

**Rapport Etude Churn**

« ABC Bank Multinational »

**Module** : Entrepôt de données

**Professeur** : Eliana RAAD LOMBARDI

**Réaliser par :**

* TAFER Abdelkrim
* MADANI Riad

1. Le scénario
2. Qui sont mes clients finaux ?

Les clients finaux sont le Directeur Commercial, Marketing, et l’équipe CRM.

1. Quelle est mon équipe de projet ?

L’équipe du projet c’est ce constitue d’un data analyste, d’un spécialiste Marketing, d’un commercial et le responsable CRM.

1. Pourquoi ce projet existe-t-il ?

L’objectif du projet et d’étudier le Churn des clients au niveau de la banque pour mettre en place des actions de retentions afin de fidéliser nos clients.

En effet, le phénomène de perte de clientèle (Churn ou **Attrition)** est généralement mesuré sous forme de taux. On parle alors de **taux d’attrition de la clientèle** (Customer churn rate en anglais). Il correspond au niveau de **perte de clients** par rapport à la clientèle globale sur une période donnée.

1. Quels sont les enjeux qu’il adresse ?

Le **taux d’attrition client** est un indicateur à surveiller régulièrement et de très près, tout comme le pourcentage de rétention. Ces ratios constituent de bons indices sur la rentabilité d’une société. En effet, augmenter le **niveau de fidélisation de sa clientèle** permet d’accroître son chiffre d’affaires.

Le cout d’attraction de nouveaux clients est beaucoup plus élevé que celui de la fidélisation.

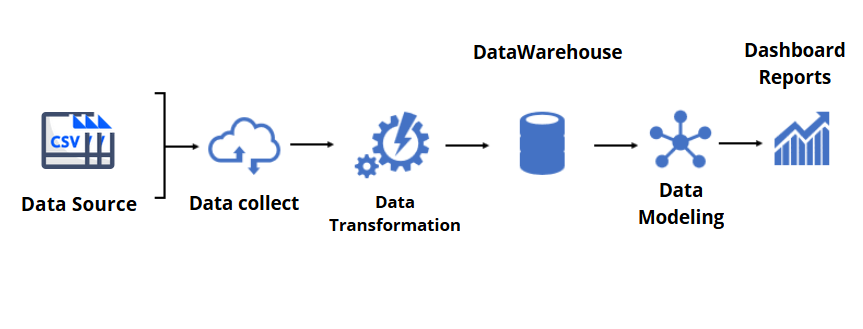
**Limiter le churn rate** à un faible niveau garantit une sécurité du chiffre d’affaires.

Il est donc particulièrement important de développer une approche qui permette d’anticiper à l’avance le départ des clients.

1. Quelles sont les fonctionnalités qui doivent être livrées en sortie de projet ?

À partir du Dashboard Power BI générés, l’utilisateur final (*Décideur*) doit pouvoir analyser les informations et en tirer des conclusions, à travers les visualisations suivantes :

1. Une vue globale sur le Churn a l’international de l’entreprise ;
2. Le profiling des clients fidèles et Churners ;
3. Une analyse du Churn par région.
4. La démarche (*les différentes phases du projet*)



* 1. Collecte des données :

La première étape consiste à collecter les données nécessaires à notre analyse, à travers l’entrepôt de donnée (DWH, service Cloud, …) de l’entreprise, après modélisation du besoin en données avec les différentes équipes lié aux métier (CRM, Marketing, …).

* 1. Traitement et Préparation de la donnée :

La préparation et le nettoyage de la donner est une étape cruciale qui vas s’impacter sur la qualité de notre Dashboard et sa véracité.

* 1. Modélisation :

Vue les différentes sources de données et leurs différentes natures, nous avons pu mettre en place un model en étoile nous permettant de faire une analyse multidimensionnelle.

* 1. Analyse de la donnée :

En étroite collaboration avec les différents corps métier lié aux projets, nous avons pu développer des mesures et des analyses, répondant à leurs besoins.

