

Master en Sciences et Techniques (MST)

Filière : Systèmes d'information Décisionnels et Imagerie (SIDI)

Module : Système d'information d'aide à la décision

Année universitaire : 2022/2023

Projet BI sous le thème :

La consommation d'alcools chez les étudiants

Réaliser par :

SIHAME BOUTAHRI

OMAYMA SENHOU

Encadré par :

Pr. YOUSSEF FARHAOUI

Table des matières

Introduction	4
Chapitre 1 : Présentation du projet.....	5
1. Vue général sur le projet.....	5
2. Objectif de projet	5
Chapitre 2 : Théorie d'informatique décisionnelle	6
1. L'informatique décisionnelle.....	6
2. L'environnement des logiciels.....	9
Chapitre 3 : Mise en œuvre du projet	10
1. La gestion du projet	10
2. Réalisation du projet.....	10
Conclusion	20

Table des figures

Figure 1 :OLAP	8
Figure 2 :le reporting	8
Figure 3 :Diagramme de GANTT	10
Figure 4 :Le schéma de nettoyage avec Pantaho	11
Figure 5 :Renommer les colonnes	11
Figure 6 :Suppression des colonnes.	11
Figure 7 :Suppression des lignes.	12
Figure 8 :Visualisation des données final.	12
Figure 9 : Schéma en étoile du projet.	13
Figure 10 :Construire les tables de dimensions avec Talend.	13
Figure 11 : TMap des tables de dimensions.	14
Figure 12 : construire la tables de fait.	14
Figure 13 : TMap de table de fait.	15
Figure 14 : Table dimension_Etude.....	15
Figure 15 :Table dimension_Alcool.....	15
Figure 16 : Table dimension_Activite.....	15
Figure17 :Tabledimension_Etudiant.	15
Figure 18 : Consommation par rapport au sexe.....	16
Figure 19 : Consommation au weekend par rapport au Age.	16
Figure 20 : Consommation au Travail par rapport au Age.	17
Figure 21 : Consommation par rapport au Garant	17
Figure 22 : Consommation par rapport au travail du père et de mère	17
Figure 23 : Consommation par rapport aux activités parascolaires.....	18
Figure 24 : Consommation par rapport au statut des parent.	18
Figure 25 : Consommation par rapport au Qualité de relation familiales.	19
Figure 26 : L'état de santé par rapport au consommation d'alcool.....	19

Introduction

Les entreprises utilisent une myriade d'applications et d'outils pour gérer leur activité au quotidien (CRM, ERP, Suite Office). Si chacune de ces applications permet de stocker, analyser, ou modifier certains types de données, ces dernières ne sont pas nécessairement compatibles entre elles. Plus encore, chaque service, équipe ou département peut utiliser un panel d'applications, parfois différents des autres entités de l'entreprise. La volumétrie de données collectées ou créées, l'absence d'uniformisation et la multiplicité des applications utilisées rendent difficile l'exploitation et l'analyse globale des données par les décideurs de l'entreprise, c'est là que l'informatique décisionnelle intervient afin d'exploiter tout le potentiel des technologies et des utilisateurs pour collecter et analyser ces données qui seront ensuite utilisées dans le processus stratégiques des organisations et leurs prises de décisions au quotidien.

Ce projet sous le thème La consommation d'alcools chez les étudiants a pour objet d'appliquer le processus d'informatique décisionnelle afin d'avoir des informations pertinentes qui vont nous aider à prendre des décisions plus avisées.

Ce rapport comporte trois chapitres complémentaires :

Chapitre 1 : Présentation du projet.

Chapitre 2 : Théorie d'informatique décisionnelle.

Chapitre 3 : Mise en œuvre du projet.

Chapitre 1 : Présentation du projet

1. Vue général sur le projet

L'alcool est une substance psychoactive capable d'entraîner la dépendance, largement utilisée dans de nombreuses sociétés depuis des siècles. L'usage nocif d'alcool est parmi les principaux problèmes de santé publique mondiaux. Il entraîne une charge de morbidité ainsi qu'un fardeau économique et social important.

Les problèmes liés à l'alcool et en particulier la consommation excessive d'alcool touchent toutes les catégories sociales notamment la population estudiantine. Cette population composée essentiellement de jeunes adultes soustraits au contrôle parental et intégrant un nouvel environnement représente une classe particulière de la population qui vit une période marquée par de nombreux changements où l'alcool est souvent présent. En effet, en comparaison de la population générale, les étudiants sont particulièrement exposés aux risques d'une consommation abusive d'alcool. De plus, les études épidémiologiques indiquent qu'une initiation précoce à l'alcool et une consommation excessive chez les jeunes sont des facteurs de risques majeurs d'usages problématiques ultérieurs [4-6].

Cela renforce l'intérêt de ce travail ayant pour objectifs de déterminer la prévalence de la consommation d'alcool chez les étudiants, et d'identifier les facteurs qui lui sont associés.

2. Objectif de projet

L'objectif de ce projet est :

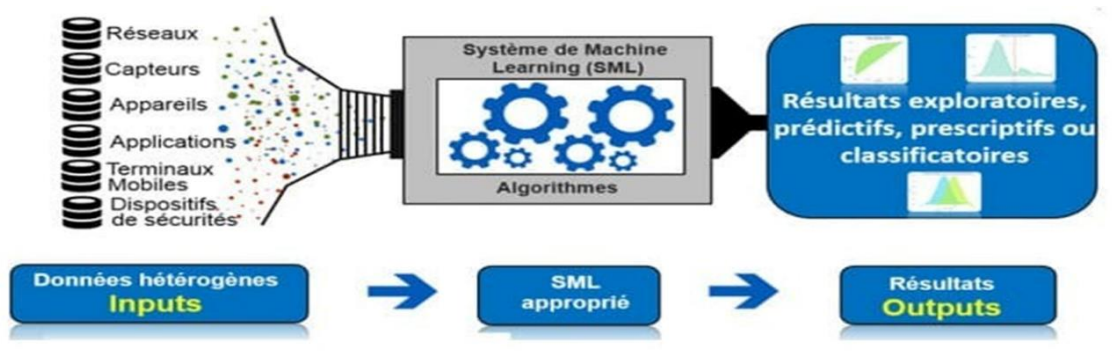
- ✓ **Découvrir la tranche d'âge qui consomme le plus d'alcool.**
- ✓ **Quel genre des étudiants le plus consomme d'alcool.**
- ✓ **Comment les activités d'un étudiant contribuent-elles à la consommation d'alcool.**
- ✓ **L'effet de la situation familiale sur la consommation d'alcool.**
- ✓ **L'impact de la consommation d'alcool sur les études.**

