

BUSINESS INTELLIGENCE ET ENTREPOT DE DONNEES

La vie en ville n'est pas la tasse de thé de tout le monde. Les villes les plus peuplées du monde, en particulier, peuvent être agitées, bruyantes et occupées. Pourtant, malgré le chaos de la vie urbaine, les villes offrent aux habitants un certain nombre de confort et de commodités qui sont plus difficiles à trouver dans les petites villes. C'est pourquoi plus de gens s'installent dans les régions urbaines du monde. Mais ces commodités reflètent-elles la qualité de vie des gens ?

Selon notre étude nous allons découvrir que plus c'est gros, moins c'est bon, du moins en ce qui concerne la taille de la population.

Anas ELKACEMI

Elève ingénieur des données et connaissance,

ESI, Rabat, Maroc

anas.elkacemi@esi.ac.ma

Salma BENZAKOUR

Elève ingénieur des données et connaissance,

ESI, Rabat, Maroc

salma.benzakour@esi.ac.ma

La taille de la population urbaine influence-t-elle sur la qualité de vie ?

The BIGGER the BETTER ?

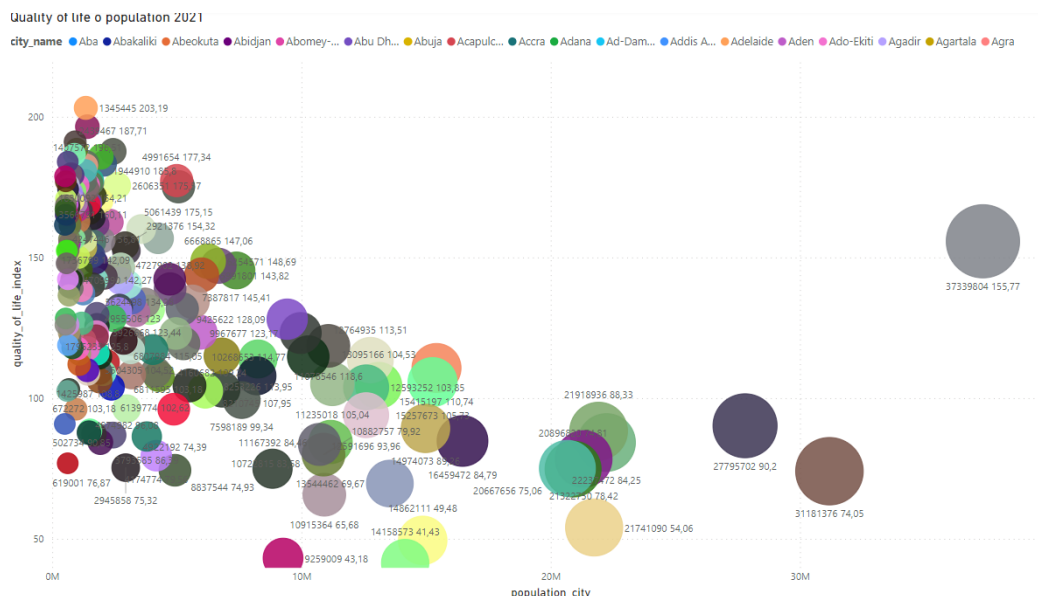


Table des matières

Introduction	2
KPIs.....	3
Conception:.....	4
Modélisation :	5
Création du Datawarehouse sur SQLServer.....	7
Analyse et Reporting.....	9
Prédiction	13

Introduction

Cette étude porte sur les aspects de la qualité de vie urbaine dans les villes du monde entier. A cette fin, L'enquête fournit les données de 251 villes, ce qui permet d'analyser l'interrelation entre les caractéristiques des citoyens, les contextes des quartiers et des villes et la satisfaction de la vie dans une ville. L'étude a analysé les dimensions suivantes potentiellement liées à la satisfaction de la vie dans une ville : L'indice de qualité de vie, l'indice du pouvoir d'achat, l'indice de pollution, rapport entre le prix des logements et le revenu, l'indice du coût de la vie, l'indice de sécurité, l'indice des soins de santé, l'indice du temps de trajet en voiture et l'indice climatique.

Lorsque les citoyens se sentent en sécurité et satisfaits de leur lieu de vie, ils sont également plus susceptibles d'être satisfaits de la vie dans une ville. Enfin, les villes où le pourcentage de personnes satisfaites de la sécurité dans une ville est élevé tendent à être celles où les citoyens sont également plus satisfaits de la vie dans une ville.

Description source de données :

Rapport sur la qualité de vie dans les villes du monde entier - Fournit 9 ensembles de données où chacun représente une année à partir de 2013 jusqu'à 2021, il contient des données relatives à la qualité de vie à travers les villes et la contribution de certains facteurs tels que le pouvoir d'achat et la sécurité sur le score fourni.

Quality of Life Index by City 2021:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2021-mid
Quality of Life Index by City 2020:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2020-mid
Quality of Life Index by City 2019:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2019-mid
Quality of Life Index by City 2018:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2018-mid
Quality of Life Index by City 2017:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2017-mid
Quality of Life Index by City 2016:	https://www.numbeo.com/quality-of-life/rankings.jsp?title=2016-mid

Nettoyage des données :

```
let
Source = Web.Page(Web.Contents("https://www.numbeo.com/quality-of-life/rankings.jsp?title=2020")),
Data0 = Source{0}[Data],
#"Type modifié" = Table.TransformColumnTypes(Data0,{{"Rank", type text}, {"City", type text}, {"Quality of Life Index", type text}, {"Pu
#"Colonnes supprimées" = Table.RemoveColumns(#"Type modifié",{"Rank"}),
#"Colonne personnalisée ajoutée" = Table.AddColumn(#"Colonnes supprimées", "Country", each let splitCity = List.Reverse(Splitter.SplitTe
#"Colonnes permutées" = Table.ReorderColumns(#"Colonne personnalisée ajoutée",{"City", "Country", "Quality of Life Index", "Purchasing P
#"Colonnes renommées" = Table.RenameColumns(#"Colonnes permutées",{{"City", "City and country"}}),
#"Texte inséré avant le délimiteur" = Table.AddColumn(#"Colonnes renommées", "City", each Text.BeforeDelimiter([City and country], ","),
#"Colonnes permutées1" = Table.ReorderColumns(#"Texte inséré avant le délimiteur",{"City and country", "City", "Country", "Quality of Li
in
#"Colonnes permutées1"
```

KPIs

Description des indicateurs clés de performance (KPI)

On a utilisé les données de l’indice de qualité de vie 2021 de Numbeo pour comparer la qualité de vie dans près de 200 villes du monde. Pour les besoins de cette recherche, on a utilisé des villes comptant plus de 500 000 habitants dans les régions métropolitaines.

L’indice mesure la qualité de vie à l’aide de huit paramètres différents :

- Coût de la vie
- Pouvoir d’achat
- Ratio prix-revenu de la propriété
- Pollution
- Temps de déplacement
- Sécurité
- Soins de santé
- Climat

La majorité des paramètres (six sur huit) semblaient correspondre à la taille de la population, ce qui donne à penser que plus la population d’une ville est grande, moins la qualité de vie est élevée.

