

Rapport Technique : Projet d'Analyse et de Simulation de la Capacité Hospitalière (PSL-CFX)

Introduction

Ce document détaille l'approche technique utilisée pour analyser la résilience de l'hôpital PSL-CFX (Pitié Salpêtrière - Charles Foix) face aux crises sanitaires, en s'appuyant sur les données de la période 2012-2021. L'objectif était de modéliser la capacité et l'activité de l'hôpital, de simuler l'impact d'une crise majeure comme la COVID-19, et d'identifier les risques de saturation.

1. Sources des Données

L'analyse repose sur le croisement de plusieurs fichiers de données pour obtenir une vision complète, à la fois au niveau local, régional et national.

- **PS_CF_data.xlsx** : Données internes à l'hôpital PSL-CFX.
 - **Description** : Fichier clé mais parcellaire, contenant des indicateurs d'activité (nombre de séjours MCO, hospitalisations complètes, places en ambulatoire) et de capacité (nombre de lits) pour les années **2012, 2015 et 2016** uniquement.
 - **Utilisation** : Sert de points de référence (vérité terrain) pour la reconstitution de la chronique d'activité.
- **hospitalisation_capacity.xlsx** : Données sur la capacité hospitalière au niveau national.
 - **Description** : Contient les statistiques annuelles sur le nombre de lits et de places en France, ainsi que l'activité MCO (Médecine, Chirurgie, Obstétrique).
 - **Utilisation** : Fournit les **taux de variation annuels nationaux**, utilisés pour estimer l'évolution de l'hôpital PSL-CFX durant les années où les données locales sont manquantes.
- **Passages_aux_urgence_data.xlsx** : Données haute fréquence sur les passages aux urgences.
 - **Description** : Série temporelle journalière du nombre de passages aux urgences.
 - **Utilisation** : Permet d'analyser et de visualiser la **saisonnalité** de l'activité hospitalière (pics hivernaux, etc.), information cruciale pour une gestion opérationnelle.
- **passage_urg_vague_covid.xlsx** : Données d'impact des vagues COVID-19 sur les urgences.
 - **Description** : Compare le nombre de passages aux urgences durant les vagues de COVID-19 à une période de référence, par département. Fournit des **pourcentages de variation** de

l'activité.

- **Utilisation** : Sert à quantifier le "choc" de demande exogène lié à la crise sanitaire pour la simulation. Les données des départements **75 (Paris)** et **94 (Val-de-Marne)** sont utilisées comme proxy pour l'hôpital.
- **hospitalisation_covid.xlsx** et **hospitalisation_par_hab.xlsx** : Données contextuelles sur la pandémie.
 - **Description** : Données hebdomadaires nationales et régionales sur les hospitalisations, soins critiques et décès liés à la COVID-19.
 - **Utilisation** : Permettent une analyse descriptive générale de la dynamique des vagues pandémiques en France, validant l'hétérogénéité temporelle et géographique de la crise.

2. Traitements des Données

La phase de traitement a été essentielle pour construire un jeu de données exploitable à partir de sources hétérogènes.

1. Nettoyage et Standardisation :

- Renommage des colonnes pour plus de clarté (ex: Nouvelles hospitalisations -> Hospitalisations).
- Conversion des formats de date (issus d'Excel) en objets `datetime` exploitables par Python.
- Remplissage de valeurs manquantes simples, par exemple en calculant le total des séjours MCO comme la somme des hospitalisations complètes et de l'ambulatoire lorsque nécessaire.

2. Reconstitution Chronologique (Data Augmentation) :

- **Problème** : Les données locales de PSL-CFX n'existent que pour 3 années. Il est impossible d'entraîner un modèle de série temporelle sur des données aussi discontinues.
- **Solution** : Une méthode de **reconstitution** a été mise en place pour les années 2013, 2014, et 2017-2019. L'algorithme part d'une valeur réelle connue (ex: 2012), puis applique le taux de variation national (issu de `hospitalisation_capacity.xlsx`) pour estimer la valeur de l'année suivante. Ce processus est répété jusqu'à la prochaine année réelle connue, où le modèle se "recalcule" sur la valeur auditée.
- **Résultat** : Création d'une série temporelle annuelle **complète et cohérente** de l'activité et de la capacité de l'hôpital de 2012 à 2019.

3. Analyse de Saisonnalité :

- Une **moyenne mobile sur 15 jours** a été appliquée à la série des passages aux urgences pour lisser le bruit journalier et faire ressortir les tendances saisonnières de fond.

3. Choix et Explication des Modèles de Données

L'analyse combine des modèles descriptifs, prédictifs et de simulation pour répondre aux objectifs du projet.

1. Analyse Descriptive et Visualisation :

- **Outils** : pandas pour la manipulation des données, matplotlib et seaborn pour les visualisations.
- **Objectif** : Comprendre les dynamiques de base des données.
- **Exemples** :
 - **Courbes d'évolution** des hospitalisations COVID pour identifier les différentes vagues.
 - **Heatmap (carte de chaleur)** de l'impact régional pour montrer les disparités géographiques de la crise.
 - **Graphiques en barres** pour comparer l'offre et la demande.

2. Modélisation Prédictive : Le Scénario "Business as Usual" :

- **Modèle** : **SARIMA** (Seasonal AutoRegressive Integrated Moving Average).
- **Justification** : Ce modèle est un standard pour les séries temporelles. Il a été choisi (et recommandé dans le cahier des charges du projet) car il permet de capturer les tendances de fond (partie ARIMA) et potentiellement la saisonnalité (partie S).
- **Application** :
 - Le modèle SARIMA est entraîné sur la série **reconstituée** de l'activité MCO de 2012 à 2019.
 - Il est ensuite utilisé pour **projeter l'activité attendue pour 2020 et 2021**. Cette prédiction représente la **"baseline"**, c'est-à-dire le scénario de référence si la crise COVID-19 n'avait pas eu lieu.

3. Simulation de Crise : L'Effet Ciseaux :

- **Modèle** : Approche par **simulation de choc**.
- **Justification** : Plutôt que d'intégrer le choc comme une variable externe dans un modèle SARIMAX (plus complexe), une approche plus simple et interprétable a été choisie.
- **Application** :
 - Le **pourcentage moyen de variation** de l'activité des urgences pendant la première vague COVID dans la région de l'hôpital (-31.65%, reflétant une chute de l'activité non-COVID mais une surcharge des services dédiés) est extrait de `passage_urg_vague_covid.xlsx`.
 - Ce facteur de choc est appliqué à la **prévision "baseline" SARIMA** pour 2020 afin de simuler la demande réelle en situation de crise.
 - Le résultat est comparé à la capacité d'accueil projetée de l'hôpital, qui, elle, est sur une tendance baissière. La visualisation de cet écart met en évidence un **"effet ciseaux"**, où

la demande explose alors que l'offre se contracte, créant un **"gap capacitaire"**.

4. Analyse de Saturation Structurelle :

- **Modèle** : Calcul d'un **Taux d'Occupation (TO) théorique**.
- **Justification** : Permet de quantifier la tension sur l'hôpital avant même la crise.
- **Application** : En se basant sur une durée moyenne de séjour (DMS) estimée, le modèle calcule le nombre total de "journées-patient" demandées sur l'année et le compare au nombre de "journées-lit" disponibles. Le résultat a montré un TO théorique **supérieur à 85%** pour 2020, même dans le scénario hors-crise, signalant une **saturation structurelle** qui a rendu l'hôpital particulièrement vulnérable au choc pandémique.