

Projet PE : Pandas

Astrid Allegretti - Vicente Spada

Sommaire :

I - Etude du personnel

1) Description

2) Moral du personnel (par secteur, par évènement)

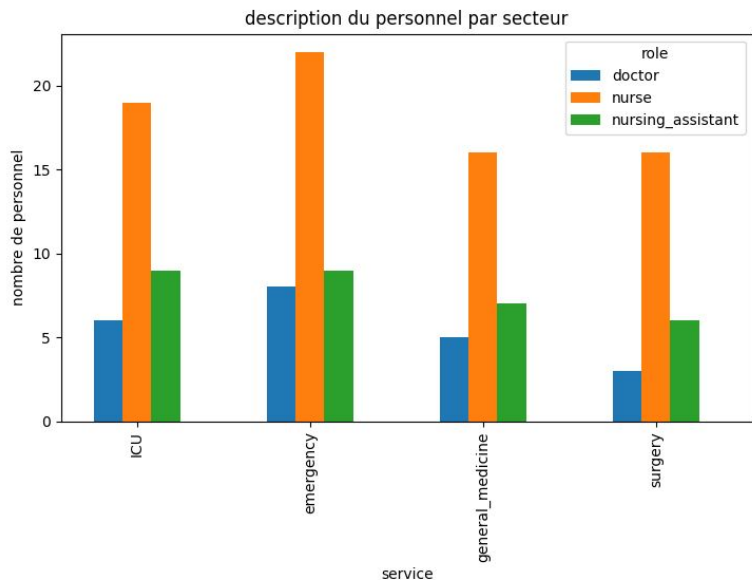
II - Etude des hospitalisations

III - Accessibilité

I - Etude du personnel

1) Description :

```
staff_schedule.groupby(by=['service','role'])['staff_id'].nunique().unstack(1).plot.bar(  
    ylabel='nombre de personnel',title='description du personnel par secteur')
```



Descriptif :

ICU : 6 docteurs, 19 infirmiers, 9 assistants infirmiers

Urgences : 8 docteurs, 22 infirmiers, 9 assistants infirmiers

Chirurgie : 3 docteurs, 16 infirmiers, 6 assistants infirmiers

Médecine générale : 5 docteur, 16 infirmiers, 7 assistants infirmiers

Analyse :

Services des urgences et ICU plus développés (plus de personnel dans toutes les catégories qu'en chirurgie et médecine générale).

Effectifs en médecine générale et chirurgie comparables mais nombre d'assistants et docteurs plus faible dans le premier secteur.

I - Etude du personnel

2) Moral du personnel :

a) Par secteur:

```
# moral moyen du personnel par secteur
df_mormoy = dfservices[['service','staff_morale','patient_satisfaction']].groupby(by = 'service').mean()
df_mormoy.plot.bar()
df_mormoy.describe()
df_mormoy
```

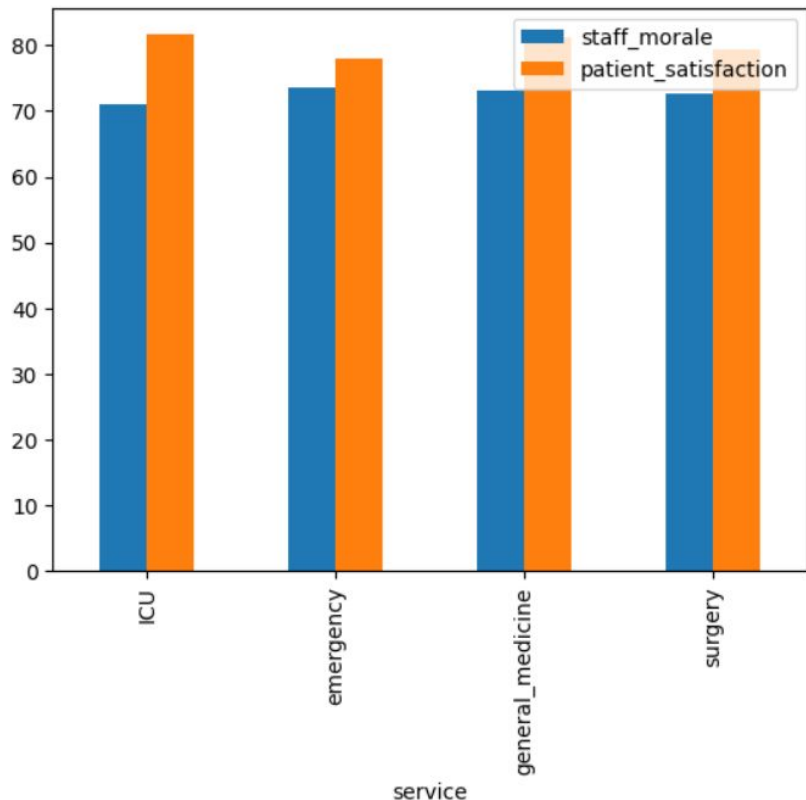
Description programme :

Regroupement des données par service et calcul de la moyenne de la satisfaction des patients et du moral du personnel

	staff_morale	patient_satisfaction
service		
ICU	70.980769	81.615385
emergency	73.557692	77.884615
general_medicine	73.096154	81.230769
surgery	72.634615	79.269231

I - Etude du personnel

2) Moral du personnel :



Analyse :

Faible de variation du moral moyen en fonction du secteur (autour de 72%), valeurs moyennes variables de 2,5 %. Pas de réelle tendance visible permettant de qualifier un 'meilleur' ou 'moins bon' secteur.

Service avec le moral le plus bas: ICU (~71 %)

Service avec le moral le plus haut: Urgences (~73,6 %)

Valeurs de moral plutôt bonnes: l'établissement ne semblent pas impactés par une crise hospitalière qui pourrait diminuer significativement le moral du personnel.

Limite : Ces valeurs moyennes sont-elles représentatives de l'état de moral de chaque secteur ? Quelle fiabilité leur accorder ?