Voici un regard sur la liste des villes incluses dans l’étude, ainsi que leurs scores globaux de qualité de vie et leurs populations de la région métropolitaine :

city_name	country_name	quality_of_life_index	population_city
Tokyo	Japan	155,77	37339804
Delhi	India	74,05	31181376
Shanghai	China	90,20	27795702
Sao Paulo	Brazil	84,25	22237472
Mexico City	Mexico	88,33	21918936
Dhaka	Bangladesh	54,06	21741090
Cairo	Egypt	78,42	21322750
Beijing	China	74,81	20896820
Mumbai	India	75,06	20667656
Karachi	Pakistan	84,79	16459472
Istanbul	Turkey	110,74	15415197
Buenos Aires	Argentina	105,73	15257673
Kolkata	India	89,26	14974073
Lagos	Nigeria	49,48	14862111
Manila	Philippines	41,43	14158573
Rio de Janeiro	Brazil	69,67	13544462
Lahore	Pakistan	104,53	13095166
Bangalore	India	113,51	12764935
Moscow	Russia	103,85	12593252

Indicateur	Description
I1	La qualité de vie par pays dans le monde entier dans une année précise
I2	La qualité de vie des villes par population dans le monde entier dans une année précise
I3	La qualité de vie pour les 10 villes les plus populationnels dans une année précise
I4	Les différents indicateurs en rapport avec la qualité de vie dans une ville précise
I5	Comparaison des indicateurs de qualité de vie à travers les années d'une ville précise

Formule pour calculer l'indice de qualité de vie à travers les autres indices :

$$\text{index. main} = \text{Math.max}(0, 100 + \text{purchasingPowerInclRentIndex} / 2.5 - (\text{housePriceToIncomeRatio} * 1.0) - \text{costOfLivingIndex} / 10 + \text{safetyIndex} / 2.0 + \text{healthIndex} / 2.5 - \text{trafficTimeIndex} / 2.0 - \text{pollutionIndex} * 2.0 / 3.0 + \text{climateIndex} / 3.0);$$

Conception:

1- Dimension "dim_institution_of_index"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>ID Institution</u>	INT	NOT NULL	IDENTITY(1,1)
<u>Institution Name</u>	NVARCHAR(150)		

2- Dimension "Dim_Geo"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>ID Country</u>	INT	NOT NULL	
<u>CountryName</u>	NVARCHAR(50)		
GDP	NVARCHAR(150)		
Continent	NVARCHAR(150)		
<u>ID Date</u>	NVARCHAR(150)		

3- Dimension "Dim_City"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>ID City</u>	INT	NOT NULL	IDENTITY(1,1)
<u>CityName</u>	NVARCHAR(50)	NOT NULL	
<u>ID Date</u>	<u>int</u>	NOT NULL	

4- Dimension "Dim_City_Indicators"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>ID_City_Indicator</u>	INT	NOT NULL	
Population	NVARCHAR(150)		
<u>DateKey</u>	NVARCHAR(150)		
<u>geographyKey</u>	NVARCHAR(150)		

5- Dimension "Dim_Date"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>DateKey</u>	INT	NOT NULL	
<u>Year</u>	NVARCHAR(150)	<u>int</u>	

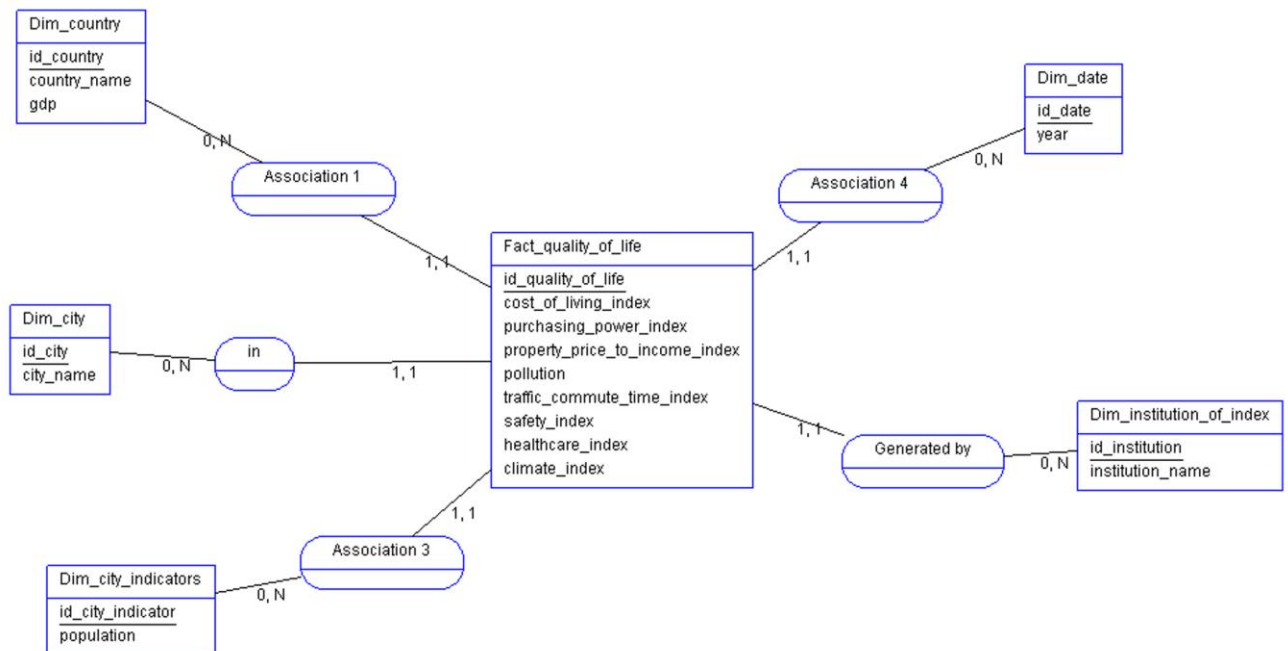
6- Dimension "Fact_Quality_of_life"

Nom de la colonne	Type de donnée	<u>Nullability</u>	Remarques
<u>ID_QualityOfLife</u>	INT	NOT NULL	
<u>QualityOfLifeIndex</u>	<u>Float</u>		Default 0
<u>CostOfLiving</u>	<u>Float</u>		Default 0
<u>PurchasingPower</u>	<u>Float</u>		Default 0
<u>PropertyPriceToIncome Ratio</u>	<u>Float</u>		Default 0
Pollution	<u>Float</u>		Default 0
<u>TrafficCommuteTime</u>	<u>Float</u>		Default 0
<u>Safety</u>	<u>Float</u>		Default 0
Healthcare	<u>Float</u>		Default 0
<u>Climate</u>	<u>Float</u>		Default 0

Modélisation :

