

MASTER 2 ECONOMETRIE ET STATISTIQUE
APPLIQUEE

DATAMINING

Résumé : La situation de covid en Corée du SUD est préoccupante puisqu'il est dans la zone rouge

ENSEIGNANT :

Dieudonné SONDJIO

ETUDIANTS :

Achille ADJIKPE

Fataï OSSENI

ANALYSE ET
MODELISATION
DE LA PANDEMIE
DE COVID-19 EN
COREE DU SUD

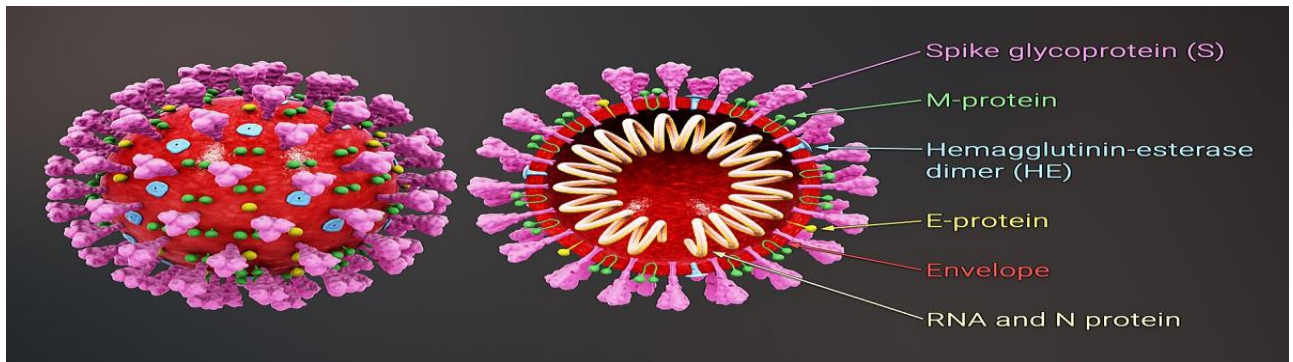
Table des matières

1	INTRODUCTION	3
2	ANALYSE EXPLORATOIRE DE LA BASE	4
2.1	DESCRIPTION DE LA BASE	4
2.2	PRESENTATION DES VALEURS MANQUANTES	4
2.3	REPARTITION DES STATUTS DES PATIENTS DU KOREE DU SUD	5
2.4	REPARTITION PAR AGE DES STATUTS DES PATIENTS DU KOREE DU SUD	5
2.5	REPARTITION DU TAUX DES CAS CONFIRMES PAR AGE DE LA POPULATION.....	6
2.6	REPARTITION COMPAREE DES CAS CONFIRMES ET TAUX DES CAS CONFIRMES	6
2.7	EVOLUTION DES CAS CONFIRMES EN FONCTION DES DATES	7
2.8	EVOLUTION DES CAS DE DECES PAR AGE DE LA POPULATION	8
2.9	EVOLUTION DU TAUX DE DECES PAR AGE DE LA POPULATION	8
2.10	REPARTITION PAR SEXE DES STATUTS DES PATIENTS DU KOREE DU SUD	9
2.11	REPARTITION PAR REGION DES STATUTS DES PATIENTS DU KOREE DU SUD	10
2.12	REPARTITION PAR DIAGRAMME CIRCULIRE DES CAS CONFIRMES UNIQUEMENT	10
2.13	REPARTITION CUMMULEE DES CAS CONFIRMES.....	11
2.14	REPARTITION CHIFFREE DES STATUTS PAR REGION	11
2.15	DISRIBUTION PAR AGE DE LA POPULATION.....	12
2.16	PROPORTION DES PERSONNES AGEES PAR REGION EN KOREE DU SUD	12
2.17	PROPORTION DES PERSONNES AGEES PAR REGION ET PAR CAS CONFIRMES	13
2.18	REPARTITION DES CAS CONFIRMES PAR SEXE A TRAVARES UN DIAGRAMME CIRCULAIRE.....	13
2.19	REPARTITION DES CAS DECEDES PAR SEXE A TRAVERS UN DIAGRAMME CIRCULAIRE.....	14
2.20	REPARTITION DU SEXE DES SEXE CONFIRMES PAR REGION	14
3	LES CHAINES DE CONTAMINATION DU COVID EN KOREE DU SUD	15
3.1	VOIE DE TRANSMISSION DE LA MALADIE EN JUIN 2020	15
3.2	RESUME DES LIEUX ET CHEMIN DE CONTAMINATION	15
3.3	LES DIFFERENTS TYPES DE CONTAMINTIONS DE COVID-2019.....	16
3.4	ANALYSE METEOROLOGIQUE DU JEU DE DONNEE en 2016	16
3.4.1	MOYENNE DES TEMPERATURES DANS CHAQUE REGIONS.	16
3.4.2	LA VITESSE MAXIMALE DU VENT.....	17
3.5	ANALYSE METEOROLOGIQUE DU JEU DE DONNEE EN TEMPS DE COVID-19.....	17
3.5.1	La TEMPERATURE MOYENNE	17
3.5.2	LA VITESSE MAXIMALE DU VENT.....	18
3.6	EVOLUTION DU NOMBRE MOYEN DE CONTACT PAR PATIENT CONFIRME	18
3.7	ANALYSE GENERALE DE LA POPULATION FLOTTANTE	19
3.7.1	LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR SEXE	19
3.7.2	LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR AGE.....	19
3.7.3	LA STRUCTURE FLOTTANTE DE LA POPULATION PAR VILLES	20
3.8	ANALYSE DES POLITIQUES GOUVERNEMENTALES CONTRE COVID-19	20
3.9	LES ACTIONS MENEES PAR LE GOUVERNEMENT DANS LES DOMAINES CIBLES.....	21

3.10	TENDANCES DES RECHERCHES SUR LES THEMES LIES AUX MALADIES RESPIRATOIRES AVANT ET APRES LE PREMIER CAS DE COVID EN KOREE DU SUD	21
3.10.1	EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DURANT TOUTE LA PERIODE DU JEU DE DONNEE.	21
3.10.2	EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DEPUIS L'ANNONCE DU PREMIER CAS DANS LE MONDE.....	22
3.10.3	EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DEPUIS L'ANNONCE DU PREMIER CAS EN KOREE DU SUD	22
4	ANALYSE GEOSPATIALE DE LA PANDEMIE DE COVID-19	23
4.1	ANALYSE GEOSPACIALE DES STATUTS DES PATIENTS DE COVID-2019 EN KOREE DU SUD.....	23
4.2	ANALYSE DE LA CHRONOLOGIE DE L'EVOLUTION DU COVID-2019 EN KOREE DU SUD	24
4.3	ANALYSE DES CAS CONFIRMES A TRAVERS UN TREEMAP	24
4.4	ANALYSE DES CAS DELIVRES A TRAVERS UN TREEMAP	25
4.5	ANALYSE DES CAS DE DECES A TRAVERS UN TREEMAP	25
4.6	ANALYSE COMPARATIVE DES CAS DE COVID EN KOREE DU SUD ET AU NIVEAU MONDIALE.	26
4.6.1	POSITION GEOSPATIALE DE L'EVOLUTION DU COVID-2019 AU NIVEAU MONDIALE	26
4.6.2	GEOLOCALISATION DES PAYS LES PLUS TOUCHES DU MONDE.....	26
5	PREDICTION DES CAS DE COVID EN KOREE DU SUD	27
5.1	NATURE DE RELATION ENTRE LES TENDANCES DE RECHERCHES ET LES CAS CONFIRMES DE COVID	27
5.2	EXTRACTION DE LA LINEARITE DANS LES DEUX DONNEES (LES RESIDUS)	27
5.3	REGRESSIONN LINEAIRE	28
5.3.1	SCORE COMPARE DES DONNEES TRAIN/TEST SUR LES DONNEES DE GOOGLE.....	28
5.4	SCORE COMPARE TRAIN/TEST SUR LES DONNEES DE NAVER.....	28
5.4.1	PREDICTION REELLE DES CAS CONFIRMES SUR LES DONNEES DE GOOGLR/NAVER	29
5.5	PASSAGE AU METHODE DE PREDICTION NAIVE ET AU METHODE DE PREDICTION PAR MOYENNE MOBILE.	29
5.6	COMPARAISON DES DIFFERENTS MODELES ENTRAINES POUR PREDIRE LE NOMBRE DE CAS CONFIRMES DE COVID	30
6	RECOMMANDATION DES INITIATIVES POUR CASSER LES CHAINES DE CONTAMINATION EN KOREE DU SUD	30
6.1	POLITIQUE ENVERS LES GROUPES D'AGE EN KOREE DU SUD.	30
6.2	LES ACTIONS D'ORDRE REGIONAL	30
6.3	POLITIQUE ENVERS LES RECHERCHES SUR LES MALADIES RESPIRATOIRES	30

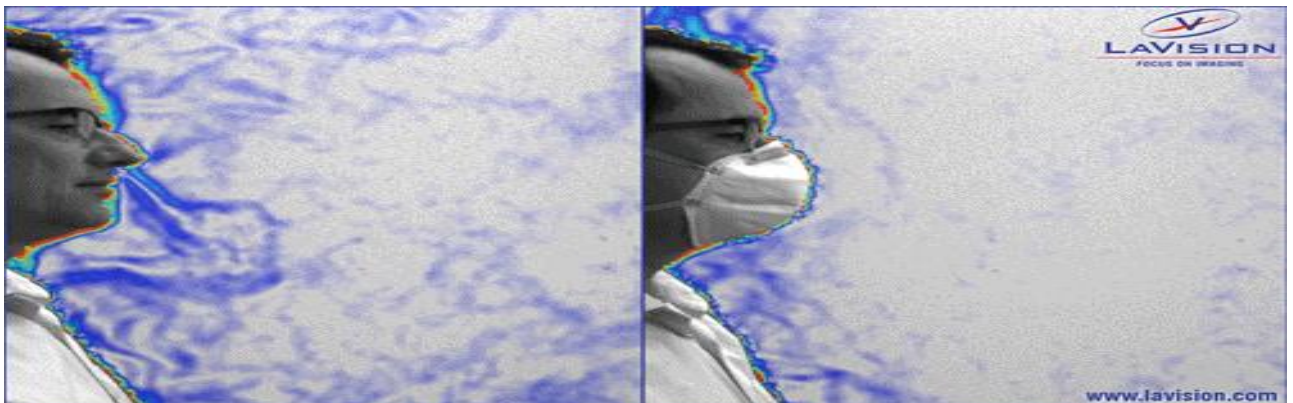
1 INTRODUCTION

La maladie à coronavirus 2019 (COVID-19) est une maladie infectieuse causée par le coronavirus 2 du syndrome respiratoire aigu sévère (SRAS-CoV-2). Elle a été identifiée pour la première fois en décembre 2019 à Wuhan, Hubei, en Chine, et a donné lieu à une pandémie en cours.



Source : Wikipedia.org

PROPAGATION DE LA PANDEMIE DU COVID-19



Source : Source: LAvision

Le virus se transmet principalement par de petites gouttelettes provenant de la toux, des éternuements et de la parole. Ces gouttelettes ne sont généralement pas aéroportées, mais les personnes se trouvant à proximité peuvent les inhaler et être infectées. Les personnes peuvent également être infectées en touchant une surface contaminée puis en se touchant le visage. La transmission peut également se faire par les aérosols qui peuvent rester en suspension dans l'air pendant de longues périodes dans des espaces clos.

L'objectif principal de ce projet est d'analyser et de modéliser le COVID-19 en Corée du Sud. De façon spécifique, il s'agira de :

- Faire une analyse géo spatiale de l'épidémie
- Faire une étude sur les chaînes de contamination et de réaliser sur la base des informations un modèle de prédiction et de la contamination du covid19 à partir des symptômes observés.

2 ANALYSE EXPLORATOIRE DE LA BASE

2.1 DESCRIPTION DE LA BASE

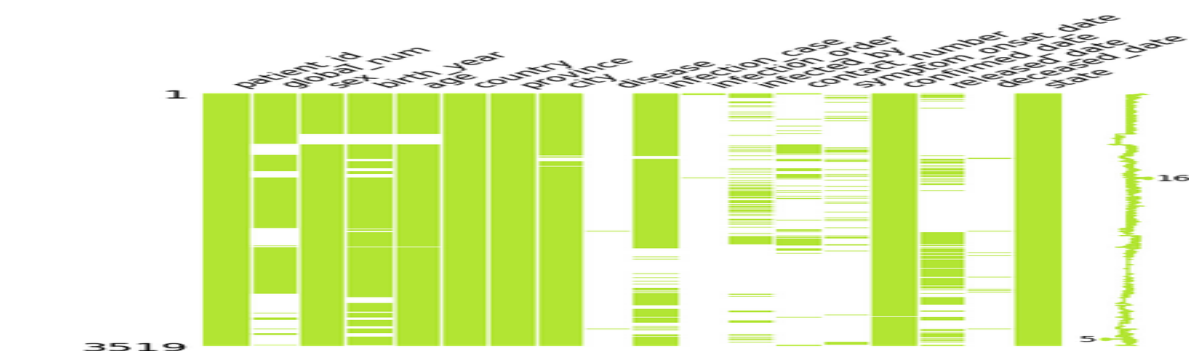
	Les_bases	Nombre_de_ligne	Nombre_de_colonne	valeurs_manquantes	Nombre_valeurs_manquantes	Noms_colonnes_manquantes
0	TimeAge	1089	5	0	0	
1	TimeProvince	2771	6	0	0	
2	Region	244	12	0	0	
3	TimeGender	242	5	0	0	
4	Time	63	7	0	0	
5	Case	174	8	0	0	
6	Weather	26271	10	81	6	avg_temp, min_temp, max_temp, max_wind_speed, most_wind_direction, avg_relative_humidity
7	PatientInfo	3519	18	24037	14	global_num, sex, birth_year, age, city, disease, infection_case, infection_order, infected_by, contact_number, symptom_onset_date, confirmed_date, released_date, deceased_date
8	PatientRoute	7279	8	3245	1	global_num
9	SeoulFloating	1084800	7	0	0	
10	SearchTrend	1642	5	0	0	
11	Policy	61	7	39	2	detail, end_date

Source : Réalisée par les auteurs de ce projet

Tableau1 : présentation des bases

Le tableau montre les différentes bases de ce projet ainsi que leurs descriptions. On remarque que la base PatientInfo est celle qui contient assez de valeurs manquantes et est d'ailleurs coloriée en noire. Le tableau montre également que la variable Weather contient aussi assez de valeurs manquantes.

2.2 PRESENTATION DES VALEURS MANQUANTES

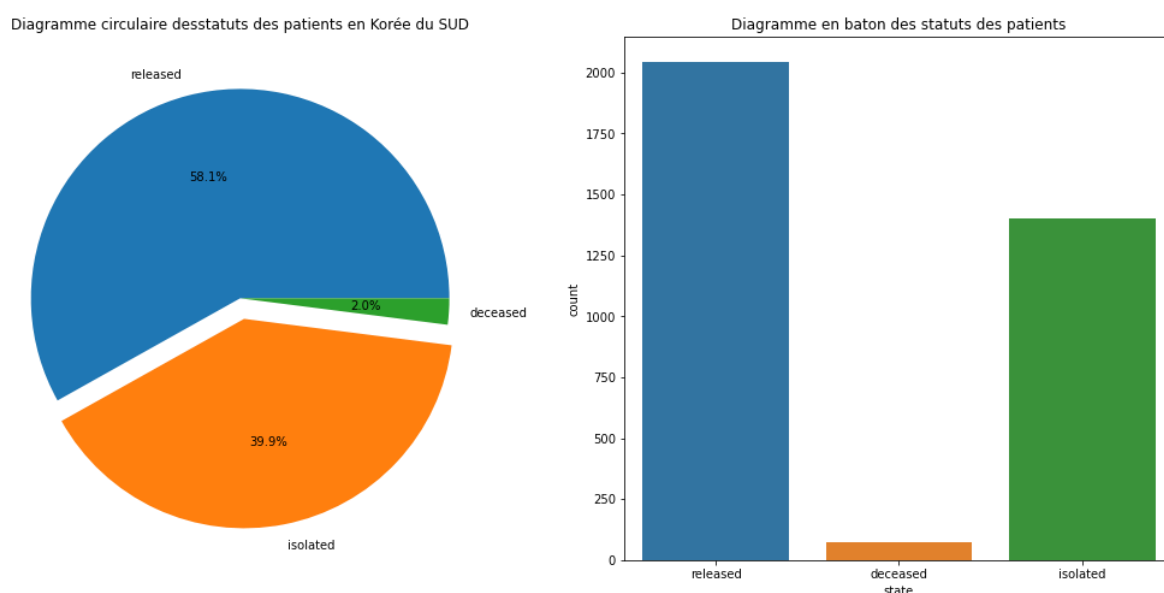


Source : Réalisée par les auteurs de ce projet

Figure1 : valeurs manquantes de la base PatientInfo

L'analyse de cette figure montre que les variables disease, infection_order , infected_by, deceased_date sont constituées respectivement de 99.49% ,99.12%,76.24,98.24 de valeurs manquantes et pleines d'autres variables contiennent également des valeurs manquantes d'après l'observation de la figure ci-dessus.