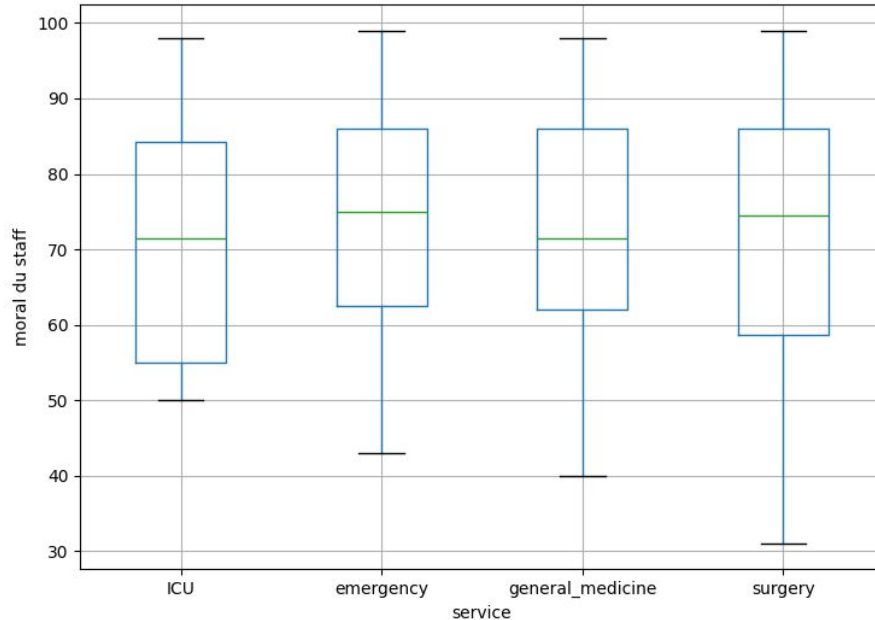
I - Etude du personnel

On regarde la description statistique du moral du personnel par secteur:

2) Moral du personnel :

```
services_weekly.boxplot(column='staff_morale', by='service', ylabel='moral du staff')  
plt.suptitle('répartition du moral du staff par service')
```

répartition du moral du staff par service



Analyse:

Valeur minimale de moral la plus basse pour le service de chirurgie (31 pts). Cependant, la répartition du premier et 3e quartile à 59 et 85 points de satisfaction semble indiquer qu'il s'agit d'une réponse singulière.

Les services ICU et chirurgie sont ceux avec les plus grands écarts entre 1er et 3e quartile (écart de 30 pts). Les valeurs de moral au sein de ces services ont donc tendance à être plus variables et non concentrées autour de la valeur médiane.

Médecine générale: La barre médiane est plus proche du 1er quartile que du 3e. Il semblerait que les valeurs de moral basses soient plus rapprochées de la médiane que les valeurs de moral élevées.

I - Etude du personnel

2) Moral du personnel :

Conclusion sur la validité des valeurs moyennes de moral par secteur:

Le graphe des valeurs moyennes de moral par secteur dissimule la **répartition des valeurs** et certaines **différences entre secteurs**. Au sein de chaque service se trouvent des membres du personnel **au moral faible** (valeurs minimales entre 50 et 30 pts en fonction du service). Cependant, on constate des valeurs **plus variables** pour le service de **chirurgie** (peuvent être dues à des situation de travail plus difficilement supportables (patients gravement blessées,...) et plus homogène pour le service ICU (soins intensifs, travail plus régulier au contact du patient).

De façon plus générale, la moitié des effectifs de chaque secteur se situe à des valeurs de moral comprises entre 85 et 60 pts et se répartit également au dessus et en dessous de 72 pts (valeur médiane globale).

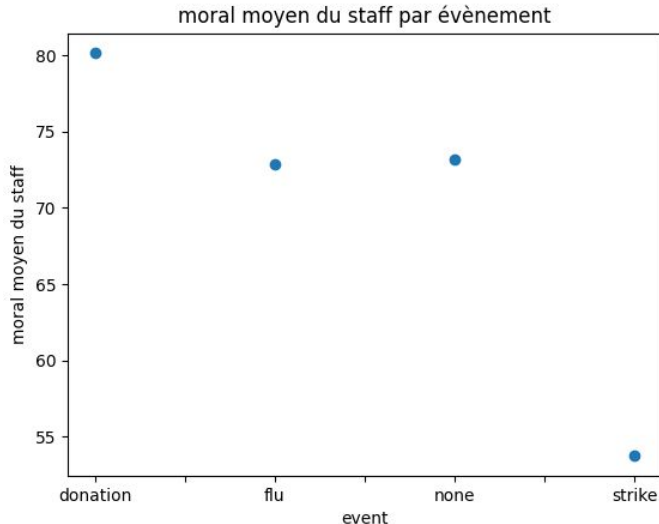
On constate un écart entre les différentes médianes de 4 pts, ce qui confirme l'**homogénéité** des **secteurs** en terme de moral. On retrouve une cohérence des valeurs de moral médianes avec celles moyennes autour de 72 pts, ce qui permet d'accorder une **validité de représentation** aux valeurs moyennes du graphique précédent.

I - Etude du personnel

2) Moral du personnel :

b) Par évènement:

```
services_weekly.groupby(by='event')['staff_morale'].mean().plot(  
style='o', ylabel='moral moyen du staff', title='moral moyen du staff par évènement')
```



Analyse et interprétation:

Moral du personnel assez haut (80 pts) lors de périodes de dons (événement qui montre le soutien public au secteur hospitalier, encourage le personnel dans son travail et améliore leur moral).

Moral plus bas (73 pts) en absence d'événements ou en période de grippe. La maladie ne semble pas spécialement affecter leur moral, ce qui peut s'expliquer par le fait qu'il soit dans leur côtoyer la maladie.

En période de grève : moral significativement plus bas (53 pts). Cohérent avec les revendications et le mécontentement exprimés par le personnel en période de grève.

I - Etude du personnel

2) Présence :

```
## Taux de présence de chaque membre du personnel:
```

```
nb_week = int(dfedt.week.max())  
df_pres = dfedt[['staff_name', 'present']].groupby(by = 'staff_name').sum()/nb_week  
df_recap = df_pres.merge(dfedt[['staff_id', 'staff_name', 'role']], on = 'staff_name', how = 'left')  
df_final = df_recap.drop_duplicates()  
df_final
```

Description :

Calcul du taux de présence pour chaque membre du personnel puis ajout de sa fonction et son identifiant en le joignant avec le DataFrame de départ.

I - Etude du personnel

2) Présence :

	staff_name	present	staff_id	role
0	Aaron Wise	0.596154	STF-5697a657	doctor
52	Abigail Shaffer	0.634615	STF-d114bde8	doctor
104	Aimee Montoya	0.596154	STF-8d812ffa	nursing_assistant
156	Allison Hill	0.576923	STF-b77cdc60	doctor
208	Amber Kidd	0.538462	STF-5157d809	doctor
...
6292	Whitney Peters	0.596154	STF-4f1df27e	nurse
6344	William Baker	0.576923	STF-e0c7759c	nurse
6396	William Herrera	0.596154	STF-05591498	nurse
6448	Zachary Ferrell	0.596154	STF-a493e575	nurse
6500	Zachary Hicks	0.634615	STF-d1d59a8b	nurse

Description statistique:

	present
count	126.000000
mean	0.599817
std	0.026726
min	0.538462
25%	0.576923
50%	0.596154
75%	0.615385
max	0.653846

Analyse :

Présence moyenne des membres du personnel de 60 % et faible écart-type des pourcentages de présence (0,027).