Chapitre 2 : Théorie d'informatique décisionnelle

1. L'informatique décisionnelle

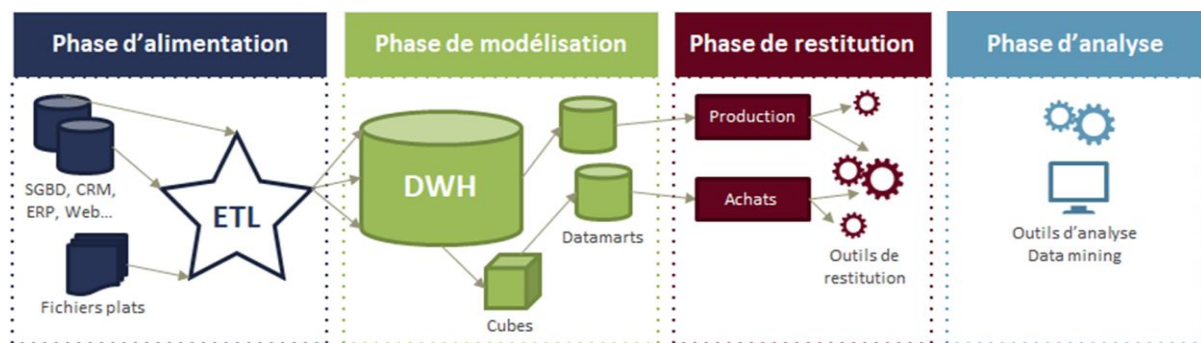
1.1 Définition

La Business Intelligence (BI) désigne les moyens, outils et méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre à un décideur d'avoir une vue d'ensemble de l'activité traitée.



1.2 La chaîne d'information décisionnelle

La chaîne d'information décisionnelle comprend différentes phases :



➤ La phase d'alimentation :

L'extraction de données de différentes sources de données, leur transformation et leur chargement à l'aide d'ETL.

➤ La phase de modélisation :

Rendre l'accès à la décision beaucoup plus facile et Immédiat grâce à une présentation graphique parlante.

➤ **La phase de restitution :**

Stocker les données de façon structurée pour des fins analytiques. Elles sont contenues dans un entrepôt de données qui est chargé de les centraliser.

➤ **La phase d'analyse :**

C'est la dernière phase. Elle concerne les utilisateurs finaux qui consultent juste les tableaux ou les graphiques qui ont été conçus au niveau de la phase de restitution.

1.3 Les notions clés de l'informatique décisionnelle

Pour comprendre la définition de l'informatique décisionnelle, il faut commencer par comprendre les notions clés.

1.3.1 La source de données

Une source de données est, comme son nom l'indique, l'endroit d'où provient la data utilisée dans le cadre de la Business Intelligence. Une source peut être un fichier, une base de données particulière sur un SGBD, ou même un flux de données transmises en direct. Enfin, il peut s'agir de l'endroit où les données ont été créées, mais aussi de celui où les informations physiques ont été numérisées.

1.3.2 Data Warehouse

Lieu de stockage de toutes les données utilisées par le système d'information décisionnel. Il permet aux applications d'aide à la décision de bénéficier d'une source d'information homogène, commune, normalisée et fiable. De plus, le data warehouse assure une étanchéité entre le système opérationnel et le système décisionnel ; et donc réduit le risque que les outils décisionnels affectent les performances du système opérationnel en place.

1.3.3 Datamart

Petits « magasins » de données dont l'ensemble forme le datawarehouse. Ils sont un sous-ensemble du data warehouse et suivent donc les mêmes principes clés. La différence entre les deux est que le datamart répond à un besoin métier plus spécifique que le data warehouse.

1.3.4 OLAP (Online Analytical Processing)

Permet aux utilisateurs d'analyser des données présentes de plusieurs systèmes de bases de données en même temps. Les données OLAP sont multidimensionnelles, ce qui signifie que l'information peut être comparée de nombreuses façons différentes. Par exemple, une entreprise peut comparer ses ventes d'ordinateurs en juin avec ses ventes en juillet, puis comparer ces résultats avec les ventes d'un autre endroit, qui pourraient être stockées dans une base de données différente.

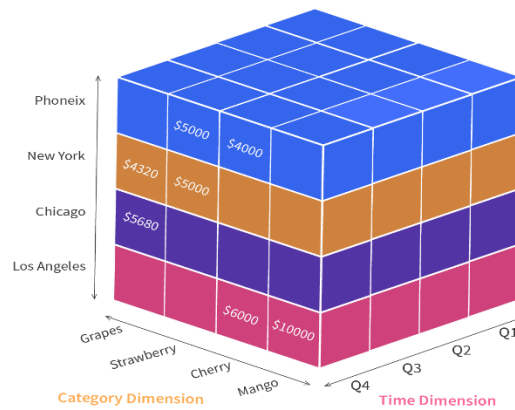


Figure 1 :OLAP

1.3.5 Le reporting

Le reporting, traduit en français communication de donnée, permet de mettre en scène des données récupérées sur une période souhaitée et de les présenter de manière claire afin qu'elles puissent être analysées et exploitées par une tierce personne. Il permet de rendre compte périodiquement des indicateurs de performance. Rapport d'activité, le reporting est un outil d'aide à la prise de décision. Le reporting constitue donc un système de données vérifiées et vérifiables. Il contient des données réelles qui auront préalablement été récoltées et triées.



Figure 2 :le reporting

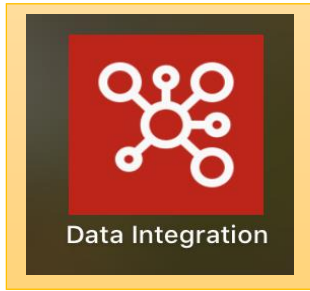
1.3.6 Les Tableaux de bords (Dashboards)

Un dashboard (tableau de bord) est un ensemble de rapports visuels qui affichent des indicateurs de performance (KPI) importants, généralement en temps réel. Les dashboards vous aident à visualiser et à approfondir rapidement les informations contenues dans vos données afin de répondre à vos questions business.

1.3.7 Les indicateurs

Les indicateurs clés de performance sont les indicateurs sur lesquels les collaborateurs d'une entreprise se basent afin de savoir si leurs objectifs ont bien été atteints.

