



Informatique décisionnelle et datamining

Projet de demi module

Réalisée par : El Belhadji Soumaya

Master : MBD

Année universitaire : 2021-2022

Le but de projet :

Le projet consistait à concevoir et réaliser une solution décisionnelle complète permettant d'analyser facilement les indicateurs de performance de la faculté des sciences est techniques.

On s'intéressera tout particulièrement à trois thèmes : inscription, réussite et Lauréats.

L'objectif est de générer des tableaux de bord qui offrent une meilleure visibilité pour les dirigeants.

Les données à analyser sont à extraire à partir d'un fichier Excel tiré du système apogée de la faculté de sciences et de techniques.

Environnement de travail :

La génération des rapports se fait avec Tableau et Microsoft Power Bi.

Analyse de datamining se fait par le clustering et les arbres de décisions pour la prédiction en utilisant SQLServer et Visual Studio.



Tableau Desktop est un logiciel pour votre ordinateur sous Windows, ou votre Mac, qui vous permet de créer des tableaux de bord à partir de n'importe quelles bases de données, qu'elles soient locales ou en ligne (dans le cloud).

Entièrement configurables, vous pouvez y afficher les données de votre choix sous la forme qui vous parle le plus afin d'avoir toujours les bonnes informations sous la main quand vous en avez besoin.



Microsoft Power BI est une solution d'analyse de données de Microsoft. Il permet de créer des visualisations de données personnalisées et interactives avec une interface suffisamment simple pour que les utilisateurs finaux créent leurs propres rapports et tableaux de bord.

L'utilisation d'un tel outil est primordiale afin que les données internes et externes de l'entreprise puissent être comprises et analysées. Analyser des données dans une grille est difficile, mettre en forme pour mettre en valeur certaines informations permet en quelques secondes d'obtenir la donnée importante.



SSDT (SQL Server Data Tools) est un outil de développement moderne permettant de générer des bases de données relationnelles SQL Server, des bases de données dans SQL Azure, des modèles de données AS (Analysis Services), des packages IS (Integration Services) et des rapports RS (Reporting Services).

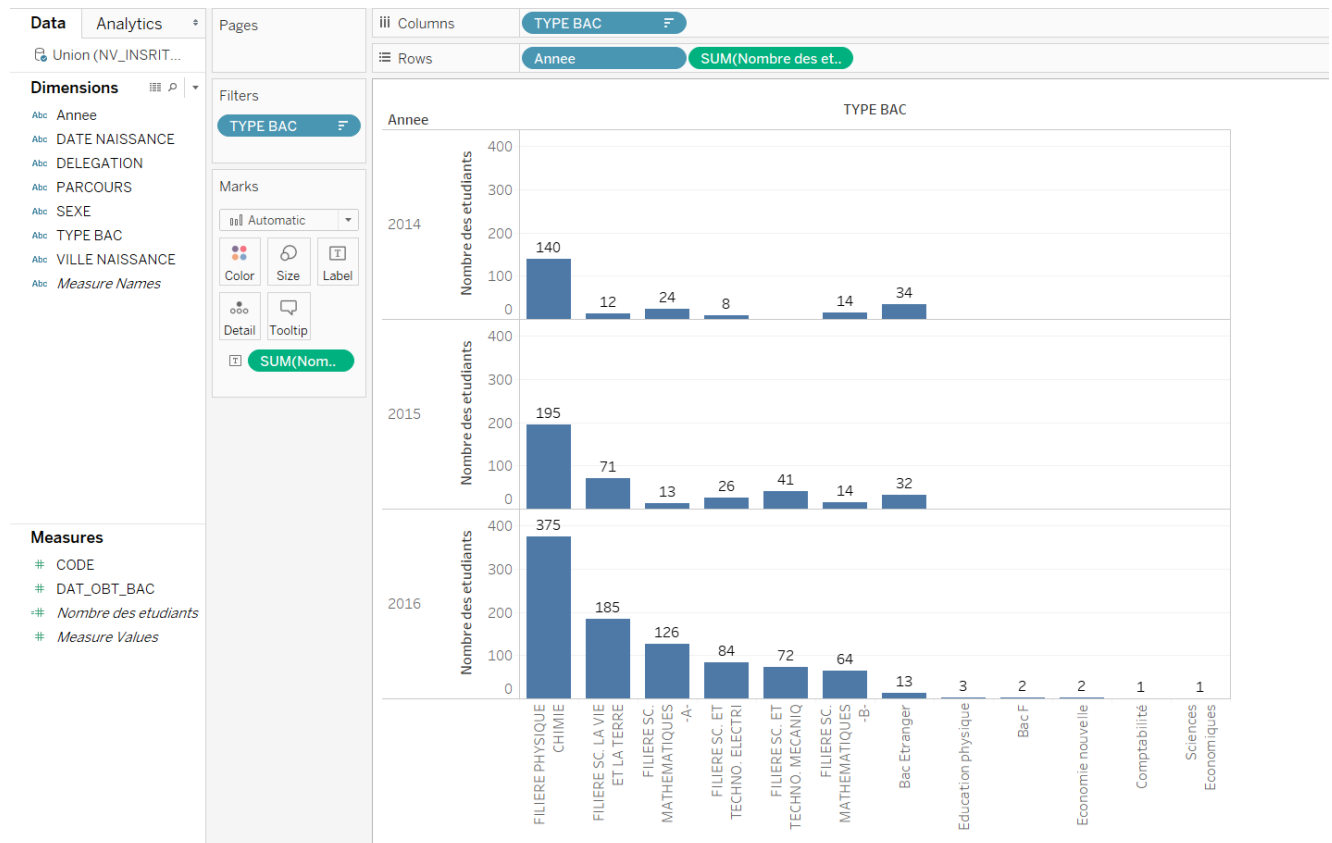
Microsoft SQL Server Analysis Services (SSAS) est un outil de traitement analytique en ligne et d'exploration de données dans Microsoft SQL Server. SSAS est utilisé comme un outil par les organisations pour analyser et donner un sens aux informations éventuellement réparties sur plusieurs bases de données, ou dans des tables ou des fichiers disparates. Analysis Services comprend un groupe de fonctionnalités OLAP et d'exploration de données et se décline en deux versions : multidimensionnelle et tabulaire.

