

Exploration des données du COVID-19 (SarS-CoV-2)

Nom / Prenom	Numéro Etudiant
Sheikh Rakib	11502605
Daudin Louise	11606555

Remerciements

Nous tenons principalement à remercier le gouvernement sud-coréen pour la mise à dispositions des données de leur patients en open data qui est disponible (à cette adresse) [à cette adresse](#) et [à cette adresse officiel](#)

Introduction

Ceci est un rapport dans un contexte du devoir universitaire pour l'Institut Galilée pour le module nommé Visualisation de données supervisé par Monsieur GROZAVU Nistor, enseignant rattaché au laboratoire LIPN.

Les données utilisées proviennent de la mise à disposition en open data des données détaillées et anonymisés de la pandémie actuelle nommée Covid-19 (Virus SarS-Cov-2) plus particulièrement connu sous le nom médiatique "Corona Virus".

Une série de graphique sera introduit au fur et à mesure de l'analyse de ces données suite à l'exploration des données par l'outil KNIME.

Etant donnée de l'évolution rapide de l'épidémie, dans le cas où le lecteur tente de reproduire les résultats obtenus, la date de la production des données pour le-dit projet est le 1er Juin 2020.

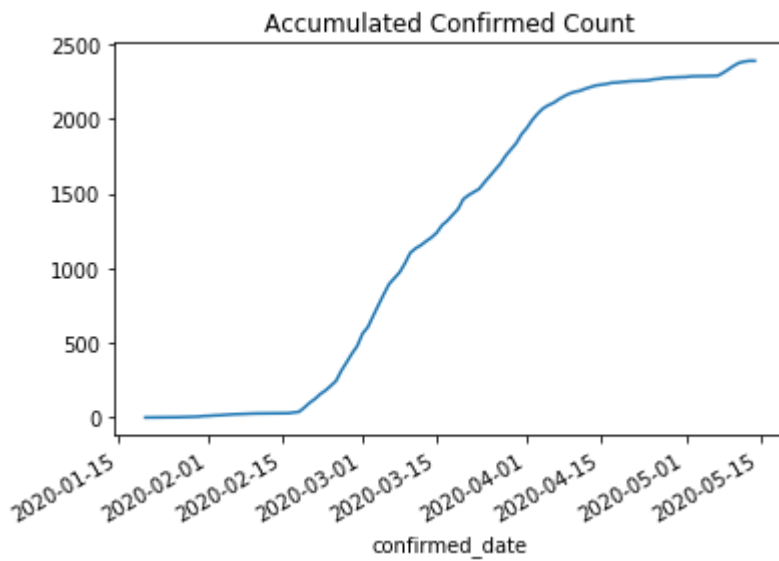
Au mois de Décembre 2019, une épidémie d'une maladie émergente (Covid-19) due au SARS-CoV-2 a débuté à Wuhan, en Chine, puis s'est rapidement propagée en Chine, et a été déclarée pandémie le 12 mars 2020 par l'OMS. Actuellement, le taux global de mortalité est d'environ 2,3 % en Chine, ce qui est probablement une surestimation car la plupart des patients présentent des symptômes bénins et ne sont donc pas testés. Ce virus s'est propagé dans le monde entier au fil du temps et les Etat-Unis est le pays devenu le cluster principal de la pandémie, suivi de l'Europe. La Chine à reconstruit une forte baisse de nouveau cas de la pandémie à partir du 18 Février 2020.

Les symptômes du Covid-19 sont identifiés principalement par les infections des voies respiratoires, les symptômes sont divers et variés. Les patient ont principalement de la fièvre supérieure à 39°C, toux, courbatures, asthénie, céphalées, myalgies, rhinorrhée, dysfonction gustative, perte de l'odorat, obstruction nasale.

Ce devoir pour le cours de Visualisation de Données se concentrera sur l'évolution de la pandémie au sein de la Corée du Sud avec les données de la KCDC (Korea Centers of Disease Control & Prevention).

Exploration initiale des données et interpretation graphique

Nous allons commencer l'analyse avec un graphe mettant en évidence l'évolution de la pandémie du nombre de cas confirmée cumulé au fil du temps.



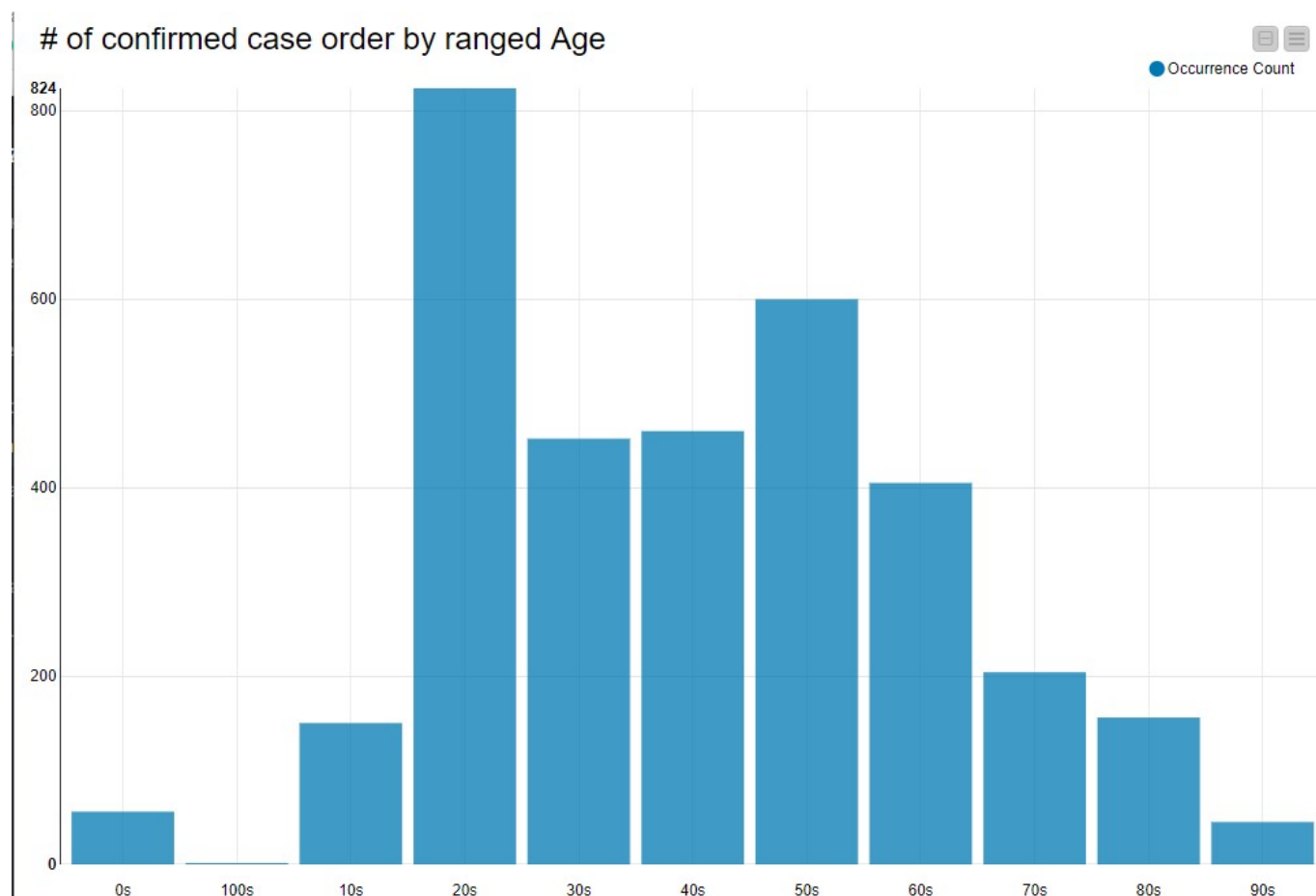
Nous remarquons qu'il y a trois stade

d'évolution selon le graphique :

- Début de l'épidémie
- Propagation rapide de l'épidémie
- Stagnation de l'épidémie

Nous souhaitons connaître le taux de cas positifs en fonction de leurs âge (par dizaine d'années) afin de nous donner une idée globale de la tranche d'âge des personnes qui pourraient être considérées les plus vulnérables.

Nous supposons tout d'abord (dans le bût de l'analyse) que la tranche d'âge les plus vulnérables soient les nourrissons (0-10 ans) et les personnes âgées (50+).



D'après le graphique de données brutes générées, nous remarquons que les personnes ayant une tranche d'âge de 20 ans sont les plus touchées par cette pandémie avec un chiffre total de 824. De plus, le nombre total des personnes infectées pour les plus de 50 ans est de 1366 avec le calcul suivant

$$600 + 405 + 204 + 156 + 1 = 1366$$

Tandis que pour les moins de 50 ans, ce chiffre s'élève à 1942 par le calcul suivant :

$$56 + 150 + 824 + 452 + 460 = 1942$$

Ces chiffres démontrent que la contamination du virus ne dépend absolument pas de la tranche d'âge liée à l'immunologie de ces personnes.

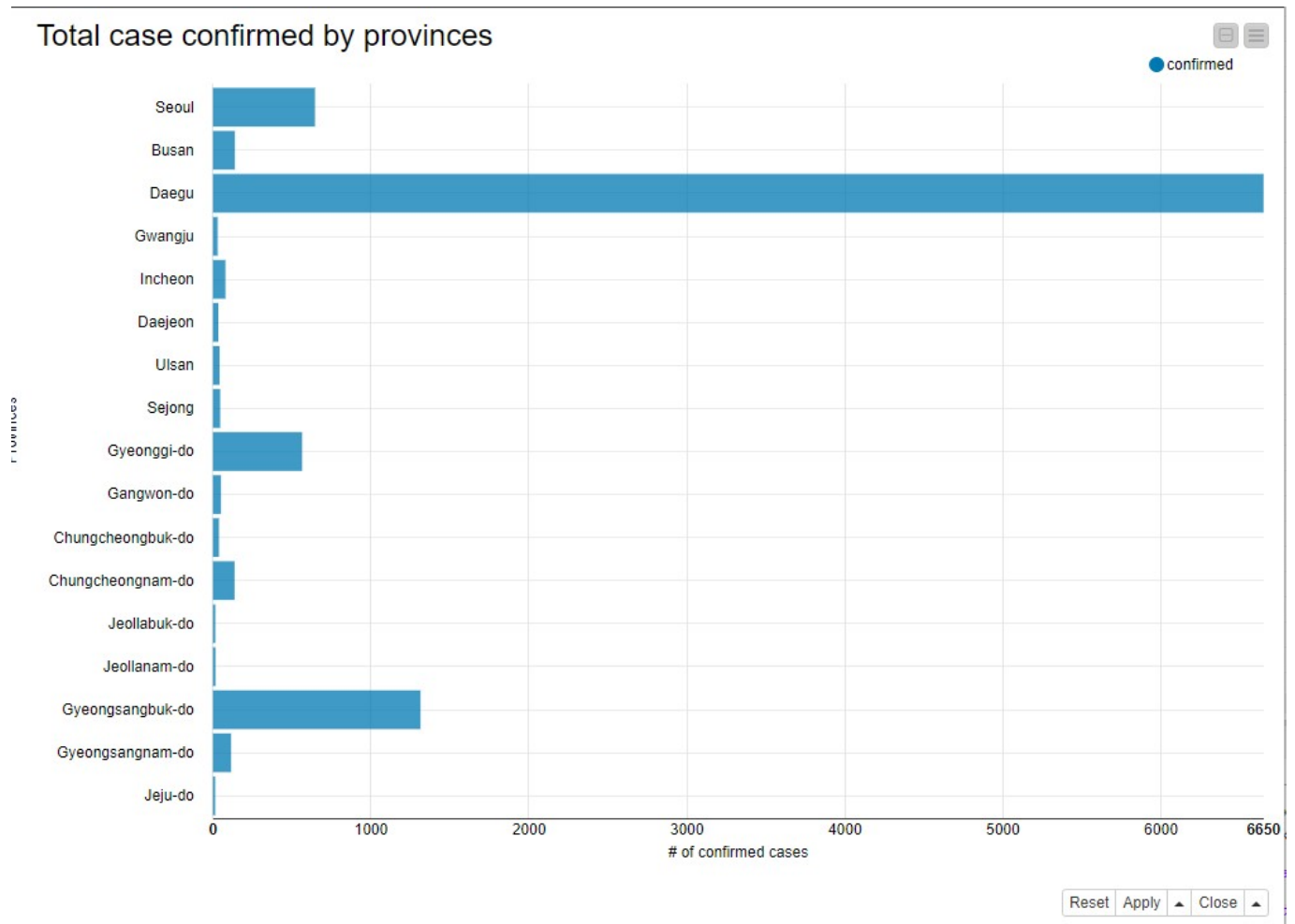
Il est important de noter que le nombre de nourissons infectés (tranche d'âge 0s) est quand même de 56. Nous nous intéressons donc aux dates auxquelles ces personnes ont été déclarées positives au virus.

A titre de référence d'après les données, la Corée du Sud a déclaré le premier patient positif le 23 janvier 2020.

Le premier cas du nourisson a été déclaré le 24 Février 2020 avec le motif suivant : Contact avec un patient. (Patient : 2000000027)

En remontant la trace de ce patient, le motif principal de cette infection est dû à l'Eglise Shincheonji Church par l'intermédiaire de deux hôtes. (Le premier hôte 2000000013 sûrement la mère du nourrisson, classé 30s en âge, et le deuxième 1200000031 classé 60s en âge, issu de l'église en question).

- Nous souhaitons connaître également la répartition géographique par province de l'épidémie des cas confirmés.



Graph 1 : Total case confirmed by provinces

D'après les résultats de cette analyse, nous remarquons que la province de Daegu est la plus touchée avec un nombre total de 6650 cas confirmés. (Données récoltées et générées le 1er juin 2020).

La province de Daegu est composée de 2,463 millions d'habitant (2017) avec une densité de 2812/km² (pour une superficie de 883.5 km²). Ce qui pourrait supposer la propagation du virus favorable à une vitesse considérable (exponentielle).

Nous pouvons supposer que c'est essentiellement aux personnes qui pourraient s'enfuir de la Chine ou bien des expatriés coréens qui sont de retour. Le point de contact principal probable serait l'aéroport International de Daegu de avec les contacts des surfaces et aussi l'échange verbal entre autrui.

Nous disposons néanmoins des données sur la source de la contamination pour les provinces et on peut affirmer que les sources de contamination pour la province de Daego sont les suivantes :

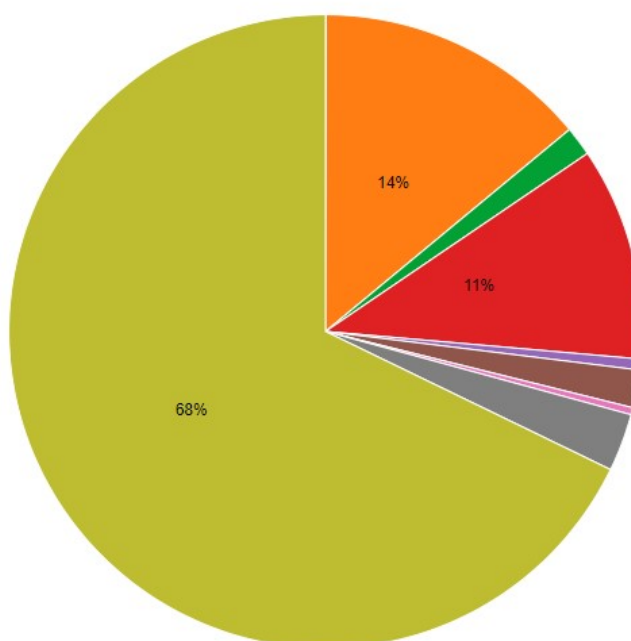
- L'Eglise de Shincheonji
- Deuxième Hopital de Mi-Ju
- L'Hopital Hansarang Convalescent

- L'Hopital Daesil
- Fatima Hospital
- Cheongdo Daenam Hospital
- Etrangers de passages
- Contact Patient

Afin de confirmer ou non notre supposition, nous nous intéressons sur la répartition du nombre de cas confirmés entre ces sources de contamination.