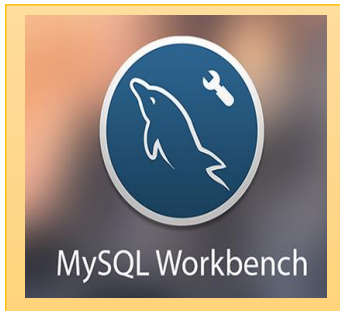
2. L'environnement des logiciels



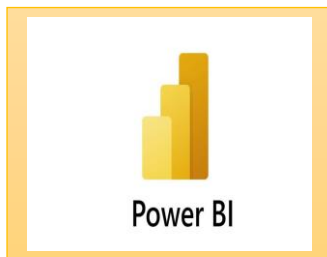
Pentaho Data Intégration (PDI), longtemps connu sous le nom de Kettle, est un ETL open source qui permet de concevoir et d'exécuter des opérations de manipulation et de transformation de données. Au moment où nous écrivons ces lignes, Pentaho Data Intégration est disponible dans sa version 5.0.



Talend Open Studio for Data Integration est un logiciel open source de type **ETL (Extract Transform Load)** développé par la société Talend. Cet outil permet l'échange inter-applicatif de données et facilite la migration d'outils métier. Est un outil permettant la gestion des données hétérogènes ou homogènes au sein d'un système d'information.



MySQL Workbench est un logiciel de gestion et d'administration de bases de données MySQL. Il permet, entre autres, de créer, modifier ou supprimer des tables, des comptes utilisateurs, et d'effectuer toutes les opérations inhérentes à la gestion d'une base de données. Pour ce faire, il doit être connecté à un serveur MySQL.



Power BI est une solution de Business Intelligence (BI) développée par Microsoft pour permettre aux entreprises de consolider, d'analyser, de visualiser et diffuser leurs données.

Chapitre 3 : Mise en œuvre du projet

1. La gestion du projet

La gestion de projet, ou le management de projet, consiste à organiser le déroulement d'un projet d'A à Z, de sa phase d'analyse à sa phase finale.

Dans notre projet nous avons segmenté le travail à plusieurs tâches pour faciliter sa réalisation.

Le diagramme de GANTT suivant présente les différentes tâches de notre projet :

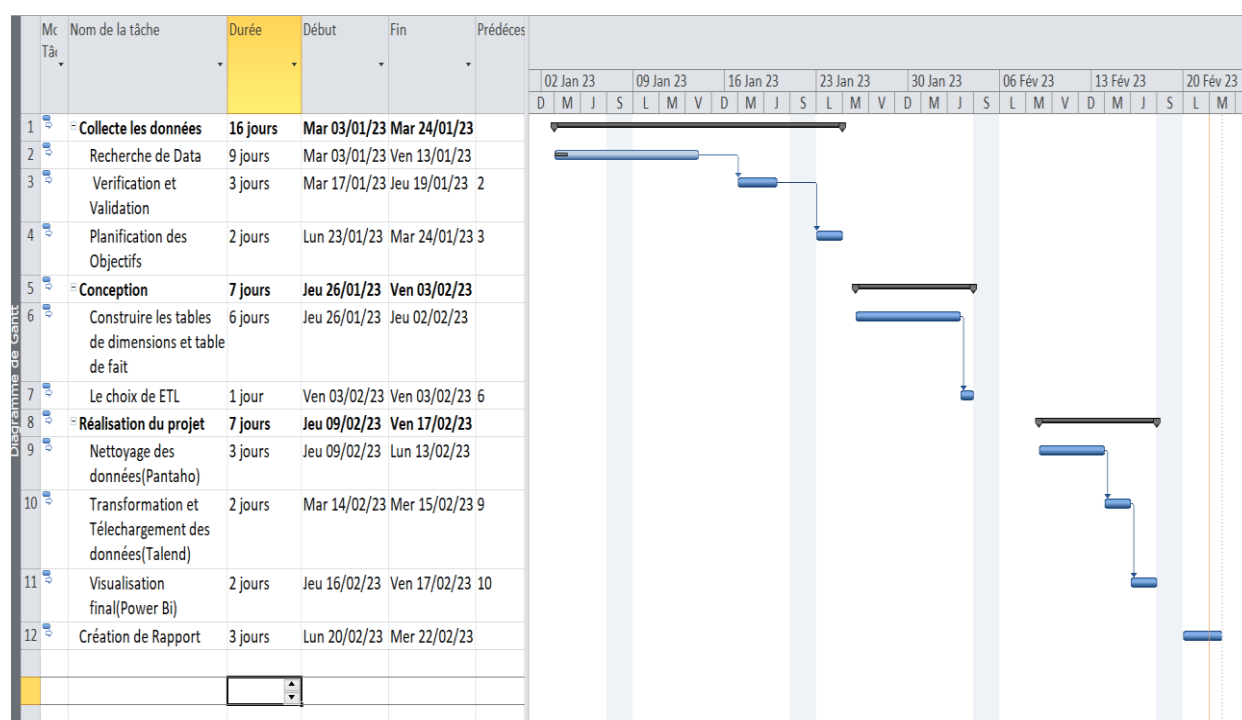


Figure 3 :Diagramme de GANTT

2. Réalisation du projet

2.1 Nettoyage des données

Les données sous forme d'un seul fichier CSV qui contient 33 colonnes et environ 400 données sur la consommation d'alcool chez les étudiants.

Comme première étape, utilisation de Pentaho spoon data integration a pour but d'effectuer les différentes transformations, par exemple (Le changement du nom des colonnes, annulation des valeurs des nulles, la suppression des colonnes inutiles...)