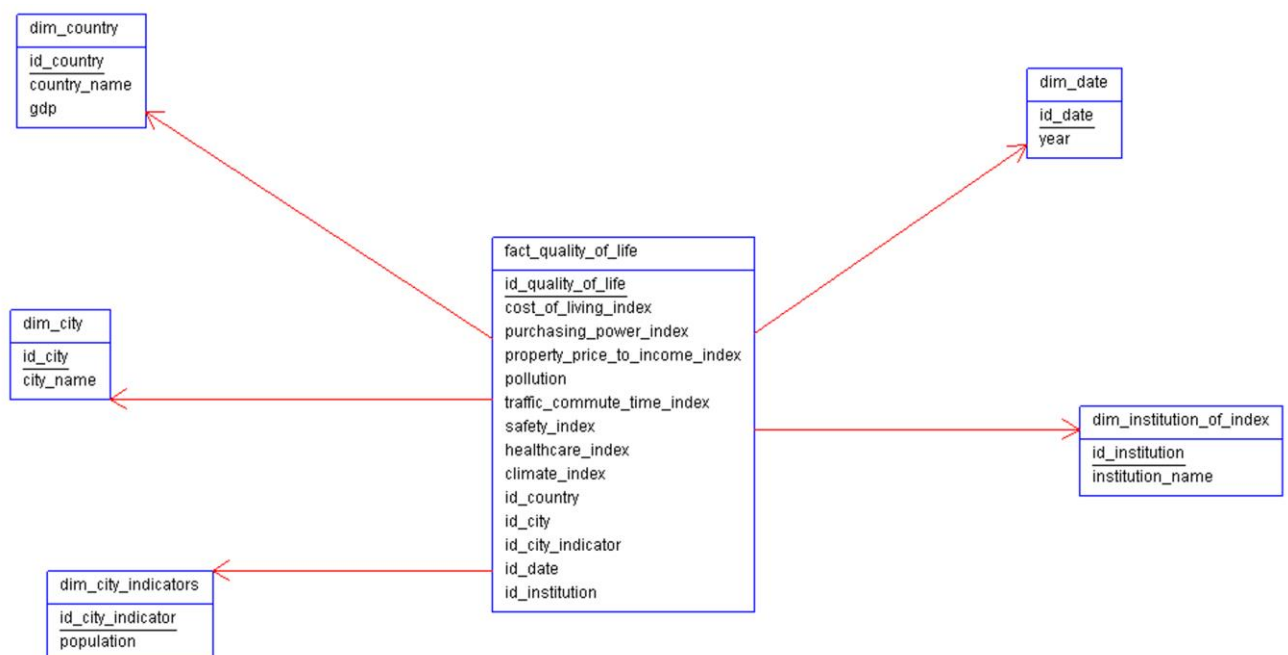
Le modèle de données conceptuel MCD

Le modèle conceptuel ne traite pas de la manière dont la base de données réelle sera mise en œuvre. Il fonctionne à un niveau très élevé, traduisant les problèmes du monde réel et les exigences commerciales en un cadre conceptuel facile à comprendre et qui, pour cette raison, peut également être communiqué aux non-experts. Il se concentre simplement sur l'identification des entités qui feront partie de la base de données et les relations entre elles, c'est-à-dire les opérations ou les associations qui existent entre elles. Il est très efficace pour comprendre et définir les processus métiers



Le modèle de données logique MLD

Dès lors, la différence entre le modèle de données conceptuel et logique est claire. Le modèle de données logique est une représentation abstraite d'une implémentation possible, sans être lié à une implémentation spécifique, tandis que le modèle de données conceptuel est une représentation de haut niveau des exigences métier et des ensembles de données et relations connectés. Le modèle logique – à travers un processus de normalisation – donne une représentation plus structurée des entités et de leurs relations et détaille leurs relations et caractéristiques : les redondances, les relations plusieurs-à-plusieurs, les ambiguïtés et les incertitudes d'attribution des entités sont résolues et un schéma directeur d'une éventuelle implémentation de base de données émerge évidemment.



Création du Datawarehouse sur SQLServer

Dans cette partie du projet, nous mettrons en œuvre l'entrepôt de données sur SSMS sous le nom qualityOfLife2, nous utiliserons le modèle en étoile

Table quality_of_life.

```
CREATE TABLE fact_quality_of_life
(
    id_quality_of_life INT NOT NULL IDENTITY(1,1),
    id_country INT NOT NULL,
    id_city INT NOT NULL,
    id_city_indicator INT NOT NULL,
    id_date INT NOT NULL,
    id_institution INT NOT NULL,
    cost_of_living_index float NOT NULL,
    purchasing_power_index float NOT NULL,
    property_price_to_income_index float NOT NULL,
    pollution float NOT NULL,
    traffic_commute_time_index float NOT NULL,
    safety_index float NOT NULL,
    healthcare_index float NOT NULL,
    climate_index float NOT NULL,
    CONSTRAINT PK_fact_quality_of_life PRIMARY KEY (id_quality_of_life)
);
```

Table Country.

```
CREATE TABLE dim_country
(
    id_country INT NOT NULL IDENTITY(1,1),
    country_name NVARCHAR(50) NULL,
    CONSTRAINT PK_dim_country PRIMARY KEY (id_country)
);
```

Table institution_of_index.

```
CREATE TABLE dim_institution_of_index
(
    id_institution INT NOT NULL IDENTITY(1,1),
    institution_name NVARCHAR(150) NULL,
    CONSTRAINT PK_dim_institution_of_index PRIMARY KEY (id_institution)
);
```

Table dim_city.

```
CREATE TABLE dim_city
(
    id_city INT NOT NULL IDENTITY(1,1),
    city_name NVARCHAR(50) NULL,
    CONSTRAINT PK_dim_city PRIMARY KEY (id_city)
);
```

Table dim_city_indicators.

```
CREATE TABLE dim_city_indicators
(
    id_city_indicator INT NOT NULL IDENTITY(1,1),
    city_name NVARCHAR(50) NULL,
    population_city BIGINT,
    CONSTRAINT PK_dim_city_indicators PRIMARY KEY (id_city_indicator)
);
```

Table dim_date.

```
CREATE TABLE dim_date
(
    id_date INT NOT NULL IDENTITY(1,1),
    year INT NULL,
    CONSTRAINT PK_dim_date PRIMARY KEY (id_date)
);
```