Génération des rapports:

1-1 Partie Tableau :

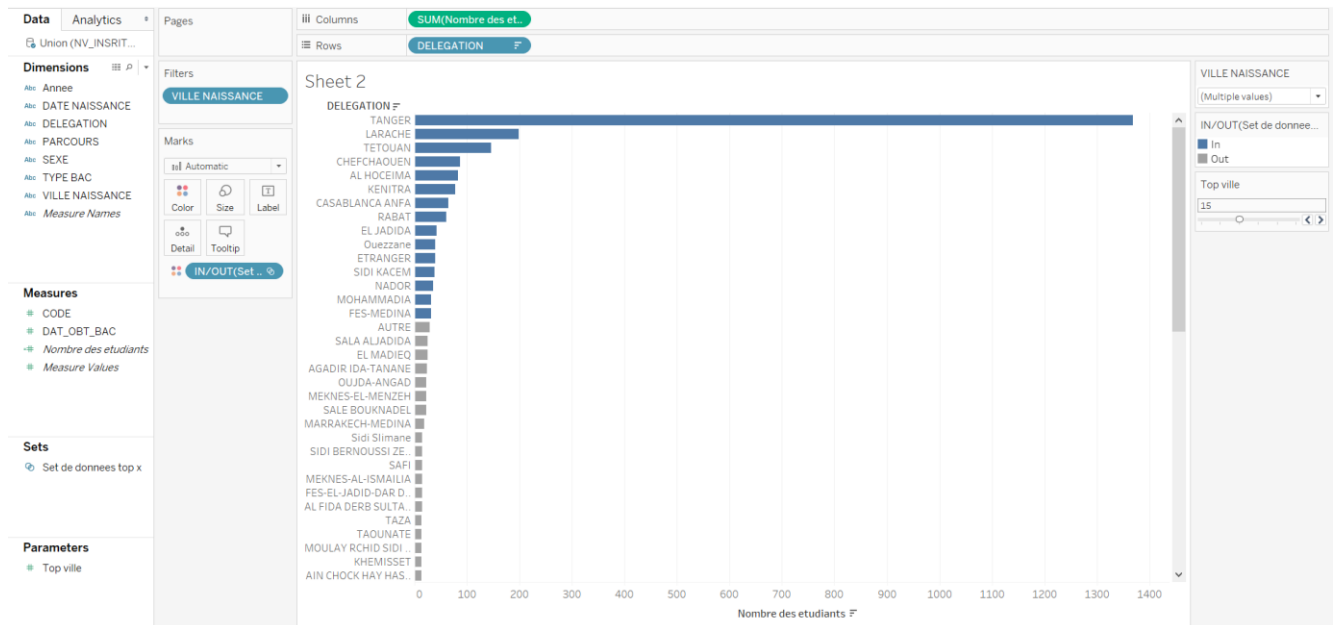
D'abord on va commencer par créer ces visualisations en utilisant Tableau Desktop :

