

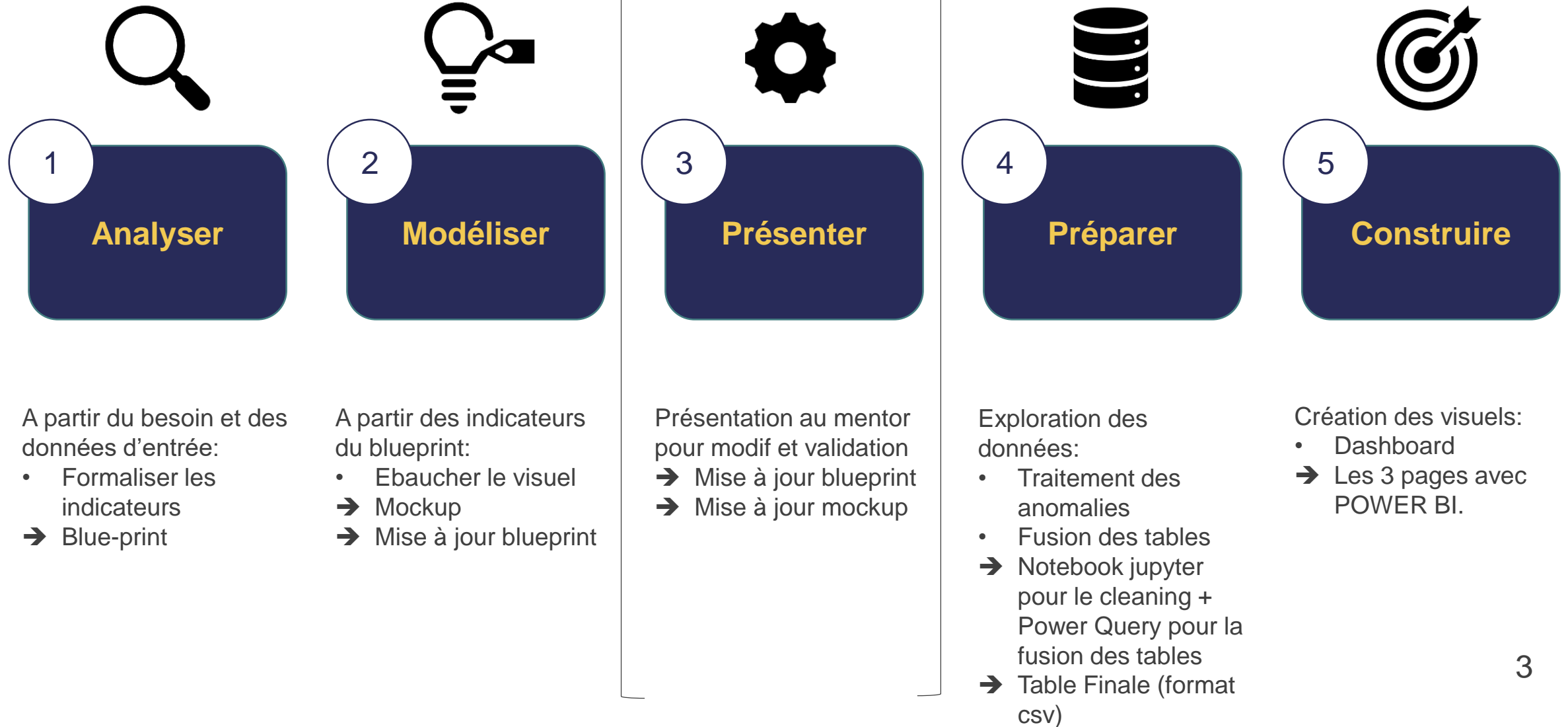
Une étude sur l'eau potable dans le Monde



Déroulé

1. Les éléments du contexte
2. Le pré-traitement des données
3. La pertinence de l'outils de visualisation retenu

Le processus de création du dashboard



Extrait Blueprint

Blueprint

Le tableau ci-dessous reprend les détails essentiels nécessaires pour le tableau de bord.

