

# CONCEPTION ET IMPLEMENTATION D'UN Data warehouse



Préparé par :

- MAURICE Carl-Emmanuel
- MAGNIMA Koumba Oxy Marcelle
- BOULAICH Kholoud

Encadré par :

- Pr. BENNANI A.

# Table des matières

Table des matières.....	2
Listes des figures et des tableaux .....	3
Résumé.....	4
Introduction.....	5
PARTIE I : CONCEPTION ET IMPLEMENTATION DU DATA WAREHOUSE.....	6
I.    Présentation de la source de données .....	6
II.   Besoins et attentes du projet .....	6
a)  Objectif de l'analyse.....	6
b)  Questions relatives à l'analyse .....	7
III.  Architecture du Datawarehouse .....	7
IV.   Modélisation des données .....	8
c)  Modèle du Datamart <b>Customer Profile</b> .....	8
V.    Extract/Transform/Load (ETL) .....	10
d)  Extraction des données.....	10
e)  Transformations des données.....	11
f)  Chargement des données .....	11
PARTIE II : VISUALISATION DES DONNEES .....	14
I.    Prétraitement des données pour les visuels.....	14
II.   Mise en forme des données quantitatives .....	16
a)  Business Query .....	17
Conclusion .....	20
REFERENCES .....	21
ANNEXES .....	22

# Listes des figures et des tableaux

Figure 1 : Architecture du Datawarehouse Tourisme 2023 .....	7
Figure 2 : Modèle du Datamart Customer Profile .....	8
Figure 4 : Création de la base de données .....	11
Figure 5 : Connection au fichier Power BI .....	12
Figure 6 : Exportation des données vers SQL Tables .....	12
Figure 7 : Propriétés du server abritant la base de données .....	13
Figure 8 : Vue des tables exportées dans SQL Server .....	13
Tableau 1 : Aperçu des sources de données .....	6
Tableau 2 : Fait et dimensions du Datamart 1.....	9
Tableau 4 : Vue des colonnes du dataset Customer Profile.	10
Tableau 6 : Vue des transformations .....	11
Tableau 7 : Données non encore catégorisées.....	14
Tableau 8 : Données catégorisées .....	15
Tableau 9 : Changement de type de données et de points en virgules.....	16
Tableau 10 : Sommation des données quantitatives.....	17
Tableau 11 : Tableau de bord présentant les données.....	18

## Résumé

Ce projet se concentre sur la création d'un entrepôt de données pour le secteur touristique britannique, une industrie économiquement significative. Les données extraites, provenant des voyageurs, sont transformées puis stockées dans une base de données, afin d'être utilisées pour améliorer l'expérience des voyageurs et promouvoir le tourisme au Royaume-Uni. Une visualisation des données est également conçue pour faciliter la prise de décision des gestionnaires.

## Introduction

L'intelligence des affaires ou Business Intelligence se réfère aux stratégies et aux processus mis en place par les entreprises ou les organisations afin de transformer leurs données en informations utiles pour une prise de décision éclairée et une exploitation optimale des opportunités présentes sur le marché.

Seulement, à l'ère du Big Data, les entreprises peuvent se retrouver très vite déboussolées face au volume et à la variété de ces données affluent de diverses sources. C'est dans ce cadre qu'est introduit le concept de data warehousing ou entreposage des données. Cette pratique se ramène à designer un **Single Source of Truth**, c'est-à-dire, une source unique de vérité, plus concrètement un référentiel central intégrant des données provenant de plusieurs sources internes ou externes à l'entreprise, qui seront utilisées par l'entreprise à des fins stratégiques ou opérationnels.

Dans le cadre de ce projet, nous proposons donc un entrepôt de données à taille réduite pour le secteur touristique de la Grande-Bretagne. Selon l'Organisation de Coopération et de Développement Economiques, le tourisme occupe une place importante dans l'économie britannique, et les données provenant de cette industrie ont une taille exponentielle.

Pour parfaire ce travail, nous avons extrait des ensembles de données dont les sources sont précisées par la suite, portant sur les informations des voyageurs. Nous avons par la suite transformé ces données afin de les rendre utiles aux preneurs de décision dans le cadre du secteur touristique. Après cette étape, nous avons stocké ces données dans une base de données qui pourra servir pour améliorer l'expérience du voyageur et promouvoir le tourisme en Grande-Bretagne.

Finalement, nous avons conçu une visualisation des données afin d'éclairer la prise de décision des managers.

# PARTIE I : CONCEPTION ET IMPLEMENTATION DU DATA WAREHOUSE

## I. Présentation de la source de données

Dans le cadre de ce projet, nous avons utilisé une seule source de données, dont la description est indiquée dans la présente section.

Source <sup>1</sup>	Type	Description
<b>Office for National Statistics</b>	Structuré	<p>Nous avons extrait de cette source trois ensembles de données portant sur les informations des voyageurs (les britanniques allant à l'étranger et les étrangers venant en Grande-Bretagne). Ces données portent aussi bien sur la durée de leur séjour que sur leur sexe, leur âge et bien d'autres caractéristiques déclinées dans la suite du rapport.</p> <p>La période s'étend de 2020 à 2023.</p>

Tableau 1 : Aperçu des sources de données

## II. Besoins et attentes du projet

Nous énoncerons dans la suite l'objectif poursuivi dans le cadre de cette analyse, mais également les interrogations auxquelles nous apporterons un éclairage lors de la visualisation des données.

### a) Objectif de l'analyse

Le but visé par cette analyse est de parvenir à fournir des informations utiles pouvant servir à accroître le potentiel touristique de la Grande-Bretagne. Pour y parvenir, nous analyserons particulièrement le profil des touristes qui visitent la Grande-Bretagne.

---

<sup>1</sup> Les liens des sources de données seront disponibles dans la section des références.

### b) Questions relatives à l'analyse

1. De quel pays proviennent les touristes qui affluent en grand nombre en Grande-Bretagne ?
2. Quelles sont les fréquentes raisons des arrivées des touristes en Grande-Bretagne ?
3. Quelles sont les tranches d'âge des touristes qui dépensent le plus ?
4. Quel est le mode de voyage le plus utilisé ?

## III. Architecture du Datawarehouse

Pour la conception du présent datawarehouse, nous avons opté pour l'approche dite Bottom-up, promue par Ralph KIMBALL. Ce dernier soutient qu'un entrepôt de données est une combinaison<sup>2</sup> de compartiments dénommés "Data Mart". Les étapes de l'élaboration de notre datawarehouse sont déclinées dans l'image suivante.