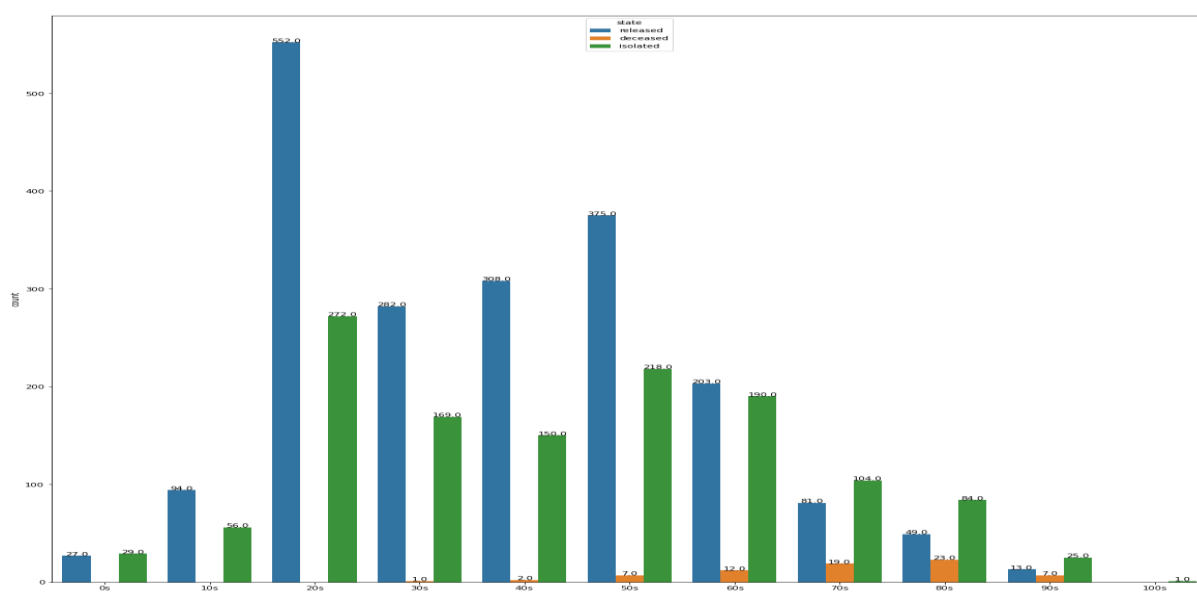
2.3 REPARTITION DES STATUTS DES PATIENTS DU KOREE DU SUD



Source : Réalisée par les auteurs de ce projet

Figure2 : Répartition des statuts

2.4 REPARTITION PAR AGE DES STATUTS DES PATIENTS DU KOREE DU SUD

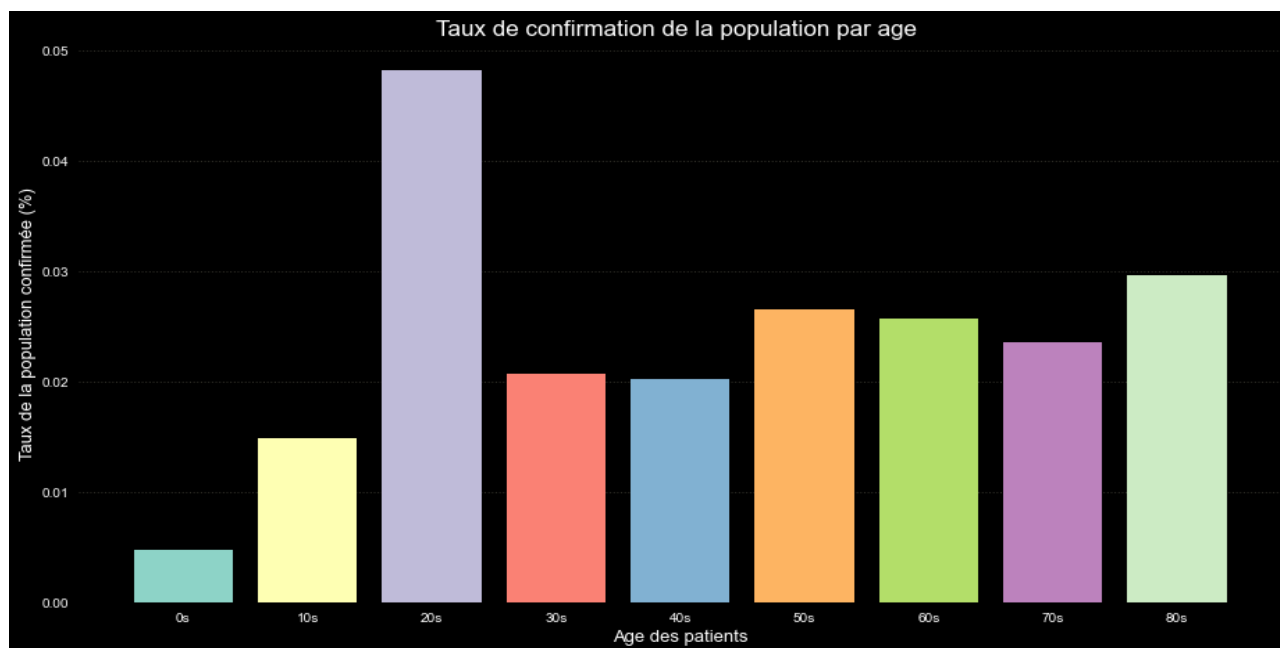


Source : Réalisée par les auteurs de ce projet

Figure3 : REPARTITION PAR AGE DES STATUTS

La figure montre que presque tous les patients de moins de 20ans sont guéris et que la probabilité d'être mort augmente à partir de 40ans. De plus des cas d'isolement deviennent élevés à partir de 20ans.

2.5 REPARTITION DU TAUX DES CAS CONFIRMES PAR AGE DE LA POPULATION

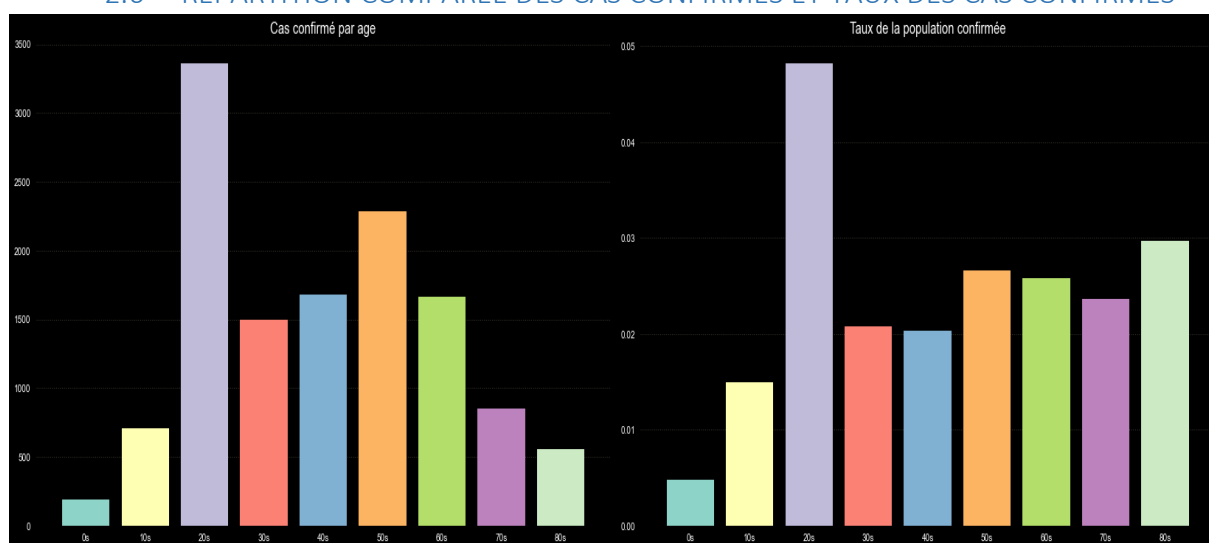


Source : Réalisée par les auteurs de ce projet

Figure4 : Cas confirmés par âge

La figure montre que les groupes de 20ans sont largement représentés dans les cas confirmés comparativement aux autres catégories d'âge. Une raison probable est que cette catégories d'âge néglige l'importance de cette maladie ou qu'elle exerce plusieurs activités nécessitant de plusieurs rencontres.

2.6 REPARTITION COMPAREE DES CAS CONFIRMES ET TAUX DES CAS CONFIRMES

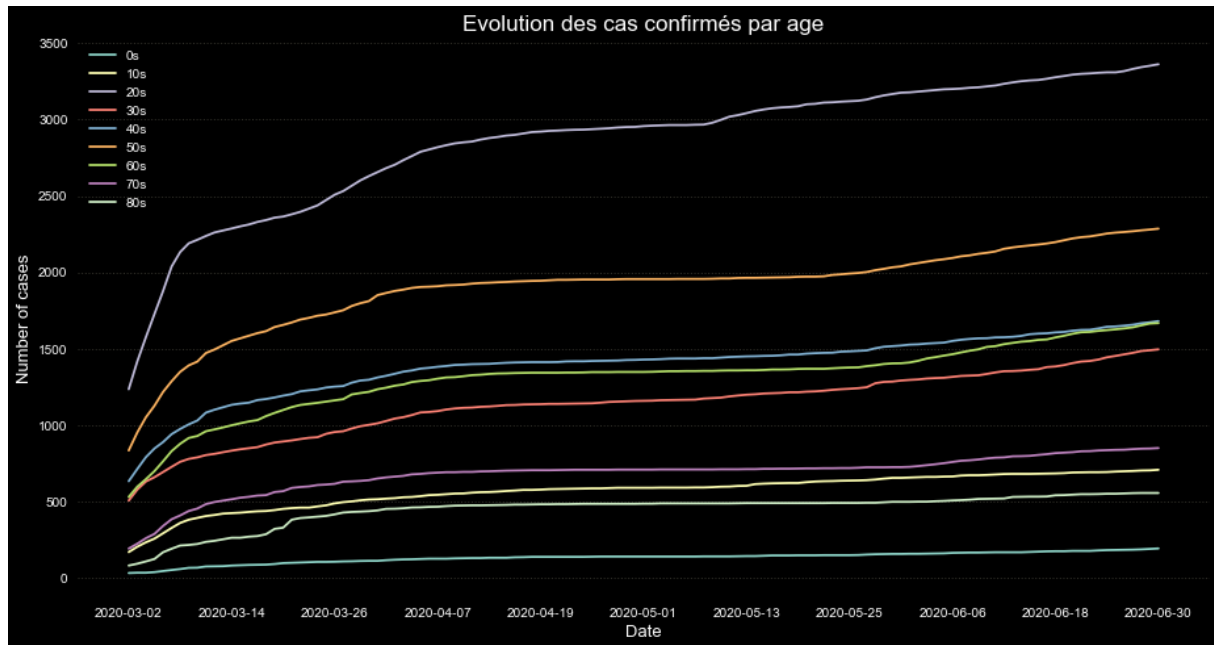


Source : Réalisée par les auteurs de ce projet

Figure5 : REPARTITION COMPAREE DES CAS CONFIRMES ET TAUX DES CAS CONFIRMES

La figure confirme encore une forte représentativité des jeunes dans les cas confirmés et signale qu'on est susceptible d'être infecté à mesure que l'âge des patients augmente.

2.7 EVOLUTION DES CAS CONFIRMES EN FONCTION DES DATES



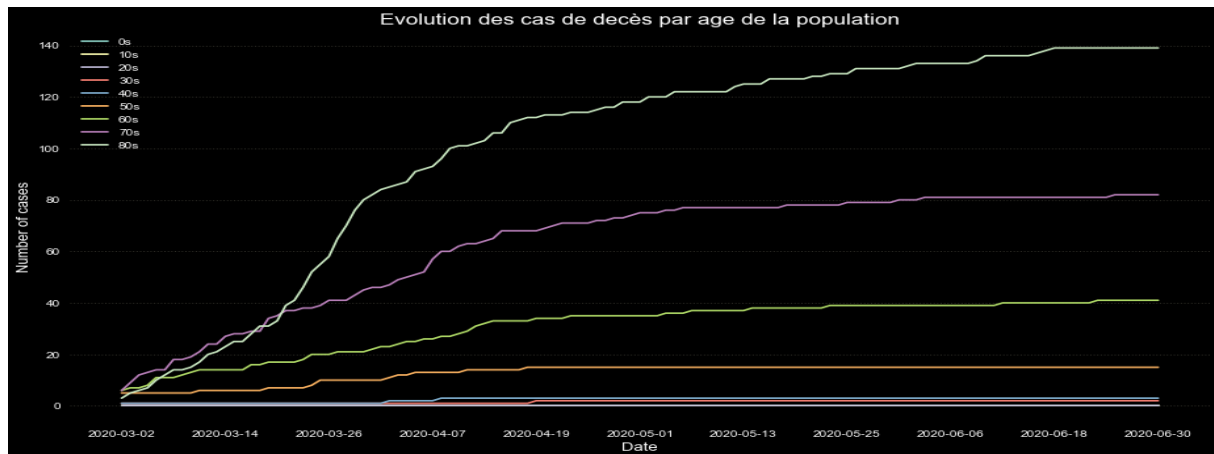
Source : Réalisée par les auteurs de ce projet

Figure6 : Evolution des cas confirmés par âge

La figure ci-dessus nous permet de faire des observations suivantes :

- L'écart entre les 20 ans et les autres groupes d'âge s'est creusé jusqu'en 2020-03-08
- Après cela, la pente de tous les groupes s'adoucit.
 - Mais celle des 80 ans recommence à s'accroître vers 2020-03-18
 - -Celui des 20 ans a augmenté plus que les autres (à partir de 2020-03-23 et 2020-05-07)
- A partir du 2020-06-30, la plupart des groupes présentent des courbes ascendantes (sauf pour les 0, 10 et 80).