- **Besoin utilisateurs** : Décrit brièvement les interactions des utilisateurs avec les données pour cette exigence (par exemple, les filtres nécessaires, si une visualisation est fixe ou interactive.).
- **Mesures spécifiques à utiliser** : Il s'agit de la liste des paramètres et de tous les paramètres calculés qui seront utilisés pour cette exigence (par exemple, le coût réel).
- **Visualisations** : Le type de visualisation qui pourrait être utilisé pour cette exigence (par exemple, un diagramme à barres)

Besoin utilisateurs	Mesures spécifiques à utiliser	Visualisation	Page/Onglet/Vue*
État des lieux mortalité dû à de l'eau insalubre	Taux de mortalité dû à l'eau (chps calculé)	CHIFFRES	Vue mondiale
Visualisation fixe	Filtre Granularity = Total		
Evolution pop dans le temps	Filtre Year = 2016		
Etats des lieux sur accessibilité à l'eau	Nb de morts dus à l'eau	CHIFFRES	Vue mondiale
Etats des lieux sur les servies basiques et services de qualités	Filtre Granularity = Total		
Instabilité politique	Filtre Year = 2016		
	Population mondiale en 2018	CHIFFRES	Vue mondiale
	Taux d'accès à l'eau en 2017	CHIFFRES	Vue mondiale
	Stabilité Politique en 2018	CHIFFRES	Vue mondiale

	Nb de morts dus à l'eau		Vue continentale/choix du continent
	Echelle Region		
	Taux d'accès à l'eau	Chiffres	Vue continentale/choix du continent
	Echelle Region	Filtre Continent	
	Population du continent	Lineplot par continent	Vue continentale/choix du continent
	Evolution dans le temps		
	Stabilité potlitique vs Taux d'accès à l'eau	2 Line plot	Vue continentale/choix du continent
	Population rurale et population urbaine	Donut piechart	Vue continent / Choix du continent
	En 2017		
	Echelle continent		
	Evolution Pop urbaine dans le temps	2 Line Plot	Vue continent / Choix du continent
	Evolution Pop rurale dans le temps		
	Tx service basiques vs qualité	Grouped bar plot	Vue continent / Choix du continent
		Valeur pull exclu	

Pré- traitement des données

fx

Table.TransformColumnTypes("#Colonnes permutées",{{"Pourcentage Pp Urbaine", type number}})

A ⁰ Region	A ⁰ Pays	1 ² Année Stabilité	1.2 Stabilité Politique	1 ² Année Population	1 ² Total Pop	1 ² Population Hommes	1 ² Population Femmes	1 ² Population Rural	
1	Eastern Mediterranean	Afghanistan	2016	-2,67	2016	35383032	18186994	17196034	2598
2	Eastern Mediterranean	Afghanistan	2015	-2,57	2015	34413603	17686166	16727437	2536
3	Eastern Mediterranean	Afghanistan	2005	-2,07	2005	25654277	13239684	12414590	1937
4	Eastern Mediterranean	Afghanistan	2012	-2,42	2012	31161376	15964874	15196504	2328
5	Eastern Mediterranean	Afghanistan	2000	-2,44	2000	20779953	10689508	10090449	1565
6	Eastern Mediterranean	Afghanistan	2017	-2,8	2017	36296113	18651985	17644126	2655
7	Eastern Mediterranean	Afghanistan	2014	-2,41	2014	33370794	17138803	16232001	2470
8	Eastern Mediterranean	Afghanistan	2013	-2,52	2013	32269589	16554277	15715314	2399
9	Eastern Mediterranean	Afghanistan	2003	-2,2	2003	23680871	12214634	11466237	1790
10	Eastern Mediterranean	Afghanistan	2006	-2,22	2006	26433049	13623896	12809162	1996
11	Eastern Mediterranean	Afghanistan	2002	-2,04	2002	22600770	11642106	10958668	1708
12	Eastern Mediterranean	Afghanistan	2010	-2,58	2010	29185507	14935756	14249755	2196
13	Eastern Mediterranean	Afghanistan	2007	-2,41	2007	27100536	13938970	13161572	2046
14	Eastern Mediterranean	Afghanistan	2009	-2,71	2009	28394813	14544055	13850751	2141
15	Eastern Mediterranean	Afghanistan	2008	-2,69	2008	27722276	14225725	13496556	2092
16	Eastern Mediterranean	Afghanistan	2011	-2,5	2011	30117413	15415361	14702050	2259
17	Eastern Mediterranean	Afghanistan	2004	-2,3	2004	24726684	12763726	11962963	1869
18	Europe	Albania	2004	-0,43	2004	3104892	1563966	1540927	168
19	Europe	Albania	2015	0,35	2015	2890513	1471456	1419068	124
20	Europe	Albania	2008	-0,03	2008	3002678	1508919	1493764	149
21	Europe	Albania	2007	-0,2	2007	3033998	1523893	1510100	154
22	Europe	Albania	2002	-0,29	2002	3126187	1582936	1543247	176
23	Europe	Albania	2011	-0,28	2011	2928592	1478059	1450542	136
24	Europe	Albania	2014	0,49	2014	2896305	1472095	1424212	127
25	Europe	Albania	2012	-0,14	2012	2914096	1474211	1439880	133
26	Europe	Albania	2016	0,34	2016	2886438	1470548	1415879	121
27	Europe	Albania	2013	0,09	2013	2903790	1472670	1431118	130
28	Europe	Albania	2006	-0,51	2006	3063021	1538721	1524300	159
29	Europe	Albania	2017	0,38	2017	2884169	1469588	1414581	119
30	Europe	Albania	2003	-0,31	2003	3118023	1574841	1543176	172
31	Europe	Albania	2005	-0,51	2005	3086810	1551975	1534835	164
32	Europe	Albania	2009	-0,05	2009	2973048	1495558	1477486	144
33	Europe	Albania	2010	-0,19	2010	2948023	1485068	1462961	140
34	Europe	Albania	2000	-0,54	2000	3129243	1583506	1545740	181
35	Africa	Algeria	2012	-1,33	2012	37383895	18871563	18512336	1167
36	Africa	Algeria	2004	-1,36	2004	32692163	16542652	16149500	1212
37	Africa	Algeria	2017	-0,92	2017	41389189	20906756	20482418	1154
38	Africa	Algeria	2000	-1,43	2000	31042235	15731099	15311139	1249

