



Rapport de Stage Technicien

réalisé à :



Elaboré par :

Meriem ELLEUCH

Développement d'une Solution BUSINESS INTELLIGENCE

Encadrement :

Mr.Achref KAROUI

Encadreur industriel

Mr.Bachar ZOUARI

Encadreur Académique

Année Universitaire : 2021/2022

Remerciement

*Je tiens à remercier et à témoigner mes profonds respects et reconnaissances à Monsieur **Achref Karoui**, pour l'expérience enrichissante et pleine d'intérêt qu'il m'a fait vivre durant ce mois, pour ses conseils précieux et la qualité de l'encadrement dont il m'a fait bénéficier tout le long de ce projet.*

*J'adresse mes plus sincères remerciements à Monsieur **Bachar Zouari** qui a fait en sorte que ma formation soit fructueuse en me transmettant son savoir-faire.*

SOMMAIRE

REMERCIEMENTS.....	2
I) Présentation Générale de la société.....	4
1) Présentation de la société.....	4
2) Slogan.....	4
3) Fiche signalétique.....	5
4) Organigramme de la Société	5
II) Description du sujet.....	6
1) l'Informatique Decisionnelle.....	7
2) L'apport de la Business Intelligence	8
3) Les Notions de Base de BI	8
a) Le datawarehouse	9
b) SQL Server Integration Services (SSIS).....	12
c) SQL Server Analysis Services (SSAS).....	14
d) Base de données multidimensionnelle OLAP.....	15
e) Base de données relationnelle.....	15
f) Cube.....	16
g) Datamart.....	16
h) Datawarehouse.....	16
i) Dimension.....	17
j) ERP.....	17
k) ETL.....	17
l) Mesure.....	18
m) Modele en etoile.....	18
n) Multidimensionnel.....	18
o) OLAP.....	18
p) Reporting.....	19
q) Saas (Software as a Service):PourPowerBI.....	19
r) Table de faits.....	19
III) REALISATION DU PROJET: Mes Missions.....	20
VI) CONCLUSION.....	28
V) BIBLIOGRAPHIE ET WEBOGRAPHIE.....	29

I. Présentation Générale de la société :

1. Présentation de la société :

ACTEOL, Une agence CRM (gestion de la relation client, **C**ustomer **R**elationship **M**anagement en Anglais) primée, spécialisée dans le regroupement de données complexes et disparates pour créer une vue client unique puissante qui peut générer des rendements significatifs sur les résultats.

Le résultat final signifie que chaque client a une meilleure connaissance et une meilleure compréhension de ses clients, ce qui lui permet de prendre de meilleures décisions en matière de marketing et de vente.

2.Slogan :



Figure 2: Slogan Société ACTEOL

3. Fiche signalétique :

Adresse : Route l'Afrane km7, Sfax 3093

Horaires : 8h30 – 18h30

Téléphone : +216 74 616 185

E-mail : info@acteol.com

Site web : <https://www.acteol-crm.com/>

Fb : <https://www.facebook.com/Acteol/>

LinkedIn : <https://www.linkedin.com/company/acteol/>

4. Organigramme de la société :

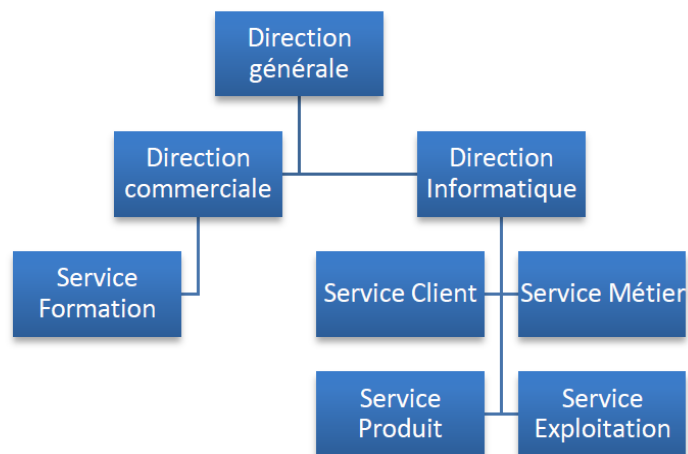


Figure 3 : organigramme de la société acteol

Acteol comprend 25 salariés

J'ai réalisé mon stage au sein du service Produit qui est en charge de la réalisation des différentes solutions commercialisées par Acteol.

II.Description du sujet :

1.Business Intelligence:

« L'informatique décisionnelle (ou BI pour Business Intelligence) désigne les moyens, les outils et les 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 aux responsables de la stratégie d'entreprise d'avoir une vue d'ensemble de l'activité traitée. »

Voici la définition que l'on retrouve généralement lorsque l'on parle d'informatique décisionnelle. Une entreprise est généralement composée de plusieurs services tels que les ressources humaines, les services comptabilité, marketing, commercial, technique... Tous conservent des informations propres à leurs fonctions : listes des clients, des employés, chiffres, emplois du temps...

L'accumulation de ces données nécessite donc leur sauvegarde dans le but d'une future exploitation. On constate ainsi régulièrement que chaque service possède son tableau de bord, ce qui lui permet de mesurer les indicateurs de performance de l'entreprise (chiffre d'affaires, calculs de bénéfices à l'année...). Cependant, chaque service a bien souvent sa façon de stocker ses informations (par exemple dans un fichier Excel, une **base de données relationnelle**...), et sa manière de calculer les indicateurs, avec sa vérité et ses critères.

Ainsi, si l'on veut considérer les données de l'entreprise dans son ensemble, la tâche s'avère rude voire parfois impossible. Pourtant, cela constituerait une utilité évidente et un réel apport à la société. En effet, une mise en relation et une analyse de toutes les données permettraient de réaliser des études et des prévisions sur le comportement et la " santé " de l'entreprise.

L'informatique décisionnelle désigne les moyens permettant de rassembler, intégrer, analyser et partager des données de l'entreprise afin d'optimiser la prise de décision. Par extension, BI désigne les solutions logicielles combinant à des fins décisionnelles des fonctions d'interrogation de bases de données, de reporting, d'analyse

multidimensionnelle (ou OLAP), de data mining et de visualisation des données.

