

BookSure application

Une application qui vous aide à gérer vos réservations



Réalisé par :

Baya BENOUREDJA
Lucas COLARD
Rodolphe GABA
Clara RAKOTOARISOA
Lucile SAILLANT

Sous la direction de :

Morgan COUSIN
Amor KEZIOU
Arona DIENE

2024 - 2025

Découvrir



Table des matières

REMERCIEMENTS	4
INTRODUCTION	5
I. PRESENTATION DE L'APPLICATION	6
A. ENJEUX ET OBJECTIFS	6
B. PUBLIC CIBLE DE L'APPLICATION	7
II. LES DONNEES	8
A. SOURCE ET DESCRIPTIF DE LA BASE DE DONNEES	8
B. STRUCTURE DES DONNEES	8
C. LES VARIABLES	10
1. LA VARIABLE D'INTERET	10
2. LES VARIABLES EXPLICATIVES	10
III. NETTOYAGE ET PREPARATION DES DONNEES	12
A. TRAITEMENT DES DONNEES	12
1. VALEURS MANQUANTES	12
2. FILTRAGES DES INCOHERENCES, DOUBLONS ET TRANSFORMATION DES VARIABLES	12
3. GESTION DES VALEURS ABERRANTES	12
B. ANALYSES DES DONNEES AVANT ET APRES NETTOYAGE	13
C. ÉQUILIBRE DE LA BASE	14
IV. MODELE DE PREDICTION	16
A. SELECTION DE VARIABLES	16
B. MODELES DE PREDICTION	17
1. PRINCIPES GENERAUX	17
2. EVALUATION DES MODELES	18
3. CHOIX DU MODELE FINAL	19
V. CONCEPTION DE L'APPLICATION	21
A. ARCHITECTURE GENERALE	21
1. ACCUEIL	21
2. « À PROPOS »	22
3. FORMULAIRE	23
4. RECAPITULATIF	25
5. RESULTAT DE LA PREDICTION	25
B. ÉVOLUTION DE L'APPLICATION	27
1. UNE PREMIERE ESQUISSE DE L'APPLICATION	27
2. LA CONFECTION DU DESIGN DE L'APPLICATION	27

3.	LA GESTION DES OUBLIS ET DES ERREURS DE SAISIES ET REINITIALISATION DES CHOIX	28
4.	UN FORMULAIRE CONCIS ET MINIMALISTE	30
5.	UNE PAGE D'INFORMATION « À PROPOS »	30
6.	UNE PAGE RESULTAT COMPLETE	31
7.	DERNIERES AMELIORATIONS DE L'EXPERIENCE UTILISATEUR	32
VI.	ENVIRONNEMENT DE DEVELOPPEMENT	34
A.	LANGAGE DE PROGRAMMATION	34
B.	REPOT GIT	35
VII.	LIMITES ET OUVERTURES	36
A.	LIMITES DE L'APPLICATION	36
B.	OUVERTURES POSSIBLES	36
VIII.	COLLABORATION ET SPRINTS SUIVANT LA METHODE AGILE	38
A.	SPRINT 1	38
B.	SPRINT 2	39
C.	SPRINT 3	40
D.	SPRINT 4	41
E.	SPRINT 5	42
IX.	ANNEXES	43
A.	DICTIONNAIRE DES VARIABLES	43
B.	ROADMAP DU PROJET	45
C.	ANALYSE DES CORRELATIONS DES VARIABLES DEUX A DEUX	45
D.	PERFORMANCES DU MODELE DE REGRESSION LOGISTIQUE	46
E.	EVALUATION DU MODELE FINAL : RANDOMFOREST	46
F.	VERSIONS DE L'APPLICATIONS	47
1.	VERSION 0	47
2.	VERSION 1	49

Remerciements

Nous souhaitons exprimer notre profonde gratitude à Monsieur Morgan COUSIN, qui a assuré le suivi de notre projet tout au long des sprints. Son expertise en organisation et méthodologie de travail, ainsi que sa direction avisée, ont été déterminantes pour structurer et mener à bien chaque étape de cette application.

Nous remercions également chaleureusement Monsieur KEZIOU Amor, qui a partagé avec nous ses connaissances approfondies dans le domaine des modèles de prédiction. Son enseignement a enrichi notre compréhension et permis d'intégrer des approches analytiques solides dans notre projet.

Enfin, nous adressons nos sincères remerciements à Monsieur DIENE Arona, pour son accompagnement précieux dans l'apprentissage et l'application de VBA, un outil essentiel à la réussite technique de notre travail.

Grâce à leur encadrement et à leurs conseils, ce projet a été une expérience enrichissante et formatrice que nous sommes fiers de présenter.

Introduction

Le secteur hôtelier traverse une période de transformation marquée par des défis importants, notamment en raison de l'évolution des comportements des consommateurs. L'annulation des réservations hôtelières devient ainsi un enjeu majeur, impactant la gestion des revenus, l'optimisation des ressources et la satisfaction des clients. Les conséquences varient selon les types d'hôtels, les périodes de réservation et les conditions d'annulation.

Les annulations de dernière minute perturbent la planification financière des hôtels, notamment en haute saison, et conduisent souvent à des pertes de revenus. Comme la plupart des hôtels s'appuient sur des prévisions d'occupation pour ajuster leurs prix et allouer leurs ressources, chaque annulation affecte directement leur taux d'occupation réel. Cela les oblige à réagir rapidement pour limiter les pertes. Par ailleurs, la gestion du personnel est également impactée, car elle dépend des prévisions d'occupation. Une vague d'annulations peut entraîner un surdimensionnement des équipes, tandis qu'une occupation plus élevée que prévue peut conduire à une pénurie de personnel.

Le surbooking est une stratégie courante pour compenser les annulations, permettant aux hôtels de maximiser leur taux d'occupation. Toutefois, si les annulations sont faibles, cela peut conduire à une surréservation et à des situations où les clients doivent être relocalisés, ce qui nuit à leur satisfaction et à la réputation de l'hôtel.

Les plateformes de réservation en ligne comme Booking et Expedia ont transformé le comportement des consommateurs, rendant l'annulation ou le changement de réservation plus faciles. Les clients, souvent enclins à réserver plusieurs hôtels avant de se décider, contribuent ainsi à l'augmentation des annulations. Cette évolution complique la gestion de la demande pour les hôtels, renforçant la nécessité de politiques d'annulation plus adaptées.

I. Présentation de l'application

A. Enjeux et objectifs

Les taux d'annulation de réservation représentent une perte importante de revenus pour les hôtels. Selon une étude de D-EDGE¹, en 2023, les annulations de réservations effectuées directement auprès des hôtels ont représenté 18 % de leur chiffre d'affaires en Europe. De même, les annulations des réservations effectuées via Booking² ont représenté 42 % du chiffre d'affaires de la plateforme en Europe pour la même année (Figure 1). Face à ces taux élevés d'annulation, il devient crucial pour les hôteliers de pouvoir anticiper et gérer efficacement les annulations.