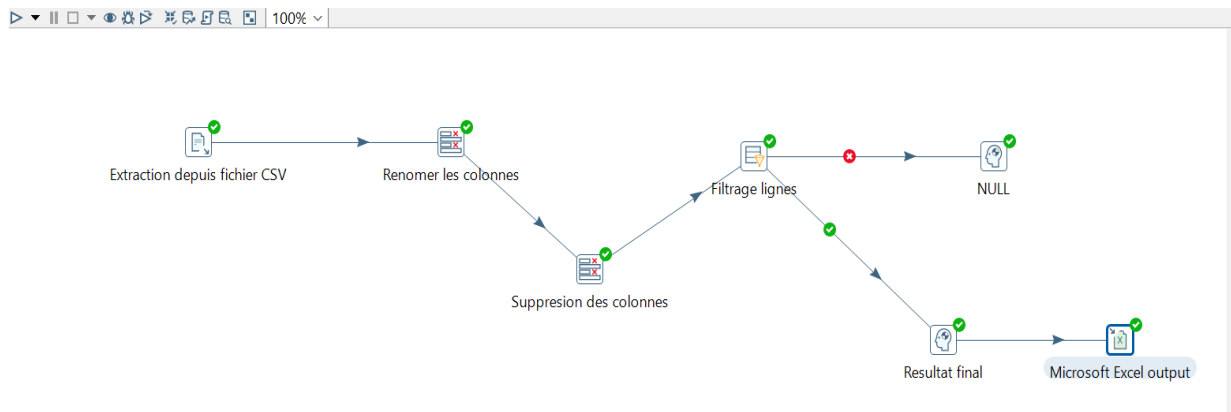


Figure 4 :Le schéma de nettoyage avec Pantaho

Premièrement nous avons renommé les colonnes de la langue anglais au français en utilisant option de « Altération structure flux »

Nom étape: Renommer les colonnes

#	Nom champ	Renommer en	Type	Longueur	Précision	Binaires à Normal	Format	Interpréter une date incorrecte	Date Locale	Date Time Zone
1	school	école	None	2		N		N		
2	sex	sexe	None	1		N		N		
3	age	Age	None	15	0	N		N		
4	address	Adress	None	1		N		N		
5	famsize	Taille de famille	None	3		N		N		
6	Pstatus	Statut des parent	None	1		N		N		
7	Medu	Education de mère	None	15	0	N		N		
8	Fedu	Education du père	None	15	0	N		N		
9	Mjob	Travail de mère	None	8		N		N		
10	Fjob	Travail du père	None	8		N		N		
11	reason	raison de choix de l'école	None	10		N		N		
12	guardian	Tuteur	None	6		N		N		
13	traveltime	Temps de trajet	None	15	0	N		N		
14	studytime	Temps d'étude	None	15	0	N		N		
15	failures	Echecs de classe	None	15	0	N		N		
16	schoolsup		None	3		N		N		
17	famsup		None	3		N		N		
18	paid		None	3		N		N		
19	activities	Activités parascolaires	None	3		N		N		
20	nursery		None	3		N		N		
21	higher	L'étude sup	None	3		N		N		
22	internet	Accès internet	None	3		N		N		
23	romantic		None	3		N		N		
24	famrel	Qualité de relation familiales	None	15	0	N		N		
25	freetime	Temps libre	None	15	0	N		N		
26	onout	Sortir avec les amis	None	15	0	N		N		

Récupérer champs

Figure 5 :Renommer les colonnes .

Puis nous avons utilisé aussi l'option de « Altération structure flux »,pour supprimer les colonnes inutiles .

Nom étape: Suppression des colonnes

#	Nom champ
1	schoolsup
2	famsup
3	paid
4	nursery
5	romantic

Figure 6 :Suppression des colonnes.

Nous avons utilisé le « Contrôle des flux » a pour but de faire filtrage des lignes qui contient des valeurs nulles.

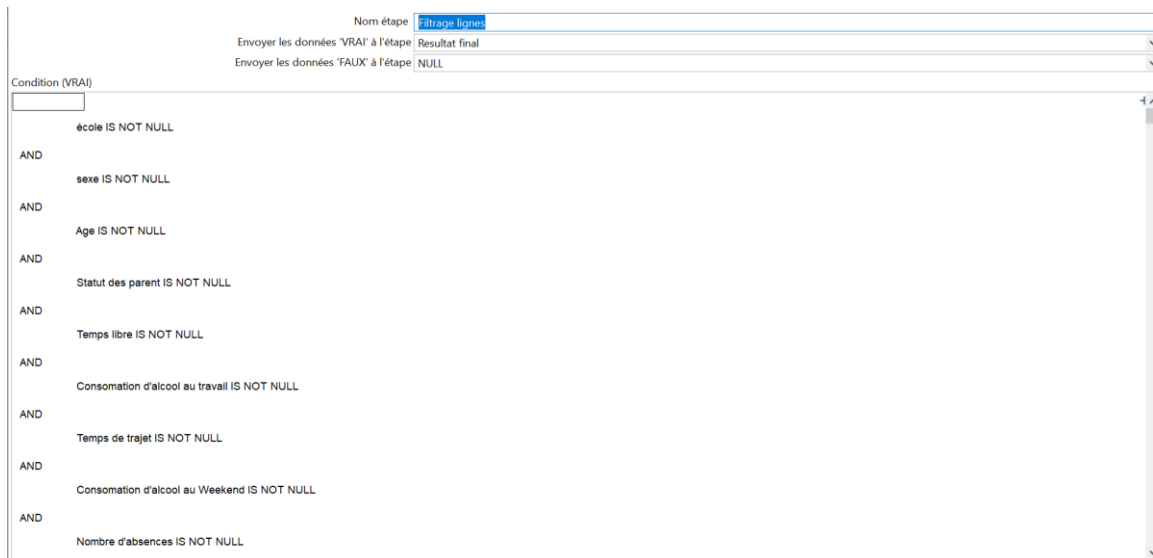


Figure 7 :Suppression des lignes.

Après nous avons stocker les données obtenus dans un fichier Excel.