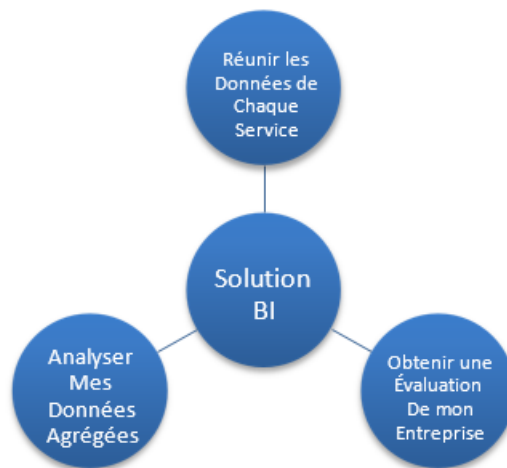


Figure 4 : la solution BI

2.L'apport de BI:

Le but de la BI est d'apporter une vision globale des données de l'entreprise, afin de répondre aux problématiques de celle-ci ou, tout simplement, afin de l'évaluer. Pour y arriver, SQL Server 2019 met donc à disposition trois plateformes qui illustrent ce cheminement.

- SQL Server Integration Services (SSIS), qui permet d'intégrer des données provenant de différentes sources pour les ranger dans un **entrepôt central (datawarehouse)**.

- SQL Server Analysis Services (SSAS), qui permet d'analyser les données, agrégées lors de SSIS, grâce à des fonctions d'analyse **multidimensionnelle**.
- SQL Server Reporting Services (SSRS), qui permet de créer, gérer et publier des rapports résultant des analyses réalisées lors de SSAS.

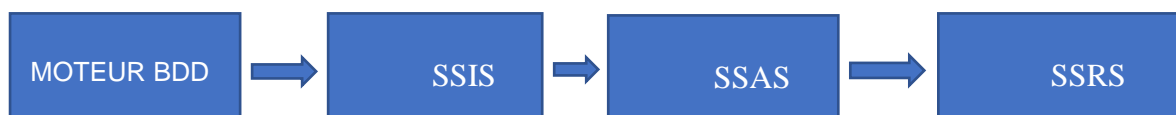


Figure 5 : Composants de SQL Server 2019

a) Le datawarehouse

La première étape d'un projet BI est de créer un entrepôt central pour avoir une vision globale des données de chaque service. Cet entrepôt porte le nom de datawarehouse. On peut également parler de **datamart**, si seulement une catégorie de services ou métiers est concernée. Par définition, un datamart peut être contenu dans un datawarehouse, ou il peut être seulement issu de celui-ci.

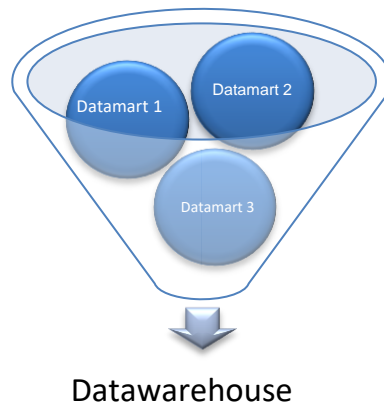


Figure 6 : Datawarehouse et Datamart

Un datawarehouse représente une base de données : les données étant intégrées, non volatiles, et historisées. Grâce à la plateforme SQL Server Integration Services (SSIS), cet entrepôt central sera rempli. Mais avant, il est indispensable de définir sa structure.

Avant de remplir le datawarehouse dans SSIS, la conception de celui-ci s'impose. Viennent donc les notions de table de faits et table de dimension :

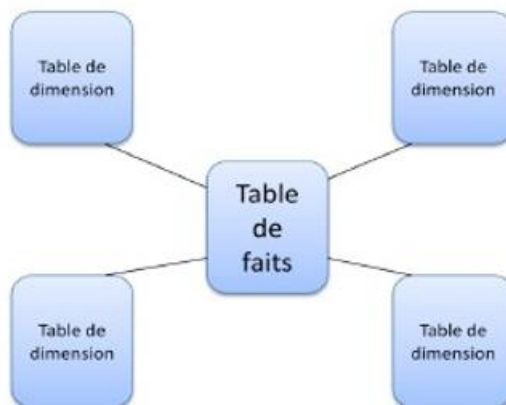


Figure 7 : Schéma en étoile

Une table de dimension contient des colonnes, chaque colonne correspondant à un attribut. Une dimension organise ainsi les données contenues dans la table en fonction d'un domaine d'intérêt. Exemples de dimension : le temps, le lieu, le type de produit...

Les colonnes d'une table de faits représentent les faits ou les mesures. Ses données sont généralement numériques, quantifiables et agrégeables. Exemples : un montant, une quantité de produits...

Les mesures sont en fait les critères ou indicateurs que l'on veut étudier en fonction de différents axes ou dimensions.

Une table de faits contient donc les mesures dont on a besoin, mais aussi les identifiants qui font référence aux tables de dimension situées tout autour de la table de faits. On parle alors de schéma en étoile, et parfois de schéma en flocon de neige quand le schéma s'étend encore plus.

Une fois les dimensions et les faits identifiés, on crée les tables directement sur SQL Server, en tant que simples tables. Pour les différencier, on peut spécifier dans le nom s'il s'agit d'une table de faits ou de dimension. Le schéma du datawarehouse est, en fait, une simple base de données avec nos tables de faits et de dimensions.

Cette partie de conception demande une connaissance pointue des besoins et attentes de l'entreprise concernée. Il faut tout prendre en compte et poser les bonnes questions : Arrivera-t-il que les dirigeants de l'entreprise changent d'avis sur la périodicité des rapports de leur activité (toutes les semaines au lieu de tous les mois par exemple) ?

Au lieu d'avoir un compte rendu statique, basé sur des critères fixes, le but ici est d'apporter une plus grande liberté qui reposerait sur les désirs et les besoins des dirigeants.

Après la conception du datawarehouse, vient son « remplissage » avec Integration Services.

c) SQL Server Integration Services (SSIS)

SQL Server Integration Services est donc la première plateforme à aborder. Une fois la structure du datawarehouse définie, les données doivent être insérées. L'outil qui va permettre le remplissage de notre base est l'**ETL** (Extract-Transform-Loading).

Comme son nom l'indique, il commence par extraire les données provenant de différentes sources (Excel, CVS, MySQL...), les transforme si besoin est, puis les charge dans le datawarehouse.

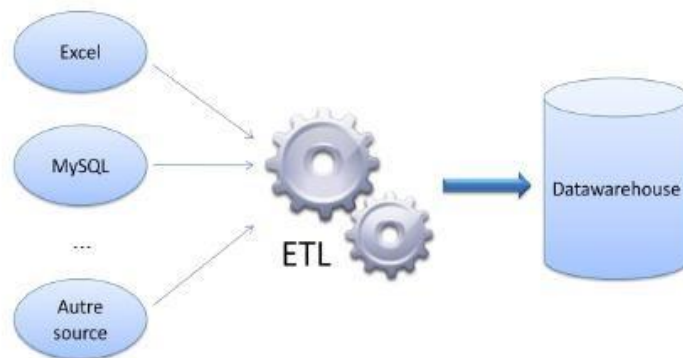


Figure 8: ETL

Ce processus se déroule donc en trois étapes :

- Extraction des données à partir d'une ou plusieurs sources de données.
- Transformation des données agrégées.
- Chargement des données dans la banque de données de destination (datawarehouse).

