

Analyse Des Tendances De Prénoms De Bébé

Diego Sanoja

2025-06-09

Introduction

Ce document a pour objectif d'expliquer les étapes de l'analyse, réalisée à l'aide de SQL dans la base de données de MySQL qui contient l'information des prénoms de bébés que les parents donnent à leurs bébés quand ils sont nés. De plus, Power BI a été utilisé afin de créer un rapport pour montrer les idées trouvées par les requêtes de l'analyse.

Objectif Du Projet

Le but de ce projet était d'enquêter dans les données des prénoms de bébés afin de produire des conclusions intéressantes des noms pendant les ans pour une entreprise de sites web fictifs. Après, de produire un rapport Power BI pour visualiser et partager les informations obtenues et en générer de nouvelles.

Journal de Modifications

Cette section donnera une liste des modifications faites aux données. Remarque: pour les colonnes et les mesures créées, un document PDF qui contient les codes DAX utilisés pour créer les sera présenté dans un fichier séparé.

Version 1.0.0 (04-06-2025)

- Effectuer une analyse exploratoire des données dans les 2 tables de la base de données. Pour faire ça, des requêtes qui renvoyant des différentes informations ont été écrites:
 - Pour la table Noms, des requêtes ont été écrites afin de trouver:
 - * Le plage d'années dans le tableau.
 - * Les différentes valeurs des colonnes Genre et État.
 - * Le nombre total de prénoms différents pour les garçons et les filles.
 - * La présence de valeurs manquantes dans les colonnes.
 - Pour la table Régions, des requêtes ont été écrites afin de trouver:
 - * Les différentes valeurs de la colonne État pour les comparer avec celles de la même colonne dans le tableau Noms.
 - * L'état manquant dans cette table (en utilisant une JOIN).
- Écrire des requêtes pour découvrir les informations suivantes dans l'ordre donné:

- Les prénoms de filles et de garçons les plus populaires, ainsi qu’une colonne indiquant leur popularité (à l’aide de la fonction de fenêtre ROW_NUMBER).
- L’évolution de la popularité des prénoms de l’année 1980 en comparaison à celle des mêmes noms en 2009.
- Les 3 prénoms de filles et de garçons les plus populaires par année et par décennie, avec une colonne indiquant le nombre de bébés qui ont reçu chaque prénom.
- Le nombre total de bébés nés par région au cours des 3 décennies de la base de données.
- Les 10 prénoms les plus populaires utilisés par les deux, garçons et filles.
- Les prénoms courts et longs les plus populaires.
- Le pourcentage total de bébés qui ont reçu le prénom Chris par état.

Version 1.1.0 (05-06-2025)

- Importer les 3 tables dérivées de la base de données à l’aide des requêtes dans un nouveau fichier Power BI Desktop pour faire d’analyse plus approfondie. Ces tableaux sont les suivants:
 - Régions (Regions) qui contient les noms des états et de leurs régions.
 - Bébés Par Année (Babies Per Year) qui contient les 5 colonnes suivantes: Année (Year), Genre (Gender), Prénom (Name), Bébés_totale (Total_babies), et Popularité (Popularity).
 - Bébés Par État (Babies Per State) qui contient les 5 colonnes suivantes: State (État), Genre, Prénom, Bébés_totale et Popularité.
- Créer 2 pages de rapport avec les titres ‘Bébés Par Année’ et ‘Bébés Par État’ dans le fichier Power BI.
- Ajouter un nouvelle table pour stocker les mesures calculées pour le rapport.
- Ajouter les mesures suivantes dans le nouveau tableau:
 - Bébés Total (Total Babies) pour calculer la somme des bébés nés.
 - % de Bébés (% of Babies) pour évaluer le % de bébés lors de l’utilisation de filtres.
 - Total Des Hommes (Total Males) pour calculer le nombre de garçons.
 - Total Des Femmes (Total Females) pour évaluer le nombre de filles.
 - Total Des Prénoms Différents (Total Different Names) pour calculer le nombre de différents prénoms donnés aux bébés.
- Stocker les mesures précédentes dans un dossier appelé ‘Bébés Par Année’.
- Ajouter dans le dossier ‘Bébés Par Année’ les 2 mesures calculées suivantes:
 - Popularité Masculine (Male Popularity) pour calculer la popularité d’un prénom de garçon spécifique par année.
 - Popularité Feminine (Female Popularity) pour évaluer la popularité d’une prénom de fille spécifique par année.
- Ajouter les visuels suivants à la page ‘Bébés Par Année’:
 - Trois filtres (slicers) à droite pour filtrer les données par genre, année et prénom.
 - Une carte en haut à gauche pour afficher les valeurs des 5 premières mesures calculées du dossier de cette page.