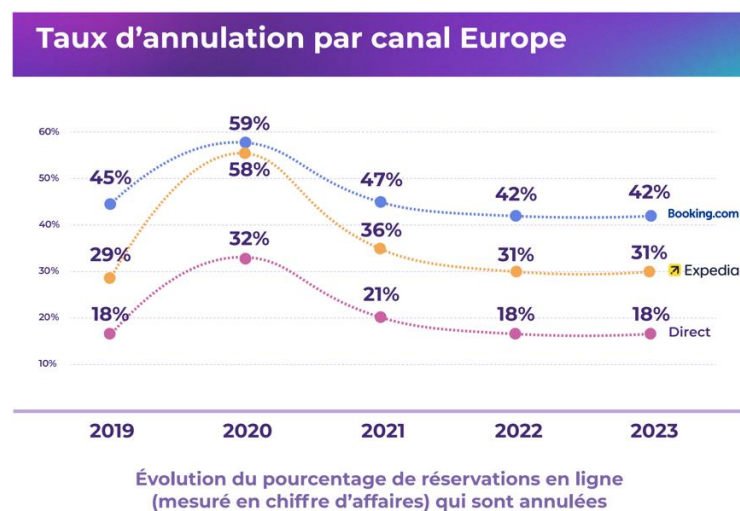


Figure 1 : Représentation des taux d'annulation dans les hôtels

Source : D-EDGE

Les annulations de dernière minute peuvent perturber la gestion des ressources humaines et opérationnelles, rendant difficile la planification des équipes et la gestion des tarifs. Cette incertitude complique également la gestion des revenus, car chaque chambre inoccupée représente une perte significative, les coûts fixes restant constants. En effet, les coûts tels que les salaires, les charges d'entretien et la gestion des équipements demeurent inchangés, même lorsque des chambres restent vacantes.

En outre, la flexibilité des options d'annulation pour les clients rend la prévision de la demande encore plus complexe. La difficulté pour les hôteliers de prévoir leur taux de remplissage réel entraîne des perturbations dans leur stratégie tarifaire et la gestion de l'offre. Il devient donc primordial de disposer d'outils permettant de mieux comprendre et anticiper ces fluctuations.

¹ D-EDGE : entreprise technologique française spécialisée dans les solutions de gestion hôtelière.

² Booking : plateformes de réservation d'hébergements. Une des plus populaires au monde.

B. Public cible de l'application

L'application BookSure a été développée pour répondre aux défis rencontrés par les hôteliers dans la gestion des annulations de réservations. Son objectif principal est de prédire le statut futur d'une réservation, c'est-à-dire de déterminer si celle-ci sera annulée ou non. Cette fonctionnalité permettra aux hôteliers de mieux anticiper les annulations et de réduire l'impact de celles-ci sur leurs revenus.

Pour effectuer cette prédiction, l'utilisateur (l'hôtelier) doit fournir certaines informations relatives à la réservation qu'il souhaite évaluer. Il peut remplir un formulaire pour entrer manuellement les détails de chaque réservation, tels que la date, la source de la réservation, le type de chambre, etc. Une fois ces informations fournies, notre système analyse les données et génère une prédiction concernant la probabilité d'annulation de la réservation.

Grâce à la prédiction des annulations, les hôtels pourront mieux anticiper leur taux d'occupation réel. Cela leur permettra de mieux allouer les ressources et de réduire les dépenses inutiles. En effet, les annulations affectent la planification des équipes et des ressources (ménage, personnel d'accueil, restauration), ce qui entraîne une baisse des besoins en main-d'œuvre et en ressources. L'anticipation des annulations permet ainsi d'optimiser les coûts opérationnels.

Cela permettra également d'appliquer des stratégies d'overbooking de manière plus sereine. En acceptant des réservations supplémentaires pour compenser les annulations probables, les hôteliers maximisent leurs opportunités de revenus tout en limitant les risques associés à cette pratique. L'application BookSure offre donc aux hôteliers un outil puissant pour prendre des décisions éclairées et optimiser la gestion de leurs réservations.

II. Les données

A. Source et descriptif de la base de données

Nous allons travailler avec des données sur la demande hôtelière. Ce jeu de données contient des informations relatives à un hôtel de campagne et à un hôtel de ville au Portugal ; des informations telles que le moment de la réservation, la durée du séjour, le nombre d'adultes, d'enfants et/ou de bébés et le nombre de places de parking disponibles, entre autres.

Les données proviennent d'un article de recherche : Nuno Antonio et al., "Hotel booking demand datasets", Data in Brief, Volume 22, 2019.

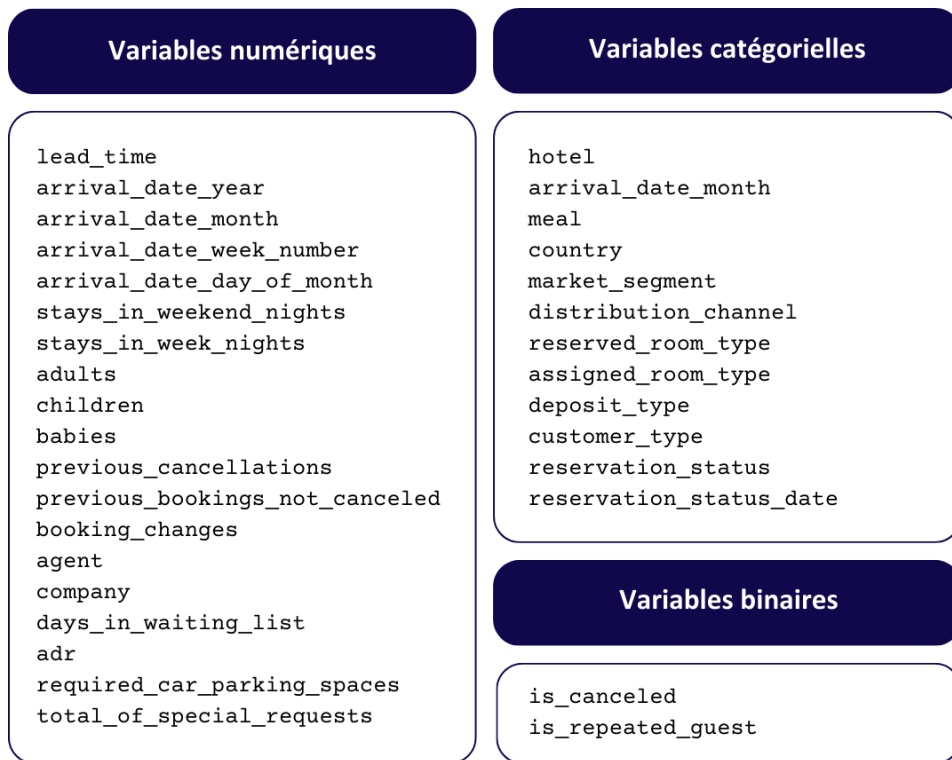
Composé de 31 variables, ce jeu de données décrit 119 390 observations. Chaque observation représente une réservation d'hôtel. Il s'agit de réservations qui devaient arriver entre le 1er juillet 2015 et le 31 août 2017, y compris les réservations qui sont effectivement arrivées et celles qui ont été annulées.