Paramètres d'une requête

PROPRIÉTÉS

Nom

Table finale

Toutes les propriétés

ÉTAPES APPLIQUÉES

Source

stability développé

Requêtes fusionnées

population développé

Lignes filtrées

Colonne dynamique

Lignes triées

Requêtes fusionnées1

mortality développé

Lignes filtrées1

Colonnes renommées

Lignes filtrées2

Colonnes supprimées

Colonnes renommées1

Requêtes fusionnées2

water développé

Colonnes renommées2

Valeur remplacée

Valeur remplacée1

Colonnes renommées3

Lignes filtrées3

Colonne dynamique1

Lignes triées1

Colonnes supprimées1

Lignes filtrées4

Lignes filtrées5

Colonnes renommées4

Colonnes supprimées2

Requêtes fusionnées3

mortality développé1

Lignes filtrées6

Colonnes supprimées3

Colonnes renommées5

Personnalisée ajoutée

Pré- traitement des données

1. Population

```
In [48]: population = pd.read_csv('Population.csv')
population.head()

Out[48]:
```

	Country	Granularity	Year	Population
0	Afghanistan	Total	2000	20779.953
1	Afghanistan	Male	2000	10689.508
2	Afghanistan	Female	2000	10090.449
3	Afghanistan	Rural	2000	15657.474
4	Afghanistan	Urban	2000	4436.282

```
In [49]: population['Population'] = population['Population']*1000
population['Population'] = population['Population'].astype("int64")
# changement d'unité et de type de la variable unité

Out[49]:
```

	Country	Granularity	Year	Population
0	Afghanistan	Total	2000	20779953
1	Afghanistan	Male	2000	10689508
2	Afghanistan	Female	2000	10090449
3	Afghanistan	Rural	2000	15657474
4	Afghanistan	Urban	2000	4436282

```
In [50]: population.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20914 entries, 0 to 20913
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Country      20914 non-null   object
1   Granularity   20914 non-null   object
2   Year          20914 non-null   int64
3   Population    20914 non-null   int64
dtypes: int64(2), object(2)
memory usage: 653.7+ KB

In [51]: population.isna().sum()

Out[51]: Country      0
Granularity    0
```

