

**Université Hassan Premier**  
**Faculté des Sciences et Techniques**  
**Département Mathématiques et Informatique**

**Mémoire de Fin d'Etude**

Pour l'obtention du  
**Master en Sciences et Techniques**  
**Réseaux et Systèmes Informatiques**

*Conception d'une solution pour l'exploitation des données relatives  
aux entreprises en vue de faciliter la prise de décision*

**Présenté par : EL MALKI ZINEB**  
*Soutenu le 20 Juin 2023 devant les membres du jury composé de :*

<i>Professeur</i>	<i>Abderrahim Marzouk</i>	<i>FST Settati</i>	<i>Président</i>
<i>Professeur</i>	<i>Abdellah EZZATI</i>	<i>FST Settati</i>	<i>Encadrant</i>

**Année Universitaire : 2022-2023**



## DÉDICACE

C'est avec profonde gratitude et sincères mots,  
Que je dédie ce projet de fin d'études à

Mes chers parents,  
Qui n'ont jamais cessé, de formuler des prières à mon égard, de me soutenir, qui ont  
sacrifié leur vie pour ma réussite et m'ont éclairé le chemin par leurs conseils  
judicieux.

Mes chers sœurs,  
Pour ses soutiens moraux et leurs conseils précieux tout au long de nos études.

Toute ma famille  
Votre amour et votre soutien me sont indispensables.

Mes chers professeurs,  
Ceux qui se dévouent sans cesse pour nous éclaircir la voie et les immenses  
horizons du savoir et dont leurs efforts méritent largement nos respects, nous vous  
remercions énormément.

Mes chers amis et tous ceux qui ont collaboré de près ou de loin à l'élaboration de ce  
travail.

Je vous remercie du fond de mon cœur.

*EL MALKI ZINEB*

## REMERCIEMENT

Après avoir rendu grâce à DIEU le tout puissant et le miséricordieux, je tiens à remercier les personnes qui m'ont permis d'effectuer ce travail ainsi que ceux qui m'ont permis d'en faire un moment profitable.

Je commence d'abord par exprimer ma profonde gratitude envers mon encadrant interne Monsieur **EZZATI Abdellah** pour avoir acceptée d'encadrer mon projet de fin d'études.

Je tiens ensuite à exprimer mon remerciement à Monsieur **ALLA Said** pour m'avoir fait confiance et ainsi m'offrir la possibilité de vivre cette expérience professionnelle, pour ses qualités humaines, et aussi pour son accompagnement durant les étapes de l'élaboration du projet.

Je remercie aussi mon encadrant externe Monsieur **RHATOUS Abdessalam** pour leur aide afin de surpasser les difficultés rencontrées durant toute ma période de stage.

Je ne laisserai pas passer l'occasion sans remercier tout le personnel d'ISICOD pour leur accueil, aide, soutien et sympathie.

Mes remerciements vont aussi aux membres de jury Monsieur **EZZATI Abdellah** et Monsieur **Marzouk Abderrahim** pour avoir accepté de juger ce présent travail.

Veuillez trouver ici le témoignage de mon respect le plus profond.

Finalement, j'exprime mes vifs et sincères remerciements à toute personne ayant participé, de près ou de loin à la réalisation de ce projet.

Merci à tous et toutes.

## **AVANT-PROPOS**

Ce rapport a été réalisé dans le cadre d'un stage PFE de 6 mois au sein de l'entreprise ISICOD en vue de l'obtention d'un Master en Réseaux et Système Informatiques (MST-RSI) à la Faculté des Sciences et Techniques de Settat (FSTS).

Sur le plan académique, cette expérience est extrêmement enrichissante d'un point de vue d'apprentissage, tant sur le plan des connaissances théoriques et pratiques que sur les relations directes avec le personnel, qui nous a offert une sociabilité rapide et satisfaisante.

## RÉSUMÉ

Le présent document constitue le fruit de travail réalisé au sein de la société ISICOD dans le cadre d'un projet de fin d'étude, l'enjeu de ce projet est de développer une solution pour notre clientèle d'entreprise afin de créer une plateforme avec un aspect décisionnel à travers l'exploitation des diverses données disponibles dans la base de données du système de production, ainsi, l'extraction des données requises à partir des procès-verbaux, en outre, ressortir et comparer les différents indicateurs de performances des entreprises dans les marchés publiques et privées.

Notre démarche consiste à construire un système BI qui garantit une durée de vie importante et une performance élevées en manière de traitement des données. Cette solution vise à générer des rapports et des tableaux de bord prenant en considération l'ensemble des indicateurs implémentés, et permettant aux décideurs d'avoir une vision globale de leurs activités et les soutenir à améliorer la qualité de décisions dans leurs stratégies en intégrant l'intelligence artificielle.

Durant ce projet nous nous basons sur la plateforme offreonline.com et sa base de données qui est en place depuis 2012.

Le projet respecte les étapes du cycle de vie des données, La première phase concerne l'étude du système existant, et la compréhension des besoins, l'étape suivante est l'extraction des indicateurs et axes d'analyse pour aborder la conception et la modélisation, la séquence suivante est de nettoyer les données à travers Jupyter notebook ,L'alimentation a été effectuée en implémentant des mécanismes permettant l'audit et le contrôle des données via la pile ELK qui consiste à aborder la collecte, traitement et analyse, La dernière étape concerne l'exploitation de cette base de données pour réaliser les tableaux de bord en utilisant Kibana.

Mots-clés : système décisionnel, ELK, exploitation des données, DLM, indicateurs de performances des entreprises.

## SUMMARY

The present document is the outcome of work conducted within the company ISICOD as part of a final study project. The aim of this project is to develop a solution for our corporate client by creating a platform with a decision-making aspect through the utilization of diverse data available in the production system's database.

Furthermore, it entails extracting the essential data from the minutes and conducting an analysis to compare the performance indicators of companies operating in both public and private markets.

Our approach consists of building a BI system that ensures a long lifespan and high performance in terms of data processing. This solution aims to generate reports and dashboards that take into account all implemented indicators, enabling decision makers to have a comprehensive view of their activities and support them in improving the quality of their decision-making in their strategies by integrating artificial intelligence.

During this project, we rely on the platform [offreonline.com](https://offreonline.com) and its database, which has been in place since 2012.

The project follows the stages of the data lifecycle. The first phase involves studying the existing system and understanding the requirements. The next step is extracting the indicators and analysis dimensions to address the design and modeling process.

The subsequent sequence involves cleaning the data using Jupyter Notebook.

During the data ingestion process, we implemented mechanisms to ensure data auditing and control using the ELK stack. The ELK stack encompasses data collection, processing, and analysis, allowing us to effectively manage and monitor the data flow. In the final stage of the project, we utilize this database to create interactive and visually appealing dashboards using Kibana. These dashboards provide a comprehensive view of the data, enabling users to explore and analyze key insights in an intuitive and efficient manner.

Keywords: decision support system, ELK, data exploitation, DLM, company performance indicators.

## LISTE DES ABRÉVIATIONS

<i>Abréviation</i>	<i>Désignation</i>
BI	Business Intelligence
AO	Appel d'offre
DCE	Dossier de consultation des entreprises
PV	Procès-verbal
ELK	Elasticsearch, Logstash, et Kibana
DLM	Data lifecycle management
SQL	Structured Query Language
NOSQL	not only SQL
JSON	JavaScript Object Notation
SGBDR	système de gestion de Base de données relationnelles



## TABLE DES FIGURES

<b>Figure 1</b> - Domaines d'activités d'ISICOD -----	4
<b>Figure 2</b> - Clients d'ISICOD-----	6
<b>Figure 3</b> - Exemple d'une annonce d'un AO sur offresonline.com -----	7
Figure 4- Cycle de vie des données -----	9
<b>Figure 5</b> - interface Trello -----	10
<b>Figure 6</b> - Diagramme de GANTT -----	10
<b>Figure 7</b> - base NOSQL -----	14
<b>Figure 8</b> - Flux informationnel lié au processus ELK -----	15
<b>Figure 9</b> - Processus fonctionnel-----	17
<b>Figure 10</b> - Diagramme de bête à cornes -----	18
<b>Figure 11</b> - la suite ELK -----	25
<b>Figure 12</b> - jupyter notebook logo-----	27
<b>Figure 13</b> - Microsoft SQL server -----	27
<b>Figure 14</b> - logo Python -----	27
<b>Figure 15</b> - logo aqua data studio-----	27
<b>Figure 16</b> - modèle logique de l'activité des marchés-----	31
<b>Figure 17</b> -le schéma des documents -----	31
<b>Figure 18</b> -logstash pipeline-----	36
<b>Figure 19</b> - pipeline filter -----	36
<b>Figure 20</b> - pipeline output -----	37
<b>Figure 21</b> - pipeline output -----	37
<b>Figure 22</b> - les documents sur kibana -----	38
<b>Figure 23</b> - jupyter notebook-----	40
<b>Figure 24</b> - connexion avec Microsoft SQL Server -----	40
<b>Figure 25</b> - chargement des données-----	41
<b>Figure 26</b> - les données en dataframe-----	41
<b>Figure 27</b> - gérer les doublons -----	42
<b>Figure 28</b> - gérer les valeurs aberrantes -----	42
<b>Figure 29</b> - gérer les valeurs manquantes -----	42

<b>Figure 30-</b> gérer la forme des valeurs-----	43
<b>Figure 31-</b> Estimation financière par année pour une entreprise -----	43
<b>Figure 32-</b> Evolution des cautions par année pour une entreprise-----	44
<b>Figure 33-</b> le taux de sélectivité par année pour une entreprise-----	45
<b>Figure 34-</b> le taux de succès pour une entreprise en 2012-----	45
<b>Figure 35-</b> nombre des appels d’offres par secteur d’activité -----	46
<b>Figure 36-</b> Estimation financière par secteur d’activité -----	46
<b>Figure 37-</b> la moyenne des cautions par secteur d’activité-----	47
<b>Figure 38-</b> Nombre des appels d’offres par ville-----	47
<b>Figure 39-</b> Origine des échecs dans l’année 2020 -----	48
<b>Figure 40-</b> les secteurs d’activités pour la ville Agadir -----	48
<b>Figure 41-</b> Taux de sélectivité par type appel d’offre -----	49

## LISTE DES TABLEAUX

<b>Tableau 1-</b> Livrables du projet .....	11
<b>Tableau 2-</b> Axes d'analyse.....	19
<b>Tableau 3-</b> Description des clés .....	20
<b>Tableau 4-</b> Description des indicateurs .....	21
<b>Tableau 5-</b> Matrice de croisement Indicateurs/axes .....	23
<b>Tableau 6-</b> Les critères des outils technologiques.....	26
<b>Tableau 7-</b> Description de la table MARCHES .....	32
<b>Tableau 8-</b> Description de la table LOTS.....	33
<b>Tableau 9-</b> Description de la table VILLE .....	33
<b>Tableau 10-</b> Description de la table PAYE.....	34
<b>Tableau 11-</b> Description de la table Cat_organismes .....	34
<b>Tableau 12-</b> Description de la table organismes .....	34
<b>Tableau 13-</b> Description de la table secteur_activite.....	34

# TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE.....	1
CHAPITRE 1.....	3
1. Présentation de l'organisme d'accueil.....	4
1.1 Description .....	4
1.2 Métiers d'ISICOD .....	4
1.3 Références .....	6
2. Présentation du client : Offresonline .....	6
3. Présentation du projet .....	8
3.1 Problématique .....	8
3.2 Objectifs attendus .....	8
4. Planification et conduite du projet .....	8
4.1 Cycle de vie des données .....	8
4.2 Gestion de projet .....	9
4.3 Diagramme de GANTT .....	10
4.4 Livrables .....	10
5. Conclusion .....	11
CHAPITRE 2.....	12
1. Définitions .....	13
2. Modélisation .....	14
2.1 Base NOSQL orientée clé valeur.....	14
2.2 Base de données relationnelle .....	15
3. Processus d'un projet ELK .....	15
4. Conclusion .....	15
CHAPITRE 3.....	16

<b>1. Analyse et expression des besoins</b>	<b>17</b>
<b>1.1 Description de l'existant</b>	<b>17</b>
<b>1.2 Expression des besoins</b>	<b>18</b>
<b>2. Cadrage fonctionnel</b>	<b>19</b>
<b>2.1 Description des axes d'analyse</b>	<b>19</b>
<b>2.2 Recensement des indicateurs</b>	<b>21</b>
<b>2.3 Matrice de croisement Indicateurs/axes</b>	<b>23</b>
<b>3. Spécifications techniques</b>	<b>25</b>
<b>4. Conclusion</b>	<b>28</b>
<b>CHAPITRE 4</b>	<b>29</b>
<b>1. Modélisation</b>	<b>30</b>
<b>1.1 Le modèle logique des données</b>	<b>30</b>
<b>1.2 Le schéma des documents</b>	<b>31</b>
<b>2. Dictionnaire de données</b>	<b>32</b>
<b>2.1 Les tables de l'activité des marchés</b>	<b>32</b>
<b>3. Nettoyage des données</b>	<b>35</b>
<b>4. La collecte et le traitement des données</b>	<b>36</b>
<b>5. La visualisation des données</b>	<b>37</b>
<b>6. Conclusion</b>	<b>38</b>
<b>CHAPITRE 5</b>	<b>39</b>
<b>1. Nettoyage de données</b>	<b>40</b>
<b>2. Restitution des données</b>	<b>43</b>
<b>2.1 Présentation des indicateurs générales relatifs aux entreprises</b>	<b>43</b>
<b>2.2 Présentation des indicateurs générales relatifs aux marchés</b>	<b>46</b>
<b>CONCLUSION GÉNÉRALE</b>	<b>50</b>
<b>BIBLIOGRAPHIE/WEBOGRAPHIE</b>	<b>51</b>

## INTRODUCTION GÉNÉRALE

Les modes de passation d'un marché public ou privé contiennent la passation suite à des appels d'offres restreints, ouverts, avec présélection ou après un examen.

L'appel d'offre constitue une méthode par laquelle un commanditaire sélectionne l'entreprise chargée d'effectuer l'exécution de travaux, la livraison des fournitures ou la réalisation d'un service. L'objectif est de susciter la concurrence entre plusieurs entreprises en vue de fournir un produit ou un service.

L'activité des marchés est un océan rouge où la concurrence bat son plein, et chaque information peut donner un avantage concurrentiel qui aidera l'entreprise à remporter le marché. Et comme étant un secteur où la transparence est primordiale, les AOs et leurs résultats sont d'accès public. On se retrouve donc avec tout un historique des données accumulées sur des années mais non utilisées jusqu'à aujourd'hui.

Ces données sont des mines d'or d'information pour les entreprises si elles sont bien traitées, d'où ce projet qui consiste à développer une solution décisionnelle pour notre plateforme en se basant sur la plateforme « offresonline » qui réalise une veille sur tous les appels d'offres publics ou privés lancés. Ceci permettra aux dirigeants des entreprises adhérentes d'en tirer des décisions fondées sans pour autant avoir une connaissance préalable au domaine de l'analyse des données.

Mon champ d'intervention se focalisera donc en premier temps à bien comprendre le marché des appels d'offres au Maroc pour pouvoir conceptualiser des indicateurs métiers pertinents, ensuite sur le nettoyage et l'analyse des données et sur le développement des rapports permettant une prise de décision fondée sur des données.

Ce rapport présente l'ensemble des phases suivies lors de la réalisation du projet. Il est ainsi organisé en cinq chapitres.