Percentage of confirmed case of the Daegu province source origin

● Cheongdo Daenam Hospital
 ● contact with patient
 ● Daesil Convalescent Hospital
 ● etc
 ● Fatima Hospital
 ● Hansarang Convalescent Hospital
 ● overseas inflow
 ● Second Mi-Ju Hospital
 ● Shincheonji Church



Graphe 2 : Percentage of confirmed case of the Daegu province source origin

Avec 68 % du nombre de personnes infectée ayant comme source d'infection l'Eglise de Shincheonji, on peut affirmer que c'est l'Eglise qui est la principale source d'infection au sein de la province de Daegu.

Et c'est seulement avec 0.3 % du nombre de personnes infectées par une entrée sur le territoire depuis l'extérieur ce qui contredit absolument la supposition posée concernant l'entrée sur le territoire sur la raison de la propagation, même si cela reste la raison principale de l'importation de celle-ci.

Nous voulons maintenant savoir si la propagation du virus par l'église Shincheonji à joué un rôle principal sur le temps.

On remarque une entrée importante de cas positive sur la période du 18 Février 2020 au 15 Avril 2020. Les valeurs d'entrée ont une fluctuation de 18 à 83 personne positif par jour (avec une moyenne de 39 positive par jour sur 57 jours).

En reprenant les données du *Graphe 2* et celle du *Tableau 1*, la région qui recense le moins de cas positif de la pandémie est Jeju-do. Le nombre de cas positive est évaluée à 14.

province	Sum(confirmed)
Busan	139
Chungcheongbuk-do	38
Chungcheongnam-do	137
Daegu	6650
Daejeon	34
Gangwon-do	51
Gwangju	30
Gyeonggi-do	564
Gyeongsangbuk-do	1314
Gyeongsangnam-do	115
Incheon	79
Jeju-do	14
Jeollabuk-do	15
Jeollanam-do	16
Sejong	46
Seoul	646
Ulsan	42

Tableau 1: Nombre de cas confirmé par province

Intéressons-nous sur l'évolution de la pandémie avant que le nombre de cas augmente de manière exponentiel suite au cluster liée à l'Eglise de Shincheonji.

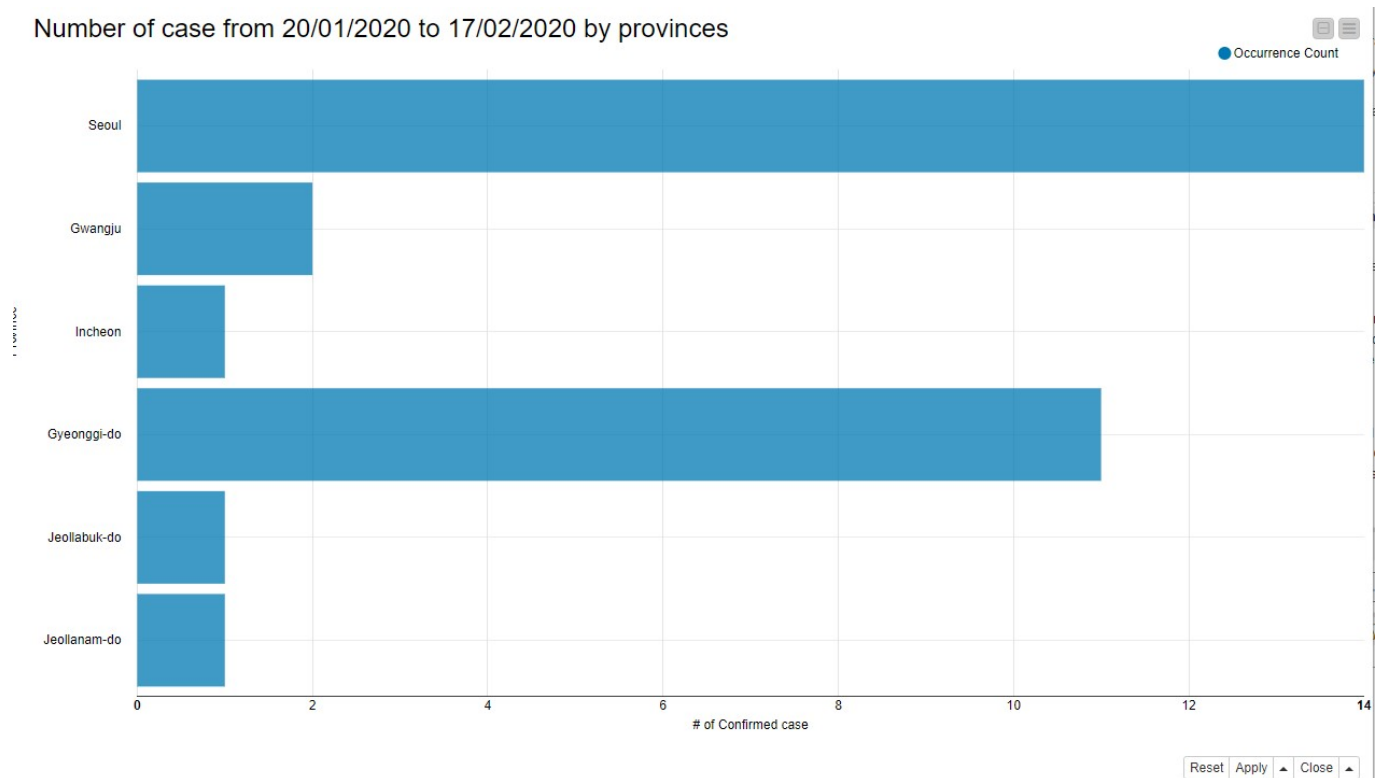
On rappelle les deux dates clés suivantes :

- La date du premier cas recensé sur le territoire coréen est le 20 Janvier 2020
- La date du premier cas recensé suite au cluster de l'Eglise de Shincheonji est le 18 Fevrier 2020.

L'analyse suivante se portera donc sur l'évolution entre 20 Janvier 2020 et le 18 Fevrier 2020.

Durant cette période, il y a 39 cas confirmés liée à la pandémie dont 5 personnes directement liée au cluster du Shincheonji (cf *Tableau 2 en Annexe*). La date du 18 Fevrier sera donc exclu, et l'analyse se portera donc sur la période du 20 Janvier 2020 au 17 Février 2020.

Comme avec l'analyse sur les données générales brute en début de ce rapport, nous devons connaitre quel est la province la plus touchée durant cette période.



Graphe 3: Number of case from 20/01/2020 to 17/02/2020 by provinces

Nous remarquons que 2 province se distingue des autres province à savoir :

- Seoul avec 14 cas confirmées
- Gyeonggi-do avec 11 cas confirmées.

Nous remarquons clairement que la province de Daegu ne figure pas sur le *Graphe 3* puisque son nombre de cas est nulle.

- Comment se fait-il que la province de Seoul soit à 646 patients confirmées étant donnée qu'elle fait partie du cluster sur la période 20/01/2020 - 17/02/2020 ?

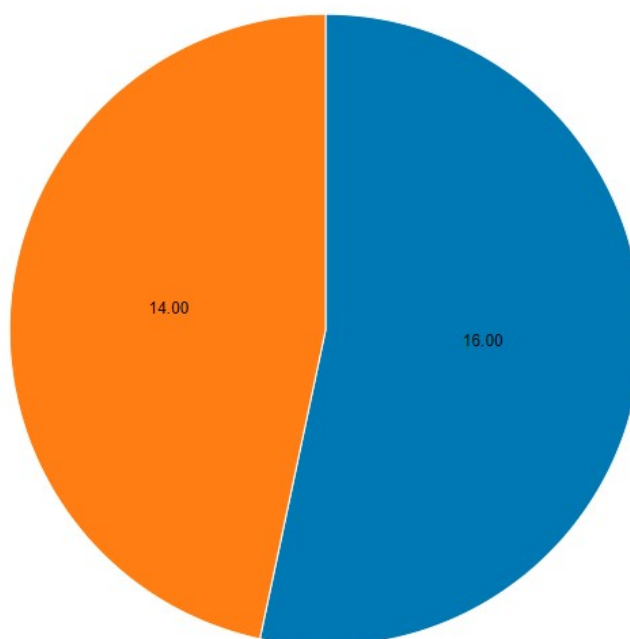
Tout d'abord, il faut étudier le motif de la contamination de cette période pour pouvoir comprendre la provenance de la pandémie avec tout les province d'une part, Seoul d'autre part.

Infection case repartition between 20/01/2020 - 17/02/2020

All provinces



● contact with patient ● overseas inflow

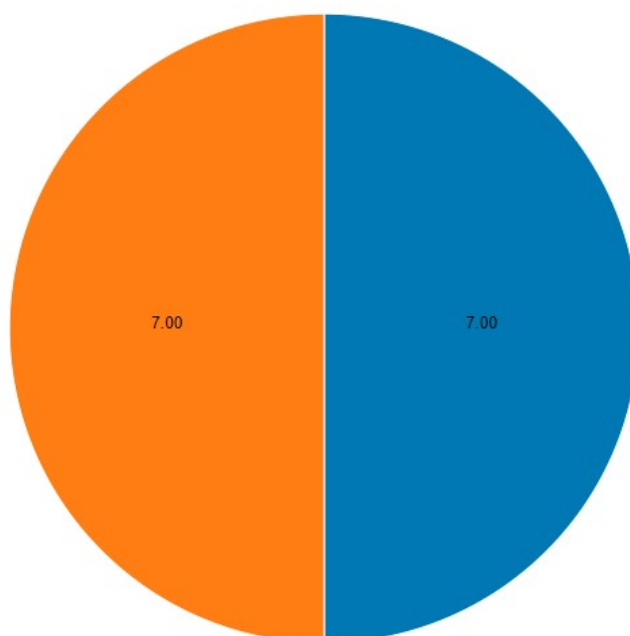


Infection case repartition between 20/01/2020 - 17/02/2020

Seoul



● contact with patient ● overseas inflow



Nous remarquons que seule 2 motifs de contamination sont mentionnés :

- Entrée sur le territoire depuis l'exterieur.
- Contact par patient.

On s'interesse à la propagation du virus au sein de la provence de Seoul. Durant cette période du 20/01/2020 au 18/02/2020, la provence de Seoul comptabilise 14 cas positifs au virus. Elle sont réparties sous formes de

deux motifs que l'on a énoncé ci-dessus (à savoir, Flux étrangers et Contact Patient), avec 7 cas respectivement dans ces deux motifs (*Graph 5*).

En analysant les données données par le *Tableau 2*, nous remarquons que les autorités ont codifié les identifiants des patients selon la provenance de la province. Ici nous nous intéressons à celui de Seoul qui est codifié par 1 000 xxx xxx.

Nous cherchons à trouver une chaîne de contamination la plus importante au sein de cette province. Nous remarquons donc en triant par ordre croissant du *tableau 2* en omettant les patients issus de l'Eglise Shincheonji, sur les valeurs de *infected_by*, le Patient 1 000 000 003 est mentionné 3 fois.

En mettant en lien les sources de contamination au niveau +1 à Seoul. 3 patients ont été contaminés par l'identifiant 1 000 000 003 issu du contact avec le patient lui-même identifié par 2 002 000 001 (*Annexe du tableau 3*).

L'identifiant 2 002 000 001 est en fait une erreur d'entrée qui représente finalement l'identifiant 2 000 000 001. Ce patient fait partie du groupe de contamination de la province du Gyeonggi-do issu d'entrée sur le territoire.

Nous disposons des données d'historique de lieux visités. Ce patient a visité l'Hôpital de Seoul, situé à Gangnam-gu, un magasin dans la même ville et a ensuite fait un déplacement vers la ville de Goyang-si, en Gyeonggi-do dans un restaurant.

Nous pensons que ce patient a contaminé le patient 1 000 000 003 par l'intermédiaire des déplacements en transport au commun de type métro (Ligne 3) au sein de la ville de Seoul durant le déplacement de l'hôpital de Gangnam vers la ville de Goyang-si.

Pour arriver à ce résultat, nous avons dû placer ces points sur une carte, et les superposer sur une carte de métro. Et nous remarquons qu'en effet, la ligne 3 passe près de deux points de contact, l'hôpital de Gangnam et le restaurant situé à Jongno-ju. Les points qui sont donnés dans la base de données sont des coordonnées GPS latitude/longitude.

Nous pouvons dorénavant affirmer que la transmission s'est faite en déplacement par utilisation des transports en communs.

Cependant nous n'avons pas d'information si le patient 2 000 000 001 est passé par l'Aéroport ou bien par un port naval. En effet, le motif de ce patient a été déclaré en tant que flux étrangère (Entrée sur le territoire). On peut seulement supposer que c'est le cas.

En reprenant les données de la répartition des cas d'infection du *Graph 4* et en les mettant en lien avec la trace des lieux visités (*Tableau 4 en annexe*). Les 16 patients ont été contaminés par l'intermédiaire des services hospitaliers qui ont accueilli les patients du flux étrangers.

- Quel est l'évolution de la pandémie pour la province de Gyeonggi-do durant cette période ?

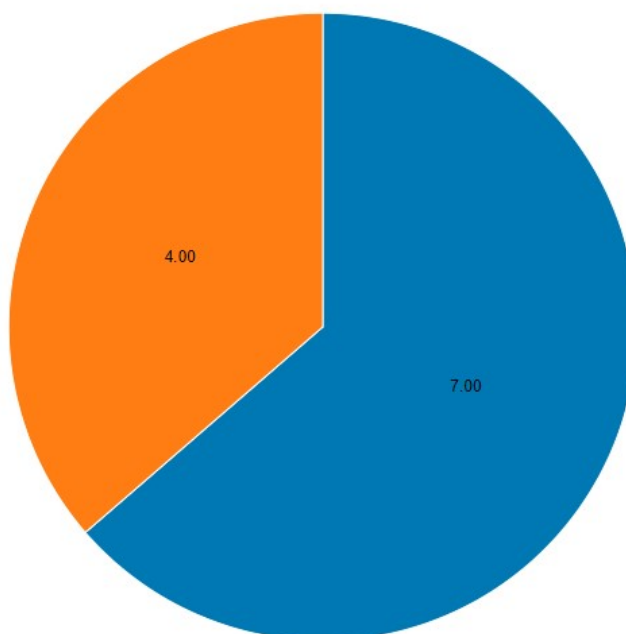
La province de Gyeonggi compte 11 patients positifs au virus pour la période du 20/01/2020 au 17/02/2020. Elle se répartit sur deux motifs de contamination qui sont les mêmes que ceux cités pour Séoul (à savoir Flux étrangers, Contact avec Patient).

Repartition motif de contamination

Gyeonggi-do



● contact with patient ● overseas inflow



On remarque qu'il y a plus de patient contaminé par contact patient que par les flux étrangers.

En appliquant le même procédé de l'analyse effectuée sur Seoul, nous remarquons qu'il y a une chaîne de contamination issue directement du flux étranger chinois.

Il s'agit du patient 2 000 000 010 en provenance de la Chine qui a contaminé deux autres patients par contact patient. (Tableau 5).

2000000008	25	2/8/2020	Gyeonggi-do	Siheung-si	hospital	37.44455	126.7895
2000000008	25	2/9/2020	Gyeonggi-do	Seongnam-si	hospital	37.35207	127.1232
2000000009	26	#####	Incheon	Namdong-gu	airport	37.45626	126.7052
2000000009	26	2/7/2020	Gyeonggi-do	Siheung-si	hospital	37.44455	126.7895
2000000009	26	2/7/2020	Gyeonggi-do	Siheung-si	store	37.43814	126.7964
2000000009	26	2/8/2020	Gyeonggi-do	Siheung-si	hospital	37.44455	126.7895
2000000009	26	2/9/2020	Gyeonggi-do	Anseong-si	hospital	37.01722	127.2604
2000000010	27	#####	Incheon	Namdong-gu	airport	37.45626	126.7052
2000000010	27	#####	Incheon	Namdong-gu	airport	37.45626	126.7052
2000000010	27	2/3/2020	Gyeonggi-do	Siheung-si	restaurant	37.43234	126.8099
2000000010	27	2/3/2020	Gyeonggi-do	Siheung-si	etc	37.43234	126.8099
2000000010	27	2/5/2020	Gyeonggi-do	Siheung-si	hospital	37.44455	126.7895

Tableau 6 : Mise en évidence du lieu de contamination formée par la chaîne 20-8 20-9 et 20-10

Le lieu de la contamination est encore une fois le service hospitalier en faisant la trace des lieux visités.