* 1. Visualisation :

Grace a l’outil Power BI nous avons mis en place un Dashboard répondant aux besoins de notre client interne.

1. Conception et Analyse :
2. Description des sources de données :

Contient les données client des titulaires de compte chez ABC Multinational Bank et le but des données sera d’analyser le taux de désabonnement des clients (Taux de Churn).

**TABLE DATE** {

**Date** : représente la date d’entrée du client

**Année** : représente l’année de la date

**Trimestre**: Le trimestre de l’année dans la date d’entrée du client

**Mois** : Le mois de la date d’entrée

**Jour**: représente le jour

**MoisAlpha** : correspond au nom alphabétique du mois

**YearMonth** : correspond au nom du mois + l’année

}

**Table GENDER** {

**Gender\_id** : une colonne qui correspond à un entier, 1 : masculin, 2 : féminin

**Gender\_category** : représente le type du sexe (Masculin, Féminin)

}

**Table EXIT\_CUSTOMER** {

**Exit\_id** : 0 pour client fidélisé, 1 pour churner

**Exit\_category** : représente le type de catégorie du client (Exit, Retain)

}

**Table ACTIVE\_CUSTOMER** {

**Active\_id** : clé primaire représente l’activité du client

**Active\_category** : correspond au fait que le client est un membre actif « 1 » ou non actif « 0 »

}

**Table CREDIT\_CARD** {

**credit\_id** : reprenste l’id de la catégorie du client (Porteur de carte de crédit ou non)

**credit\_category** : correspond a la catégorie du client (1 : porteur de carte de crédit, 0 non porteur)

}

**Table CUSTOMER\_INFO** {

**Customer\_id** : contient des valeurs aléatoires qui incrémente a chaque nouvel ajout d’un client

**Surname**: c’est le nom du client

**Age** : l’age du client

**Tenure** : fait référence au nombre d'années que le client a été client de la banque.

**Balance** : correspond au solde des comptes de clients

**Estimated\_salary** : Le salaire estimé du client

**Credit\_score** : représente le score du client au sein de la banque

}

**Table GEOGRAPHI** {

**Geo\_id** : fait référence a une région

**Location** : chaine de caractère qui correspond aux régions que le client fait partie

}

**Table BANK\_CHURN** {

**Id**: fait référence aux id des clients

**Date** : date à laquelle le client s’est associé ou à rejoindre la banque

**Product\_id** : fait référence au nombre de produits qu’un client a achetés auprès de la banque

**Exit\_id** : fait référence à l’Id du client qui a quitté ou non la banque

**Active\_id** : l’activité du client (le fait qu’il est actif ou non)

**Credit\_id** : si le client a une carte de crédit ou non

**Customer\_id** : contient des valeurs aléatoires qui incrémente à chaque nouvel ajout d’un client

**Geo\_id** : fait référence a la table Geography

**Mesures**

**Category\_age** : la tranche d’âge des clients de la banque

**Score\_type** : les différentes catégories qui correspondent aux intervalles des scores (Excellent, Très Bien, Bien, Equitable, Médiocre)

**Taux\_attrition :** le taux de churn qui correspond au nombre des clients perdus divisé par le nombre total des clients de la banque