- Un graphique linéaire au milieu à gauche pour montrer le nombre total de garçons et de filles nés par année.
- Une matrice en bas à gauche pour afficher le nombre de prénoms donnés aux garçons et aux filles et leur popularité par année.
- Répéter toutes les tâches effectuées pour les mesures calculées et les visuels dans la page ‘Bébés Par Année’ pour remplir la page ‘Bébés Par État’ tandis qu’en effectuant les modifications suivantes:
 - Utiliser les colonnes du tableau ‘Bébés Par État’ pour créer les mesures calculées.
 - Changer le nom du dossier pour celui de la nouvelle page.
 - Ajouter V2 à la fin des nouveaux calculs.
 - Remplacer le graphique linéaire par un graphique à barres côte à côte.
 - Changer la colonne Année dans le filtre par la colonne Région column de la table Régions. Similairement, remplacer la colonne Année dans la matrice et dans le graphique à barres par la colonne État du tableau ‘Bébés Par État’.

Version 1.2.0 (06-06-2025)

- Utiliser la fonction DISTINCT afin de créer une nouvelle table appelée ‘Longueur De Prenoms’ (Names Length) qui stocke les différents prénoms de bébé.
- Ajouter un colonne à la nouvelle table à l’aide de la fonction LEN pour montrer le nombre de lettres de chaque nom.
- Connecter la nouvelle table avec les tables Bébés Par Année et Par État en utilisant les colonnes Prenoms.
- Ajouter un nouveau filtre (slicer) pour filtrer par le nombre de lettres du prénom dans la page ‘Bébés Par Année’.

Informations Sur Les Données

Cette section présentera la liste des informations fournies par l’analyse effectuée dans MySQL et les visuels créés dans les pages du rapport Power BI.

MySQL

Cette sous-section présentera chacune des informations obtenues à chaque phase de l’analyse effectuée a l’aide de MySQL.

Analyse Exploratoire Initiale Des Données

Le objectif de cette étape était d’apprendre plus des tableaux qui seraient utilisés après pour l’analyse et vérifier s’il y avait des erreurs et des fautes de frappe qui pourraient générer de résultats inexacts dans les étapes futures du projet. Sur cette étape, les informations suivantes ont été trouvées:

Pour le tableau Noms:

- Le plage d’années s’étend de 1980 à 2009.
- Il n’y a pas de fautes de frappe dans les colonnes État et Genre.

- Il n'y aucune valeur manquante dans les colonnes.
- Il y a un total de 14474 et 9730 différents prénoms pour les filles et les garçons respectivement mais le total de différents prénoms est de 22240 (qu'est inférieur à la somme des 2 nombres précédents). Alors, certains noms dans le tableau sont utilisés par les deux, filles et garçons.
- Il y a 51 différents états dans le tableau.

Pour le tableau Régions:

- Il y a 50 différents états dans le tableau (un de moins que dans la table Noms).
- L'état manquant dans cette table est MI qui devrait être situé dans la région du Midwest.