Ajout des clés étrangères.

```
ALTER TABLE fact_quality_of_life ADD CONSTRAINT
FK_fact_quality_of_life_dim_city FOREIGN KEY(id_city)
REFERENCES dim_city (id_city);

ALTER TABLE fact_quality_of_life ADD CONSTRAINT
FK_fact_quality_of_life_dim_date FOREIGN KEY(id_date)
REFERENCES dim_date (id_date);

ALTER TABLE fact_quality_of_life ADD CONSTRAINT
FK_fact_quality_of_life_dim_city_indicators FOREIGN KEY(id_city_indicator)
REFERENCES dim_city_indicators (id_city_indicator);

ALTER TABLE fact_quality_of_life ADD CONSTRAINT
FK_fact_quality_of_life_dim_institution_of_index FOREIGN KEY(id_institution)
REFERENCES dim_institution_of_index (id_institution);

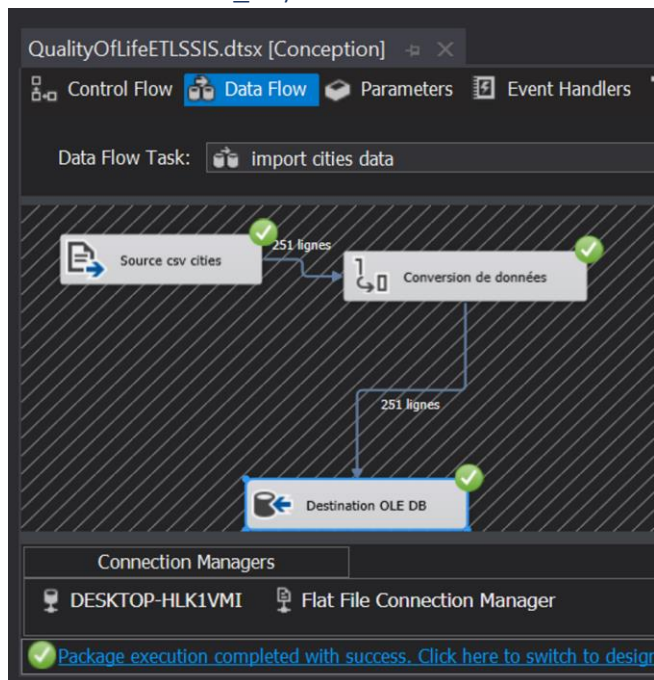
ALTER TABLE fact_quality_of_life ADD CONSTRAINT
FK_fact_quality_of_life_dim_country FOREIGN KEY(id_country)
REFERENCES dim_country (id_country);
```


Pipeline ETL

Population des dimensions

Ensuite on transforme les données issues des fichiers plats, et on les charge dans les tables de dimensions.

Alimentation Dim_city



SQLQuery1.sql - DE...HLK1VMI\user (64))

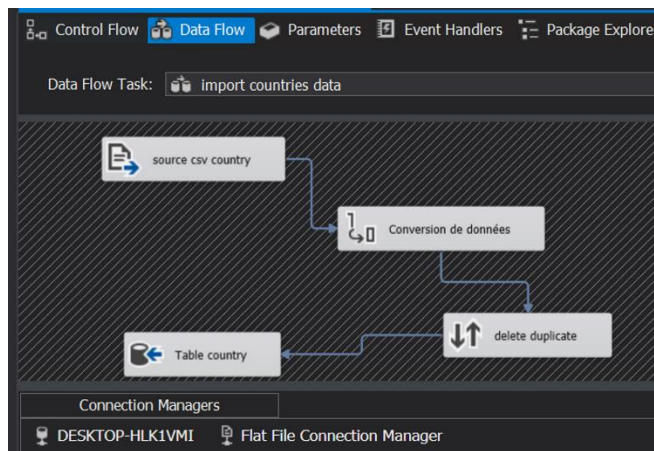
```
/****** Script de la commande SelectTopNRows *****  
SELECT TOP (1000) [id_city]  
,[city_name]  
FROM [QualityOfLife2].[dbo].[dim_city]
```

Résultats

	id_city	city_name
1	3010	Adelaide
2	3011	Canberra
3	3012	Wellington
4	3013	Raleigh
5	3014	Zurich
6	3015	The Hague (Den Haag)
7	3016	Madison
8	3017	Columbus
9	3018	Austin
10	3019	Brisbane
11	3020	Eindhoven
12	3021	Victoria
13	3022	Charlotte
14	3023	Vienna
15	3024	Geneva
16	3025	Richmond
17	3026	San Diego
18	3027	Perth

Exécution de requête réussie. DESKTOP-HLI

Alimentation Dim_country

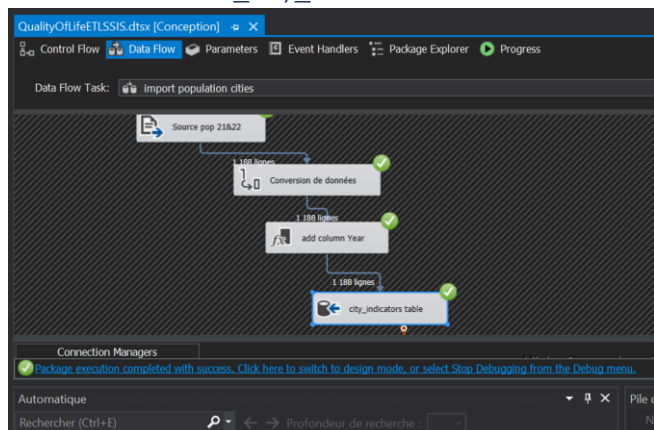


Résultats

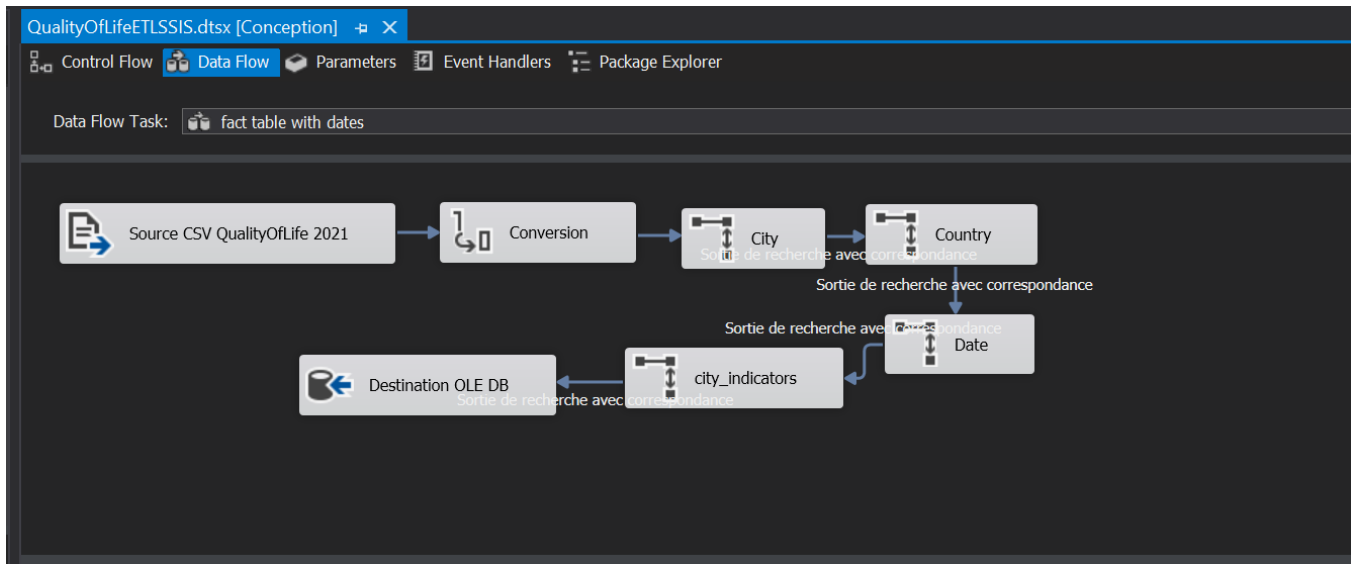
	id_country	country_name
1	972	Albania
2	973	Argentina
3	974	Armenia
4	975	Australia
5	976	Austria
6	977	Azerbaijan
7	978	Bahrain
8	979	Bangladesh
9	980	Belarus
10	981	Belgium
11	982	Bosnia And Herzegovina
12	983	Brazil
13	984	Bulgaria
14	985	Canada
15	986	Chile
16	987	China
17	988	Colombia
18	989	Costa Rica

Exécution de requête réussie.