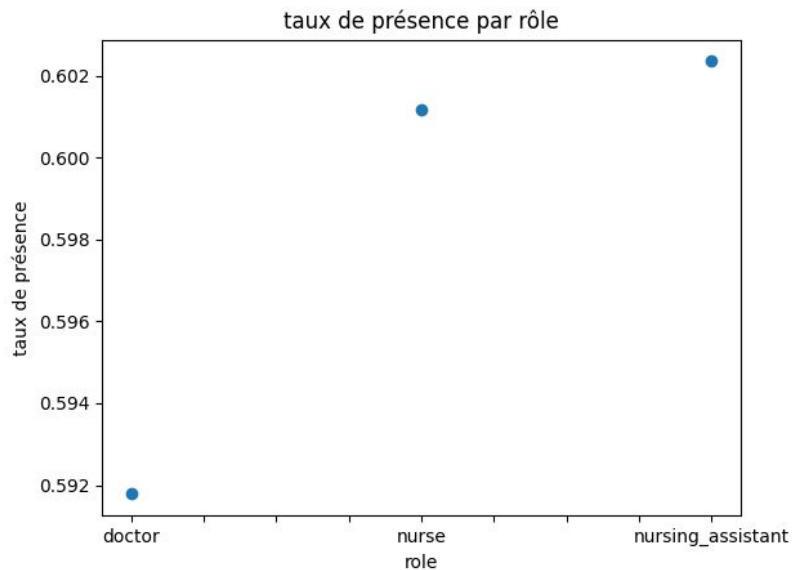
Il semble que les membres du personnel soient présents de façon régulière, sans excès ou absence total au travail (valeurs minimale et maximale de 0,538 et 0,654).

I - Etude du personnel

2) Présence par métier :

Taux de présence
des employés ~
60 %

present	
role	
doctor	0.591783
nurse	0.601159
nursing_assistant	0.602357



Analyse :

Faible différence de taux de présence entre les différents métiers (0,01% de différence maximale).

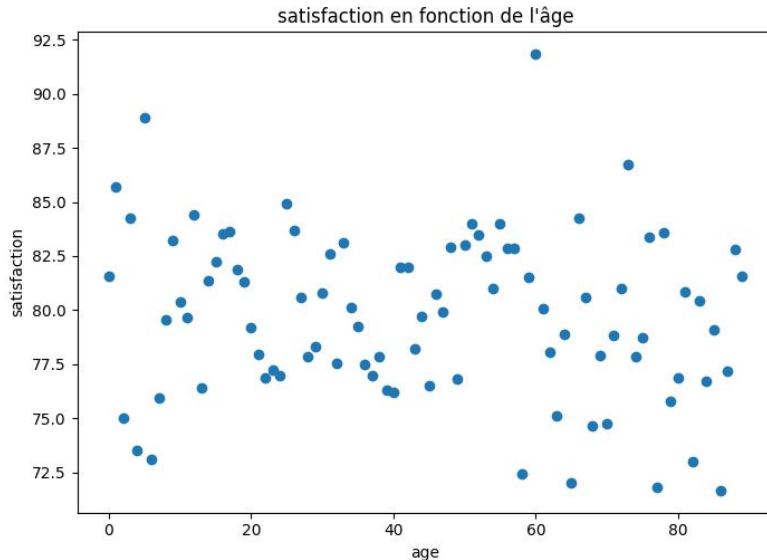
Les infirmiers assistant sont légèrement plus présents que les infirmiers, eux-mêmes plus présents que les docteurs. Ces différences de présence peuvent s'expliquer par la différence de salaires (les assistants ou infirmiers chercheront à travailler plus pour augmenter leurs revenus) et par la différence d'implication nécessaire au bon fonctionnement du service (un docteur réalisera principalement des consultations et visites de contrôle et interviendra plus rarement lors d'opérations. A l'inverse, les infirmiers effectuent un suivi continu des patients et sont présents tout au long des opérations.)

II - Etude des hospitalisations

1) Satisfaction en fonction de l'âge

```
patients["tranche_age"] = pd.cut(
patients.age, bins=[0,20,60,patients.age.max()], labels=['young','adult','elderly'])
```

```
patients.groupby(by='tranche_age')['satisfaction'].mean().plot(
style='o', ylabel='satisfaction', title='satisfaction en fonction de la tranche d\'âge')
```

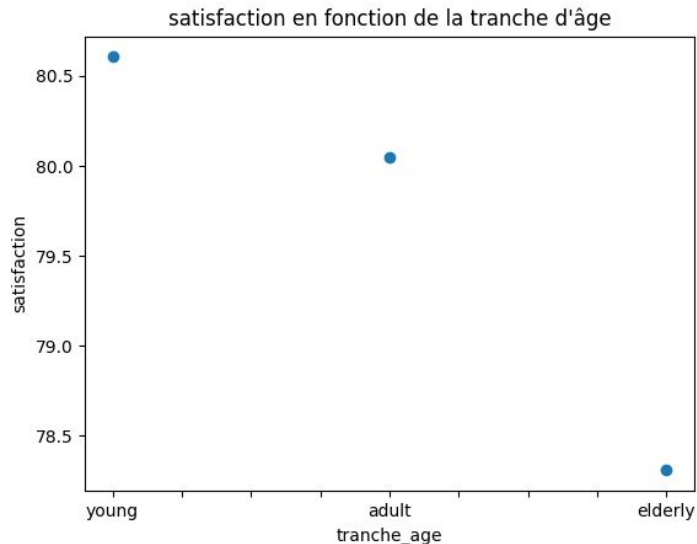


Analyse :

On réalise un premier graphique représentant la satisfaction des patients en fonction de leur âge. Le nuage de points ne permet pas de distinguer de réelle tendance permettant de corrélérer satisfaction et âge du patient. Le choix a alors été fait de moyenner les résultats de satisfaction par tranche d'âge et de réorganiser ainsi les résultats de satisfaction selon des catégories d'âge prédéfinies.

II - Etude des hospitalisations

1) Satisfaction en fonction de l'âge



Limite : Ces résultats dépendent évidemment de la définition des catégories d'âge et il faut également relativiser la validité de représentation de la moyenne, les effectifs dans les différentes catégories pouvant varier fortement.