2.8 EVOLUTION DES CAS DE DECES PAR AGE DE LA POPULATION



[Les récents décès enregistrés dans la base]

	0	1	2	3	4	5	6	7	8
age	0s	10s	20s	30s	40s	50s	60s	70s	80s
deceased	0	0	0	2	3	15	41	82	139

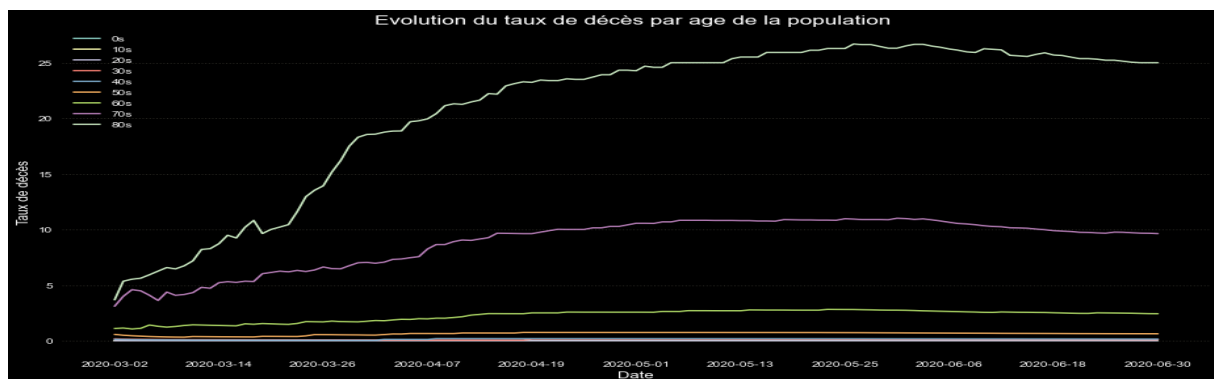
Source : Réalisée par les auteurs de ce projet

Figure7 : Evolution des cas de décès en fonction des dates

La figure ci-dessus nous permet de faire les déductions suivantes :

- Nettement différent des confirmés, les 80 et 70 sont aux places les plus élevées
- Les 20 ans n'ont aucun cas décédé alors qu'il y a environ 3 500 infections parmi eux.
- À partir de 2020-06-30, tous les âges montrent des courbes qui s'aplatissent, même si les infections continuent à se produire.

2.9 EVOLUTION DU TAUX DE DECES PAR AGE DE LA POPULATION



Source : Réalisée par les auteurs de ce projet

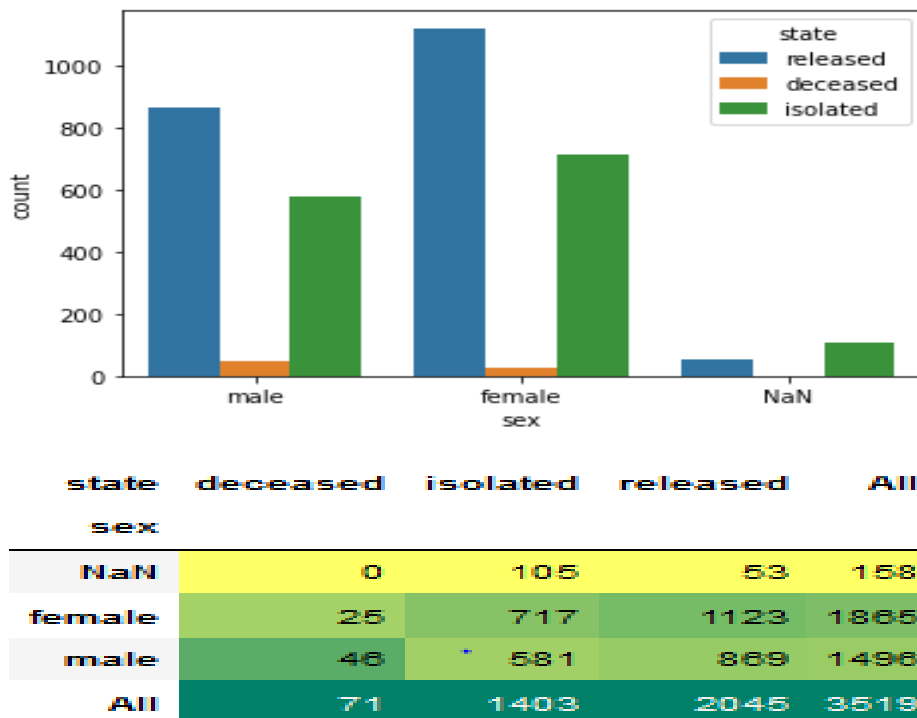
Figure8 : Taux de décès par âge

La figure nous permet de faire des déductions suivantes :

- Le taux de mortalité des groupes de 80ans est plus de deux fois supérieur à celui des groupes de 70ans.

- Les taux de décès des groupes de 80ans et 70 se comportent de manière décroissante à la fin du mois de mai (plus de cas, moins de décès).
- Les autres groupes ont montré des taux stables jusqu'à présent.

2.10 REPARTITION PAR SEXE DES STATUTS DES PATIENTS DU KOREE DU SUD



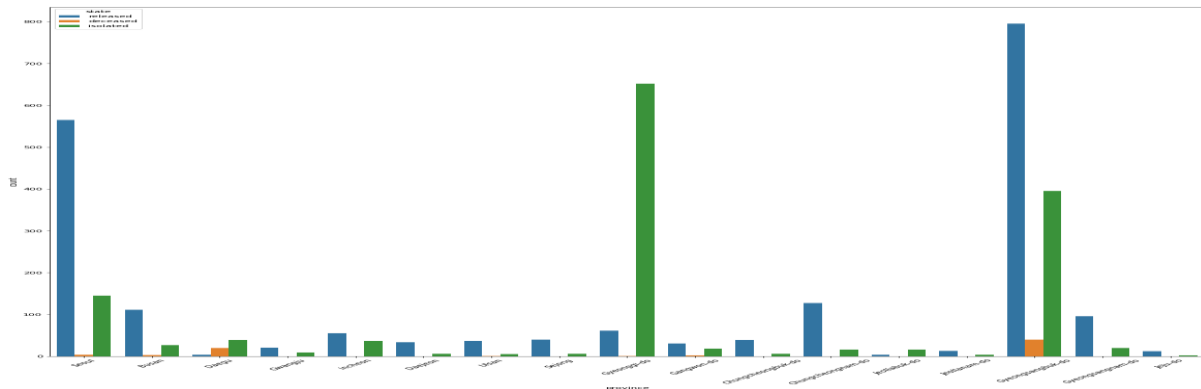
Source : Réalisée par les auteurs de ce projet

Figure9: REPARTITION PAR SEXE DES STATUTS

La figure ci-dessus nous permet de faire les remarques suivantes :

- Il y a une répartition proportionnelle des cas délivrés et ceux des cas isolés. Ce résultat est conforté par le tableau ci-dessus.
- Les femmes semblent plus susceptibles d'être infectées (isolées de la population) (51,10 % contre 41,41 %).
- Les hommes semblent plus enclins à être décédés (Taux de mortalité homme=64.78 % contre Taux de mortalité femme=35,21 %).

2.11 REPARTITION PAR REGION DES STATUTS DES PATIENTS DU KOREE DU SUD

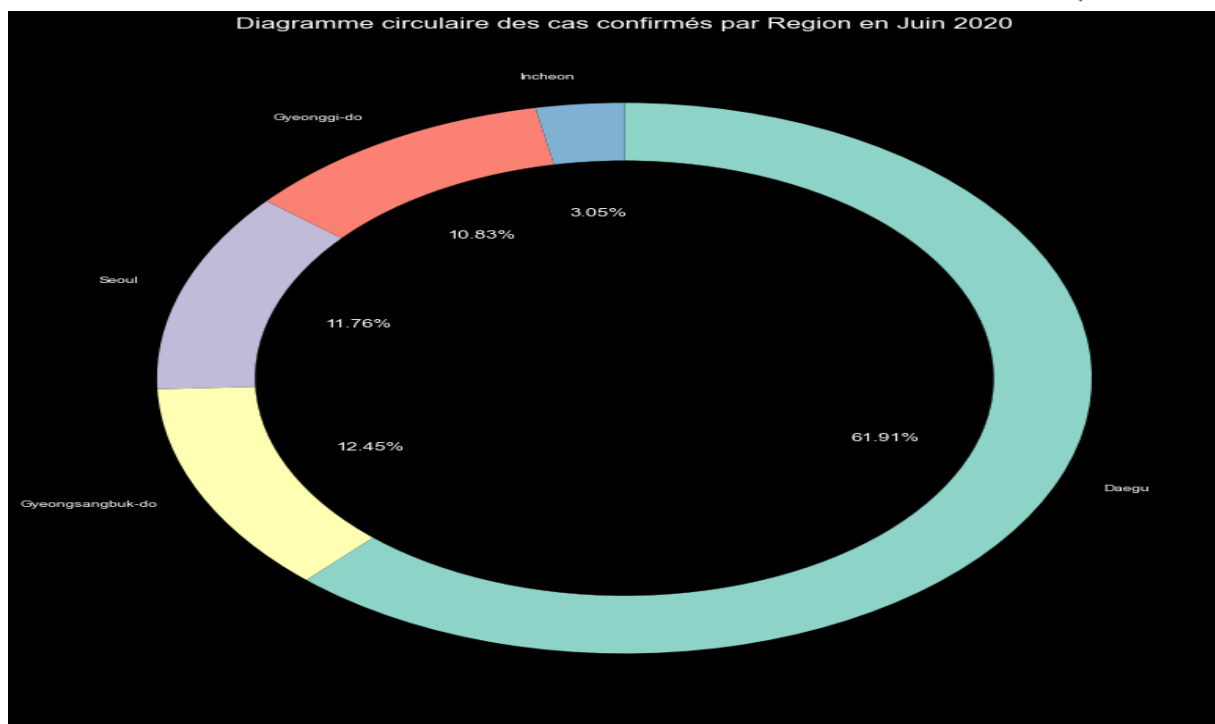


Source : Réalisée par les auteurs de ce projet

Figure10 : REPARTITION PAR REGION DES STATUTS

La figure montre que les régions à forte isolements sont Gyeongsangbuk-do et Gyeonggi-do. Les cas de mort ne sont pas assez.

2.12 REPARTITION PAR DIAGRAMME CIRCULAIRE DES CAS CONFIRMES UNIQUEMENT



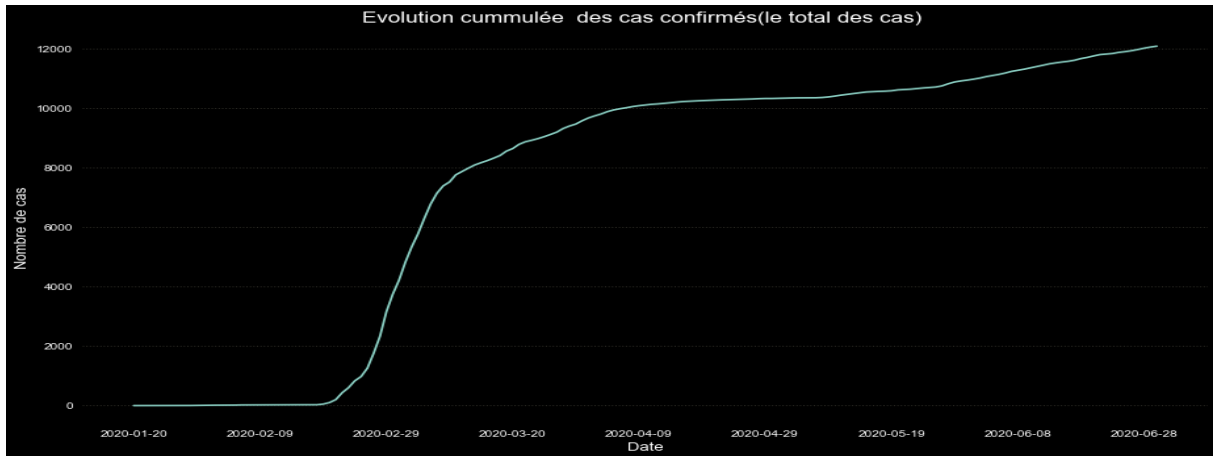
Source : Réalisée par les auteurs de ce projet

Figure11 : Diagramme circulaire des cas confirmés

La figure ci-après nous permet de faire des conclusions suivantes :

- Daegu et Gyeongsangbuk-do comptent environ 70 % de tous les cas confirmés.
- En dehors de ces régions, c'est la région métropolitaine de la capitale qui compte le plus de cas (Séoul, Gyunggi-do, Incheon).

2.13 REPARTITION CUMMULEE DES CAS CONFIRMES



Source : Réalisée par les auteurs de ce projet

Figure12: REPARTITION CUMMULEE DES CAS CONFIRMES

La figure nous permet de faire les déductions suivantes après quelques calculs sur le nombre de cas durant les mois de 2020 :

- Moins de 100 cas pour le premier mois (du 1er cas le 2020-01-20 au 104ème le 2020-02-20).
- Une forte croissance linéaire s'est produite à partir de là pendant environ 3 semaines (avec un dépassement de 8000 cas le 14 mars 2020).
- La courbe s'est aplatie jusqu'au 2020-05-08 (puis a recommencé à monter).

2.14 REPARTITION CHIFFREE DES STATUTS PAR REGION

state province	deceased	isolated	released	All
Busan	3	27	111	141
Chungcheongbuk-do	0	6	39	45
Chungcheongnam-do	0	16	127	143
Daegu	20	39	4	63
Daejeon	0	6	34	40
Gangwon-do	2	18	31	51
Gwangju	0	9	21	30
Gyeonggi-do	1	652	61	714
Gyeongsangbuk-do	40	395	795	1230
Gyeongsangnam-do	0	20	96	116
Incheon	0	37	55	92
Jeju-do	0	2	12	14
Jeollabuk-do	0	16	4	20
Jeollanam-do	0	4	13	17
Sejong	0	6	40	46
Seoul	4	145	565	714
Ulsan	1	5	37	43
All	71	1403	2045	3519

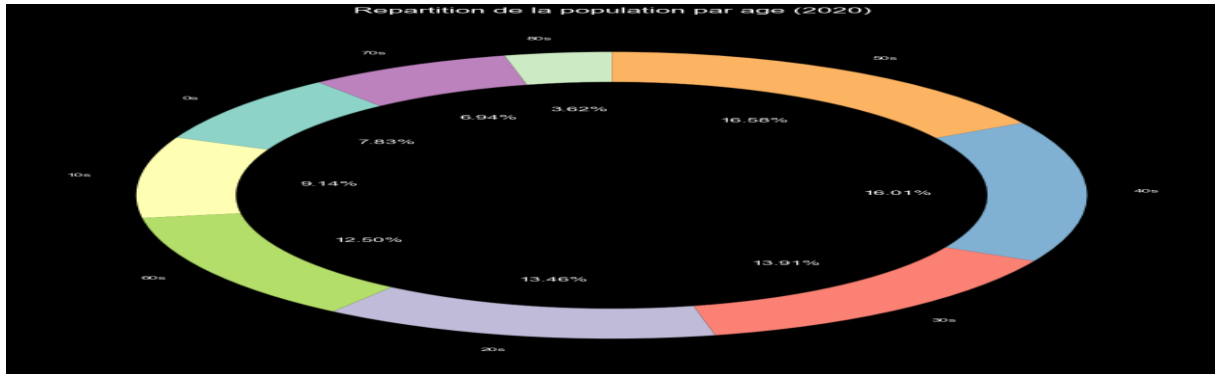
Source : Réalisée par les auteurs de ce projet

Tableau2 : Représentation chiffrée des statuts par région

Ce tableau confirme les résultats de la figure précédente et montre que les régions telles Gyeongsangbuk-do et Gyeonggi-do ont respectivement 395 et 652 de cas isolés. Retenons que le nombre de mort enregistré n'est pas assez (71 morts).

2.15 DISTRIBUTION PAR AGE DE LA POPULATION

A ce niveau, nous avons complété la base de donnée par la variable population qui prend en compte le nombre de population par Age de la population en Corée du Sud (Source : [KOSIS - Korean Statistical Information Service](#)).

