

Power BI

*Data visualisation sur les datasets :
Covid-19, Ventes, Marketing*

Auteur
GHOUIBI Ghassen

Encadré par
Mme Rakia Jaziri

Table des matières

I Introduction	2
I.1. Power BI	2
I.2. Objectif	2
II Mise en place	2
II.1. Machine Virtuelle	2
II.2. Installation	3
III Jeu de données	4
III.1. Chargement de données	4
III.2. Nettoyage et transformation de données	4
III.3. Dashbord intelligent	5
III.4. Business Analytique	6
III.5. Business Intelligence	6
IV Conclusion	6

I Introduction

Ce rapport contient les étapes réaliser pour répondre au question du projet détaillé en section, il sera accompagné du projet exporter.

I.1 Power BI

Power BI¹ est un service d'analyse commerciale de Microsoft. Il vise à fournir des visualisations interactives et des capacités de Business Intelligence avec une interface suffisamment simple pour que les utilisateurs finaux puissent créer leurs propres rapports et tableaux de bord.

I.2 Objectif

Le choix de travailler sur le covid en respectant les consignes du rapport notre dataset est sous ce format (20000,6). Notre objectif c'est prédire des résultats sur quelques jours à travers notre dataset.


II Mise en place

Après faire le choix sur l'outil qu'on va utiliser dans ce rapport, il est nécessaire quelques installation à faire avant d'installer Power BI.

II.1 Machine Virtuelle

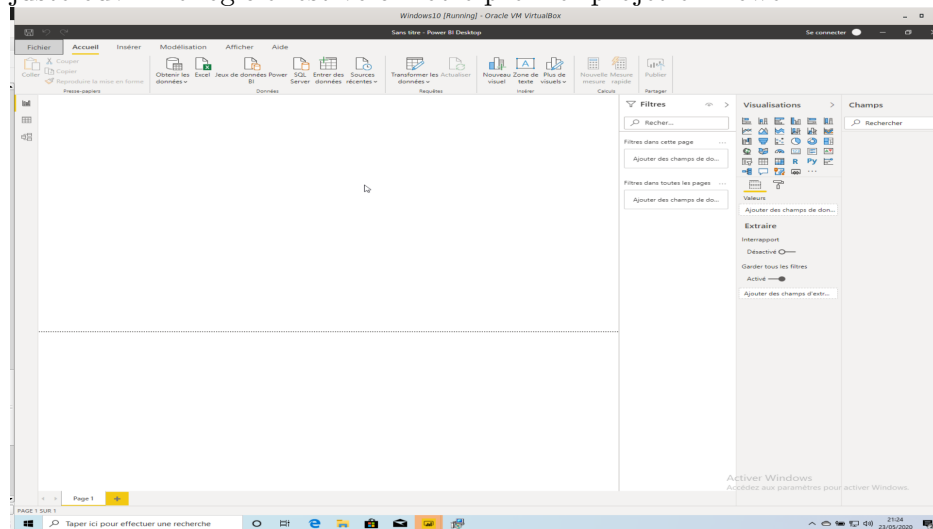
On fait les choix de configuration suivante après avoir télécharger Windows 10 grâce au lien suivant.

1. Cette définition est prise de wikipedia https://en.wikipedia.org/wiki/Power_BI

New Settings Discard Show		Preview
General Name: Windows10 Operating System: Windows 10 (64-bit) Settings File Location: /home/hope/VirtualBox VMs/Windows10		
System Base Memory: 5173 MB Boot Order: Floppy, Optical, Hard Disk Acceleration: VT-x/AMD-V, Nested Paging, Hyper-V Paravirtualization		
Display Video Memory: 128 MB Graphics Controller: VBoxSVGA Remote Desktop Server: Disabled Recording: Disabled		
Storage Controller: SATA SATA Port 0: Windows10.vdi (Normal, 114,83 GB) SATA Port 1: [Optical Drive] Win10_1909_French_x64.iso (Inaccessible)		
Audio Host Driver: PulseAudio Controller: Intel HD Audio		
Network Adapter 1: Intel PRO/1000 MT Desktop (NAT)		
USB USB Controller: OHCI Device Filters: 0 (0 active)		
Shared folders None		
Description None		

II.2 Installation

Après avoir installer notre Windows 10 sur notre machine virtuelle et passer l'étape de configuration de windows on peut télécharger Power BI. Soit on sur le Microsoft store, soit à traves ce lien. Maintenant il faut juste ouvrir le logiciel est voici notre premier projet en Power BI.