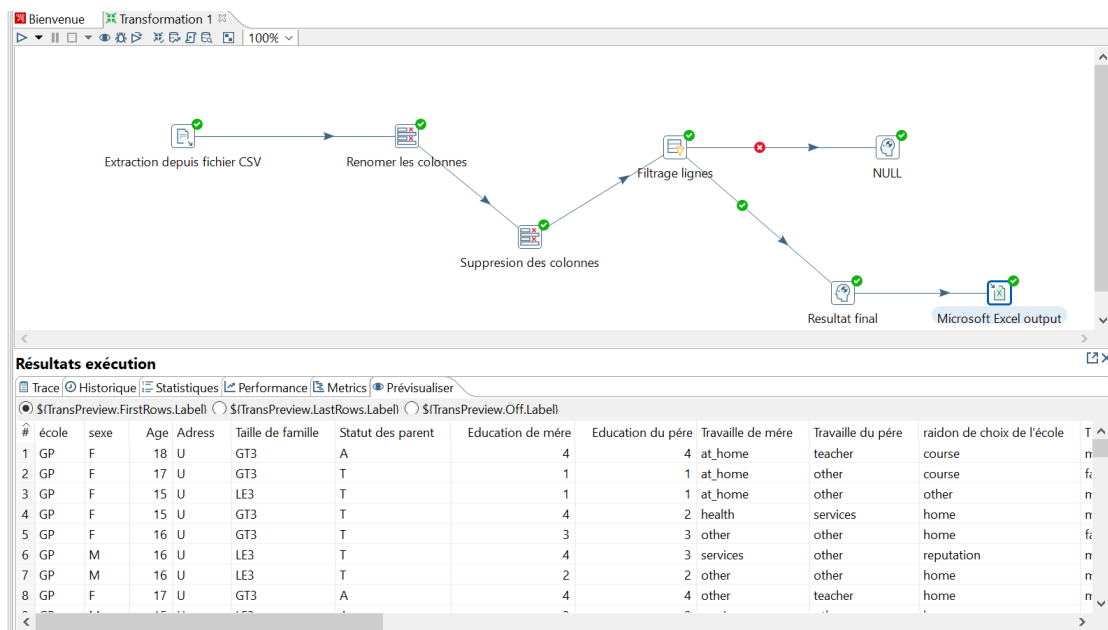


Figure 8 :Visualisation des données final.

2.2 Conception de base de données

2.2.1 Schéma en étoile

Le schéma en étoile est une façon de mettre en relation les dimensions et les faits dans Un entrepôt de données. Le principe est que les dimensions sont directement reliées à un fait (schématiquement, ça fait comme une étoile).

Cette figure ce dessous montre les tables des dimensions qui sont les axes d'analyse sur lesquels on veut étudier nos données observables :

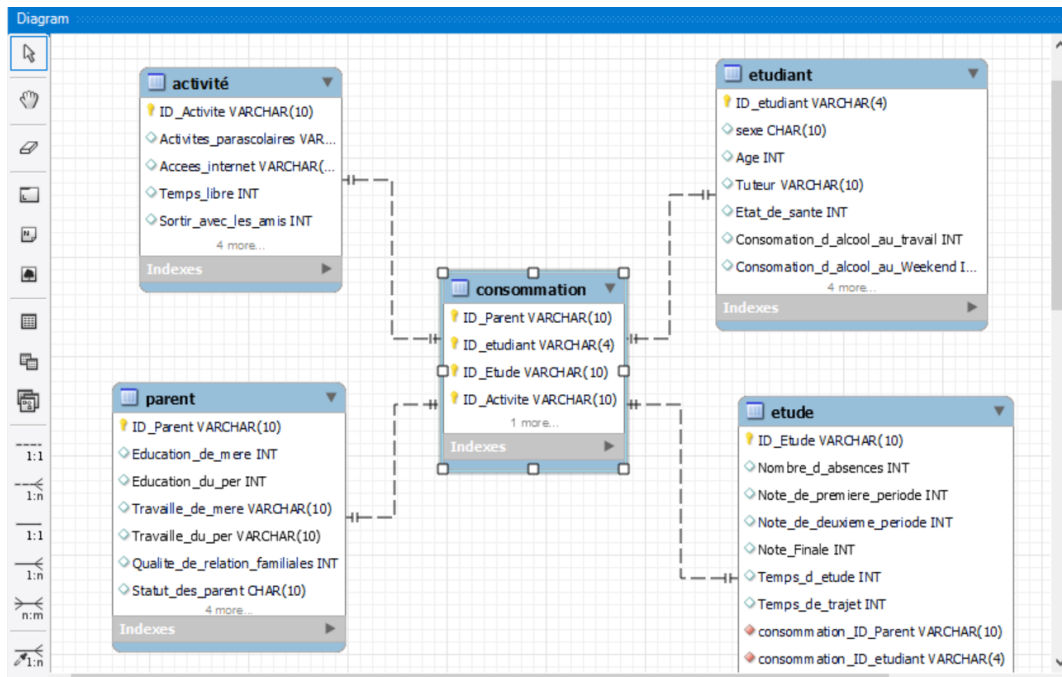


Figure 9 : Schéma en étoile du projet.

2.3 Réalisation du Data Warehouse

Après avoir collecter et le nettoyer de données. On l'utilise sur logiciel Talend comme une source de donnée pour but de la création du jobs et le chargement de Data dans MySQL Workbench.

❖ Création de Job qui construire les tables dimension

Après la création de job et la donnée un nom, nous avons importé le fichier des données afin de le glisser ,et nous avons utilisé TMap pour construire les tables de dimensions et stocker dans la base de données (Après la liaison avec notre ase de données)

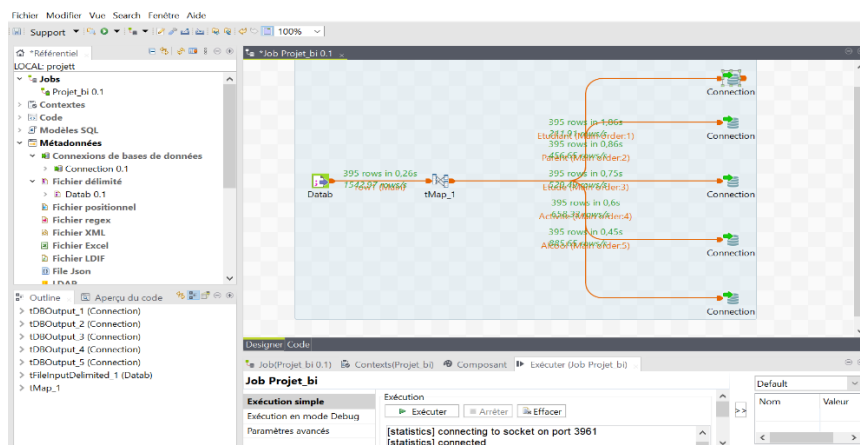


Figure 10 :Construire les tables de dimensions avec Talend.