Le premier chapitre présente le contexte général du projet, il porte sur la présentation de l'organisme d'accueil Isicod. Ensuite présente le client de la solution, la description du projet et à la fin l'approche adoptée pour la conduite du projet.

Le deuxième chapitre présentera les notions et les techniques d'informatique décisionnelle utilisées.

Le troisième est consacré à la phase d'analyse et des spécifications, il présente les besoins de l'utilisateur final en les analysant pour en dégager les mesures et les axes d'analyse, et finalement il présente les outils technologiques utilisés dans la réalisation du projet.

Le quatrième chapitre porte sur la modélisation du système décisionnel et l'étude technique.

Quant au cinquième chapitre, il présente la mise en œuvre de la solution en se basant sur la phase d'analyse et de conception. En premier lieu, la phase de nettoyage des données est présentée, puis la mise en œuvre du reporting est abordée.

## **CHAPITRE 1**

---

### **CONTEXTE GÉNÉRAL DU PROJET**

Ce chapitre est consacré au contexte général du projet en présentant en premier lieu l'organisme d'accueil, ses métiers et ses références. La deuxième section est dédiée à la présentation du client. Ensuite la description du projet, sa problématique et ses objectifs ainsi que la présentation de l'approche adoptée pour la conduite du projet.



## 1. Présentation de l'organisme d'accueil

### 1.1 Description

*ISICOD MAROC*, Ingénierie des Systèmes Informatiques et de Formation Conception Organisation et Développement. Est une société marocaine dont la mission est de comprendre les besoins de ses clients et de leur apporter les opportunités présentées par les nouvelles technologies de l'information et des télécommunications dans un environnement commercial complexe et changeant.

Créée en 2002, ISICOD compte une cinquantaine d'employés et son siège social est situé à Rabat.

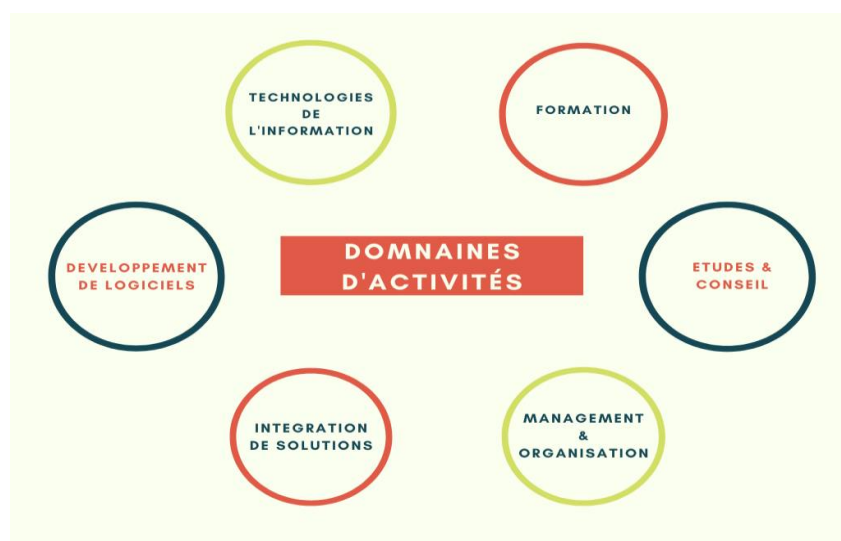
ISCOD a été créée pour aider à promouvoir le développement :

- Des entreprises marocaines, qu'elles soient privées ou publiques, en accompagnant leur transformation afin qu'elles restent performantes dans un monde économique ouvert et compétitif.
- Des partenaires étrangers dans le cadre du redéploiement, de la délocalisation et de la recherche d'efficacité et de compétitivité.

Partenaire officiel Darktrace, leader mondial de la cyber sécurité basée sur l'intelligence artificielle.

### 1.2 Métiers d'ISICOD

ISICOD est structurée en six domaines d'activité afin de répondre au mieux aux exigences de ses clients.



*Figure 1 - Domaines d'activités d'ISICOD*

## ***TECHNOLOGIES DE L'INFORMATION***

L'offre d'ISICOD couvre plusieurs prestations dans le domaine des technologies de l'information et de la communication :

- Conseil.
- Formation.
- Développement Web et mobile.
- Mise en place des solutions intranet.
- Développement de logiciels de gestion.

## ***DEVELOPPEMENT DE LOGICIELS***

La société accompagne ses clients dans la conception et le développement d'applications informatiques spécifiques en utilisant des outils et des moyens techniques appropriés pour assurer une productivité élevée et une bonne ergonomie.

## ***INTEGRATION DE SOLUTIONS***

ISICOD fournit à ses clients des solutions complètes qui correspondent à leurs besoins et à leurs stratégies. L'expérience de son équipe pluridisciplinaire lui permet de prendre en compte les circonstances techniques et économiques dans lesquelles s'intègrent les solutions proposées.

## ***FORMATION***

L'entreprise fait de la formation l'un des axes principaux de ses activités.

L'expérience pédagogique et la qualité des relations de ses conseillers lui permettent d'offrir une gamme complète de services.

## ***ETUDES & CONSEILS***

Grâce à la recherche et à l'analyse des besoins, les consultants et partenaires d'ISICOD peuvent déterminer la méthode de mise en œuvre du projet. Ils peuvent tenir compte de la spécificité de l'organisation cliente et de sa stratégie à moyen et long terme.

## ***MANAGEMENT & ORGANISATION***

À cette fin, ISICOD utilise une approche qui permet de considérer l'organisation comme un système complexe, dynamique et évolutif. Elle étudie également l'impact

des technologies de l'information, en particulier Internet, sur la performance des entreprises.

### 1.3 Références

ISICOD a pu mener à bien de multiples projets pour ses clients :



Figure 2- Clients d'ISICOD

## 2. Présentation du client : Offresonline

Le maître d'ouvrage du projet est la société Offresonline qui est une société marocaine dont le domaine d'activité principal est le commerce électronique et le commerce des matériels et services informatiques et électriques.

L'une des activités de la société est centrée autour de la plateforme « offresonline.com », qui a été lancée avec succès et qui fournit à ses clients une panoplie de services y compris :

- Service des appels d'offres :
  - Publics et privés : Offresonline réalise une veille de tous les appels d'offres publics et privés au niveau national et africain plus généralement. Les équipes cherchent, trient et classent la totalité des appels d'offres avec une mise à jour quotidienne pour permettre à ses adhérents une recherche facile, rapide et efficace.
  - Archives : Les appels d'offres sont mémorisés après expiration de la date limite afin de permettre de consulter, et à tous moments, les marchés clôturés.

- Résultats : Offresonline met à la disposition de ses adhérents la quasi-totalité des résultats des marchés en publiant le PV des séances publiques contenant toutes les informations nécessaires : concurrents du marché, concurrents évincés à cause du dossier administratif et technique, à cause de l'offre technique et à cause de l'offre financière, le résultat final du marché, etc.
- Programmes prévisionnelles : Les programmes prévisionnelles bénéficient les clients d'une longueur d'avance sur les concurrents en terme de préparation des marchés annoncés (contact / négociations fournisseurs, prévisions d'achat et gestion des stocks. . .).
- Service des ventes aux enchères : Ce service permet de consulter une base de données régulièrement actualisée des biens mis en ventes par l'État. De même que pour les appels d'offres, il est possible d'effectuer une recherche multicritère des avis de ventes aux enchères (activité, ville, organisme et date limite).

Détails d'appel d'offre N° 29AOM/SNRT/2023 publié le 21/04/2023

Ajouter aux favoris

1 :: ☒ NUMERO D'ORDRE : 1207911

**AVIS DE REPORT**

Type avis	Objet avis	Date avis
➔ Report	La date d'ouverture des plis est reportée du 24/05/2023 au 29/05/2023 12:30 ( consulter l'avis de report au niveau du DCE)	18/05/2023
➔ Report	la date d'ouverture des plis est reportée du 15/05/2023 au 24/05/2023 à 11h et les prospectus et documents techniques ne sont pas exigés dans le cadre de cet appel d'offre.	13/05/2023

[SOUSSION PAR VOIE ÉLECTRONIQUE](#)

REFERENCE	29AOM/SNRT/2023
ORGANISME	S.n.r.t
SOUS ORGANISME	Societe nationale de radiodiffusion et de television

**Vérification de la conformité et accompagnement à l'homologation de la sécurité des systèmes d'information en conformité avec loi n° 05-20 relative à la cybersécurité**

CAUTION	13 100,00 DH
CLASSIFICATION	
LOT	LOT UNIQUE
ADRESSE	DIRECTION DES ACHATS , SOCIÉTÉ NATIONALE DE RADIODIFFUSION ET DE TÉLÉVISION, SIÈGE RABAT
MONTANT DE RETRAIT	

*Figure 3 - Exemple d'une annonce d'un AO sur offresonline.com*

### **3. Présentation du projet**

#### **3.1 *Problématique***

Après avoir répondu et obtenu l'attribution d'un marché, les entreprises se trouvent souvent confrontées à une contrainte de temps pour évaluer adéquatement leur succès, identifier les causes des échecs, etc. Pourtant, ces indicateurs revêtent une importance cruciale pour assurer une gestion efficace et favoriser la croissance continue de l'activité de l'entreprise.

Dans le but d'offrir aux entreprises adhérentes une assistance et aide dans la prise de bonnes décisions stratégiques pour souscrire à un marché, et permettre aux responsables d'avoir une vue d'ensemble sur l'activité des marchés, Offresonline a jugé important de développer une solution pour l'aide à la décision.

#### **3.2 *Objectifs attendus***

Comme mentionné précédemment, le projet vise à développer d'un module décisionnel basé sur le calcul des indicateurs de performance spécifiques.

Un tel système d'information décisionnel permettra aux clients finaux de s'aligner avec leurs stratégies visant à remporter le plus grand nombre des marchés auxquels ils ont souscrit.

Concrètement, pour mesurer la performance d'un client on devra :

- Analyser ses propres résultats selon différentes dimensions.
- Mener des analyses relatives aux marchés.
- Mener des analyses par rapport à la concurrence.

Le projet a donc comme objectifs :

- La conception et l'alimentation d'un entrepôt de données pour la gestion de l'activité des marchés.
- La production des rapports et tableaux de bord décrivant cette activité.

### **4. Planification et conduite du projet**

#### **4.1 *Cycle de vie des données***

La gestion du cycle de vie des données (DLM) est une approche de la gestion des données tout au long de leur cycle de vie, de leur entrée à leur exploitation et utilisation. Les données sont divisées en phases en fonction de différents critères, et

elles passent d'une phase à l'autre à mesure qu'elles accomplissent différentes tâches ou répondent à certaines conditions.

Je vais aborder ce projet par cette approche car elle est basée entièrement sur les besoins des clients.

Techniquement, les données utilisées vont passer par le cycle de vie classique qui permet de les collecter, les traiter et analyser, les stocker, les archiver, et qui comporte à la fin une couche de présentation adaptée à une utilisation par des profils non informaticiens.

### CYCLE DE VIE DES DONNÉES

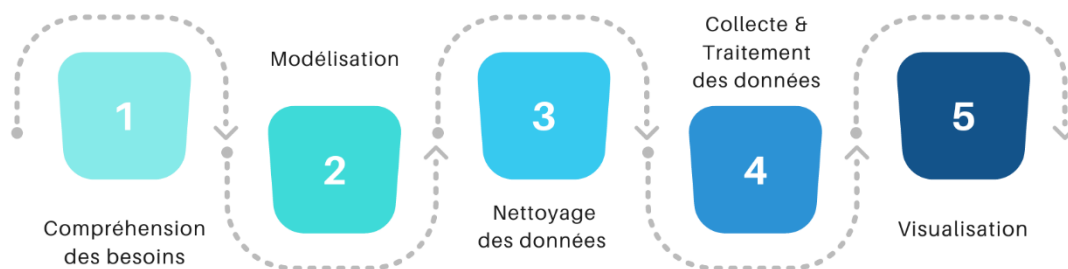


Figure 4- Cycle de vie des données

#### 4.2 Gestion de projet

Pour l'organisation et la gestion de projet, nous avons utilisé l'outil Trello.

Trello est un outil de gestion de projet en ligne, Il repose sur une collaboration des projets en planches listant des cartes, chacune représentant des tâches. Les cartes sont assignables à des utilisateurs et sont mobiles d'une planche à l'autre, traduisant leur avancement.

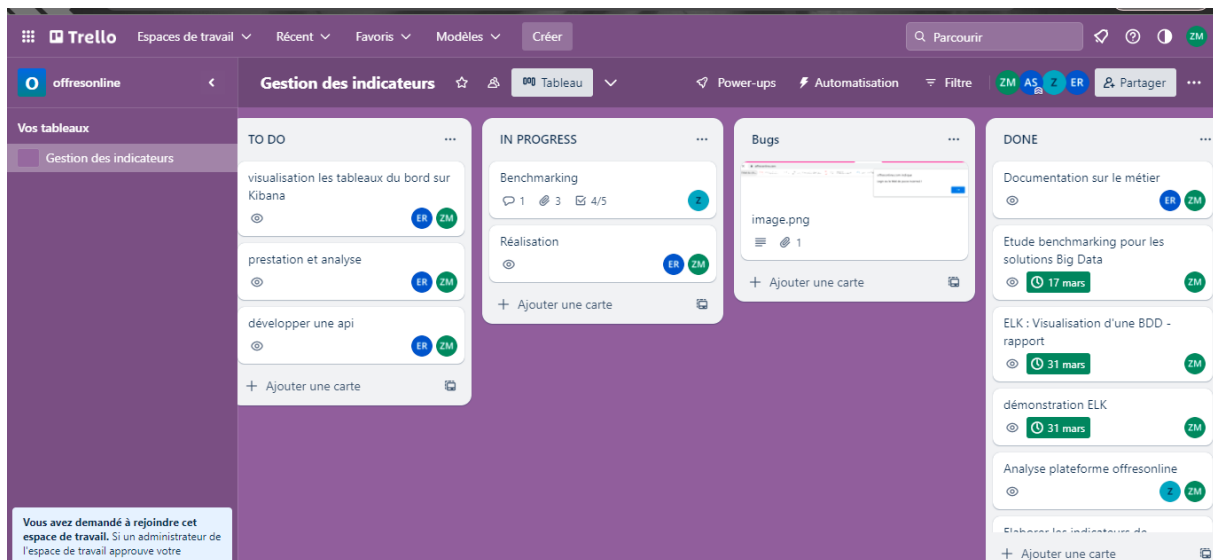


Figure 5- interface Trello

### 4.3 Diagramme de GANTT

Le diagramme de GANTT suivant présente les différentes phases du projet :

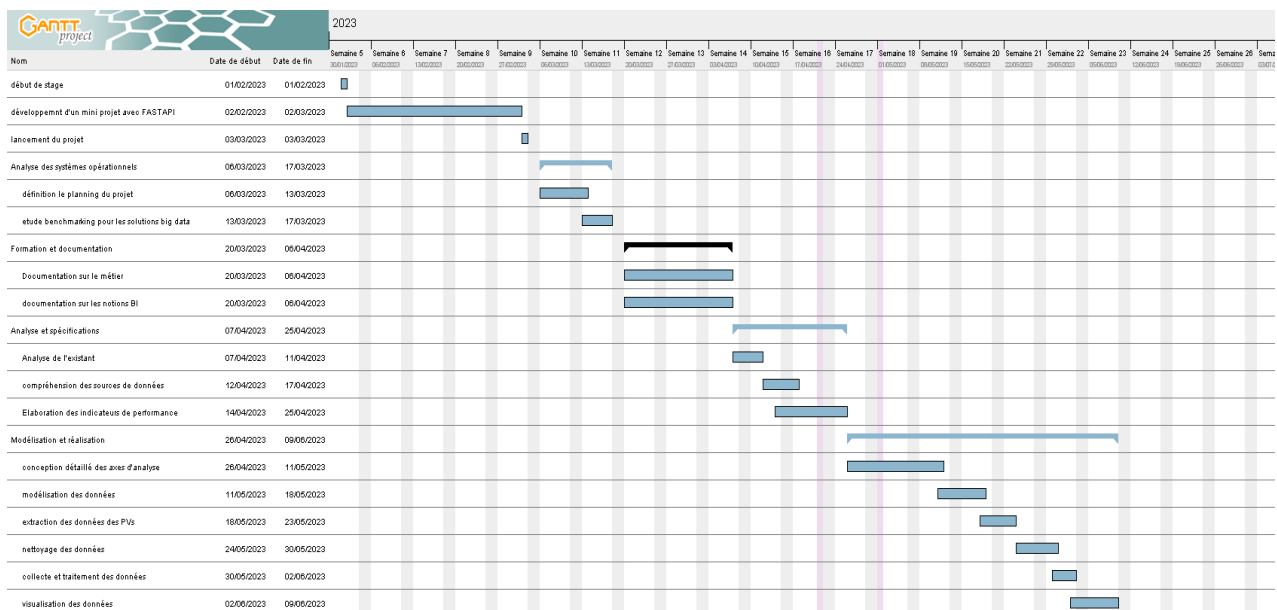


Figure 6- Diagramme de GANTT

### 4.4 Livrables

A la fin de chaque phase du projet, un ou plusieurs livrables sont établis pour mettre en valeur les résultats atteints, avoir une meilleure visibilité sur le projet et contribuer à la rédaction du rapport.