→ **Taux d'inscription par rapport au type de bac et l'année :**



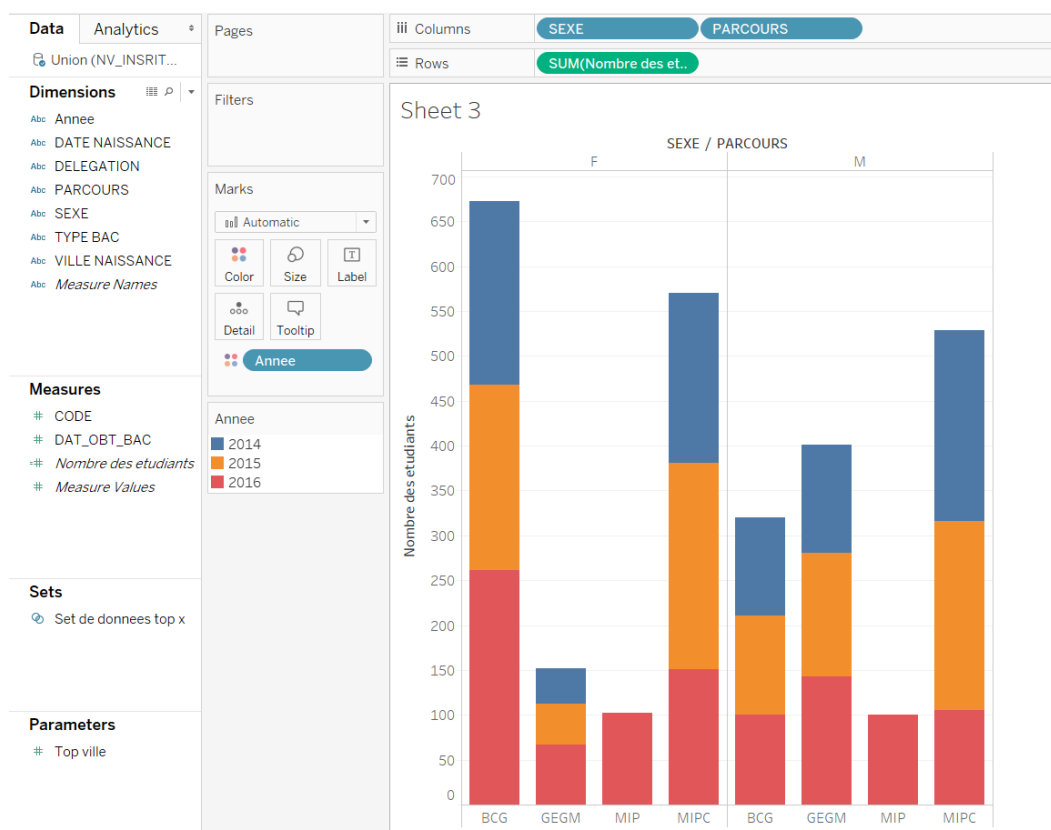
Ce graph montre que le taux d'inscription représenté par la somme de nombre des étudiants augmente généralement chaque année, avec la dominance de filière physique chimique comme série de bac.

→ **Taux d'inscription par rapport aux villes :**



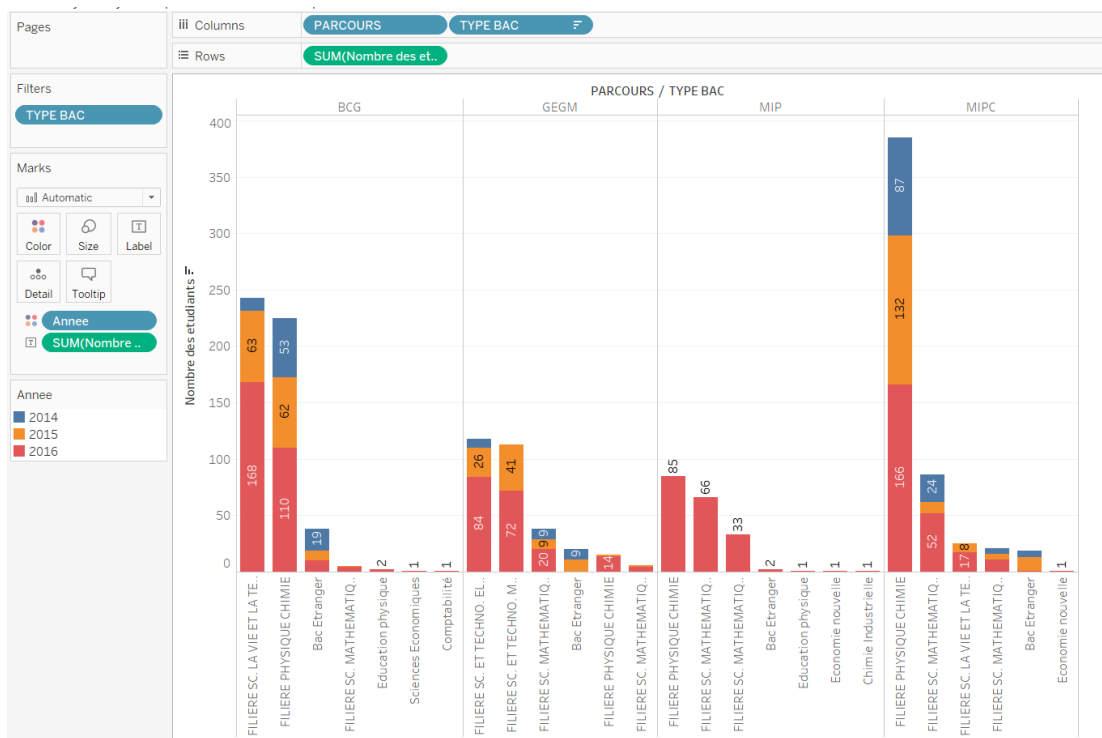
Le top 15 des villes inscrit est illustré dans ce graph, on remarque la dominance de Tanger et les villes du nord en général.