III Jeu de données

III.1 Chargement de données

Après avoir lancer **Power BI** on va cliquer sur *Obtenir les données* ensuite mettre le choix sur excel vu que nos données sont du format xlsx, après cliquer sur tranformer les données et voici un exemple :

	Province_State	Country_Region	Date	ConfirmedCases	Fatalities
1		Afghanistan	22/01/2020	0	0
2		Afghanistan	23/01/2020	0	0
3		Afghanistan	24/01/2020	0	0
4		Afghanistan	25/01/2020	0	0
5		Afghanistan	26/01/2020	0	0
6		Afghanistan	27/01/2020	0	0
7		Afghanistan	28/01/2020	0	0
8		Afghanistan	29/01/2020	0	0
9		Afghanistan	30/01/2020	0	0
10		Afghanistan	31/01/2020	0	0
11		Afghanistan	01/02/2020	0	0
12		Afghanistan	02/02/2020	0	0
13		Afghanistan	03/02/2020	0	0
14		Afghanistan	04/02/2020	0	0
15		Afghanistan	05/02/2020	0	0
16		Afghanistan	06/02/2020	0	0
17		Afghanistan	07/02/2020	0	0
18		Afghanistan	08/02/2020	0	0
19		Afghanistan	09/02/2020	0	0
20		Afghanistan	10/02/2020	0	0
21		Afghanistan	11/02/2020	0	0
22		Afghanistan	12/02/2020	0	0
23		Afghanistan	13/02/2020	0	0
24		Afghanistan	14/02/2020	0	0
25		Afghanistan	15/02/2020	0	0
26		Afghanistan	16/02/2020	0	0
27		Afghanistan	17/02/2020	0	0
28		Afghanistan	18/02/2020	0	0
29		Afghanistan	19/02/2020	0	0
30		Afghanistan	20/02/2020	0	0
31		Afghanistan	21/02/2020	0	0
32		Afghanistan	22/02/2020	0	0
33		Afghanistan	23/02/2020	0	0
34		Afghanistan	24/02/2020	1	0
35		Afghanistan	25/02/2020	1	0
36		Afghanistan	26/02/2020	1	0
37		Afghanistan	27/02/2020	1	0
38		Afghanistan	28/02/2020	1	0
39		Afghanistan	29/02/2020	1	0
40		Afghanistan	01/03/2020	1	0
41		Afghanistan	02/03/2020	1	0
42		Afghanistan	03/03/2020	1	0
43		Afghanistan	04/03/2020	1	0
44		Afghanistan	05/03/2020	1	0
45		Afghanistan	06/03/2020	1	0
46		Afghanistan	07/03/2020	1	0

Nos données sont importer dans **Power Query**, On effectue un analyse rapide sur notre dataset. On remarque que les **id** sont des int, **Province.State** est du text qui représente la ville des cas, **Country_Region** représente les pays, les cas confirmées sont représenté par **Confirmed Cases** et les cas de décès par **Fatalities** on remarque aussi qu'il y a une corrélation entre **Id** et **Date**. Pour résumé on a un data des cas confirmées et cas de décès tout les jour à partir du 22 Janvier 2020 jusqu'a 31 Mars notre études ce basé sur cette période.

III.2 Nettoyage et transformation de données

Comment préciser dans la section d'avant, en effet notre dataset est propre est prêt sauf qu'il manque quelques ville sur les informations mais ce n'est très dérangement dans notre cas.

On clique sur appliquer est ferme dans **Power Query** et maintenant on peut regarder notre dataset après la transformation si jamais on a effectuer quelques opération comme supprimer les doubles et les valeurs abérantes.