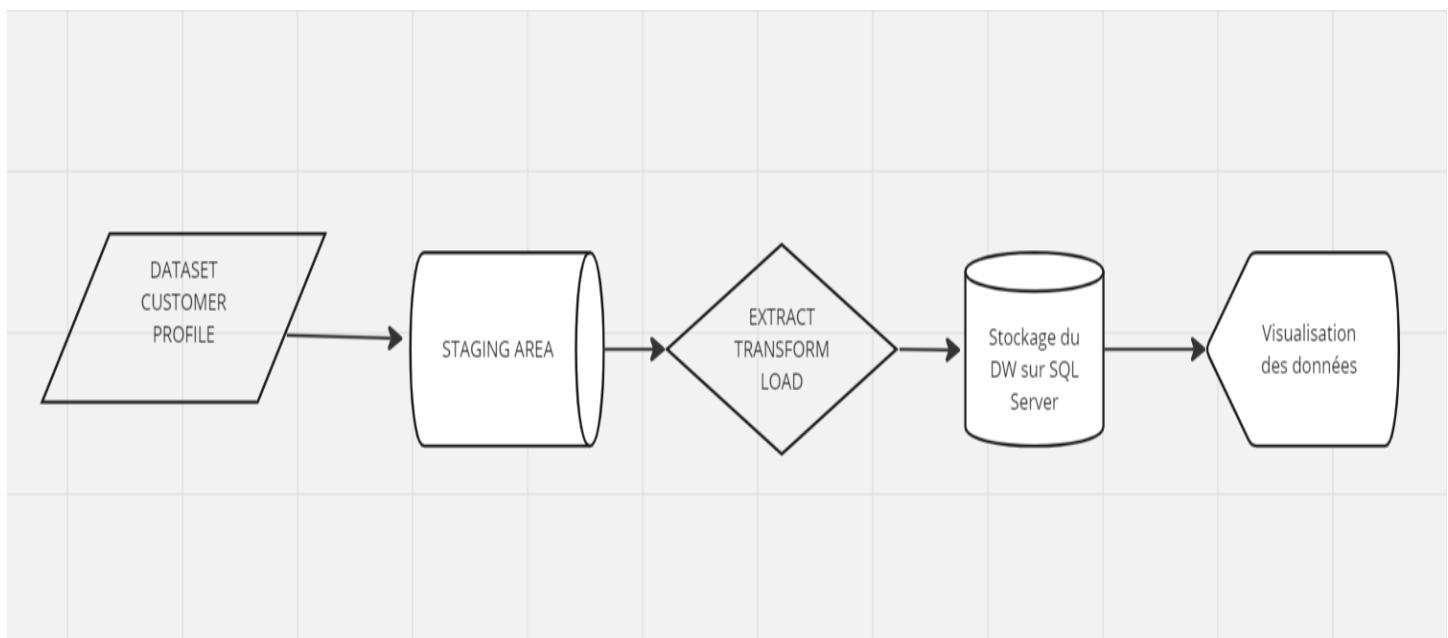


Figure 1 : Architecture du Datawarehouse Tourisme 2023

<sup>2</sup> <https://www.cartelis.com/blog/data-warehouse-modelisation-etoile/>

## IV. Modélisation des données

Afin de réduire la complexité de la compréhension des données, ainsi que faciliter les requêtes portant sur ces données, nous avons opté pour une approche dite dénormalisée en utilisant la modélisation en Etoile. Cette dernière nous permet de présenter les données sous la forme d'une étoile, plus concrètement, nous avons au centre une table dite de "faits<sup>3</sup>" et tout autour de cette dernière un ensemble de tables dites de "dimension"<sup>4</sup> liées à la table de faits par des relations<sup>5</sup>.

Nous présenterons donc de manière distincte les modèles relatifs aux différents ensembles de données.