Le figure suivant montre le contenu de TMap pour construire les tables de dimensions :

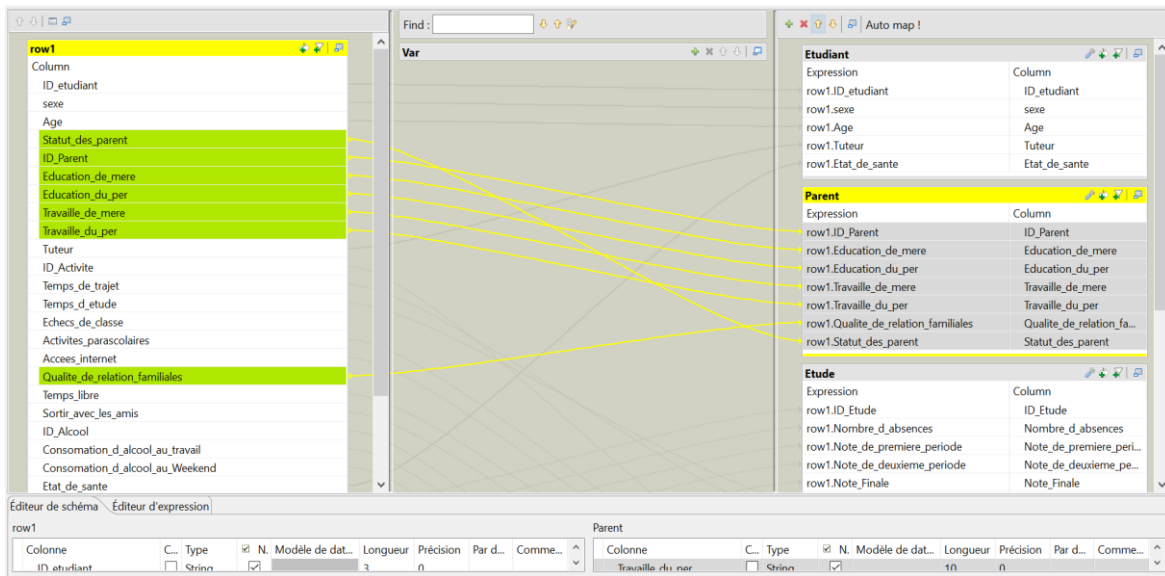


Figure 11 : TMap des tables de dimensions.

❖ La création de table de fait

Avant de créer la table de fait nous avons récupéré les schémas depuis la base de données, puis nous avons les glisser et utiliser TMap.

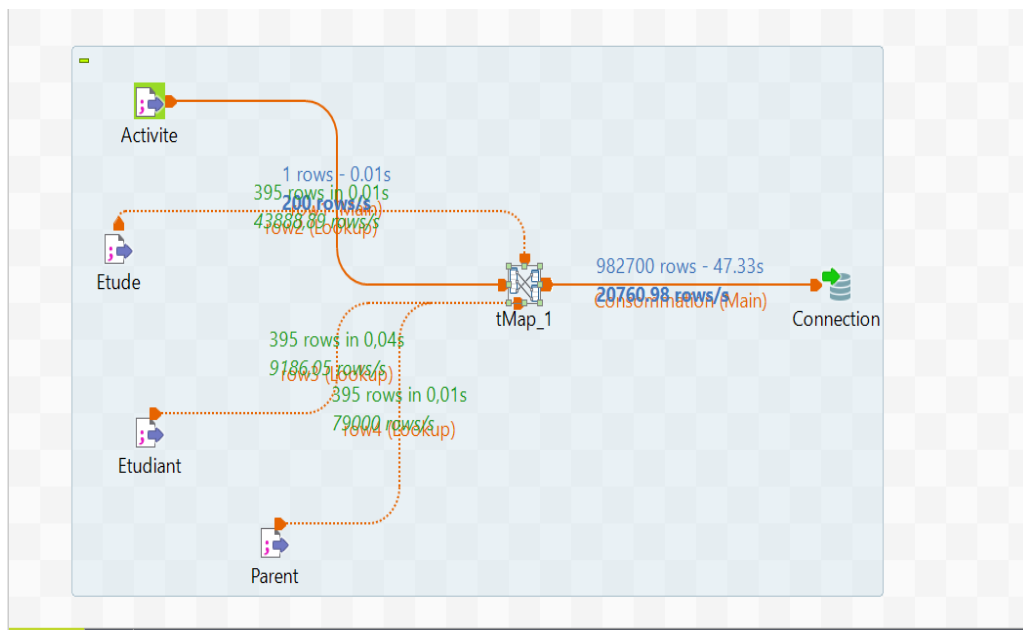


Figure 12 : construire la tables de fait.

Le figure suivant montre le contenu de TMap pour construire la table de fait :

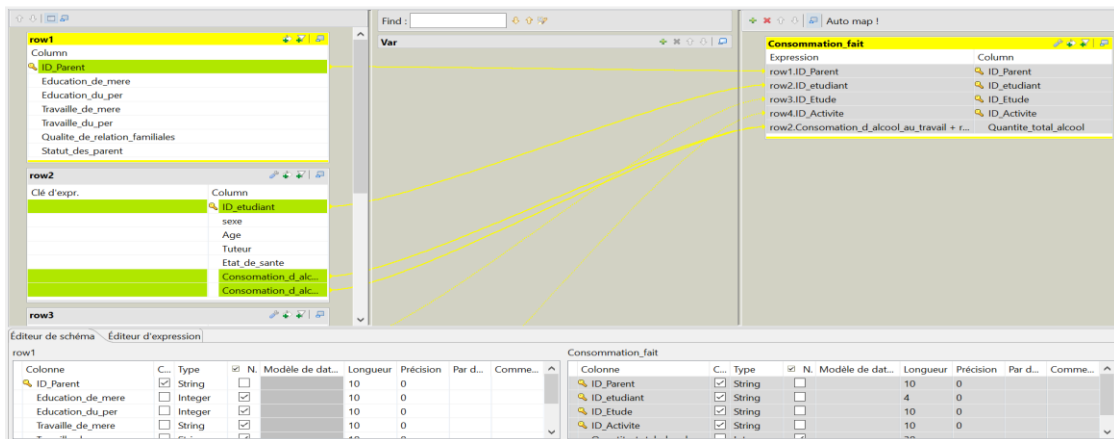


Figure 13 : TMap de table de fait.

Voilà l'ensemble des tables de dimensions et table de fait :

Figure 14 : Table dimension_Etude.

Figure 15 : Table dimension_Alcool.

Figure 16 : Table dimension_Activite.

Figure 17 : Table dimension_Etudiant.

2.4 Visualisation

La visualisation (Reporting), c'est la dernière étape dans la réalisation du projet qui s'intéresse à l'analyse des données sous forme des graphiques ou tabulaire. En se base sur notre Data Warehouse pour faire la visualisation des solutions à l'aide de Power BI.

Après l'importation de Data nous avons fait la visualisation de la consommation d'alcool en travail et en weekend par rapport à des facteurs.