Analyse :

Les résultats de satisfaction sont discriminés selon 3 catégories : personnes âgées (plus de 60 ans), adultes (entre 20 et 60 ans) et jeunes (moins de 20 ans).

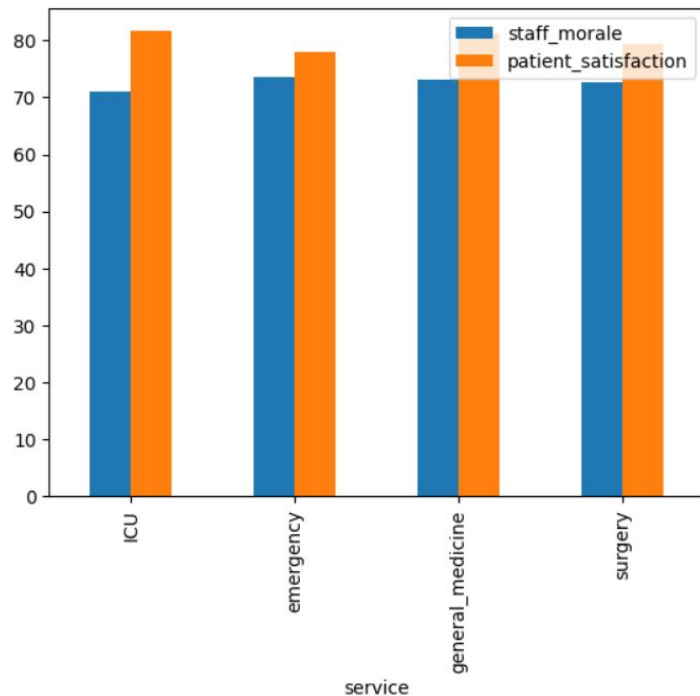
La satisfaction des patients semble alors décroître en fonction de l'âge. Cette distinction semble plus prononcée pour les personnes âgées (en moyenne la satisfaction est plus basse de 2 à 1,5 points par rapport aux autres catégories). Elle est minime entre jeunes et adultes avec seulement 0,5 points d'écart entre les satisfactions moyennes.

Ce résultat pourrait traduire le fait que des personnes plus âgées, ayant plus d'expérience, tendent à moins tolérer certains aspects négatifs de leur période de soin. Cependant, cette baisse de la satisfaction ne peut être directement reliée à des pathologies plus lourdes lors du vieillissement, le graphe II.7 laissant suggérer que l'absence de lien direct entre vieillissement et augmentation de la durée d'hospitalisation.

II - Etude des hospitalisations

2) Satisfaction en fonction du service:

Programme : cf moral des employés par secteur I.2)



	staff_morale	patient_satisfaction
service		
ICU	70.980769	81.615385
emergency	73.557692	77.884615
general_medicine	73.096154	81.230769
surgery	72.634615	79.269231

Analyse :

Secteur avec le plus de satisfaction : ICU (81 %)
Secteur avec le moins de satisfaction : Urgences (78 %)
Ecart assez faibles (de l'ordre de 2 % entre les services).
La satisfaction générale des patients reste assez bonne (autour de 79 pts)

II - Etude des hospitalisations

2) Satisfaction en fonction du service:

Mise en perspective des résultats :

Ces différences de résultats peuvent s'expliquer selon des critères de durée passée au sein du service ainsi que de disponibilité de ces services.(graphes II.6 et III.2)

En effet, le service de soins intensifs (ICU) accompagne des patients dont un organe est défaillant, nécessitant un soutien et un accompagnement sur une durée généralement plus longue et prévue par le service. Ce service présentant de plus une meilleure disponibilité (graphe III.2), les patients peuvent ressentir plus de confort lors de leur passage.

A l'inverse, les patients passant par les urgences restent sur de moins longues durées et dans un service surchargé (graphe II.2), ce qui peut leur laisser l'impression d'avoir été moins bien accompagné au cours de leur séjour. Il faut ajouter à cela que les hospitalisations aux urgences sont pour la plupart imprévues et ne sont pas vécues comme un événement positif pour les patients devant s'y rendre, ce qui peut influencer sur leur note de satisfaction.

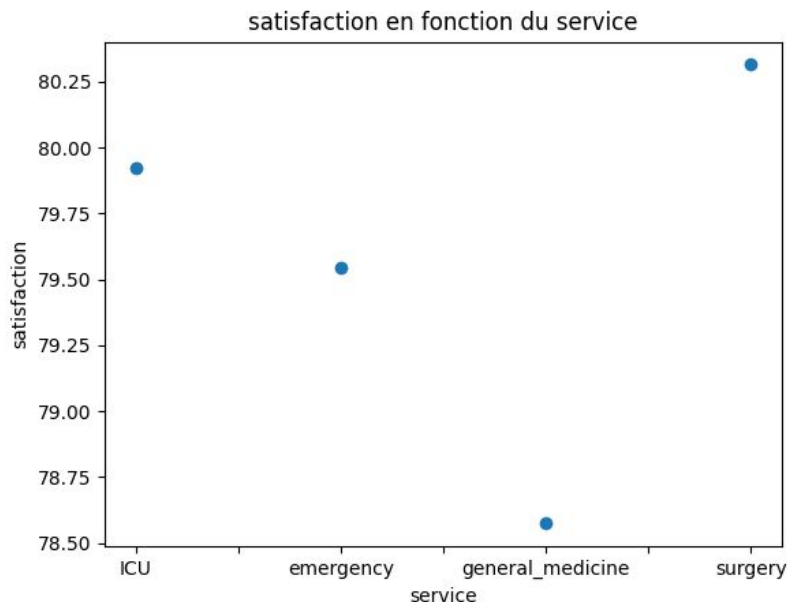
Le résultat assez élevé du service de médecine générale peut quant à lui s'expliquer par des durées de séjour très faibles (graphe II.6), des hospitalisations prévues et anticipées par le service, dont la disponibilité reste convenable.

Enfin, le résultat plus faible du service de chirurgie peut être expliqué par la prise en charge de pathologies engendrant de vives douleurs, au sein d'un service assez demandé et dans le cas d'hospitalisations n'étant pas nécessairement anticipées (transferts fréquents de patients depuis les urgences vers le service de chirurgie), ce qui peut contribuer à une diminution de la satisfaction des patients.

II - Etude des hospitalisations

2) Satisfaction en fonction du service :

```
patients.groupby(by='service')['satisfaction'].mean().plot(  
style='o', ylabel='satisfaction', title='satisfaction en fonction du service')
```



Description : La satisfaction la plus faible est pour la médecine générale proche de 78,6 et elle est maximale en chirurgie à 80,35.