### c) Modèle du Datamart *Customer Profile*

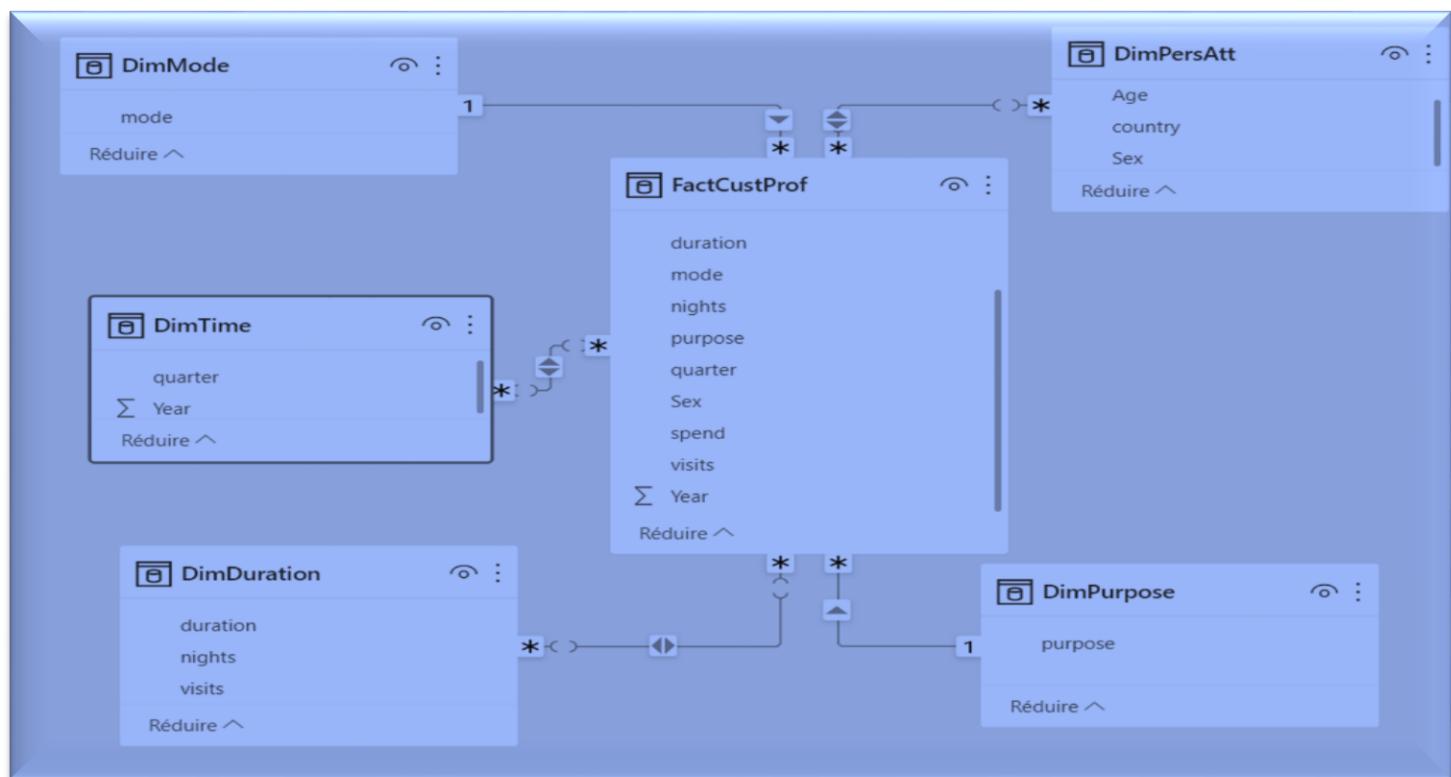


Figure 2 : Modèle du Datamart *Customer Profile*

<sup>3</sup> Une table de faits contient des colonnes clés de dimension qui se rapportent aux tables de dimension et des colonnes de mesures numériques.

<sup>4</sup> Les tables de dimension contiennent précisément des colonnes plus descriptives des entités de l'entreprise ou de l'organisation.

<sup>5</sup> Les relations se rapportent au lien qui unissent les tables de dimension à la table de faits, nous avons entre autres : plusieurs à un (\*: 1) ; un à plusieurs (1 : \*) ; un à un (1 : 1) et plusieurs à plusieurs (\* : \*).

Table	Type	Description
CustProf	Fait	Cette table présente toutes les informations relatives aux voyageurs rendues disponibles via la source que nous avons énoncée plus haut.
Mode	Dimension	Il s'agit du moyen de transport utilisé par le voyageur, il y a entre autres : l'avion, la mer ou le tunnel
PersAtt	Dimension	Cette table référence les informations portant sur les attributs propres du voyageur tels que son genre, son âge (ou plutôt intervalle d'âge) mais également son pays d'origine.
Duration	Dimension	Les informations présentes dans cette table se rapportent à la durée du séjour du voyageur ou encore au nombre de nuits et de visites qu'il a effectué en Grande-Bretagne.
Purpose	Dimension	La raison du séjour du voyageur est déclinée dans cette table ; que ce soit pour des vacances, des études, le travail ou autres.
Time	Dimension	Indispensable dans toute modélisation, cette table se réfère à la vision temporelle, que ce soit annuel ou trimestriel.

Tableau 2 : Fait et dimensions du Datamart 1

## V. Extract/Transform/Load (ETL)

Cette étape revêt une importance cruciale dans la conception d'un datawarehouse. Elle nous garantit une intégration des données provenant de plusieurs sources, des données propres, fiables et utiles à l'analyse. En effet, au cours de ce processus, les données sont extraites de leur source d'origine, nettoyées puis chargées dans un espace de stockage qui abritera l'entrepôt de données<sup>6</sup>.

### d) Extraction des données

Nous présenterons un échantillon de l'ensemble de données sur lequel a porté notre analyse.

Year	quarter	ukos	mode	country	purpose	package
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent
2020	Jan-Mar	UK residents	Air	Austria	Holiday	Independent

Age	Sex	duration	12 visits	12 nights	12 spend	sample
25-34	Female	1-3 nights	2223,711	4447,422	560375,211	
25-34	Female	4-13 nights	5027,223	34225,923	1075193,969	
35-44	Male	4-13 nights	27709,513	157528,438	20077008	
35-44	Female	1-3 nights	591,172	1182,344	88675,802	
35-44	Female	4-13 nights	14706,683	71838,369	5978837,83	
35-44	Female	14-27 nights	664,589	9304,241	1482697,223	
45-54	Male	1-3 nights	1302,04	3195,253	260408,097	

Tableau 3 : Vue des colonnes du dataset Customer Profile.

<sup>6</sup> Cet espace de stockage est généralement une base de données OLAP gérée par un Système de Management de Bases de Données (DBMS).

### e) Transformations des données

Nous présenterons dans le tableau qui suit les transformations incluant le nettoyage des données opérées dans le cadre cette analyse.

Dataset	Transformations
Customer Profile	<ul style="list-style-type: none"><li>-Concaténation des ensembles de données de 2020 à 2023</li><li>-Suppression des colonnes <b>Sample</b> et <b>Package</b></li><li>-Suppression des éléments vides</li><li>-Suppression des valeurs manquantes</li></ul>

Tableau 4 : Vue des transformations

### f) Chargement des données

Afin de procéder au chargement des données dans la base de données de SQL Server , nous devons dans un premier temps créer une base de données à l'aide de l'outil SQL Server Management Studio (SSMS).

Pour ce faire, nous écrivons la requête suivante :

```
CREATE DATABASE TourismDatawarehouse2023  
GO
```

Figure 3 : Création de la base de données

Pour exporter les données depuis Power BI jusqu'à SQL Server, nous utiliserons un outil appelé DAX studio. Pour cela, nous devons ouvrir nos fichiers sur Power BI et lancer l'ouverture de l'outil DAX Studio. Plus concrètement, DAX Studio se connectera automatiquement à notre fichier Power BI déjà ouvert et nous offrira l'option d'export de données, puis nous proposera de nous connecter à la base de données que nous avons précédemment créée sur SQL Server, ceci en renseignant les propriétés relatives à la base de données. Les images qui suivent seront plus explicites concernant les étapes énoncées :