Dans SSIS, on appelle package l'environnement dans lequel on travaille. On peut construire plusieurs packages Integration Services.

L'organisation des flux était comme si dissous:

Chaque package contient son flux, sa propre connexion fichier plat et sa connexion OLE DB Destination

Prenons l'exemple de la table de dimension AgeBand:

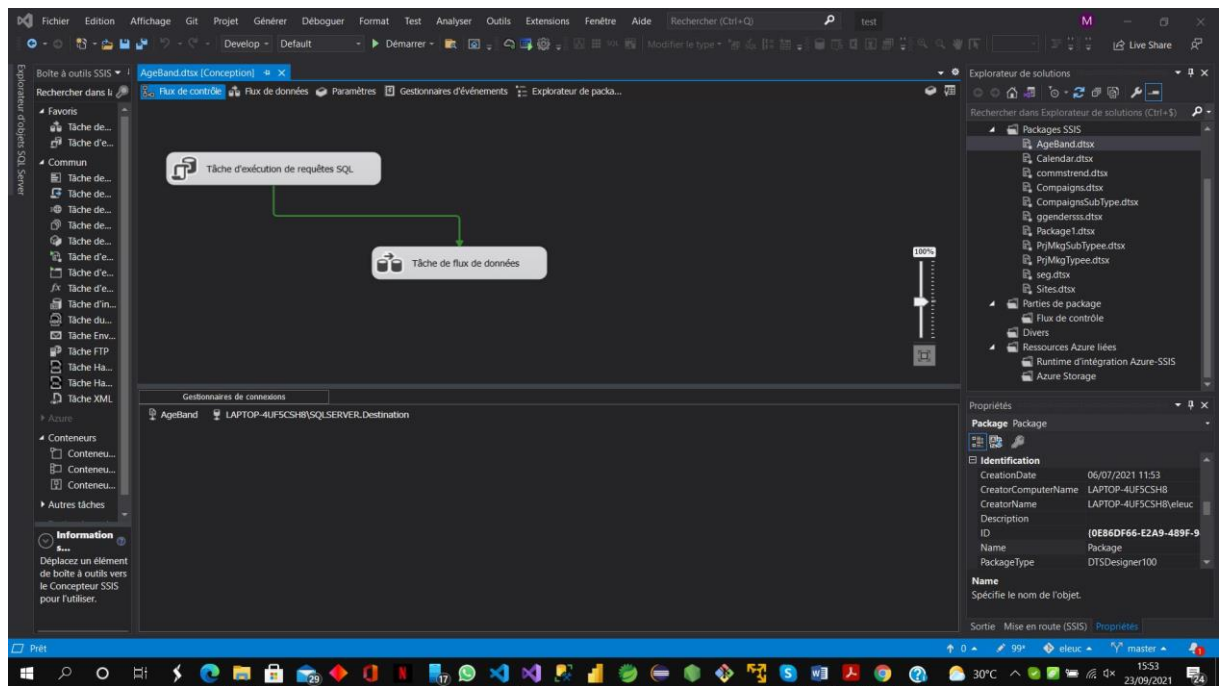


Figure 8 : Interface du package ID report dans SSIS

Chaque package contient toutes les tâches d'intégration. L'enchaînement des tâches d'un package est orchestré par le flux de contrôle. Lorsqu'une tâche a pour objectif d'assurer la transformation des données, elle est nommée « tâche de flux de données ». A l'intérieur de cette tâche se trouve un flux de données contenant au minimum une source, une transformation et une destination.

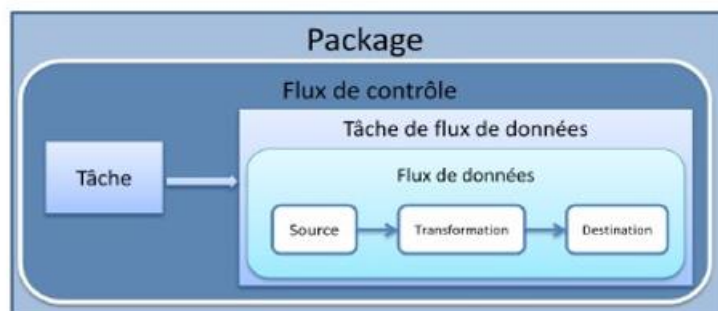


Figure 9: notion de Package

d) SQL Server Analysis Services (SSAS)

SQL Server Analysis Services est la plateforme qui permet de créer et gérer des structures multidimensionnelles. Pour cela, SSAS fournit des fonctions **OLAP** (On Line Analytical Processing) qui permettent, à partir des entrepôts de données, d'analyser l'activité de l'entreprise grâce à des statistiques : moyennes mobiles, coefficients de corrélation, valeurs cumulées...

L'élément principal de l'infrastructure OLAP est le **cube**. C'est en fait une **base de données multidimensionnelle**, qui permet l'analyse de ces données. Un cube reprend les mesures de la table de faits que l'on a pu établir lors de la conception du datawarehouse, et s'en sert pour effectuer des calculs, les mesures étant des données quantitatives.