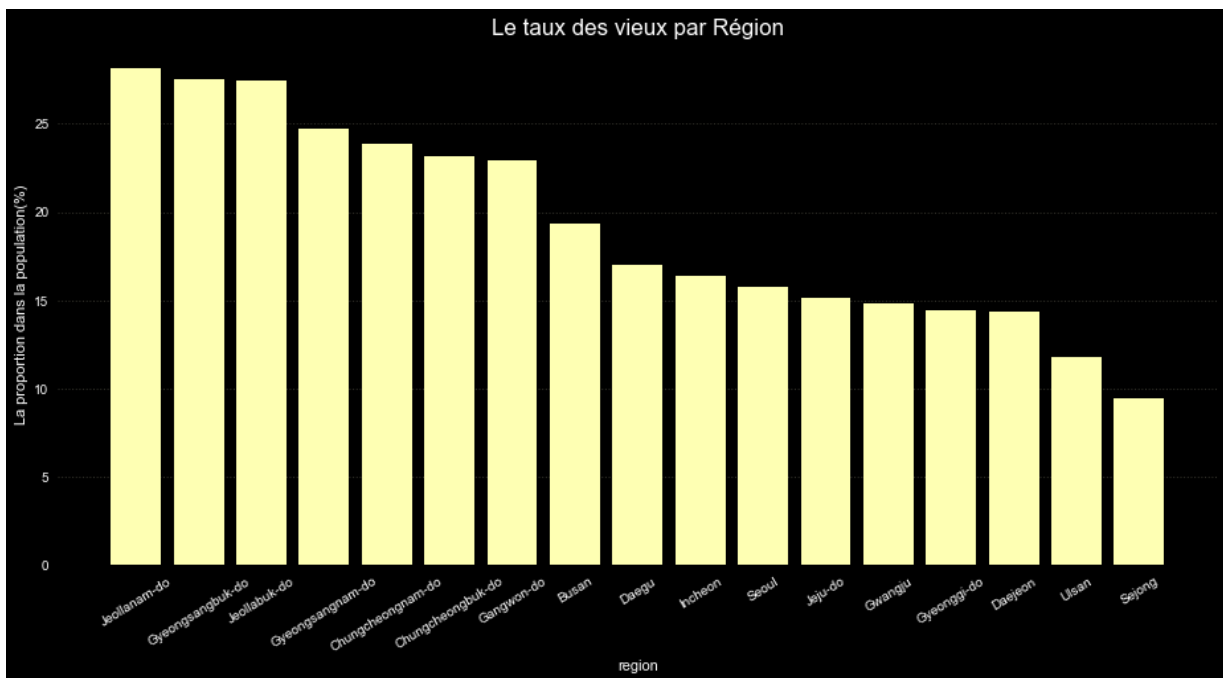


Source : Réalisée par les auteurs de ce projet

Figure13 : Répartition par Age de la population

La figure montre que la population est constituée en majorité des groupes de 40ans et 50ans. Elle montre qu'il y a équiprobabilité entre le groupes de 30ans et de 20ans.

2.16 PROPORTION DES PERSONNES AGEES PAR REGION EN KOREE DU SUD



Source : Réalisée par les auteurs de ce projet

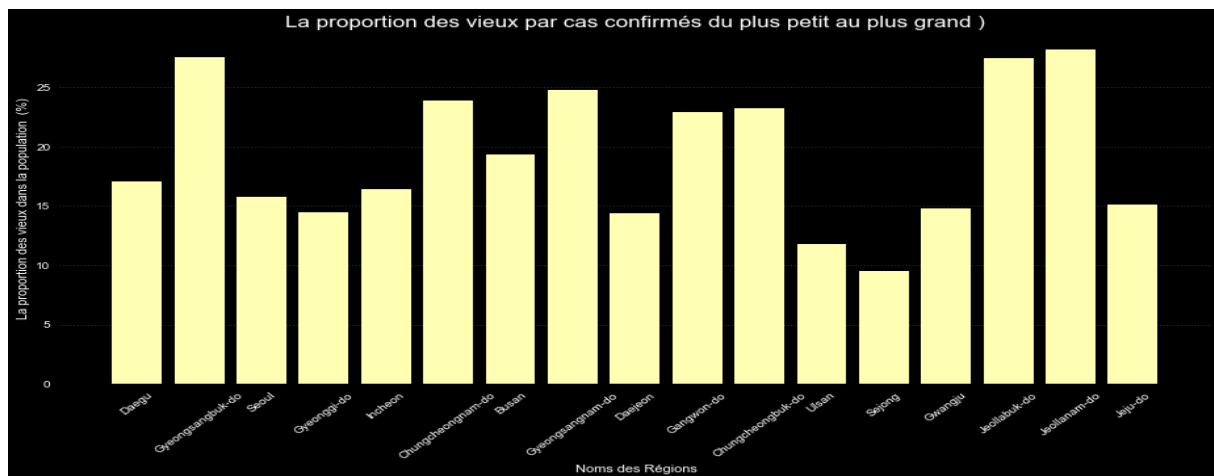
Figure14: PROPORTION DES PERSONNES AGEES PAR REGION

La figure nous permet de faire des remarques suivantes :

- Aucune région particulière ne présente une composition exceptionnellement élevée ou faible de personnes âgées.

- L'épicentre de l'épidémie, Daegu, se trouve à la 9e place de la liste et n'est pas tellement plus âgée que les autres villes à droite.

2.17 PROPORTION DES PERSONNES AGEES PAR REGION ET PAR CAS CONFIRMES

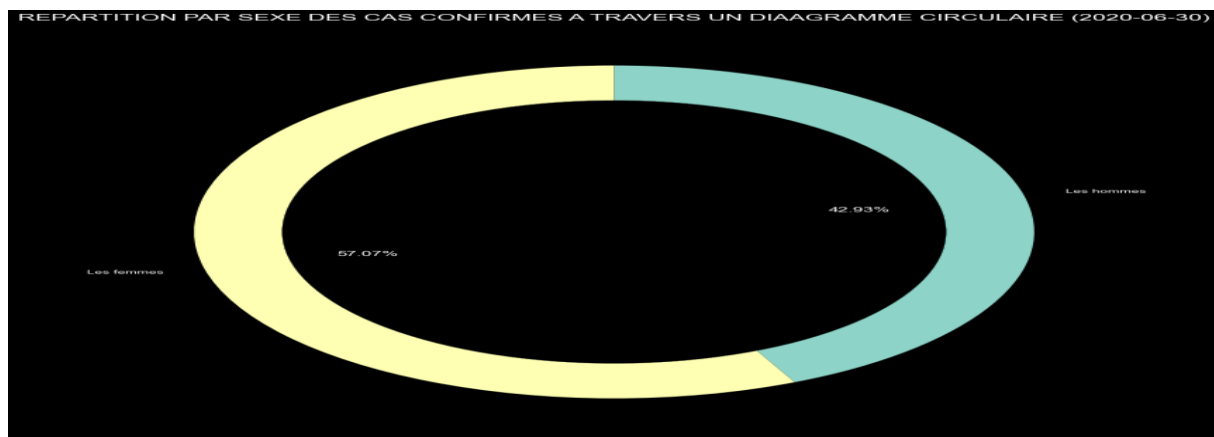


Source : Réalisée par les auteurs de ce projet

Figure15: PROPORTION DES PERSONNES AGEES PAR REGION ET PAR CAS CONFIRMES

La figure ci-dessus nous permet de conclure qu'il n'y a pas de forte corrélation entre la proportion de personnes âgées dans la population et le nombre de cas confirmé.

2.18 REPARTITION DES CAS CONFIRMES PAR SEXE A TRAVERS UN DIAGRAMME CIRCULAIRE

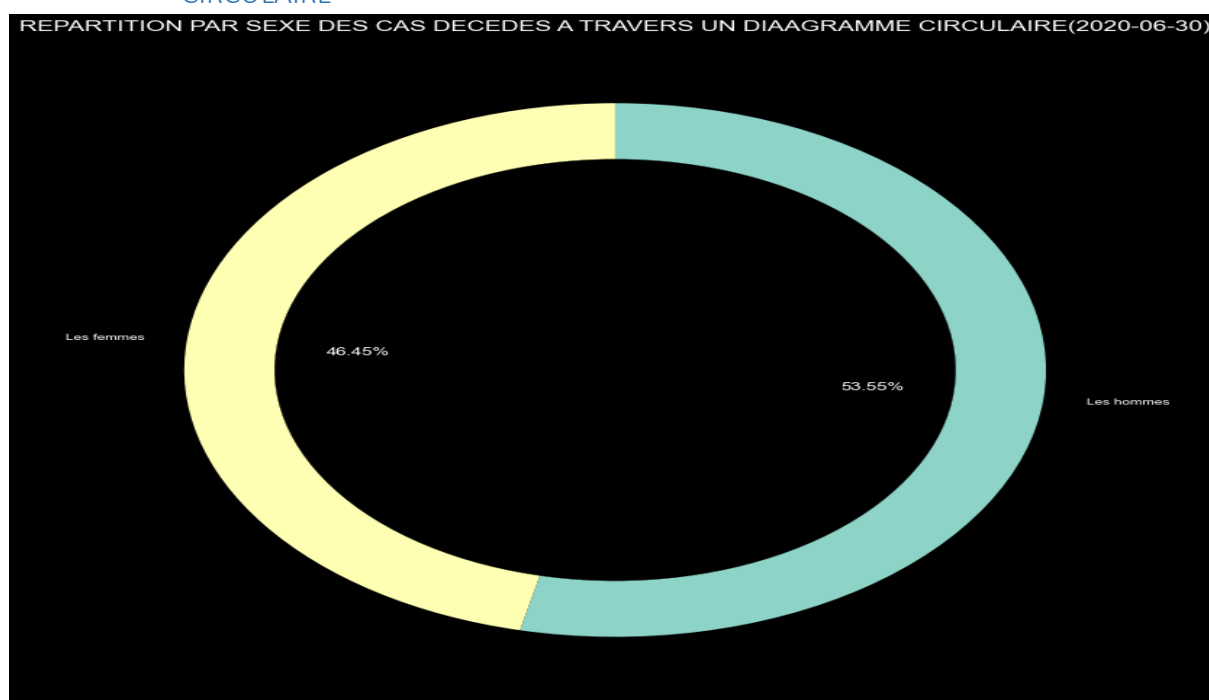


Source : Réalisée par les auteurs de ce projet

Figure16 : REPARTITION DES CAS CONFIRMES PAR SEXE A TRAVERS UN DIAGRAMME CIRCULAIRE

La figure montre qu'il y a plus de femmes infectées que les hommes

2.19 REPARTITION DES CAS DECEDES PAR SEXE A TRAVERS UN DIAGRAMME CIRCULAIRE



Source : Réalisée par les auteurs de ce projet

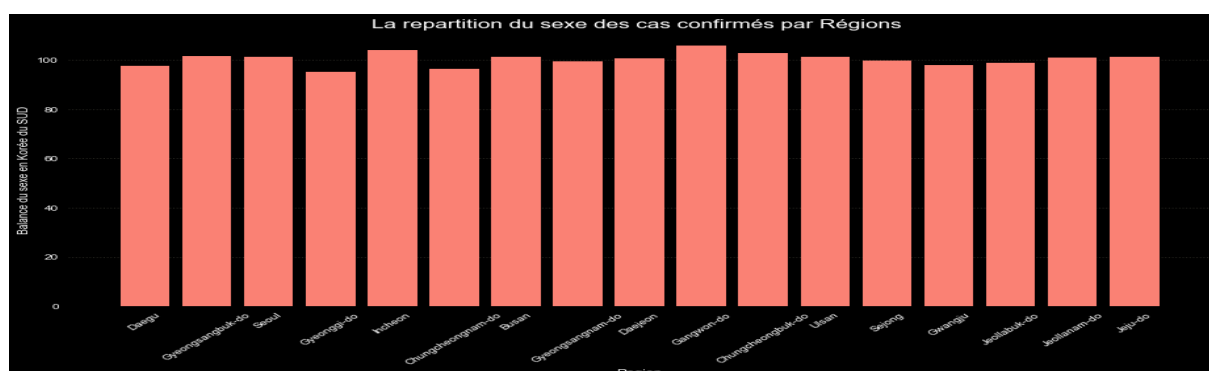
Figure17: REPARTITION DES CAS DECEDES PAR SEXE A TRAVERS UN DIAGRAMME CIRCULAIRE

La figure nous permet de faire des conclusions suivantes :

- Il y a eu plus d'hommes décédés presque tout le temps
- L'écart est plus étroit que celui des cas confirmés

2.20 REPARTITION DU SEXE DES SEXE CONFIRMES PAR REGION

A ce niveau nous avons complété la base de donnée par une variable balance qui représente le Ratio hommes/femmes (province/ville/arme/gu). Source : [KOSIS - Korean Statistical Information Service](#).



Source : Réalisée par les auteurs de ce projet

Figure18 : REPARTITION DU SEXE DES CAS CONFIRMES PAR REGION

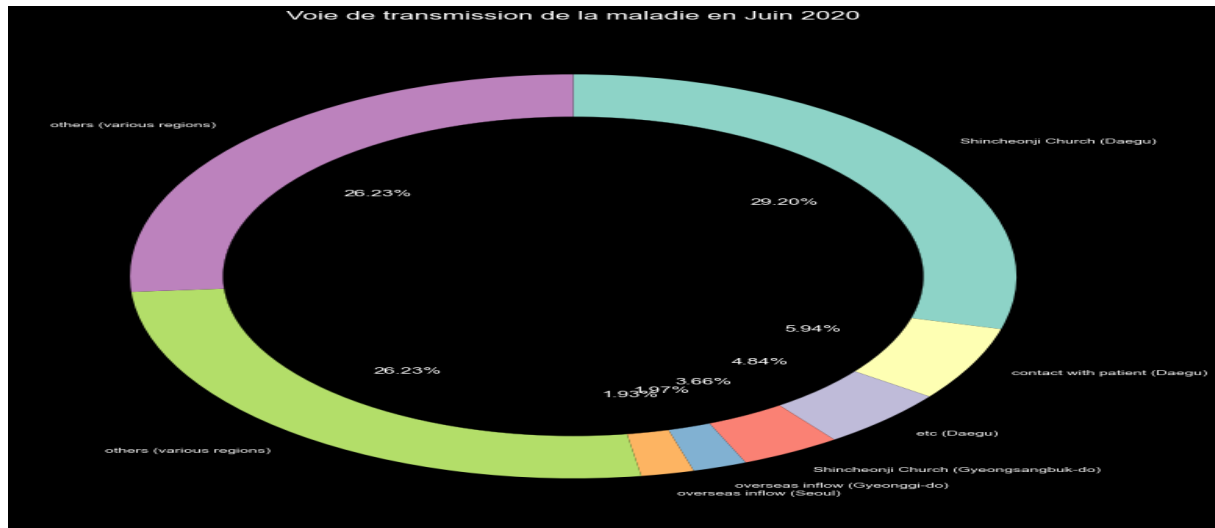
La figure ci-après nous permet de faire des remarques suivantes :

- Le sexe semble équilibré dans l'ensemble et dans chaque région
- Il ne montre aucune corrélation avec le nombre de cas confirmés

3 LES CHAINES DE CONTAMINATION DU COVID EN KOREE DU SUD.

Il s'agit dans cette partie de montrer les voies et moyens par lesquels les populations sont infectées de COVID.

3.1 VOIE DE TRANSMISSION DE LA MALADIE EN JUIN 2020



Source : Réalisée par les auteurs de ce projet

Figure19 : Voie de transmission du COVID-19.