**Taux\_de\_fidelisation** : le taux des clients qui sont encore fidèles à la banque

**Total\_client** : le nombre total des clients

**Total\_balance** : la somme des scores des clients

**Total\_tenure** : nombre totale des années que les clients ont étés des clients a la banque

**Total\_salaire** : la somme de tous les salaires estimés des clients

**Balance\_moyenne** : la somme de tous les scores divisés sur le nombre totale des clients

**Tenure\_moyen** : la moyenne du nombre d’années que le client est associé à la banque

**Salaire\_moyen** : le salaire moyen des clients

}

**Les contraintes :** la liaison entre les différentes tables

BANK\_CHURN.date > DATE.date

BANK\_CHURN.product\_id > GENDER.gender\_id

BANK\_CHURN.exit\_id > EXIT\_CUSTOMER.exit\_id

BANK\_CHURN.active\_id > ACTIVE\_CUSTOMER.active\_id

BANK\_CHURN.credit\_id > CREDIT\_CARD.credit\_id

BANK\_CHURN.customer\_id > CUSTOMER\_INFO.customer\_id

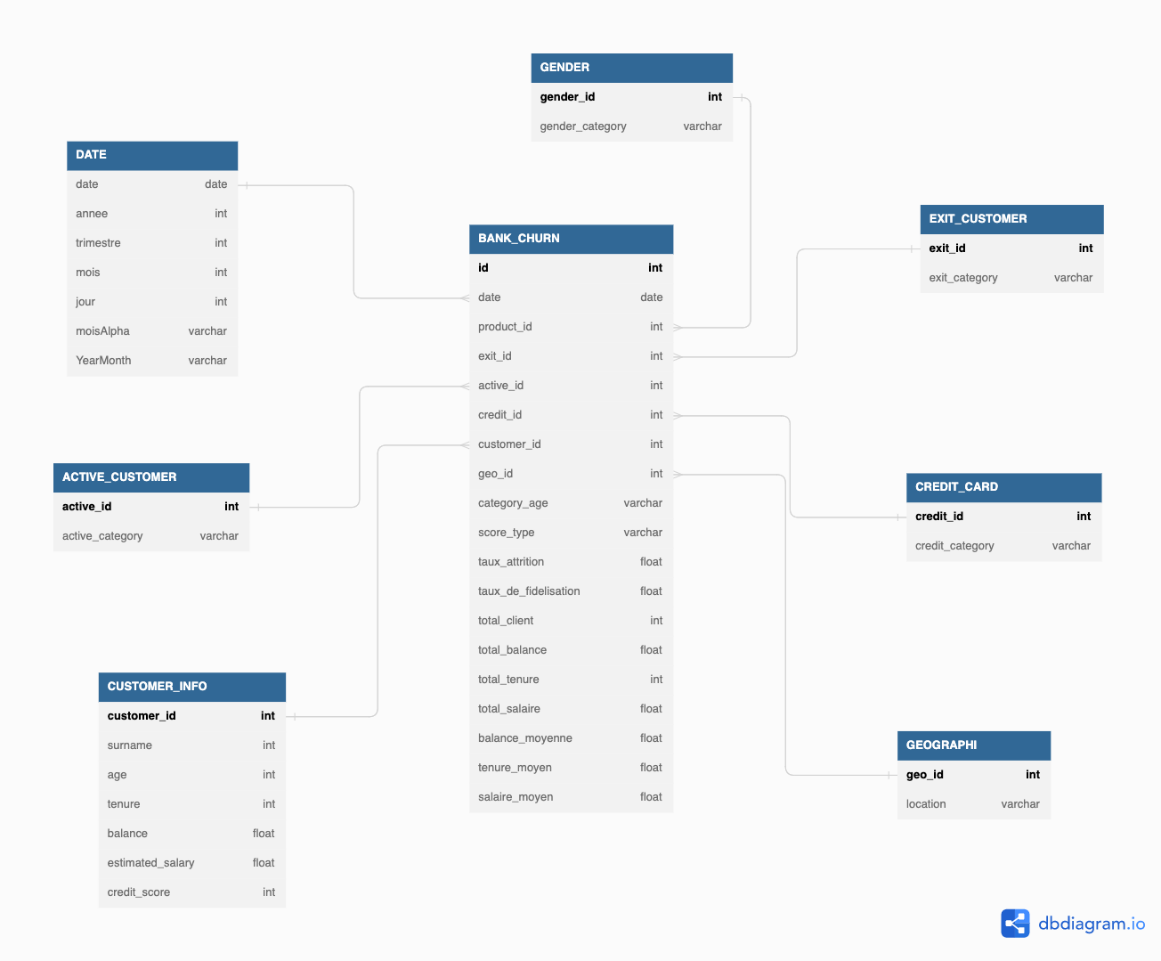
BANK\_CHURN.geo\_id > GEOGRAPHI.geo\_id

1. Les étapes de traitement/transformation de données

(A l’aide de **l’Editeur Power Query**), nous avons nettoyer et traiter le jeu de données commençant par :