Alimentation Dim_city_indicators



Alimentation Fact table



SQLQuery5.sql - DE...HLK1VM\user (69) SQLQuery4.sql - DE...HLK1VM\user (68) SQLQuery3.sql - DE...HLK1VM\user (65)

```

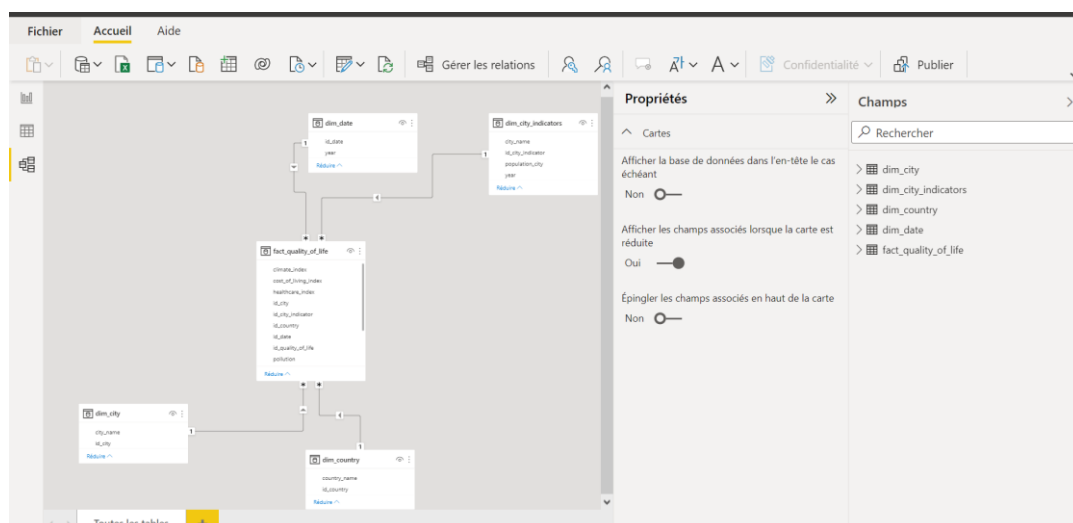
/***** Script de la commande SelectTopNRows à partir de SSMS *****/
SELECT TOP (1000) [id_quality_of_life]
, [id_country]
, [id_city]
, [id_city_indicator]
, [id_date]
, [cost_of_living_index]
, [purchasing_power_index]
, [property_price_to_income_index]
, [pollution]
, [traffic]

```

	id_quality_of_life	id_country	id_city	id_city_indicator	id_date	cost_of_living_index	purchasing_power_index	property_price_to_income_index	pollution	traffic
1	27612	975	3010	65444	18	80,4	100,23	4,51	18,11	24,02
2	27613	1051	3014	74844	18	135,49	117,91	8,01	17,31	33,78
3	27614	1058	3017	67284	18	73,13	122,05	2,45	25,32	25,56
4	27615	1058	3018	65940	18	63,04	129,63	3,59	36,27	34,12
5	27616	975	3019	66628	18	84,71	96,93	5,95	21,88	40,03
6	27617	1058	3022	67052	18	72,66	120	3,01	28,44	35,18
7	27618	976	3023	73908	18	71,31	87,36	13,12	17,76	26,51
8	27619	1051	3024	68204	18	131,75	109,73	9,45	26,44	25,48
9	27620	1058	3026	72436	18	80,22	112,72	5,24	35,04	33,8
10	27621	975	3027	71700	18	84,64	99,43	6,33	23,25	35,83
11	27622	993	3029	67316	18	97,49	84,96	8,39	21,06	28,61
12	27623	1058	3032	67452	18	64,8	141,09	2,35	41,92	35,22
13	27624	1058	3033	72660	18	89,48	121,2	5,23	28,63	41,57
14	27625	985	3035	72052	18	67,37	92,31	4,56	29,29	27,91

Analyse et Reporting

On a procédé par l'importation des données du datawarehouse dans l'onglet Modèle :



quality_of_life_index

- 41.43
- 43.18
- 49.48
- 54.06
- 55.68
- 59.67
- 74.05
- 74.39
- 74.81
- 74.93
- 75.06
- 75.32
- 76.87
- 78.42
- 79.56
- 79.92
- 83.58
- 84.25
- 84.39
- 84.46
- 84.79
- 86.39
- 96.1

24 8 8

city_name

quality_of_life_index

population_city

city_name Abu Dh... Adelaide Ahmeda... Albuquerque Almaty Amman Amster... Ankara Antwerp Athens Atlanta Auckland Austin Baku Baltimore Bangkok Barynk