La figure ci-dessus montre Shincheonji Church (in Daegu + Gyeongsangbuk-do) compte environ 45 % de toutes les infections enregistrées dans notre jeu de donnée. De plus , les gens viennent de l'extérieur pour transmettre la maladie aux population

3.2 RESUME DES LIEUX ET CHEMIN DE CONTAMINATION

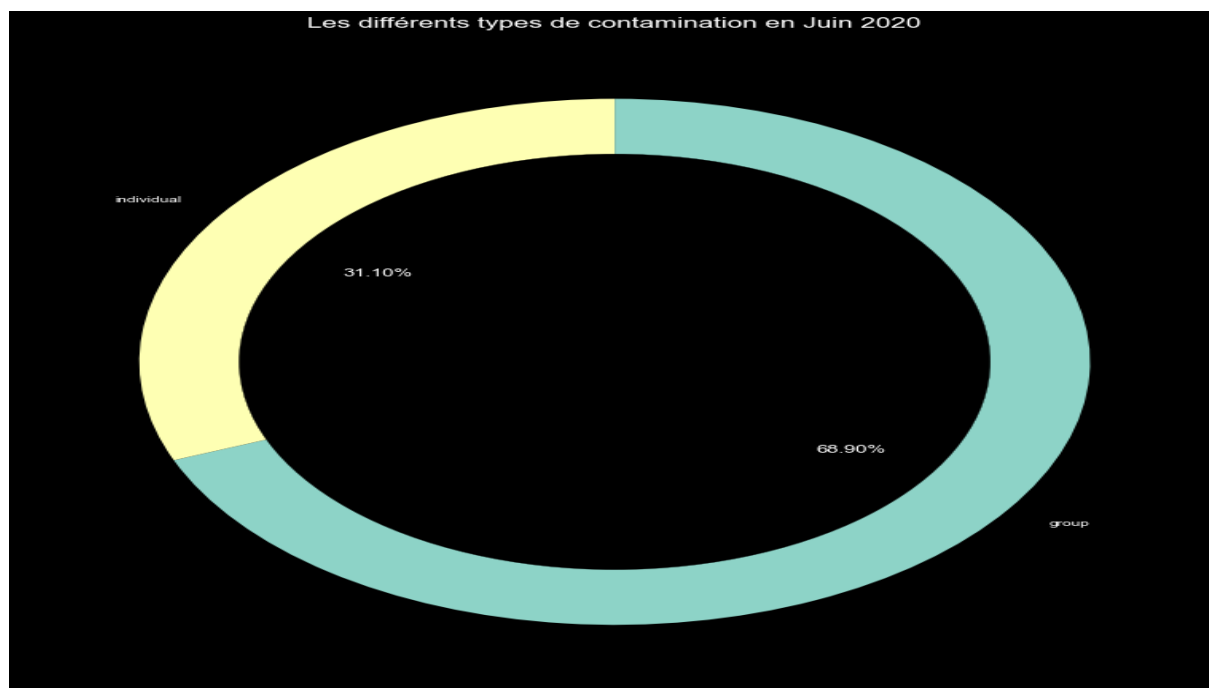


Source : Réalisée par les auteurs de ce projet

Figure20 : RESUME DES LIEUX ET CHEMIN DE CONTAMINATION

La figure montre que le virus est importé de l'extérieur (overseas inflow). Les lieux de contaminations sont essentiellement Church, Center et hospital.

3.3 LES DIFFERENTS TYPES DE CONTAMINTIONS DE COVID-2019



Source : Réalisée par les auteurs de ce projet

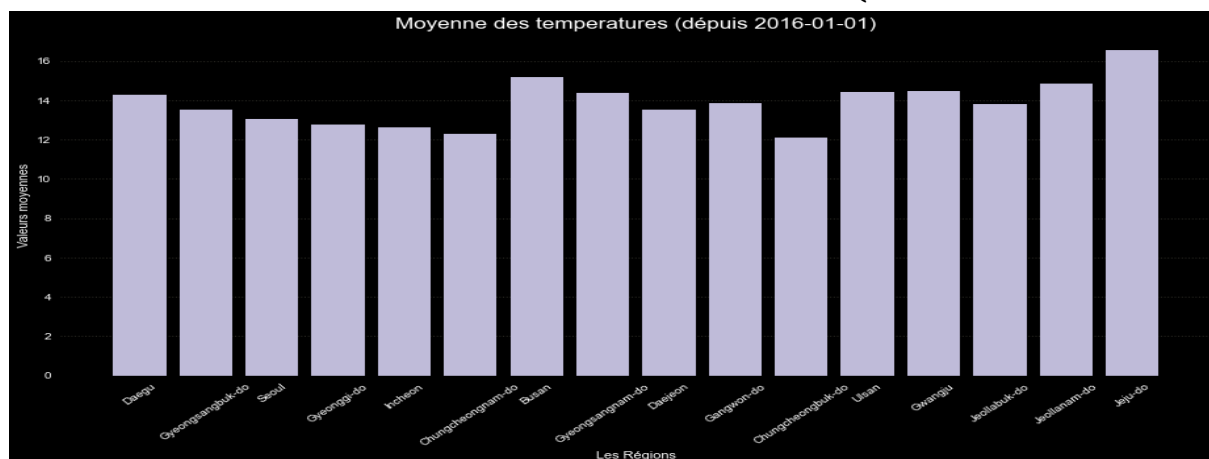
Figure21 : LES DIFFERENTS TYPES DE CONTAMINTIONS DE COVID-2019

La figure ci-dessus nous permet de faire les conclusions suivantes :

- Conformément à ce qui précède (chemin de transmission), environ 70% des transmissions ont eu lieu dans le groupe
- Ce chiffre peut être justifié par des événements de groupe comme les transmissions sur les lieux de travail et les clubs.
- La prévention doit d'abord se concentrer sur les groupes

3.4 ANALYSE METEOROLOGIQUE DU JEU DE DONNEE en 2016

3.4.1 MOYENNE DES TEMPERATURES DANS CHAQUE REGIONS.

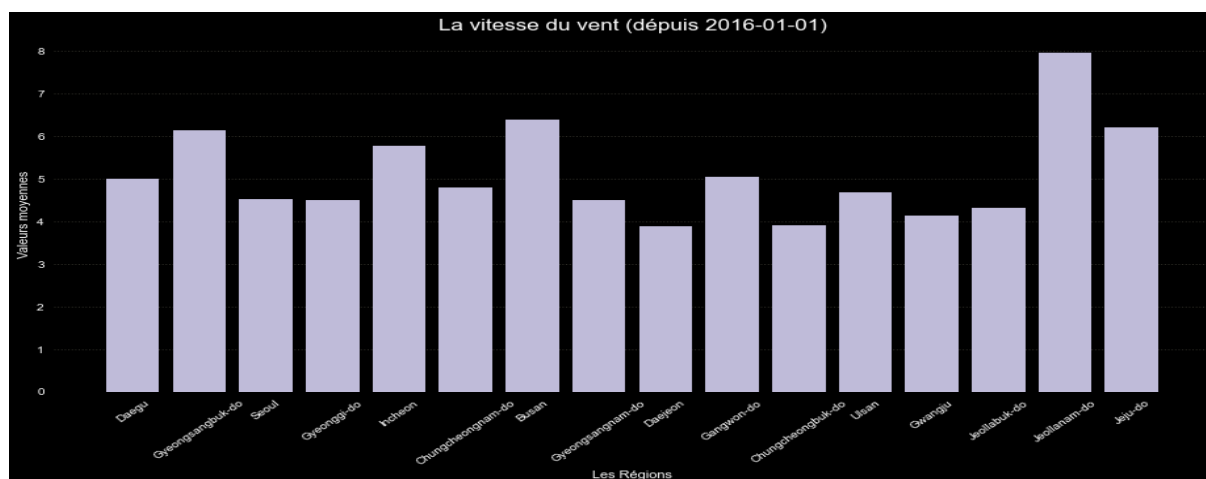


Source : Réalisée par les auteurs de ce projet

Figure22 : MOYENNE DES TEMPERATURES DANS CHAQUE REGIONS.

La figure montre qu'il n'y a pas de fortes corrélations entre les températures dans les différentes zones.

3.4.2 LA VITESSE MAXIMALE DU VENT



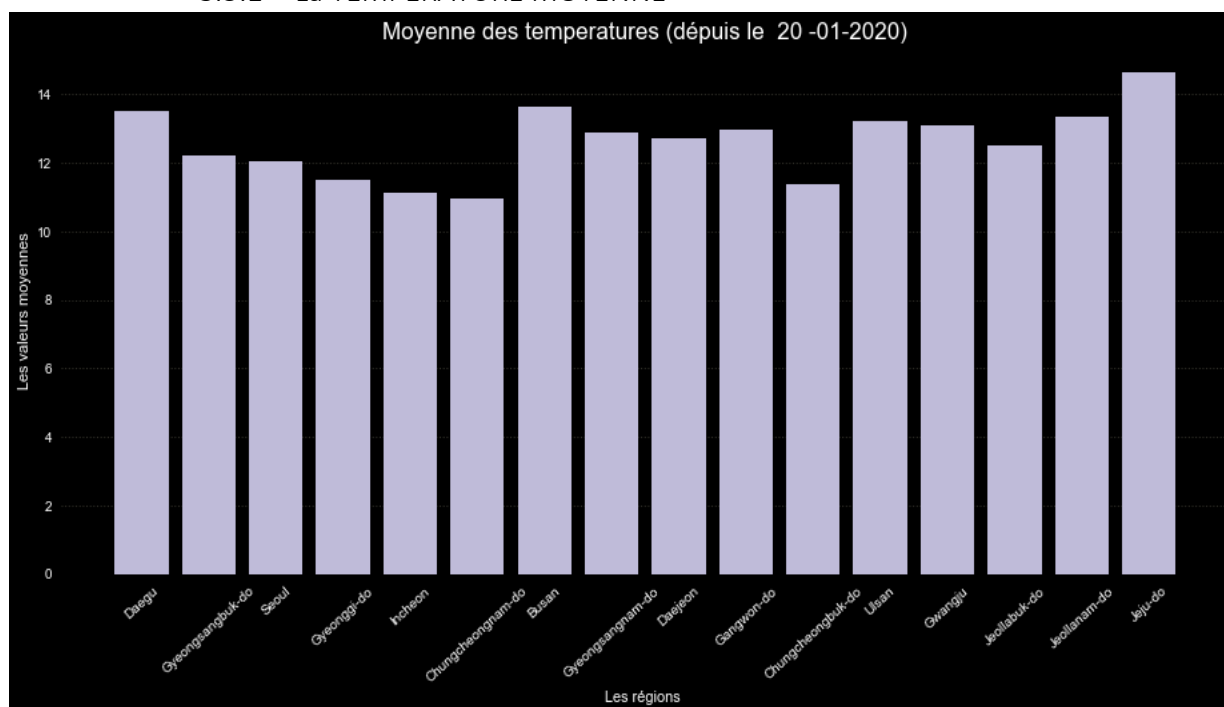
Source : Réalisée par les auteurs de ce projet

Figure23 : La vitesse du vent en Korée du SUD

La figure montre que la vitesse maximale du vent à Jeollanam-do est trop élevée comparativement aux autres régions.

3.5 ANALYSE METEOROLOGIQUE DU JEU DE DONNEE EN TEMPS DE COVID-19

3.5.1 La TEMPERATURE MOYENNE

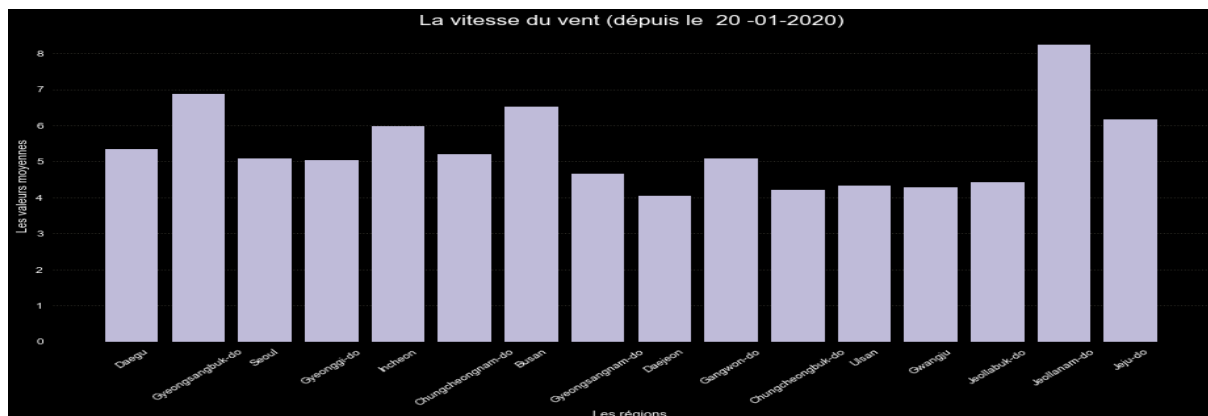


Source : Réalisée par les auteurs de ce projet

Figure24: La TEMPERATURE MOYENNE

La figure montre qu'il y a une faible corrélation négative en général (plus il fait froid, plus il y a de cas).

3.5.2 LA VITESSE MAXIMALE DU VENT

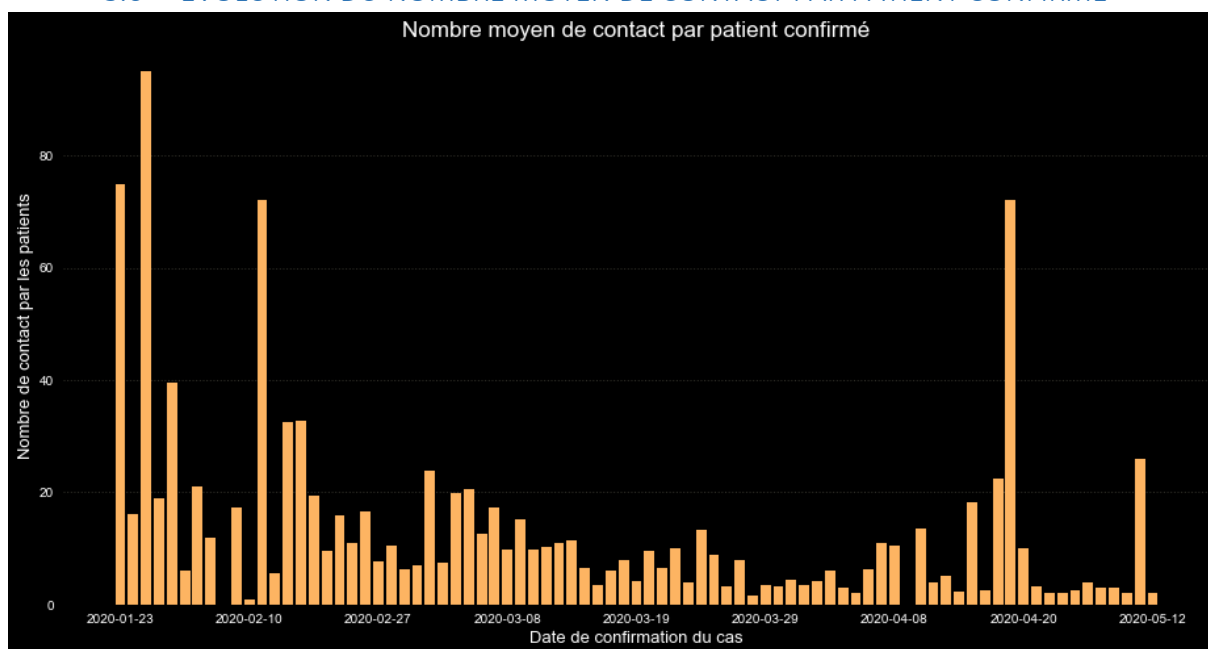


Source : Réalisée par les auteurs de ce projet

Figure25 : LA VITESSE MAXIMALE DU VENT

La figure montre une forte variation de la vitesse du vent en 2020. De plus, on remarque une corrélation positive plus faible (plus il y a de vent, plus il y a de cas).

