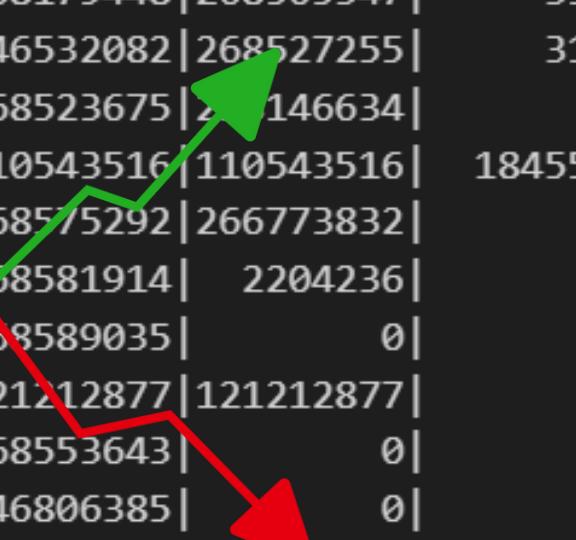
# Contexte & Motivation

- **Intérêt** pour la blockchain et le fonctionnement du Bitcoin
- **Comprendre** la création et le rôle du Bitcoin
- **Explorer** un marché financier décentralisé
- **Analyser** et interpréter le réseau via les données



# Objectifs du Projet

- Lecture de papiers de recherche sur le Bitcoin, son réseau et la théorie des graphes
- Constat : compréhension difficile du réseau pour les débutants
- Réalisation d'une analyse exploratoire du dataset
- Graphiques interactifs : top adresses, balances, répartition quotidienne
- Étude des acteurs principaux et de la concentration de richesse
- Analyse des transactions grâce aux théories des graphe pour identifier des communautés et les comparer
- Création d'un site web interactif pour aider les non-initiés à comprendre le fonctionnement du réseau Bitcoin



RC_ID	DST_ID	VALUE_SATOSHI	VALUE_USD
249381610	0	681	0.05593536
268469242	268496714	1931775	158.67037696
268179448	268505347	35457170	2912.34873344
246532082	268527255	31740845	2607.10055936
268523675	26851914	146634	187396
110543516	110543516	18455257135	1515858.9410115199
268575292	266773832	246170	20.21968768
268581914	2204236	499059	40.99125504
268589035	0	1665	0.13675825
121212877	121212877	9496	0.77977872
268553643	0	517	0.042454255
246806385	0	14664	1.20442736
268135976	0	1896	0.15569297
268584035	268584035	43848172	3601.173217280002
268590151	268537102	1000000	82.11654144
268461690	268464080	1469468617	120697.88188672
...			
268497420	32140711	6688900	549.40676096

only showing top 20 rows

# Données utilisées

## Dataset Orbitaal

Données structurées en :

- Graphes de flux (stream graph) au niveau des blocs
- Snapshots à différentes résolutions (jour, mois, année...)
- Tables de nœuds avec balances, adresses, et périodes d'activité

Données utilisées :

- Node Table :  
Informations sur les entités (balances, adresses, périodes d'activité)
- Snapshots journaliers :  
Visualisations classiques (top senders, top receivers, net balance...)
- Snapshots horaires :  
Visualisations en graphe (analyse de structure du réseau)

# Outils et technologies

## Analyse des données et Visualisation

Python  
Pyspark  
pandas // numpy  
plotly  
dash  
Matplotlib  
Jupyter

## Analyse de Graphe

Polars  
pyarrow  
Networkx  
igraph  
Scipy  
pyvis  
python-louvain  
datetime

## Developement WEB

Flask  
flask-cors  
HTML  
CSS  
JAVA SCRIPT

# Analyse Exploratoire et Etude des Acteurs

# Analyse Exploratoire & Richesse & Concentration

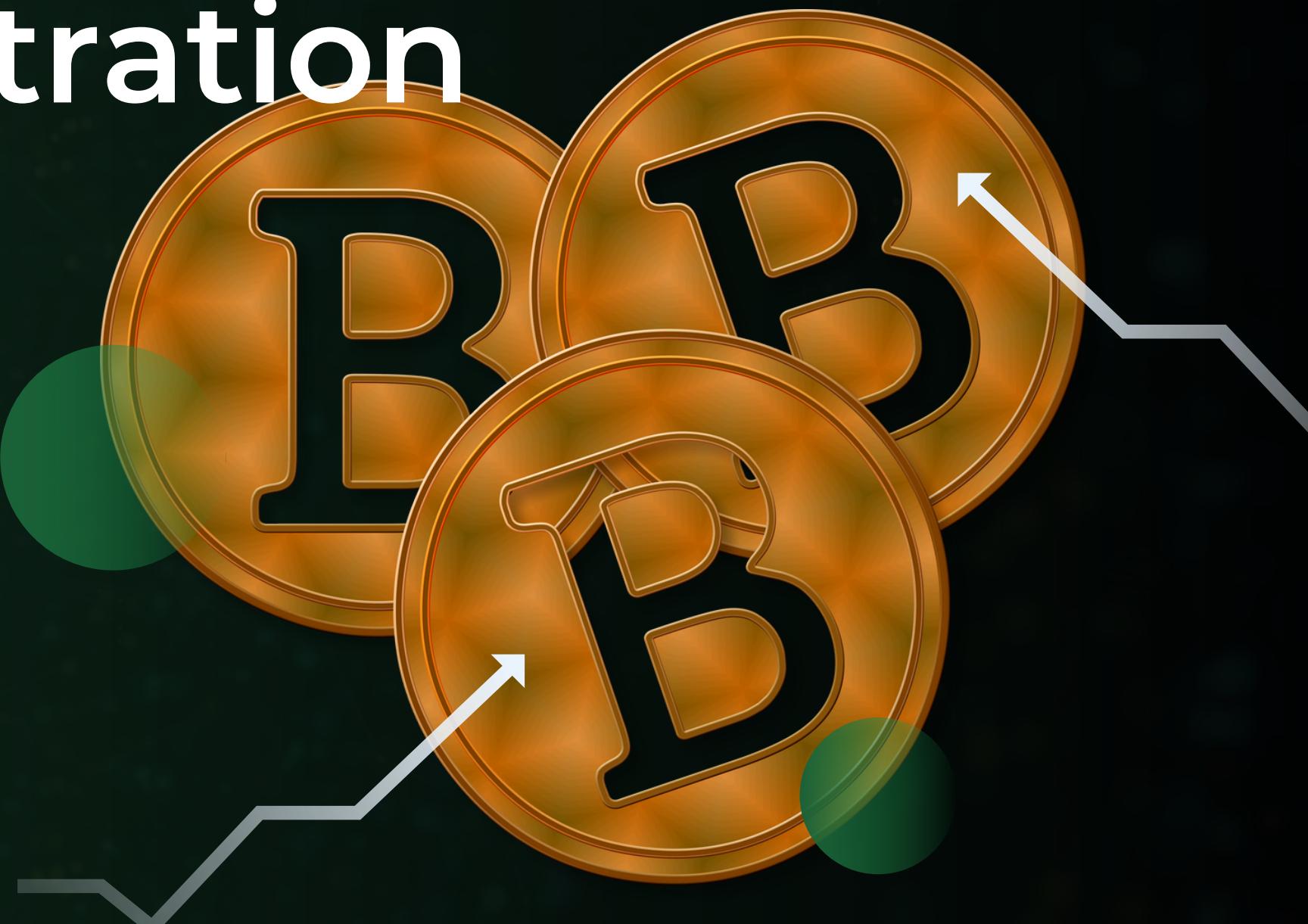
Objectif : analyser l'activité du marché via les top émetteurs/récepteurs et la répartition du bitcoin.