→ Taux d'inscription par rapport au sexe, parcours et année :



On remarque la dominance des filles dans les parcours BCG et MIPC , par contre au garçons qui inscrit plus fréquemment dans les parcours MIPC et GEGM.

→ Taux d'inscription par rapport au parcours et type de bac et l'année :

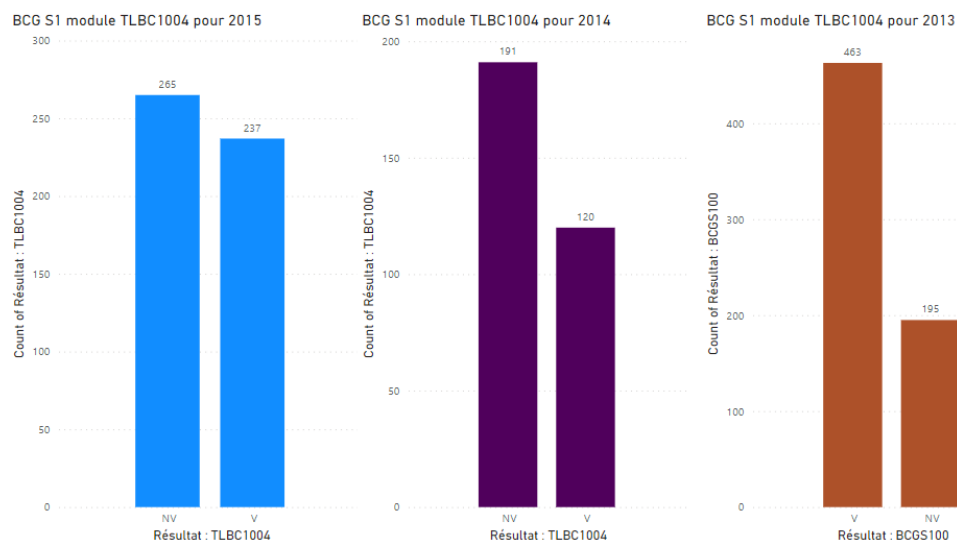


Cette visualisation affiche que les étudiants avec un bac Physique Chimique tend de s'inscrire au parcours MIPC, et les étudiants avec un bac science de la vie et de la terre tend de s'inscrire au parcours BCG.

1-2 Partie Power BI :

Maintenant on va travailler avec Microsoft Power Bi pour générer la suite des rapports :

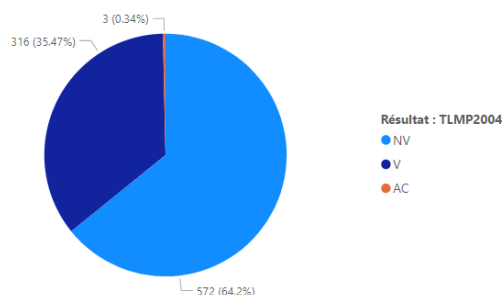
→ Taux de réussite et d'échec pour le même module à travers les années :



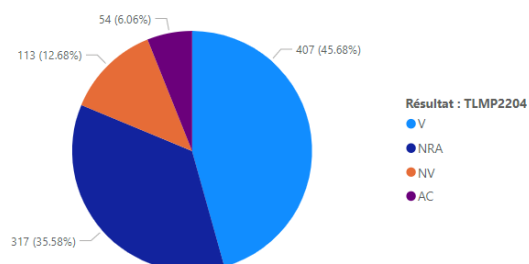
Le taux de réussite est indiqué par le nombre des validés, également pour le taux d'échec qui est le nombre des non validés.

→ Taux de réussite et d'échec et d'acquisition par rapport aux semestres de même année :

MIPC 2015 module TLMP2004 pour 2015

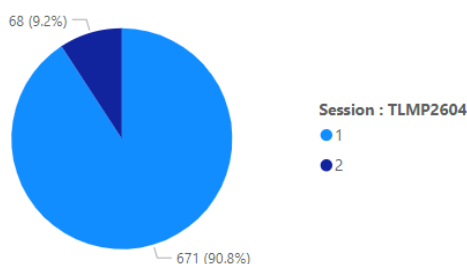


MIPC 2015 module TLMP2204 pour 2015

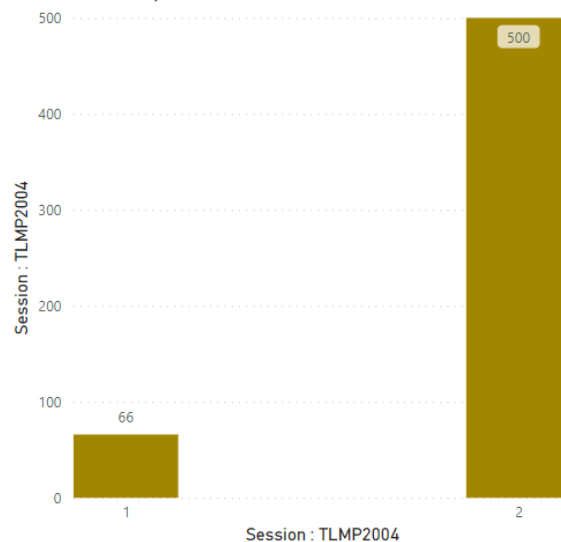


→ Taux de validation par rapport aux sessions :