Ces analyses permettent de déduire que la cause de la formation d'une chaîne de contamination est du au patient qui se retrouve au service hospitalier au contact direct avec d'autre patient. On peut supposer que le service hospitalier est soit de nature d'une consultation bénin, soit sur présentation au service des Urgences.

- Comment l'Eglise Shicheonji est devenu en une journée, un cluster ayant entraîné comme conséquence la propagation rapide au sein du pays ?

Nous savons que le 18/02/2020, 5 patients issue de l'Eglise Shincheonji sont positives au virus. L'analyse se portera donc sur ces 5 personnes particulièrement.

Nous rappelons les données qui sont à disposition pour l'analyse concernant ces patient dans le Tableau 7 suivant :

Row ID	I global_...	S sex	I birth_y...	S age	S country	S ▲ province	S city	S disease	S ▲ infection_case
1200000038	38	female	1963	50s	Korea	Daegu	Nam-gu	TRUE	?
1200000031	31	female	1959	60s	Korea	Daegu	?	?	Shincheonji Church
1200000034	34	male	1996	20s	Korea	Daegu	Jung-gu	?	Shincheonji Church
1200000035	35	female	1994	20s	Korea	Daegu	Nam-gu	?	Shincheonji Church
1200000036	36	female	1972	40s	Korea	Daegu	Nam-gu	?	Shincheonji Church
6015000002	39	female	?	60s	Korea	Gyeongsangbuk-do	Yeongcheon-si	?	Shincheonji Church

Tableau 7 : Liste des patient qui sont positive à la date de 18/02/2020 issue de l'Eglise Shincheonji.

Nous remarquons que nous avons pas d'information si ces 5 patient ont été contaminés par des patients déjà connu antérieur à cette date. En revanche, le patient 6 015 000 002 attire notre attention car elle n'est pas en provenance de la provence de Daegu mais de Gyeongsangbuk-do.

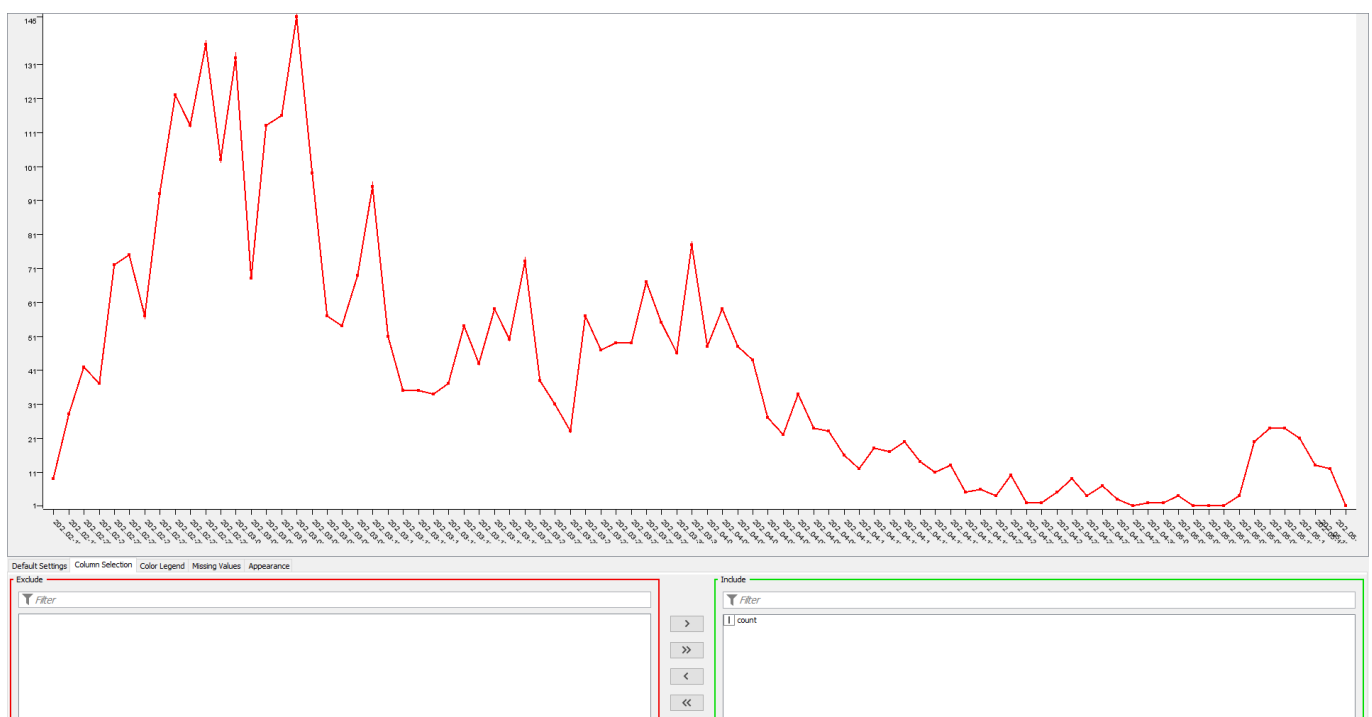
En voulant retracer l'historique des lieux visitée, nous possedons aucune information concernant celle-ci.

En regardant les numéro des patient depuis le début de l'épidémie, la patient 31 (12 000 000 031) est la première à être positive test faisant partie à l'Eglise.

Son historique des lieux visité (Tableau 8) ne correspond à aucune information qui peut être utilisée pour remonter son lieu de contamination.

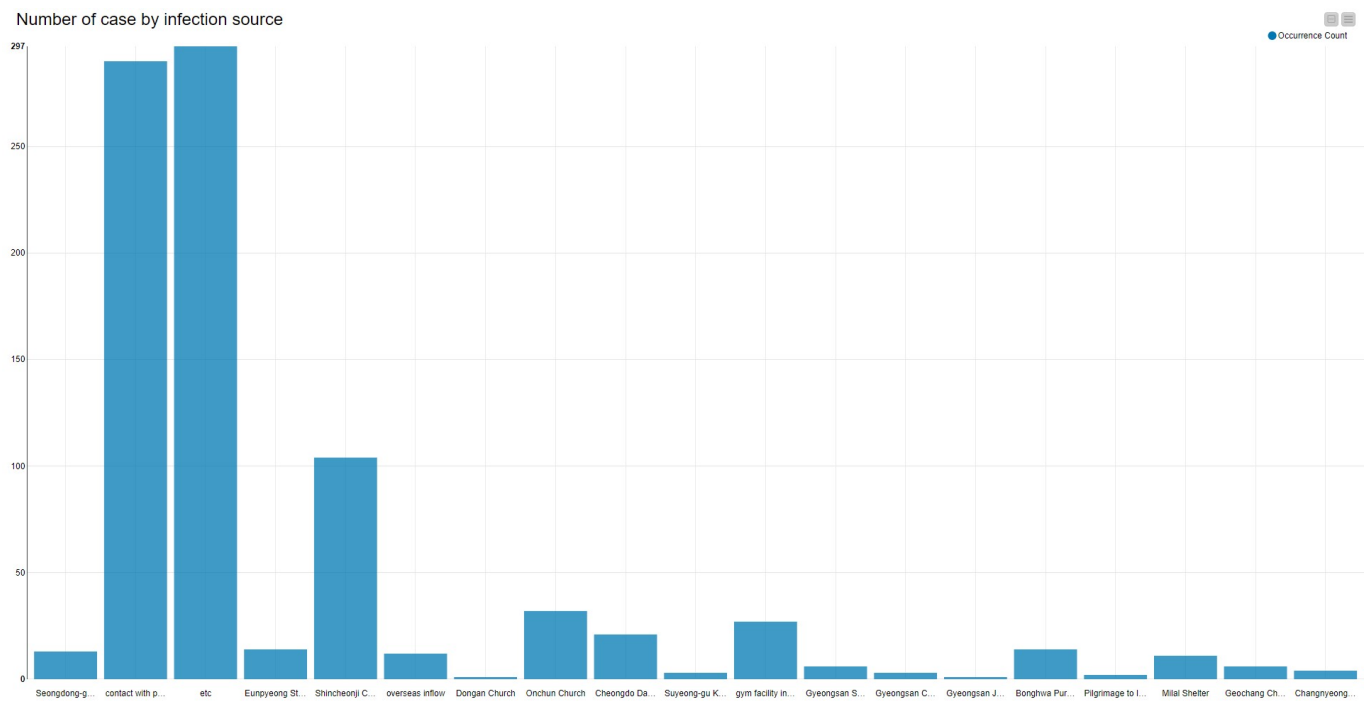
Nous somme donc toujours sur le flou sur l'origine de la contamination massive a partir de l'église. Nous savons juste que c'est à partir de la patiente 31 que la propagation s'est accelérée. Ceci c'est une situation très similaire dans le cas de la France impliquant l'Eglise de la Porte Ouverte église dans le cadre d'un rassemblement religieux.

Reprenons l'analyse de la propagation à partir du moment où l'Eglise est devenu le cluster principale le 18/02/2020.



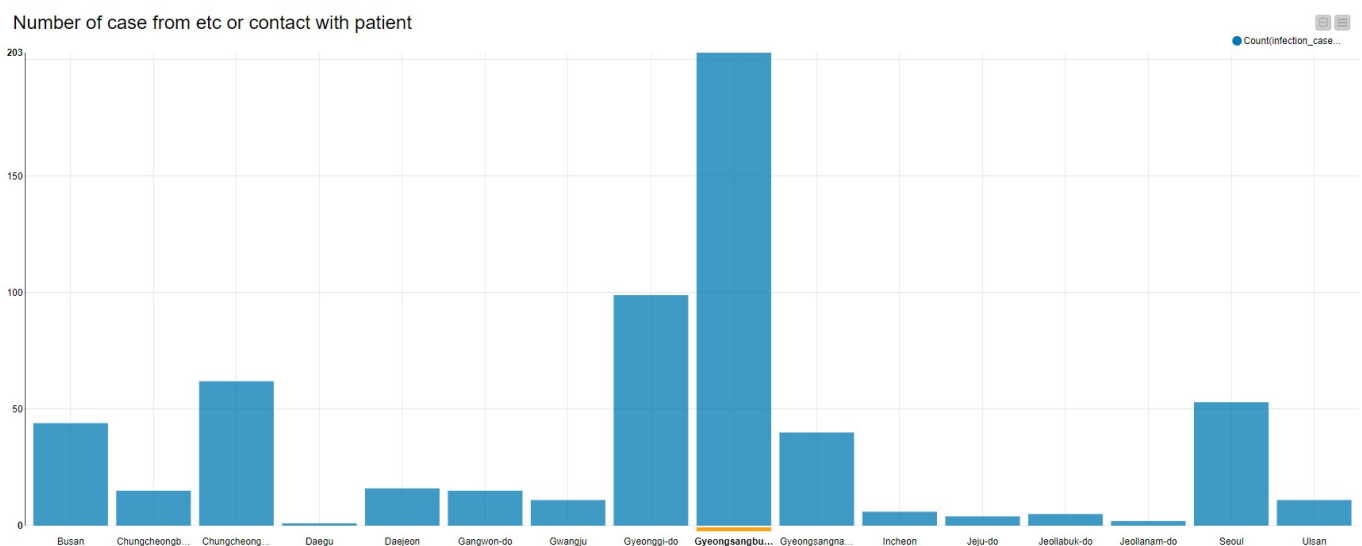
Le pic des patients positifs à été atteinte le 5 mars 2020, étudions l'évolution rapide entre la période du 18/02/2020 au 5/03/2020.

Nous remarquons qu'il y a une augmentation de 1463 de cas positives sur cette période, ce qui porte au nombre totale de 1493 de cas positif depuis le premier cas recensé.



L'augmentation considérable du nombre de cas positives est du à ces deux raisons suivantes :

- Contact avec un patient
- etc.



Les provinces qui ont contribué fortement à ces résultats sont :

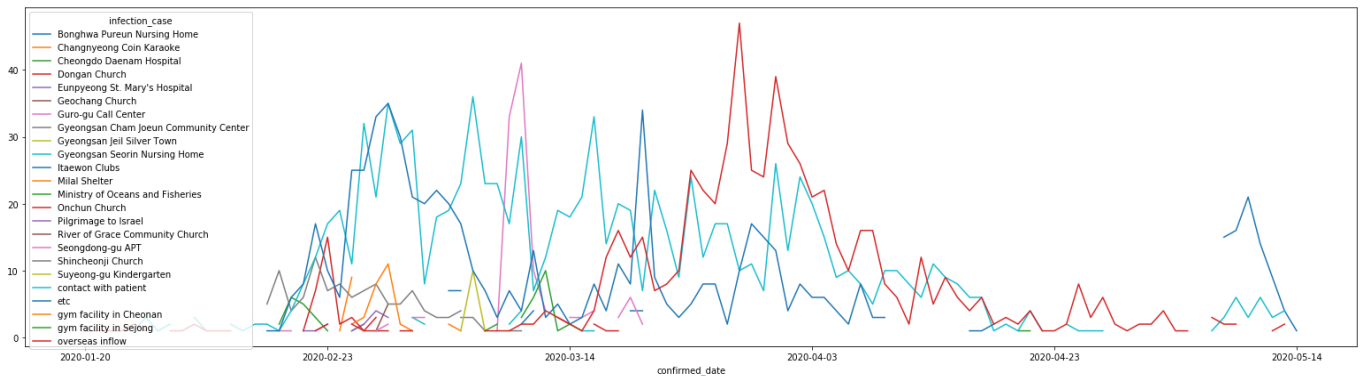
- Gyeongsangbuk-do (203)
- Gyeonggi-do (99)

Row ID	I global...	S sex	I birth_y...	S age	S country	S province	S city	S disease	S infection_case	I infectio...	D infecte...	I contact...	S sympto...	S ▲ con...
2000000012	32	female	2009	10s	Korea	Gyeonggi-do	Suwon-si	?	contact with patient	?	2,000,000,007	?	?	2020-02-18
2000000013	140	female	1988	30s	Korea	Gyeonggi-do	Gimpo-si	?	contact with patient	?	1,200,000,031	?	2020-02-20	2020-02-21
2000000014	162	male	1987	30s	Korea	Gyeonggi-do	Gimpo-si	?	contact with patient	?	2,000,000,013	?	?	2020-02-21
2000000015	246	female	1982	30s	Korea	Gyeonggi-do	Bucheon-si	?	contact with patient	?	1,200,000,031	?	?	2020-02-22
2000000016	347	male	1996	20s	Korea	Gyeonggi-do	Bucheon-si	?	overseas inflow	?	?	?	?	2020-02-22
2000000017	360	male	1959	60s	Korea	Gyeonggi-do	Icheon-si	?	contact with patient	?	?	?	?	2020-02-22
2000000018	361	male	1999	20s	Korea	Gyeonggi-do	Pocheon-si	?	etc	?	?	?	?	2020-02-22
2000000019	362	male	1956	60s	Korea	Gyeonggi-do	Anyang-si	?	contact with patient	6	1,000,000,014	32	?	2020-02-22
2000000025	392	male	1966	50s	Korea	Gyeonggi-do	Gimpo-si	?	Shincheonji Church	?	?	15	?	2020-02-22
2000000029	433	female	1954	60s	Korea	Gyeonggi-do	Suwon-si	?	etc	?	?	?	?	2020-02-22

Row ID	I global...	S sex	I birth_y...	S age	S country	S province	S city	S disease	S infection_case	I infectio...	D infecte...	I contact...	S sympto...	S ▲ con...
6015000001	37	male	?	40s	Korea	Gyeongsangbuk-do	Yeongcheon-si	?	etc	?	?	?	?	2020-02-18
6015000002	39	female	?	60s	Korea	Gyeongsangbuk-do	Yeongcheon-si	?	Shincheonji Church	?	?	?	?	2020-02-18
6001000001	67	female	1991	20s	Korea	Gyeongsangbuk-do	Gyeongsan-si	?	Shincheonji Church	?	1,200,000,031	7	?	2020-02-19
6001000002	73	female	2001	10s	Korea	Gyeongsangbuk-do	Gyeongsan-si	?	?	?	?	2	?	2020-02-19
6015000003	41	female	?	70s	Korea	Gyeongsangbuk-do	Yeongcheon-si	?	etc	?	?	?	?	2020-02-19
6015000004	66	male	?	20s	Korea	Gyeongsangbuk-do	Yeongcheon-si	?	Shincheonji Church	?	?	?	?	2020-02-19
6020100054	54	male	1963	50s	Korea	Gyeongsangbuk-do	Cheongdo-gun TRUE	Cheongdo Daenam Hospital	?	?	?	?	?	2020-02-19
6020100055	55	male	1961	50s	Korea	Gyeongsangbuk-do	Cheongdo-gun TRUE	Cheongdo Daenam Hospital	?	?	?	?	?	2020-02-19

Tableau 9 et 10: motifs de contamination en début de propagation

En étudiant les données de Gyeongsangbuk-do avec tout les motifs recensés, nous remarquons que l'origine de la propagation du virus est issue de l'Eglise Shincheonji Church. Il en est de même pour Gyeonggi-do. Les contacts patient ont repris le relais pour accélérer la contamination locale. Ce qui laisse supposer que l'Eglise à du organisé un événement dans le but de rassembler les personnes venant d'autres villes. La contamination s'est donc développée rapidement sur les autres provinces.