**Tableau 1- Livrables du projet**

Phase	Livrable
Lancement du projet	Planning du projet.
Formation et documentation	Compte rendu de la compréhension du sujet. Présentation : Etude benchmarking pour les solutions Big Data. Rapport : démonstration ELK et visualisation BDD.
Analyse et spécification des besoins	Document d'analyse et de spécification des besoins découlant de l'analyse et de l'expression des besoins par le client final.
Modélisation et extraction des données.	Matrice de croisement. BDD JSON extraite et générée à partir des PVs. Une représentation graphique de la structure des données engendrée à partir de la BDD de production.
Le nettoyage des données	BDD décontaminée et corrigée.
La collecte et le traitement des données	BDD traitée en se basant sur le calcul d'un ensemble des indicateurs.
visualisation	Tableau de bord

## 5. Conclusion

Dans ce chapitre, j'ai présenté un descriptif de l'organisme d'accueil. J'ai passé par la suite à la présentation du projet et finalement l'aspect organisationnel du travail.



## **CHAPITRE 2**

---

### **NOTIONS ET TECHNIQUES**

Ce chapitre présente les notions d'informatique décisionnelle utilisées, à savoir l'architecture décisionnelle, la modélisation et le processus d'un projet BI en utilisant la suite ELK.

## 1. Définitions

### *L'informatique décisionnelle (en anglais « Business intelligence »)*

C'est une discipline de l'informatique qui cible l'exploitation des données de l'entreprise afin de faciliter la prise de décision par les responsables, c'est-à-dire la compréhension du fonctionnement actuel et l'anticipation des actions pour un pilotage éclairé de l'entreprise.

Les outils et les méthodes de la BI vise à récupérer des données brutes (sources externes provenant des clients ou des fournisseurs, données de marchés, etc.), à les transformer en informations et à les diffuser sous forme de tableaux de bord ou reporting.

### *Nettoyage des données*

Le nettoyage de données est le processus qui consiste à identifier, corriger, supprimer ou transformer les données erronées, incohérentes, dupliquées, manquantes ou aberrantes présentes dans un ensemble de données. L'objectif du nettoyage de données est d'améliorer la qualité, l'intégrité et la fiabilité des données en vue d'une utilisation efficace dans des analyses, des modèles ou d'autres applications.

### *Collecte des données en temps réel*

Logstash est l'outil principal utilisé pour la collecte des données en temps réel avec ELK à partir de différentes sources telles que les bases de données en streaming, les flux d'événements, les messages des files d'attente, les fichiers de logs, les API, les systèmes de surveillance, etc.

### *Stockage de données*

Elasticsearch est utilisé comme moteur de recherche et de stockage des données. Il permet de stocker les données collectées de manière distribuée et d'effectuer des recherches et des requêtes rapides sur ces données. Elasticsearch utilise une indexation de type clé-valeur pour organiser et rechercher les données.

### *Traitement de données*

Lorsque les données sont stockées dans Elasticsearch, elles peuvent être traitées et analysées pour extraire des informations pertinentes. Cette étape peut impliquer l'utilisation de Kibana pour créer des visualisations, des tableaux de bord et des

graphiques interactifs afin d'explorer et d'analyser les données. Kibana offre des fonctionnalités avancées de visualisation et d'analyse des données, ce qui permet aux utilisateurs de prendre des décisions éclairées.

### ***Pipeline***

Un pipeline logstash est une séquence d'étapes de traitement des données, y compris l'entrée, le filtre et la sortie, qui sont appliquées aux données collectées.

### ***Indexation***

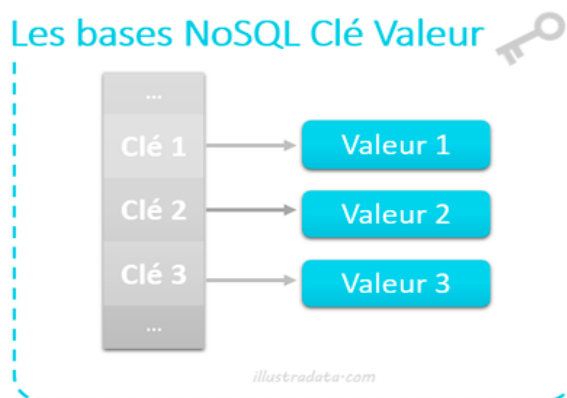
Elasticsearch organise les données en index, qui sont des collections logiques de documents.

## **2. Modélisation**

La modélisation des données est le processus de représentation graphique des flux de données. Lors de la création d'une nouvelle structure de base de données ou d'une structure alternative, le concepteur commence par créer un diagramme montrant comment les données entreront et sortiront de la base de données. Ce diagramme de flux sert à définir les caractéristiques des formats de données, des structures et des fonctions de gestion des bases de données afin de prendre en charge efficacement les besoins en flux de données.

### ***2.1 Base NOSQL orientée clé valeur***

Les bases NOSQL se basent sur le principe de stockage d'une valeur associée à une clé unique. La valeur associée à une clé peut être une simple chaîne de caractère comme un document, ou encore un objet beaucoup plus complexe pouvant contenir une multitude d'information.



***Figure 7- base NOSQL***

## 2.2 Base de données relationnelle

Dans un système de base de données relationnelles, les diagrammes sont utilisés pour représenter la structure et les relations entre les tables de la base de données. Ces diagrammes sont connus sous le nom de diagrammes de modèle relationnel ou de diagrammes de schéma relationnel.

### 3. Processus d'un projet ELK

Le cheminement depuis la donnée brute provenant des sources (BDD, fichier plat, ERP, CRM, etc.), à la production de tableaux de bord et de reportings, peut être faite de la manière suivante :



*Figure 8- Flux informationnel lié au processus ELK*

Logstash sera utilisé pour récupérer, filtrer et normaliser des informations liées aux données issues de différentes sources. Une fois cela fait, Logstash envoie ces informations dans un système de stockage ici Elasticsearch qui quant à lui se chargera d'indexer, stocker et effectuer une recherche et une analyse en temps réel de vos données. Enfin, Kibana récupère ces données afin de fournir un système de visualisation et d'exploration en plus de Logstash et Elasticsearch afin que vous puissiez facilement comprendre vos données sous forme de tableaux et de graphiques.

### 4. Conclusion

J'ai défini dans ce chapitre les notions et les technologies de la Business Intelligence employées dans mon projet. J'exposerai dans le chapitre suivant la phase d'analyse, spécifications des besoins.

## **CHAPITRE 3**

---

### **NOTIONS ET TECHNIQUES**

Ce chapitre a pour objectifs de présenter les besoins des utilisateurs finaux, de les détailler via l'analyse des besoins pour en dégager les différentes mesures et axes d'analyse et de présenter les outils technologiques utilisés.

## 1. Analyse et expression des besoins

Le client d'Isicod souhaite, par le biais de ce projet, combler une lacune importante dans la prise de décision. Ce manque se caractérise par l'absence de supports d'aide à la décision et de reporting pour les adhérents de la plateforme « Offresonline.com ».

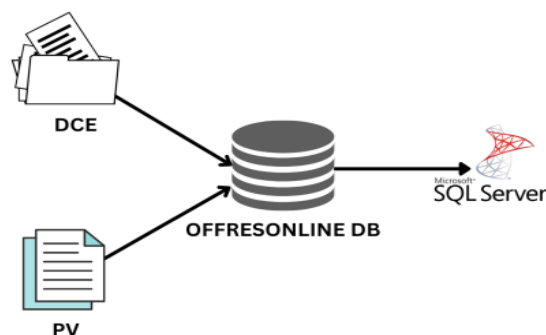
Partant de ce constat, une analyse complète de l'existant sera effectuée et présentée.

### 1.1 Description de l'existant

Cette partie présentera le principe fonctionnel général de tout le processus permettant à Offresonline de collecter les données importantes pour la mise en place du système décisionnel.

Offresonline dispose d'une plateforme qui collecte les avis des AOs lancés quotidiennement, également, les dossiers de consultation des entreprises contenant tous les documents lancés par le maître d'ouvrage et qui serviront à définir le marché (Avis d'appel d'offres, cahiers des prescriptions spéciales, plans et documents techniques, acte d'engagement, bordereau des prix et du détail estimatifs, bordereau du prix global, cadre du sous détail des prix, règlement de consultation...). Les informations importantes sont enregistrées dans la base de données OFFRESONLINE. Certaines informations sont communiquées aux adhérents et les autres sont historiées dans la BDD qui comporte 73 tables et qui est gérée par le SGBD SQL Server, et de laquelle les données utiles seront extraites en vue de les exploiter pour atteindre les objectifs désirés.

Après la conclusion du marché, les PV des séances publiques sont générés et stockés dans la BDD.



**Figure 9- Processus fonctionnel**

## 1.2 Expression des besoins

La première étape consiste à appréhender les exigences du client afin d'aboutir de manière méthodique à la solution, car un produit ne revêt pas de sens que s'il répond aux besoins pour lesquels il a été conçu.

Il convient donc de passer par l'étape d'expression des besoins dès le lancement du projet. Il s'agit d'explicitier l'exigence fondamentale qui justifie la conception d'un produit, pour cela, il faudra se poser les trois questions suivantes :

- A qui le produit rend-il service ? Il rend service bien évidemment aux décideurs des entreprises adhérentes.
- Sur quoi agit-il ? Il agit sur les données concernant les abonnés et les marchés.
- Dans quel but ?

Ces questions fondamentales ont abouti à la construction du diagramme de bête à cornes suivant :

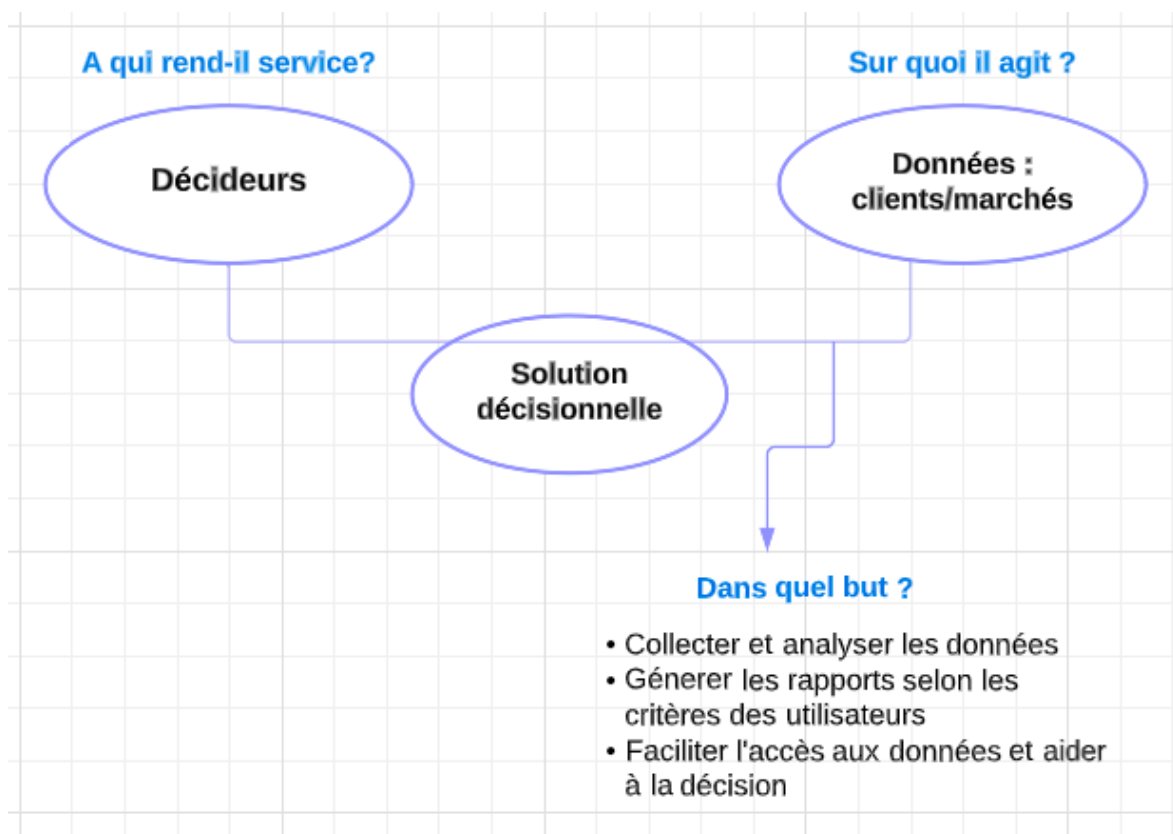


Figure 10- Diagramme de bête à cornes

Les besoins définis du client d'ISICOD sont donc :

- **Collecter et analyser les données**

Cette étape permettra de répondre aux besoins de facilité et de rapidité d'accès aux informations. Cette BDD constituera également une archive riche des données de la plateforme offresonline.

- **Générer les rapports selon les critères des utilisateurs**

Une fois les données stockées et les indicateurs sont chargés, les données seront ensuite publiées sous forme de tableaux de bord. Afin d'avoir des informations précises, les utilisateurs demandent la génération de rapports en précisant la distribution des indicateurs et des axes d'analyse sur les tableaux de bord.

- **Facilité l'accès aux données et aider à la décision**

Les tableaux de bord seront générés dans le cadre d'une application web et mobile pour aider à la décision, proposer des suggestions, et faciliter l'accès aux dirigeants des entreprises adhérentes qui sont les utilisateurs principaux de ce module décisionnel.

## **2. Cadrage fonctionnel**

Cadrer le projet sur le plan fonctionnel permet de traduire les besoins présentés précédemment en termes d'indicateurs, d'axes d'analyse et d'états de restitution.