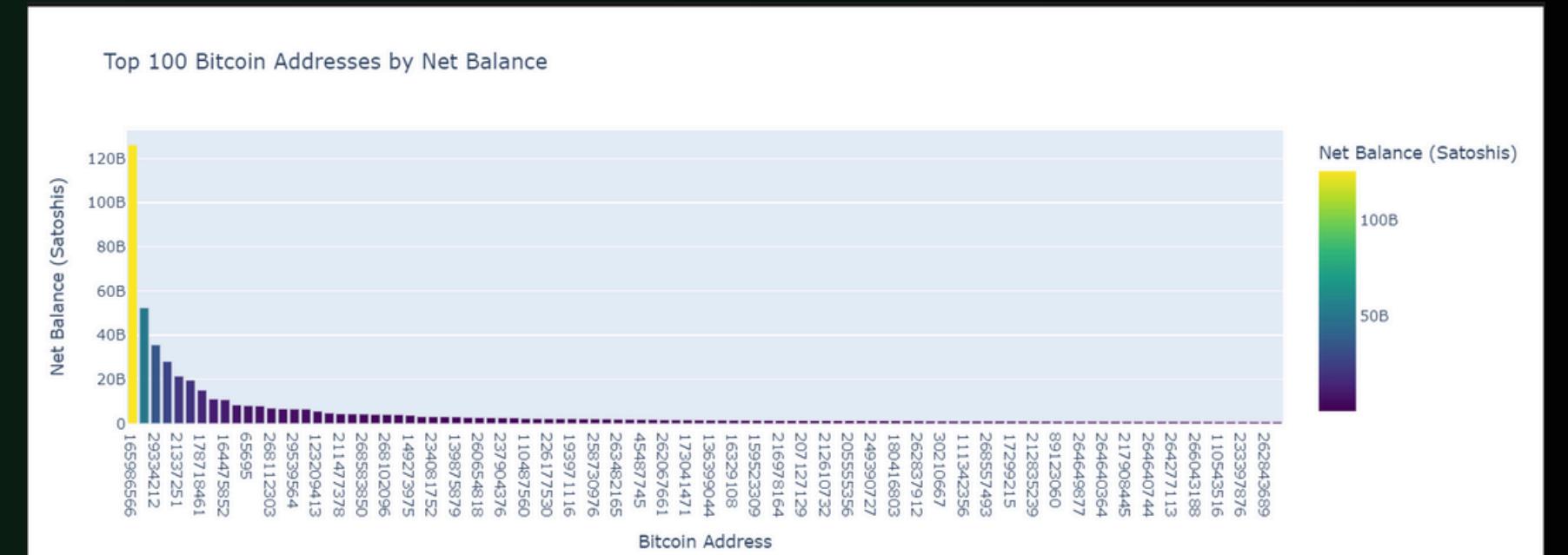
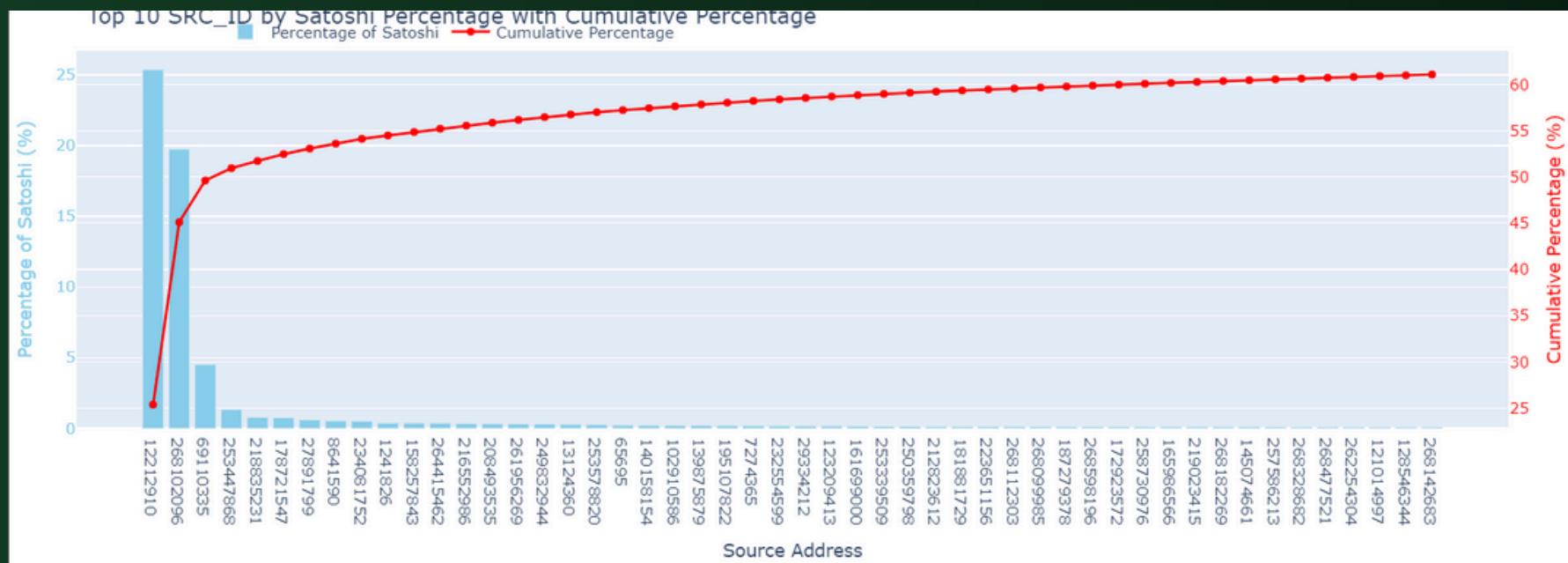