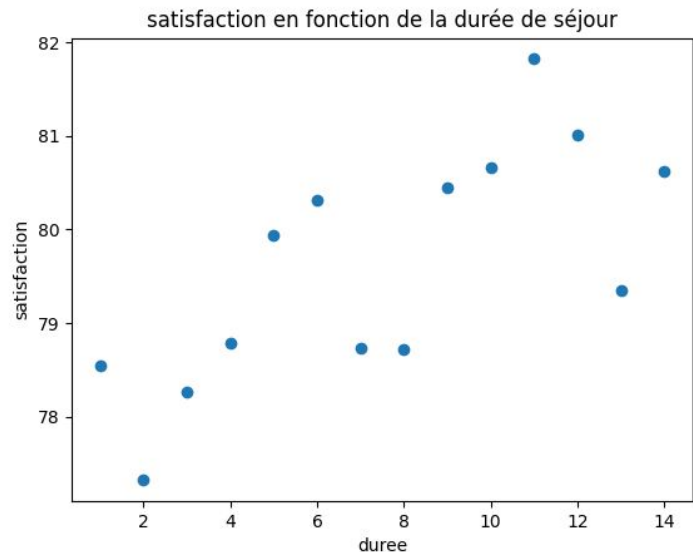
Analyse : La différence de satisfaction entre les services reste faible, ce qui indique une qualité perçue globalement homogène. Cependant, le service de médecine générale affiche une satisfaction légèrement inférieure, ce qui peut traduire une charge de travail plus élevée ou des attentes différentes des patients. À l'inverse, la chirurgie obtient la meilleure évaluation, possiblement en raison de la nature plus ponctuelle et ciblée des interventions, souvent associée à une amélioration visible de l'état du patient.

II - Etude des hospitalisations

3) Satisfaction en fonction de la durée du séjour

```
patients["duree"] = (pd.to_datetime(patients.departure_date) - pd.to_datetime(patients.arrival_date)).dt.days
```

```
patients.groupby(by='duree')['satisfaction'].mean().plot(  
style='o', ylabel='satisfaction', title='satisfaction en fonction de la durée de séjour')
```



Principe : calcul de la durée de passage des patients puis regroupement par durée des données de satisfaction et calcul de la satisfaction moyenne.

Analyse : Il semble se dégager une tendance croissante de la satisfaction en fonction de la durée d'hospitalisation. Cela peut s'expliquer par le lien créé entre patients et soignants (plus fort lorsque la durée augmente), contribuant à une plus grande satisfaction du patient, par une appréciation plus objective après plusieurs jours écoulés au sein de l'hôpital, ou par un soulagement (et donc une satisfaction) plus grand lors du départ de l'hôpital après une longue hospitalisation.

II - Etude des hospitalisations

4) Nombre d'entrée annuelle par tranche d'âge :

```
## Nombre d'entrées par tranche d'âge 0-15,15-19,20-35,36-55,56-70,>70
df_stat_age = pd.cut(dfpatients.age,bins = [0,15,20,35,55,70,max(dfpatients.age)],labels = ['enfant (moins de 15ans)', 'adolescent',
                                                                                          'jeune adulte','adulte',
                                                                                          'adulte entre 55 et 70 ans','personnage âgée'])

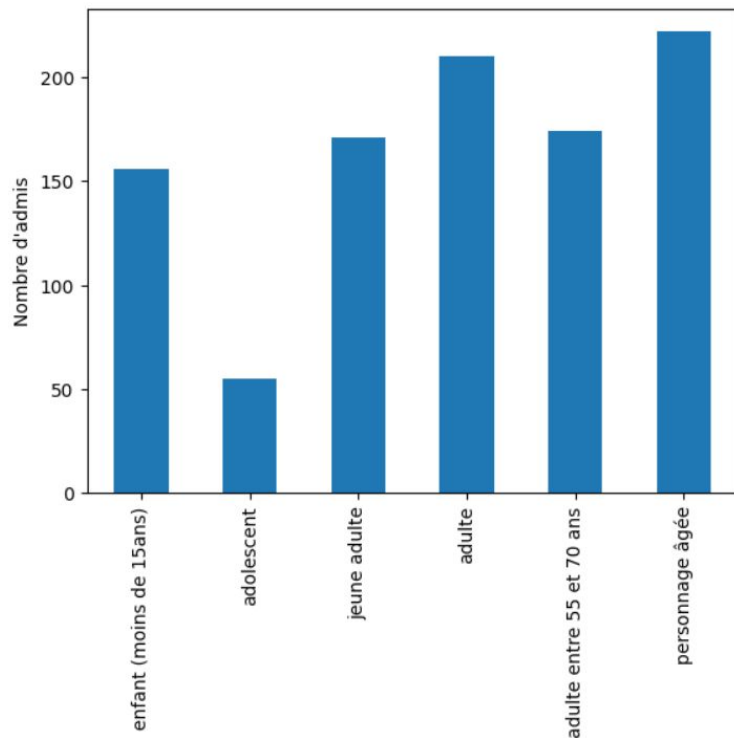
df_stat_age2=pd.DataFrame(df_stat_age)
df_stat_age2.columns = ['categories_age']
dfpatients['categories_age'] = df_stat_age2['categories_age']
df_stat_age_3 = dfpatients[['categories_age','name']].groupby(by = 'categories_age').count()
df_stat_age_3.plot.bar(ylabel = "Nombre d'admis",legend = False)
```

Description :

Création de catégories en fonction des âges des patients puis conversion du résultat en DataFrame. Ajout de la série au tableau patients, regroupement par catégorie d'âge puis décompte du nombre de patients par catégorie

II - Etude des hospitalisations

5) Nombre d'entrée annuelle par tranche d'âge :



Analyse :