Graph des motifs de contamination au fil du temps

Analysons les motifs de contamination au jusqu'au 15/05/2020. Nous partons du principe qu'à partir de 20 personnes contaminé pour le même motif et le même est considéré comme foyer de contamination.

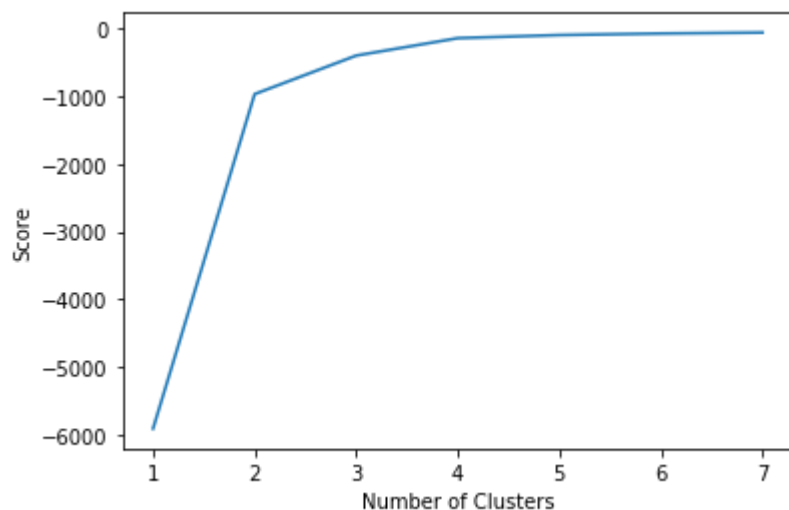
Nous pouvons donc recenser les foyers de contamination suivantes :

- Bonghwa Pureun Nursing Home
- contact with patient
- Guro-gu Call Center
- Dogan Church
- laewon Clubs

A partir de ce graphe, nous pouvons affirmer que la propagation du virus se fait principalement au sein de l'Eglise, les transports en commun (d'où la propagation vers les autres villes) et au sein des services hospitaliers.

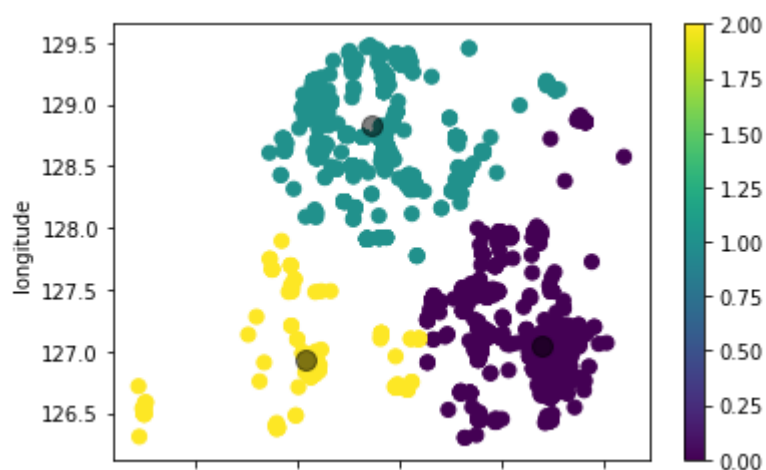
Affinons l'analyse avec la manipulation des données par python.

Nous souhaitons connaître le nombre de cluster analytique des données.



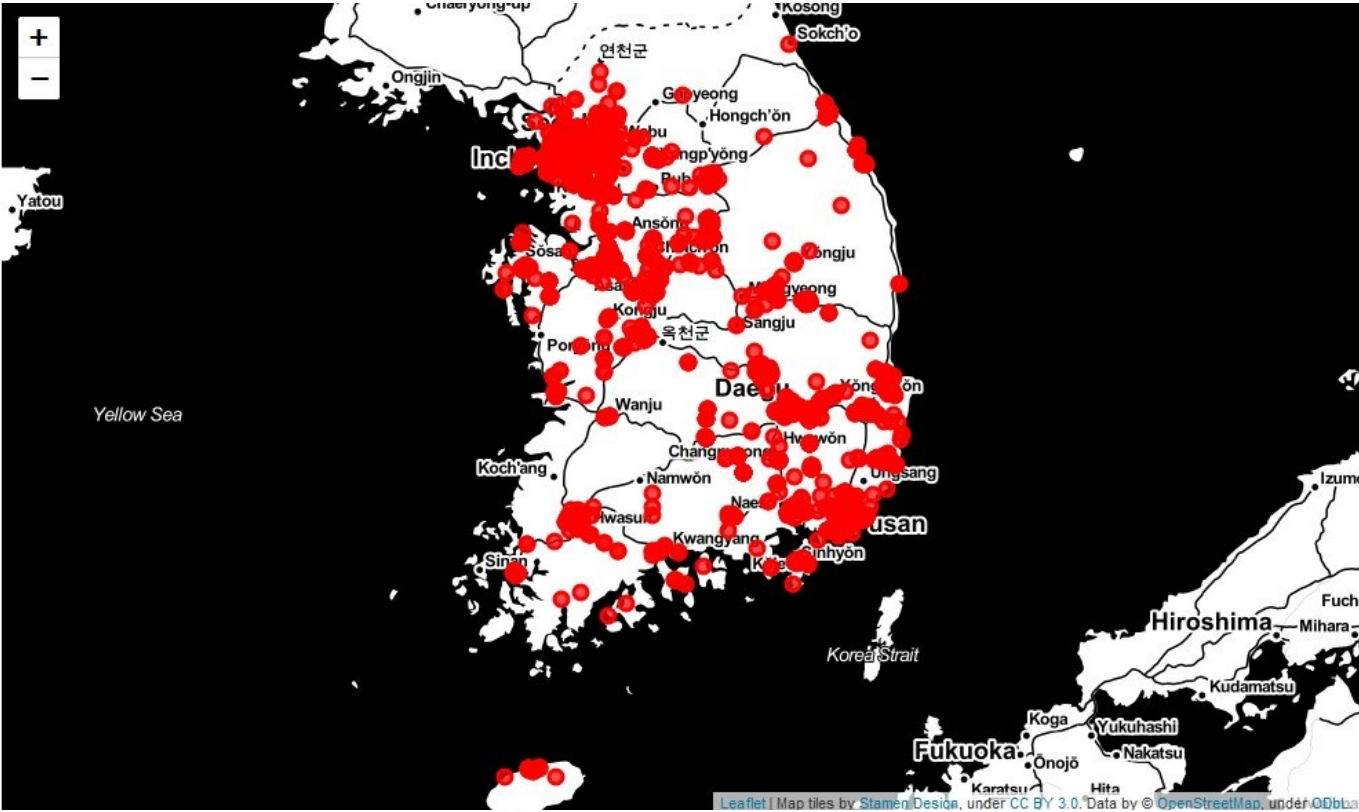
Nous remarquons qu'après le score de 3 cluster, la courbe va tendre vers 0.

Nous transposons les données de clusters sous forme d'un graphe suivant :



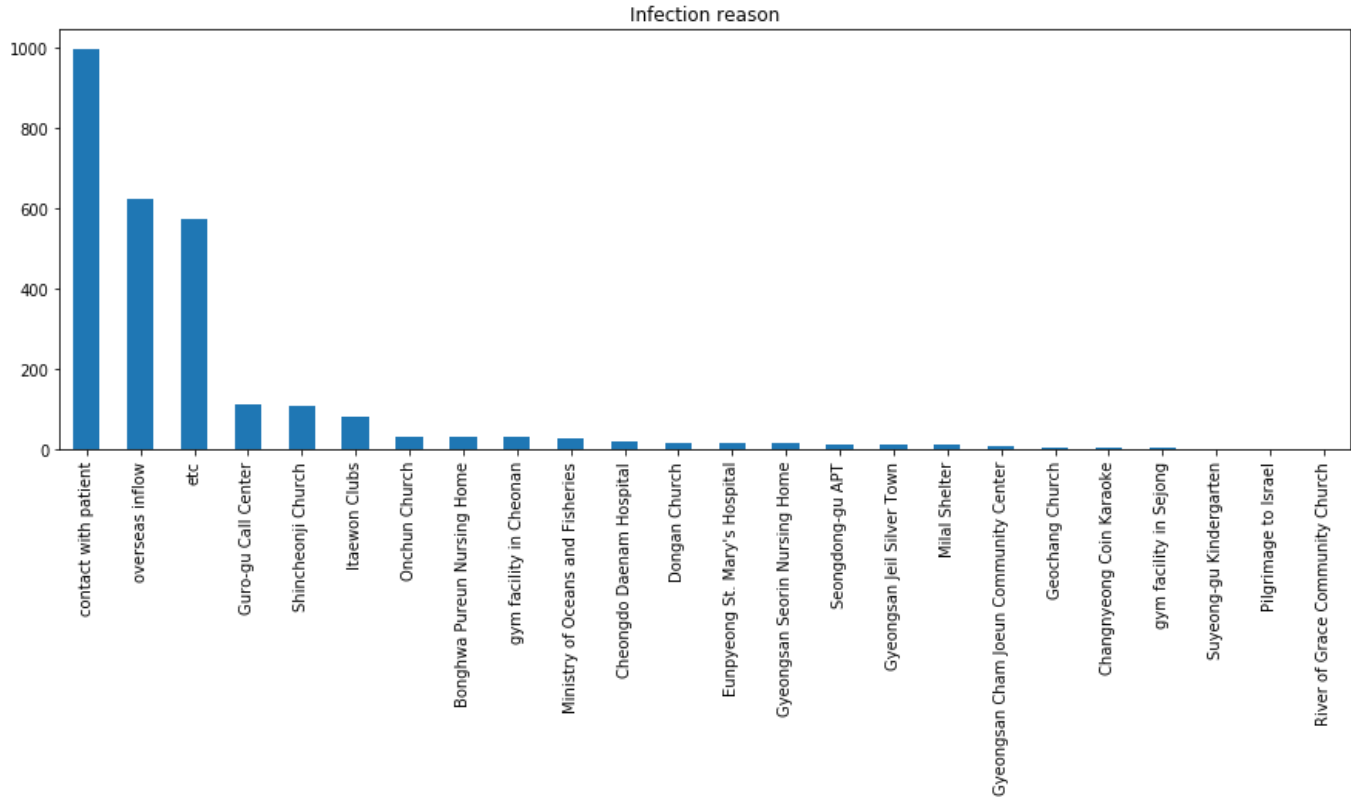
Nous remarquons bien qu'il y a 3 cluster qui sont distinguées par les 3 couleurs.

En les transposant une nouvelle fois ces données vers une carte, nous avons donc une représentation visuelle de la localisation des cas confirmés.



Carte 1 : Localisation des cas confirmée

Nous allons afficher le nombre totale des motifs de contamination afin de confirmer les 3 clusters principaux qu'ont été cité (Guro-gu Call Center, Shincheonji Church, Itaewon Clubs).

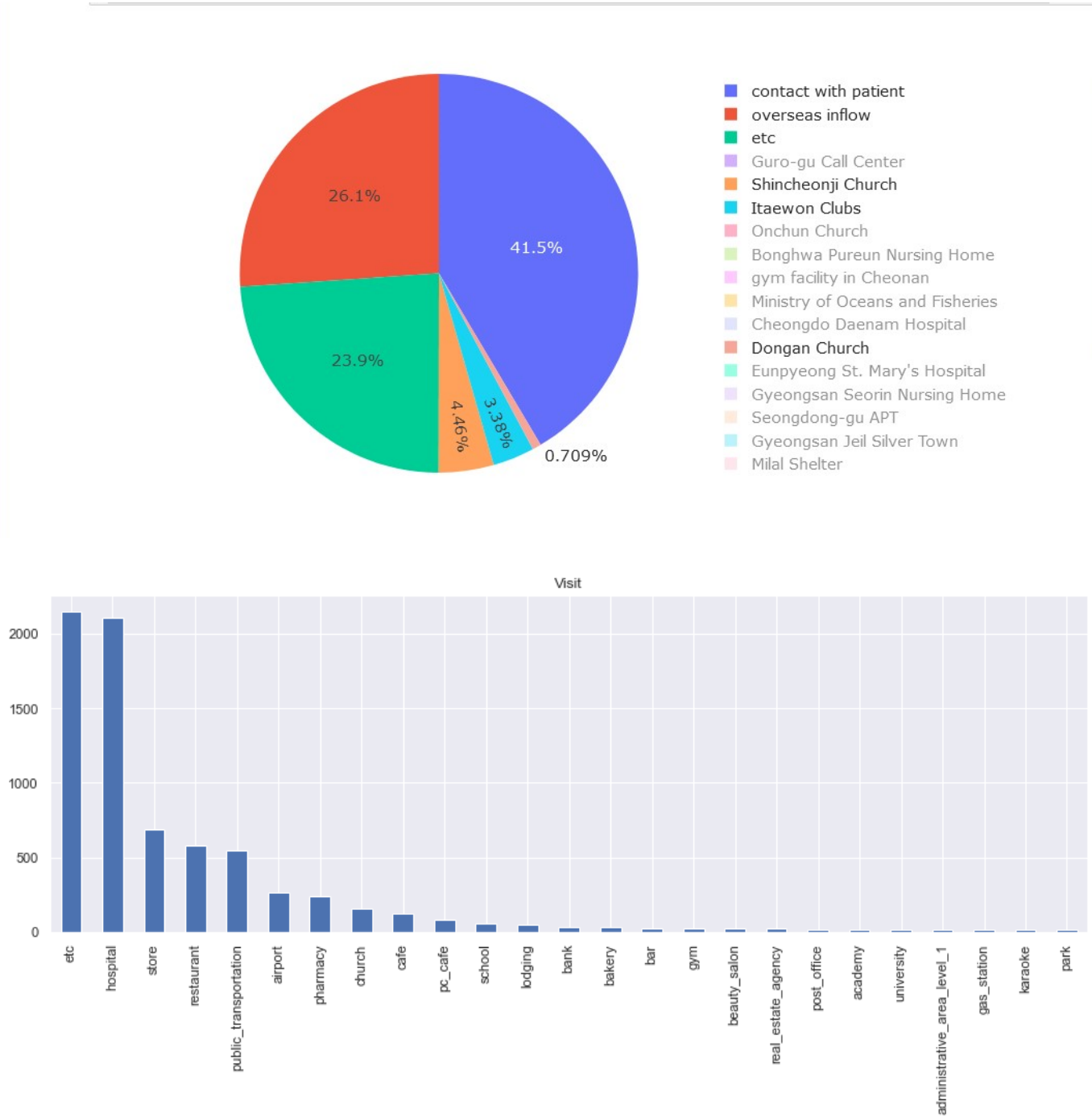


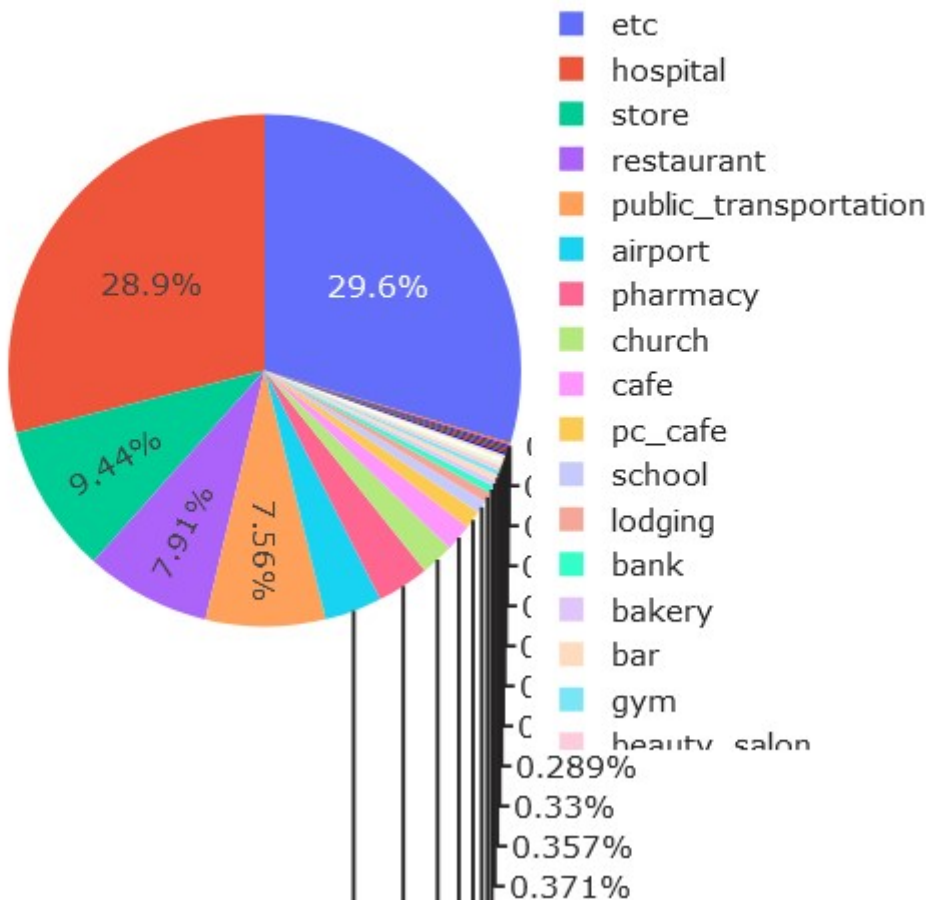
Nous remarquons que contact with patient occupe la première place, ce qui est totalement logique car la pandémie progresse par contact patient du service hospitalier.