L'objectif principal de cette base de données est d'analyser les réservations d'hôtel afin de mieux comprendre les tendances et les facteurs influençant les annulations de réservations. Plus spécifiquement, cette analyse vise à :

- Identifier les facteurs associés aux annulations : Comprendre quelles variables sont les plus susceptibles de prédire si une réservation sera annulée ou non.
- Évaluer l'impact de différents segments de marché : Examiner comment les différents segments de clients (par exemple, les voyageurs d'affaires par rapport aux vacanciers) influencent les taux d'annulation.
- Optimiser la gestion des réservations : Fournir des recommandations aux hôteliers pour améliorer la gestion des réservations et réduire le taux d'annulation.
- Développer des modèles de prédiction : Créer des modèles statistiques ou d'apprentissage automatique pour prédire la probabilité d'annulation d'une réservation en fonction des caractéristiques fournies.

B. Structure des données

Les variables du jeu de données peuvent être réparties en catégories. Nous disposons de 19 variables numériques, 12 variables catégorielles et 2 variables binaires. Un premier coup d'œil dans la composition des variables révèle la présence de valeurs manquantes dans les données.



L'analyse des variables catégorielles permet de mettre en évidence le nombre de modalités uniques dans ces variables, le mode et la fréquence à laquelle il est présent. C'est ce que nous pouvons remarquer dans la figure suivante :

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type	customer_type
count	119390	119390	119390	118902	119390	119390	119390	119390	119390	119390
unique	2	12	5	177	8	5	10	12	3	4
top	City Hotel	August	BB	PRT	Online TA	TA/TO	A	A	No Deposit	Transient
freq	79330	13877	92310	48590	56477	97870	85994	74053	104641	89613

Figure 2 : Description de quelques variables catégorielles

Nous pouvons également remarquer que le nombre d'observations dans la variable `country` est en dessous du nombre d'observations du jeu de données. Cela indique la présence de valeurs manquantes dans cette variable. Nous nous occuperons du cas des valeurs manquantes par la suite.

Passons aux variables numériques. Pour avoir un premier point de vue sur ces variables, nous allons calculer quelques statistiques descriptives.

	previous_cancellations	booking_changes	total_of_special_requests	used_company	used_agent	is_repeated_guest
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000
mean	0.087118	0.221124	0.571363	0.056931	0.863138	0.031912
std	0.844336	0.652306	0.792798	0.231712	0.343703	0.175767
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
75%	0.000000	0.000000	1.000000	0.000000	1.000000	0.000000
max	26.000000	21.000000	5.000000	1.000000	1.000000	1.000000

Figure 3 : Statistiques descriptives de quelques variables numériques

De cette figure, nous constatons que pour au moins 75% des réservations, il n'y a quasi pas d'annulations passées, pas de modifications dans la réservation et pas de requêtes spéciales.

C. Les variables

1. La variable d'intérêt

La variable d'intérêt est celle qui référence le statut de la réservation (`is_canceled`). Le statut correspondant, ici, si la réservation est annulée ou non est de type binaire et est codé en 1 si la réservation est annulée et en 0 sinon.

2. Les variables explicatives

La base de données contient plusieurs variables clés qui peuvent être utilisées pour atteindre ces objectifs. Voici les principales variables explicatives :

- `lead_time` : Nombre de jours entre la date de réservation et la date d'arrivée. Cette variable peut influencer la probabilité d'annulation.
- `arrival_date_year`, `arrival_date_month`, `arrival_date_week_number`, `arrival_date_day_of_month` : Informations sur la date d'arrivée, qui peuvent aider à analyser les tendances saisonnières ou hebdomadaires des annulations.
- `stays_in_weekend_nights` et `stays_in_week_nights` : Durée du séjour en nuits, qui peut avoir un impact sur le taux d'annulation.
- `adults`, `children`, `babies` : Démographie des clients, utile pour segmenter les analyses selon le type de voyageur.

- meal : Type de repas inclus dans la réservation, qui peut influencer l'expérience client et la satisfaction.
- country : Pays d'origine de la réservation, permettant d'analyser les différences culturelles dans les comportements de réservation.
- market_segment et distribution_channel : Indiquent comment les clients ont réservé leur séjour, ce qui est utile pour comprendre l'efficacité des différents canaux de vente.
- is_repeated_guest : Indique si le client a déjà séjourné dans l'hôtel, ce qui pourrait influencer sa décision de réserver à nouveau ou d'annuler.
- adr (Average Daily Rate) : Taux de la chambre par nuit, qui peut être un indicateur de la valeur perçue par le client.
- total_of_special_requests : Nombre de demandes spéciales faites par le client, ce qui peut être lié à la satisfaction et à la probabilité d'annulation.

III. Nettoyage et préparation des données

Le nettoyage des données est une étape importante dans le processus d'analyse des données, il permet d'assurer la qualité et la fiabilité des informations.

A. Traitement des données

1. Valeurs manquantes

Les données ont été importées et une première analyse a permis de détecter des valeurs manquantes dans plusieurs variables. En raison de la proportion importante de valeurs manquantes pour les variables `company` et `agent`, qui affichent respectivement 94,31% et 13,69 % de données manquantes, nous avons choisi de les convertir en variables binaires. En effet, pour ces variables, les données manquantes ne doivent pas être considérées comme telle mais comme une indication de non-applicabilité. Cela signifie que la réservation n'a pas été effectuée par une agence de voyage ou une compagnie, entraînant l'absence de valeur. Ces variables-ci ont été transformé en variables binaires, en remplaçant les valeurs manquantes par 0 et les autres valeurs par 1, et ont été renommé par `used_company` et `used_agent`.

2. Filtrages des incohérences, doublons et transformation des variables

Les lignes présentant des valeurs incohérentes, telles que les réservations sans client, ont été supprimées, tout comme les doublons afin d'assurer l'unicité des observation et d'éviter les biais.

Certaines variables ont été converties au type approprié afin de faciliter l'analyse, comme la variable du nombre d'enfant, convertie en entier. De plus, la variable catégorielle `meal` a été uniformisé en remplaçant la valeur `Undefined` par `SC`.

3. Gestion des valeurs aberrantes

Nous avons réalisé une analyse des valeurs aberrantes en s'appuyant sur des statistiques descriptives des variables numériques ainsi que des représentations graphiques. Des plafonds ont été fixés pour limiter l'impact des valeurs extrêmes de certaines variables. Dans certains cas, des valeurs excessives, telle que la valeur maximale observée pour la variable `adr`, ont été supprimées, car elles semblaient résulter d'une erreur de saisie.

Concernant la variable `required_car_parking_spaces`, un seuil a également été appliqué. Cette variable a été plafonnée à 5, car nous avons estimé qu'il est peu probable qu'une réservation nécessite plus de 5 places de parking.

A l'issue de ce processus de nettoyage de données, nous avons réduit la base de données de 119 390 observations à 86 761 observations.

B. Analyses des données avant et après nettoyage

Afin d'identifier l'impact du processus de nettoyage, des analyses univariées et bivariées ont été réalisées sur les données brutes ainsi que sur les données nettoyées. Ces analyses nous permettent de comparer les distributions des variables majeurs avant et après nettoyage.

Avant nettoyage, 37 % des réservations étaient annulées et 63 % non annulées. En revanche, après le nettoyage, nous constatons que la proportion de réservations annulées a diminué à 27,6 % et celles non annulées ont augmenté à 72,4 % (cf. Figure 4).