Cette section concernera ces trois volets.

### ***2.1 Description des axes d'analyse***

***Tableau 2- Axes d'analyse***

Axe d'analyse	Description
Marchés	Il contient les éléments caractérisant le marché : identifiant et numéro de référence, objet de marché, estimation financière, caution,...etc.
Lots	Il représente une quantité groupée de produits ou de service. Contient l'identifiant et l'objet des lots, La date, la ville, pays de marché ...etc.



CAT_organismes	Il contient les éléments décrivant un organisme : identifiant, nom de l'organisme et type des appels d'offres lancés par l'organisme
Organismes	Il contient les éléments décrivant les sous organismes : identifiant, nom des sous organismes, le code.
Paye	Pays dans lesquels une entreprise opère ou envisage d'opérer sur le plan commercial.
Ville	Elle contient l'identifiant, le nom de la ville et la région
Secteur d'activité	Il représente l'identifiant et le nom du secteur d'activité du marché.

L'extraction des documents PVs a abouti à la création d'une BDD NOSQL orientée documents, organisée sous la forme de paires clé-valeur. Voici sa présentation :

*Tableau 3- Description des clés*

Clé	Description
Objet	Il fait référence à la description ou à la nature d'un contrat, d'une commande, etc.
Lots	La subdivision d'un marché en parties distinctes.
Lieu d'ouverture des plis	Il s'agit d'une salle de réunion ou d'un bureau administratif où les soumissionnaires peuvent assister à l'ouverture des plis.
Liste des concurrents ayant déposé un pli	La liste des entreprises ou des fournisseurs qui ont soumis une offre.
Liste des concurrents évincés à l'issue de l'examen des dossiers administratifs et technique	La liste des concurrents évincés à l'issue de l'examen des dossiers administratifs et technique.
Liste des concurrents admissibles sans réserve	La liste des concurrents admissibles sans réserve est établie après avoir passé en

	revue les dossiers administratifs et techniques de chaque soumissionnaire.
Liste des concurrents admissibles avec réserve	La liste des entreprises ou des fournisseurs qui sont considérés comme éligibles à poursuivre dans le processus d'évaluation.
Montant des actes d'engagement des concurrents	Montant financier spécifié par chaque concurrent.
Liste des concurrents écartés	Liste des entreprises ou des fournisseurs qui ont été exclus.
Concurrent retenu	L'entreprise ou au fournisseur qui a été sélectionné.
Justification du choix de l'attributaire	L'explication détaillée des raisons pour lesquelles un concurrent spécifique a été sélectionné.

## 2.2 Recensement des indicateurs

Afin de cerner les besoins du client, plusieurs réunions ont eu lieu. Ces dernières nous ont permis de déterminer 3 catégories d'indicateurs qui vont nous aider à tirer profit des données. Ces trois catégories sont les indicateurs propres à l'entreprise, ceux relatifs aux marchés et finalement des indicateurs relatifs aux offres publiées.

**Tableau 4- Description des indicateurs**

Indicateur	Description	Axe d'analyse
Les indicateurs propres à l'entreprise		
L'estimation financière au cours des années	L'estimation des coûts par année pour l'entreprise	Marchés, Cat_organismes, Organismes, Lots, la base de données Json
Evolution des cautions au cours des années	Les garanties financières des marchés par année pour l'entreprise	
Les gains au cours des années	Les gains par année de l'entreprise	
Taux de sélectivité par année	(Nombre des AOs soumissionnés)/(Nombre	

	des AOs lancés) par année	
Taux de succès par année	Nombre d'A.O gagnés / Nombre d'A.O soumis par année	
Les indicateurs relatifs aux marchés		
Nombre des appels d’offres par secteur d’activité	Nombre totale des marchés publiés pour chaque secteur d’activité	Marchés, Organismes, cat_organismes, Lots, Secteur_activité, Ville, Paye, la base de données Json
Estimation financière par secteur d’activité	L’estimation des coûts pour chaque secteur d’activité	
Tendance centrale des cautions par secteur d’activité	La moyenne des cautions pour chaque secteur d’activité	
Nombre des appels d’offres par ville	Nombre totale des marchés publiés pour chaque ville	
Origine des échecs	Pourcentage des échecs après l’examen des dossier administratifs et techniques	
Les secteur d’activités dans une ville donnée	les secteurs d'activités dominants dans une ville	
Taux de sélectivité par type appel d’offre.	Le taux de sélectivité pour les appels d’offres privés ou publiques	
Les indicateurs relatifs aux offres publiées		
Liste des concurrents ayant déposés un pli	Les concurrents ayant soumissionnés d’un offre	La base de données Json
Liste des concurrents évincés à l’issue de l’examen des dossiers	Les concurrents évincés à l’issue de l’examen des dossiers administratifs et techniques	

administratifs et techniques		
Montant des actes d'engagement des concurrents	Les coûts proposés par les différents concurrents	
Justification du choix de l'attributaire	Explication donnée par l'entité responsable d'un appel d'offre	

### ***2.3 Matrice de croisement Indicateurs/axes***

La matrice de croisement permet de regrouper les indicateurs et les axes d'analyse pour examiner les différentes relations et interactions, En croisant les indicateurs avec les différents axes d'analyses, nous obtenons la matrice figurante dans le tableau ci-dessous.

***Tableau 5- Matrice de croisement Indicateurs/axes***

Axes Indicateurs	Marchés	Lots	Ville	Secteur d'activité	Cat_organismes	Organismes	BDD Json
Les indicateurs propres à l'entreprise							
L'estimation financière au cours des années	X	X			X	X	
Evolution des cautions per année	X	X			X	X	
Les gains par année	X	X			X	X	X
Taux de sélectivité par année	X	X			X	X	

Taux de succès par année	X	X			X	X	
Les indicateurs relatifs aux marchés							
Nombre des appels d'offres par secteur d'activité	X			X			
Estimation financière par secteur d'activité	X			X			
Moyenne des cautions par secteur d'activité	X			X			
Nombre des appels d'offres par ville	X		X				
Origine des échecs	X						X
Les secteur d'activités dans une ville donnée	X		X	X			
Taux de sélectivité par type d'offre.	X						
Les indicateurs relatifs aux offres publiées							
Liste des concurrents							

ayant déposés un pli	X						X
Liste des concurrents évincés à l'issue de l'examen	X						X
Montant des actes d'engagement des concurrents	X						X
Justification du choix de l'attributaire	X						X

### 3. Spécifications techniques

Après une analyse détaillée des besoins fonctionnels, La spécification technique de l'architecture du projet est mise en évidence dans cette partie.

Le choix des outils est basé sur un ensemble de critères listés après une étude benchmarking pour les différents solutions Big data :

#### *La suite ELK*



**Figure 11-** la suite ELK

ELK est un acronyme qui fait référence à la suite de logiciels open source Elasticsearch, Logstash et Kibana. Chaque composant de cette suite remplit un rôle spécifique dans le traitement et la visualisation des données.

**Tableau 6- Les critères des outils technologiques**

Outil technologique	critères
Elasticsearch	<p>Est un moteur de recherche et d'analyse distribué, conçu pour stocker et interroger de grandes volumes de données en temps réel. Les critères importants incluent :</p> <ul style="list-style-type: none"> <li>• Capacité d'indexation rapide et efficace de données.</li> <li>• Recherche avancée, y compris la recherche en texte intégral.</li> <li>• Possibilité de requêtes complexes et de filtres pour extraire des informations spécifiques.</li> </ul>
Logstash	<p>Est un outil de collecte, de transformation et de traitement des données, les critères importants sont :</p> <ul style="list-style-type: none"> <li>• Capacité à collecter des données de différents format et sources, y compris les fichiers journaux, les BDD, les flux réseau, etc.</li> <li>• Transformation des données en les enrichissant, les filtrant et en appliquant des opérations spécifiques.</li> </ul>
Kibana	<p>Est une interface de visualisation des données qui permet d'explorer, de rechercher et de visualiser les données stockées dans Elasticsearch. Les critères importants sont :</p> <ul style="list-style-type: none"> <li>• Tableaux de bord interactifs et personnalisables pour la visualisation des données.</li> <li>• Visualisation graphiques telles que les graphiques, des cartes, etc.</li> <li>• Outils de recherche et de filtrage pour explorer les données en temps réel.</li> </ul>

### ***Jupyter notebook***



***Figure 12- jupyter notebook logo***

La plateforme choisit pour nettoyer et prétraiter les données est jupyter notebook, il se considère comme un environnement de développement interactif et flexible pour effectuer le nettoyage des données. Il permet d'explorer les données, de prétraiter et de visualiser les données.

### ***Microsoft SQL Server***



***Figure 13- Microsoft SQL server***

La base de données de production repose sur l'utilisation de Microsoft SQL Server en tant que SGBDR, Il est spécifiquement conçu pour stocker, organiser et gérer de manière efficaces de vastes quantités de données. Le langage utilisé pour interagir avec cette base de données est celui des requêtes SQL.

### ***Python***



***Figure 14- logo Python***

La construction de la base de donnée NOSQL sera réalisé à l'aide du langage de programmation Python, permettant d'extraire les informations et les données requises à partir des PVs. Ces données seront ensuite présentées dans une BDD NOSQL orientée document, où elles seront organisées sous forme de paires clé-valeur.

### ***Aqua Data Studio***



***Figure 15- logo aqua data studio***

Aqua Data Studio est un environnement de développement intégré multiplateforme pour la gestion et l'analyse de bases de données. Il fournit une interface conviviale pour les développeurs et les administrateurs de bases de données, leur permettant d'interagir avec différentes bases de données à l'aide du langage SQL.



#### **4. Conclusion**

Dans ce chapitre j'ai présenté les besoins des utilisateurs finaux en les détaillant, ensuite j'ai dégagé les différentes axes d'analyse et j'ai présenté les outils technologiques utilisés.

## **CHAPITRE 4**

---

### **MODÉLISATION ET ETUDE TECHNIQUE**

Ce chapitre consiste à exposer d'une partie importante du volet décisionnel du projet. Il s'agit de la modélisation des axes d'analyse pour l'activité des marchés, le nettoyage des données, la collecte, le traitement, et la visualisation. Cette partie est d'une telle importance car elle sert principalement à examiner la chaîne BI en cours de construction et obtenir, par la suite, grâce aux techniques de reporting des rapports d'activités qui vont servir d'entrée dans une partie ultérieure.

## 1. Modélisation

Grâce à l'analyse des besoins du client effectuée dans le chapitre précédent, des informations cruciales ont été recueillies pour orienter la conception de la base de données. Dans ce but, la modélisation des volets de la base de la base de données a été utilisée en raison de ses nombreux avantages.

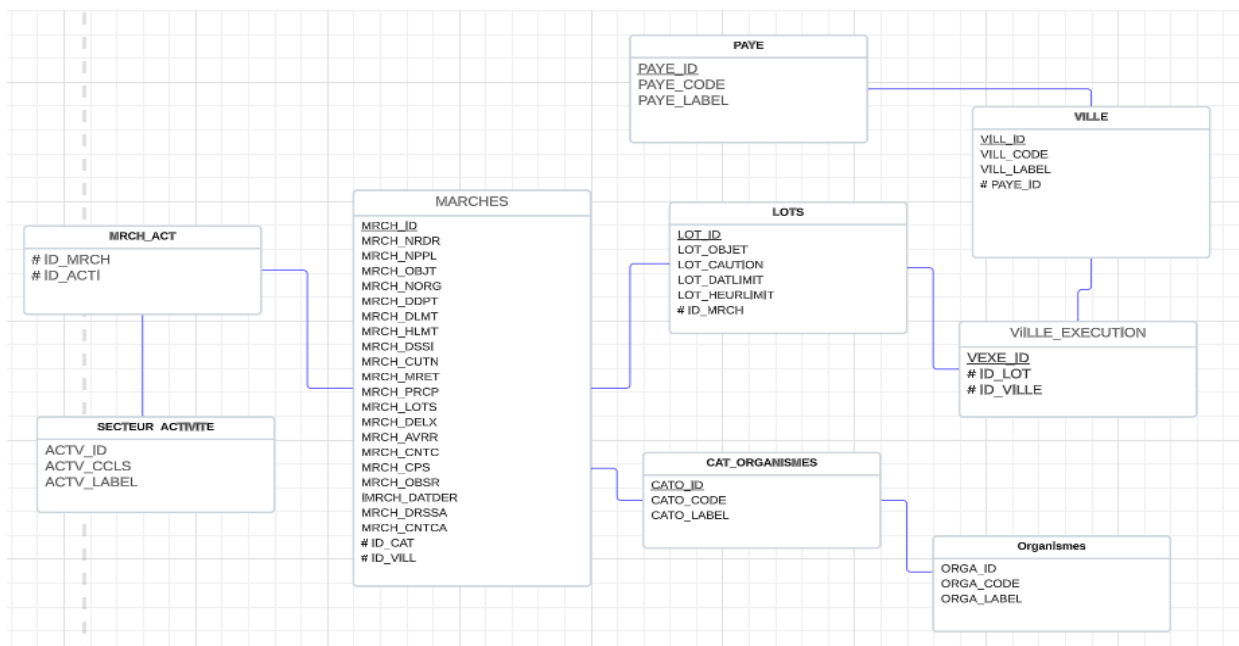
En utilisant une sélection d'indicateurs pertinents, il a été possible de filtrer la base de données de production et de choisir le schéma contenant les informations nécessaires. De plus, ces indicateurs ont permis d'exploiter les données importantes extraites de PVs.

### *1.1 Le modèle logique des données*

Le diagramme représentant le modèle logique de l'activité des marchés s'articule autour des tables suivantes.

- MARCHES
- LOTS
- VILLE
- VILLE\_EXECUTION
- ACTIVITE
- MRCH\_ACT
- CAT\_ORGANISMES
- ORGANISMES

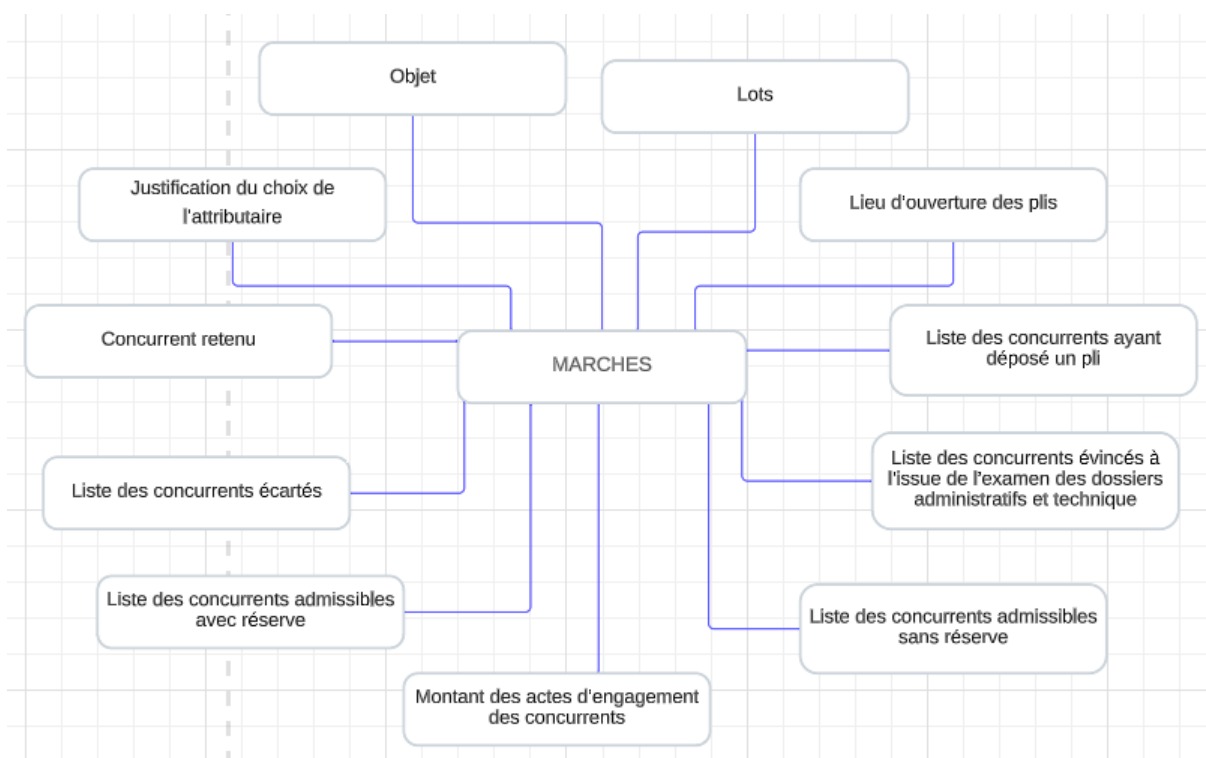
La figure suivante illustre le modèle logique de l'activité des marchés.