De plus, 600 personnes de flux étrangers ont été recensées ce qui est également logique avec le retour massive depuis la ville de Wuhan avant son lockdown (et aussi venant d'autres pays).

Le motif etc regroupe tout les cas en cours d'investigation par les autorités sanitaires, celle ci peuvent correspondre aux transport en commun principalement.

On retrouve finalement nos trois clusters principaux avec ce graphe, ce qui confirme bien qu'il sont des cluster principaux.





Ce graphe

recense les lieu les plus visités avant la déclaration d'un patient confirmée. On remarque que c'est:

- L'Hopital (29.6 %)
- Les magasins (28.9 %)
- Restaurant (9.44%)
- Les tranports en commun (7.56 %)

qui revient très fréquemment.



On a ici une répartition de tout les patient regroupé par la cause de l'infection en fonction de leur groupe d'age. On peut remarquer que les lieux de travail tel que les hopitaux touchent principalement les adultes. Tandis que les églises touche les adultes et les enfants.

Le virus a une létalité de 71 personnes parmi les personnes infectées, ce qui donne un ration de 0.01% de taux de mortalité. Cependant, les chiffres ne sont pas exactes lorsque l'on compare au données gouvernementale. Nous suspectons qu'il y a un problème de mise à jour de leur part, ce qui explique la différence entre les données disponibles et les données communiquées par la voie de la presse. Le chiffre de mortalité attendu est de 262 pour 11037 cas confirmées (0.02%).

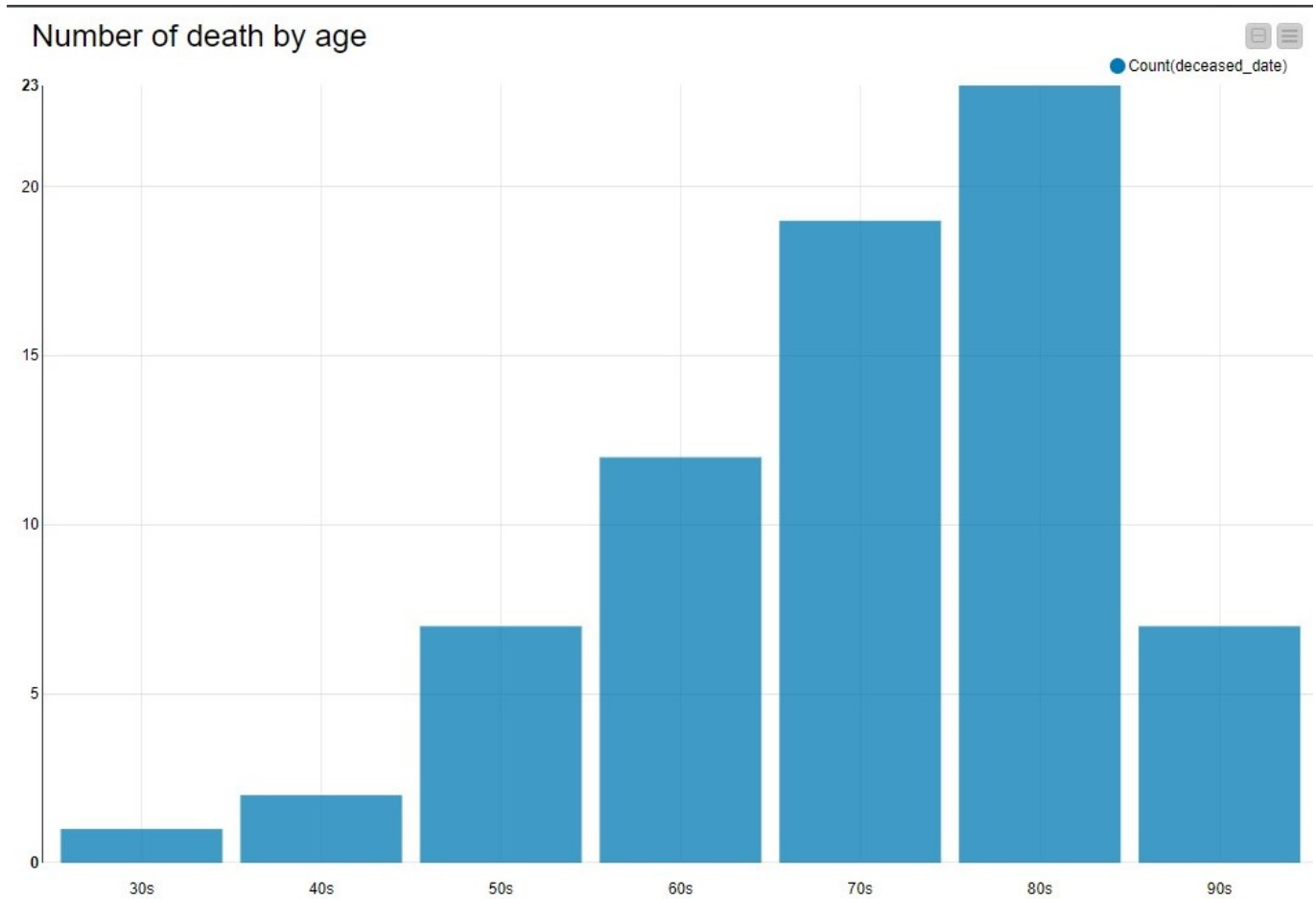


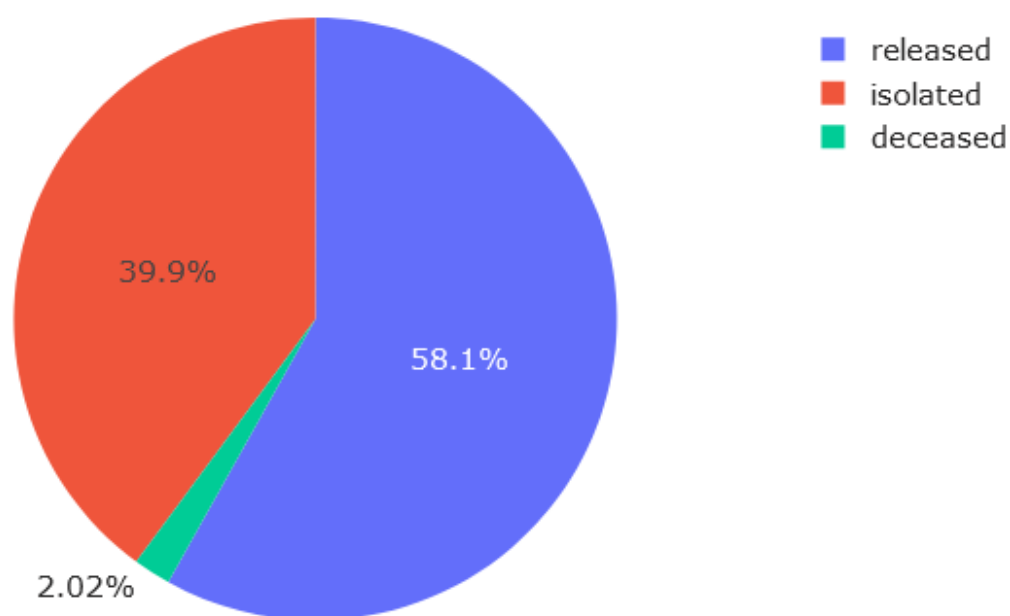
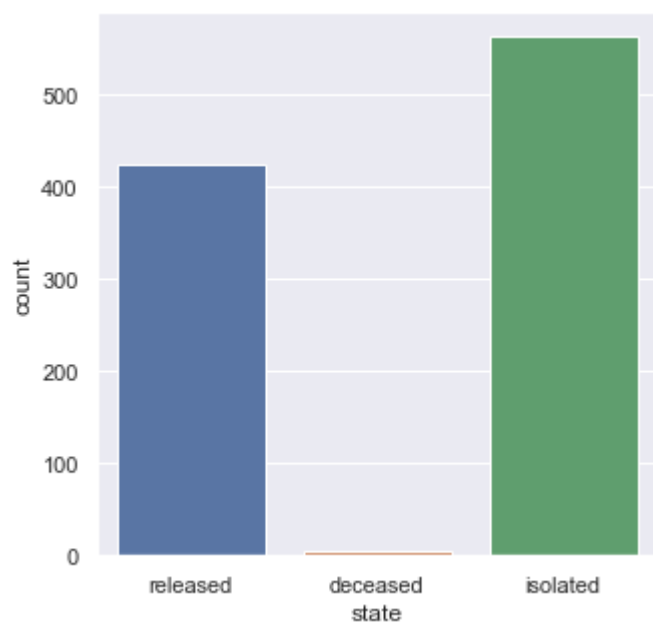
Chart 10: Number of death grouped by age

0	10	20	30	40	50	60	70	80	90	100	Total
0	0	0	0.22	0.43	1.18	3.05	10	17.29	18.42	0	50.59

Tableau 11: Taux de mortalité en fonction de l'age exprimé en pourcentage

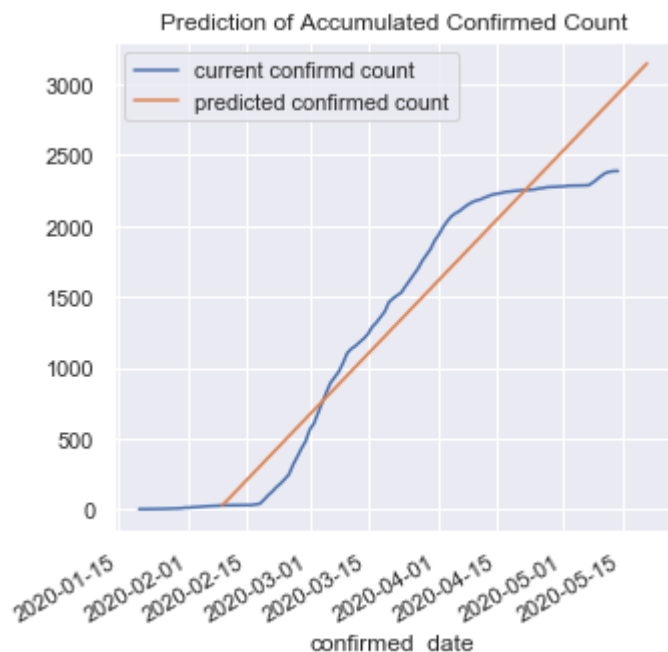
Nous pouvons remarquer qu'au niveau de la mortalité, les personnes âgées de plus de 50 ans sont les plus touchés par cette pandémie avec 23 patients âgés dans la tranche de 80 ans, 19 pour la tranche de 70 ans et 7 pour la tranche de 90. Ce qui représente 17.29 % (10 % et 18.42 % respectivement) de taux de mortalité pour cette âge avec un totale de 50.59 %.

Nous allons démontrer que le taux de mortalité est quand même relativement faible.



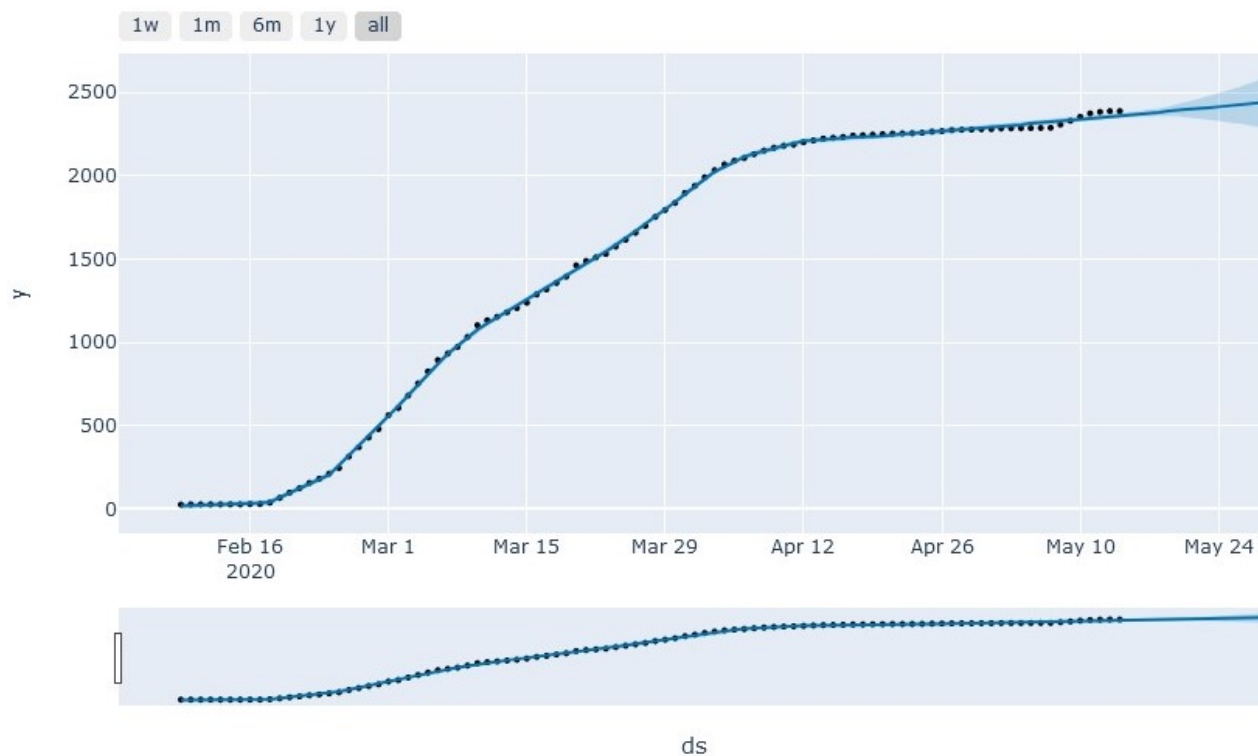
Puisque les données que l'on possède ne sont pas à jour, nous avons 71 morts, ce qui représente un taux de 2.02 %. Notons néanmoins que 58.1 % des patient qui étaient infecté ne le sont plus (ou bien déchargé des services hospitaliers) et que 39.9 % sont en confinement.

- Simulation de l'évolution de la pandémie.