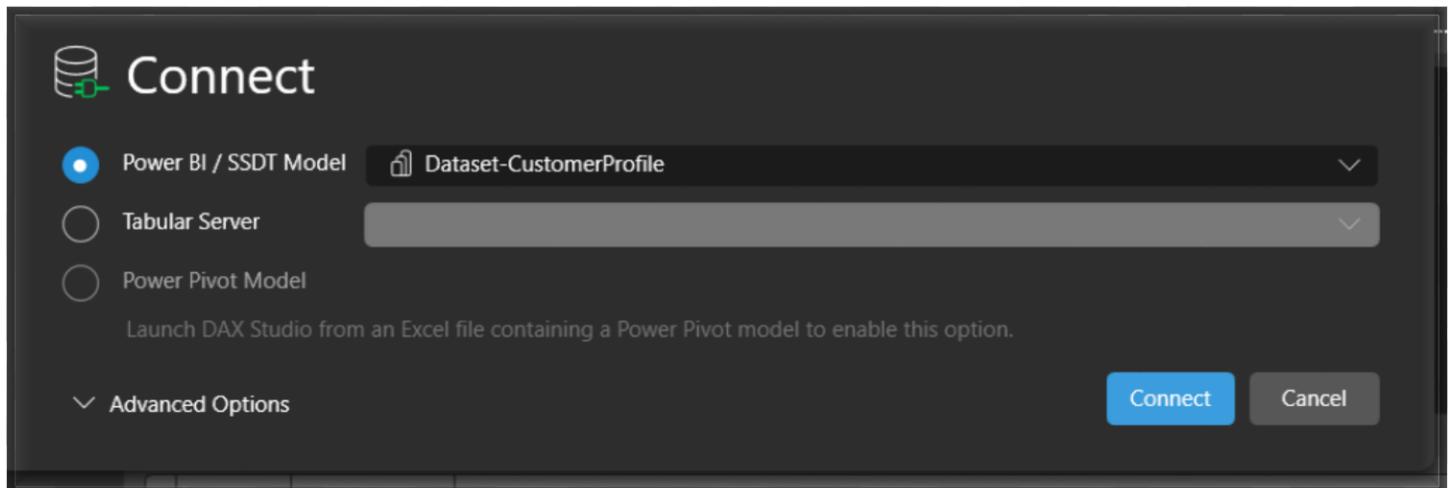


Figure 4 : Connection au fichier Power BI

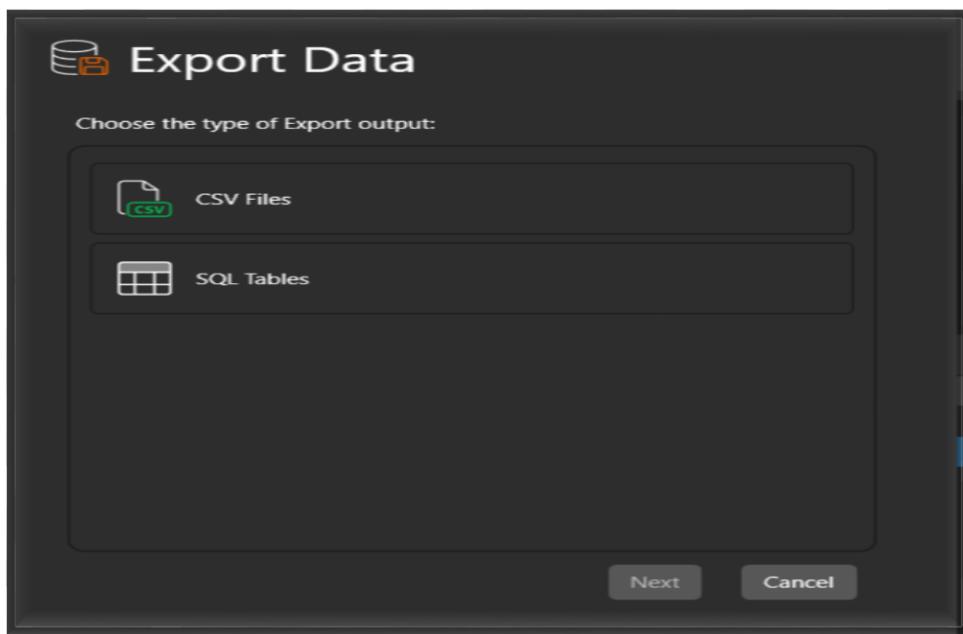


Figure 5 : Exportation des données vers SQL Tables

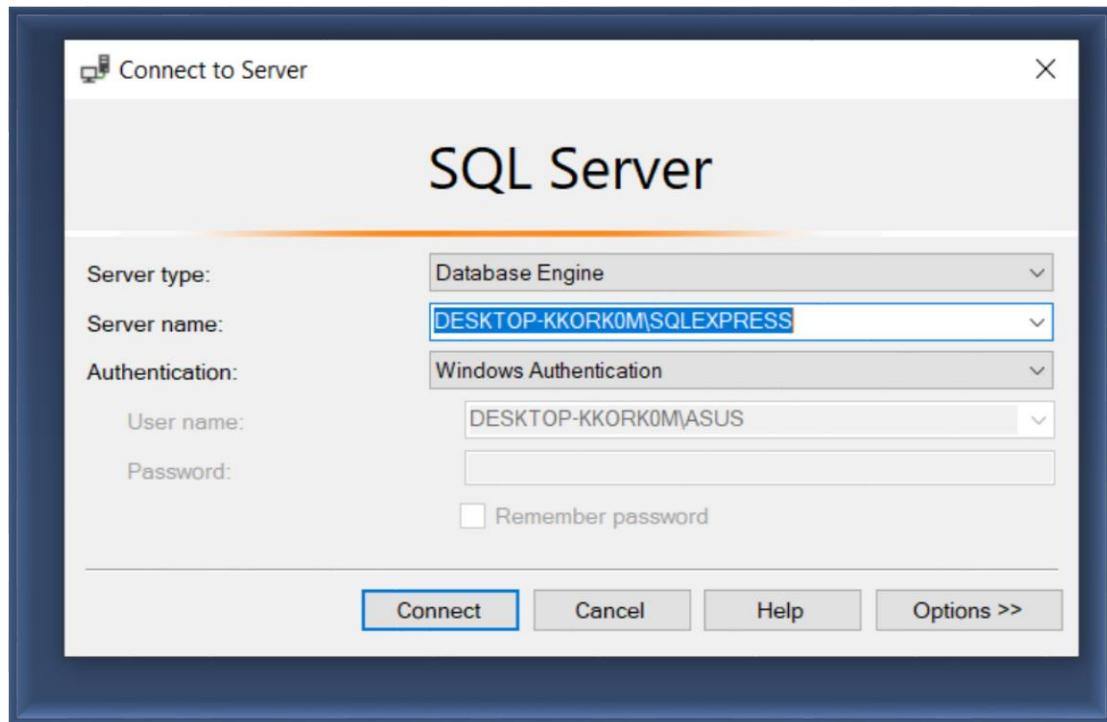


Figure 6 : Propriétés du server abritant la base de données

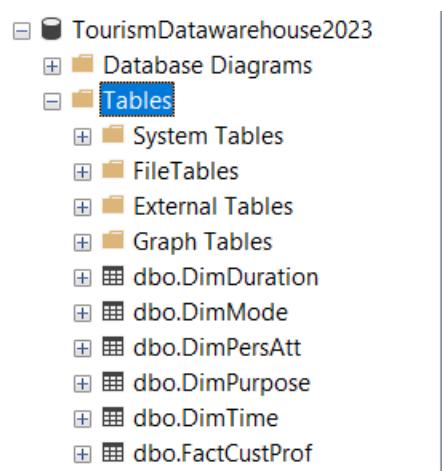


Figure 7 : Vue des tables exportées dans SQL Server

## PARTIE II : VISUALISATION DES DONNEES

### I. Prétraitement des données pour les visuels

Dans le but de transformer les données d'une manière à être convenables pour la visualisation, nous avons changé premièrement les types de données pour chacune des colonnes, le cas échéant, ensuite nous avons altéré, si besoin est, leurs propriétés. Ce qui nous a permis de faire la distinction entre les diverses catégories de données. Les figures ci-dessous en illustrent.