**Figure 16-** modèle logique de l'activité des marchés

## 1.2 Le schéma des documents

Le schéma ci-dessous représente le modèle des documents structurés sous la forme des paires clé-valeur.



**Figure 17-**le schéma des documents

## 2. Dictionnaire de données

Après avoir conclu la phase de la modélisation des données, la conception physique de la base de données sera entamée, dans cette étape, le contenu de chaque table des schémas de modélisation sera détaillé, en dressant un dictionnaire de données complet qui définira de manière précise la signification de chaque champ.

### 2.1 Les tables de l'activité des marchés

Dans ce qui suit, le contenu des tables sera présenté.

#### *La table MARCHES*

**Tableau 7- Description de la table MARCHES**

Champ	Type	Description
MRCH_ID	Nombre entier (int)	L'identifiant d'un marché
MRCH_NRDR	Nombre entier (int)	Le numéro d'ordre d'un marché
MRCH_NPPL	Chaine de caractère (Varchar(200))	Le référence d'un marché
MRCH_OBJET	Chaine de caractère (Varchar(2500))	La description et la nature de l'achat de ce marché
MRCH_CUTN	Nombre décimal	La caution et la garantie financière déposée par l'entrepreneur.
MRCH_MRET	Nombre décimal	Le montant de retrait, il s'agit de la valeur monétaire qui peut être retirée.
MRCH_CNTCA	Nombre décimal	L'estimation financière
MRCH_LOTS	Chaine de caractère (Varchar(1))	Indique que la marché contient Lot unique U ou plusieurs P.
MRCH_DDPT	Datetime	Date de début
MRCH_DLMT	Datetime	Date limite
MRCH_DELX	Datetime	Date d'exécution
ID_CAT	Nombre entier (int)	

ID_VILL	Nombre entier (int)	Ces clés sont des clés étrangères issues de la table Ville et Cat_organismes.
---------	---------------------	---

### ***La table LOTS***

***Tableau 8- Description de la table LOTS***

Champ	Type	Description
LOT_ID	Nombre entier (int)	L'identifiant d'un Lot
LOT_OBJET	Chaine de caractère (Varchar(2500))	La description d'un Lot
LOT_CAUTION	Nombre décimal	la garantie financière déposée pour un Lot
LOT_DATLIMIT	Datetime	Date limite
ID_MRCH	Nombre entier (int)	clé étrangère issue de la table MARCHES

### ***La table VILLE***

***Tableau 9- Description de la table VILLE***

Champ	Type	Description
VILL_ID	Nombre entier (int)	L'identifiant d'une ville
VILL_CODE	Chaine de caractère (Varchar(5))	Le code postale
VILL_LABEL	Chaine de caractère (Varchar(40))	La ville
PAYE_ID	Nombre entier (int)	clé étrangère issue de la table PAYE

### ***La table PAYE***

***Tableau 10- Description de la table PAYE***

Champ	Type	Description
PAYE_ID	Nombre entier (int)	L'identifiant d'un pays
PAYE_CODE	Chaine de caractère (Varchar(5))	Le code unique à un pays
PAYE_LABEL	Chaine de caractère (Varchar(40))	Le pays

### ***La table CAT\_ORGANISMES***

***Tableau 11- Description de la table Cat\_organismes***

Champ	Type	Description
CATO_ID	Nombre entier (int)	L'identifiant d'une organisme
CATO_CODE	Chaine de caractère (Varchar(5))	Le code de l'organisme
CATO_LABEL	Chaine de caractère (Varchar(40))	Le nom de l'organisme

### ***La table ORGANISMES***

***Tableau 12- Description de la table organismes***

Champ	Type	Description
ORGA_ID	Nombre entier (int)	L'identifiant du sous-organisme
ORGA_CODE	Chaine de caractère (Varchar(5))	Le code du sous-organisme
ORGA_LABEL	Chaine de caractère (Varchar(40))	Le nom du sous-organismes

### ***La table SECTEUR\_ACTIVITE***

***Tableau 13- Description de la table secteur\_activite***

Champ	Type	Description
ACTV_ID	Nombre entier (int)	L'identifiant de secteur d'activité

ACTV_CCLS	Chaîne de caractère (Varchar(5))	Classification standardisée des secteurs d'activité
ACTV_LABEL	Chaîne de caractère (Varchar(40))	Le secteur d'activité

### 3. Nettoyage des données

La définition de dictionnaire de données marquera le début de la phase de nettoyage des données. Cette étape représente une tâche courante dans le processus de préparation des données.

Jupyter est un environnement de développement interactif largement utilisé pour l'analyse et le nettoyage des données en utilisant le langage Python.

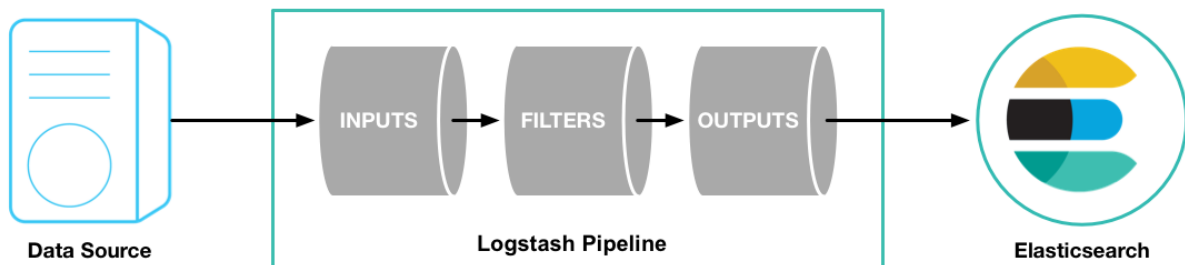
Les étapes pour effectuer le prétraitement des données sont représentées comme suit.

- Importation des bibliothèques : pour faire la connexion à la base de données Microsoft SQL Server, nous utilisons la bibliothèque Pymssql dans le but d'effectuer les opérations de manipulation et d'interrogation des données.
- Chargement des données : Pour importer les données à partir d'une base de données, utilisation de la bibliothèque Pandas en complément de Pymssql est recommandée, Pandas est une bibliothèque puissante en Python, qui offre des fonctionnalités avancées pour le traitement et l'analyse des données.
- Exploration des données : en effectuant une analyse des informations sur les colonnes, nous pouvons vérifier la présence de valeurs manquantes dans les données.
- Nettoyage des valeurs aberrantes : identifier les valeurs aberrantes ou incohérentes et choisir la meilleure approche pour les traiter.
- Normalisation et transformation des données : standardiser la forme des valeurs pour unifier l'affichage en effectuant des opérations de normalisation et de transformation sur les données
- Exportation des données : une fois le nettoyage de données terminé, nous pouvons exporter les données pour une utilisation ultérieure.



#### 4. La collecte et le traitement des données

La collecte des données avec la pile ELK est effectuée en utilisant l'approche des logstash pipelines, dans le but de définir les flux de traitement personnalisés pour collecter, transformer et acheminer les données vers Elasticsearch



*Figure 18-logstash pipeline*

Le processus de fonctionnement de pipelines logstash se définit comme suit.

- Configuration des pipelines : la première étape consiste à configurer les pipelines en définissant des fichiers de configuration spécifiques pour chaque flux de données à traiter, nous spécifions les sources de données, les filtres de transformation et les destinations de données.
- Collecte des données : une fois les pipelines configurés, logstash se charge de collecter les données à partir des sources spécifiques. Logstash récupère les données entrantes et les achemine vers les étapes suivantes du pipeline.
- Traitement des données : les pipelines permettent de définir des filtres pour transformer et traiter les données en fonction des besoins.

```
filter {
  aggregate {
    task_id => "%{MRCH_ID}"
    code => "
      map['MRCH_ID'] ||= event.get('MRCH_ID')
      map['lots'] ||= []
      map['lots'] << {

        'lot_id' => event.get('LOT_ID'),
        'lot_objet' => event.get('LOT_OBJET'),
        'lot_caution' => event.get('LOT_CAUTION')
      }

      "
    push_previous_map_as_event => true
    timeout => 3
  }
}
```

*Figure 19- pipeline filter*

Les besoins spécifiques sont obtenus en utilisant des opérations et des requêtes SQL.

```
input { jdbc {
  jdbc_connection_string => "jdbc:sqlserver://192.168.14.253;databaseName=MARCHE_TEST;integratedSecurity=false;"
  jdbc_user => "user_mssql"
  jdbc_password => "123456789"
  jdbc_driver_class => "com.microsoft.sqlserver.jdbc.SQLServerDriver"
  statement => "SELECT
M.MRCH_ID,
L.ID_MRCH,
M.MRCH_NRDR,
M.MRCH_NPPL,
M.MRCH_OBT,
M.MRCH_NORG,
M.MRCH_DDPT,
M.MRCH_DLMT,
M.MRCH_HLMT,
M.MRCH_DSSI,
M.MRCH_CUTN,
M.MRCH_MRET,
M.MRCH_PRCP,
M.MRCH_LOTS,
M.MRCH_DELX,
M.MRCH_AVRR,
M.MRCH_CNTC,
M.MRCH_CPS,
M.MRCH_OBSR,
```

*Figure 20- pipeline output*

- Sortie des données : Acheminement des données vers Elasticsearch. Une fois les données collectées et transformées, Logstash les envoie vers les destinations spécifiées dans la configuration du pipeline.

```
output {
  elasticsearch {
    hosts => ["http://localhost:9200"]
    index => "offresenligne"
  }
}
```

*Figure 21- pipeline output*

## 5. La visualisation des données

Une fois les calculs effectués pour obtenir les indicateurs élaboré, les informations nécessaires peuvent être visualisées sur Kibana.

Nous pouvons explorer et afficher les données sur Kibana en utilisant ses fonctionnalités de visualisation avancées pour représenter graphiquement les résultats obtenus.

Pour accéder aux données,

- On crée tout d'abord une vue de données (data view) dans Kibana.

La création d'une vue de données permet de spécifier les sources de données, les filtres et les agrégations à appliquer pour obtenir une vue personnalisée dans Kibana.

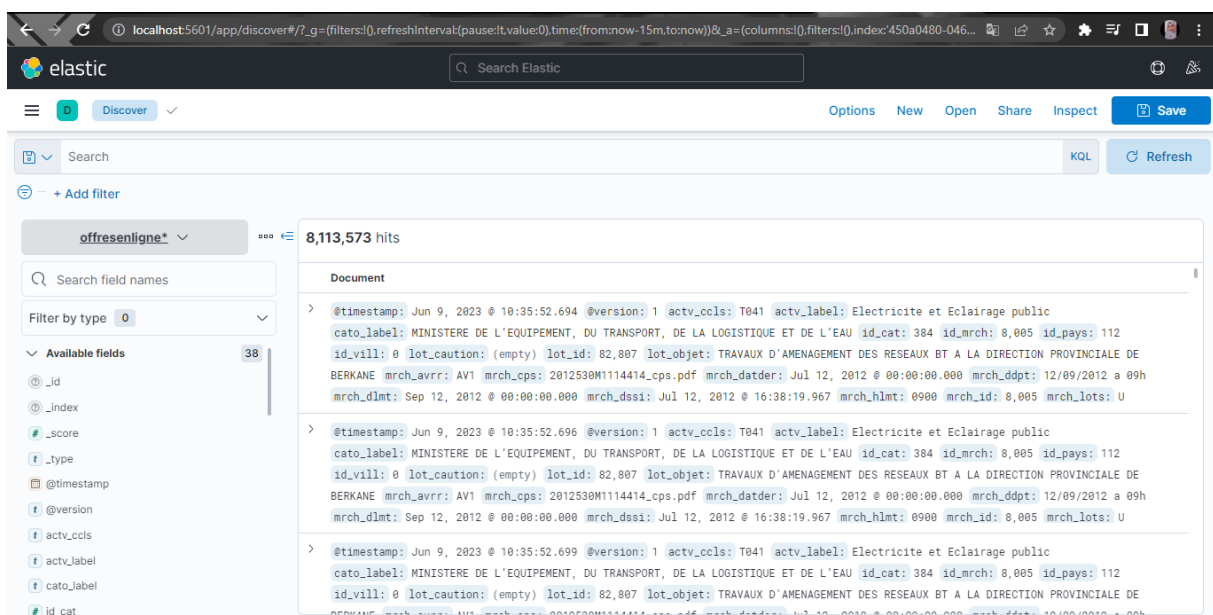
- Après avoir créé la vue de données, nous accédons à l'onglet « Discover » dans Kibana.

L'onglet Discover permet d'explorer les données en affichant une vue tabulaire des documents indexés dans Elasticsearch.

- Ensuite, nous sélectionnons le « index pattern » dans Kibana.

L'index pattern est utilisé pour définir la structure des données indexées dans Elasticsearch et permet à Kibana de comprendre comment interagir avec ces données.

Les données s'afficheront dans Kibana conformément à la configuration de l'index pattern sélectionné.



*Figure 22- les documents sur kibana*

## 6. Conclusion

Ce chapitre a couvert les différentes étapes de la conception, du nettoyage, de la collecte, du traitement et de la visualisation des données, en mettant l'accent sur l'importance de chaque étape pour obtenir des résultats précis et exploitables.