➤ La consommation d'alcool en travail et en weekend par rapport au sexe :

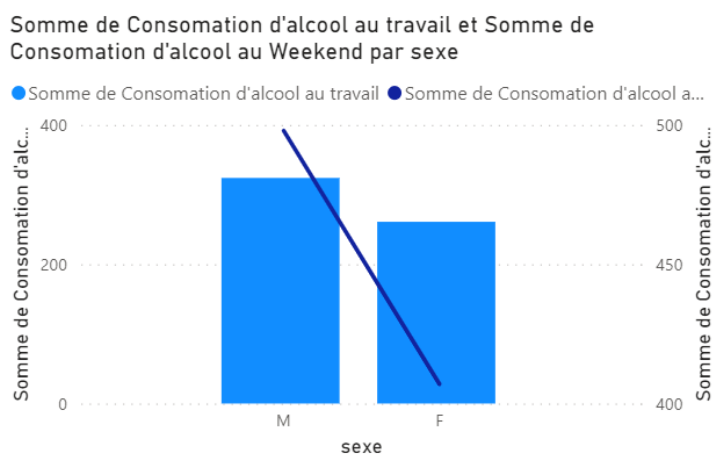


Figure 18 : Consommation par rapport au sexe.

➤ La consommation d'alcool en travail et en weekend par rapport à l'âge :

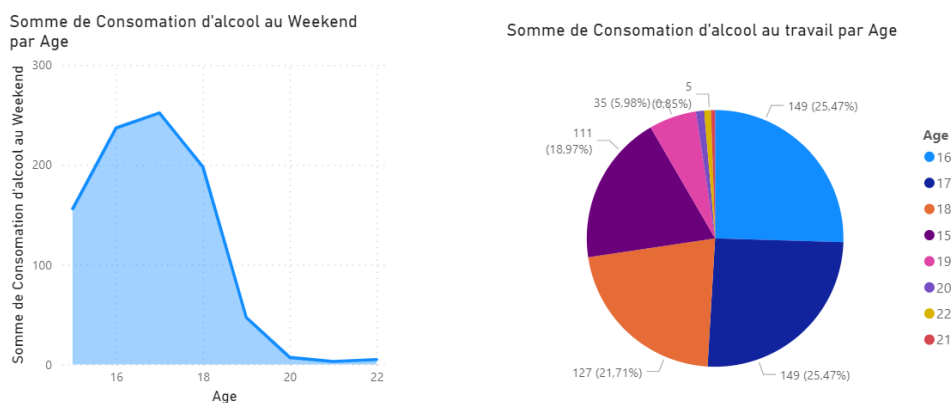


Figure 19 : Consommation au weekend par rapport au Age.

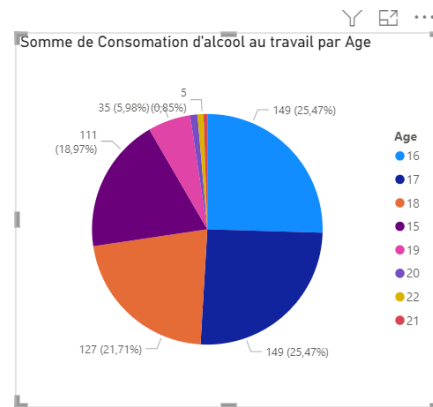
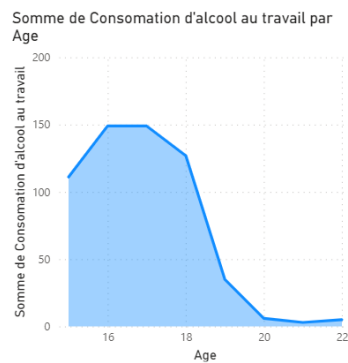


Figure 20 : Consommation au Travail par rapport au Age.

➤ **La consommation d'alcool en travail et en weekend par rapport au tuteur(garant) :**

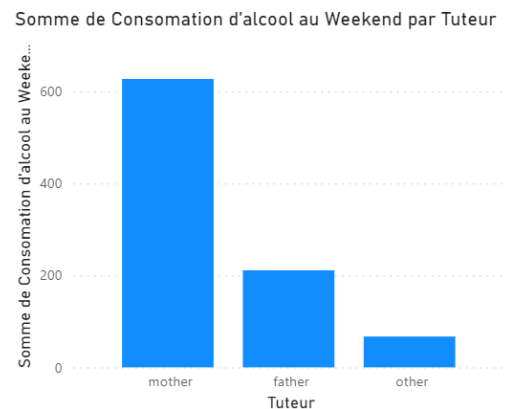
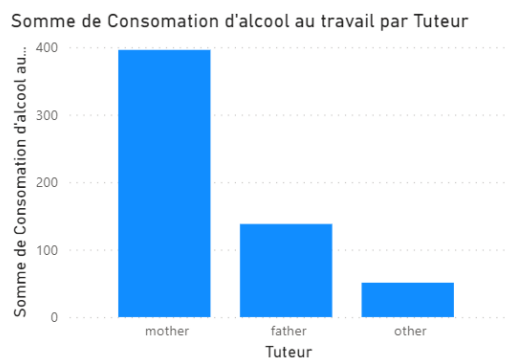


Figure 21 : Consommation par rapport au Garant .

➤ **La consommation d'alcool en travail et en weekend par rapport au travail du père et de la mère :**

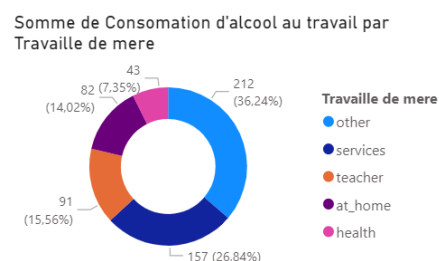
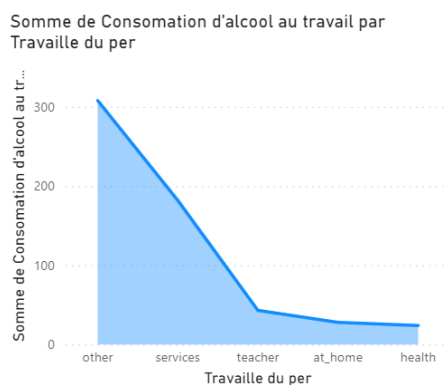


Figure 22 : Consommation par rapport au travail du père et de mère .