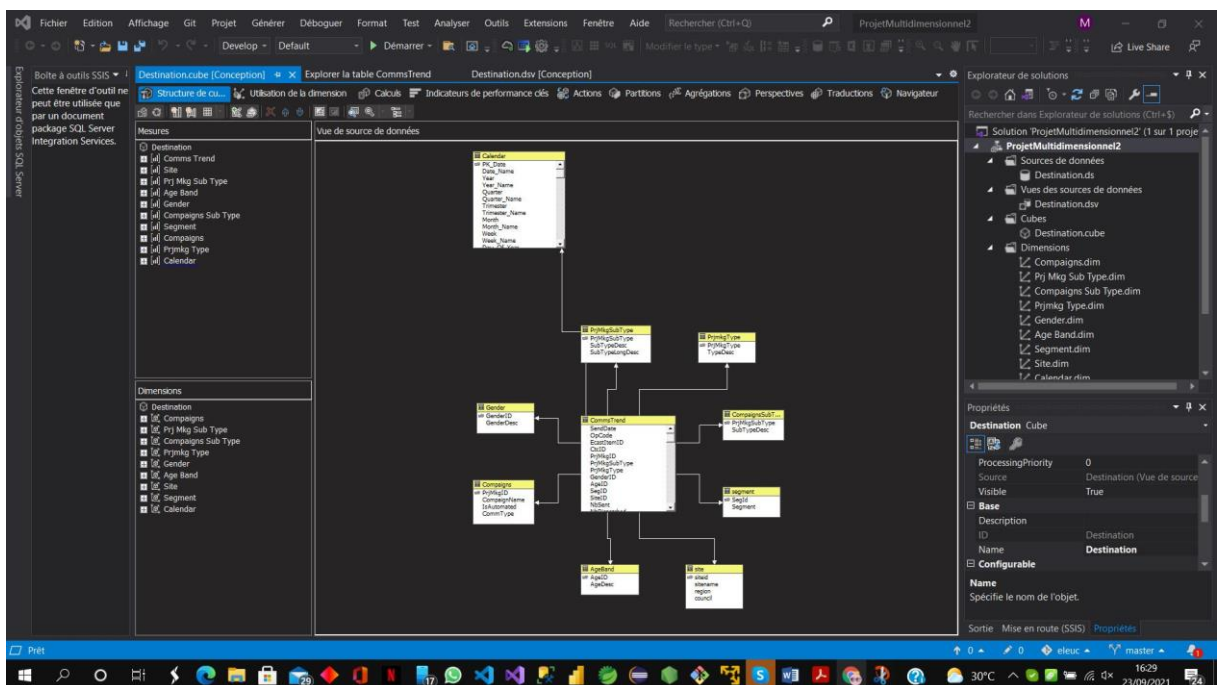


Figure 10 : Diagramme des tables

Base de données multidimensionnelle OLAP : Modèle de base de données traitant les données non comme des tables et des colonnes relationnelles, mais en tant que cubes d'informations dont les cellules comportent des données de synthèse et de dimension. Chaque cellule est fonction d'un ensemble de coordonnées qui précisent sa position dans les dimensions de la structure.

Base de données relationnelle : Ensemble d'informations organisées sous forme de lignes et de colonnes dans des tables. Chaque table détermine une classe d'objets pour l'organisation concernée. Les requêtes peuvent exploiter les données d'une table pour rechercher des données associées dans d'autres tables. Les liens entre les tables sont établis à l'aide de jointures entre les champs clés.

Cube : Ensemble des données organisées et synthétisées dans une structure multidimensionnelle définie par un ensemble de dimensions et de mesures.

Datamart : Sous ensemble d'un datawarehouse lié à un métier de l'entreprise (finance, marketing, RH, etc...) et conçu pour répondre aux besoins d'un groupe spécifique d'utilisateurs en respectant les exigences de sécurité de l'entreprise. L'entreprise peut construire des datamarts « Ventes », « Finances » en ayant l'assurance que les utilisateurs n'ont accès qu'aux données qui les concernent.

Datawarehouse : Entrepôt de données, isolé des systèmes opérationnels, permettant d'agréger des données thématiques, intégrées, non volatiles et historiées, dans un but de faciliter la prise de décision.

Dimension : Attribut structurel d'un cube constituant une hiérarchie organisée de catégories qui décrivent les données d'une table de faits. Ces catégories décrivent généralement un ensemble identique de membres sur lesquels les utilisateurs souhaitent fonder une analyse. Par exemple, une dimension géographique peut inclure des niveaux Pays, Région, Département et Ville.

ERP (Enterprise Resource Planning) : Progiciel de gestion intégré. L'ERP regroupe tout ou partie des applications nécessaires à la gestion de l'entreprise. Que ce soit des applications horizontales (comptabilité, paie, facturation) ou verticales (gestion de production, gestion des stocks par secteur d'activité).

ETL (Extract, Transform, Load) : Outils destinés à l'extraction, à la transformation et au chargement des données dans un datawarehouse. Il permet de communiquer des données de systèmes d'informations différents ce qui représente bien les besoins actuels en terme de système d'informations, de transfert de données et permet de garder une cohérence. L'objectif principal est d'importer les données d'un fichier plat généré par l'application de l'entreprise, dans la base de données.

Mesure : Dans un cube, ensemble des valeurs, généralement numériques basées sur une colonne dans la table de faits du cube. Les mesures sont des valeurs centrales qui sont agrégées et analysées.

Modèle en étoile : Arrangement de tables dans une base de données relationnelle. Au centre, on trouve la table de faits ; les branches de l'étoile qui rayonnent à partir de la table de faits correspondent aux dimensions.

Multidimensionnel : Structure de données ayant au moins trois dimensions indépendantes.

OLAP (On Line Analytical Processing) : Technologie utilisant des structures multidimensionnelles pour offrir un accès rapide aux données en vue d'une analyse. Les données sources OLAP sont souvent stockées dans les magasins de données d'une base de données relationnelle.

Reporting : Un compte rendu souvent appelé en anglais reporting est l'opération consistant, pour une entreprise, à faire un rapport de son activité. Il s'agit d'un outil décisionnel.

SaaS (Software as a Service) : SaaS est un concept consistant à proposer un abonnement d'un logiciel plutôt que l'achat d'une licence.

Table de faits : Table centrale dans un schéma de magasin de données composées de mesures numériques et de clés associant des faits à des tables de dimension. Les tables de faits renferment des données qui décrivent des événements inhérents à une activité commerciale, tels que des transactions bancaires ou des ventes de produits.

III.Mes Missions:

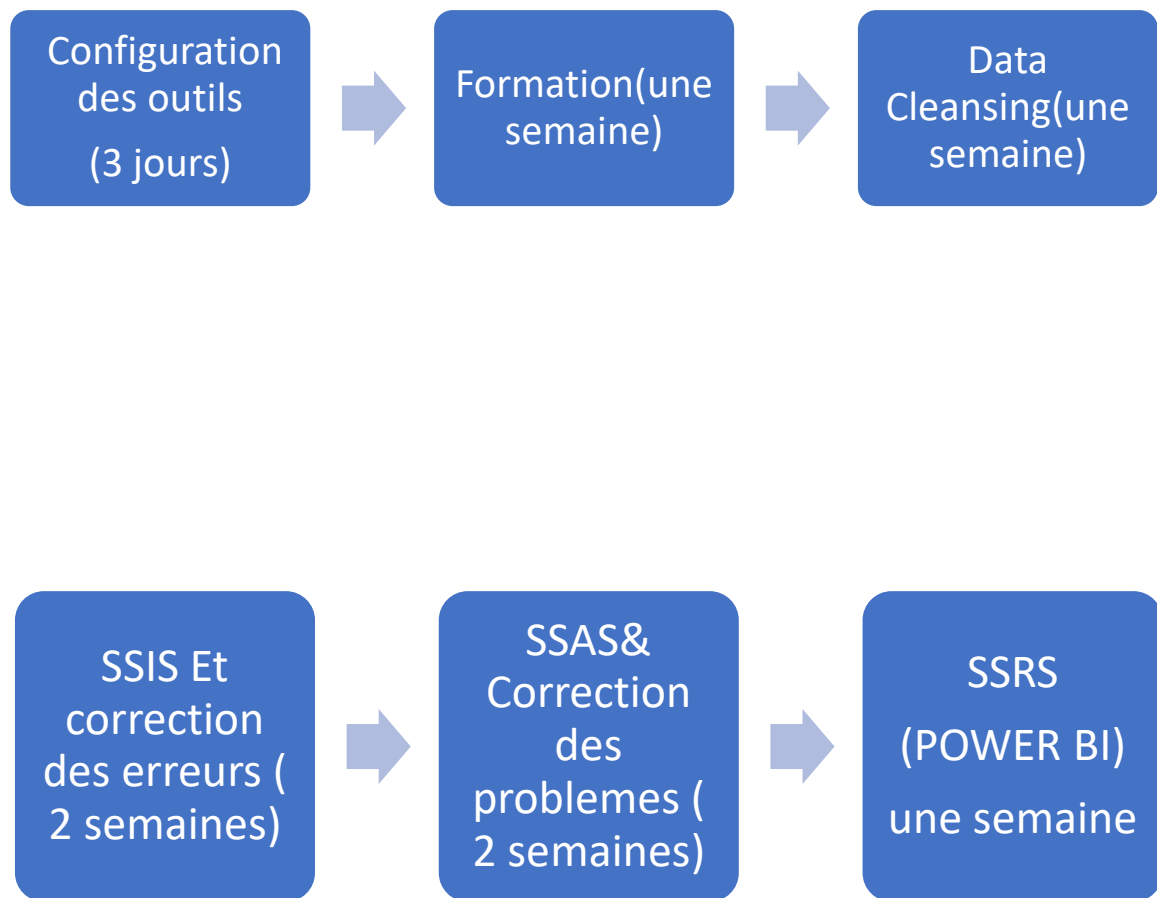


Figure 11 : Planning du déroulement des missions

Lors des 2 premières semaines de formation il m'a fallu acquérir les notions du BI et du reporting tout en apprenant à en utiliser les outils. Ce fut la partie du stage la plus difficile étant donné la complexité des notions.

Configuration:

Quelque captures lors de la phase de configuration de ssms et visual Studio:

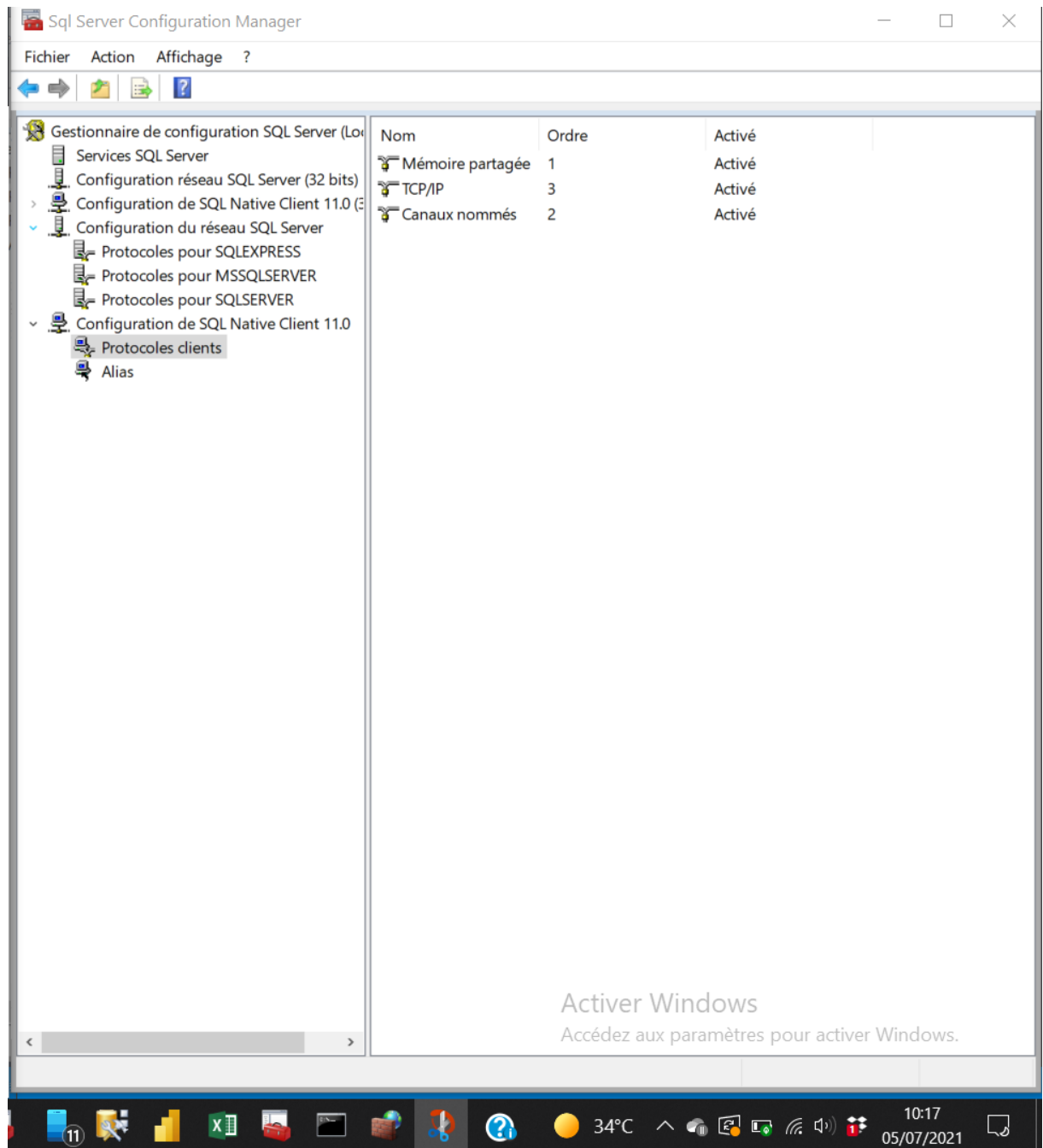


Figure 11 : Configuration avec SSCM

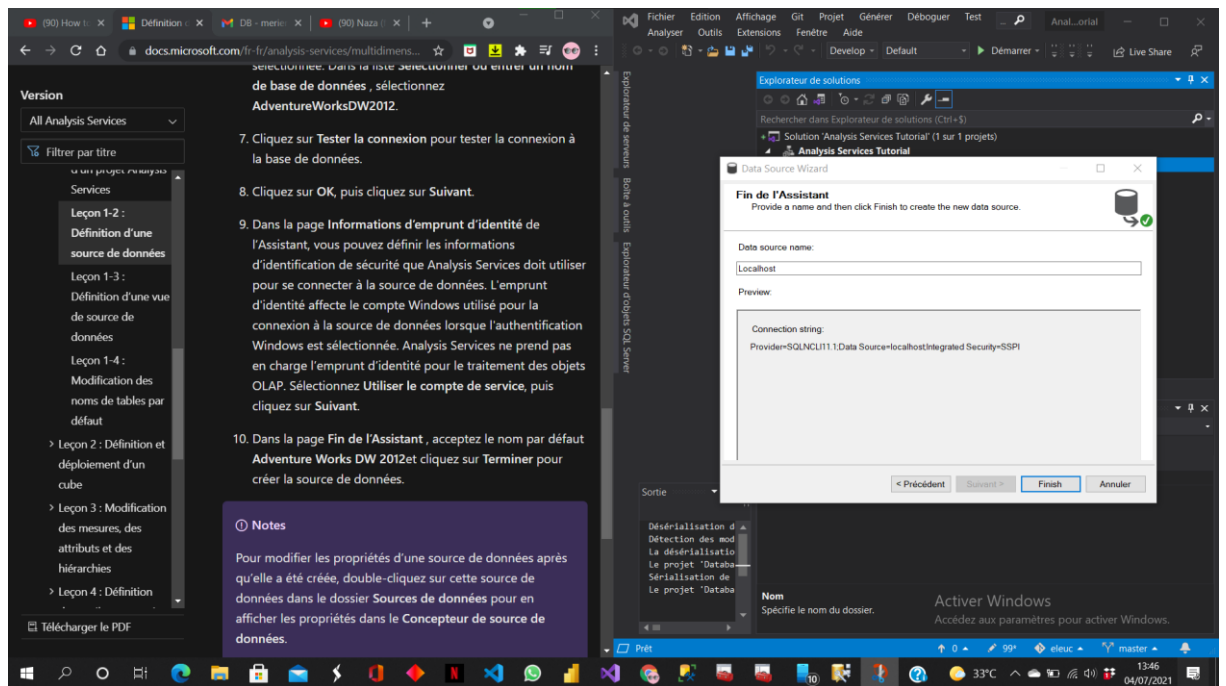


Figure 12 : Configuration au niveau de Visual Studio

Data Cleansing : prenons l'exemple de la table CommsTrend :

Avant:

Figure 13 : fichier CommsTrend.CSV avant Data Cleansing

Après:

SendDate	OpCode	EcastItemID	CtlcID	PrjMkgID	PrjMkgSubTyp	PrjMkgType	GenderID	AgeID	SegID	SiteID	NbSent	NbDispatched	NbDeliv	ErrorCode	LastError	NbTrial	NbViews
30/01/2020	93720	18470	2757555	11402	3	1	1	3	7	97	1	1	1				1
30/01/2020	93720	16664	2664530	11402	3	1	2	2	6	104	1	1	1				1
01/02/2020	93655	6880	24098	11382	2	1	2	5	6	3	1	1	1				1
01/02/2020	93655	11157	5315679	11382	2	1	2	2	6	169	1	1	1				1
01/02/2020	93655	6983	41415	11382	2	1	1	1	6	9	1	1	1				1
01/02/2020	93655	7288	88940	11382	2	1	1	7	3	3	1	1	1				1
02/02/2020	93740	1676	8101379	11406	5	1	2	3	4	220	1	1	1				1
02/02/2020	93740	1699	8133453	11406	5	1	1	2	3	220	1	1	1				1
02/02/2020	93740	1705	8146066	11406	5	1	1	3	3	220	1	1	1				1
02/02/2020	93740	1724	8166547	11406	5	1	2	4	3	220	1	1	1				1
02/02/2020	93740	1731	8182145	11406	5	1	2	2	3	220	1	1	1				1
01/02/2020	93655	11443	5619565	11382	2	1	2	4	6	169	1	1	1				1
01/02/2020	93655	7298	91232	11382	2	1	1	5	3	3	1	1	1				1
01/02/2020	93655	7304	92476	11382	2	1	2	3	3	57	1	1	1				1
01/02/2020	93655	7314	94914	11382	2	1	1	4	6	106	1	1	1				1
01/02/2020	93655	7320	96341	11382	2	1	1	2	6	3	1	1	1				1
02/02/2020	93655	7340	100613	11382	2	1	2	2	5	9	1	1	1		INT002: Domi	3	1
01/02/2020	93655	7365	103954	11382	2	1	1	1	4	3	1	1	1				1
01/02/2020	93655	7372	104307	11382	2	1	2	4	3	3	1	1	1				1
29/01/2020	93680	25573	2454203	11393	2	1	2	6	5	79	1	1	1				1
29/01/2020	93680	29217	4465748	11393	2	1	2	2	7	51	1	1	1				1
29/01/2020	93680	19224	1040850	11393	2	1	2	3	6	51	1	1	1				1
29/01/2020	93680	24047	2127405	11393	2	1	2	4	6	79	1	1	1				1
29/01/2020	93675	2103	2269080	11392	2	1	2	5	6	81	1	1	1				1
29/01/2020	93680	25785	2550438	11393	2	1	2	3	7	91	1	1	1				1
29/01/2020	93680	24442	2181539	11393	2	1	1	3	6	79	1	1	1				1
30/01/2020	93715	138875	4056115	11401	3	1	2	2	7	18	1	1	1				1
30/01/2020	93715	138888	4027001	11401	3	1	1	4	6	18	1	1	1				1
30/01/2020	93715	138907	4027759	11401	3	1	1	4	5	18	1	1	1		INT002: Domi	4	1

Figure 14: l'après

SSIS:

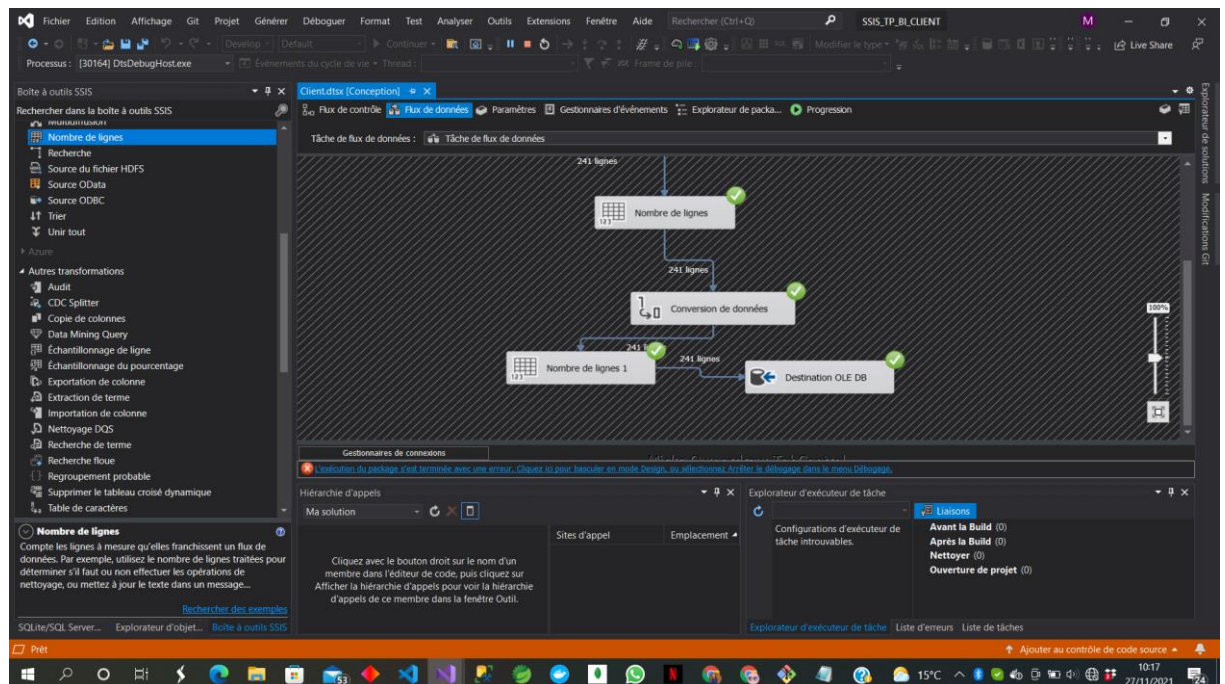


Figure 15 : package Client dans SSIS

SSAS:

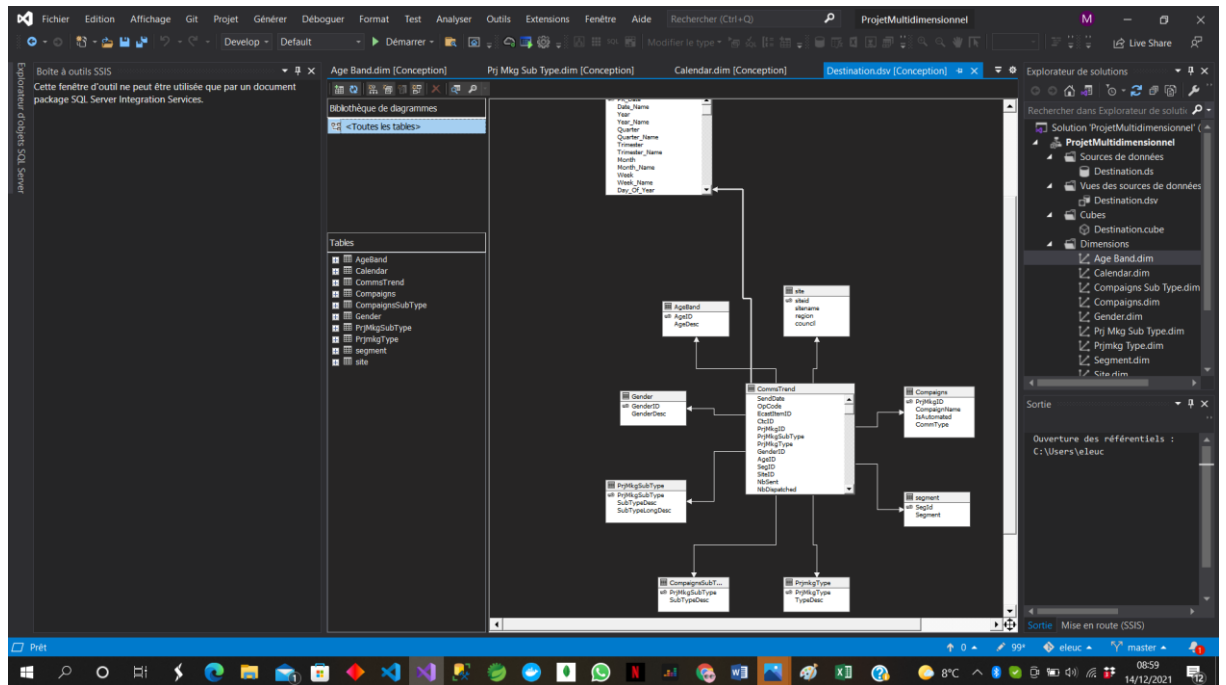


Figure 16: Modele de notre Solution sans SSAS

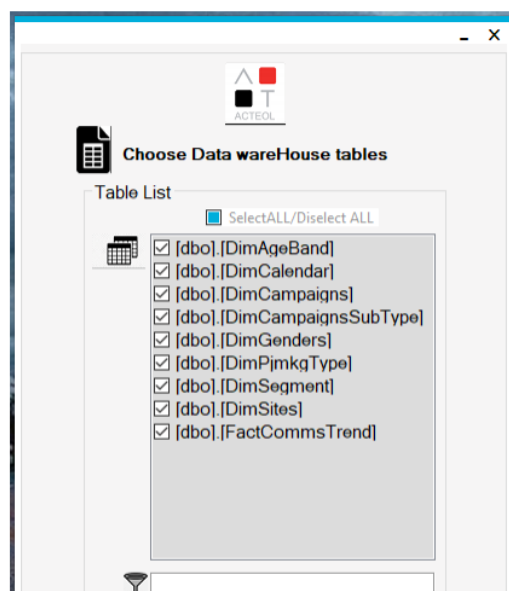


Figure 17: Choisir les tables pour general le cube

Dimension	Hierarchy	Operator	Filter Expression	Pa
<Select dimension>				
AgeDesc	CampaignName	GenderDesc	NbDeliv	NbContacts
16-24	Blackbrook Lei...	Unknown	3	0
16-24	Dolphin Weeke...	Female	298	0
16-24	Dolphin Weeke...	Male	289	0
16-24	Dolphin Weeke...	Unknown	15	0
16-24	East Feb 20 G...	Female	607	0
16-24	East Feb 20 G...	Male	585	0
16-24	East Feb 20 G...	Unknown	25	0
16-24	Middlesbrough...	Female	356	2
16-24	Middlesbrough...	Male	372	3
16-24	Middlesbrough...	Unknown	9	0
16-24	Stradbroke Po...	Female	31	23
16-24	Stradbroke Po...	Male	42	32
16-24	Stradbroke Po...	Unknown	3	1
16-24	Stradbroke Po...	Female	31	0
16-24	Stradbroke Po...	Male	42	0
16-24	Stradbroke Po...	Unknown	3	0
16-24	Swim Lesson	Female	1226	0
16-24	Swim Lesson	Male	599	0

Figure 18: Cube OLAP

Reporting:

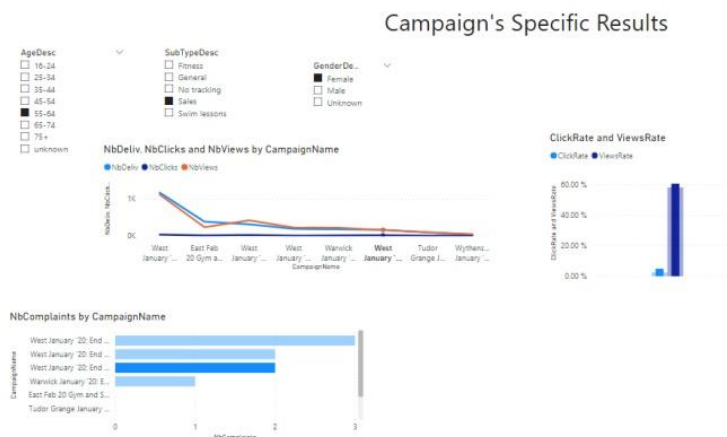


Figure 19 : tableau de bord des resultats Specifiques des offres des ventes

Campaign's General Results

CampaignName	Region	NbDeliv	UniqueClicks	UniqueViews	Date_Name
Wythenshawe January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	534.00	30.00	261.00	Wednesday, January 29 2020
West January '20: End Of Month NIF Fitness Sales Email 290120	SE	1.00	1.00	1.00	Wednesday, January 29 2020
West January '20: End Of Month NIF Fitness Sales Email 290120	WEST	1,870.00	39.00	916.00	Wednesday, January 29 2020
West January '20: End Of Month NIF + 20% Off Annual Fitness Sales Email 290120	SE	1.00	1.00	1.00	Wednesday, January 29 2020
West January '20: End Of Month NIF + 20% Off Annual Fitness Sales Email 290120	WEST	197.00	4.00	95.00	Thursday, January 30 2020
West January '20: End Of Month NIF + 20% Off Annual Fitness Sales Email 290120	WEST	4,058.00	167.00	2,533.00	Wednesday, January 29 2020
West January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	SE	1.00	0.00	0.00	Wednesday, January 29 2020
West January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	142.00	1.00	65.00	Friday, January 31 2020
West January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	380.00	7.00	149.00	Thursday, January 30 2020
West January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	15,733.00	334.00	8,225.00	Wednesday, January 29 2020
West January '20: End Of Month NIF + 140 Off Fitness Sales Email 290120	WEST	1,808.00	41.00	1,102.00	Wednesday, January 29 2020
Warwick January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	34.00	1.00	22.00	Thursday, January 30 2020
Warwick January '20: End Of Month NIF + 1MONF Fitness Sales Email 290120	WEST	2,472.00	74.00	1,514.00	Wednesday, January 29 2020
Tudor Grange January 2020 Swimming Newsletter Service 290120	EAST	3.00	0.00	3.00	Wednesday, January 29 2020
Tudor Grange January 2020 Swimming Newsletter Service 290120	WEST	0.00	0.00	0.00	Thursday, January 30 2020
Total		134,793.00	2,248.00	38,628.00	

95.66 %

DelivRate

28.66 %

ViewsRate

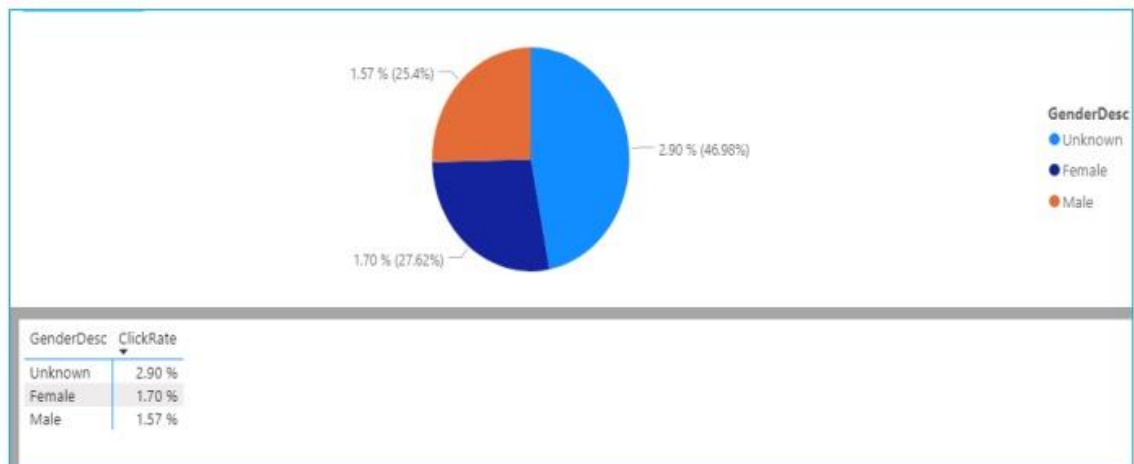
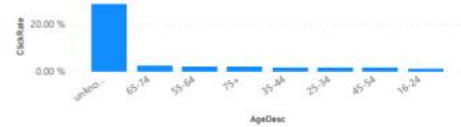
1.67 %

ClickRate

ClickRate by GenderDesc



ClickRate by AgeDesc



Dans notre cas on a :

Le total du pourcentage des hommes qui ont fait un click est **1.57%**.

Le total du pourcentage des femmes qui ont fait un click est **1.70%**.

Le total du pourcentage du genre inconnu qui a fait un click est **1.70%**.

Figure 20 : charte graphique de classement par genre

Conclusion

Ces 7 semaines de stage m'ont permis de me placer dans un contexte professionnel et de travailler en équipe sur un projet d'une grande envergure.

Mon stage s'est avéré particulièrement formateur du point de vue technique. J'ai renforcé mes bases dans le langage SQLet, surtout, j'ai découvert le monde du Business Intelligence J'y ai trouvé un grand intérêt car ce sont des connaissances qui ne nous sont pas toutes enseignées à l'université (en 2eme année Genie Informatique)et qui viennent compléter ma formation.

Sur le plan personnel, j'ai compris que les petites structures se révèlent très dynamiques, conviviales et travailleuses. En effet, j'ai pu constater que les salariés sont très impliqués dans la vie de l'entreprise et du fait qu'ils se connaissent bien, la communication au sein de la société est réellement avantagée.

Bibliographie et Webographie

[1] BERTRAND Burquier. « Business intelligence avec SQL server 2008 », Paris, Dunod, 2009.

[2]<http://www710.univ-lyon1.fr/~elghazel/BI/presentation.html>

[3] <http://www.dotnet-france.com/Documents/SQLServer/BI/Introduction%C3%A0%20la%20BI%20avec%20SQL%20Server%202008.pdf>