## **CHAPITRE 5**

---

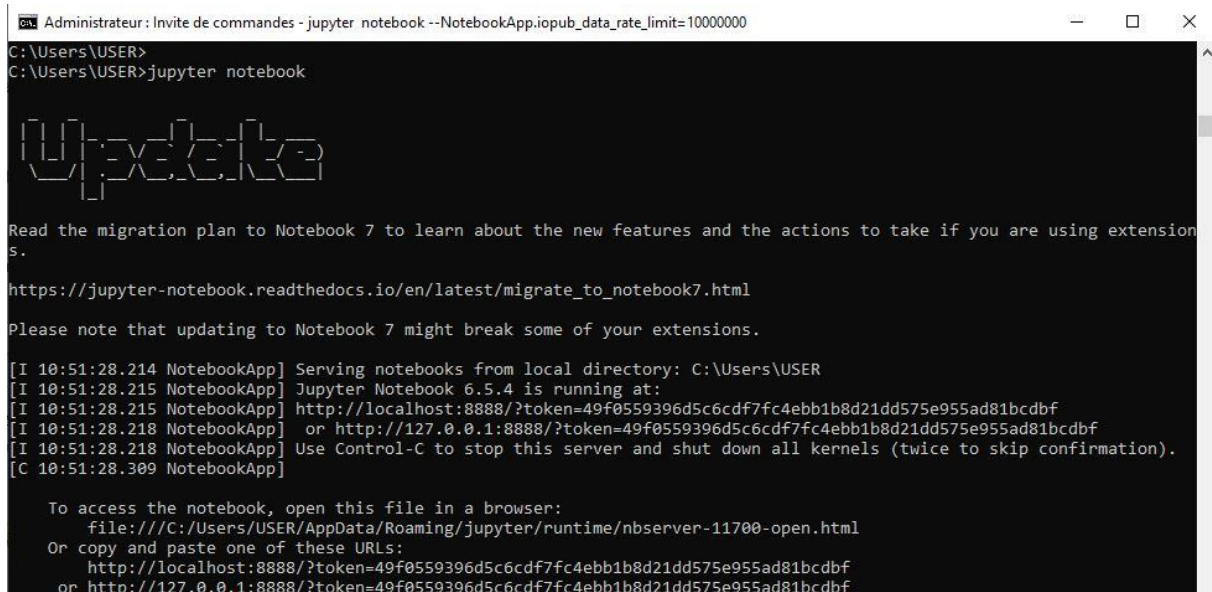
### **MISE EN ŒUVRE DE LA SOLUTION**

Après avoir défini l'architecture globale de la solution dans le chapitre précédent, ce chapitre se concentrera sur la description détaillée de chaque étape de la solution, accompagnée d'aperçus des interfaces de développement correspondantes. Cela permettra d'illustrer le travail effectué et de donner un aperçu concret de chaque partie de la solution.

## 1. Nettoyage de données

Pour obtenir une base de données nettoyée et prête à être utilisée à l'aide de Jupyter, les étapes suivantes sont généralement appliquées

- Lancement de logiciel Jupyter notebook avec la commande : **jupyter notebook**



```
Administrateur : Invite de commandes - jupyter notebook --NotebookApp.iopub_data_rate_limit=10000000
C:\Users\USER>jupyter notebook

Jupyter Notebook

Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions.
https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html

Please note that updating to Notebook 7 might break some of your extensions.

[I 10:51:28.214 NotebookApp] Serving notebooks from local directory: C:\Users\USER
[I 10:51:28.215 NotebookApp] Jupyter Notebook 6.5.4 is running at:
[I 10:51:28.215 NotebookApp] http://localhost:8888/?token=49f0559396d5c6cdf7fc4ebb1b8d21dd575e955ad81bcdbf
[I 10:51:28.218 NotebookApp] or http://127.0.0.1:8888/?token=49f0559396d5c6cdf7fc4ebb1b8d21dd575e955ad81bcdbf
[I 10:51:28.218 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 10:51:28.309 NotebookApp]

To access the notebook, open this file in a browser:
file:///C:/Users/USER/AppData/Roaming/jupyter/runtime/nbserver-11700-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=49f0559396d5c6cdf7fc4ebb1b8d21dd575e955ad81bcdbf
or http://127.0.0.1:8888/?token=49f0559396d5c6cdf7fc4ebb1b8d21dd575e955ad81bcdbf
```

*Figure 23- jupyter notebook*

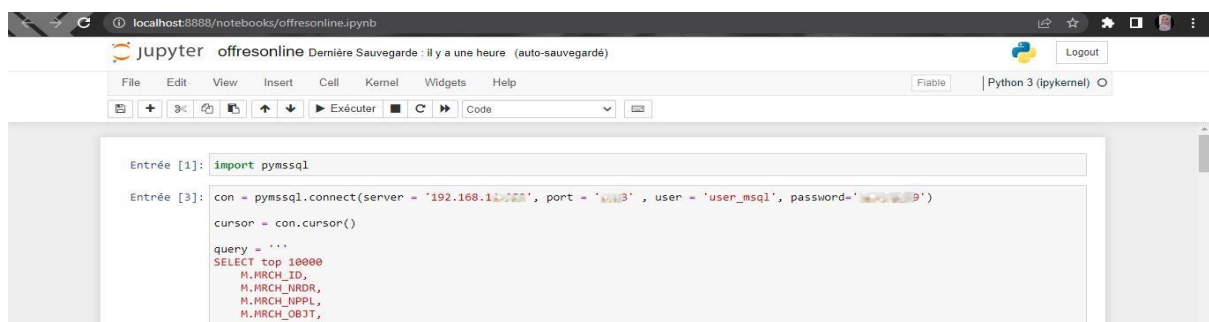
- Importation des bibliothèques Pymssql, et Pandas

Pymssql est une bibliothèque pour effectuer les opérations de manipulation et d'interrogation des données.

Pandas est une bibliothèque Python, qui offre des fonctionnalités avancées pour le traitement et l'analyse des données

- Connexion avec Microsoft SQL Server

Les informations de connexion appropriées, telles que l'hôte, le nom de la base de données, le nom d'utilisateur et le mot de passe sont nécessaires pour établir une connexion avec la base de données SQL Server



```
Entrée [1]: import pymssql

Entrée [3]: con = pymssql.connect(server = '192.168.1.100', port = '1433', user = 'user_mssql', password='password')
cursor = con.cursor()

query = '''
SELECT top 10000
M_HRCH_ID,
M_HRCH_NDR,
M_HRCH_NPPL,
M_HRCH_OBT,
M_HRCH_NORG.
```

*Figure 24- connexion avec Microsoft SQL Server*

- Chargement des données

```

cursor.execute(query)
data = cursor.fetchall()

for item in data:
    print(item)

con.close()

```

(3025, 506, '05/2012', 'TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE RESEAU ROUTIER DE LADITE DIRECTION', 'dp de zagora', None, datetime.datetime(2012, 3, 21, 0, 0), '', datetime.datetime(2012, 2, 20, 12, 36, 23, 577000), '7000', None, 'N', 'U', None, '', None, '', None, 384, 0, 112, datetime.datetime(2012, 2, 20, 0, 0), None, '0101', 'RABAT', 'MINISTERE DE L'EQUIPEMENT, DU TRANSPORT, DE LA LOGISTIQUE ET DE L'EAU', 'M1110001', 'DIRECTION DES ROUTES', 'T031', 'TRAVAUX Routiers')

(3025, 506, '05/2012', 'TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE RESEAU ROUTIER DE LADITE DIRECTION', 'dp de zagora', None, datetime.datetime(2012, 3, 21, 0, 0), '', datetime.datetime(2012, 2, 20, 12, 36, 23, 577000), '7000', None, 'N', 'U', None, '', None, '', None, 384, 0, 112, datetime.datetime(2012, 2, 20, 0, 0), None, '0101', 'RABAT', 'MINISTERE DE L'EQUIPEMENT, DU TRANSPORT, DE LA LOGISTIQUE ET DE L'EAU', 'M1110001', 'DIRECTION DES ROUTES', 'T031', 'TRAVAUX Routiers')

**Figure 25-** chargement des données

Une fois la connexion établie, nous pouvons exécuter des requêtes SQL en utilisant la méthode **cursor()** pour créer un curseur et la méthode **execute()** pour exécuter des requêtes.

- Représentation de données sous la structure d'un dataframe

La représentation des données sous la structure d'un dataframe est une façon courante et efficace de manipuler et d'analyser des données en utilisant la bibliothèque Pandas en Python.

```

Entrée [4]: import pandas as pd

Entrée [5]: df = pd.DataFrame(data)
            print(df.head())

```

	0	1	2	3
0	3025	506	05/2012	TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE ...
1	3026	507	05/2012	CONTRÔLE DES TRAVAUX DE RECHARGEMENT DES ACCOT...
2	3027	508	05/2012	TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE ...
3	3028	509	05/2012	ACHAT DE PENSEMENTS ET ACCESSOIRES DE MEDECINE...
4	3029	510	05/2012	REALISATION DES ACTIONS DE FORMATION AU PROFIT...

**Figure 26-** les données en dataframe

- Gestion des doublons

Nous utilisons la méthode **drop\_duplicates()** pour supprimer les lignes dupliquées du dataframe.

```

Entrée [6]: df_cleaned = df.drop_duplicates()

Entrée [7]: print(df_cleaned)

```

	0	1	2	3
0	3025	506	05/2012	TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE ...
2	3026	507	05/2012	CONTRÔLE DES TRAVAUX DE RECHARGEMENT DES ACCOT...
3	3027	508	05/2012	TRAVAUX D'ENTRETIEN DES OUVRAGES D'ART SUR LE ...
4	3028	509	05/2012	ACHAT DE PENSEMENTS ET ACCESSOIRES DE MEDECINE...
6	3029	510	05/2012	REALISATION DES ACTIONS DE FORMATION AU PROFIT...

**Figure 27-** gérer les doublons

- Gestion des valeurs aberrants ou incorrectes

Nous créons une fonction lambda qui définit les règles de validation ou de correction des valeurs aberrantes ou incorrectes. Cette fonction prendra une valeur en entrée et retournera la valeur corrigée.

```

Entrée [23]: df[0] = df[0].apply(lambda x: x if x > 0 else 0)

Entrée [24]: print(df[0])

```

	0
0	3025
1	3026
2	3027
3	3028
4	3029
...	...
9995	406553
9996	406554
9997	406555
9998	406556
9999	406557

Name: 0, Length: 10000, dtype: int64

**Figure 28-** gérer les valeurs aberrantes

- Suppression des lignes avec les valeurs manquantes

La suppression des lignes contenant des valeurs manquantes dans un dataframe est effectuée en utilisant la méthode **dropna()** de Pandas

```

Entrée [21]: df_cleaned = df.dropna(subset=[0])

Entrée [22]: print(df_cleaned)

```

	0	1	2	3
0	3025	506	05/2012	travaux d'entretien des ouvrages d'art sur le ...
1	3026	507	05/2012	contrôle des travaux de rechargement des accot...
2	3027	508	05/2012	travaux d'entretien des ouvrages d'art sur le ...
3	3028	509	05/2012	achat de pensements et accessoires de medecine...
4	3029	510	05/2012	realisation des actions de formation au profit...

**Figure 29-** gérer les valeurs manquantes

- Standardiser la forme des valeurs

Pour standardiser la casse des valeurs dans une colonne, nous utilisons les méthodes **str.upper()** ou **str.lower()**.

Pour supprimer les espaces en début et fin de chaîne dans une colonne, nous utilisons la méthode **str.strip()**

Nous pouvons remplacer certaines valeurs par d'autres pour standardiser la forme des données. En utilisant la méthode **replace()**

```
Entrée [20]: df[24] = df[24].str.upper()
df[24] = df[24].str.strip()
print(df[24])

0      RABAT
1      RABAT
2  DRILOUCH
3  MOHAWMEDIA
4      OUJDA
...
9995    RABAT
9996  NOUACER
9997  CASABLANCA
9998    AGADIR
9999  CASABLANCA
Name: 24, Length: 10000, dtype: object
```

**Figure 30-** gérer la forme des valeurs

## 2. Restitution des données

La phase de restitution de données permet de donner un aperçu global sur les rapports générés.

Après avoir généré des rapports sous l'outil de visualisation et de reporting Kibana maintenant il est temps de les présenter tout en spécifiant leurs graphes et leur utilité.

L'intérêt de ces rapports est de permettre aux contrôleurs de gestion d'accéder de manière interactive aux informations.

### 2.1 Présentation des indicateurs générales relatifs aux entreprises

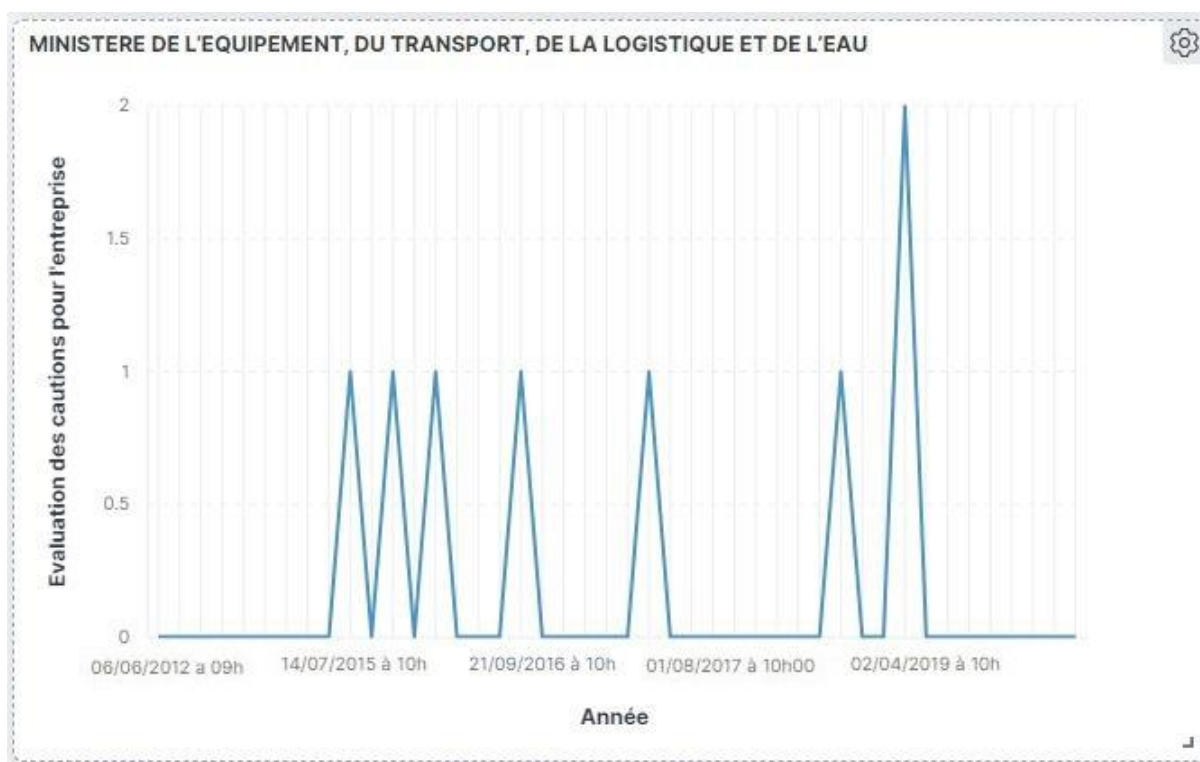
- Estimation financière au cours des années pour les entreprises afin de prévoir le cout de la réalisation