3.6 EVOLUTION DU NOMBRE MOYEN DE CONTACT PAR PATIENT CONFIRMÉ



Source : Réalisée par les auteurs de ce projet

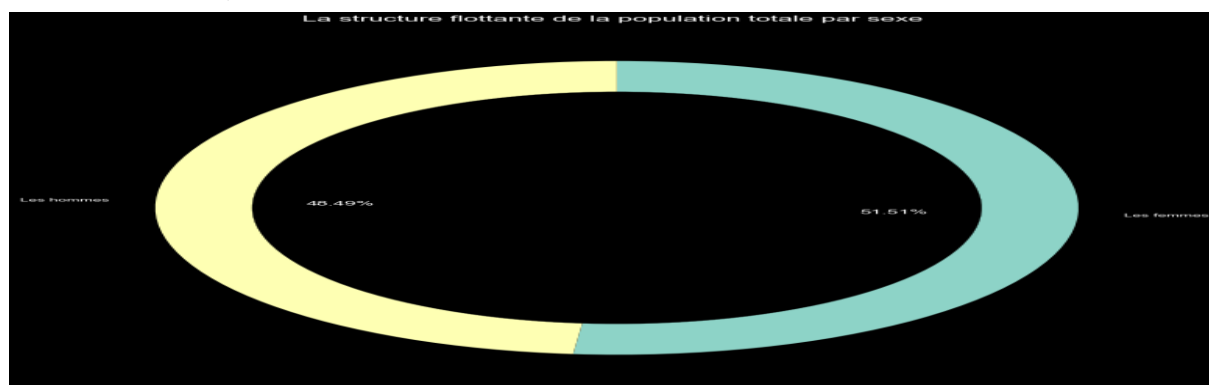
Figure26: EVOLUTION DU NOMBRE MOYEN DE CONTACT

La figure montre qu'en dehors des pics exceptionnels, on remarque une corrélation négative entre les patients confirmés et le nombre de contacts. Cela voudra dire que plus une personne a été confirmée récemment, moins elle a eu de contacts avant le diagnostic.

3.7 ANALYSE GENERALE DE LA POPULATION FLOTTANTE

Il s'agit à ce niveau de déterminer la structure de la population en fonction des heures, des jours des régions et en fonctions d'autres caractéristiques flottante.

3.7.1 LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR SEXE

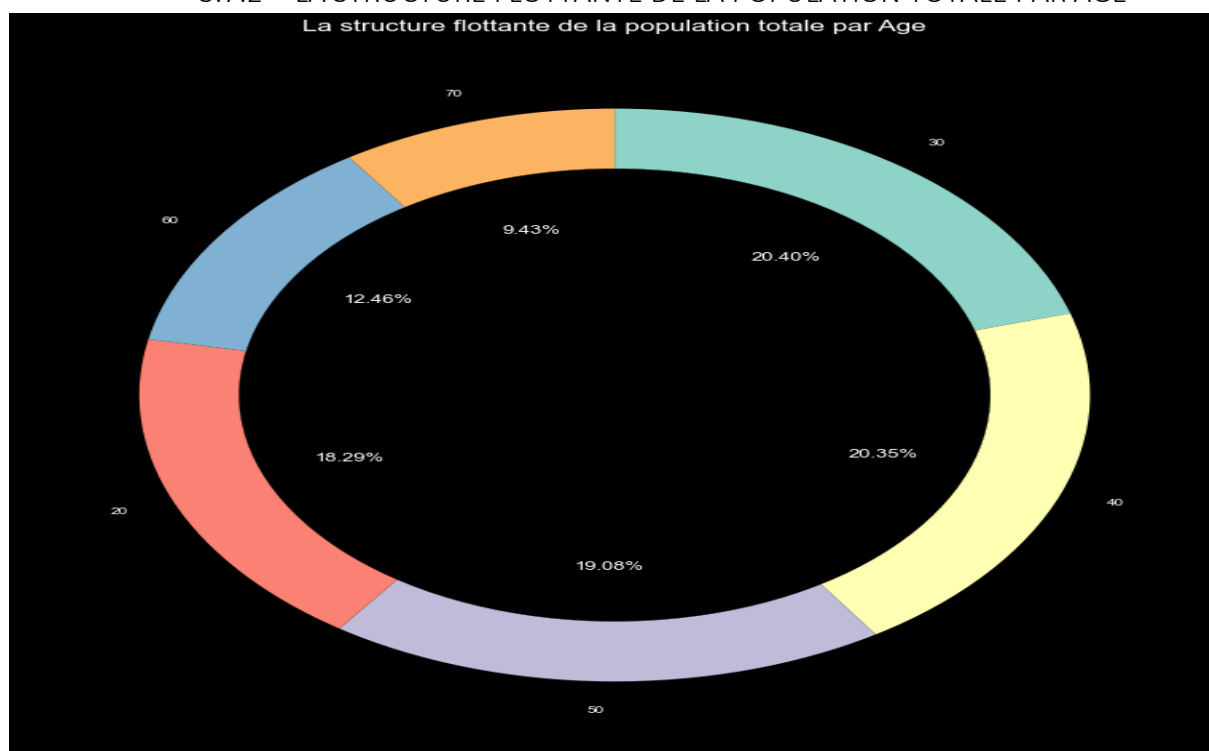


Source : Réalisée par les auteurs de ce projet

Figure27 : LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR SEXE

La figure montre qu'il y a plus de femmes que d'hommes dans la population flottante. Les femmes représentent 51,51 de la population flottante.

3.7.2 LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR AGE

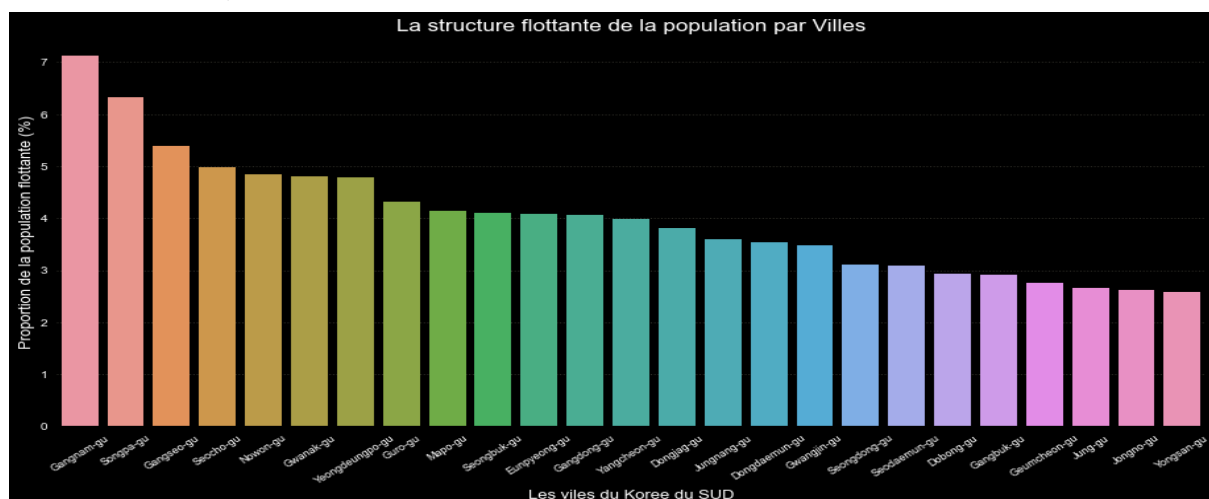


Source : Réalisée par les auteurs de ce projet

Figure28: LA STRUCTURE FLOTTANTE DE LA POPULATION TOTALE PAR AGE

La figure montre que la population flottante commence par 20ans et se termine par 70ans dont 30ans,40ans et 30ans sont plus représentés.

3.7.3 LA STRUCTURE FLOTTANTE DE LA POPULATION PAR VILLES

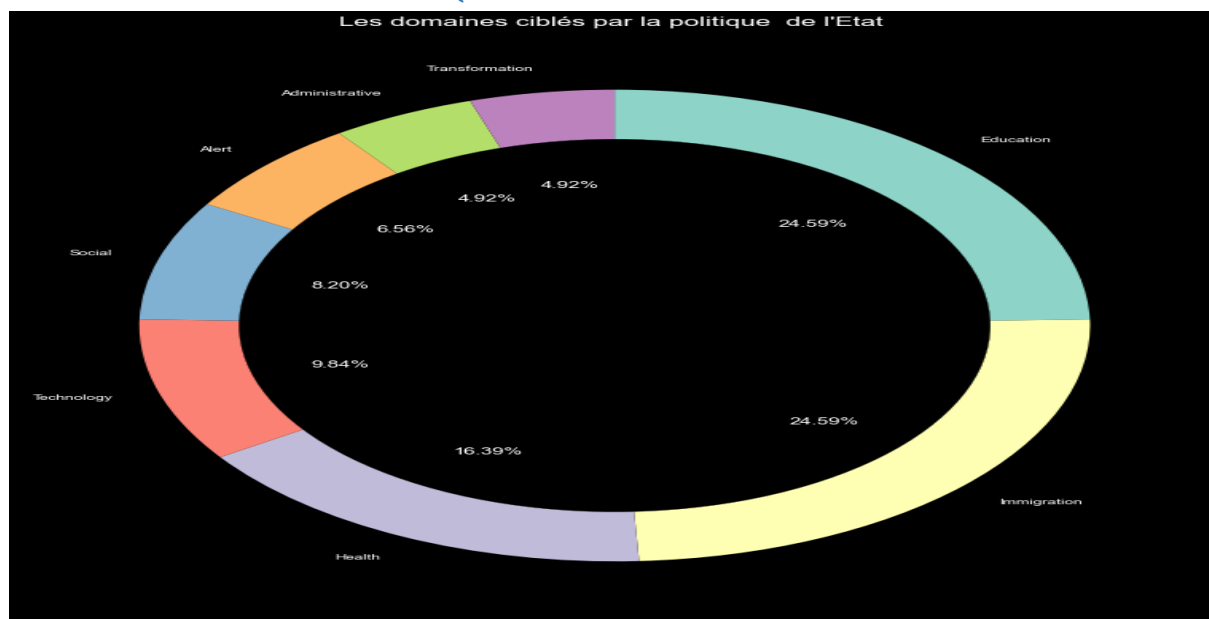


Source : Réalisée par les auteurs de ce projet

Figure29 : LA STRUCTURE FLOTTANTE DE LA POPULATION PAR VILLES

La figure montre qu'il y a une relative corrélation positive entre la circulation dans les villes et le risque d'être infecté. C'est-à-dire que plus il y a de personnes en circulation, plus il peut y avoir d'infections.

3.8 ANALYSE DES POLITIQUES GOUVERNEMENTALES CONTRE COVID-19



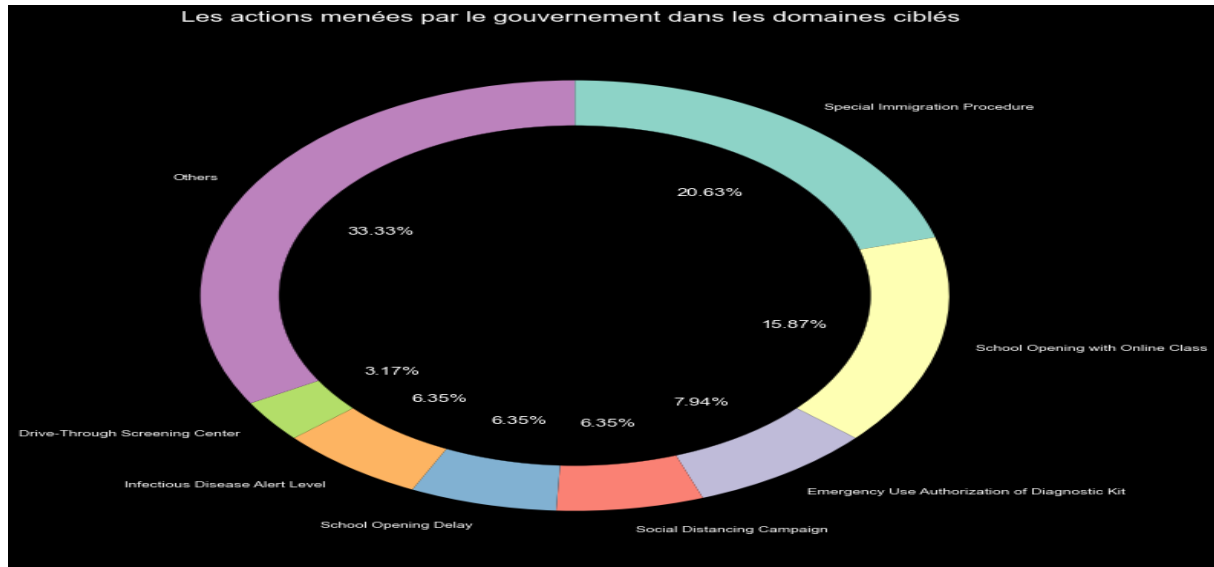
Source : Réalisée par les auteurs de ce projet

Figure30 : Analyse des domaines ciblés par l'Etat Koréen.

La figure nous permet de faire des déductions suivantes :

- Les politiques les plus fréquemment mises en œuvre concernent l'immigration
- Un autre type de politique est celui de l'éducation
- Les troisièmes concernent la santé publique

3.9 LES ACTIONS MENEES PAR LE GOUVERNEMENT DANS LES DOMAINES CIBLES



Source : Réalisée par les auteurs de ce projet

Figure31: LES ACTIONS MENEES PAR LE GOUVERNEMENT DANS LES DOMAINES CIBLES

La figure ci-après nous permet de faire des conclusions suivantes :

- Les politiques d'immigration sont spécifiques à chaque pays.
- Les politiques dans l'éducation concernent les retards de l'ouverture des écoles et ses alternatives (cours en ligne).
- Les politiques dans la santé concernent l'autorisation pour le kit de diagnostic, les campagnes de distanciation sociale et les distributions de masques.

3.10 TENDANCES DES RECHERCHES SUR LES THEMES LIES AUX MALADIES

RESPIRATOIRES AVANT ET APRES LE PREMIER CAS DE COVID EN KOREE DU SUD

A ce niveau, nous avons consacré notre temps à l'analyse des recherches effectuée en Corée du Sud sur certaines maladies respiratoires en utilisant la base searchTrends dans nos jeux de données (Fournisseur de données : Naver (l'un des moteurs de recherche les plus influents en Corée du Sud, Source : [Search engine market share](#)).

3.10.1 EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DURANT TOUTE LA PERIODE DU JEU DE DONNEE.

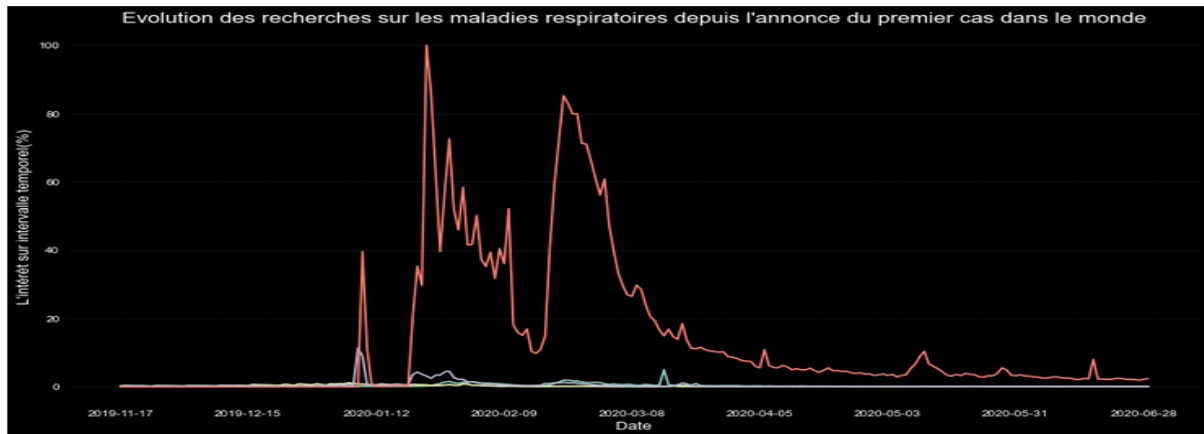


Source : Réalisée par les auteurs de ce projet

Figure32 : Evolution des recherches sur les maladies respiratoires durant toute la période du jeu de donnée.

La figure ci-après montre qu'il y a des pics d'intérêt des recherches sur les maladies respiratoires avant COVID.