* Modification des types de quelques colonnes
  + **CustomerID (int)** : en Chaine de caractère,
  + **Balance, EstimateSalary (Str)** : en Décimal avec deux nombres après la virgule
* Création de nouvelles colonnes
  + **Score Type** : chaine de caractère (Médiocre, Equitable, Bien, Très Bien, Excellent)
  + **Catégorie d’âge** : [15-20] Adolescent, [21-30] Jeune, [31-50] Adulte, [51-99] Aînés
  + **Année, Trimestre, Mois, Jour**
  + **MoisAlpha** : chaine de caractère qui fait référence aux noms des mois
  + **YearMonth**: chaine de caractère qui contient le nom du mois plus l’année
* Suppression des colonnes qu'on n'a pas besoin au cours de notre analyse (RowNumber).
* Suppression des valeurs nulles
* Changement des noms de colonnes

1. La modélisation décisionnelle (*l’analyse multi-dimensionnelle*)



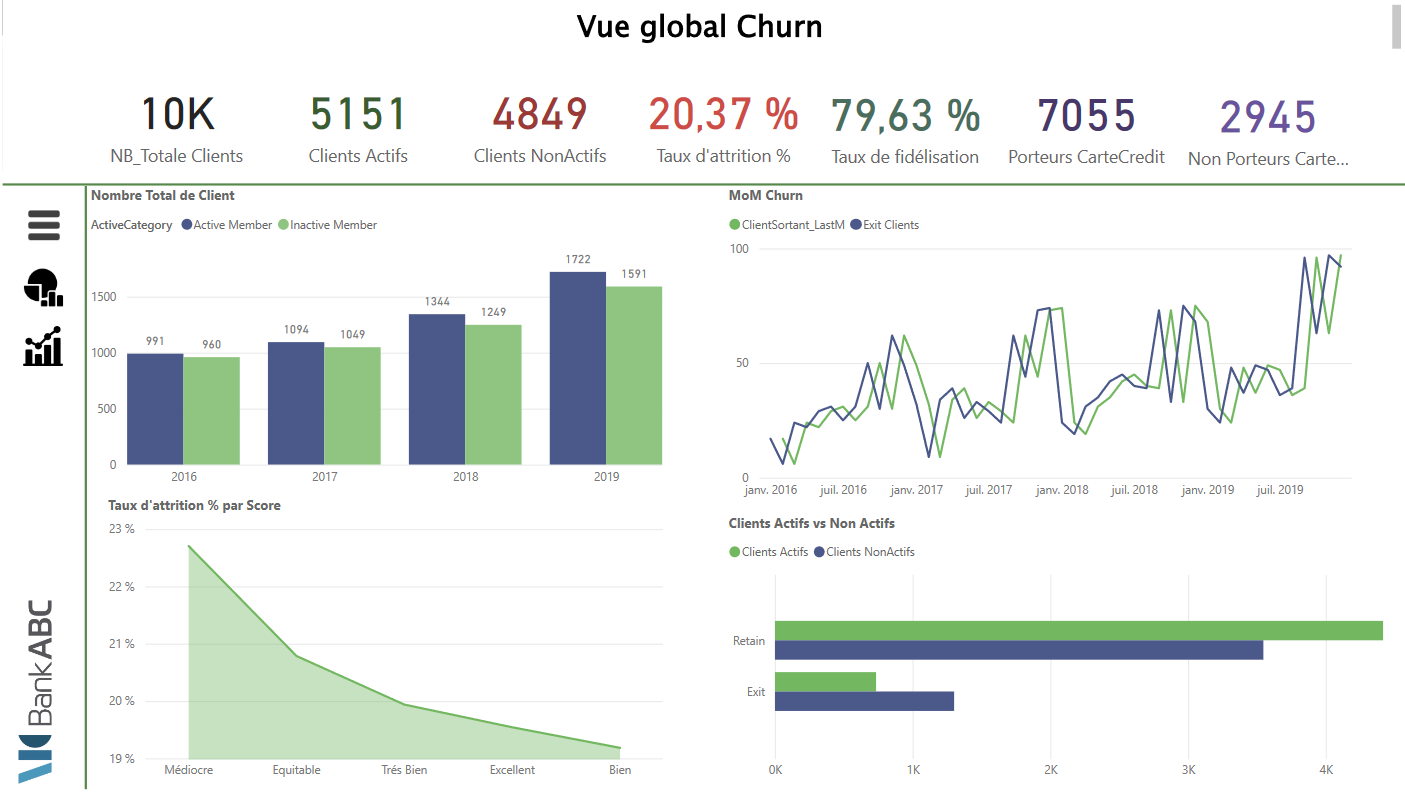
1. Les requêtes d’analyse (*les questions à résoudre*)