The screenshot shows the Power BI Desktop interface with the 'Dataset-CustomerProfile' dataset open. The 'Outils de colonne' (Column Tools) ribbon tab is active. The main workspace displays a table with 2869 rows and 12 columns. The columns are: Year, quarter, category, mode, country, purpose, age, sex, duration, visits, nights, and spend. The 'Données' pane on the right lists dimensions and facts, with 'Year' currently selected. The status bar at the bottom indicates the table has 2869 rows and provides download options.

Tableau 5 : Données non encore catégorisées

Fichier Accueil Aide Outils externes

Dataset-CustomerProfile - Power BI Desktop Se connecter

Outils de table

Nom FactCustProf Marquer comme table de dates Gérer les relations Nouvelle mesure Nouvelle mesure rapide colonne table

Structure Calendriers Calculs

**Données**

Rechercher

Year quarter Category mode country purpose Age Sex duration visits nights spend

Year	quarter	Category	mode	country	purpose	Age	Sex	duration	visits	nights	spend
2020	Jan-Mar	Overseas residents	Air	Bulgaria	VFR	45-54	Female	4-13 nights	1597.501473176574	12780.011785412593	1314743.7124243204
2020	Jan-Mar	Overseas residents	Air	Czech Republic	VFR	45-54	Male	4-13 nights	1246.9800847191211	6988.41055809758	66550.9063599097
2020	Jan-Mar	Overseas residents	Air	Czech Republic	VFR	45-54	Female	4-13 nights	1901.2981252913671	11407.788751748203	749111.4613647986
2020	Jan-Mar	Overseas residents	Air	Cyprus EU	VFR	45-54	Male	4-13 nights	1030.3863331215614	5151.931665607807	530648.9615576041
2020	Jan-Mar	Overseas residents	Air	France	VFR	45-54	Female	4-13 nights	11112.267196042336	49436.63288098629	2411499.1645355206
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Male	4-13 nights	11460.262380565251	59621.11357180085	4267284.5701714065
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Female	4-13 nights	1764.437306740105	9075.81258485234	278450.853054532
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Female	4-13 nights	209.02036815799377	1045.101840789969	133564.01525295802
2020	Jan-Mar	Overseas residents	Air	Gibraltar	VFR	45-54	Female	4-13 nights	368.4872287087054	4790.333897321317	79224.75291723717
2020	Jan-Mar	Overseas residents	Air	Greece	VFR	45-54	Female	4-13 nights	2107.004879836775	13057.60359654964	512351.7853066982
2020	Jan-Mar	Overseas residents	Air	Hungary	VFR	45-54	Male	4-13 nights	2178.1770712639645	12560.035250014913	881811.3194642459
2020	Jan-Mar	Overseas residents	Air	Irish Republic	VFR	45-54	Male	4-13 nights	3753.3236837742907	26445.349876609547	1044465.5646512171
2020	Jan-Mar	Overseas residents	Air	Irish Republic	VFR	45-54	Female	4-13 nights	20106.237314860024	158981.24890514987	8770724.724549713
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Male	4-13 nights	1425.167539799113	6953.902051613679	537685.6213902178
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Female	4-13 nights	3429.0662115779105	20085.061131146864	929141.337751882
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Female	4-13 nights	476.48434831189047	3811.8747864951238	97679.29140393755
2020	Jan-Mar	Overseas residents	Air	Latvia	VFR	45-54	Male	4-13 nights	2175.3422602127625	15227.395821489337	160975.32725574443
2020	Jan-Mar	Overseas residents	Air	Lithuania	VFR	45-54	Female	4-13 nights	506.28369677831574	3543.98587744821	271874.3451699556
2020	Jan-Mar	Overseas residents	Air	Malta	VFR	45-54	Male	4-13 nights	1755.3910624521536	8776.955312260769	88474.59460405461
2020	Jan-Mar	Overseas residents	Air	Malta	VFR	45-54	Female	4-13 nights	838.3158710351933	6706.526968281546	222153.7058243262
2020	Jan-Mar	Overseas residents	Air	Netherlands	VFR	45-54	Male	4-13 nights	2616.829715348792	15472.71948712366	513928.4720677808
2020	Jan-Mar	Overseas residents	Air	Norway	VFR	45-54	Female	4-13 nights	3948.7378870082607	15794.951548030343	2194751.78584926
2020	Jan-Mar	Overseas residents	Air	Poland	VFR	45-54	Male	4-13 nights	1976.2201235458829	12845.663627314636	951236.2963006891
2020	Jan-Mar	Overseas residents	Air	Poland	VFR	45-54	Female	4-13 nights	1975.7544750130887	17123.205450113437	649035.3450417997
2020	Jan-Mar	Overseas residents	Air	Portugal	VFR	45-54	Male	4-13 nights	1925.6011601735815	14131.160321234878	569796.410772065
2020	Jan-Mar	Overseas residents	Air	Portugal	VFR	45-54	Female	4-13 nights	15434.739542486375	123477.916339891	231521.09313729563
2020	Jan-Mar	Overseas residents	Air	Russia	VFR	45-54	Male	4-13 nights	869.4634995762168	6086.244497033517	4360359.450374727
2020	Jan-Mar	Overseas residents	Air	Spain	VFR	45-54	Male	4-13 nights	10970.929911825473	73095.26041956079	3271463.5830537924

Tableau : FactCustProf (2 869 lignes)

Mise à jour disponible (cliquez pour la télécharger)

Windows Search Home Internet Mail File Power BI Data Flow Power BI DAX Power BI Visual Studio 19°C FRA 29/11/2023

Tableau 6 : Données catégorisées

## II. Mise en forme des données quantitatives

Pour faciliter l'utilisation des données quantitatives et les calculs agrégés y relatifs, nous avons simplifiées leurs écritures. Les deux figures qui suivent montrent les étapes.