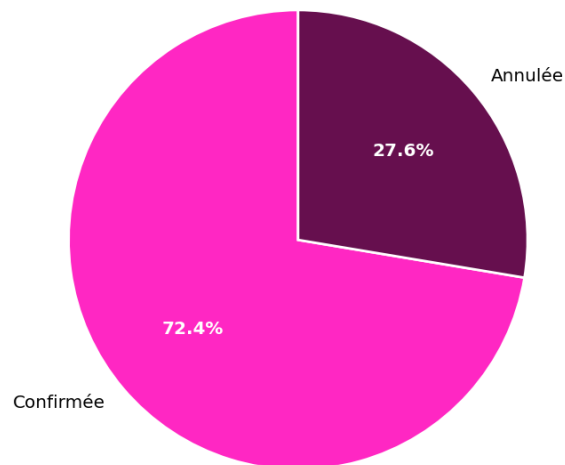


Figure 4 : Statut de la réservation après nettoyage

Ensuite, nous nous sommes intéressé la répartition des types d'hôtels. Les données avant nettoyage montraient qu'une majorité des réservations concernaient les complexes hôteliers

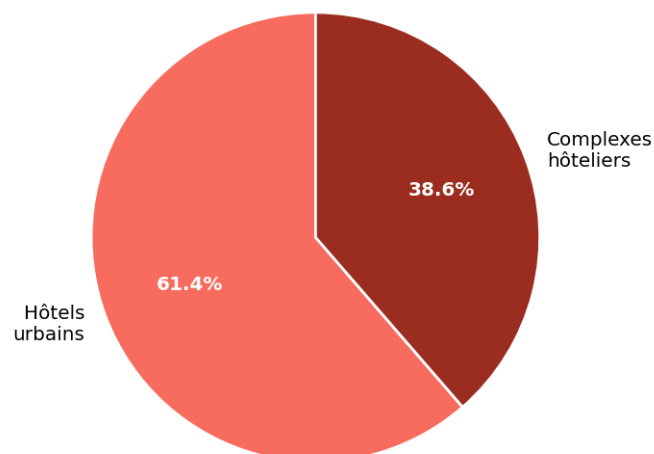


Figure 5 : Répartition des types d'hôtels après nettoyage

(66,4 %), tandis qu'après le nettoyage, cette proportion a légèrement diminué à 61,4 %. En revanche la répartition des hôtels en milieu urbain a augmenté à 38,6 % (cf. Figure 5).

Après avoir examiné les données à travers des analyses univariées, nous poursuivons avec les analyses bivariées, en comparant les données initiales et nettoyées. Cette étape nous permet d'explorer les relations entre deux variables.

En analysant le statut d'annulation des réservations en fonction du type d'hôtel, avant le nettoyage des données, les hôtels urbains présentaient un taux d'annulation de 41,73 % contre 27,76 % pour les complexes hôteliers. À l'inverse, 58,27 % des hôtels situés en ville n'ont pas été annulés, comparé à 72,24 % pour les complexes hôteliers.

Après le nettoyage de données (cf. Figure 6), une baisse d'environ 10 % est observée au niveau des taux d'annulation pour chaque type d'hôtel. Les hôtels situés en ville enregistrent désormais un taux d'annulation de 30,09 % contre 23,71 % pour les complexes hôteliers.

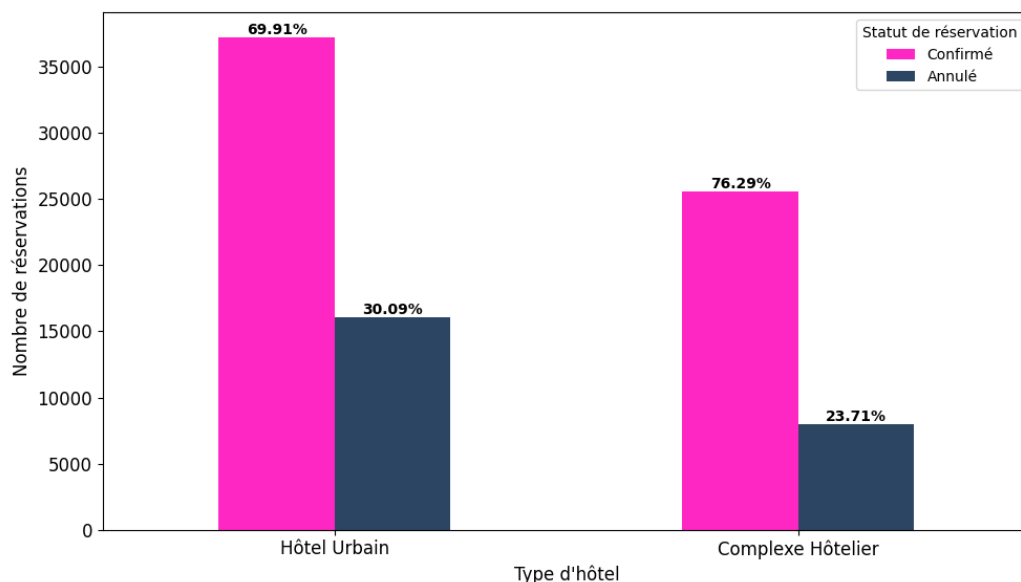


Figure 6 : Statut d'annulation selon le type d'hôtel après nettoyage

C. Équilibre de la base

L'analyse exploratoire a mis en évidence un déséquilibre accentué principalement au niveau de la variable cible statut de la réservation. Afin de traiter ce déséquilibre, une pondération des classes a été envisagée, complétée par un rééchantillonnage des données pour équilibrer les classes et rendre la base plus représentative. Ce déséquilibre peut affecter la performance des modèles, qui tendent à favoriser les classes majoritaires.

Parmi l'ensemble des techniques existantes pour résoudre ce problème, nous avons les techniques liées aux données. Le **sur-échantillonnage de la classe minoritaire** peut être une

option. Il s'agit de créer des copies supplémentaires des données de la classe minoritaire pour augmenter son poids. Pour ce faire, l'on peut avoir recours à des techniques comme SMOTE (*Synthetic Minority Oversampling Technique*) qui génèrent des exemples synthétiques de la classe minoritaire. Le **sous-échantillonnage de la classe majoritaire** est aussi un moyen de pallier ce problème. En effet, cette méthode réduit la taille de la classe majoritaire en échantillonnant un sous-ensemble de ses données. Il est également possible de ces deux techniques pour faire un **rééchantillonnage combiné**.

Dans le cadre de la réalisation de notre application, nous avons opté pour un sous-échantillonnage de la classe majoritaire, à savoir les réservations non annulées.

IV. Modèle de prédiction

A. Sélection de variables

Parmi toutes les variables dont nous disposons, nous ensuite avons sélectionné les variables utiles par la méthode de recherche qui se base sur un algorithme génétique.

L'objectif de cette sélection vise à identifier un sous-ensemble optimal de variables explicatives pour améliorer les performances du modèle, réduire la complexité du modèle ou encore éviter le problème de surajustement.