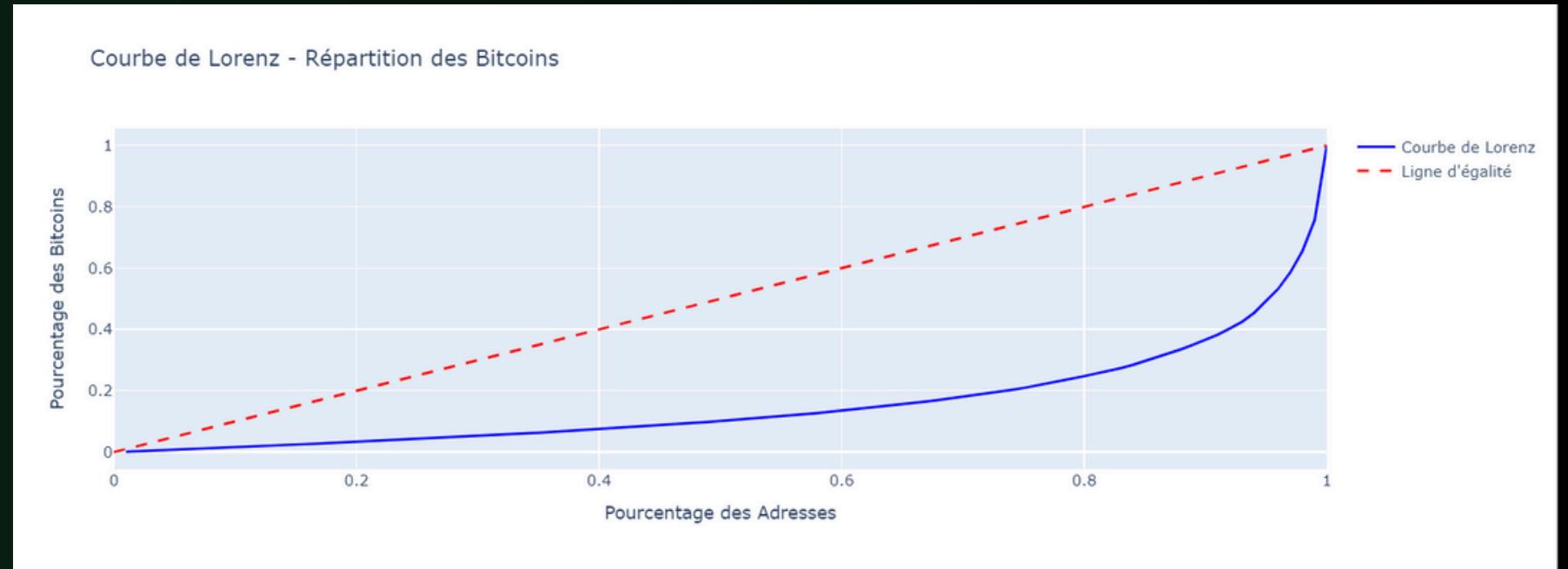
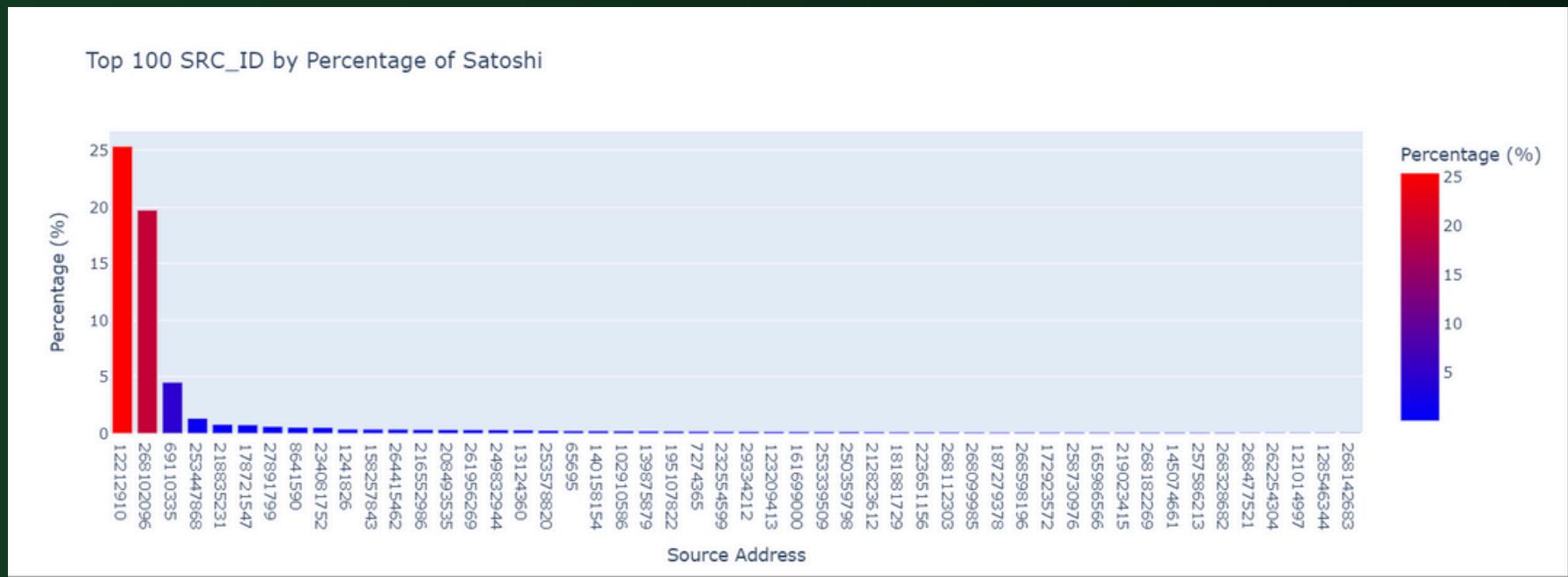
Étapes :