1345445 203,19

917620 191,12

97900 188,46

2069067 187,71

63094 174,34

4991654 177,34

5061439 175,15

1317168 169,28

3566791 160,11

1047613 150,13

6668865 147,06

7387817 145,41

5624498 134,28

5215747 131,75

9425622 128,09

12764935 113,51

15415197 110,74

12591696 93,96

14974073 89,26

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

12764935 113,51

15237673 105,73

16459472 84,79

21918936 88,33

22237472 84,25

27795702 90,2

31181376 74,05

21741090 54,06

14862111 49,48

9259009 43,18

10915364 65,68

10722815 83,58

13544462 69,67

11167392 84,46

12591696 93,96

12535018 105,04

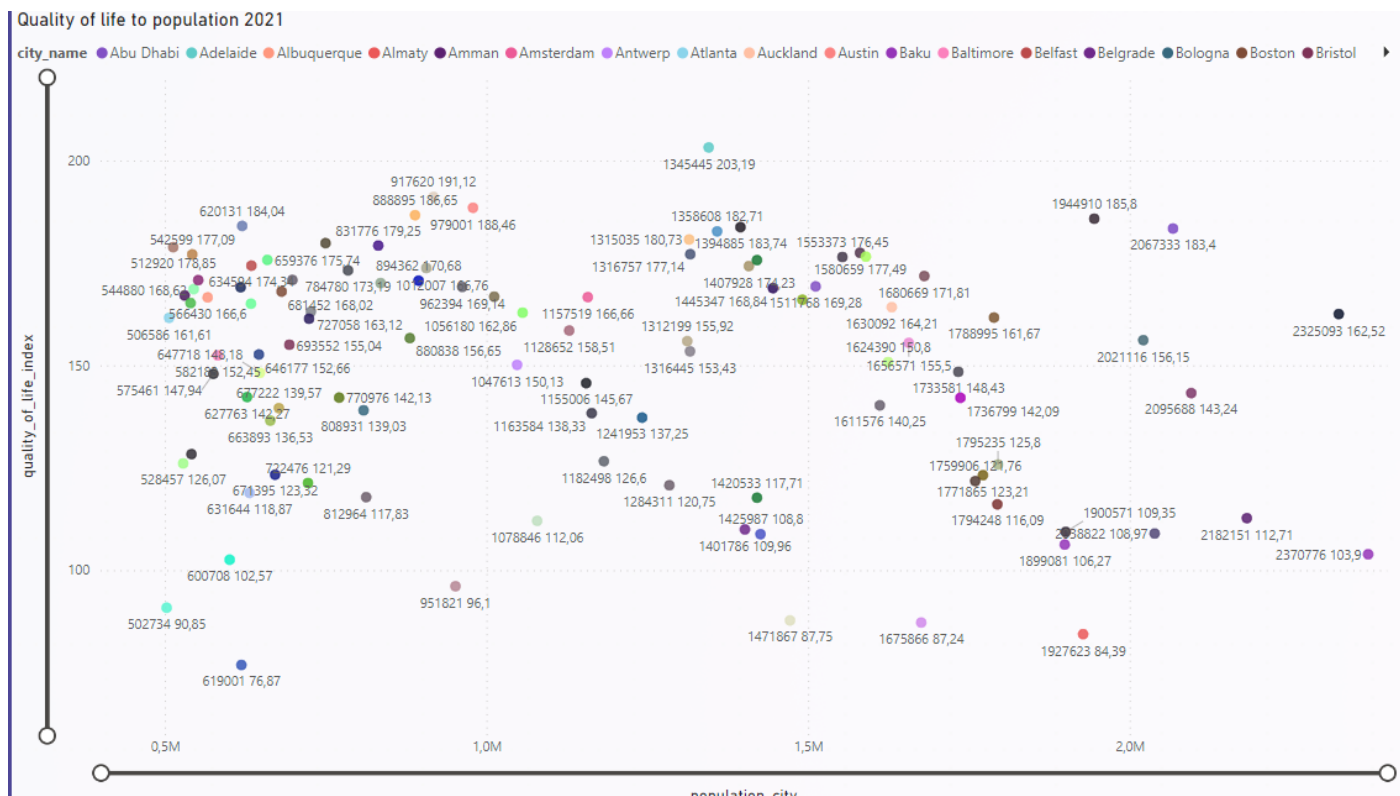
12764935 113,51

15237673 105,73

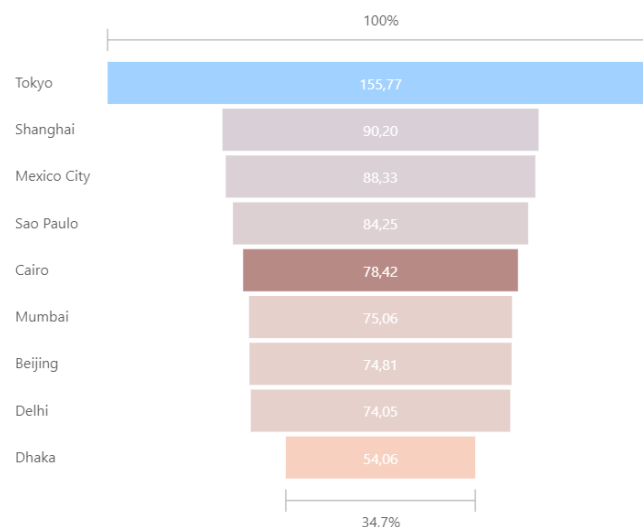
16459472 84,79

21918936 88,33

Cependant, Tokyo est une anomalie : sa qualité de vie est bien supérieure à ce que sa population métropolitaine pourrait prévoir.



La qualité de vie pour les 10 villes avec la plus grande population dans une année précise - 2021 -



Zoom sur Tokyo

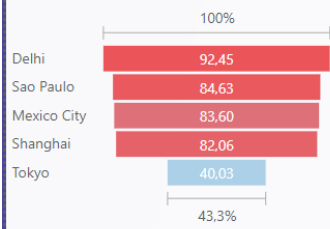
Bien qu'elle soit la plus grande zone urbaine du monde, Tokyo se compare favorablement aux autres grandes mégapoles de l'indice. Classée parmi les cinq principales mégapoles, elle se classe au premier rang pour le pouvoir d'achat, la sécurité, le temps de déplacement, le coût de la vie, les soins de santé et la pollution.

Qu'y a-t-il de si spécial à Tokyo ? Une chose importante pour la ville japonaise est son système de transport public impeccable. Le transport de la ville est si efficace que le service ferroviaire de la ville a déjà présenté des excuses publiques après le départ d'un train 25 secondes plus tôt que prévu.

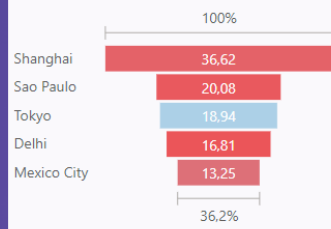
Un autre facteur qui rend Tokyo si vivable est son logement relativement abordable, au moins par rapport à d'autres grandes villes comme New York et Hong Kong. Cela s'explique en partie par la souplesse du système de zonage des terres de la ville, qui permet aux promoteurs de construire relativement facilement des logements et des collectivités polyvalentes.