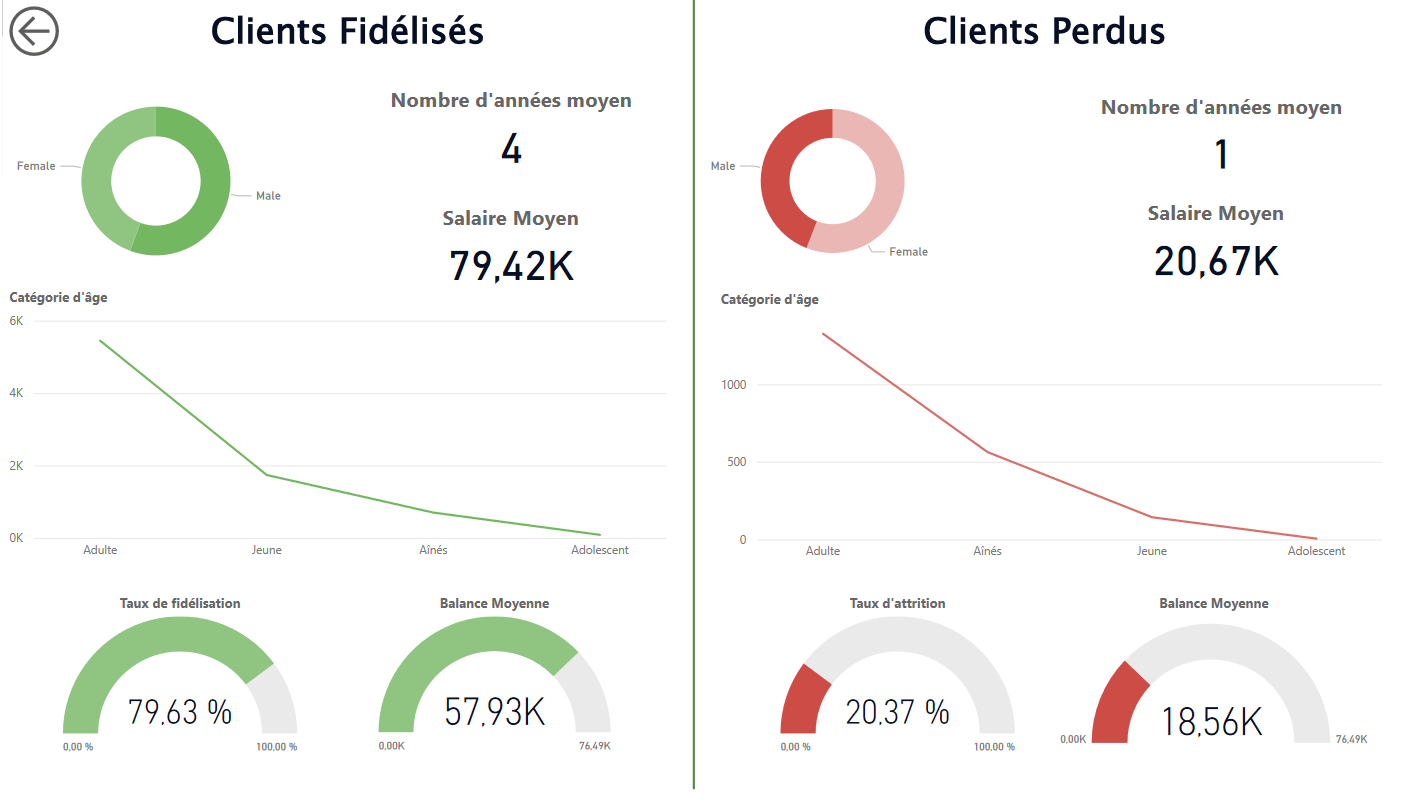
* Quel est le nombre total des clients par catégorie d’activité au cours des 2 dernières années
* Quel est le pourcentage de désabonnement des clients par région
* Quelles sont les catégories d’âge les plus susceptibles de quitter la banque.
* Quel est le taux d’attrition et le nombre des clients perdus durant une année précise
* Lesquelles entre les clients actifs et non actifs sont moins susceptibles de quitter la banque.
* Est-ce que l’emplacement d’un client peut influer sur sa décision de quitter la banque.
* Une comparaison entre le nombre des clients fidélisés & les clients perdus ainsi que les taux de fidélisation et d’attrition par leurs type (Actif ou non Actif) au cours des années