nbr des validés par session module TLMP2604 MIPC



nbr des validés par session module TLMP2004 MIPC



On remarque que le taux de réussite dans la deuxième session est plus élevé pour le parcours MIPC.

→ Taux d'échec par rapport aux semestres :

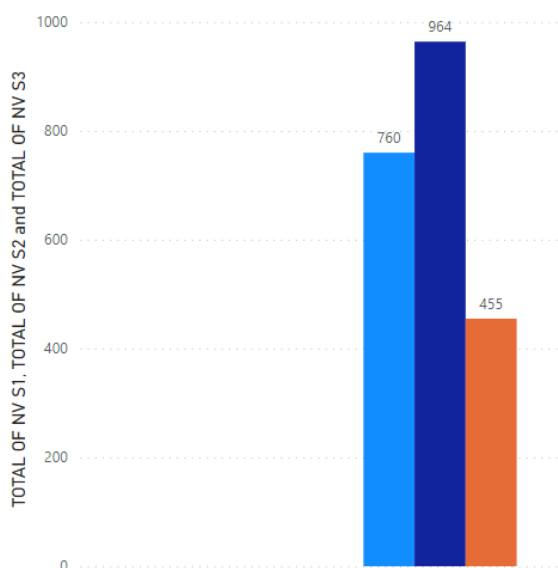
Pour cela nous avons créé une mesure appelée TOTAL OF NV qui calcule le nombre des non validé (NRA +NV) pour tous les modules de chaque semestre :

```
1 TOTAL OF NV = 'BCG S3 2013'[Count of NV1]+'BCG S3 2013'[Count of NV2]+'BCG S3 2013'[Count of NV3]+'BCG S3 2013'[Count of NV4]+'BCG S3 2013'[Count of NV5]
```

```
1 Count of NV1 = CALCULATE(COUNT('BCG S3 2013'[COD_ETU]),FILTER('BCG S3 2013','BCG S3 2013'[Résultat : BCGS300]="NV")) + CALCULATE(COUNT('BCG S3 2013'[COD_ETU]),FILTER('BCG S3 2013','BCG S3 2013'[Résultat : BCGS300]="NRA"))
```

TOTAL OF NV S1, TOTAL OF NV S2 and TOTAL OF NV S3

● TOTAL OF NV S1 ● TOTAL OF NV S2 ● TOTAL OF NV S3



On remarque que le nombre des NV dans la deuxième semestre est plus élevé que les autres semestres de l'année 2013 BCG.

Analyse Datamining :

Pour l'analyse des données, nous avons tout d'abord importé le fichier excel concernant les données d'inscription pour l'année 2014 au SQL Server , pour qu'elle serve comme la source des données de notre SSAS . Après la création d'un projet «Analysis Services Multidimensional and Data Mining » et le configurer proprement, on peut maintenant établir les analyses qu'on souhaite.

Object Explorer

Connect - [Server] [Database] [Tables] [Views] [External Resources] [Synonyms] [Programmability] [Service Broker] [Storage] [Security]

DESKTOP-A8B42ET (SQL Server 15.0.2000.5 - sa)

Databases

System Databases

Database Snapshots

mini-projet

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.'2014\$'

dbo.'2015\$'

dbo.'2016\$'

Views

External Resources

Synonyms

Programmability

Service Broker

Storage

Security

Security

Server Objects

Replication

PolyBase

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent

XEvent Profiler

SQLQuery1.sql - DES...ini-projet (sa (63))