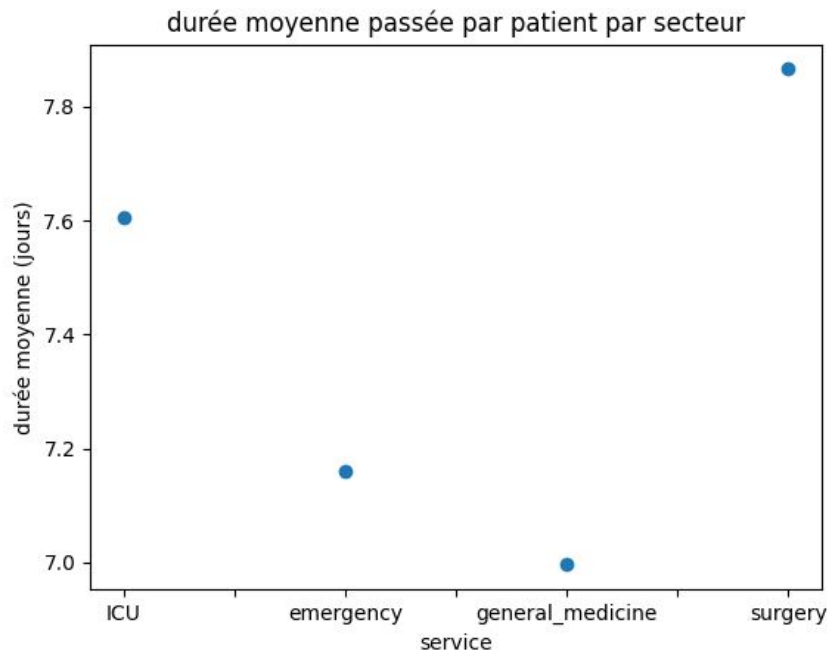
Tendance variable selon la définition des classes d'âge (en particulier, la classe adolescent recouvre une faible tranche d'âge et il serait possible d'inclure les 55 - 70 ans à une autre catégorie d'âge).

Pas de tendance significative observable mais il semble que le nombre d'entrée augmente selon l'âge des admis (150 admis chez les enfants, 160 chez les jeunes adultes, 210 chez les adultes et 220 chez les personnes âgées). Cela serait cohérent avec une dégradation de la santé lors du vieillissement.

II - Etude des hospitalisations

6) Durée moyenne passée par patient par secteur

```
patients.groupby(by=['service'])['duree'].mean().plot(  
style='o', ylabel='durée moyenne (jours)', title='durée moyenne passée par patient par secteur')
```



Description:

Médecine générale : service à la durée d'hospitalisation la plus faible (1 semaine).

Viennent ensuite les urgences (durée d'hospitalisation moyenne de 1 semaine et 5 heures), puis le service de soins palliatifs (7 jours et 17 heures) et enfin la chirurgie, à la durée d'hospitalisation moyenne la plus longue (1 semaine et 1 jours).

Analyse :

Peu d'écart entre les valeurs moyennes des différents services (au plus un jour de différence entre 2 services). Néanmoins, il est possible d'interpréter les différences de valeurs moyennes par considération de la nature des opérations réalisées dans chaque secteur.

II - Etude des hospitalisations

6) Durée moyenne passée par patient par secteur

Service	Type de soin	Conséquence(s)
ICU (soins palliatifs)	Prise en charge de patients pour pallier au dysfonctionnement d'un organe.	Prise en charge plus longue et éventuellement régulière des patients. Durée d'hospitalisation plus longue.
Chirurgie	Opérations spécifiques nécessitant une anesthésie et un accompagnement en amont et en aval de l'hospitalisation.	Durée d'hospitalisation éventuellement très longue pour des interventions lourdes. Plus de temps au sein du service pour préparer le patient à l'opération et lui apporter le suivi nécessaire une fois celui ci terminé.
Urgence	Prise en charge de tout type de problème de santé imprévu et subi par le patient.	Séjour plus court dans le service dû à de potentielles réorientation vers d'autres services une fois l'urgence traitée.
Médecine générale	Opérations plus globales et anticipées.	Hospitalisation nécessitant éventuellement moins de suivi, ce qui diminuerait le temps de séjour.

II - Etude des hospitalisations

7) Durée moyenne passée par patient en fonction de l'âge:

```
dfpatients['arrival_date'] = pd.to_datetime(dfpatients['arrival_date'],format = '%Y-%m-%d')
dfpatients['departure_date'] = pd.to_datetime(dfpatients['departure_date'],format = '%Y-%m-%d')
dfpatients['duree_hosp'] = dfpatients['departure_date'] - dfpatients['arrival_date']
df_tpsmoy = dfpatients[['categories_age','duree_hosp']].groupby(by = 'categories_age').mean()
df_tpsmoy.plot.bar()
```

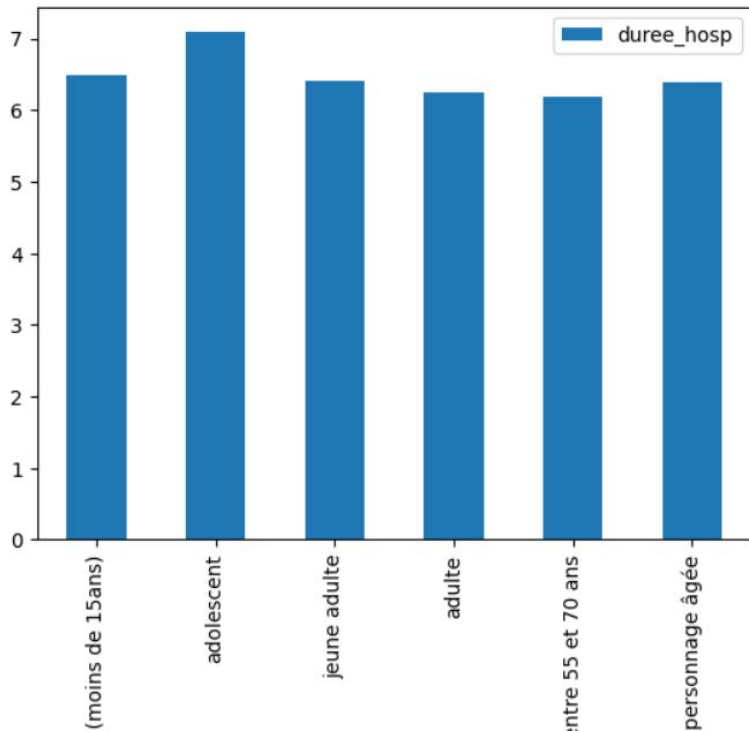
	duree_hosp
categories_age	
enfant (moins de 15ans)	7 days 12:09:13.846153846
adolescent	8 days 04:48:00
jeune adulte	7 days 09:57:53.684210526
adulte	7 days 05:22:17.142857142
adulte entre 55 et 70 ans	7 days 03:35:10.344827586
personnage âgée	7 days 09:24:19.459459459

Description :

Conversion des dates en datetime, calcul du temps d'hospitalisation pour chaque patient, regroupement par catégorie et calcul de la moyenne.

II - Etude des hospitalisations

8) Durée moyenne passée par patient en fonction de l'âge:



Description :

Temps d'hospitalisation qui semble plus élevé chez les adolescents (1 semaine en moyenne). On constate très peu de différence entre les durées d'hospitalisation moyenne des différentes classes d'âge (durée d'hospitalisation moyenne de 6 jours et demi).