Zoom in TOP 5 POP

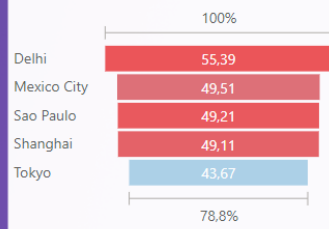
POLLUTION



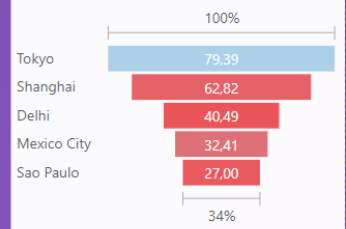
PROPERTY PRICE



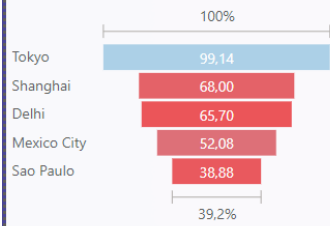
TRAFFIC COMMUTE TIME



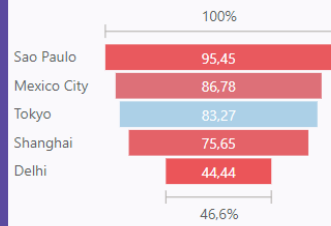
SAFETY



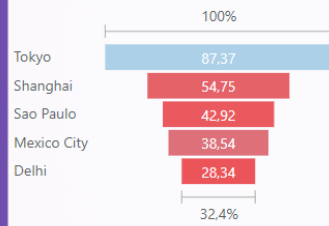
PURCHASING POWER



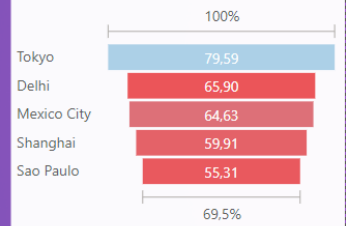
CLIMATE



COST OF LIVING

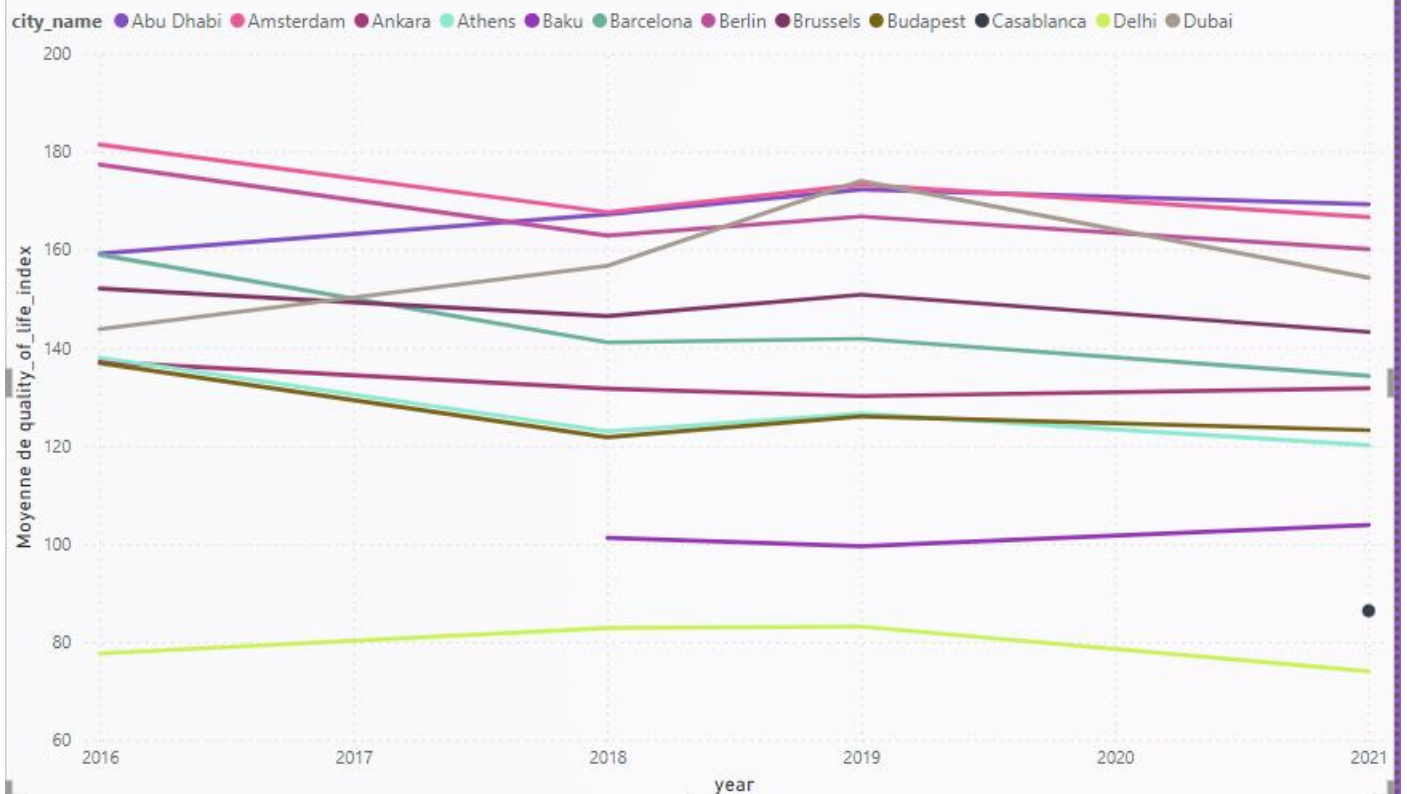


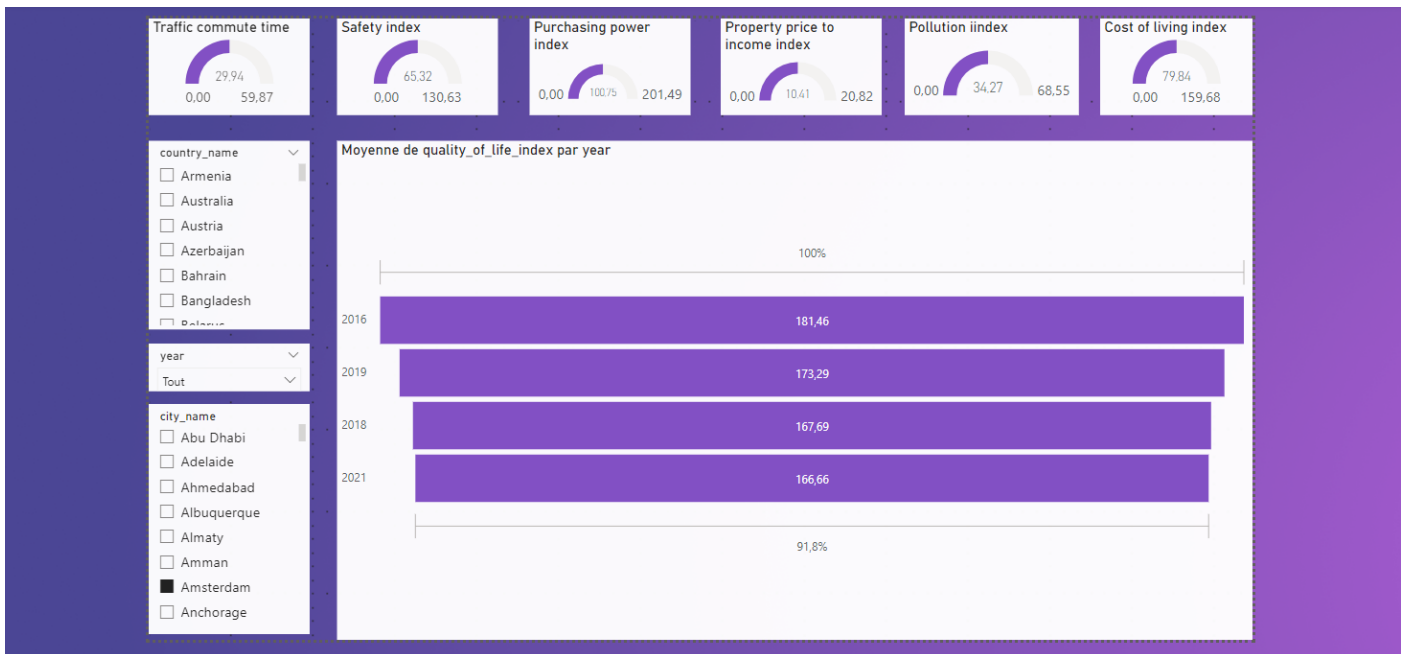
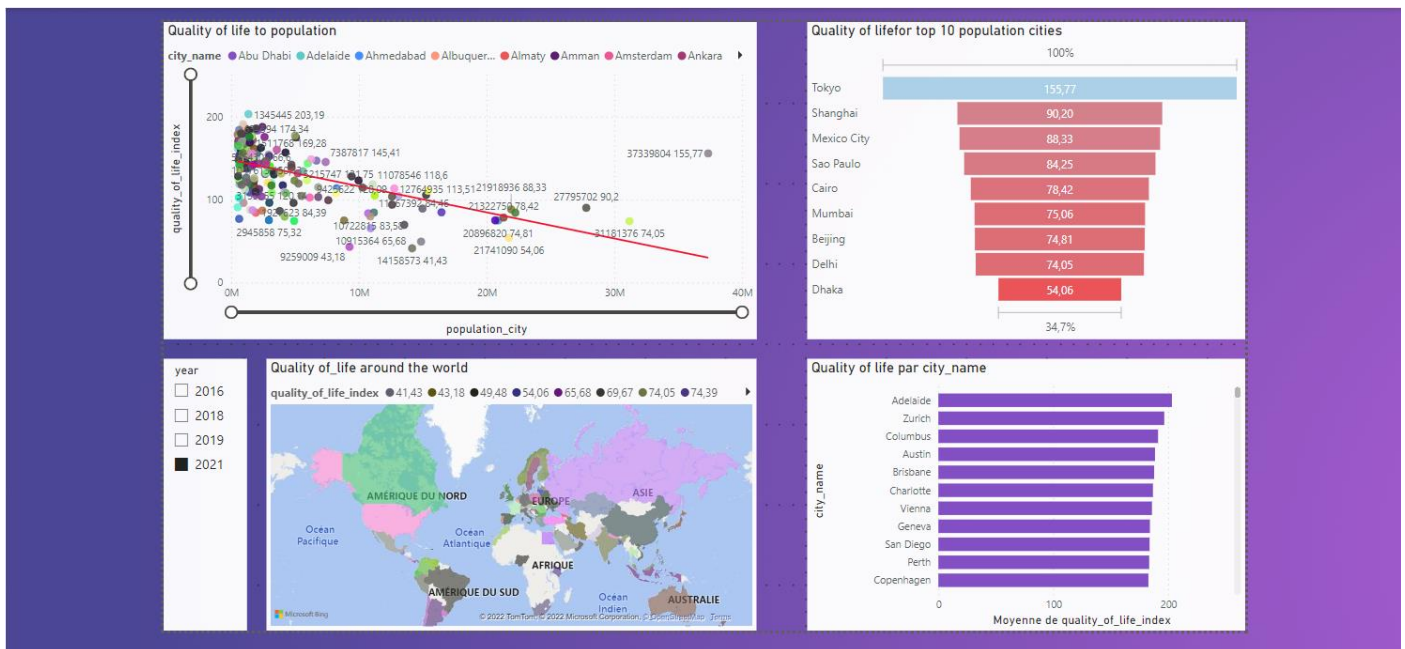
HEALTHCARE