	Age	Sex	duration	visits	nights	spend
1	16-24	Male	4-13 nights	971.57347863232544	8744.161307690389	694675.0372220619
2	16-24	Female	4-13 nights	369.8041732252901	1479.2166929011605	301390.40117861144
3	25-34	Male	1-3 nights	1170.0416584760462	3172.28883267874	606287.6511289267
4	25-34	Male	4-13 nights	344.1382130917782	2408.9674916424474	91196.62646932123
5	25-34	Female	1-3 nights	260.0742871850886	780.222861552658	133938.25790032063
6	35-44	Male	4-13 nights	447.035000392798	1788.14001571192	216811.97519050702
7	35-44	Female	1-3 nights	260.0742871850886	780.222861552658	198956.8296965928
8	45-54	Male	1-3 nights	737.2517319448696	2211.751598346085	195371.10896539043
9	35-44	Female	1-3 nights	975.5539325952748	1951.1078651905495	209744.09550798408
10	45-54	Male	1-3 nights	475.1961788741093	950.3923577482186	40391.67520429929
11	45-54	Male	4-13 nights	393.7406703001546	1574.9626812006184	793387.4506548115
12	0-15	Female	4-13 nights	376.755061070761	2637.2854627495326	43326.832602313756
13	16-24	Male	4-13 nights	737.2517319448696	2949.0069277794782	84783.94917366
14	16-24	Female	4-13 nights	5173.844530920439	24021.820743990224	2535137.1876109913
15	25-34	Male	4-13 nights	376.755061070761	4897.81585939199	43326.832602313756
16	25-34	Female	4-13 nights	248.854925721729	995.419702886916	16175.570171912384
17	35-44	Male	4-13 nights	129.9362849967067	649.6814249835335	73414.00102313928
18	35-44	Female	4-13 nights	376.755061070761	2637.2854627495326	62164.5859076756
19	55-64	Male	4-13 nights	1107.0559051701598	5907.440313581799	200918.77156376978
20	65 & over	Female	4-13 nights	737.2517319448696	2949.0069277794782	84783.94917366
21 is	0-15	Female	4-13 nights	18653.476690408766	130574.33683286136	20724012.60304414
22 is	25-34	Male	1-3 nights	432.2954585330235	1296.886397559907	101589.43447552605
23 is	35-44	Male	1-3 nights	129.9362849967067	389.8088549901200	21439.487024456605
24 is	45-54	Male	1-3 nights	6991.523917846757	6991.523917846757	104872.85876770136
25 is	35-44	Male	1-3 nights	832.2055663157817	1664.411326315633	455216.4447747326
26	25-34	Female	1-3 nights	975.5539325952748	2926.661797785824	624354.5168609759
27	35-44	Male	1-3 nights	628.7940481766826	1257.5880963533652	381049.1931950697
28	35-44	Female	1-3 nights	1861.4179252508711	3722.8358505017422	368560.7491996725

Tableau 7 : Changement de type de données et de points en virgules

Dataset-CustomerProfile - Power BI Desktop

Fichier Accueil Aide Outils externes Outils de table Outils de colonne

Nom : spend Format : Nombre décimal Type de données : Nombre décimal

Résumé Somme Catégorie de données : Aucune catégorie Trier par colonne Propriétés Gérer les relations Nouvelle colonne

Structure Mise en forme

**Outils de colonne**

Year	quarter	Category	mode	country	purpose	Age	Sex	duration	visits	nights	spend
2020	Jan-Mar	Overseas residents	Air	Bulgaria	VFR	45-54	Female	4-13 nights	1597,50	12780,01	1314743,71
2020	Jan-Mar	Overseas residents	Air	Czech Republic	VFR	45-54	Male	4-13 nights	1246,98	6988,41	665502,91
2020	Jan-Mar	Overseas residents	Air	Czech Republic	VFR	45-54	Female	4-13 nights	1901,30	11407,79	749111,46
2020	Jan-Mar	Overseas residents	Air	Cyprus EU	VFR	45-54	Male	4-13 nights	1030,39	5151,93	530648,96
2020	Jan-Mar	Overseas residents	Air	France	VFR	45-54	Female	4-13 nights	11112,27	49436,63	2411499,16
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Male	4-13 nights	11460,26	59621,11	4267284,57
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Female	4-13 nights	1764,44	9075,81	278450,85
2020	Jan-Mar	Overseas residents	Air	Germany	VFR	45-54	Female	4-13 nights	209,02	1045,10	133564,02
2020	Jan-Mar	Overseas residents	Air	Gibraltar	VFR	45-54	Female	4-13 nights	368,49	4790,33	79224,75
2020	Jan-Mar	Overseas residents	Air	Greece	VFR	45-54	Female	4-13 nights	2107,00	13057,60	512351,79
2020	Jan-Mar	Overseas residents	Air	Hungary	VFR	45-54	Male	4-13 nights	2178,18	12560,04	881811,32
2020	Jan-Mar	Overseas residents	Air	Irish Republic	VFR	45-54	Male	4-13 nights	3753,32	26445,35	1044465,56
2020	Jan-Mar	Overseas residents	Air	Irish Republic	VFR	45-54	Female	4-13 nights	20106,24	158981,25	8770724,72
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Male	4-13 nights	1425,17	6953,90	537685,62
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Female	4-13 nights	3429,07	20085,06	929141,34
2020	Jan-Mar	Overseas residents	Air	Italy	VFR	45-54	Female	4-13 nights	476,48	3811,87	97679,29
2020	Jan-Mar	Overseas residents	Air	Latvia	VFR	45-54	Male	4-13 nights	2175,34	15227,40	160975,33
2020	Jan-Mar	Overseas residents	Air	Lithuania	VFR	45-54	Female	4-13 nights	506,28	3543,99	271874,35
2020	Jan-Mar	Overseas residents	Air	Malta	VFR	45-54	Male	4-13 nights	1755,39	8776,96	88474,59
2020	Jan-Mar	Overseas residents	Air	Malta	VFR	45-54	Female	4-13 nights	838,32	6706,53	22153,71
2020	Jan-Mar	Overseas residents	Air	Netherlands	VFR	45-54	Male	4-13 nights	2616,83	15472,72	513928,47
2020	Jan-Mar	Overseas residents	Air	Norway	VFR	45-54	Female	4-13 nights	3948,74	15794,95	2194751,79
2020	Jan-Mar	Overseas residents	Air	Poland	VFR	45-54	Male	4-13 nights	1976,22	12845,66	951236,30
2020	Jan-Mar	Overseas residents	Air	Poland	VFR	45-54	Female	4-13 nights	1975,75	17123,21	649035,35
2020	Jan-Mar	Overseas residents	Air	Portugal	VFR	45-54	Male	4-13 nights	1925,60	14131,16	569796,41
2020	Jan-Mar	Overseas residents	Air	Portugal	VFR	45-54	Female	4-13 nights	15434,74	123477,92	231521,09
2020	Jan-Mar	Overseas residents	Air	Russia	VFR	45-54	Male	4-13 nights	869,46	6086,24	4360359,45
2020	Jan-Mar	Overseas residents	Air	Spain	VFR	45-54	Male	4-13 nights	10970,93	73095,26	3271463,58