Suivre Les Changements De Popularité Des Prénoms

Le but de cette partie était de découvrir les informations suivantes: les prénoms les plus populaires pour les garçons et les filles pendant les 3 décennies de la base de données, le changement de popularité de ces noms pendant les années, et quels prénoms présents dans la première et la dernière année avaient la plus forte variation de popularité. Cinq requêtes ont été rédigées pour répondre les questions précédentes et elles ont révélé les informations suivantes:

- Le prénom le plus populaire pour garçons était Michael avec un total de 1376418 bébés recevant ce nom.
- Le prénom le plus populaire pour filles était Jessica avec un total de 863121 bébés recevant ce nom.
- De 1980 à 1997, Jessica était entre les 3 prénoms le plus utilisés pour les filles avec au moins de 20000 bébés recevant ce nom chaque année mais après cela, sa popularité a diminué plus chaque année.
- La popularité globale du prénom Michael pour garçons est restée en haut élevée puisqu'il a été le plus populaire pendant les premières 19 années et le deuxième le plus populaire pendant les 10 années suivantes.
- Les 5 prénoms qui ont eu la plus forte augmentation de popularité pendant la période de 30 années sont Colton, Aidan, Rowan, et Skylar.
- Les 5 prénoms qui ont eu la plus forte diminution de popularité pendant la période de 30 années sont Rusty, Tonia, Cherie, Kerri et Tara.

Comparer La Popularité À Travers Des Décennies

Le objectif de cette étape était de découvrir les 3 prénoms les plus utilisés par les garçons et les filles par année et par décennie. Pour cela, une requête distincte a été écrite pour chaque cas et ses résultats ont été triés pour obtenir chaque conclusion. Les informations suivantes ont été obtenues à partir de ces requêtes:

- Pour les femmes, quelques noms populaires autres que Jessica (qui était le plus utilisé par filles dans les années 80 et 90) au cours du 3 décennies de la base de données ont été:
 - Jennifer (le plus populaire de 1980 à 1984 et le deuxième le plus utilisé les années 80).
 - Amanda (le deuxième ou le troisième plus utilisé pendant 7 ans des années 80 et le deuxième plus populaire de la décennie).
 - Ashley (parmi les premiers 3 noms chaque année de 1984 à 1997 et le deuxième plus populaire des années 90).

- Emily (parmi les premiers 3 noms chaque année de 1994 à 2008 et le plus populaire des années 2000).
- Pour les hommes, quelques noms populaires autres que Michael (qui était le plus utilisé par garçons dans les années 80 et 90 et le deuxième dans les années 2000) au cours du 3 décennies de la base de données ont été:
 - Christopher (le deuxième plus utilisé de 1980 à 1994 et le deuxième plus populaire pendant ces décennies).
 - Matthew (soit comme le deuxième ou le troisième plus populaire chaque an des années 80 et 90 et le deuxième plus utilisé pendant ces décennies).
 - Jacob (le nom le plus populaire de 1999 à 2009).
 - Joshua (le troisième plus populaire chaque année de 2002 à 2006 et le troisième plus utilise au cours de la décennie correspondante).
- Il y avait plus de changements de popularité pour les 3 prénoms le plus utilisés pour les filles que pour les garçons dans le 3 décennies.

Comparer La Popularité À Travers Des Régions

Le but de cette partie était de découvrir les 3 prénoms les plus populaires pour les garçons et les filles par région et le nombre total de bébés nés dans ces régions au cours des 3 décennies. Pour cela, une requête distincte a été écrite pour chaque cas et ses résultats ont été triés pour obtenir chaque conclusion. Ces requêtes ont révélé les informations suivantes:

- Les régions avec le plus grand et le plus petit nombre bébés nés étaient le Sud et la Nouvelle-Angleterre avec un total de 34219920 et 4269213 respectivement.
- Pour les femmes:
 - Jessica était le prénom le plus populaire par chaque région sauf le Sud.
 - Ashley était entre les 3 noms les plus utilisés par chaque région sauf la Nouvelle-Angleterre.
 - Jennifer était le troisième prénom le plus populaire dans le centre de l'Atlantique et le Sud, et le deuxième dans le Pacifique.
 - Sarah était le troisième prénom le plus utilisé dans le centre de l'Ouest et les Montagnes, et le deuxième dans la Nouvelle-Angleterre.
- Pour les hommes:
 - Michael était le prénom le plus populaire par chaque région sauf le Sud.
 - Christopher était entre les 3 noms les plus utilisés par chaque région sauf le centre de l'Ouest.
 - Matthew était le deuxième prénom le plus populaire dans la Nouvelle-Angleterre, le centre de l'Ouest et le centre de l'Atlantique.
 - Joshua était le troisième prénom le plus utilisé dans le centre de l'Ouest et le Sud, et le deuxième dans les Montagnes.