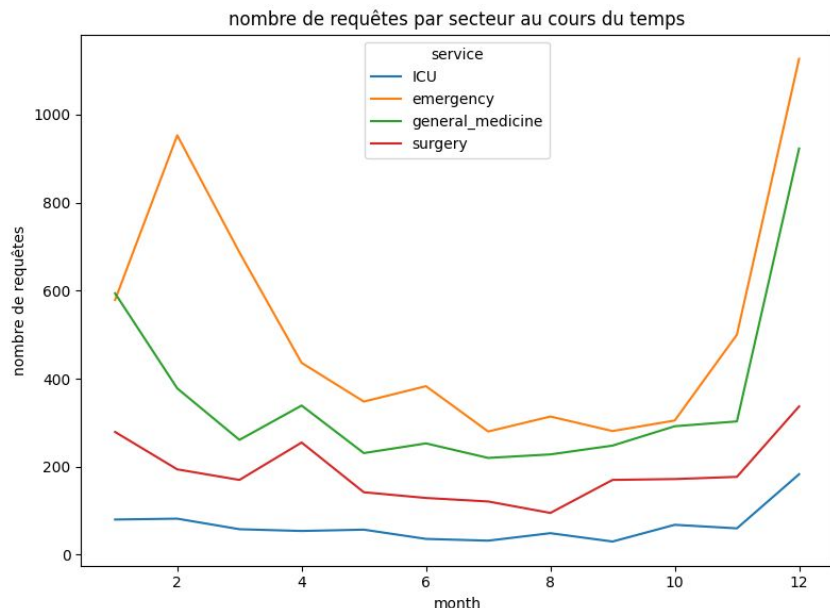
Analyse :

La valeur plus élevée chez les adolescents peut s'expliquer par la définition des classes d'âge. Peu d'individus hospitalisés se trouvant dans cette catégorie, il suffit que quelques individus aient subi des opérations nécessitant une plus longue hospitalisation pour faire augmenter significativement la moyenne. Cette valeur n'est donc pas indicatrice d'une quelconque tendance. On peut alors conclure qu'il ne semble pas y avoir de réelles corrélation entre l'âge et la durée d'hospitalisation des patients.

III - Accessibilité

1) Demande au cours du temps

```
services_weekly.groupby(by=['month','service'])['patients_request'].sum().unstack(1).plot(  
    ylabel='nombre de requêtes', title='nombre de requêtes par secteur au cours du temps')
```



Observations :

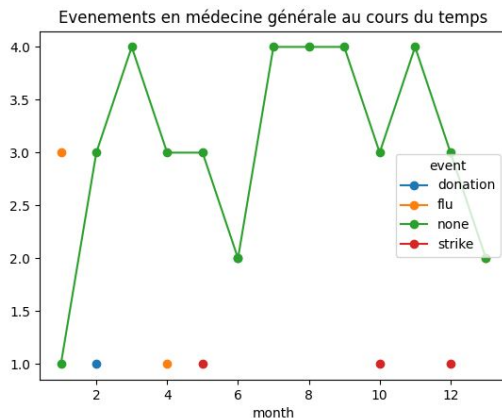
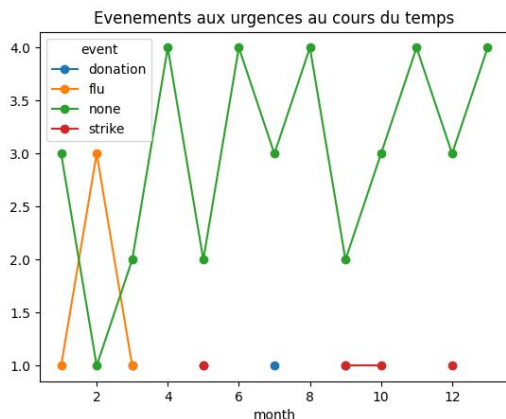
Pics de requêtes dans les secteurs de médecine générale et chirurgie aux mois de janvier et avril.

Un pic de requête aux urgences au mois de Février.

Tous les services semblent présenter un pic de requêtes au mois de décembre.

III - Accessibilité

1) Demande au cours du temps



Analyse :

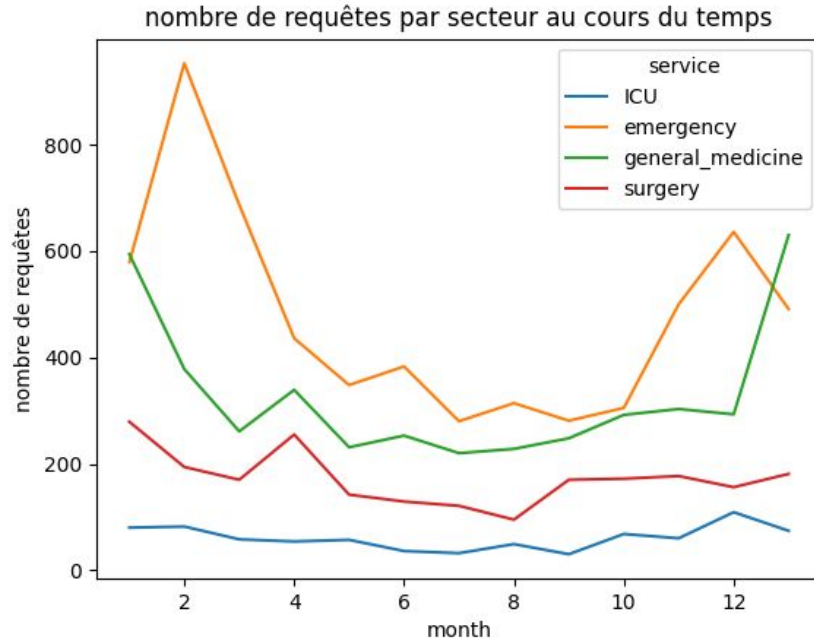
On pourrait s'attendre que le nombre de requête par secteur soit lié à des événements particuliers dans ces secteurs.

- Au urgences et en médecine générale cela coïncide bien avec une grippe en février mais ce n'est pas le cas du mois de décembre.

On a remarqué une anomalie dans les données avec un mois de décembre de 8 semaines probablement dû à un système basé sur des mois de 4 semaines, les jours "en trop" se sont fait labellisé en décembre. On a donc rajouté un 13ème mois sur les graphiques pour avoir une cohérence temporelle.

III - Accessibilité

1) Demande au cours du temps



Ainsi, le pic de décembre disparaît, c'était juste un mois qui comptait pour deux ...