Table : FactCustProf (2 869 lignes) Colonne : spend (2 740 valeurs distinctes)

Mise à jour disponible (cliquez pour la télécharger)

Données

Rechercher

DimDuration DimMode DimPersAtt DimPurpose DimTime FactCustProf

Age Category country duration mode nights purpose quarter Sex

spend visits Year

Tableau 8 : Sommation des données quantitatives

### a) Business Query

On ne pourrait parler de visualisation des données sans parler d'un objectif précis à atteindre, ou d'une réponse à trouver ou encore d'une problématique à résoudre. Pour cela, il faudrait choisir judicieusement les types de visuels à insérer dans le tableau de bord. En ce qui concerne notre cas, nous nous sommes basés sur les réponses essentielles que cherchaient les preneurs de décisions dans le cadre du secteur touristique. Nous pouvons examiner la figure suivante.

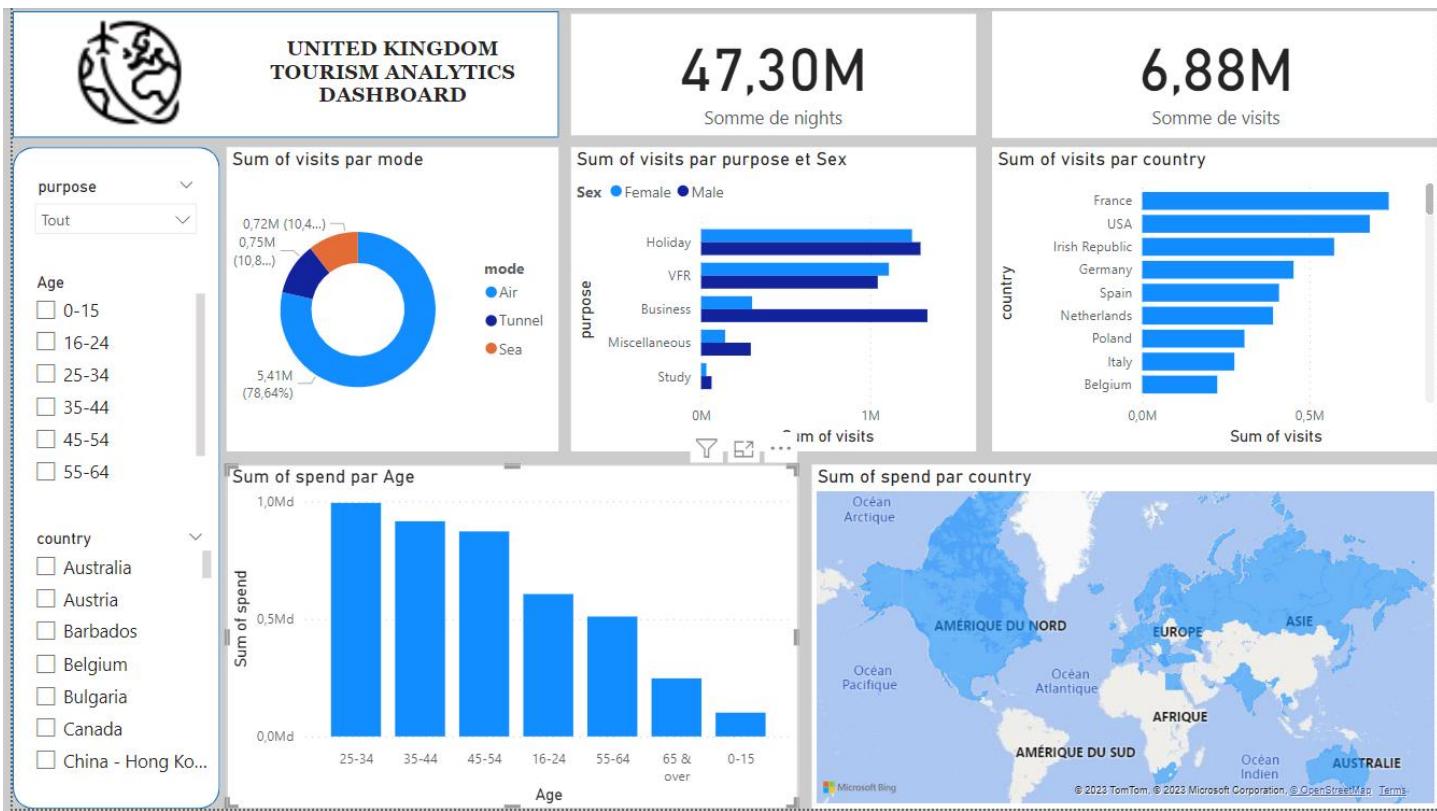


Tableau 9 : Tableau de bord présentant les données

Comme nous pouvons le constater, ce tableau de bord peut aider les preneurs de décisions à évaluer et prendre les mesures nécessaires en vue d'atteindre leurs objectifs. Il leur permet de trouver des réponses aux diverses questions comme par exemple :

### 1. De quel pays proviennent les touristes qui affluent en grand nombre en Grande-Bretagne ?

La visualisation nous indique que les visiteurs français, américains et irlandais sont les plus nombreux en Grande Bretagne. Pour faciliter leur accueil et même augmenter leur nombre, la Grande-Bretagne peut prendre plusieurs mesures. Par exemple, elle peut mettre en place des campagnes de marketing ciblées dans les pays d'origine des visiteurs pour promouvoir les attractions touristiques du pays. Elle peut également simplifier les procédures d'obtention de visa pour les visiteurs en provenance de ces pays. En outre, elle peut améliorer les infrastructures touristiques pour offrir une expérience de voyage plus agréable aux visiteurs. Enfin, elle peut former les professionnels du tourisme pour mieux accueillir les visiteurs étrangers et leur offrir un service de qualité.