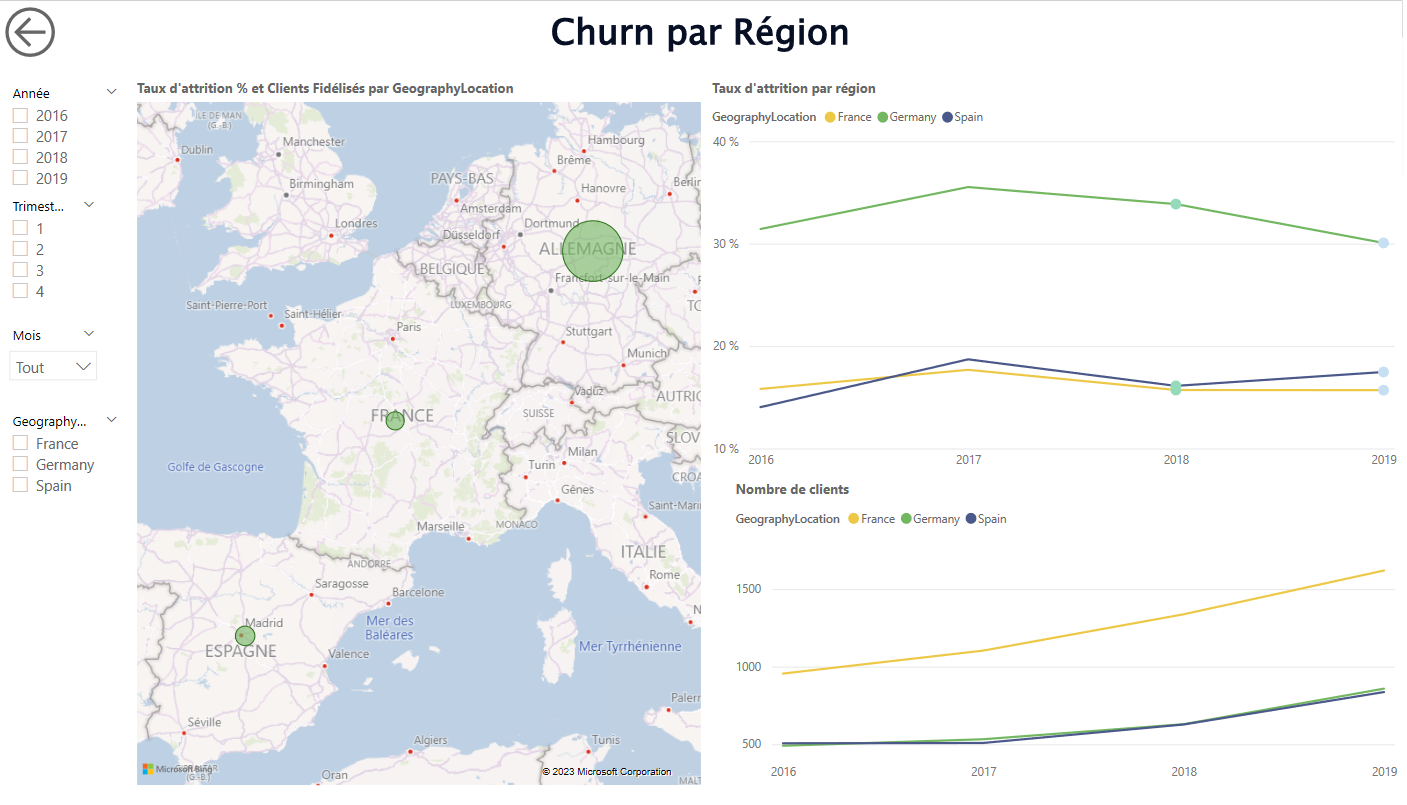
1. Le Dashboard
2. Justification du choix de l’outil (Power BI Desktop)