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [CODE]
,[DATE_NAISSANCE ]
,[VILLE_NAISSANCE]
,[SEXE]
,[PARCOURS]
,[DAT_OBT_BAC]
,[TYPE_BAC]
,[DELEGATION]
FROM [mini-projet].[dbo].[ '2014$' ]

```

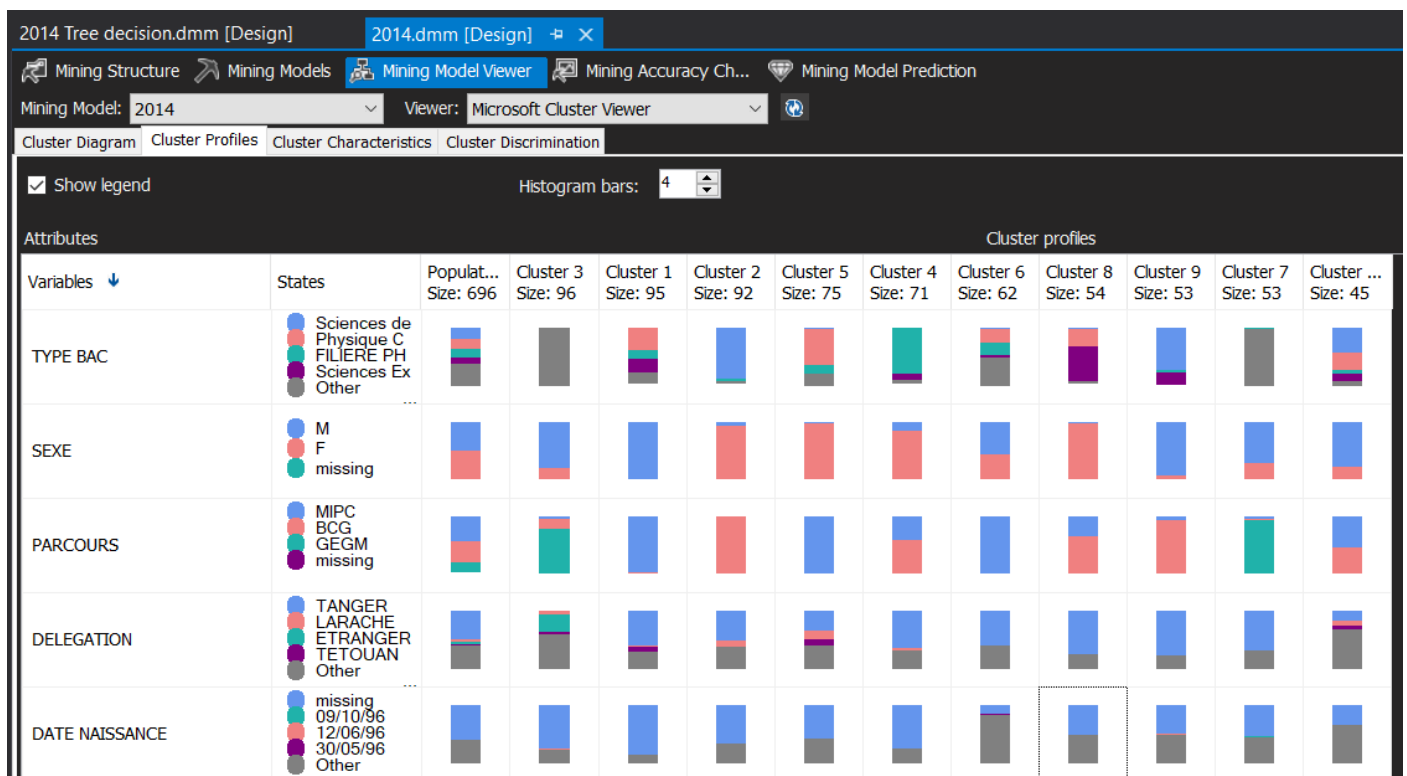
100 %

Results Messages

	CODE	DATE_NAISSANCE	VILLE_NAISSANCE	SEXE	PARCOURS	DAT_OBT_BAC	TYPE_BAC	DELEGATION
1	14007278	22/09/96	CASABLANCA	F	BCG	2014	Sciences Expérimentales	CASABLANCA ANFA
2	14007230	12/05/94	TMARA	F	MIPC	2014	FILIERE SC. MATHÉMATIQUES -A-	RABAT
3	14007228	18/03/95	TANGER	F	BCG	2014	Sciences Expérimentales	TANGER
4	14007131	18/09/94	RWANDA	M	GEGM	2014	Bac Etranger	AUTRE
5	14007110	17/01/95	TANGER	F	BCG	2014	Sciences Expérimentales	TANGER
6	14007105	27/06/94	KOUMONDE	M	BCG	2014	Bac Etranger	ETRANGER
7	14007104	14/02/97	EL JADIDA	F	GEGM	2014	FILIERE SC. ET TECHNO. ELECTRI	EL JADIDA
8	14006708	28/01/93	TANGER	M	MIPC	2014	Sciences Mathématiques (A)	TANGER
9	14006702	30/09/96	TANGER	M	BCG	2014	Sciences Expérimentales	TANGER
10	14006616	29/07/95	JAMAAT ZAAROURA	F	BCG	2014	Sciences de la Vie et de la Terre	TANGER
11	14006562	28/03/96	TOLAGNARO	F	BCG	2014	Bac Etranger	ETRANGER
12	14006561	31/10/95	TANAMBAO TOAMASINA	F	BCG	2014	Bac Etranger	ETRANGER
13	14006560	01/03/96	TOAMASINA	F	BCG	2014	Bac Etranger	ETRANGER

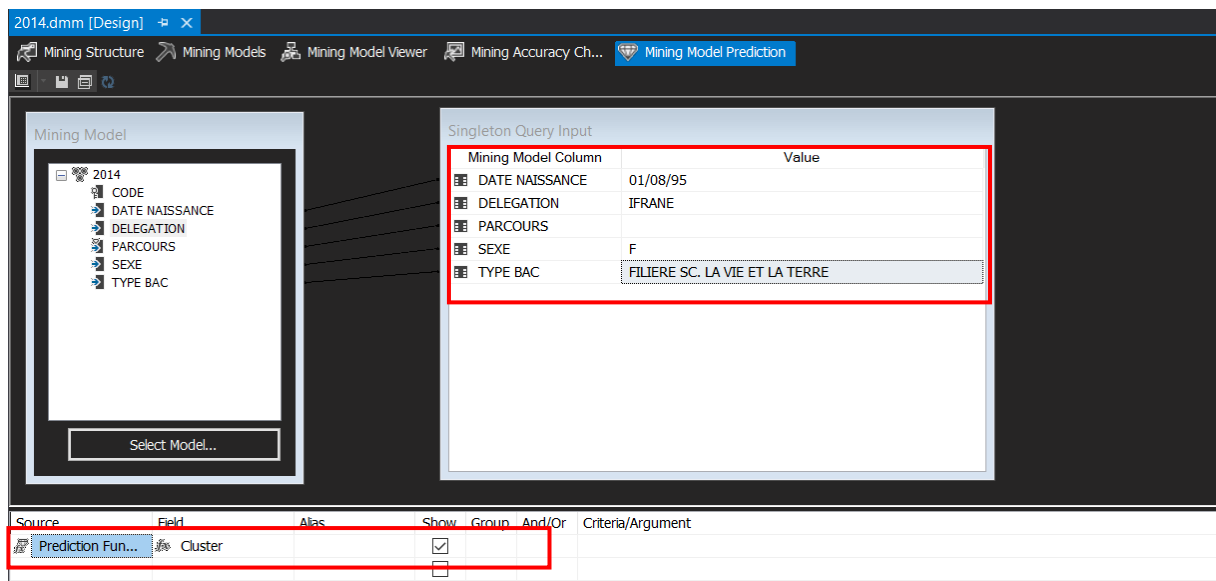
1-1 Recherche des clusters :

- 2 Nous avons choisi de créer des clusters entre le type de bac, sexe, parcours, délégation et la date de naissance, le système a généré 10 Clusters :