- Exclusion des mineurs pour ne garder que les transactions entre entités.
- Focus sur les 100 premières adresses (60 % des bitcoins détenus par les 15–20 premières).
- Création de visualisations interactives pour illustrer ces dynamiques.

Repere : Des Courbes Mathématiques

- Courbe de Lorenz + INDICE DE GINI





# Etude des Acteurs du Réseau

Objectif : Identifier les adresses et leur attribuer des étiquettes pour mieux comprendre les acteurs du réseau.

Étapes :

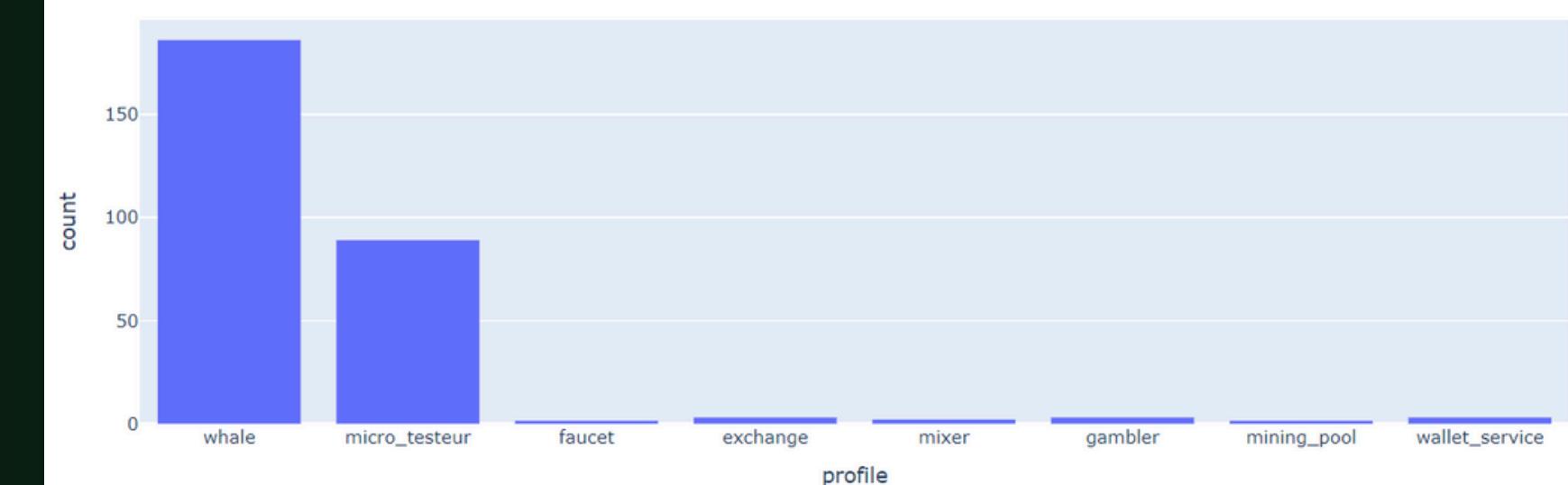
- Recherche des ID connues dans le fichier d'adresses (via WalletExplorer).
- Attribution de profils : Whales, Wallet Services, Drug Buyers, Gamblers, Faucets, etc.
- Analyse des transactions : Top utilisateurs, transaction moyenne, interactions entre profils, transactions atypiques.



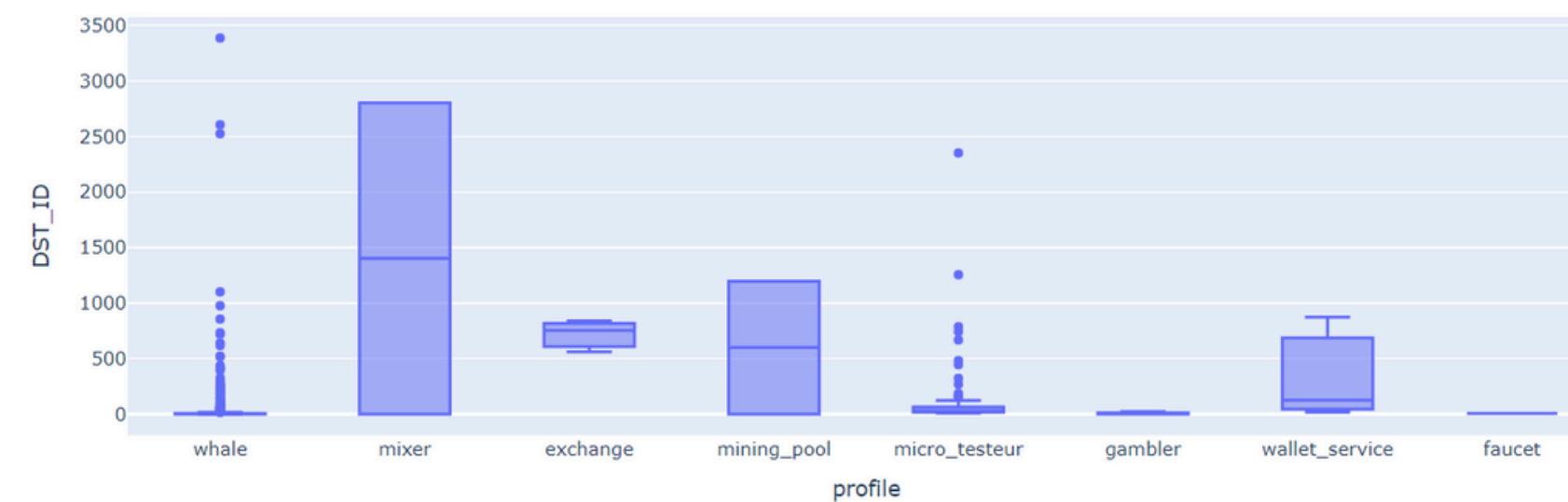
Valeur moyenne des transactions par profil (USD)