Les algorithmes génétiques permettent d'explorer efficacement un espace de recherche complexe, notamment dans le cas où le nombre de variables est élevé et que toutes les combinaisons possibles seraient coûteuses à évaluer. Dans ce principe, une population de chromosomes est générée aléatoirement. Chaque chromosome représente une combinaison de variables. Une fonction est également définie pour évaluer la qualité de chaque solution. Dans le cas de la régression logistique, la fonction d'évaluation utilisée peut être la log-vraisemblance du modèle, le critère AIC ou BIC (des critères d'information), la capacité prédictive (la justesse de prédiction, la précision du modèle, l'aire sous la courbe ROC). Après un processus itératif, le sous-ensemble de variables avec la meilleure qualité est sélectionné. Après la mise en œuvre de cette méthode de sélection de variables, nous avons retenus les variables suivantes :

```
lead_time
stays_in_week_nights
adults
children
previous_cancellations
previous_bookings_not_canceled
booking_changes
adr
used_company
customer_type
deposit_type
reserved_is_assigned
required_car_parking_spaces
total_of_special_requests
reserved_is_assigned
```


B. Modèles de prédiction

Pour la suite de notre analyse, nous allons comparer cinq (05) modèles de prédictions, à savoir : la régression logistique, le modèle d'analyse discriminante linéaire, le modèle des K plus proches voisins, le random Forest et le XGBoost. Avant la mise en place des modèles, l'analyse des variables retenues révèle que la variable relative au nombre de places de parking exigé par un client constitue un biais particulièrement élevé dont nous voudrions nous passer. En effet, dans notre jeu de données, toutes les réservations qui ont été annulées n'avaient exigés aucune place de parking. C'est pour cela que la variable « **required_car_parking_spaces** » ne fait plus partie du processus de modélisation.

1. Principes généraux

La **régression logistique** est un modèle statistique qui estime la probabilité qu'un échantillon appartienne à une classe (0 ou 1). Elle repose sur une fonction logistique (ou sigmoïde) pour transformer une combinaison linéaire des variables explicatives (X) en une probabilité entre 0 et 1. Si $P(y = 1 | X) > 0.5$, on prédit $y = 1$, sinon $y = 0$.

L'**analyse discriminante linéaire** suppose que les données de chaque classe suivent une distribution gaussienne (normale) avec une variance-covariance identique entre les classes. Le modèle maximise la séparation entre les deux classes en trouvant une frontière linéaire. Ce modèle est performant si les hypothèses de normalité et de covariance sont respectées.

Le modèle des **K plus proches voisins** est un modèle basé sur les observations. Pour prédire la classe d'un nouvel échantillon, il identifie les K observations les plus proches dans l'espace des caractéristiques (selon une métrique, souvent la distance euclidienne). C'est un modèle simple qui fait un calcul de distance. Il ne formule aucune hypothèse sur la distribution des données.

Random Forest est un modèle d'ensemble basé sur des arbres de décision. Il combine plusieurs arbres de décision entraînés sur des sous-échantillons aléatoires des données et des variables pour produire une prédiction robuste. Ce modèle gère les interactions complexes entre les variables.

XGBoost est un algorithme de boosting basé sur des arbres de décision. Il construit les arbres de manière séquentielle, chaque nouvel arbre corrigeant les erreurs des arbres précédents en optimisant une fonction de perte. Ce modèle est généralement très performant et limite le sur-apprentissage grâce à la régularisation qui y est intégrée.

2. Evaluation des modèles

Par validation croisée, nous avons déterminé le modèle « optimal » en termes de paramètres du modèle et de justesse de prédiction pour chacun des modèles précédemment cités. Le figure 7 présente le résultat obtenu.

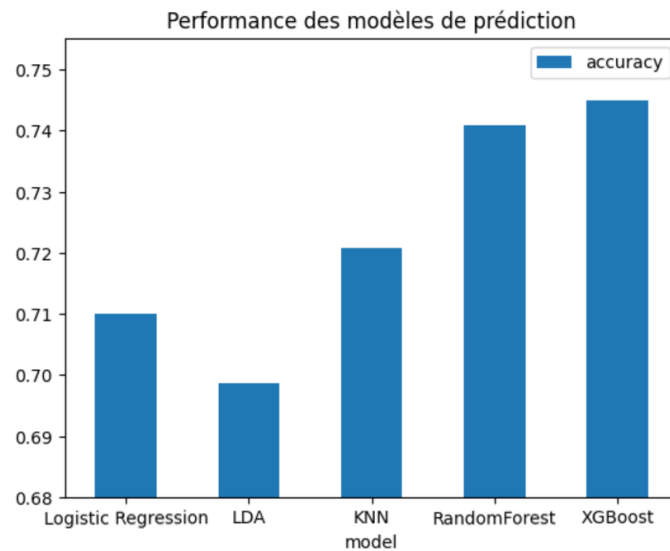


Figure 7 : Comparaison de la justesse des modèles de prédiction

Ce graphique met en évidence la performance de nos modèles de prédiction. L'ordonnée de ce graphique représente la fiabilité de prédiction. Nous pouvons remarquer que deux modèles font la course en tête. Il s'agit notamment du modèle Random Forest et du modèle XGBoost. Dans l'optique de mieux départager les modèles, des capacités prédictives autres que l'accuracy (la justesse ou la fiabilité) doivent être prises en compte.

C'est ce que nous avons essayé de faire avec la figure qui suit. Cette courbe présente la courbe ROC de chacun des modèles. Le score AUC de la courbe ROC traduit la capacité d'un modèle à faire mieux ou pas par rapport à une prédiction aléatoire.

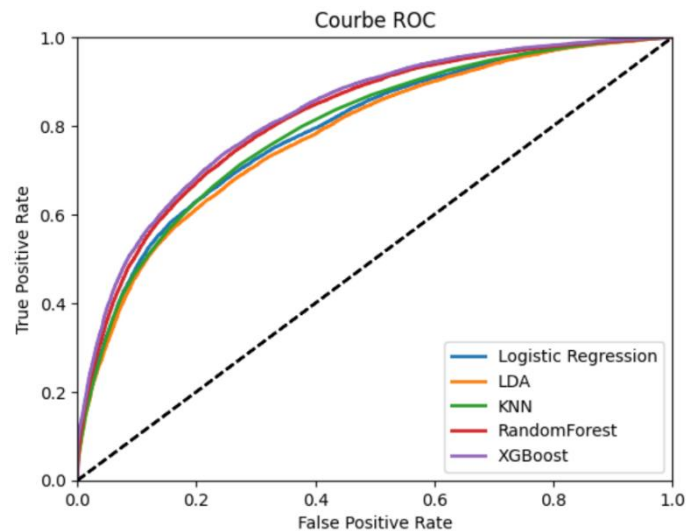


Figure 8 : Courbes ROC des modèles de prédiction

La diagonale sur cette figure représente un modèle qui prédit aléatoirement. Alors, un modèle fait mieux que l'aléatoire si sa courbe est au-dessus et relativement éloignée de cette diagonale. Plus la courbe ROC est éloignée de la diagonale, plus le modèle est bon. Nous constatons qu'en termes de répartition du taux de vrai positif par rapport au taux de faux positif, les modèles Random Forest et XGBoost sont relativement comparables. Le critère de la courbe ROC ne nous permet pas d'avoir un choix final.