Id	Province_State	Country_Region	Date	ConfirmedCases	Fatalities
2		Afghanistan	2020-03-22	0	0
201		Albania	2020-03-22	0	0
202		Algeria	2020-03-22	0	0
302		Andorra	2020-03-22	0	0
402		Angola	2020-03-22	0	0
502		Antigua and Barbuda	2020-03-22	0	0
602		Argentina	2020-03-22	0	0
702		Armenia	2020-03-22	0	0
802		Austria	2020-03-22	0	0
1702		Azerbaijan	2020-03-22	0	0
1802		Bahamas	2020-03-22	0	0
1902		Bahrain	2020-03-22	0	0
2002		Bangladesh	2020-03-22	0	0
2102		Barbados	2020-03-22	0	0
2202		Belarus	2020-03-22	0	0
2302		Belgium	2020-03-22	0	0
2402		Belize	2020-03-22	0	0
2502		Benin	2020-03-22	0	0
2602		Bhutan	2020-03-22	0	0
2702		Bolivia	2020-03-22	0	0
2802		Bosnia and Herzegovina	2020-03-22	0	0
2902		Brazil	2020-03-22	0	0
3002		Brunei	2020-03-22	0	0
3102		Bulgaria	2020-03-22	0	0
3202		Burkina Faso	2020-03-22	0	0
3302		Burundi	2020-03-22	0	0
3402		Cambodia	2020-03-22	0	0
3502		Cameroon	2020-03-22	0	0
3602		Central African Republic	2020-03-22	0	0
3702		Chad	2020-03-22	0	0
3802		Chile	2020-03-22	0	0
3902		Colombia	2020-03-22	0	0
4002		Congo (Brazzaville)	2020-03-22	0	0
4102		Congo (Kinshasa)	2020-03-22	0	0
4202		Costa Rica	2020-03-22	0	0
4302		Cote d'Ivoire	2020-03-22	0	0
4402		Croatia	2020-03-22	0	0
4502		Cuba	2020-03-22	0	0
4602		Cyprus	2020-03-22	0	0
4702		Czechia	2020-03-22	0	0
4802		Denmark	2020-03-22	0	0
4902		Dominican Republic	2020-03-22	0	0
5002		Dominica	2020-03-22	0	0

Dans cette partie on choisit de regrouper par pays en plus additionner tout les nombre de cas confirmé et cas de décès mais néanmoins l'opération va fausser nos résultats le nombre de cas représente une accumulation de nombre de cas.

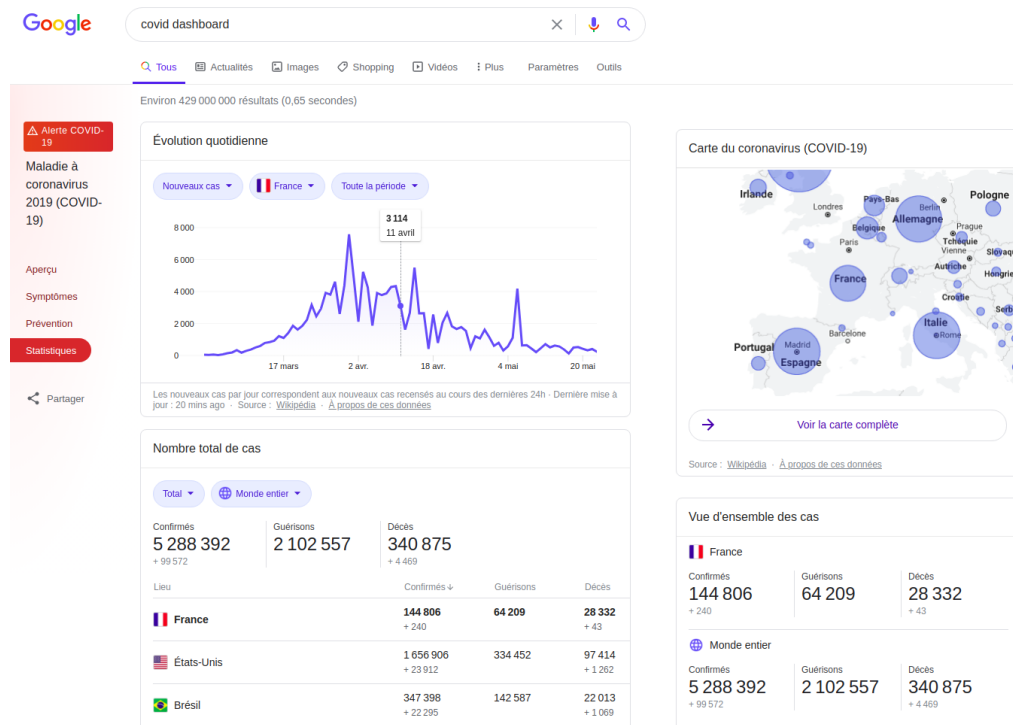
On va essayer de séparer la date pour faciliter la visualisation par jours avec un clique gacuhe sur la colone Date on ajouter une colone à partir de cet exemple, ensuite le même processus pour le mois et l'année.