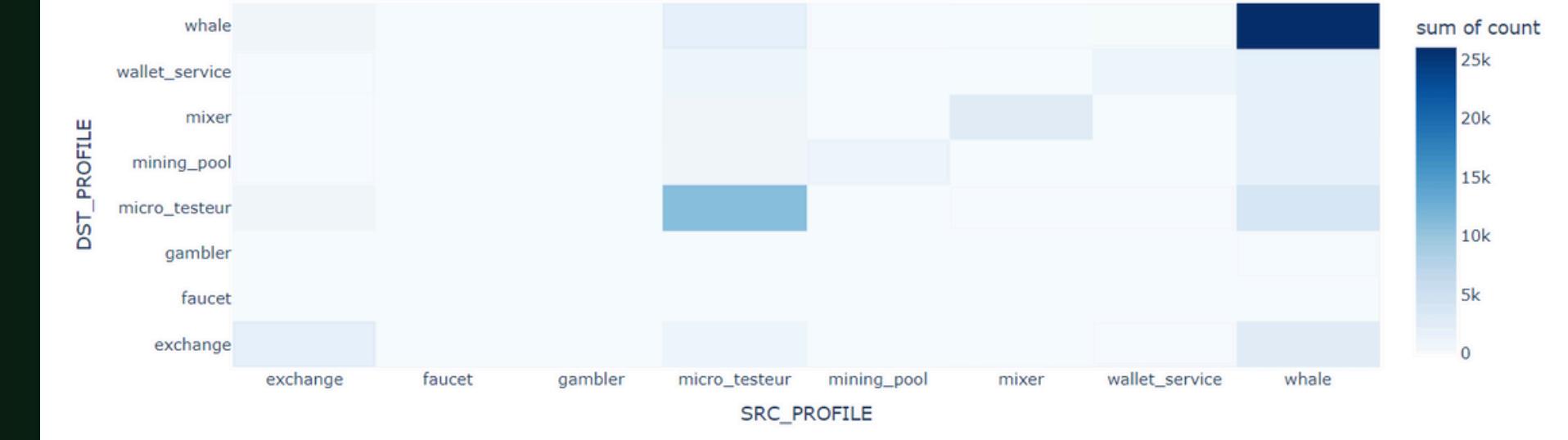
Nombre de transactions internes (SRC = DST) par profil

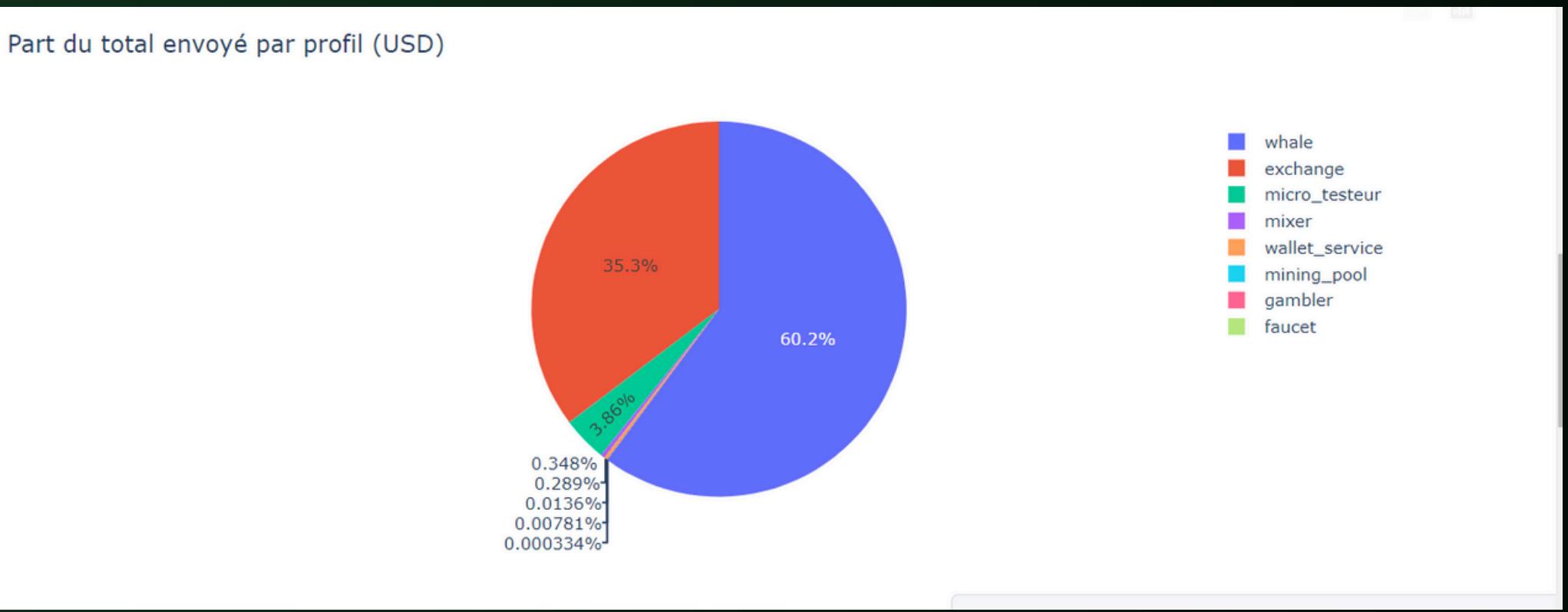


Nombre de destinataires uniques par adresse (par profil)