3. Choix du modèle final

Le modèle retenu est le modèle Random Forest. Ce modèle présente plusieurs avantages majeurs. Tout d'abord, il permet une interprétabilité des résultats grâce à la mesure de l'importance de chaque variable dans le processus de prédiction. Cette caractéristique est essentielle pour comprendre quelles variables influencent le plus les annulations de réservations et pour adapter les stratégies de prévention en conséquence.

De plus, Random Forest est un modèle robuste, capable de gérer les interactions complexes entre les variables sans nécessiter d'hypothèses strictes sur la distribution des données, contrairement à des modèles comme l'analyse discriminante linéaire. Il a également montré une grande capacité à éviter le sur-apprentissage, ce qui garantit la généralisation du modèle aux nouvelles données.

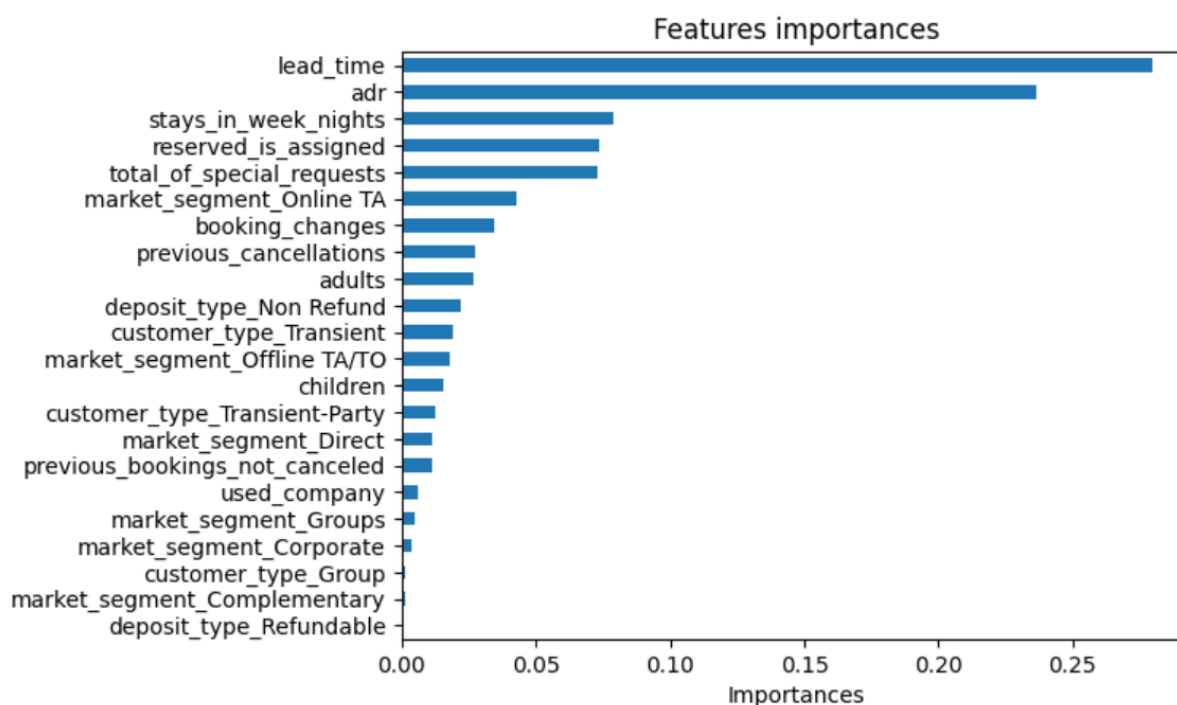


Figure 9 : Niveau d'importance des variables dans le modèle RandomForest

La figure 9 présente l'importance des variables dans le processus d'entraînement du modèle de prédiction. Ces scores reflètent l'impact relatif de chaque variable dans la construction et la performance globale du modèle. Chaque fois qu'un arbre de décision utilise une variable pour effectuer un split (une division), il mesure combien cette division améliore l'homogénéité des sous-ensembles de données (par exemple, en réduisant l'entropie ou le critère de Gini dans notre problème de classification). Cette mesure est attribuée à la variable utilisée pour le split. La contribution de cette variable est ensuite agrégée sur tous les arbres de la forêt, puis normalisée.

Si une variable a une importance élevée, cela signifie que le modèle l'a fréquemment utilisée pour effectuer des divisions significatives dans les arbres, et qu'elle contribue fortement à la précision du modèle. Dans notre cas, nous avons l'exemple du « *Délai entre la réservation et le séjour* » et du « *Prix moyen journalier* », avec les scores d'importances les plus élevés.

Si une variable a une importance faible ou nulle, cela signifie qu'elle n'a presque pas été utilisée dans les arbres ou qu'elle n'apporte pas d'information utile pour la prédiction. Il s'agit ici par exemple des variables « *Segment de marché complémentaire* » et « *Type de dépôt remboursable* », avec les scores d'importances les plus faibles.

V. Conception de l'application

Nous avons développé notre application avec VBA³ sur Excel. Ce choix s'explique par la facilité d'accès et d'utilisation d'Excel, qui est largement répandu dans les entreprises et permet une prise en main rapide pour les utilisateurs non techniques. VBA a permis de concevoir une interface utilisateur intuitive à l'aide de formulaires personnalisés, ainsi que d'automatiser le traitement des données pour produire des prédictions. Nous avons utilisé la fonctionnalité UserForm pour réaliser les formulaires. Dans cette partie, nous allons donc décrire la conception de l'application ainsi que son évolution au fil du projet.

A. Architecture générale

1. Accueil

Notre application s'ouvre en lançant le fichier *Booksue_app.xlsm*. On arrive alors sur notre page d'accueil (cf. Figure 10) qui permet d'orienter l'utilisateur vers les fonctionnalités de notre application. Elle présente l'objectif de l'application ainsi qu'une petite section « Le saviez-vous ? » destinée à donner une information contextuelle et qui permet de montrer l'enjeu et l'intérêt de notre application. On retrouve également notre logo avec un message de bienvenue pour donner une image accueillante et chaleureuse. On a ensuite deux boutons qui vont permettre de naviguer dans l'application. Le premier bouton « Commencer » permet de lancer le formulaire, le remplir et ainsi obtenir la prédiction. Le bouton « À propos » amène sur une nouvelle page qui va donner un peu plus d'informations sur les contours de l'application.

³ Visual Basic for Applications : langage de programmation intégré aux applications Microsoft Office permettant, entre autres, de créer des fonctionnalités personnalisées et de développer des interfaces interactives directement dans les applications.



Figure 10 : Page d'accueil de l'application (version finale)

2. « À propos »

La page « À propos » (cf. Figures 11 et 12) a différents objectifs. Tout d'abord, elle communique des informations sur l'application, son contexte de création, ses créateurs ainsi que leur contact s'il y a besoin de signaler d'éventuels bugs ou tout simplement de poser des questions. Cela permet d'humaniser l'interface, renforçant ainsi la confiance des utilisateurs. Ensuite, nous présentons brièvement les données sur lesquelles nous avons travaillé, les variables utilisées pour prédire le statut futur de la réservation ainsi que la méthode de prédiction utilisée dans l'application.