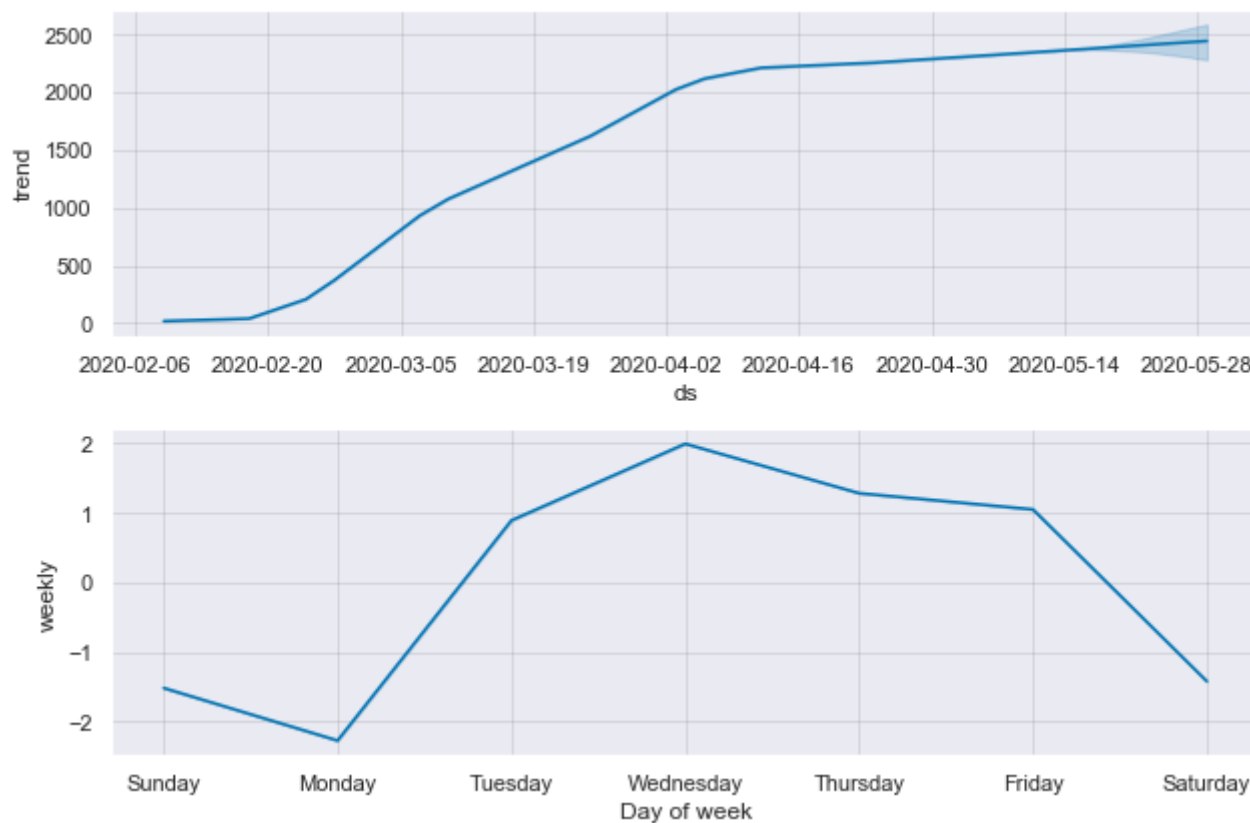


Nous avons utilisé un modèle de régression des données avec un réseau neural implémentaire depuis la librairie MLPRegressor (par le biais Sklearn) Nous remarquons qu'une droite de type $Ax+B$ est tracé, ce qui indique la prédiction de l'évolution de la pandémie. Cependant, La prédiction est faussé à partir du mois d'Avril puisque la courbe actuelle des cas confirmées converge en un point I.

La simulation des données par la bibliothèque prophet est une correction analytics les prédictions effectuée par MLPRegressor ce qui rend réaliste les prévisions à court terme.



La courbe bleu représente les données simulées par prophete et les points noir représente les données actuelle. La zone blue est l'intervalle de confiance probable du nombre de cas confirmée dans le court terme. On remarque que cette intervalle augmente au fil du temps.

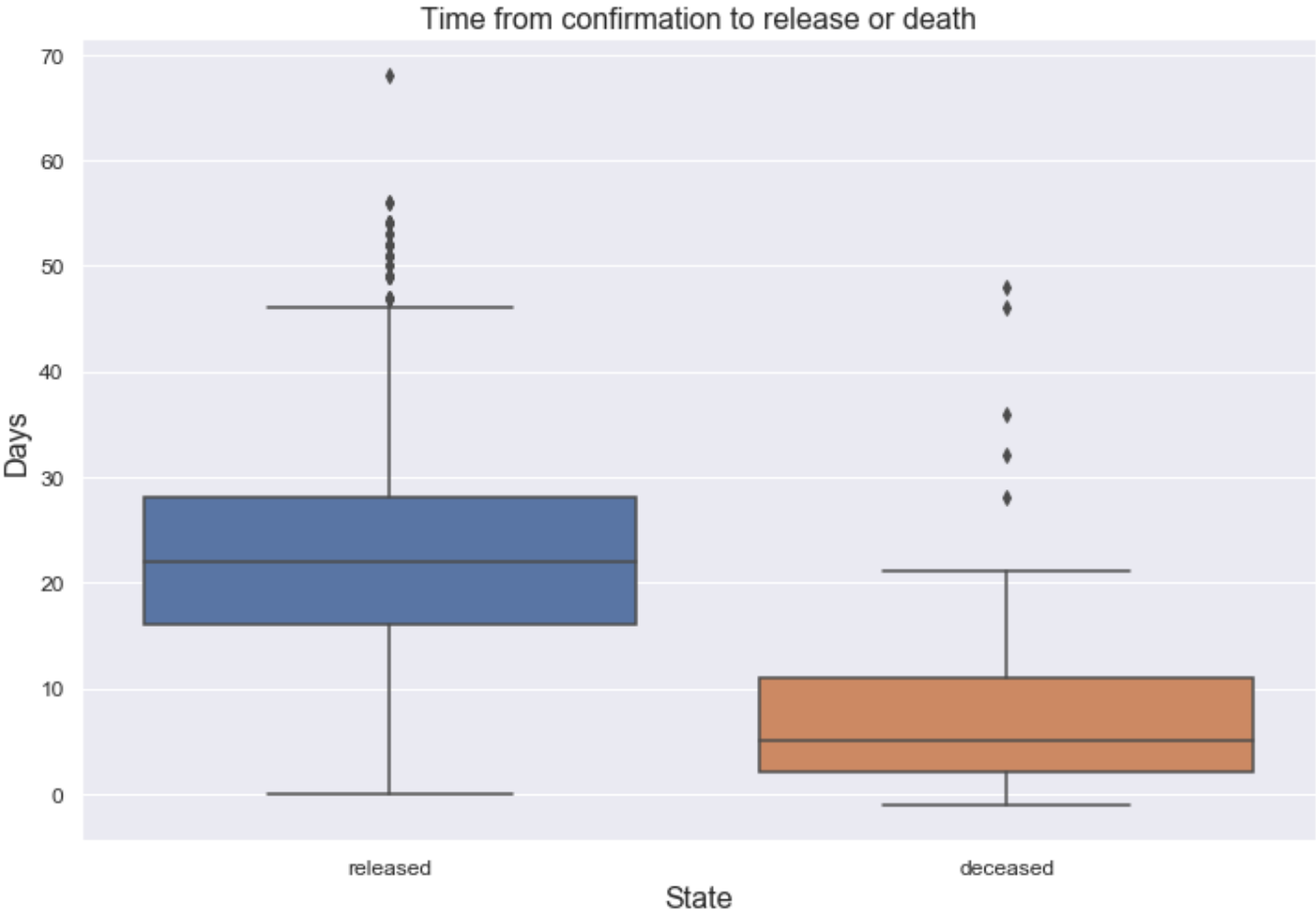


Ce graphe est la prédiction sur 1 semaine journalière de la tendance de l'épidémie à partir du 15 mai 2020. Elle annonce un pic estimé à la journée de Mercredi avec une baisse du nombre de cas.

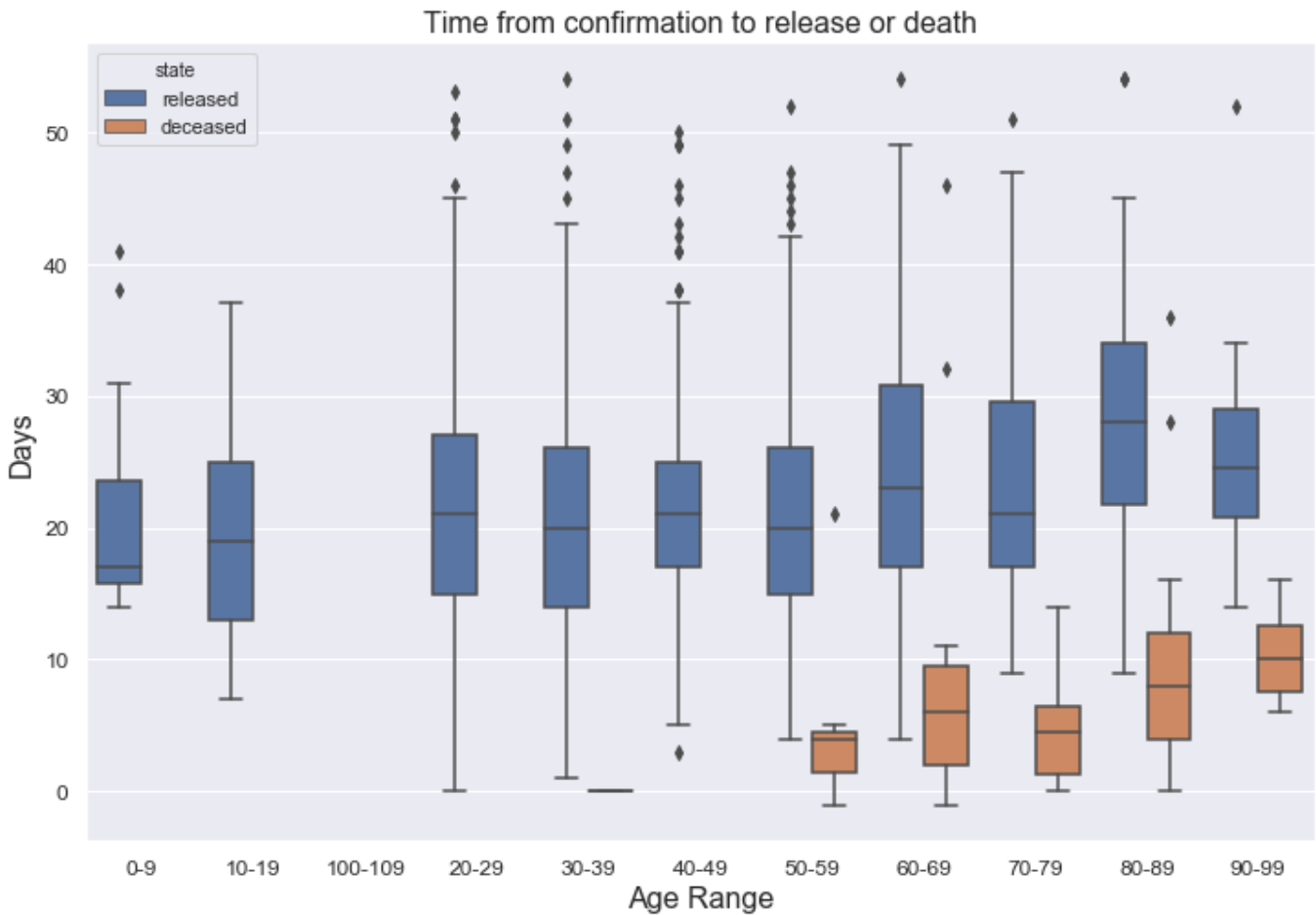
Un patient qui a été confirmé au virus peut, au fil du temps, soit :

- Guérie
- Décédé

Le graphe suivant permet de démontrer le nombre de journée qu'il faut pour qu'un patient rentre dans l'une des deux catégories citées ci-dessus

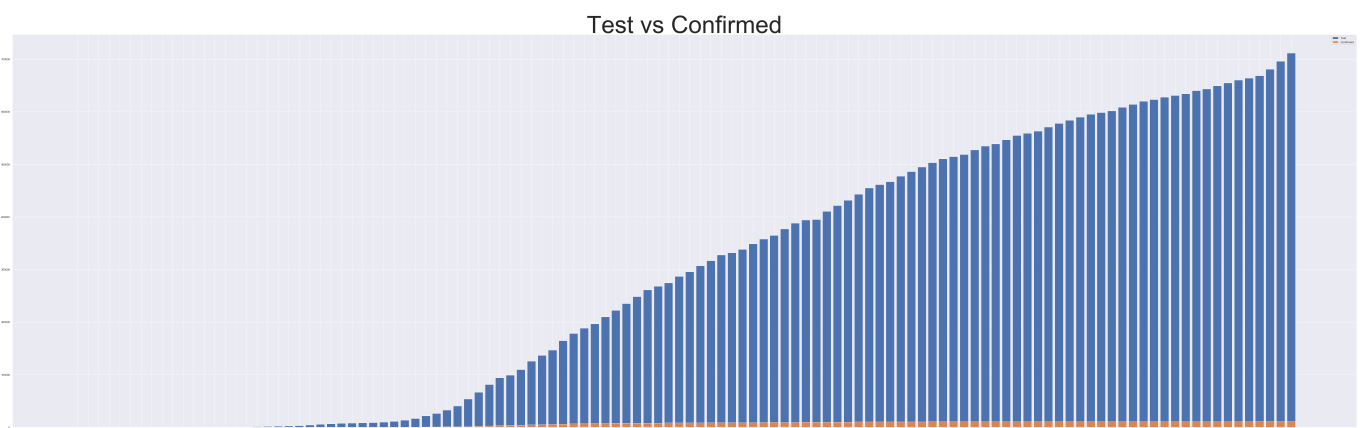


Nous remarquons qu'un patient guérie met plus temps de temps qu'un patient qui va décéder. La différence entre les deux est l'ordre de 2 jours.

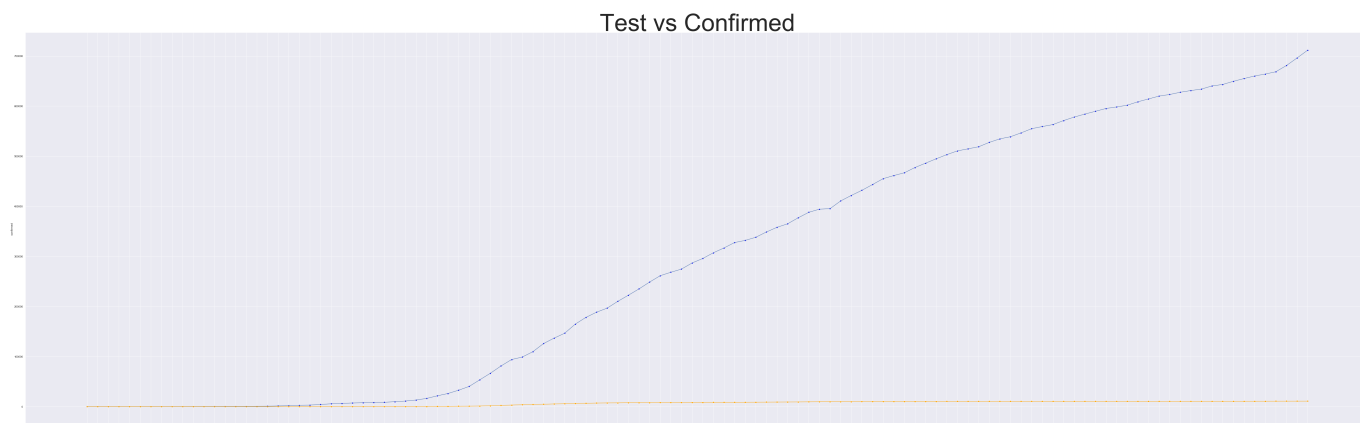


Ce graphe met en évidence sur la mortalité des personnes âgées, en effet, à partir de 50 ans, ces personnes sont plus susceptibles de mourir à cause du virus, ce qui démontre le taux de mortalité parmi les mort à 50.59%.