→ Prédiction par les clusters :

On fait entrer les valeurs de date de naissance, de délégation, de sexe et de type de bac pour prédire le parcours :



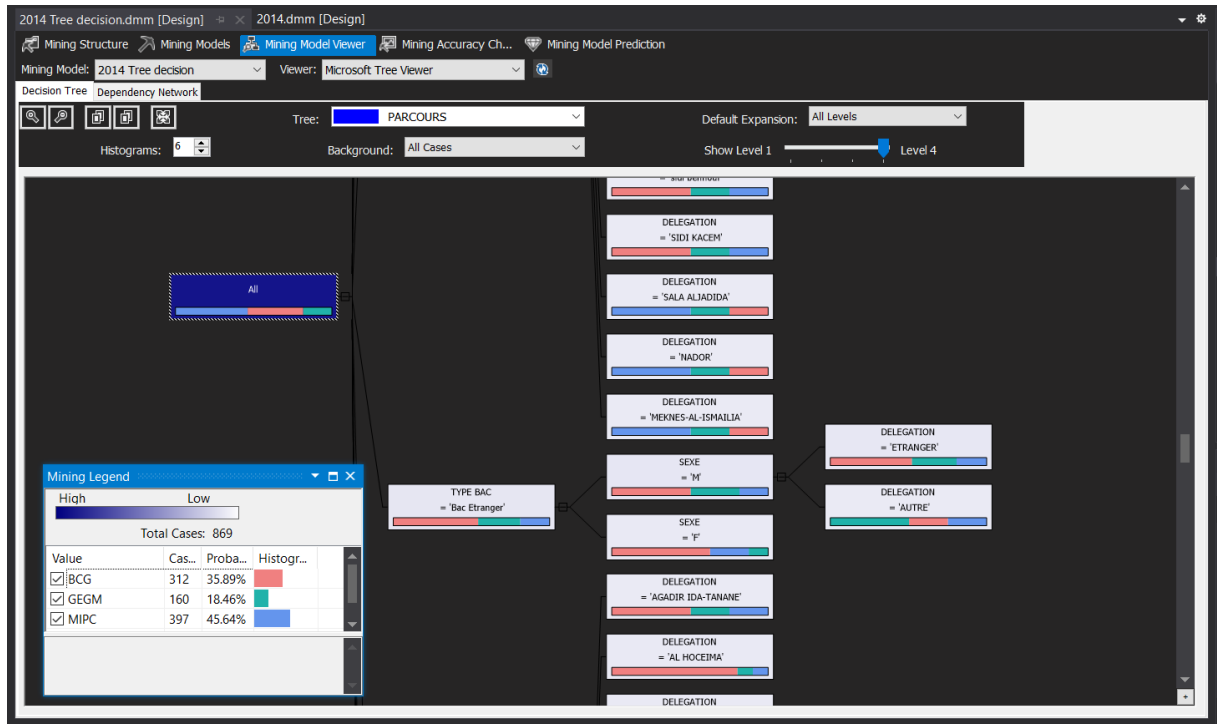
Le résultat de prédiction est le cluster 10, pour extraire la valeur de parcours prédire il faut prendre sa valeur avec la plus grande probabilité dans ce cluster, pour notre cas c'est le parcours MIPC.

The screenshot shows the 'Microsoft Cluster Viewer' window. The 'Cluster' dropdown is set to 'Cluster 10'. The 'Characteristics for Cluster 10' table lists variables, their values, and their probabilities. The row for 'PARCOURS' with value 'MIPC' is highlighted with a red box.

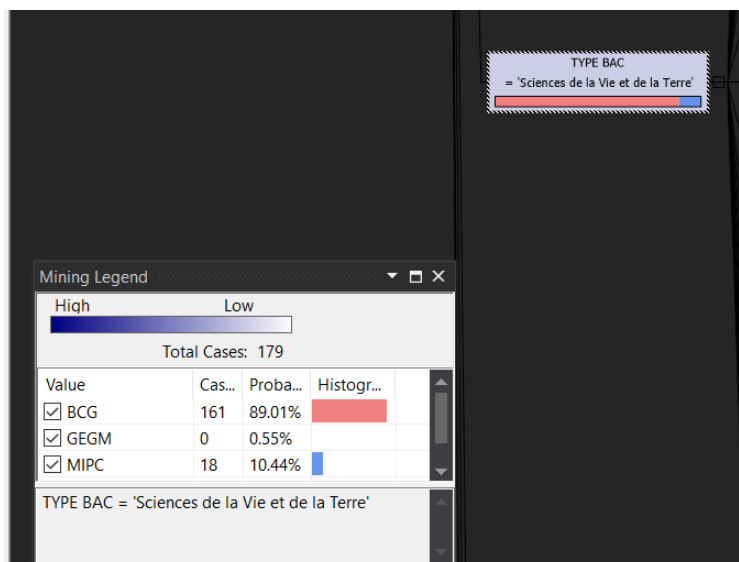
Variables	Values	Probability
SEXE	M	
PARCOURS	MIPC	
PARCOURS	BCG	
TYPE BAC	Sciences de la Vie et de la Te...	
DATE NAISSANCE	missing	
TYPE BAC	Physique Chimie	
SEXE	F	

2-1 Arbre de décisions :

Notre arbre de décision parcourt tous les attributs pour trouver le parcours le plus adéquat.



Par exemple pour le type de bac science de la vie et la terre, l'arbre a trouvé que le parcours BCG a la probabilité la plus élevée.



2-2 Prédiction :

La prédiction par les arbres de décision, se fait par donner les valeurs d'entrée : date de naissance, délégation, sexe et type de bac

The screenshot shows the 'Mining Model Prediction' window. On the left, the 'Mining Model' tree lists the model components: 2014 Tree decision, CODE, DATE NAISSANCE, DELEGATION, PARCOURS, SEXE, and TYPE BAC. The 'Singleton Query Input' table on the right contains the input values for the model. The bottom table shows the prediction results for the PARCOURS field.