3.10.2 EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DEPUIS L'ANNONCE DU PREMIER CAS DANS LE MONDE



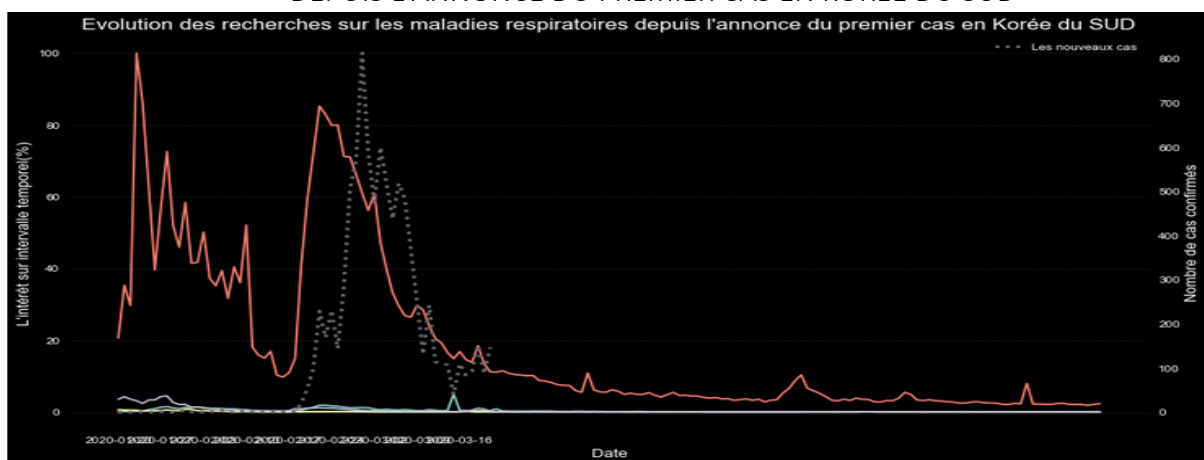
Source : Réalisée par les auteurs de ce projet

Figure33 : Evolution des recherches sur les maladies respiratoires depuis l'annonce du premier cas dans le monde.

La figure ci-après nous permet de faire des conclusions suivantes :

- On note aucun changement distinct depuis environ un mois et demi dans les recherches sur les thèmes des maladies respiratoires
- On note également une augmentation des recherches sur les maladies respiratoires dès que le premier cas de covid est annoncé dans le monde.

3.10.3 EVOLUTION DES RECHERCHES SUR LES MALADIES RESPIRATOIRES DEPUIS L'ANNONCE DU PREMIER CAS EN KOREE DU SUD



Source : Réalisée par les auteurs de ce projet

Figure34 : Evolution des recherches sur les maladies respiratoires depuis l'annonce du premier cas en Corée du Sud.

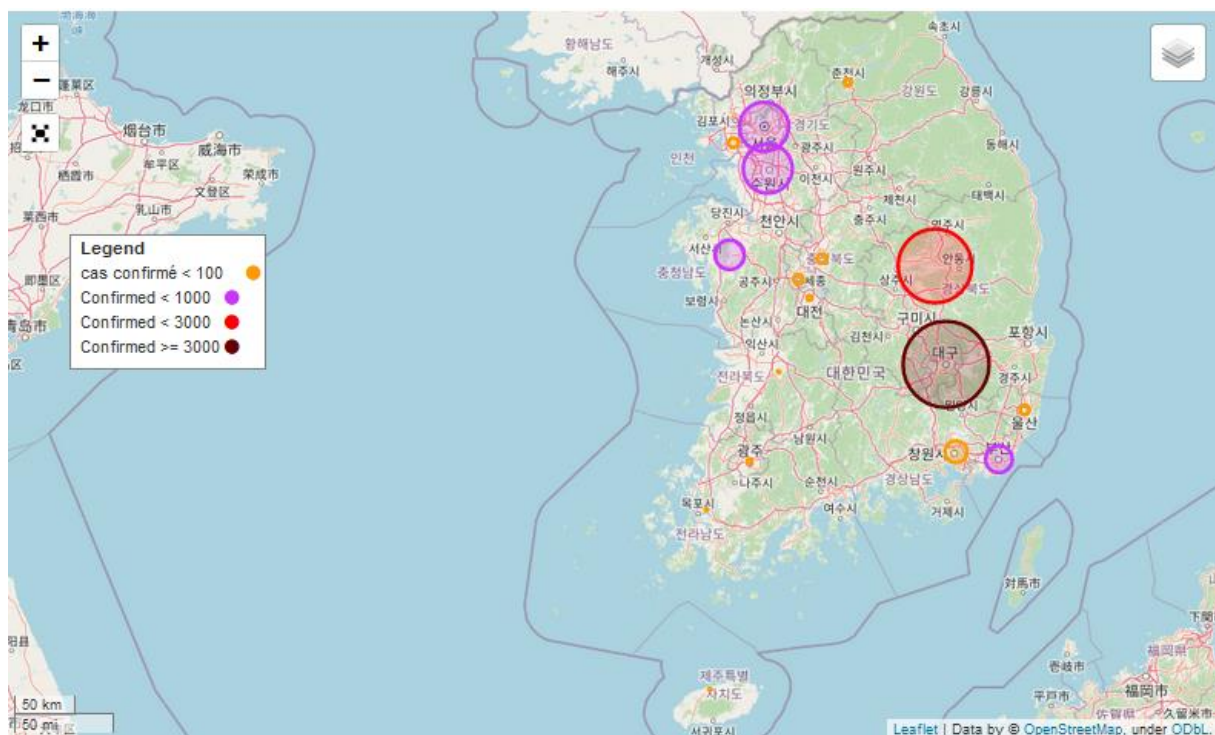
La figure ci-dessus nous permet de faire des remarques suivantes :

- On note un relâchement dans les recherches au début lié peut être à la non prise en compte du danger de corona.
- Les tendances de recherche présentent une allure impressionnante avec la forme des cas confirmés quotidiennement depuis le 21 février 2020.

4 ANALYSE GEOSPATIALE DE LA PANDEMIE DE COVID-19

Nous aimerions signaler qu'une partie de nos données provient des données mondiales sur le covid-19 pour nous aider à réussir la comparaison des tendances au niveau mondiale.

4.1 ANALYSE GEOSPACIALE DES STATUTS DES PATIENTS DE COVID-2019 EN KOREE DU SUD

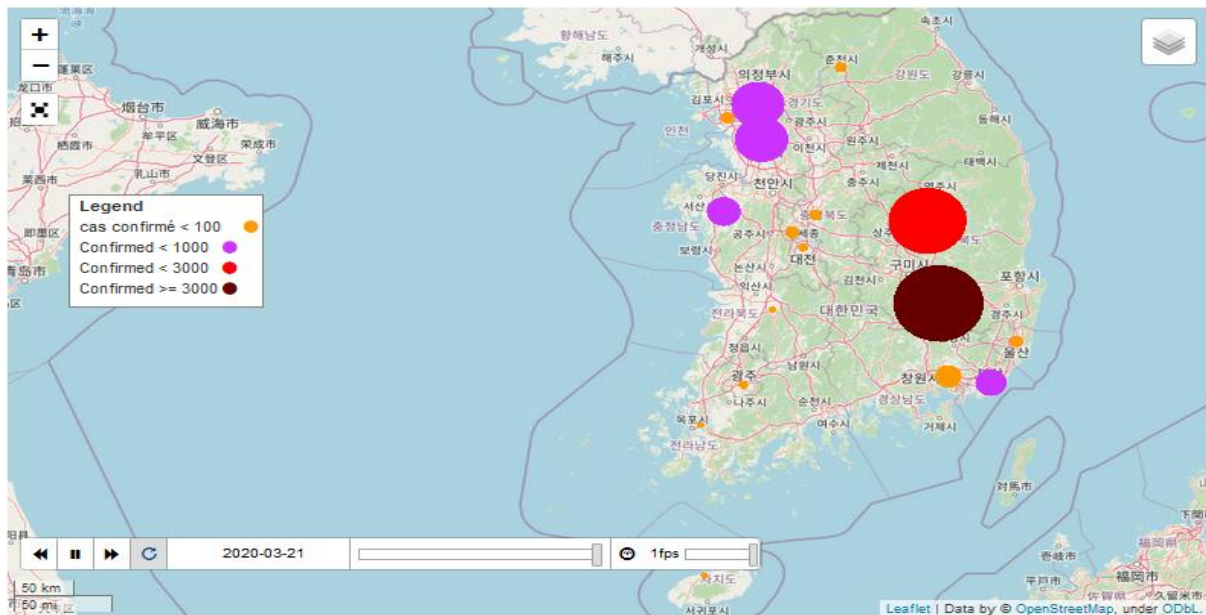


Source : Réalisée par les auteurs de ce projet

Figure35: ANALYSE GEOSPACIALE DES STATUTS DES PATIENTS DE COVID-2019 EN KOREE DU SUD

L'analyse géospatiale révèle que c'est au Daegu en Corée du SUD que le corona a commencé dans ce pays avec 6387 de cas confirmés

4.2 ANALYSE DE LA CHRONOLOGIE DE L'ÉVOLUTION DU COVID-2019 EN KOREE DU SUD



Source : Réalisée par les auteurs de ce projet

Figure36 : Suivi chronologique de l'évolution du COVID dans chaque région

Signalons que ce graphique différent du premier graphique géospatial est dynamique et suit l'évolution de la maladie dans chaque région du Korée du Sud. Cependant la capture d'écran que nous avons effectuée a désactivé l'option vidéo du graphique. C'est pourquoi nous tenons à informer tout lecteur de ce projet. En effet la zone noire montre que la situation de COVID-2019 est extrêmement grave dans la région et il faut adopter des mesures exceptionnelles pour ralentir le rythme de croissance de la maladie.

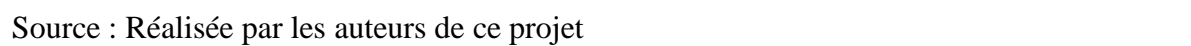
4.3 ANALYSE DES CAS CONFIRMES A TRAVERS UN TREEMAP



Source : Réalisée par les auteurs de ce projet

Figure37: ANALYSE DES CAS CONFIRMES A TRAVERS UN TREEMAP

4.4 ANALYSE DES CAS DELIVRES A TRAVERS UN TREEMAP



La figure montre que les régions telles Daegu et Gyeongsangbuk-do ont environs 79% des

4.5 ANALYSE DES CAS DE DECES A TRAVERS UN TREEMAP

Daegu 997	Gyeongsangbuk-do 363
--------------	-------------------------



Source : Réalisée par les auteurs de ce projet

Figure40: ANALYSE DES CAS DE DECES A TRAVERS UN TREEMAP

La figure montre que les régions telles Daegu et Gyeongsangbuk-do ont environs 95% des cas de décès de covid.

4.6 ANALYSE COMPARATIVE DES CAS DE COVID EN KOREE DU SUD ET AU NIVEAU MONDIALE.

4.6.1 POSITION GEOSPATIALE DE L'EVOLUTION DU COVID-2019 AU NIVEAU MONDIALE

Nous avons utilisé les données complète sur covid au niveau mondiale disponible sur le site de Kaggle.

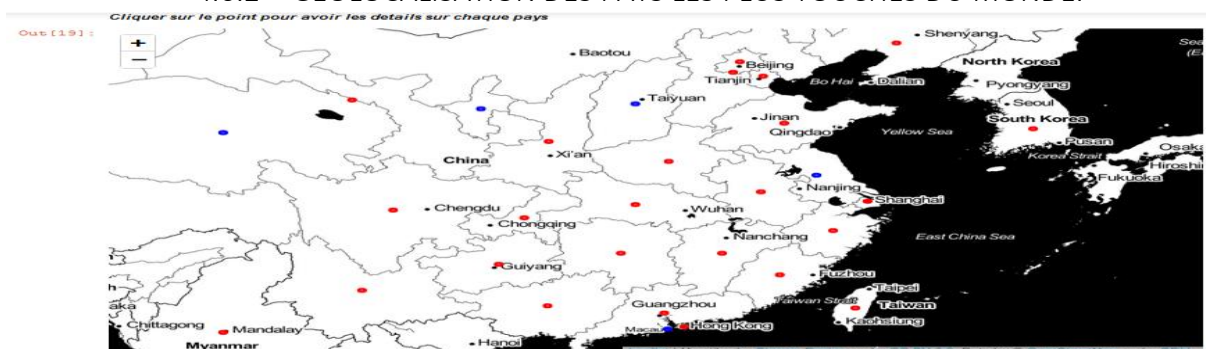


Source : Réalisée par les auteurs de ce projet

Figure41: POSITION GEOSPATIALE DE L'EVOLUTION DU COVID-2019 AU NIVEAU MONDIALE

La figure montre que le Korée du SUD fait partie des pays à haut risque ou la situation du COVID-19 est dangereuse avec 10591 de cas confirmés et 225 morts. Le rouge matérialise les cas graves dans ces pays.

4.6.2 GEOLOCALISATION DES PAYS LES PLUS TOUCHES DU MONDE.



Source : Réalisée par les auteurs de ce projet

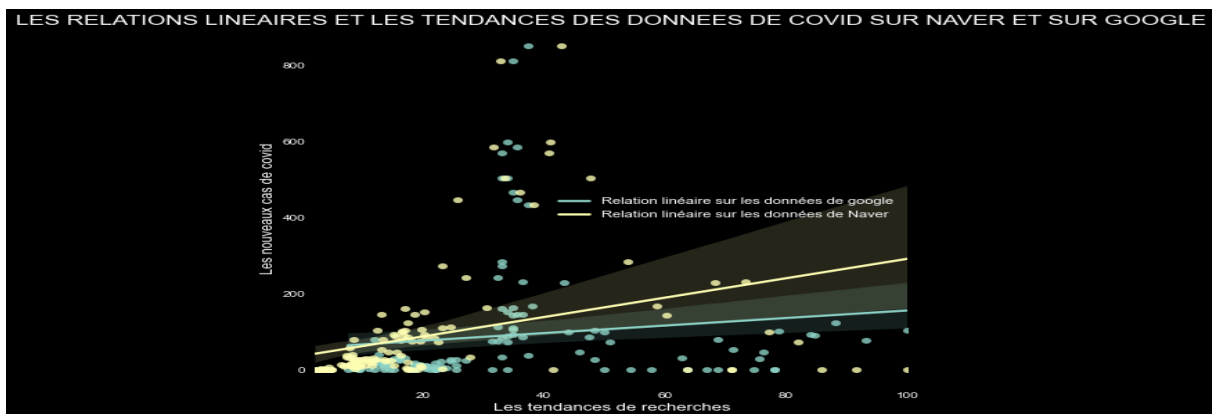
Figure42: Géolocalisation des pays les plus touchés du monde.

La figure confirme encore que la situation en Korée du sud est préoccupante par ce qu'il figure parmi les pays à hauts risques.

5 PREDICTION DES CAS DE COVID EN KOREE DU SUD

Nous avons appris dans l'analyse exploratoire que plus le nombre de cas de covid devient élevé, plus les recherches sur les maladies respiratoires s'intensifient (D'après les analyses faites sur les données fournies par le fournisseur Naver). En effet nous aimerions approfondir nos analyses sur celle des tendances de recherches fournie par le fournisseur Google (Source : [Google trends](#)).

5.1 NATURE DE RELATION ENTRE LES TENDANCES DE RECHERCHES ET LES CAS CONFIRMES DE COVID

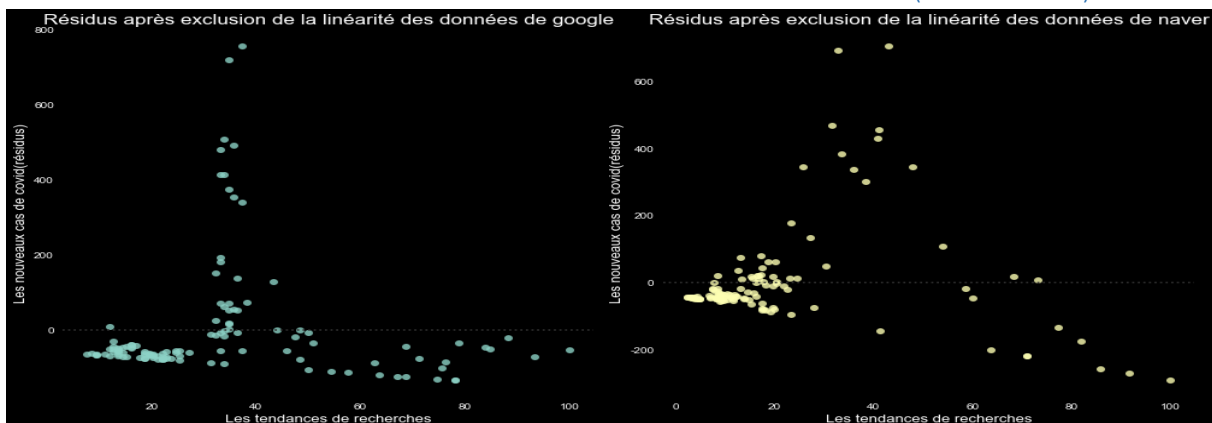


Source : Réalisée par les auteurs de ce projet

Figure43 : Nature de relation entre les cas de covid et les tendances.