Le gouvernement a décidé de tester sa population, le graphe suivant permet de mettre en évidence le ratio des tests positifs (en orange) et négatif (en bleu). Le nombre de tests est porté à environ 700 000.



Sous forme de courbe



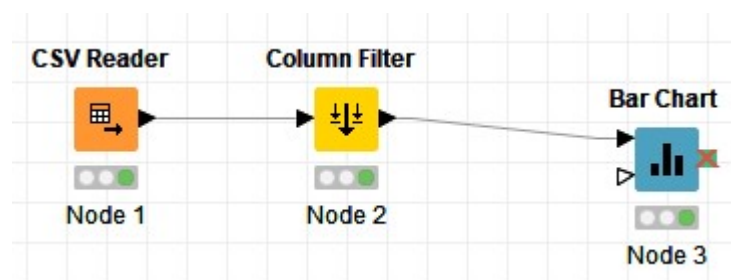
Conclusion de l'Analyse

La Corée du Sud est touché par cette pandémie quelque semaine après la Chine le 20 Janvier 2020 par l'intermediaire l'entrée sur le territoire Coréen en provenance de Chine. La propagation du virus était au ralenti puisque les contamination se fesait au niveau des services hospitaliers. La pandémie s'est accélérée au sein du pays de manière exponentiel suite à l'infection par l'Eglise de Shincheonji. La province de Daegu par la suite est devenu le cluster principal de la pandémie au sein du pays, et le reste toujours aujourd'hui. Cependant, la Corée du Sud ont réussi à contrôler la propagation du virus malgré le récent cluster de laewon Club et Dogan Church.

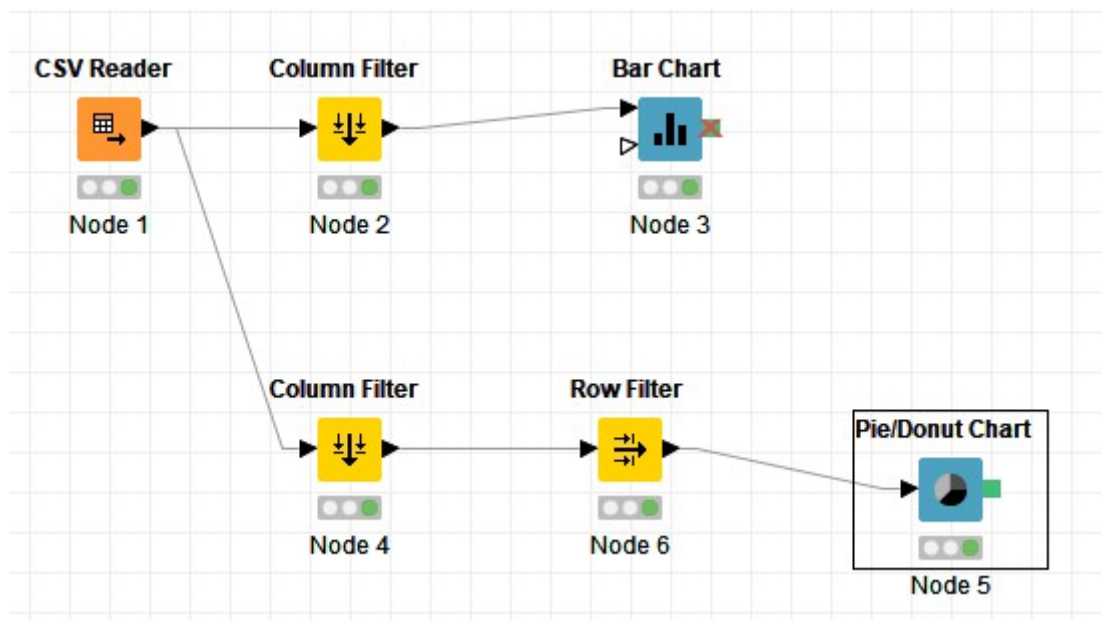
La pandémie touche toutes les tranches d'âges, plus particulièrement la tranche d'âge de 20 ans. Elle est en revanche mortelle à partir de 50 ans, avec une mortalité cumulée de 96 % parmi les mort (50.59 % pour le taux de mortalité cumulée).

La recente évolution de la pandémie laisse penser que la Corée du Sud est au stade de la décrue. En effet, la forte vague de cas positifs s'est déroulée du 18 Février au 15 avril pour une durée totale de 57 jours.

Annexes



Flow Chart du graphe 1



Flow Chart du graphe 2

Table "PatientInfo.csv" - Rows: 39 Spec - Columns: 17 Properties Flow Variables

Row ID	I global_...	S sex	I birth_y...	S age	S country	S province	S city	S disease	S infection...	I infectio...	D infecte...	I contact...	S sympto...	S ▲ con...	S release...	S t
1400000001	1	female	1985	30s	China	Incheon	etc	?	overseas inflow	?	?	?	2020-01-19	2020-01-20	2020-02-06	?
1000000001	2	male	1964	50s	Korea	Seoul	Gangseo-gu	?	overseas inflow	1	?	75	2020-01-22	2020-01-23	2020-02-05	?
2000000001	3	male	1966	50s	Korea	Gyeonggi-do	Goyang-si	?	overseas inflow	1	?	16	?	2020-01-26	2020-02-12	?
2000000002	4	male	1964	50s	Korea	Gyeonggi-do	Pyeongtaek-si	?	overseas inflow	1	?	95	?	2020-01-27	2020-02-09	?
1000000002	5	male	1987	30s	Korea	Seoul	Jongno-gu	?	overseas inflow	1	?	31	?	2020-01-30	2020-02-02	?
1000000003	6	male	1964	50s	Korea	Seoul	Jongno-gu	?	contact with ...	2	2,002,000,001	17	?	2020-01-30	2020-02-19	?
1000000004	7	male	1991	20s	Korea	Seoul	Mapo-gu	?	overseas inflow	1	?	9	2020-01-26	2020-01-30	2020-02-15	?
1000000005	9	female	1992	20s	Korea	Seoul	Seongbuk-gu	?	contact with ...	2	1,000,000,002	2	?	2020-01-31	2020-02-24	?
1000000006	10	female	1966	50s	Korea	Seoul	Jongno-gu	?	contact with ...	3	1,000,000,003	43	?	2020-01-31	2020-02-19	?
1000000007	11	male	1995	20s	Korea	Seoul	Jongno-gu	?	contact with ...	3	1,000,000,003	0	?	2020-01-31	2020-02-10	?
5000000001	8	female	1958	60s	Korea	Jeollabuk-do	Gunsan-si	?	overseas inflow	1	?	113	?	2020-01-31	2020-02-12	?
2000000003	12	male	1971	40s	China	Gyeonggi-do	Bucheon-si	?	contact with ...	2	?	422	?	2020-02-01	2020-02-18	?
1000000008	13	male	1992	20s	Korea	Seoul	etc	?	overseas inflow	1	?	0	?	2020-02-02	2020-02-24	?
2000000004	14	female	1980	40s	China	Gyeonggi-do	Goyang-si	?	contact with ...	3	2,000,000,003	3	?	2020-02-02	2020-02-18	?
2000000005	15	male	1977	40s	Korea	Gyeonggi-do	Suwon-si	?	contact with ...	2	2,000,000,002	15	2020-02-01	2020-02-02	2020-02-24	?
1300000001	16	female	?	40s	Korea	Gwangju	?	?	overseas inflow	?	?	450	2020-01-27	2020-02-03	2020-02-20	?
1000000009	19	male	1983	30s	Korea	Seoul	Songpa-gu	?	overseas inflow	2	?	68	?	2020-02-05	2020-02-21	?
1000000010	21	female	1960	60s	Korea	Seoul	Seongbuk-gu	?	contact with ...	3	1,000,000,003	6	?	2020-02-05	2020-02-29	?
1300000002	18	female	?	20s	Korea	Gwangju	?	?	contact with ...	?	1,300,000,001	8	?	2020-02-05	2020-02-20	?
2000000006	17	male	1982	30s	Korea	Gyeonggi-do	Guri-si	?	overseas inflow	2	?	290	?	2020-02-05	2020-02-12	?
2000000007	20	female	1978	40s	Korea	Gyeonggi-do	Suwon-si	?	contact with ...	3	2,000,000,005	2	2020-02-04	2020-02-05	2020-02-24	?
1000000011	23	female	1962	50s	China	Seoul	Seodaemun-gu	?	overseas inflow	1	?	23	?	2020-02-06	2020-02-29	?
5100000001	22	male	1974	40s	Korea	Jeollanam-do	etc	?	contact with ...	?	1,300,000,001	1	?	2020-02-06	2020-02-15	?
1000000012	24	male	1992	20s	Korea	Seoul	etc	?	overseas inflow	1	?	0	?	2020-02-07	2020-02-27	?
2000000008	25	female	1946	70s	Korea	Gyeonggi-do	Siheung-si	?	contact with ...	2	2,000,000,010	12	?	2020-02-09	2020-03-05	?
2000000009	26	male	1968	50s	Korea	Gyeonggi-do	Siheung-si	?	contact with ...	1	2,000,000,010	0	?	2020-02-09	?	?
2000000010	27	female	1982	30s	China	Gyeonggi-do	Siheung-si	?	overseas inflow	1	?	40	?	2020-02-09	2020-03-07	?
2000000011	28	female	1989	30s	China	Gyeonggi-do	Goyang-si	?	contact with ...	2	2,000,000,001	1	?	2020-02-10	2020-02-17	?
1000000013	29	male	1938	80s	Korea	Seoul	Jongno-gu	?	contact with ...	4	1,000,000,017	117	?	2020-02-16	?	?
1000000014	30	female	1952	60s	Korea	Seoul	Jongno-gu	?	contact with ...	5	1,000,000,013	27	2020-02-06	2020-02-16	2020-03-12	?
1200000031	31	female	1959	60s	Korea	Daegu	?	?	Shincheonji C...	?	?	1160	?	2020-02-18	?	?
1200000033	33	female	1980	40s	Korea	Daegu	Jung-gu	?	contact with ...	?	1,200,000,031	?	?	2020-02-18	?	?
1200000034	34	male	1996	20s	Korea	Daegu	Jung-gu	?	Shincheonji C...	?	?	?	?	2020-02-18	?	?
1200000035	35	female	1994	20s	Korea	Daegu	Nam-gu	?	Shincheonji C...	?	?	?	?	2020-02-18	?	?
1200000036	36	female	1972	40s	Korea	Daegu	Nam-gu	?	Shincheonji C...	?	?	?	?	2020-02-18	?	?
1200000038	38	female	1963	50s	Korea	Daegu	Nam-gu	TRUE	?	?	?	?	?	2020-02-18	?	2020
2000000012	32	female	2009	10s	Korea	Gyeonggi-do	Suwon-si	?	contact with ...	?	2,000,000,007	?	?	2020-02-18	2020-03-04	?
6015000001	37	male	?	40s	Korea	Gyeongsang...	Yeongcheon-si	?	etc	?	?	?	?	2020-02-18	2020-02-26	?
6015000002	39	female	?	60s	Korea	Gyeongsang...	Yeongcheon-si	?	Shincheonji C...	?	?	?	?	2020-02-18	2020-03-03	?

Tableau 2 : Liste des patients contaminées avec les mentions Shincheonji du 18 Février 2020.

Table "PatientInfo.csv" - Rows: 14 Spec - Columns: 17 Properties Flow Variables

Row ID	I global_...	S sex	I birth_y...	S age	S country	S province	S city	S disease	S ▲ infection_case	I infection_...	D infected_by
1000000001	2	male	1964	50s	Korea	Seoul	Gangseo-gu	?	overseas inflow	1	?
1000000002	5	male	1987	30s	Korea	Seoul	Jongno-gu	?	overseas inflow	1	?
1000000003	6	male	1964	50s	Korea	Seoul	Jongno-gu	?	contact with patient	2	2,002,000,001
1000000004	7	male	1991	20s	Korea	Seoul	Mapo-gu	?	overseas inflow	1	?
1000000005	9	female	1992	20s	Korea	Seoul	Seongbuk-gu	?	contact with patient	2	1,000,000,002
1000000006	10	female	1966	50s	Korea	Seoul	Jongno-gu	?	contact with patient	3	1,000,000,003
1000000007	11	male	1995	20s	Korea	Seoul	Jongno-gu	?	contact with patient	3	1,000,000,003
1000000008	13	male	1992	20s	Korea	Seoul	etc	?	overseas inflow	1	?
1000000009	19	male	1983	30s	Korea	Seoul	Songpa-gu	?	overseas inflow	2	?
1000000010	21	female	1960	60s	Korea	Seoul	Seongbuk-gu	?	contact with patient	3	1,000,000,003
1000000011	23	female	1962	50s	China	Seoul	Seodaemun-gu	?	overseas inflow	1	?
1000000012	24	male	1992	20s	Korea	Seoul	etc	?	overseas inflow	1	?
1000000013	29	male	1938	80s	Korea	Seoul	Jongno-gu	?	contact with patient	4	1,000,000,017
1000000014	30	female	1952	60s	Korea	Seoul	Jongno-gu	?	contact with patient	5	1,000,000,013

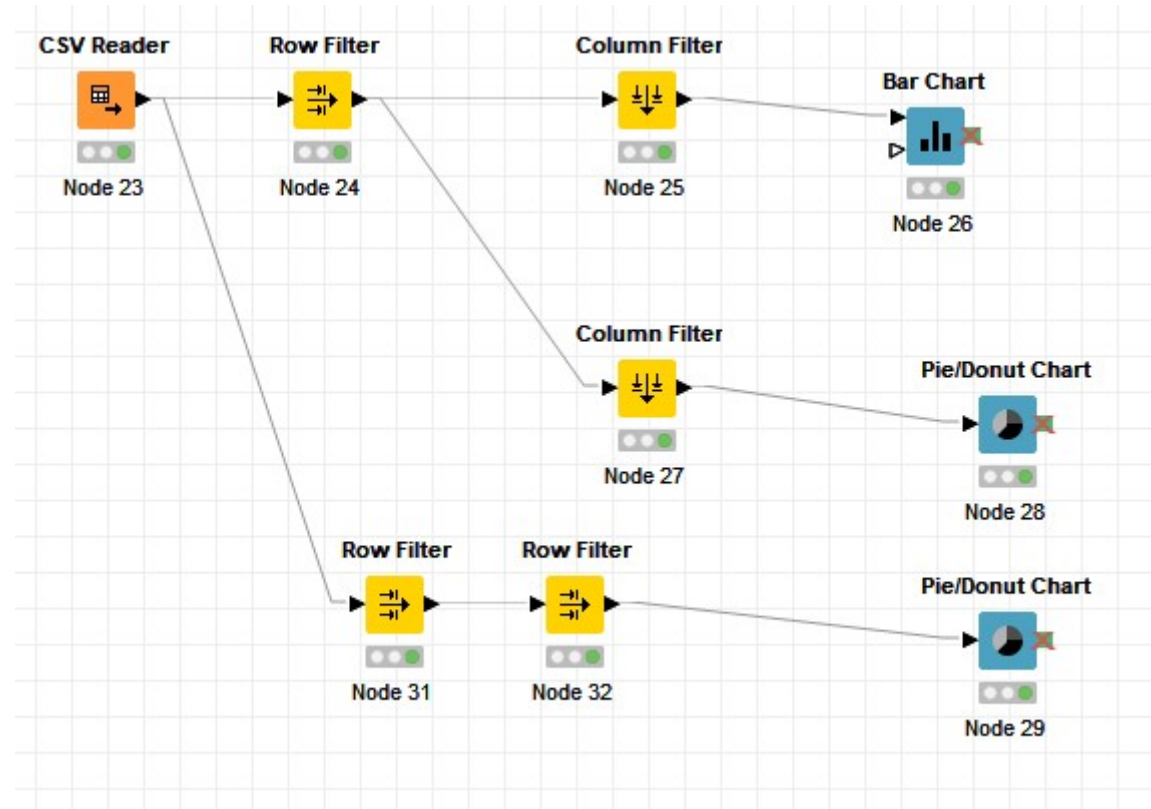
Tableau 3 : Liste des patient dont le nombre contamination par le même contact patient est le plus élevée