|  |  |  |
| --- | --- | --- |
| **Critère​** | **Description​** | **Pourquoi Power BI ?​** |
| Connectivité a la données​ | Capacité de se connecter à différentes sources et systèmes de données différentes localement ou en ligne (Cloud, Datawarehouse)​ | Il offre une connectivité à un large éventail de sources de données (tels que XML, JSON), Excel, BDD SQL/NoSQL, sources basées sur le cloud, services en ligne (Google  Analytics, Facebook, ...).​ |
| Modélisation de la données​ | Manipuler les données source, les transformer, les combiner et les modéliser à fin de les préparer​ | L’utilisation du Power Query (M) pour interroger les sources de données, ainsi que DAX pour les analyser dans Power Pivot. Ajouter à cela, la possibilité d'utiliser du Python ou R​ Il permet aussi à tirer des techniques de ML pour effectuer une analyse des sentiments, générer des prévisions, etc.​ |
| Visualisation​ | Utilise des visuels claires et concis, faciles à manipuler, à configurer, et respecte les normes de la data visualisation. Accessible sur n'importe quelle appareils (ordinateur, mobile …)​ | Il offre une large sélection de graphiques pour tout type de visualisation, accessible sur n'importe quel appareil.​  L'intégration de visuels est rendu possible soit via les différentes bibliothèques des langages Python ou R, soit en téléchargeant de nouvelles depuis le marketplace.​ |
| Serveur et Automatisation​ | Stocker ses rapports sur un serveur local ou cloud pour les rendre accessibles pour tous les utilisateurs, avec la capacité de mettre à jour automatiquement les visualisations en programmant un rafraîchissement des données.​ | Vous pouvez automatiser l’actualisation des données et publier vos rapports sur un serveur local ou un cloud, pour que tous les utilisateurs bénéficient des informations les plus récentes.​ |
| Documentation / Communauté​ | La disponibilité d'une documentation détaillée, de formations et de grande communauté en ligne ​ | Microsoft dispose d’un portail de documentation Power BI complet ainsi qu'une vaste base de connaissances en ligne et une massive communauté à travers le monde entier.​ |
| Sécurisation de la données​ | Sécurisation et protection de la données utilisées dans l'outils​  ​ | Power BI est conforme à plus de 100 normes et certifications du Centre de gestion de la confidentialité Microsoft (par exemple, IL6, FedRAMP, HIPAA et FINRA).​ |

1. Présentation du Dashboard







1. Les indicateurs clés

* Taux d’attrition
* Taux de fidélisation
* Nombre Totale des clients
* Nombre des clients fidélisés et perdus
* Nombre des clients actifs et non actifs
* Nombre d’années moyen
* Salaire moyen
* Balance moyenne

1. Les bonnes pratiques de Dataviz :

* **A qui je souhaite m’adresser ?**

Afin de faire passer le bon message à la bonne personne. Nos tableaux de bord sont adressés aux membres d’une équipe métier celle du Marketing (ou au membre du comité de direction).