La figure montre que la relation semble être plus linéaire pour les données de Naver que pour les données de Google.

5.2 EXTRACTION DE LA LINEARITE DANS LES DEUX DONNEES (LES RESIDUS)



Source : Réalisée par les auteurs de ce projet

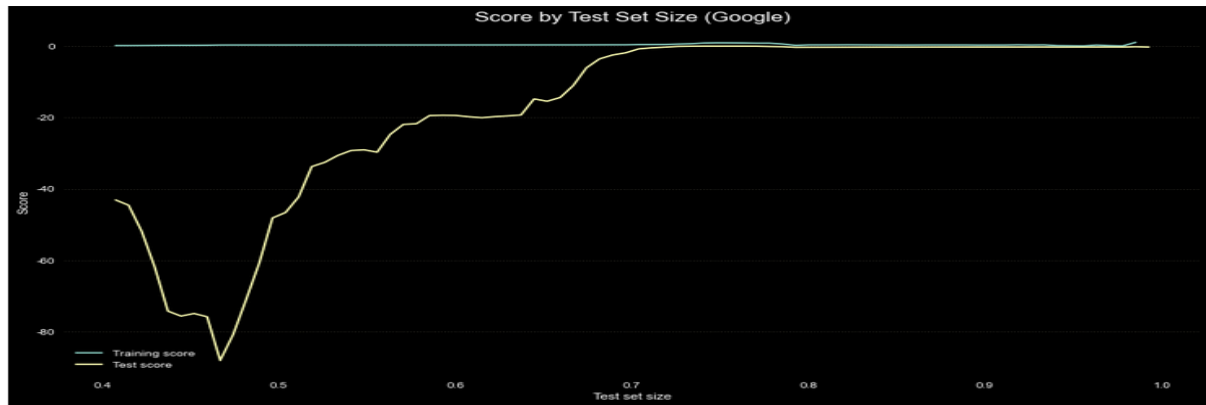
Figure44: Les résidus des deux données Google et Naver.

La figure montre que les résidus des deux données ne suivent pas une distribution gaussienne et ne sont pas stables dans le temps. Alors une simple relation linéaire ne peut pas expliquer la relation entre les tendances des recherches sur le corona et le nombre de cas confirmés.

5.3 REGRESSIONN LINEAIRE

A ce niveau, nous aimerions prédire le nombre de cas confirmés en Korée du SUD à partir des données des tendances de corona fournies par Google et Naver. La variable à expliquer est le nombre de cas de covid obtenu à différentes dates.

5.3.1 SCORE COMPARE DES DONNEES TRAIN/TEST SUR LES DONNEES DE GOOGLE

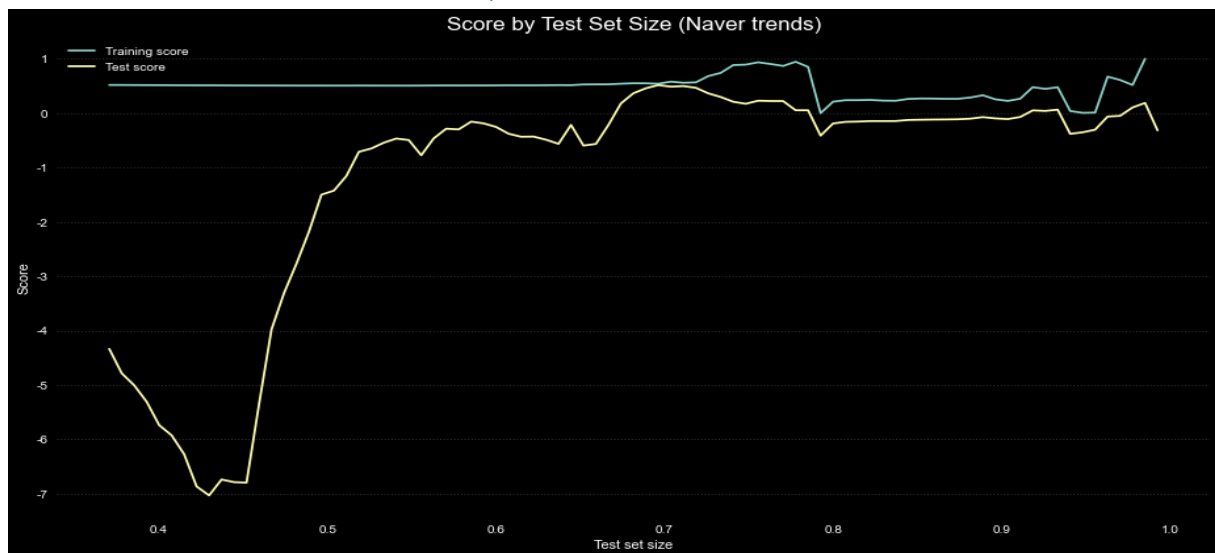


Source : Réalisée par les auteurs de ce projet

Figure45 : Score comparé sur les données de Google

La figure montre que la régression linéaire sur les données de Google sont défailantes. De plus le score (-0,0650) sur les données Test est trop faible et le modèle n'apprend presque rien.

5.4 SCORE COMPARE TRAIN/TEST SUR LES DONNEES DE NAVER

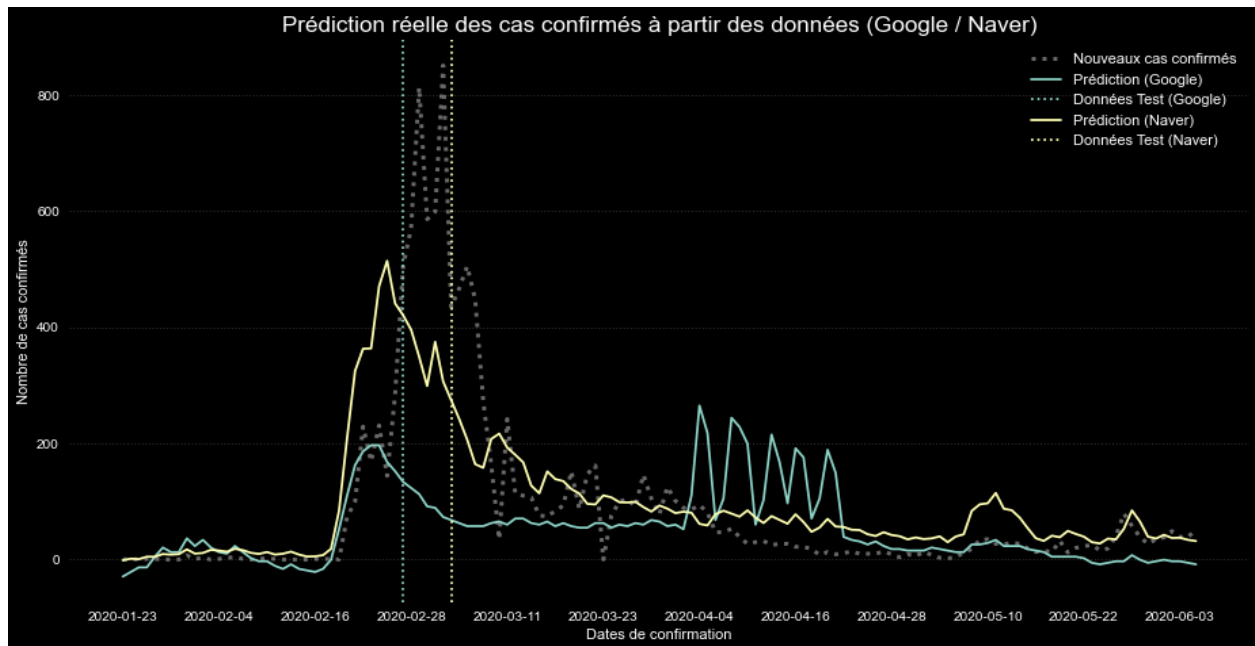


Source : Réalisée par les auteurs de ce projet

Figure46 : Score comparé sur les données de Naver

La figure montre que le score (0,5178) du test est meilleur que celui des données Google.

5.4.1 PREDICTION REELLE DES CAS CONFIRMES SUR LES DONNEES DE GOOGLR/NAVER



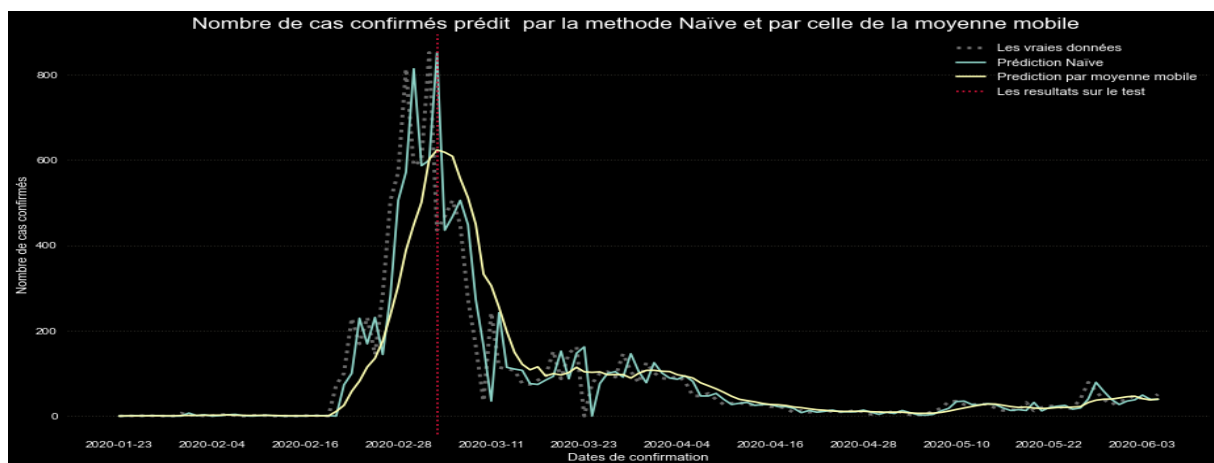
Source : Réalisée par les auteurs de ce projet

Figure47 : PREDICTION REELLE DES CAS CONFIRMES SUR LES DONNEES DE GOOGLR/NAVER

LA FIGURE CI-DESSUS NOUS PERMET DE FAIRE DES CONCLUSIONS SUIVANTES/

- Les tendances de recherche sur les données de Google rendent les prédictions non pertinentes. Cela peut être dû aux fluctuations des données sur ce site.
- Les prédictions sont meilleures en test qu'en apprentissage sur les données de Naver.

5.5 PASSAGE AU METHODE DE PREDICTION NAIVE ET AU METHODE DE PREDICTION PAR MOYENNE MOBILE.

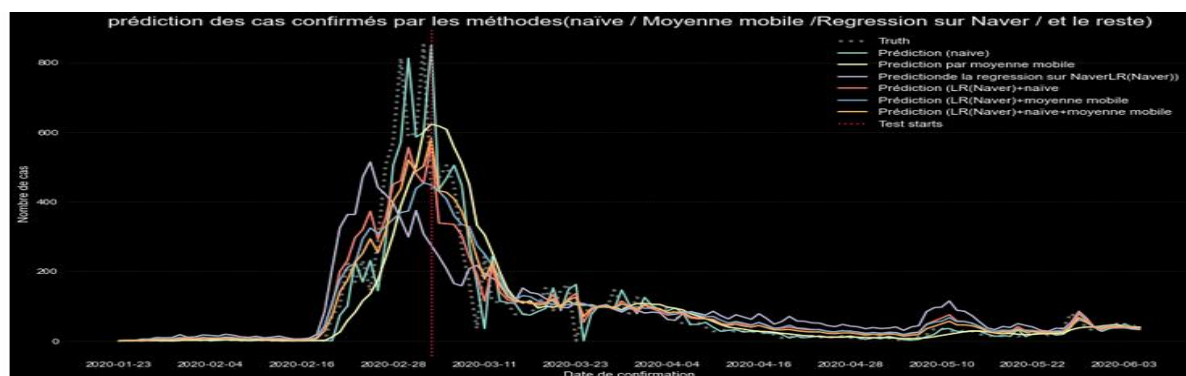


Source : Réalisée par les auteurs de ce projet

Figure48 : METHODE DE PREDICTION NAIVE ET AU METHODE DE PREDICTION PAR MOYENNE MOBILE.

La figure montre que la méthode naïve a un meilleur score (0,5989) sur les données de Test que celle de la moyenne mobile. Donc la méthode naïve prédit mieux les cas confirmés.

5.6 COMPARAISON DES DIFFERENTS MODELES ENTRAINES POUR PREDIRE LE NOMBRE DE CAS CONFIRMES DE COVID



Source : Réalisée par les auteurs de ce projet

Figure49: COMPARAISON DES DIFFERENTS MODELES ENTRAINES

La figure ci-dessus nous permet de faire des déductions suivantes :

- La méthode naïve (score sur le test=0,5989) fonctionne mieux que la méthode par moyenne mobile (score sur le test=0,5772) mais d'une moindre mesure
- Le modèle de régression linéaire sur les tendances de recherche Naver montre un score comparable à celui de la moyenne.
- Lorsque nous combinons les 3 prédictions, nous obtenons un meilleur score (score sur le test=0,8092)

6 RECOMMANDATION DES INITIATIVES POUR CASSER LES CHAINES DE CONTAMINATION EN KOREE DU SUD

6.1 POLITIQUE ENVERS LES GROUPES D'ÂGE EN KOREE DU SUD.

On a vu dans nos analyses que les jeunes de 20 ans sont le groupe d'âge le plus infecté (27,80 % du total) et que les plus de 70 ans sont le groupe d'âge le plus décédé (48,08 % du total). Il faut alors que les pays séparent les deux groupes tout en prévenant la propagation dans son ensemble.

6.2 LES ACTIONS D'ORDRE REGIONAL

Les analyses ont révélé que plus la population est nombreuse dans la région, plus il y a de cas en général. De plus, on a constaté beaucoup de cas dans les églises en Corée du SUD donc les rassemblements communautaires tels que les événements religieux sont plus risqués. Jusqu'à 60 % des cas semblent avoir été transmis par un rassemblement spécifique (l'église Sinchonji à Daegu). Il faut de façon urgente que les autorités dans ces régions préviennent les églises du danger et détectent le risque en cas de réalisation d'un cas de covid.

6.3 POLITIQUE ENVERS LES RECHERCHES SUR LES MALADIES RESPIRATOIRES

On a constaté avec nos analyses que les tendances de recherches sur les maladies respiratoires existaient avant COVID mais pas assez. En effet, nous exhortons le gouvernement Koréen à investir considérablement dans ces recherches en finançant par exemple le fournisseur des

données fiables Naver ou Google. Nous aimerions signaler au lecteur que le gouvernement a pris des mesures dans les domaines tels que l'éducation, la santé et autres qui sont salutaires. Par exemple, les mesures éducatives sont : Retard dans l'ouverture des écoles, cours en ligne. De plus, les politiques d'immigration ont été mises en œuvre dans les régions. Cependant, nos analyses ont révélé que toutes les mesures prises par le gouvernement sont effectuées sans la fermeture des frontières. En effet, nous exhortons le gouvernement Koréen à accompagner les mesures de la fermeture des frontières pour éviter les chaînes de contamination d'ordre étranger.