Alors que notre monde s'urbanise de plus en plus et que les villes du monde entier continuent d'augmenter, seront-elles en mesure d'imiter la croissance de Tokyo ? Sinon, quelles autres tendances et innovations en matière de design urbain les villes peuvent-elles utiliser pour améliorer la qualité de vie ?

Moyenne de quality_of_life_index par year et city_name





Prédiction

importer les bibliothèques nécessaires : numpy ,pandas , scikit learn , matplotlib lib

```
Entrée [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

importer la dataset

```
Entrée [11]: data = pd.read_excel('source.xlsx')
x = np.array(data['Cost of Living Index'])
y = np.array(data['Quality of Life Index'])
x = x.reshape(-1,1)
y = y.reshape(-1,1)
```

Implémentation avec scikit-learn

initialisation du modèle

```
Entrée [12]: regression_model = LinearRegression()
```

Adapter les données (entraînement du modèle)

```
Entrée [14]: regression_model.fit(x, y)
```

```
Out[14]: LinearRegression()
```

Prédiction

```
Entrée [16]: y_predicted = regression_model.predict(x)
```

Évaluation du modèle

```
Entrée [17]: rmse = mean_squared_error(y, y_predicted)
             r2 = r2_score(y, y_predicted)
```

Affichage des valeurs

```
Entrée [19]: print("Pente : ", regression_model.coef_)
             print("Ordonnée à l'origine : ", regression_model.intercept_)
             print("Racine carrée de l'erreur quadratique moyenne : ", rmse)
             print('Score R2 : ', r2)
```

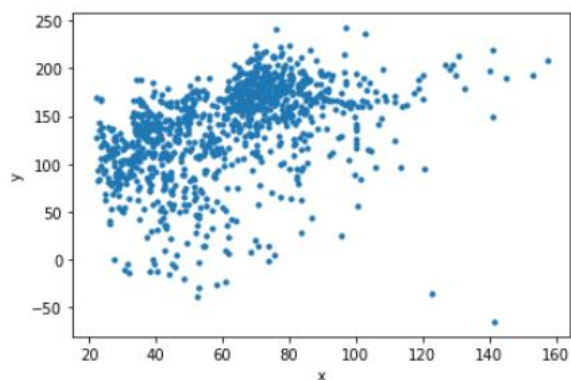
```
Pente : [[0.92863014]]
Ordonnée à l'origine : [78.95874792]
Racine carrée de l'erreur quadratique moyenne : 1986.432080662046
Score R2 : 0.18832638839130544
```

Tracée des valeurs

Points de données

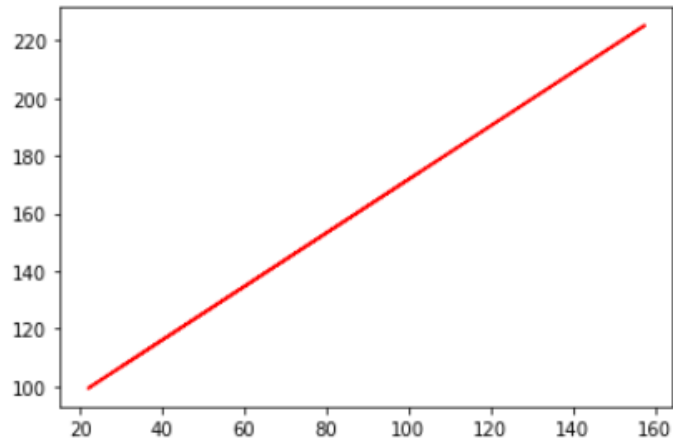
```
Entrée [21]: plt.scatter(x, y, s=10)
             plt.xlabel('x')
             plt.ylabel('y')
```

```
Out[21]: Text(0, 0.5, 'y')
```



Valeurs prédites

```
Entrée [23]: plt.plot(x, y_predicted, color='r')  
plt.show()
```



```
Entrée [24]: plt.scatter(x, y, s=10)  
plt.xlabel('x')  
plt.ylabel('y')  
plt.plot(x, y_predicted, color='r')  
plt.show()
```