➤ **La consommation d'alcool en travail par rapport aux activités parascolaires :**

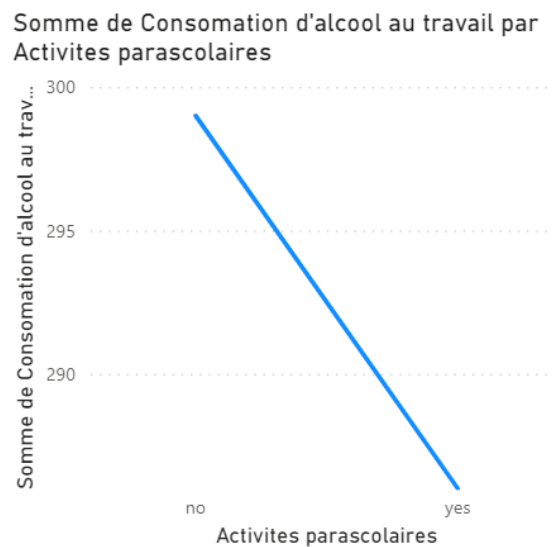


Figure 23 : Consommation par rapport aux activités parascolaires.

➤ **La consommation d'alcool en travail et en weekend par rapport au statut des parents :**

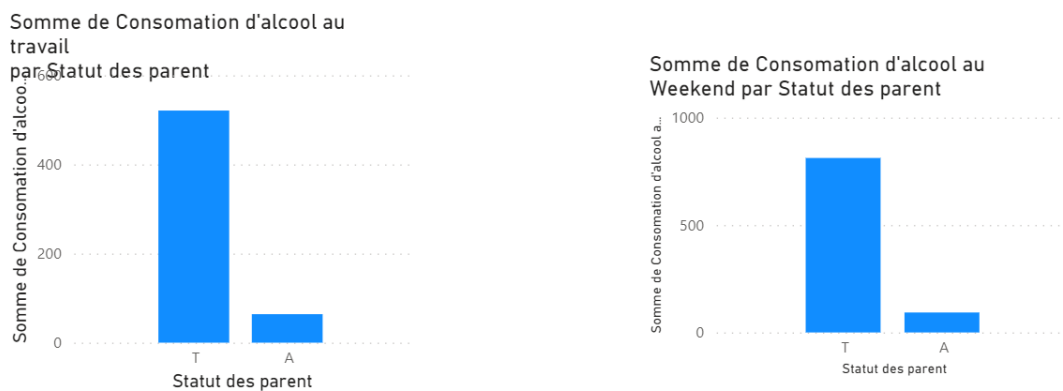


Figure 24 : Consommation par rapport au statut des parent.

➤ **La consommation d'alcool en travail et en weekend par rapport au qualité de relation familiales :**

Somme de Consommation d'alcool au travail et Somme de Consommation d'alcool au Weekend par Qualite de relation familiales

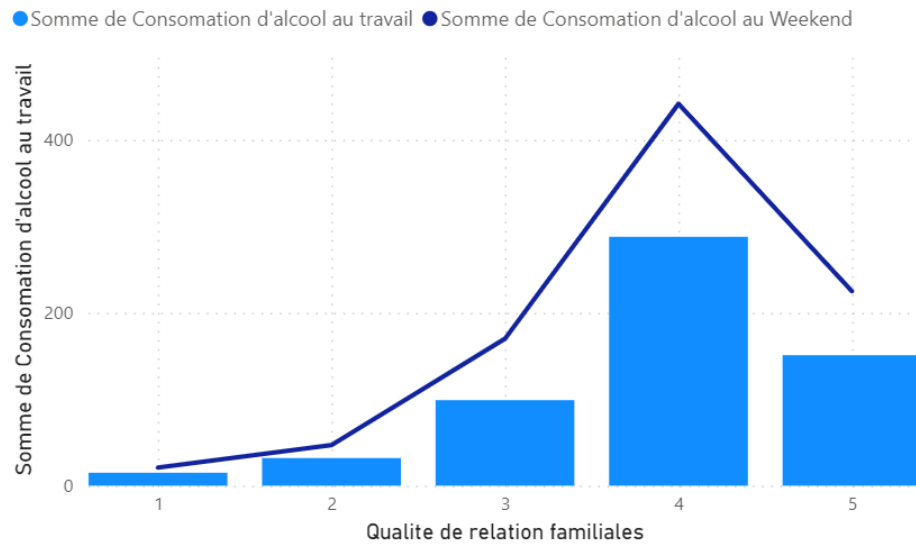


Figure 25 : Consommation par rapport au Qualité de relation familiales.

➤ L'état de santé par rapport au consommation d'alcool :

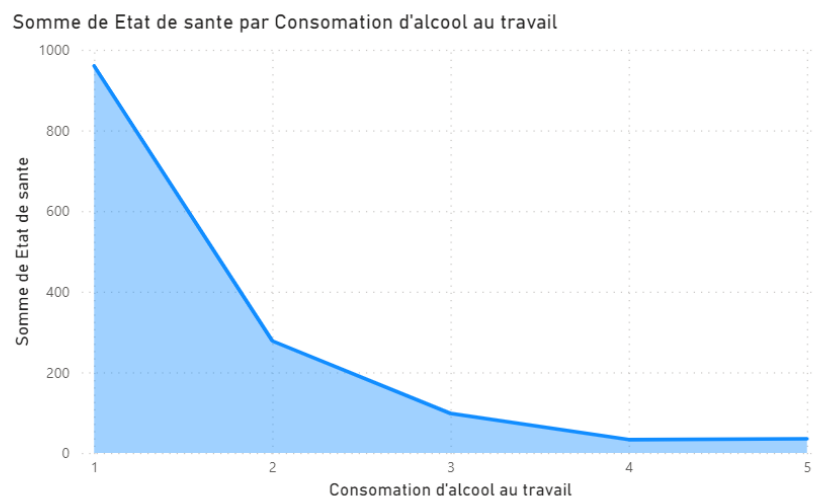


Figure 26 : L'état de santé par rapport au consommation d'alcool.

2.5 Quelques décisions :

- Les parents doivent contrôler leurs enfants dans la période [15ans-18ans].
- Le père est le meilleur garant pour leurs enfants.
- Les étudiants doivent engager dans les activités parascolaires.

Conclusion

Ce projet nous a permis de découvrir le processus qui aide les décideurs à prendre la décision la plus performante et la plus efficace. Alimenter le DataWarehouse, la création du cube OLAP ainsi que la création des différents rapports sont les étapes globales de projet.

Ce projet nous a permis d'approfondir nos connaissances acquises en matière de l'informatique décisionnelle.