Figure 11 : Page 'À propos' de l'application (page 1 - version finale)

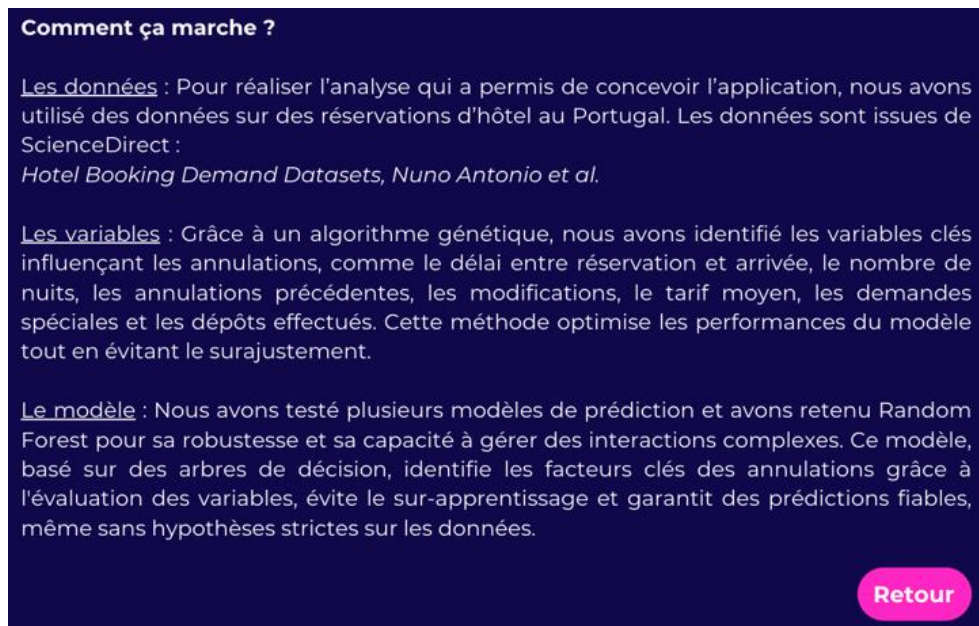


Figure 12 : Page 'À propos' de l'application (page 2 - version finale)

3. Formulaire

Le formulaire de collecte des données permet à l'utilisateur de saisir des informations relatives à la réservation d'hôtel dont il veut connaître le statut futur. Ce sont ces données qui vont être utilisées pour alimenter le modèle de prédiction. Ce formulaire est conçu pour être simple et direct, avec des champs clairs. Nous l'avons divisé en plusieurs sections pour plus de clarté et pour rendre le questionnaire plus digeste pour l'utilisateur (cf. Figures 13 et 14). Ainsi nous avons créé quatre sections distinctes :

- Informations générales sur la réservation
- Informations sur le séjour
- Informations sur le client
- Historique et préférences du client

Ces quatre sections sont réparties sur deux pages pour éviter de surcharger une seule page et améliorer l'expérience utilisateur.

Pour la saisie de l'utilisateur, nous avons fait les choix suivants : les questions pour lesquelles la réponse est oui ou non, nous avons opté pour des boutons d'options, ainsi pour l'utilisateur, il est plus simple de répondre et c'est plus efficace. Les questions pour lesquelles les modalités de réponses sont plus complexes et plus nombreuses, nous avons opté pour des listes déroulantes. Ainsi, pour ces deux types de questions, l'utilisateur ne peut pas faire d'erreur de saisie et ne peut pas sélectionner plusieurs options. Pour finir, les questions pour lesquelles l'utilisateur doit renseigner un entier ou, plus généralement, une valeur numérique, nous avons opté pour des zones de saisie de texte. Dans ce dernier cas, il est nécessaire de gérer les erreurs de saisie pouvant être faites par l'utilisateur, nous y reviendrons dans la suite.

Sur les deux pages de formulaire nous avons ajouté un bouton « Quitter », pour permettre à l'utilisateur de quitter le formulaire et de revenir à la page d'accueil. Cela lui permet d'éviter d'avoir à chercher comment sortir du formulaire et lui donner un moyen intuitif et claire de le faire grâce à un bouton qui attire l'œil. On retrouve également un bouton « Réinitialiser les choix » qui permet, comme son nom l'indique de réinitialiser les choix fait par l'utilisateur en un seul clique plutôt que de supprimer un par un chaque choix. Cela rend l'application plus simple d'utilisation et plus efficace.

Sur la première page du formulaire nous avons un bouton « Suite du questionnaire » permettant de passer à la page suivante et continuer de répondre au formulaire. Sur la seconde page nous avons un bouton « Page précédente » permettant de revenir à la première page et ainsi rectifier des données si jamais une erreur a été faites. Sur la dernière page, nous avons le bouton suivant permettant d'arriver à la page récapitulative.

The screenshot shows the first page of a questionnaire titled "Questionnaire". It is divided into two main sections: "Informations générales sur la réservation :" and "Informations sur le séjour :".

Informations générales sur la réservation :

- Type d'hôtel : [dropdown menu]
- Prix moyen de la nuitée : [text input]
- Montant : [text input]
- Type de dépôt : [dropdown menu]
- Type de marché : [dropdown menu]
- Est-ce que le client est passé par une entreprise ? [radio buttons: Oui, Non]

Informations sur le séjour :

- Nombre de nuits en semaine : [text input]
- Nombre de modifications de la réservation : [text input]
- La chambre réservée par le client est-elle la même que celle qui lui a été assignée ? [radio buttons: Oui, Non]

At the bottom, there are three buttons: "Quitter" (orange), "Réinitialiser les choix" (orange), and "Suite du questionnaire" (orange).

Figure 13 : Première page du formulaire de l'application (version finale)

The screenshot shows the second page of the questionnaire titled "Questionnaire". It is divided into two main sections: "Informations sur le client :" and "Historique et préférences du client :".

Informations sur le client :

- Nombre d'adultes : [text input]
- Nombre d'enfants : [text input]
- Type de client : [dropdown menu] [blue button with "?"]

Historique et préférences du client :

- Nombre de réservations non annulées : [text input]
- Nombre de réservations annulées : [text input]
- Nombre de demandes spéciales : [text input]
- Combien de jours avant son arrivée le client a-t-il réservé son séjour ? [text input]

At the bottom, there are four buttons: "Quitter" (orange), "Page précédente" (orange), "Réinitialiser les choix" (orange), and "Suivant" (orange).

Figure 14 : Deuxième page du formulaire de l'application (version finale)

4. Récapitulatif

Une fois les données saisies et le bouton « Suivant » sélectionné, l'utilisateur arrive sur la page récapitulative (cf. Figure 15) qui lui énumère toutes les données qu'il a pu saisir, permettant ainsi à l'utilisateur de vérifier s'il ne s'est pas trompé et de lancer la prédiction en toute confiance.

Sur cette page, les quatre sections de notre formulaire sont bien distinctes. Nous avons également le bouton « Corriger » permettant de revenir au début du formulaire pour que l'utilisateur puisse corriger les éventuelles erreurs qu'il a saisies. Le bouton « Quitter » qui va tout annuler et retourner à la page d'accueil. Et le bouton le plus important, « Lancer la prédiction », qui va permettre de communiquer les données saisies par l'utilisateur au modèle de prédiction et ainsi afficher une dernière page qui va donner le résultat de ce qui a été prédit.