## 2. Quelles sont les fréquentes raisons des arrivées des touristes en Grande-Bretagne ?

Sur le tableau, le graphique à barres empilées nous permet de connaitre que les vacances et les visites familiales sont les motifs pour lesquels il y a plus de visiteurs. De ce fait, la Grande-Bretagne peut prendre des initiatives favorables pour améliorer leur expérience. Par exemple, si les visiteurs affluent pendant les vacances, elle peut organiser des événements culturels ou des festivals pour les divertir. Si les visiteurs affluent pour des visites familiales, elle peut offrir des réductions sur les billets d'entrée pour les attractions touristiques familiales. En outre, le pays peut également fournir des informations sur les sites touristiques locaux, les restaurants et les hôtels pour aider les visiteurs à planifier leur voyage.

## 3. Quelles sont les tranches d'âge des touristes qui dépensent le plus ?

Selon l'histogramme, les tranches d'âge pour lesquelles les dépenses sont les plus élevées sont « 25-34 ; 35-44 ; 45-54 ». Alors, la Grande-Bretagne peut prendre plusieurs mesures pour encourager cette tranche d'âge à dépenser davantage. Par exemple, elle peut adapter les offres touristiques pour répondre aux besoins et aux préférences de cette tranche d'âge. Elle peut également organiser des événements qui attirent cette tranche d'âge, tels que des concerts, des festivals, des expositions, etc. En outre, elle peut offrir des réductions sur les activités et les attractions touristiques populaires pour cette tranche d'âge. Enfin, elle peut améliorer les services offerts aux touristes de cette tranche d'âge, tels que les services de transport, les services de restauration, les services d'hébergement, etc. pour leur offrir une expérience de voyage plus agréable et plus personnalisée.

## 4. Quel est le mode de voyage le plus utilisé ?

Le graphique en anneau dans le tableau nous montre que le moyen de prédilection des visiteurs est la voie aérienne. Dans ce cas, le pays peut prendre plusieurs mesures pour faciliter leur voyage. Par exemple, il peut simplifier les procédures d'obtention de visa pour les visiteurs en provenance des pays pour lesquels ils sont plus nombreux. Il peut également améliorer les infrastructures aéroportuaires pour offrir une expérience de voyage plus agréable aux visiteurs. En outre, il peut augmenter la fréquence des vols en provenance des pays avec une plus grande affluence pour faciliter les déplacements des visiteurs. Enfin, il peut offrir des réductions sur les billets d'avion pour les visiteurs en provenance de ces mêmes pays.

## Conclusion

En résumé, cette étude a souligné l'importance cruciale de la conception et de la mise en œuvre d'un entrepôt de données dans le contexte du projet présenté. Dans la première partie, nous avons détaillé les sources de données, les exigences et attentes du projet, ainsi que l'architecture globale de l'entrepôt de données. La modélisation des données à l'aide de datamarts spécifiques tels que les profils clients et l'attraction des clients a ajouté une couche supplémentaire d'efficacité à la gestion de l'information.

La deuxième partie de l'étude s'est concentrée sur la visualisation des données, soulignant l'importance du prétraitement des données pour la représentation visuelle et du formatage des données quantitatives pour répondre aux besoins spécifiques de l'entreprise. L'utilisation judicieuse des processus ETL (extraction/transformation/chargement) était essentielle pour garantir la qualité des données tout au long du cycle de vie du projet. Bref, cette approche holistique de la conception à la visualisation des données a créé un cadre solide pour décortiquer l'information et atteindre efficacement les objectifs définis au début du projet. Cette approche intégrée devrait fournir une base solide pour les futurs efforts d'analyse de données au sein de votre organisation.

## REFERENCES

Office for National Statistics, **Dataset Travelpac: travel to and from the UK**,  
<https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/datasets/travelpac>

VisitBritain, England Visitor Attractions: Archive, <https://www.visitbritain.org/research-insights/england-visitors-attractions-archive>

Abdeltawab M. A. Hendawi, Hisham El-Shishiny, **A Data Warehouse Prototype for the Tourism Industry: A Case Study from Egypt**, Cairo, Egypt.

Tamer Abdulaziz, Ibrahim F. Moawad, Wesal M. Abu-Alam, **Building Data Warehouse System for the Tourism Sector**, Egypt, Decembre 2015

Sindhujan Dhayalan, **Data Warehousing and Business Intelligence Project on Tourism sector' impact on Ireland Economy**, Novembre 2018, <https://fr.slideshare.net/SindhujanDhayalan/data-warehousing-and-business-intelligence-project-on-tourism-sectors-impact-on-ireland-economy>

Siddharth Sharma, M.Sc. in Data Analytics, National College of Ireland, **World Tourism Data Warehouse, Data Warehousing and Business Intelligence Project**, <https://github.com/sunny-NCI/Data-Warehouse>

Haciane, Gaya; Lu, Chuan Chieh; Lerdphayakkarat, Rassaniya; and Mitra, Rudraxy, "Data Warehousing Class Project Report" (2018). **Engineering and Technology Management Student Projects**. 1944.,  
[https://pdxscholar.library.pdx.edu/etm\\_studentprojects/1944](https://pdxscholar.library.pdx.edu/etm_studentprojects/1944)

## ANNEXES

Ci-dessous se trouve le code Python pour effectuer quelques transformations sur le dataset 1.

```
[ ] import pandas as pd  
  
[ ] data_2020=pd.read_excel('/content/TravelPac 2020 Q1.xls')  
  
[ ] data_2021=pd.read_excel('/content/Travelpac 2021.xlsx')  
  
[ ] data_2022=pd.read_excel('/content/Travelpac 2022.xlsx')  
  
[ ] data_2023=pd.read_excel("/content/TravelpacQ1_2023.xlsx")  
  
[ ] data_concat= pd.concat([data_2020,data_2021, data_2022,data_2023])
```

### COMBINER LES TABLES

▶ data\_concat

### SUPPRESSION DES COLONNES SAMPLE ET PACKAGE

```
[ ] data_concat.drop(columns=['package', 'sample'], inplace=True)
```

### SUPPRESSION DES COLONNES SAMPLE ET PACKAGE

```
[ ] data_concat.drop(columns=['package', 'sample'], inplace=True)
```

### SUPPRESSION DES LIGNES CONTENANT LES INFORMATIONS PORTANT SUR LES UK RESIDENTS

```
[ ] data_concat = data_concat[data_concat["ukos"].str.contains("UK residents") == False]
```

```
[ ] data_concat.to_csv('dataconcat.csv', index=False)
```