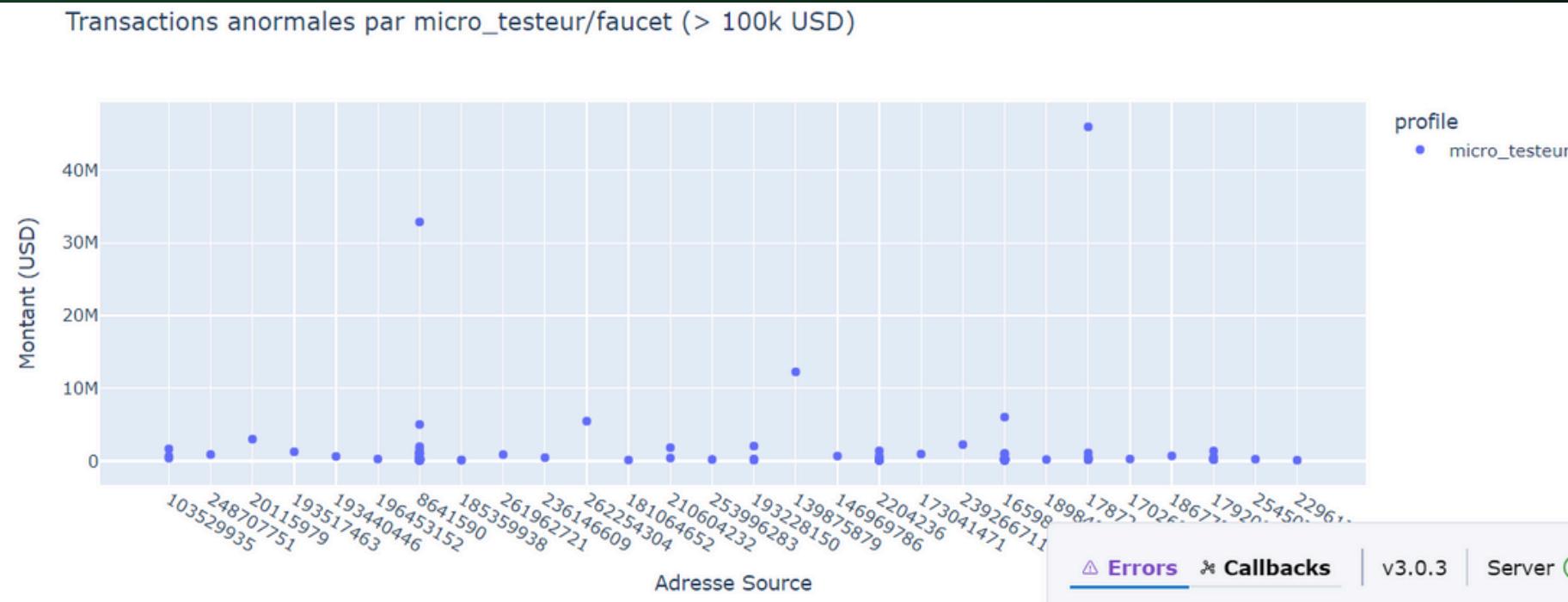


Interactions entre profils (SRC → DST)





Transactions anormales par micro\_testeur/faucet (> 100k USD)



Nombre de transactions par adresse, par profil



# Analyse Visuelle

# Analyse Visuelle



## VISUALISATION DES FLUX MONÉTAIRES:

Permettre aux utilisateurs de visualiser les transactions entre différents acteurs du réseau, identifiant ainsi les principaux flux de valeur

## IDENTIFICATION DES ACTEURS CLÉS:

Mettre en évidence les nœuds du réseau qui jouent un rôle central dans les transactions, qu'ils soient des gros dépensiers ou des gros récepteurs

## COMPARAISON TEMPORELLE

Offrir la possibilité de comparer les transactions sur différentes périodes pour observer les évolutions et les tendances au fil du temps

# Fonctionnalités Clés:

## ■ Visualization des transactions :

Les utilisateurs peuvent visualiser les transactions sous forme de graphes interactifs, où chaque nœud représente un acteur et chaque arête une transaction.

## ■ Détection des communautés

L'outil identifie automatiquement les communautés de transactions, permettant de comprendre les interactions entre groupes d'acteurs

## ■ Analyse des flux monétaires:

Les utilisateurs peuvent analyser les flux de valeur entre les acteurs, identifiant ainsi les principaux dépensiers et récepteurs

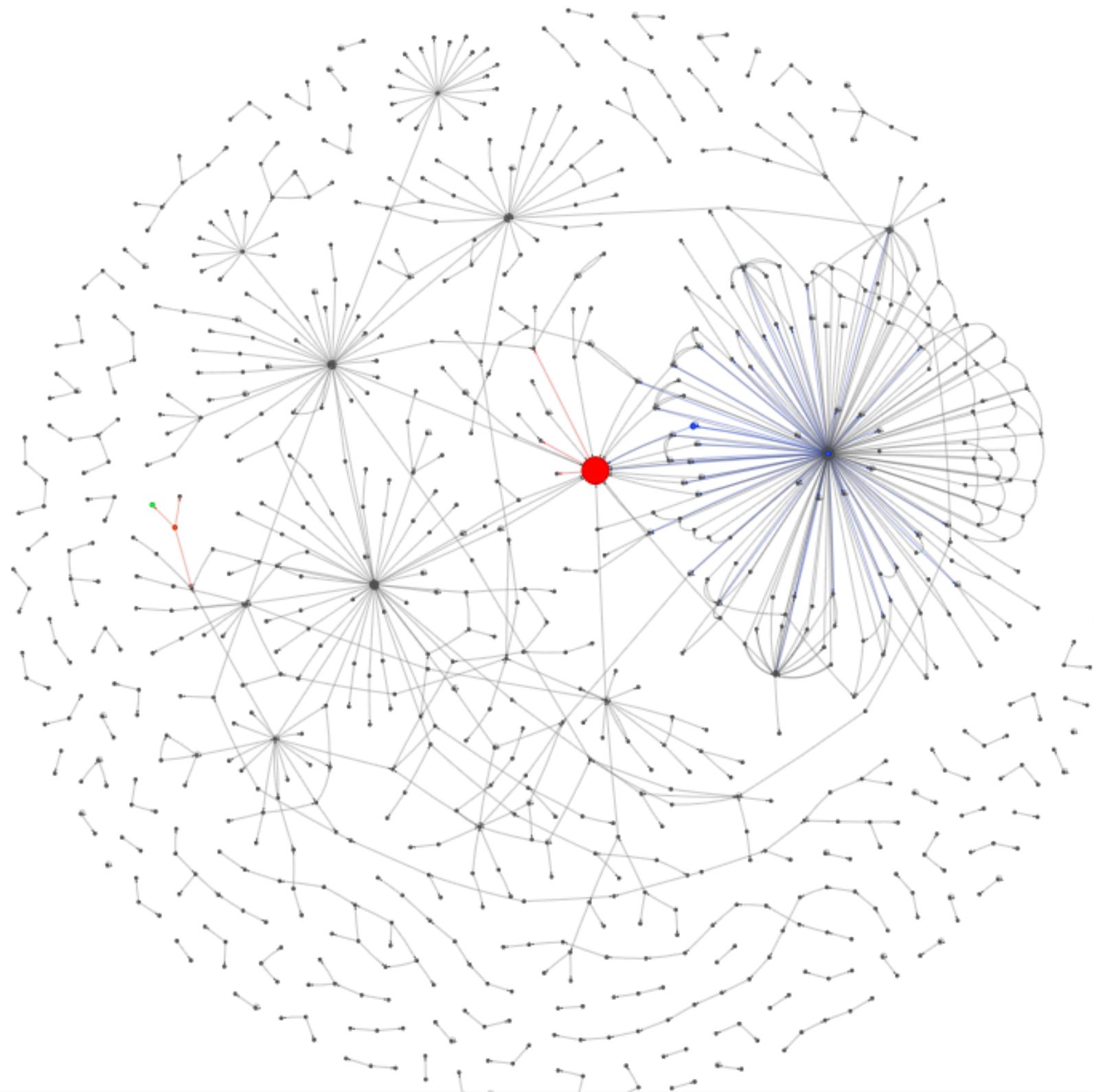
## ■ Comparaison Temporelle :

L'outil permet de comparer les transactions sur différentes périodes, mettant en évidence les évolutions et les tendances.