5. Mortality

```
In [91]: mortality = pd.read_csv('MortalityRateAttributedToWater.csv')
mortality.head()

Out[91]:
```

	Year	Country	Granularity	Mortality rate attributed to exposure to unsafe WASH services	WASH deaths
0	2016	Afghanistan	Female	15.31193	NaN
1	2016	Afghanistan	Male	12.61297	NaN
2	2016	Afghanistan	Total	13.92067	4824.353
3	2016	Albania	Female	0.12552	NaN
4	2016	Albania	Male	0.20650	NaN

```
In [92]: mortality.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 549 entries, 0 to 548
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Year        549 non-null    int64
1   Country     549 non-null    object
2   Granularity 549 non-null    object
3   Mortality rate attributed to exposure to unsafe WASH services 549 non-null    float64
4   WASH deaths 183 non-null    float64
dtypes: float64(2), int64(1), object(2)
memory usage: 21.6+ KB

In [93]: mortality.isna().sum()
# 366 valeurs manquantes sur le nombre de décès

Out[93]: Year      0
Country    0
Granularity 0
Mortality rate attributed to exposure to unsafe WASH services 0
```

2. Region

```
In [60]: region = pd.read_csv('RegionCountry.csv')
region.head()

Out[60]:
```

	REGION (DISPLAY)	COUNTRY (DISPLAY)
0	Europe	Albania
1	Europe	Andorra
2	Europe	Armenia
3	Western Pacific	Australia
4	Europe	Austria

```
In [61]: region.rename(columns={"REGION (DISPLAY)": "Region", "COUNTRY (DISPLAY)": "Country"}, inplace = True)
region.head()

Out[61]:
```

	Region	Country
0	Europe	Albania
1	Europe	Andorra
2	Europe	Armenia
3	Western Pacific	Australia
4	Europe	Austria

```
In [62]: region.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194 entries, 0 to 193
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Region      194 non-null    object
1   Country     194 non-null    object
dtypes: object(2)
memory usage: 3.2+ KB

In [63]: region.isna().sum()

Out[63]: Region      0
Country    0
dtype: int64

In [64]: region.nunique()
```

6. Water

```
In [100]: water = pd.read_csv('BasicAndSafelyManagedDrinkingWaterServices.csv')
water.head()

Out[100]:
```

	Year	Country	Granularity	Population using at least basic drinking-water services (%)	Population using safely managed drinking-water services (%)
0	2000	Afghanistan	Rural	21.61913	NaN
1	2000	Afghanistan	Total	27.77190	NaN
2	2000	Afghanistan	Urban	49.48745	NaN
3	2000	Albania	Rural	81.78472	NaN
4	2000	Albania	Total	87.86662	49.29324

```
In [101]: water.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10476 entries, 0 to 10475
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Year        10476 non-null   int64
1   Country     10476 non-null   object
2   Granularity 10476 non-null   object
3   Population using at least basic drinking-water services (%) 9415 non-null    float64
4   Population using safely managed drinking-water services (%) 3286 non-null    float64
dtypes: float64(2), int64(1), object(2)
memory usage: 409.3+ KB

In [102]: water.isna().sum()

Out[102]: Year      0
Country    0
Granularity 0
Population using at least basic drinking-water services (%) 1061
Population using safely managed drinking-water services (%) 7190
dtype: int64

In [103]: water.nunique()

Out[103]: Year      18
Country    194
Granularity 3
Population using at least basic drinking-water services (%) 7784
Population using safely managed drinking-water services (%) 2877
dtype: int64
```

4. Stabilité politique

```
In [81]: stability = pd.read_csv('PoliticalStability.csv')
stability.head()

Out[81]:
```

	Country	Year	Political_Stability	Granularity
0	Afghanistan	2000	-2.44	Total
1	Afghanistan	2002	-2.04	Total
2	Afghanistan	2003	-2.20	Total
3	Afghanistan	2004	-2.30	Total
4	Afghanistan	2005	-2.07	Total

```
In [82]: stability.drop(columns=["Granularity"], inplace=True)

In [83]: stability.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3526 entries, 0 to 3525
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Country      3526 non-null   object
1   Year          3526 non-null   int64
2   Political_Stability 3526 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 82.8+ KB

In [84]: stability.isna().sum()

Out[84]: Country      0
Year      0
Political_Stability 0
dtype: int64
```

Choix de L'outil : POWER BI

1. Skill très demandé en entreprise
2. Interface intuitive
3. Beaucoup de choix de visualisations
4. Power Query déjà utilisé pour préparer et nettoyer les données

Difficultés:

- *Bien segmenter les besoins pour chaque vue*
- *Données manquantes et conception table finale.*
- *Des années de référence différentes selon les variables*

Conclusion