Explorer Les Prénoms Unique

Le objectif de cette étape était de rechercher des informations pour découvrir les 10 prénoms androgynes (noms donnés aux 2, les garçons et les filles) les plus populaires, le longueur des noms les plus courts et les plus longs, ainsi que leur popularité. Pour trouver cela, 3 différentes requêtes ont été écrites et elles ont révélé les informations suivantes:

- Il y avait un total de 1964 prénoms androgynes et les 4 le plus utilisés sont ceux dans la section précédente pour hommes.
- Les noms les plus courts de la base de données ont 2 lettres seulement et il y a 69 prénoms différents avec le respective nombre de caractères. De ceux, le plus populaire est Ty avec 29205 bébés recevant ce nom.
- Les noms les plus longs de la base de données ont 15 lettres et il y a 4 prénoms différents avec le respective nombre de caractères: Franciscojavier (le plus populaire avec 52 bébés recevant ce nom, Ryanchristopher, Johnchristopher and Mariodelosangel.

Power BI

Cette sous-section présentera les informations supplémentaires obtenues en visualisant les noms de l'ensemble de données par année et par région sur le 2 pages du rapport Power BI.

Bébés Par Année

Cette page du rapport a été désignée pour générer des informations de la distribution des prénoms par année. Les visuels et les filtres dans cette page ont révélé les information suivantes:

- Au cours des 3 décennies d'informations contenues dans l'ensemble de données, un total de 98.73 million de bébés sont nés et ont reçu 1 de 22240 prénoms différents.
- Des bébés dans l'ensemble de données, 52.87 million (53.55%) d'entre eux sont des garçons et les autres des filles.
- Chaque année, il y avait plus des garçons que des filles. De plus, l'année avec moins garçons a eu plus des bébés que l'année avec plus filles.
- Les années où le plus et les moins nombre de naissances de bébés ont été 1990 et 1980 respectivement.
- Les décennies où le plus et les moins nombre de naissances de bébés ont été les années 90 et 80 respectivement, avec une différence de 0.97%.
- Chaque nom de l'ensemble de données a été utilisé au moins 5 fois (montré en triant la matrice par le nombre de bébés).
- Presque de 81% des bébés de l'ensemble de données (environ 79.8 million d'entre eux) portent des prénoms qui contient entre 5 et 8 caractères.

Bébés Par État

Cette page du rapport a été désignée pour générer des informations de la distribution des prénoms par région. Les visuels et les filtres dans cette page ont révélé les information suivantes:

- Les états avec le plus et les moins nombre de naissances de bébés (pour les 2 genres) sont CA (la Californie) et WY (le Wyoming) avec un total de 14.08 millions (7.39 millions d'entre eux sont des hommes) et environ de 128680 bébés (près de 74580 d'entre eux sont des hommes).
- Des 4.27 million de bébés nés dans la région de la Nouvelle-Angleterre, presque la moitié d'entre eux sont nés dans l'état MA (le Massachusetts).
- Pour la région du centre de l'Atlantique, l'état avec le plus nombre de naissances de bébés était NY (New York) qui a eu un total de 6.69 million bébés (3.55 million de garçons).
- L'état avec le plus nombre de naissances de bébés (la Californie) est situé dans la région du Pacifique et environ de 80% des bébés dans cette région ont été nés là-bas.
- Le deuxième état avec le plus nombre de naissances de bébés était TX (Texas) et est situé dans la région Sud qui a 16 états.