* **Quelle histoire je souhaite lui raconter ?**

Notre tableau de bord comportant les indicateurs les plus importants. Il est clair et impactant. Un graphique = un message ! les idées sont présentées par écran. L’idée est d’aller à l’essentiel et de sélectionner les indicateurs qui ont le plus d’enjeux pour notre sujet d’analyse. Nous pouvons obtenir une vision globale de l’activité. La création de plusieurs onglets nous permettra d’approfondir le sujet étudié et d’aller plus loin en granularité.

* **Quelles sont les informations importantes que ma cible souhaite voir tout de suite ?**

Nos tableaux de bord présentent des indicateurs mesurant la problématique de taux d’attrition (Churn).

Par exemple : une augmentation du taux d’attrition (Churn) de notre banque va générer une prise d’action immédiate du département marketing. La donnée est actionnable, elle révèle une anomalie. C’est à cela que l’on reconnaît un bon tableau de bord.

* **Je dispose de quelles données ?**

Nous disposons des données concerne les churners (le taux d’attrition et les churners sont le sujet de notre analyse)

**Les bonnes pratiques pour les choix des graphiques :**

Voici quelques exemples de graphiques de notre des tableaux de bord en fonction du besoin spécifique de notre utilisateur final.

* **Line chart :** 
  + Comparer le taux d’attrition par churners et par région
* **Bar chart :** 
  + Comparer le nombre total des clients actifs et non actifs au cours des années
  + Comparer le nombre des churners avec le nombre des clients fidèles
* **Les cartes :** La région géographie à un facteur important dans l’analyse parce qu’on voit bien la répartition du taux de churn par rapport à chaque région
* **Graphique linéaire :** 
  + Comprendre les tendances et les fluctuations de taux de churn ou de fidélisation au fil des années
  + Comprendre les tendances des taux de churn par profile du client ou par cote de score
* **Sparkilnes :** associer le nombre des churners qui a une valeur d’état actuel suivie sur une période de temps spécifique
* **Camembert :** Comparer le nombre des churners par rapport à leurs genre (Masculin ou féminin)

Conclusion :

Comme nous avons pu le voir, les clients non actifs sont naturellement plus susceptibles de quitter la banque mais ce qui est plus intéressant c'est le fait qu'il soit d'une majorité de sexe féminin et surtout jeune ou adolescent.

Effectivement l’âge reste un facteur important car la banque est moins attrayante pour les profils jeunes.

Mais surtout ont conclu sur une corrélation certaine entre l’appréciation donnée au client via le score lors de son abonnement, et la probabilité future d’un churn de ça part.

Pour finir, cette analyse nous a permis de nous rendre compte du taux élevé de churn en Allemagne, en effet à ce rythme la division allemande de la banque perdra plus de clients qu’elle en attire.

1. Les décisions à retenir pour le projet

* **Cibler les campagnes de fidélisation :** mettre en place des campagnes de fidélisation ciblant spécifiquement les femmes et les jeunes clients, en utilisant des offres et des incitations adaptées à ces groupes de clients.
* **Améliorer l'expérience client :** améliorer l'expérience client pour les jeunes en proposant des produits et des services adaptés à leurs besoins et en utilisant des canaux de communication pertinents pour ces groupes de clients.
* **Évaluer les processus d’inscription :** évaluer les processus d'inscription pour s'assurer que les clients sont correctement évalués et que les scores d'appréciation sont fiables.
* **Mettre en place des programmes de fidélité :** mettre en place des programmes de fidélité pour récompenser les clients fidèles et les inciter à rester avec la banque.