**Figure 31-** Estimation financière par année pour une entreprise

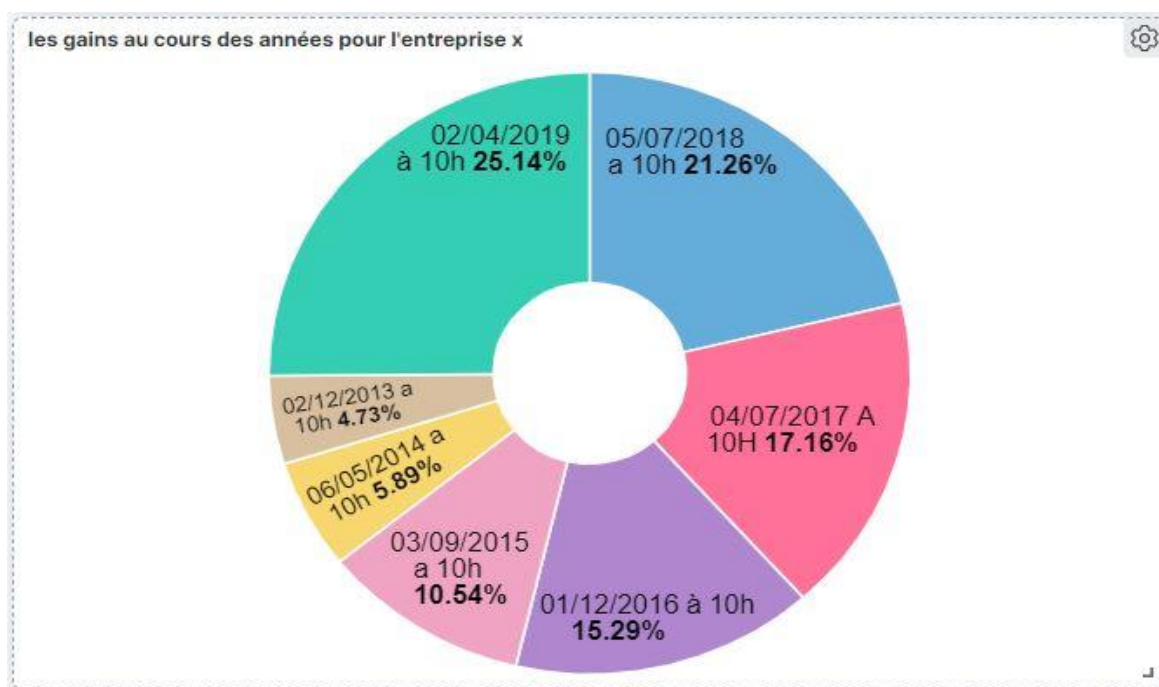


- Evolution des cautions au cours du temps pour les entreprises dans le but d'observer la garantie financière déposée pour les marchés



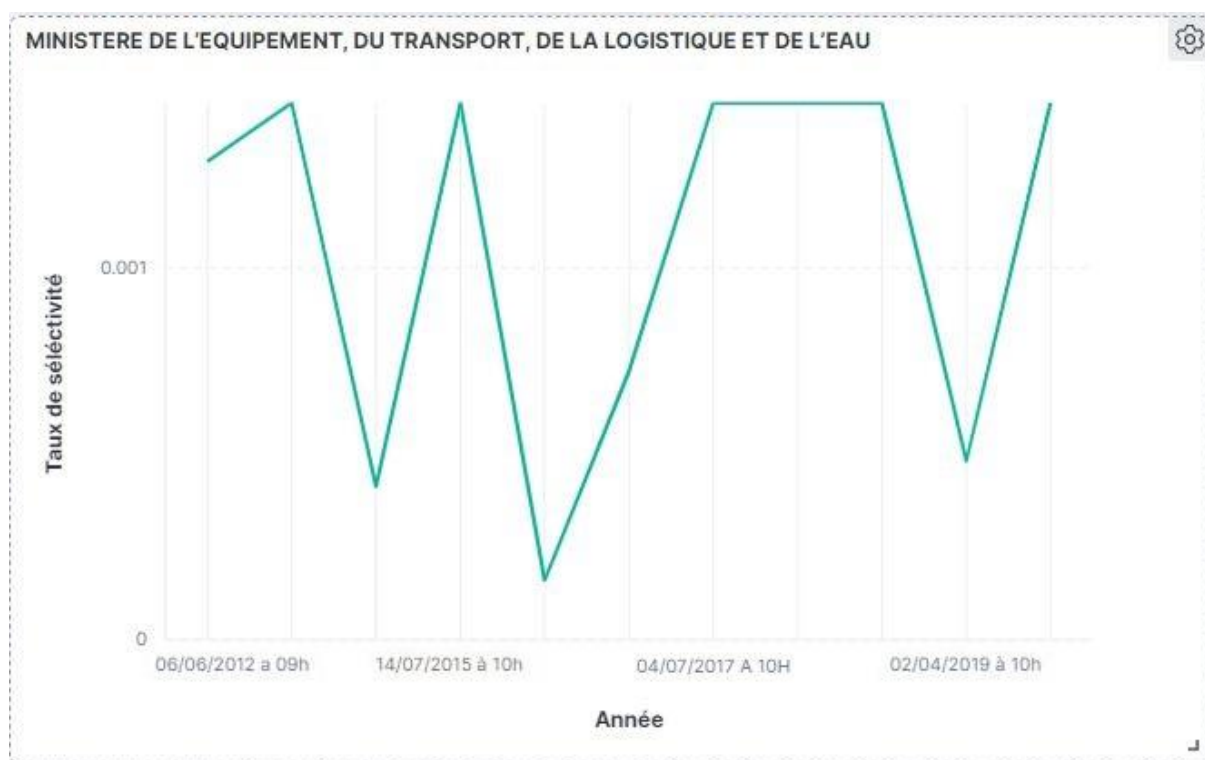
*Figure 32- Evolution des cautions par année pour une entreprise*

- Les gains au cours des années pour les entreprises



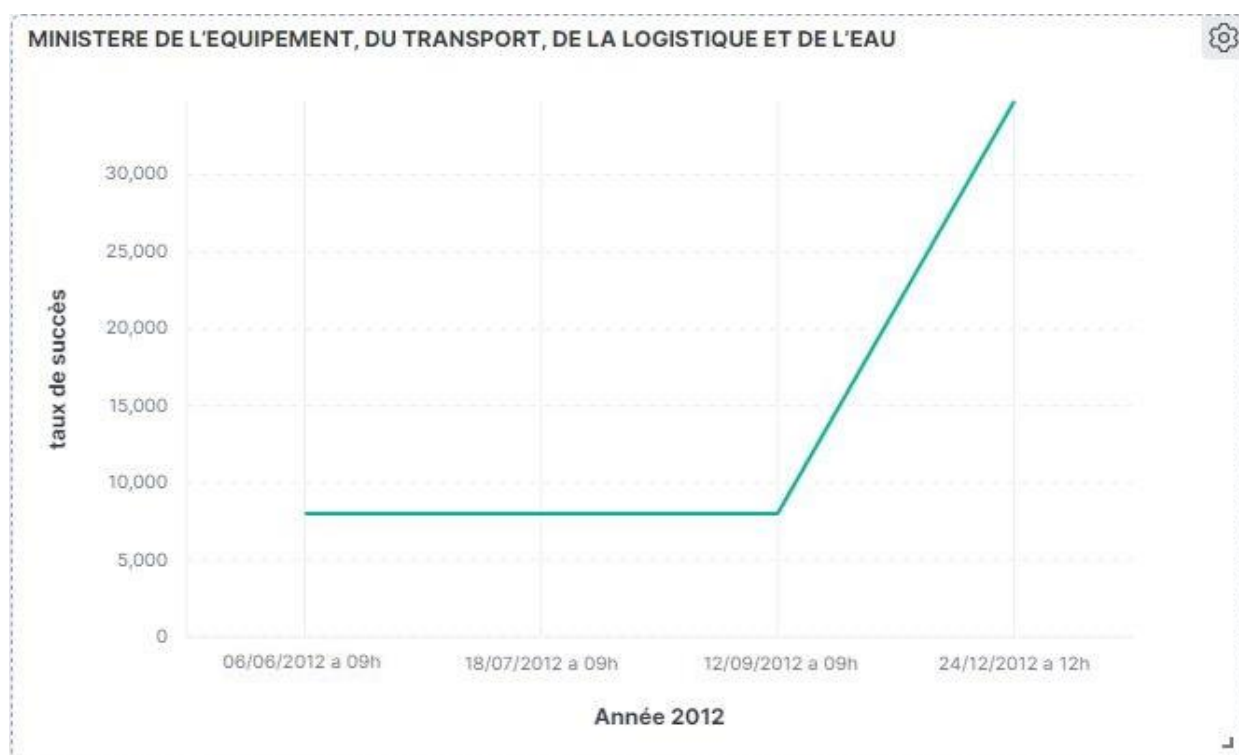
*Figure 5.11 – le pourcentage des gains par année pour une entreprise*

- Taux de sélectivité par année pour les entreprises



**Figure 33-** le taux de sélectivité par année pour une entreprise

- Taux de succès pour les entreprises pendant une année



**Figure 34-** le taux de succès pour une entreprise en 2012

## 2.2 Présentation des indicateurs générales relatifs aux marchés

- Nombre des appels d'offres par secteur d'activité dans le but de comparer les différents secteurs d'activités et d'observer le secteur le plus actif.

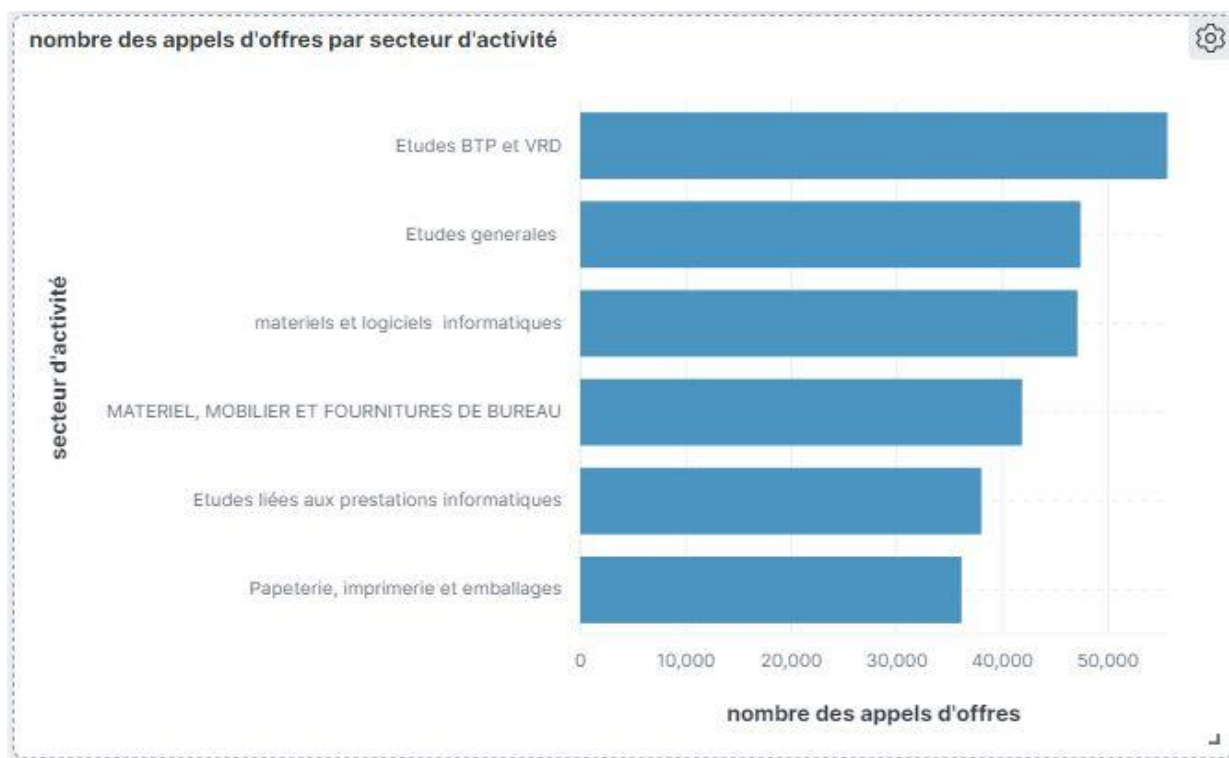


Figure 35- nombre des appels d'offres par secteur d'activité

- Estimation financière par secteur d'activité

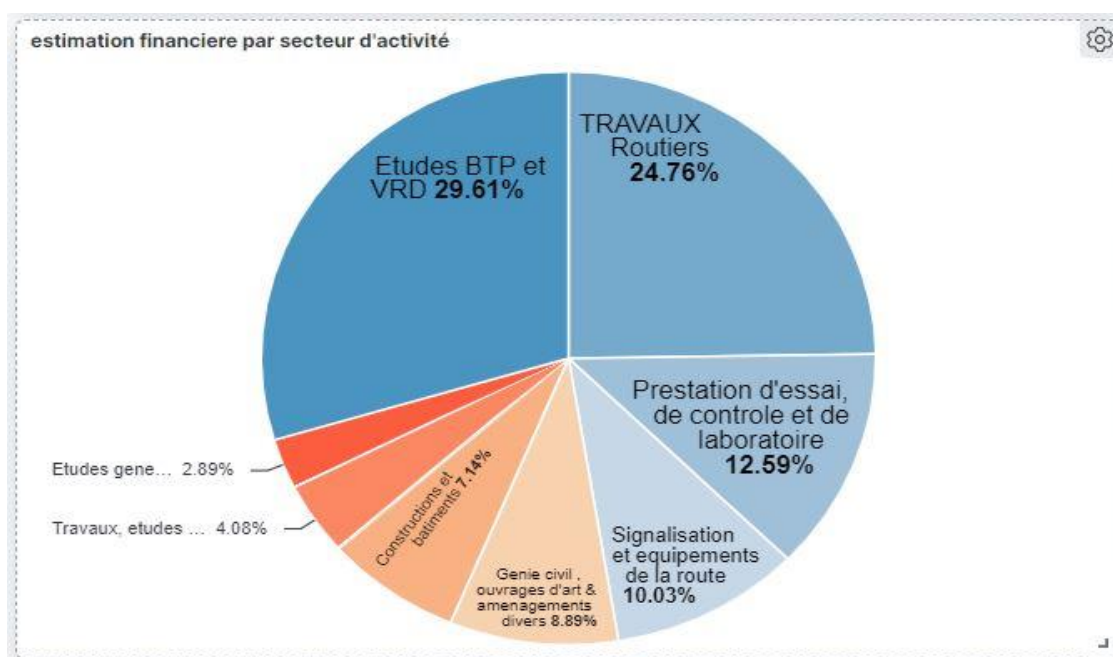
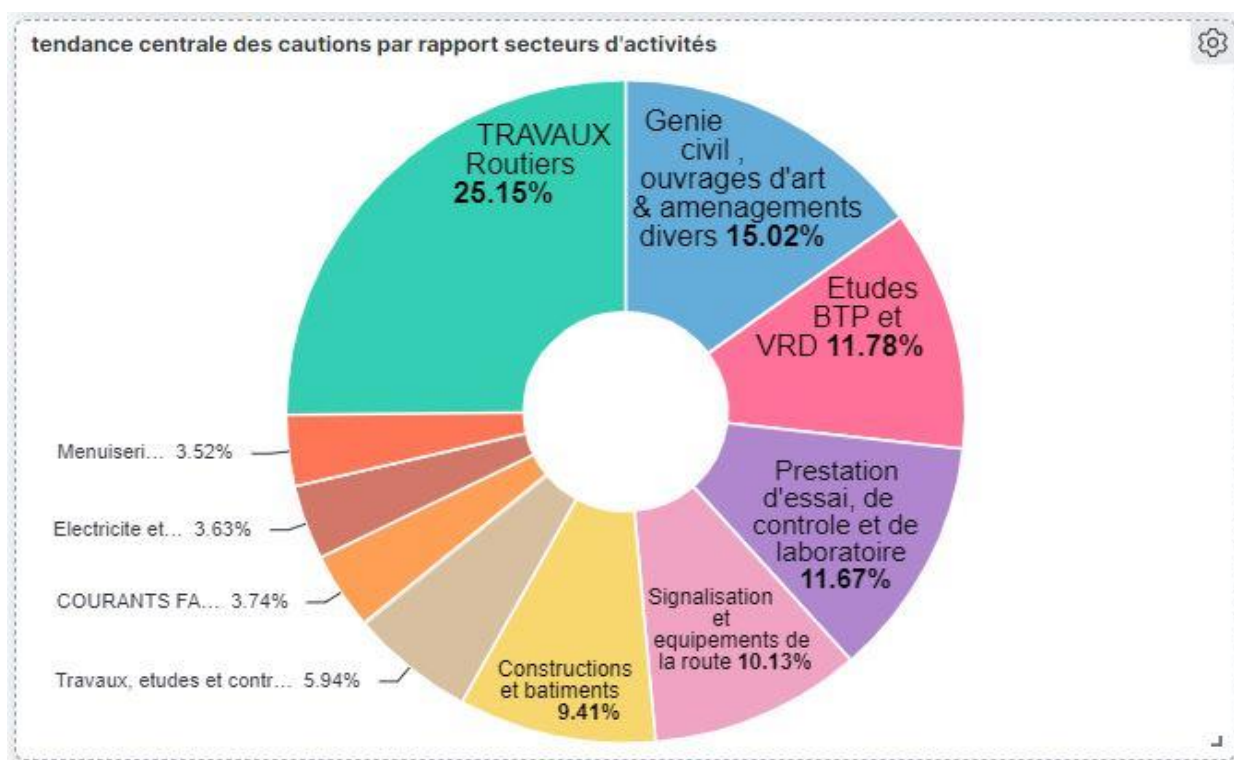