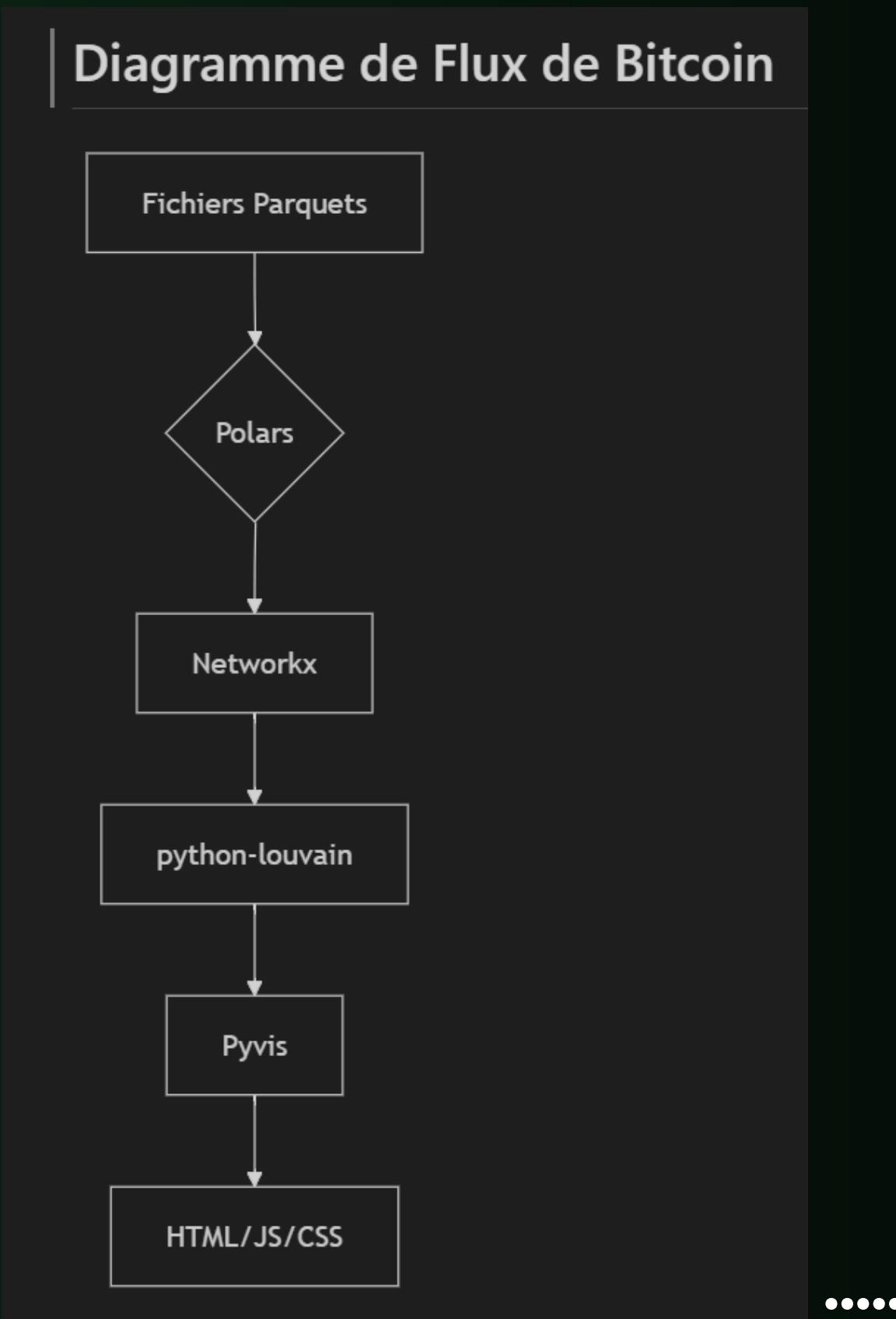
● Top Spenders

● Top Receivers

● Top Spenders & Receivers



**Et en terme de technologies:**



# Creation du Site Web



# Integration des resultats

## Creation du site Web

Preparation de la page de base en mettant en place un site web statique  
Page d'accueil  
Section d'analyse exploratoire  
Section d'analyse a base graphe

## Integrer les données

Reprise des resultats de mes collegues et les integrer en Site

# Mise en place d'un site Statique

Mise en place de l'arcihtecute

Faire le choix des technologies que je vais utilisé

Creation d'une page d'accueil avec des menu pour préparer l'arrivée de données analysé

HTML / CSS / JAVA SCRIPT /FLASK



K ProjetKafka

127.0.0.1:5000

Accueil Section 1 Section 2 Section 3

# Analyse du jeu de données ORBITAAL

Ce projet s'inscrit dans une thématique d'analyse de données massives (big data) dans le cadre des transactions Bitcoin. Dans le cadre de ce projet, nous travaillons sur la représentations de graphes dans différents types de résolutions.