Analyse :

Urgences : Pic de requêtes au mois de Février qui coïncide avec la présence d'une grippe dans le secteur au même moment. Les nombres anormalement élevés de requêtes peuvent donc être attribués à une épidémie de grippe hivernale. Le 2e pic aux mois de novembre et décembre, ne coïncidant pas véritablement avec un contexte spécifique à l'hôpital, peut quant à lui s'expliquer par la période de fêtes de fin d'années (qui connaît un nombre plus important d'incidents) ou le contexte hivernal (système immunitaire plus faible, gel sur les routes pouvant mener à des accidents de la route, accidents de skis,...).

III - Accessibilité

1) Demande au cours du temps:

Analyse (suite):

Médecine générale : Le pic au mois d'avril s'explique également par une épidémie de grippe. A l'inverse, le pic présent au mois de janvier de l'année suivante ne coïncide avec aucun événement. Il pourrait s'expliquer lui aussi par la prolifération de maladie contagieuses en hiver, créant une augmentation du nombre de requêtes pour des soins en médecine générale.

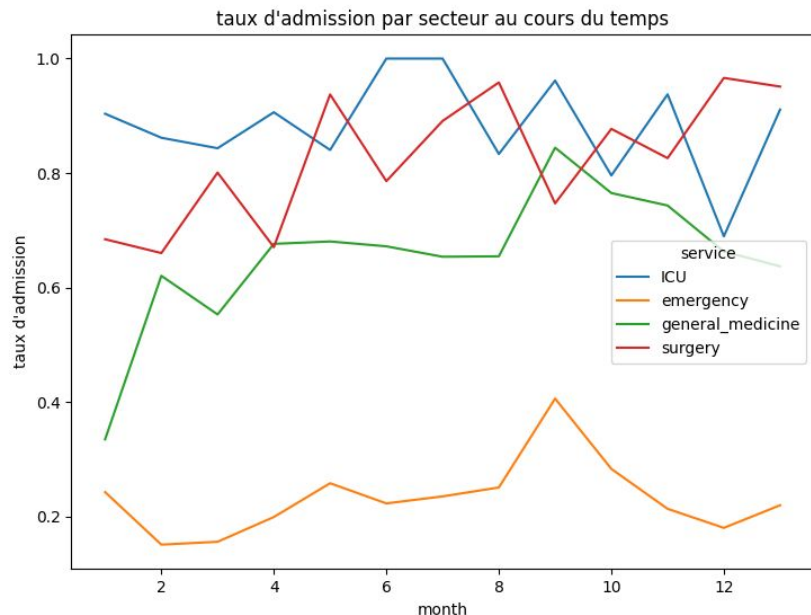
Chirurgie : On constate un pic au mois d'avril qui ne semble pas pouvoir s'expliquer par l'épidémie de grippe. En effet, les patients atteints de cette maladie se dirigeront plutôt vers les services de médecine générale ou urgences. Les hospitalisations en chirurgie étant généralement planifiées, le pic de requêtes peut être expliqué par une augmentation des disponibilités du service à cette période de l'année. (III.3)

ICU : Peu de variations au cours de l'année, cohérent avec le fait qu'il s'agisse d'un service dont les entrées sont généralement prévues et fluctuent assez peu.

III - Accessibilité

2) Taux d'admission

```
services_weekly["taux_acces"] = services_weekly['patients_admitted'] / services_weekly['patients_request']  
  
services_weekly.groupby(by=['month', 'service'])['taux_acces'].mean().unstack(1).plot(  
    ylabel='taux d\'admission', title='taux d\'admission par secteur au cours du temps')
```

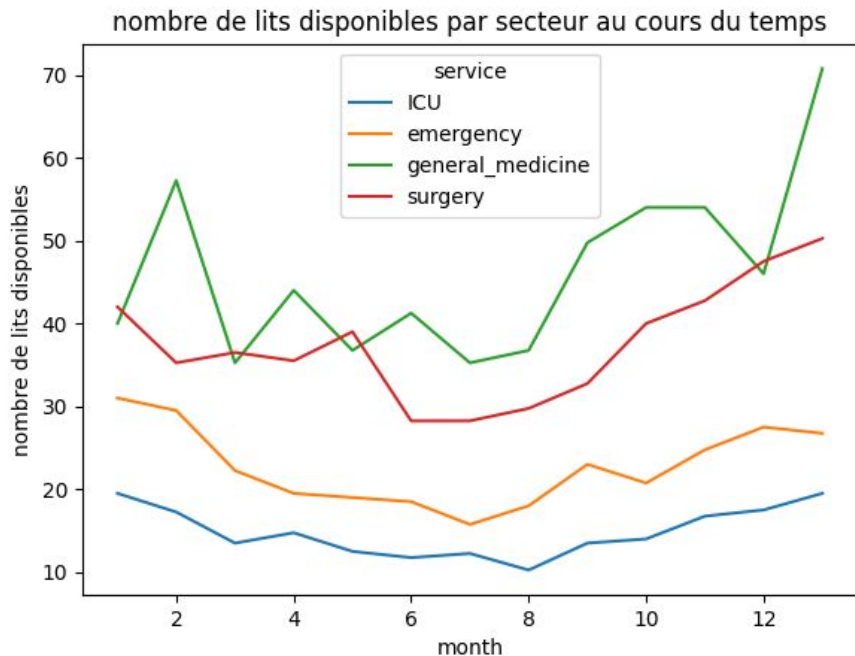


Description : Le taux d'admission est le plus élevé pour les services de réanimation (ICU) et de chirurgie, proches de 0,9 à 1 tout au long de l'année. Le service de médecine générale présente un taux intermédiaire, autour de 0,7, tandis que celui des urgences reste nettement inférieur, entre 0,1 et 0,4.

III - Accessibilité

3) Nombre de lits disponibles

```
services_weekly.groupby(by=['month','service'])['available_beds'].mean().unstack(1).plot(  
    ylabel='nombre de lits disponibles', title='nombre de lits disponibles par secteur au cours du temps')
```



Description : Le nombre de lits disponibles varie fortement selon le service. La médecine générale dispose du plus grand nombre de lits, généralement entre 35 et 70, avec un pic en fin d'année. La chirurgie suit avec des valeurs autour de 35 à 50 lits, tandis que les urgences se situent entre 15 et 30 lits. Enfin, la réanimation (ICU) présente le nombre de lits le plus faible, autour de 10 à 20, avec une légère baisse en milieu d'année avant une remontée finale.

Analyse III.2 et III.3 :

En soin intensif la demande est plus rare et peu de lits ce qui explique le taux d'admission élevé. Au urgences peu de lits mais la demande bien plus grande d'où le faible taux d'admission. En médecine générale la demande est élevée comme aux urgences mais beaucoup plus de lits ce qui permet de maintenir un taux d'admission correct autour de 0,7.