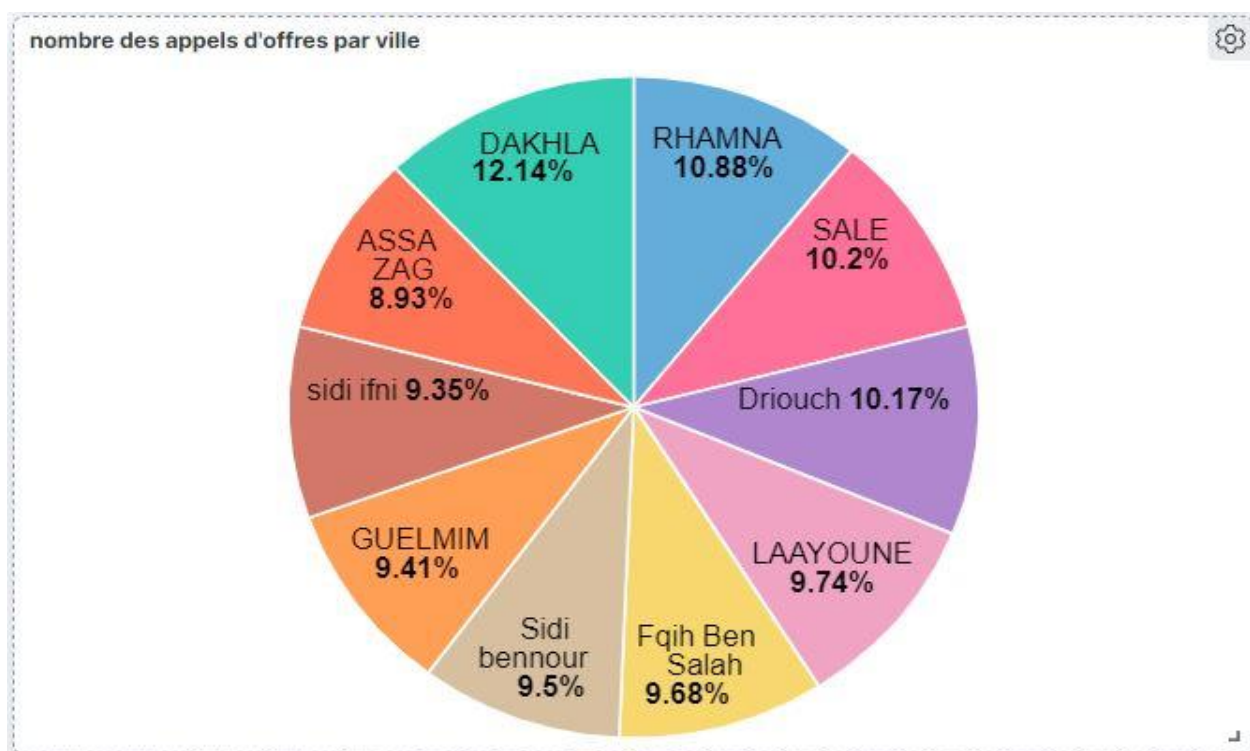
Figure 36- Estimation financière par secteur d'activité

- Tendance centrale des cautions par rapport les secteurs d'activités



**Figure 37-** la moyenne des cautions par secteur d'activité

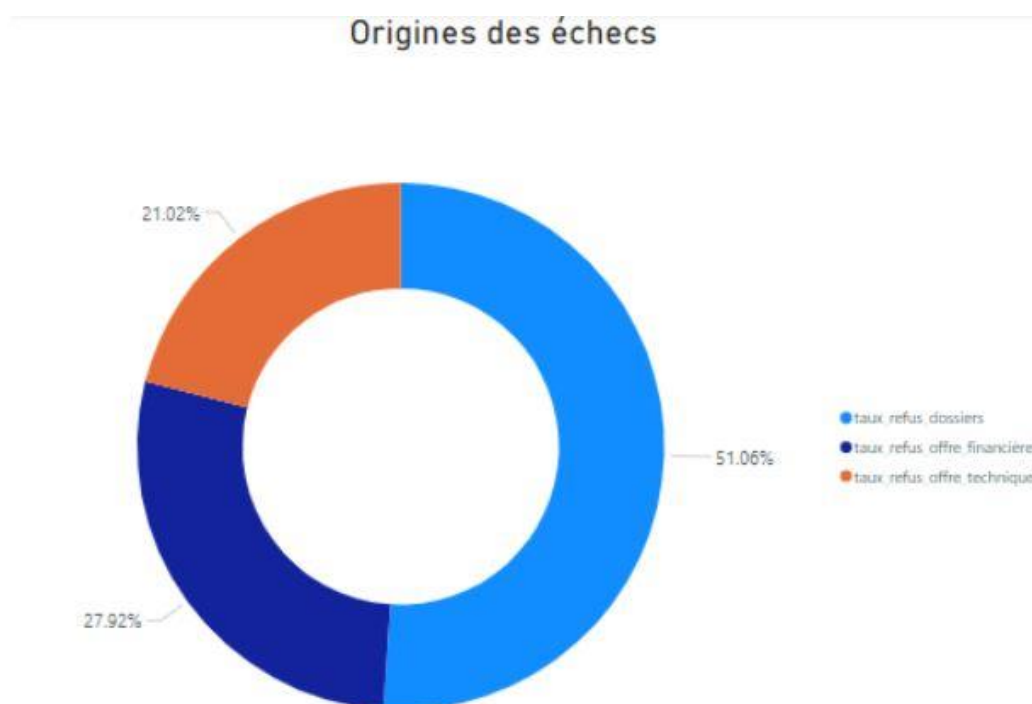
- Nombre des appels d'offres par ville afin d'observer la ville la plus active.



**Figure 38-** Nombre des appels d'offres par ville

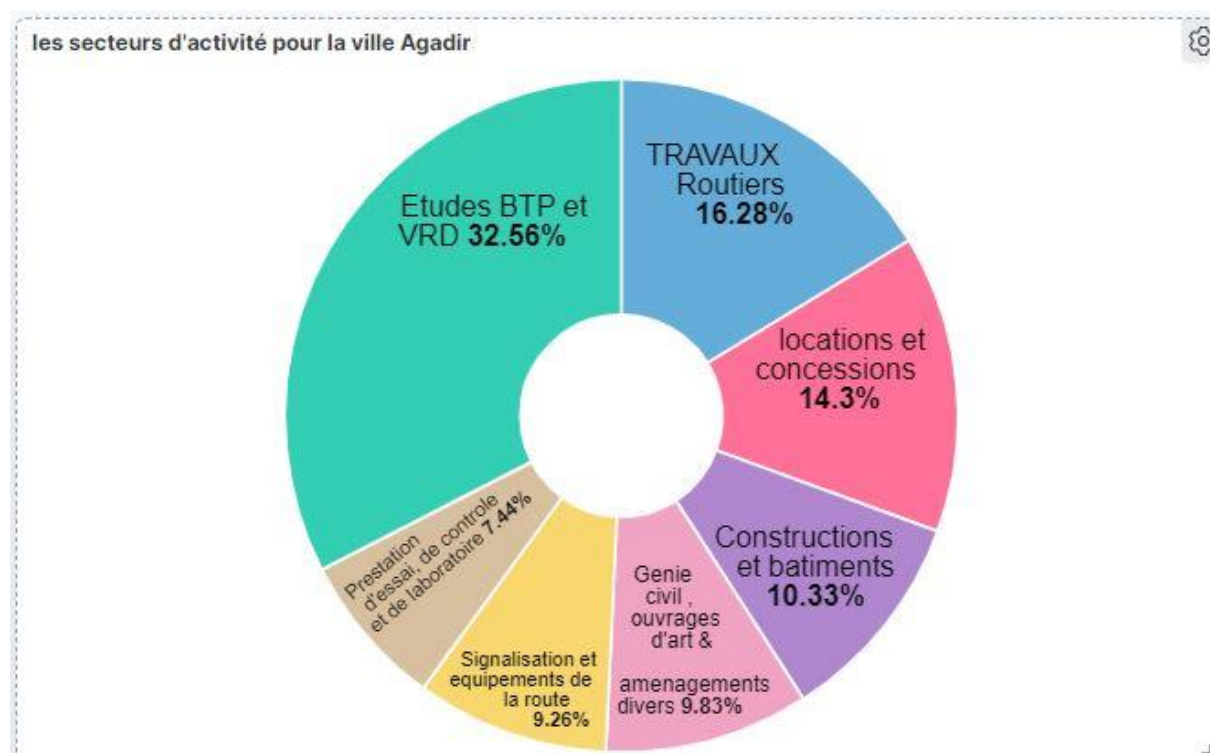


- L'origine des échecs après l'examen des dossiers administratifs, financière, et technique. Ce rapport contient un graphique représentant la répartition des origines des échecs dans l'année 2020.



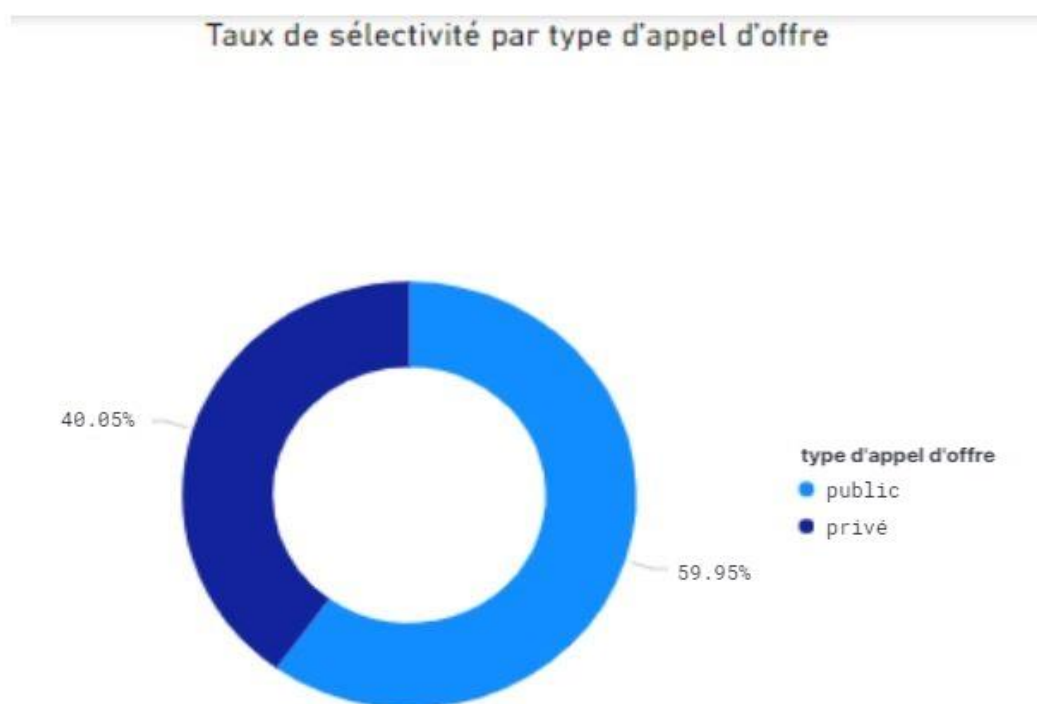
**Figure 39- Origine des échecs dans l'année 2020**

- Les secteurs d'activités dominants dans les villes



**Figure 40- les secteurs d'activités pour la ville Agadir**

- Taux de sélectivité par type appel d'offre



**Figure 41-** Taux de sélectivité par type appel d'offre

## CONCLUSION & PERSPECTIVES

Le projet consistait à mettre en œuvre un système décisionnel pour la plateforme offresonline. Il s'agit précisément de générer des tableaux de bord qui offrent une meilleure visibilité sur l'activité des marchés des adhérents et qui mettent à disposition des dirigeants l'information essentielle pour la prise de décision relativement à la soumission à un marché.

Vu le caractère décisionnel du projet, les différentes étapes du cycle de vie des données ont été implémentés, Après avoir étudié l'existant et recueilli les besoins des utilisateurs finaux, j'ai défini les indicateurs et les axes d'analyse et abouti à l'architecture technique de la solution. Grâce aux indicateurs et axes d'analyse identifiés, j'ai pu dresser une Modélisation des données, le prétraitement, la collecte, le traitement, et la présentation des tableaux de bord.

En termes de perspectives, je souhaite poursuivre le développement de la solution en mettant en place une API permettant de récupérer les indicateurs et informations nécessaires en utilisant la technologie FastAPI. Cela permettra d'offrir une interface plus dynamique et flexible pour accéder aux données.

De plus, je suis très intéressé par le domaine de l'intelligence artificielle (IA). Je souhaite explorer les possibilités d'appliquer des techniques d'IA dans le cadre du projet, notamment en développant un modèle capable de prendre des décisions et de fournir des suggestions basées sur l'analyse des données. Cela permettra d'apporter une dimension prédictive et plus avancée à la solution.

Enfin, je souhaite également me concentrer sur la présentation des résultats en développant des interfaces conviviales et intuitives à l'aide de la technologie Angular. Cela permettra aux utilisateurs finaux d'interagir facilement avec les tableaux de bord et de visualiser les informations de manière claire et compréhensible.

Par ailleurs, ce projet m'a été bénéfique aussi bien au niveau informatique qu'au niveau professionnel. Il m'a permis de raffiner mes capacités d'abstraction et de conception en modélisation décisionnelle et offert l'occasion de découvrir et manipuler plusieurs outils Open Source de la BI.

## **BIBLIOGRAPHIE/WEBOGRAPHIE**

- [1] [http ://offresonline.com/](http://offresonline.com/)
- [2] <https://www.elastic.co/fr/what-is/elk-stack>
- [3] <https://www.elastic.co/guide/index.html>
- [4] <https://docs.jupyter.org/en/latest/>
- [5] <https://docs.snowflake.com/en/user-guide/ecosystem-bi>
- [6] <http://b3d.bdpedia.fr/docstruct.html>
- [7] livre : INTRODUCTION AU COURS ELK Appartient à devopssec.fr