80 %

13:10 22/04/2025 4

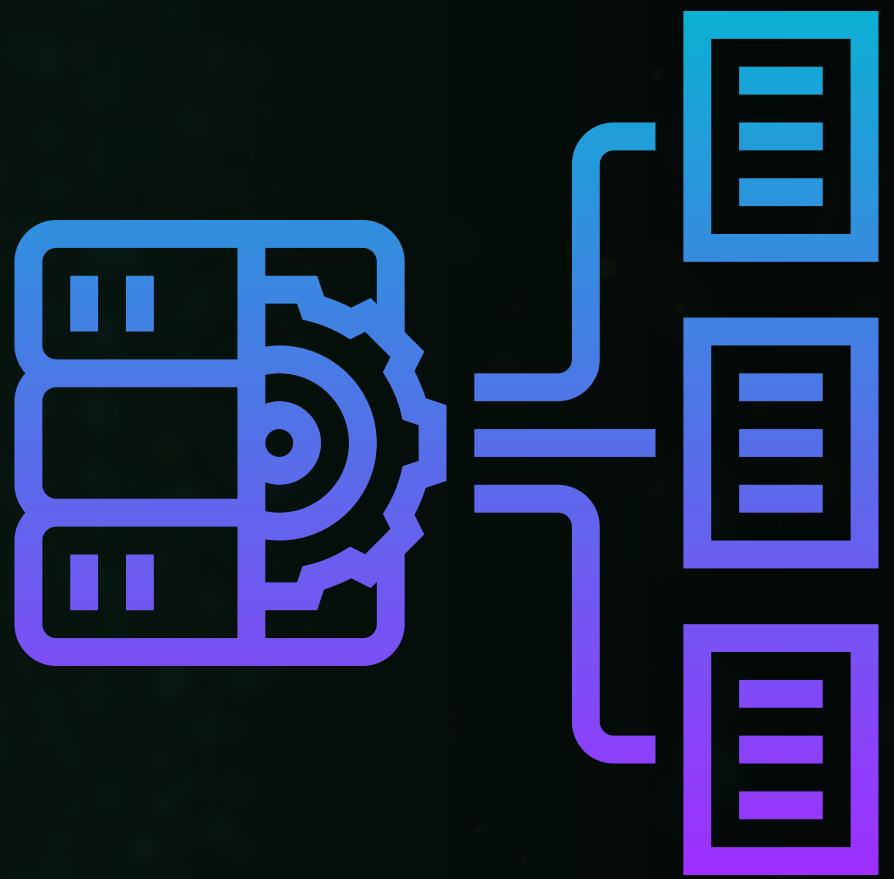
# Intégrer les données

Interaction entre les parties

- frontale (client)
- dorsale (serveur)



Gestion des parquets



# Gestion des parquets

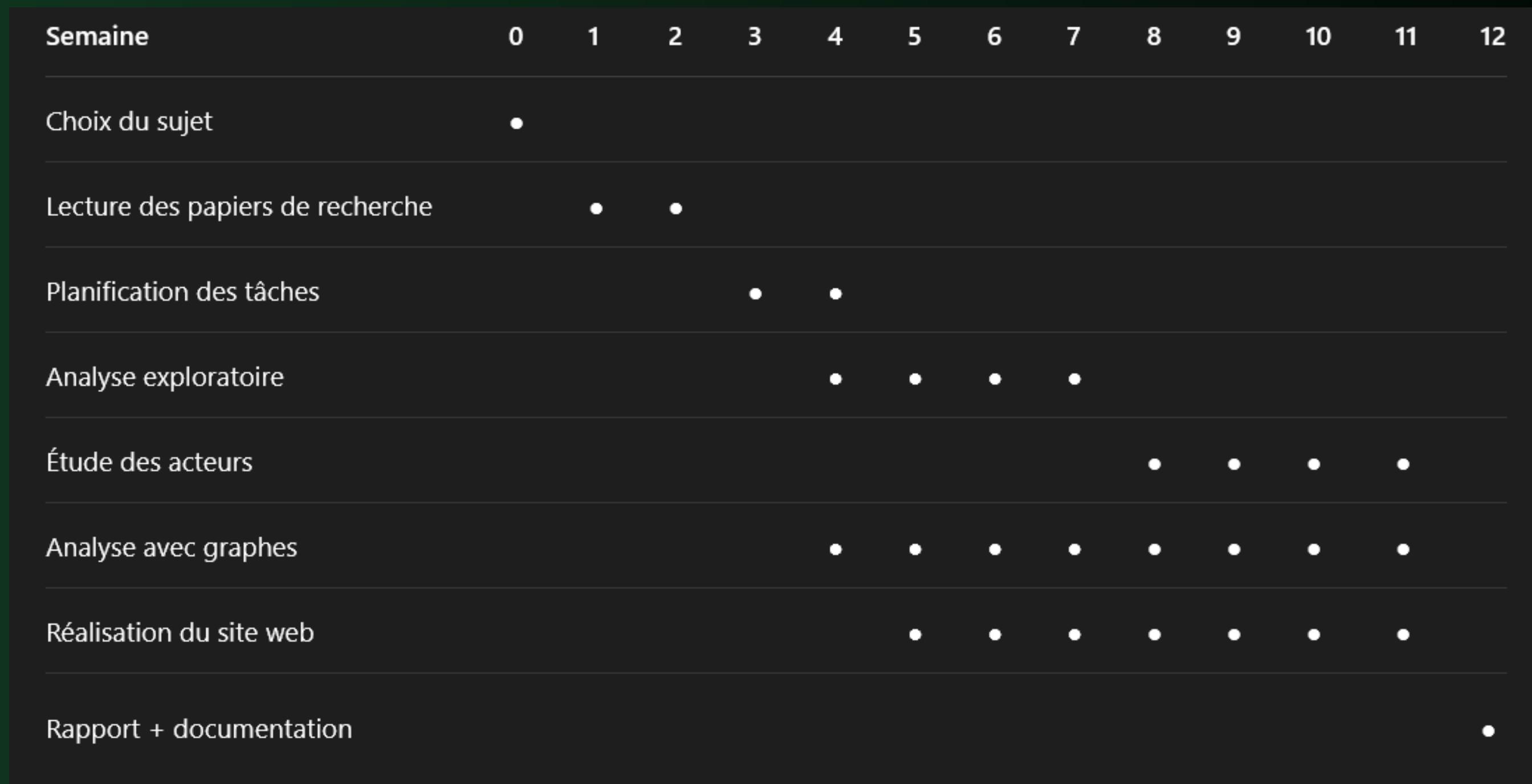
Creation de deux sections

Gestion des parquets

Interaction entre les parties

- frontale (client)
- dorsale (serveur)

# Déroulement du projet



# Succès mitigé



Non-respect total du cahier des charges

Trop d'objectifs fixés, souvent irréalisables

Difficulté à évaluer la charge réelle de chaque tâche



Volume de données

Performance

Difficulté d'aggregation



Integration des données

# Amélioration possible

Export des résultats

Analyse prédictive

Agrégation des analyses

Dashboard Intéractif

# Remercient Spéciale

Coquidé Célestin  
Cazabet Remy



# THANK YOU