1000000001	2	#####	Seoul	Jung-gu	hospital
1000000002	5	#####	Seoul	Seongbuk-gu	etc
1000000002	5	#####	Seoul	Seongbuk-gu	store
1000000002	5	#####	Seoul	Seongdong-gu	public_transpc
1000000002	5	#####	Seoul	Seongbuk-gu	public_transpc
1000000002	5	#####	Seoul	Seongbuk-gu	store
1000000002	5	#####	Seoul	Seongbuk-gu	restaurant
1000000002	5	#####	Seoul	Dongdaemun-gu	store
1000000002	5	#####	Seoul	Seongbuk-gu	etc
1000000002	5	#####	Seoul	Jungnang-gu	store
1000000002	5	#####	Seoul	Jungnang-gu	restaurant
1000000002	5	#####	Seoul	Gangnam-gu	public_transpc
1000000002	5	#####	Seoul	Jungnang-gu	hospital
1000000002	5	#####	Seoul	Jungnang-gu	hospital
1000000003	6	#####	Seoul	Jongno-gu	church
1000000003	6	#####	Seoul	Jongno-gu	restaurant
1000000004	7	#####	Seoul	Jungnang-gu	hospital
1000000005	9	#####	Seoul	Jungnang-gu	hospital
1000000006	10	#####	Gyeonggi-do	Goyang-si	beauty_salon
1000000007	11	#####	Gyeonggi-do	Goyang-si	beauty_salon
1000000008	13	#####	Seoul	Jung-gu	hospital
1000000009	19	#####	Gyeonggi-do	Seongnam-si	etc
1000000009	19	#####	Gyeonggi-do	Seongnam-si	etc
1000000009	19	#####	Seoul	Songpa-gu	etc
1000000009	19	#####	Seoul	Songpa-gu	restaurant
1000000009	19	2/1/2020	Seoul	Songpa-gu	etc
1000000009	19	2/1/2020	Incheon	Yeonsu-gu	store
1000000009	19	2/1/2020	Incheon	Yeonsu-gu	store
1000000009	19	2/3/2020	Gyeonggi-do	Seongnam-si	etc
1000000009	19	2/3/2020	Gyeonggi-do	Seongnam-si	restaurant
1000000009	19	2/5/2020	Seoul	Jungnang-gu	hospital
1000000010	21	#####	Seoul	Jongno-gu	church
1000000010	21	#####	Seoul	Jongno-gu	church
1000000010	21	2/3/2020	Seoul	Seongbuk-gu	hospital
1000000010	21	2/5/2020	Seoul	Jongno-gu	hospital
1000000011	23	2/6/2020	Seoul	Seodaemun-gu	hospital
1000000011	23	2/2/2020	Seoul	Jung-gu	store
1000000011	23	2/2/2020	Seoul	Seodaemun-gu	etc
1000000011	23	2/2/2020	Seoul	Mapo-gu	store
1000000011	23	2/2/2020	Seoul	Seodaemun-gu	etc
1000000013	29	2/5/2020	Seoul	Jongno-gu	etc
1000000013	29	#####	Seoul	Seongbuk-gu	hospital
1000000013	29	#####	Seoul	Jongno-gu	hospital
1000000013	29	2/4/2020	Gyeonggi-do	Uiyeongbu-si	public_transpc

Tableau 4 : Mise en valeur de 'L'Hopitale' comme lieu de passage principale favorisant la contamination du virus.

Row ID	I global_...	S sex	I birth_y...	S age	S country	S province	S city	S disease	S infection_case	I infectio...	D infected_by
2000000001	3	male	1966	50s	Korea	Gyeonggi-do	Goyang-si	?	overseas inflow	1	?
2000000002	4	male	1964	50s	Korea	Gyeonggi-do	Pyeongtaek-si	?	overseas inflow	1	?
2000000003	12	male	1971	40s	China	Gyeonggi-do	Bucheon-si	?	contact with patient	2	?
2000000004	14	female	1980	40s	China	Gyeonggi-do	Goyang-si	?	contact with patient	3	2,000,000,003
2000000005	15	male	1977	40s	Korea	Gyeonggi-do	Suwon-si	?	contact with patient	2	2,000,000,002
2000000006	17	male	1982	30s	Korea	Gyeonggi-do	Guri-si	?	overseas inflow	2	?
2000000007	20	female	1978	40s	Korea	Gyeonggi-do	Suwon-si	?	contact with patient	3	2,000,000,005
2000000008	25	female	1946	70s	Korea	Gyeonggi-do	Siheung-si	?	contact with patient	2	2,000,000,010
2000000009	26	male	1968	50s	Korea	Gyeonggi-do	Siheung-si	?	contact with patient	1	2,000,000,010
2000000010	27	female	1982	30s	China	Gyeonggi-do	Siheung-si	?	overseas inflow	1	?
2000000011	28	female	1989	30s	China	Gyeonggi-do	Goyang-si	?	contact with patient	2	2,000,000,001

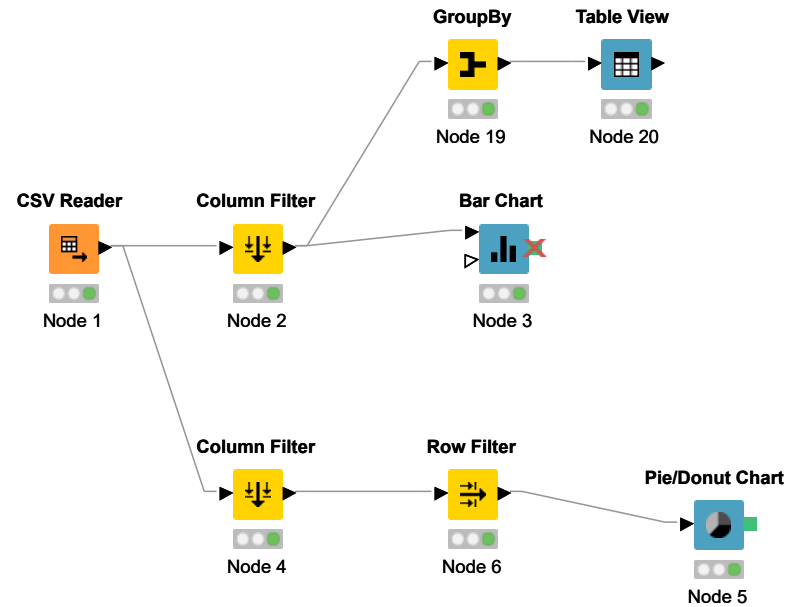
Tableau 5 : Mise en évidence d'une chaine de contamination au sein de la provenance de Gyeonggi-do avec un flux étranger chinois.

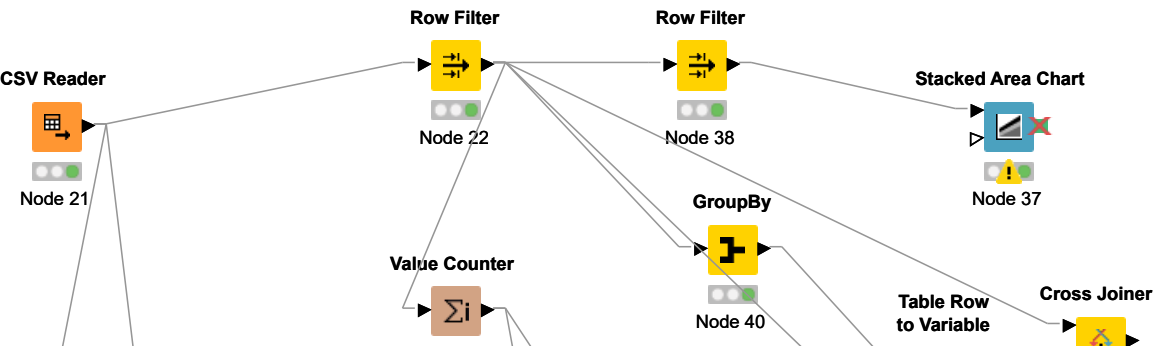
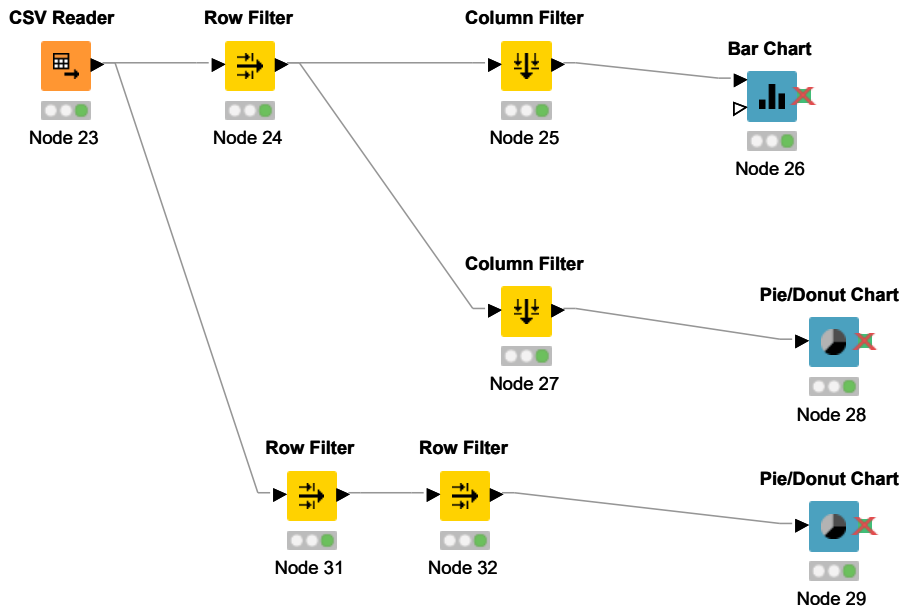
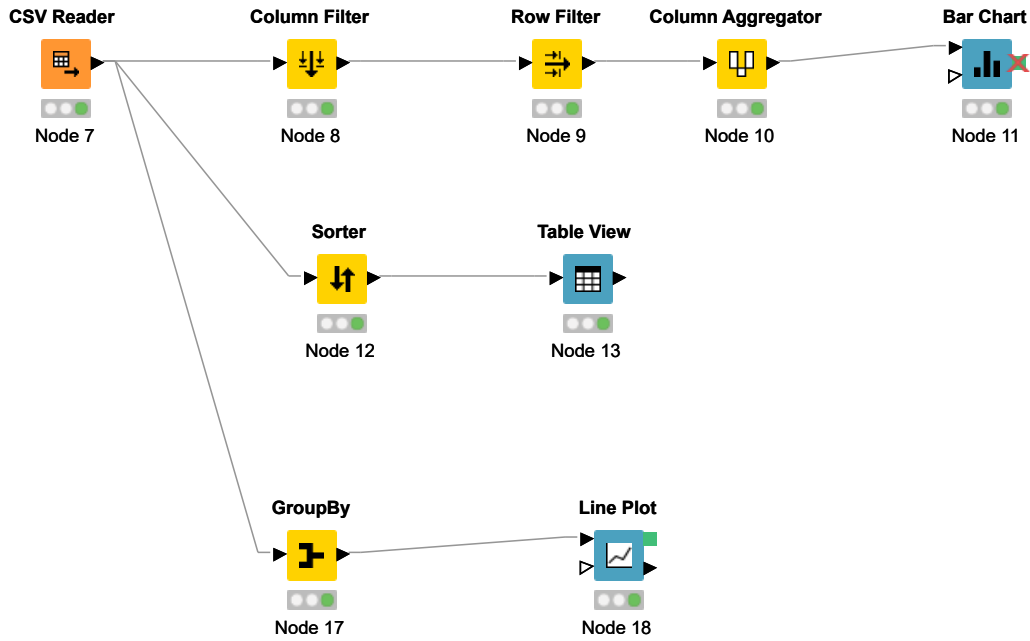


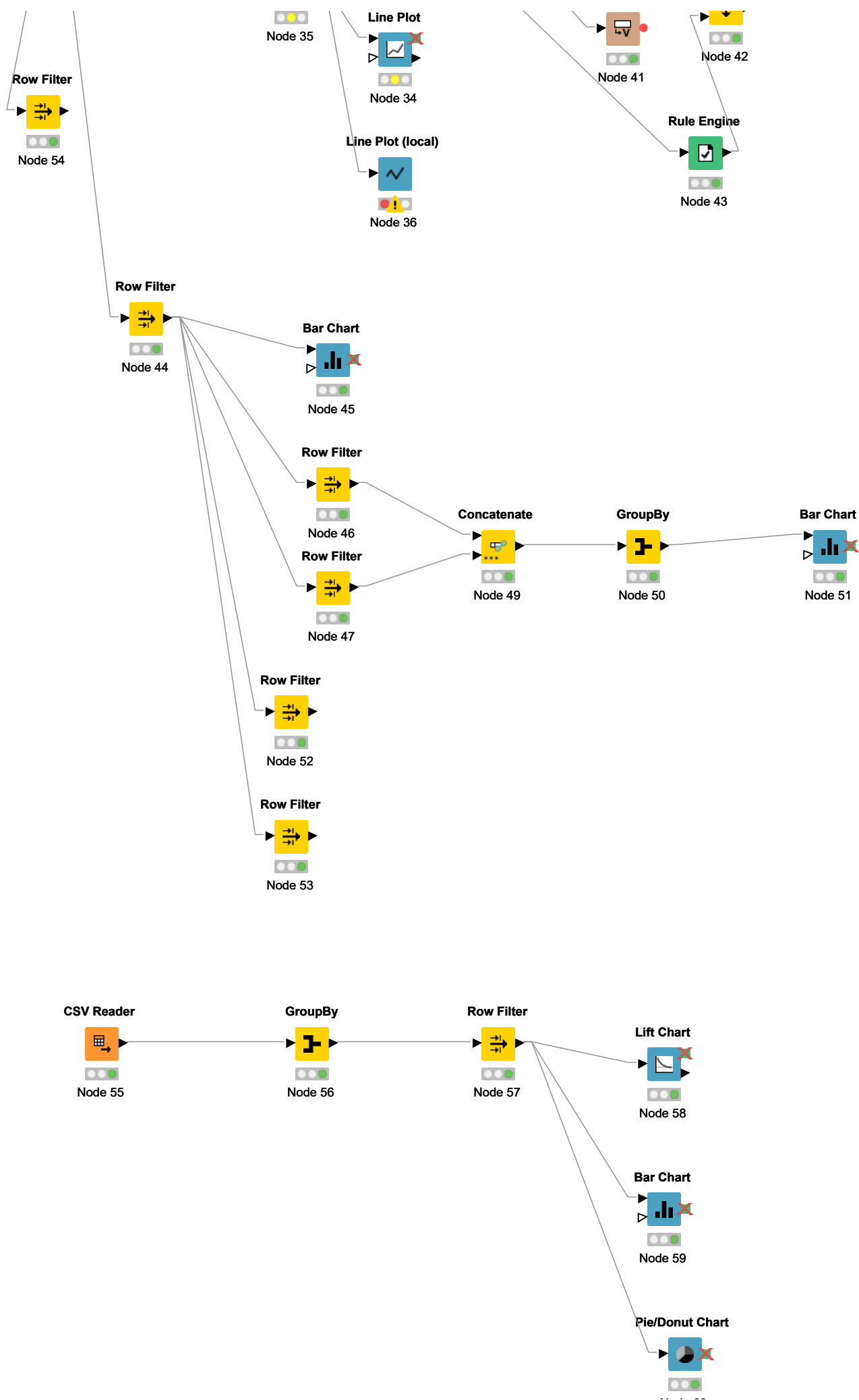
Flow Chart 3 : workflow pour les données de Seoul et Gyeonggi-do respectivement

1200000031	31	2/6/2020	Daegu	Dong-gu	etc	35.8866	128.6353
1200000031	31	2/7/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/7/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/8/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/9/2020	Daegu	Nam-gu	church	35.84121	128.5982
1200000031	31	2/9/2020	Daegu	Nam-gu	etc	35.83972	128.5666
1200000031	31	2/9/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/9/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/15/2020	Daegu	Dong-gu	lodging	35.88235	128.6621
1200000031	31	2/15/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/16/2020	Daegu	Nam-gu	church	35.84121	128.5982
1200000031	31	2/16/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/17/2020	Daegu	Suseong-gu	hospital	35.84461	128.6122
1200000031	31	2/17/2020	Daegu	Suseong-gu	hospital	35.85839	128.6352
1200000031	31	2/17/2020	Daegu	Seo-gu	hospital	35.8596	128.5406

Tableau 8 : Historique des lieux visité de la part du patient 31 (12000000031)







Node 60

Workflow : KNIME