Id	Province_State	Country_Region	Date	ConfirmedCases	Fatalities	Day
2		Afghanistan	2020-03-22	0	0	22
201		Albania	2020-03-22	0	0	22
202		Algeria	2020-03-22	0	0	24
302		Andorra	2020-03-22	0	0	25
402		Angola	2020-03-22	0	0	26
502		Antigua and Barbuda	2020-03-22	0	0	27
602		Argentina	2020-03-22	0	0	28
702		Armenia	2020-03-22	0	0	29
802		Austria	2020-03-22	0	0	30
1702		Azerbaijan	2020-03-22	0	0	31
1802		Bahamas	2020-03-22	0	0	32
1902		Bahrain	2020-03-22	0	0	33
2002		Bangladesh	2020-03-22	0	0	34
2102		Barbados	2020-03-22	0	0	35
2202		Belarus	2020-03-22	0	0	36
2302		Belgium	2020-03-22	0	0	37
2402		Belize	2020-03-22	0	0	38
2502		Benin	2020-03-22	0	0	39
2602		Bhutan	2020-03-22	0	0	40
2702		Bolivia	2020-03-22	0	0	41
2802		Bosnia and Herzegovina	2020-03-22	0	0	42
2902		Brazil	2020-03-22	0	0	43
3002		Brunei	2020-03-22	0	0	44
3102		Bulgaria	2020-03-22	0	0	45
3202		Burkina Faso	2020-03-22	0	0	46
3302		Burundi	2020-03-22	0	0	47
3402		Cambodia	2020-03-22	0	0	48
3502		Cameroon	2020-03-22	0	0	49
3602		Central African Republic	2020-03-22	0	0	50
3702		Chad	2020-03-22	0	0	51
3802		Chile	2020-03-22	0	0	52
3902		Colombia	2020-03-22	0	0	53
4002		Congo (Brazzaville)	2020-03-22	0	0	54
4102		Congo (Kinshasa)	2020-03-22	0	0	55
4202		Costa Rica	2020-03-22	0	0	56
4302		Cote d'Ivoire	2020-03-22	0	0	57
4402		Croatia	2020-03-22	0	0	58
4502		Cuba	2020-03-22	0	0	59
4602		Cyprus	2020-03-22	0	0	60
4702		Czechia	2020-03-22	0	0	61
4802		Denmark	2020-03-22	0	0	62
4902		Dominican Republic	2020-03-22	0	0	63
5002		Dominica	2020-03-22	0	0	64

On a aussi profiter d'extraire le nom du jour pour faire une visualisation sur les jours de la semaines.

III.3 Dashbord intelligent

Dans cette section on souhaite réaliser un dashbord intelligent comme celui de google.



III.4 Business Analytique

proposer et traiter un cas d'usage analytique en mettant l'accent sur l'étape de visualisation,

III.5 Business Intelligence

Création de tableau de bord pour le suivi de l'activité de l'entreprise, Reporting, création de datamart, ...

IV Conclusion