Mining Model Column	Value
DATE NAISSANCE	10/06/97
DELEGATION	TETOUAN
PARCOURS	
SEXE	M
TYPE BAC	FILIERE SC. MATHEMATIQUES -B-

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
2014 Tree deci...	PARCOURS		<input checked="" type="checkbox"/>			
Prediction Fun...	PredictHistogram		<input checked="" type="checkbox"/>			[2014 Tree decision].[PARCOURS]

Le résultat de prédiction est la valeur de parcours avec la probabilité de chaque valeur possible

The screenshot shows the 'Mining Model Prediction' window with the 'PARCOURS' field selected. The results table displays the predicted value and its probability.

PARCOURS	\$SUPPORT	\$PROBABILI...	\$ADJUSTED...	\$VARIANCE	\$STDEV
GEGM	9	0.5882352...	0.1856810...	0	0
MIPC	5	0.3529411...	0.0111069...	0	0
BCG	0	0.0588235...	0.0047309...	0	0
	0	0	0	0	0

Conclusion :

A la fin de cet atelier, on peut dire que nous avons réussi à répondre à la majorité des questions concernant le Data Mining. On a abordé plusieurs technique et approches utilisées dans le datamining qui nous permettre de bien comprendre les données massives produits à partir le service de scolarité de la faculté des sciences et technique de Tanger. Le traitement de ces données aide à a extraire des informations utiles qui va nous aider a bien évalué les étudiants, et de prédire les performances des étudiants qui vont s'inscrire dans la faculté.