Informations générales sur la réservation :	
Type d'hôtel :	Prix moyen de la nuitée :
Complexe hôtelier	138,89
Type de dépôt :	Type de marché :
Aucun dépôt	Agence de voyage hors ligne
Est-ce que le client est passé par une entreprise ?	
Oui	

Informations sur le séjour :	
Nombre de nuits en semaine :	Nombre de modifications de la réservation :
6	1
La chambre réservée par le client est-elle la même que celle qui lui a été assignée ?	
Non	

Informations sur le client :	
Nombre d'adultes :	Nombre d'enfants :
4	0
Type de client :	
Individuel	

Historique et préférences du client :	
Nombre de réservations non annulées :	Nombre de réservations annulées :
4	1
Nombre de demandes spéciales :	
0	
Combien de jours avant son arrivée le client a-t-il réservé son séjour ?	
65	

Buttons: Quitter, Corriger, Lancer la prédiction

Figure 15 : Page récapitulative de l'application (version finale)

5. Résultat de la prédiction

Une fois les données soumises, la page de résultats s'affiche et présente la prédiction générée par le modèle. Sur cette page, on distingue quatre sections. La première section correspond au résultat prédit par le modèle. Dans celle-ci, on donne le statut prédit de la réservation ainsi qu'un pourcentage qui estime le risque d'annulation donné par notre modèle. Ces deux éléments s'affichent en vert si le modèle prédit un risque d'annulation inférieur à 50 % (cf. Figure 16) et en rouge dans le cas contraire (cf. Figure 17). La seconde section est une section « Recommandation » qui va, en fonction du résultat de la prédiction, donner des conseils à l'hôtelier. Si la réservation est prédite comme confirmée, on lui propose de

recommencer avec ces autres réservations. Si la réservation est prédite comme annulée, on lui donne des conseils pour lui éviter de perdre son client ou d'anticiper en surbookant son hôtel. Ces deux sections sont donc personnalisées en fonction du résultat de la prédiction. En revanche, les deux qui suivent ne le sont pas, mais donnent des éléments sur le modèle. On a donc la section « Confiance », dans laquelle on donne la fiabilité du modèle en pourcentage. Pour finir, nous avons ajouté une section qui met en avant les principaux facteurs qui vont influencer le résultat de la prédiction, ce qui va permettre à l'utilisateur de mieux comprendre pourquoi une réservation serait annulée ou non.

Sur cette page, nous retrouvons deux boutons, l'un permettant de recommencer une nouvelle prédiction et qui renvoie à la première page du questionnaire. Le second qui permet de quitter la page résultat est renvoie à la page d'accueil.

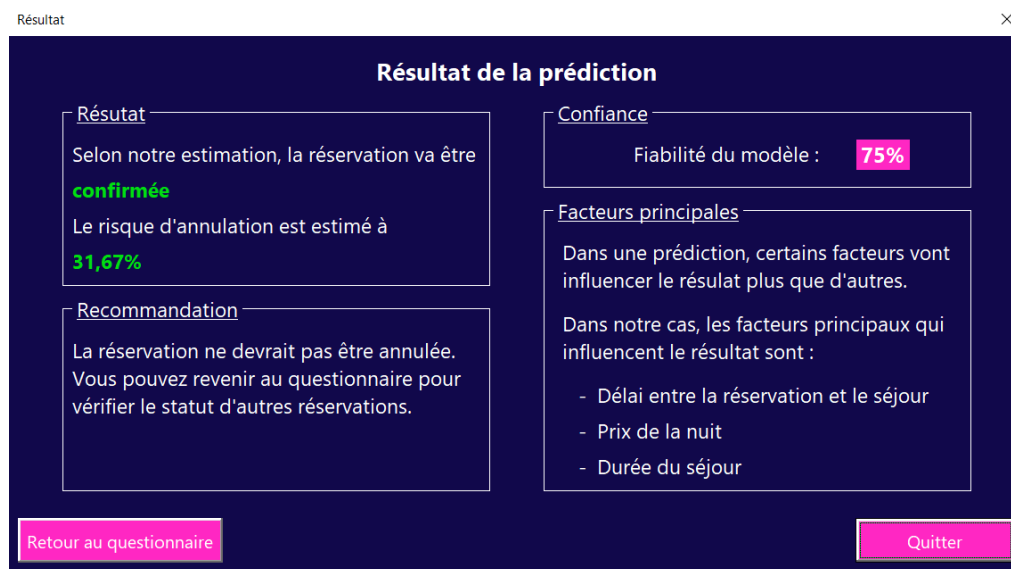


Figure 16 : Page résultat de l'application (réservation confirmée – version finale)

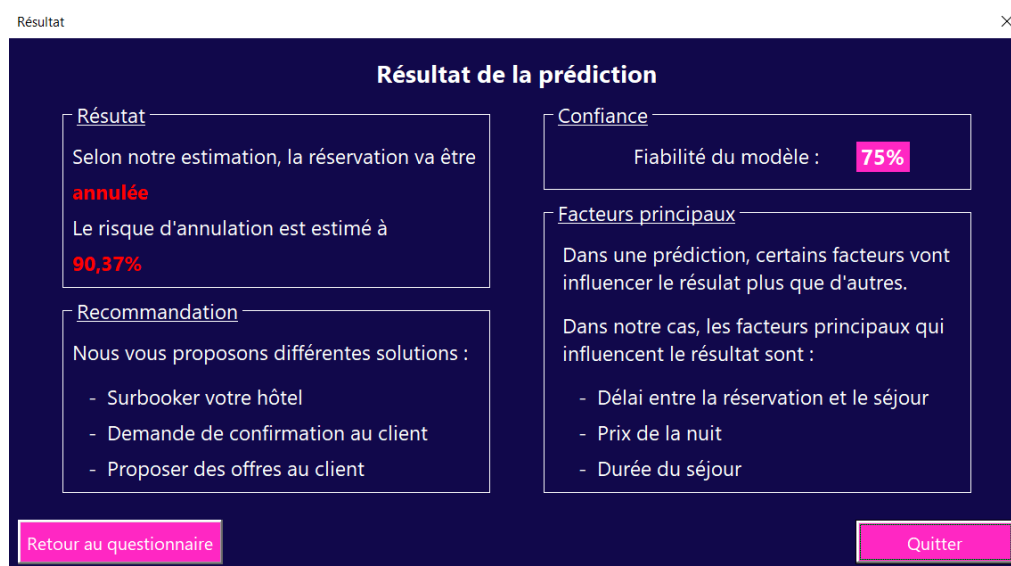


Figure 17 : Page résultat de l'application (réservation annulée – version finale)

B. Évolution de l'application

Le développement de l'application s'est appuyé sur une méthodologie de travail par sprints, inspirée des principes agiles. Pour chaque sprint, des objectifs étaient fixés afin d'avancer pas à pas dans la conception de l'application.

1. Une première esquisse de l'application

La première étape et l'objectif du premier sprint ont été de confectionner une première version simplifiée de l'application (cf. [Version 0](#)). Pour ce premier objectif, il a fallu se familiariser avec le langage VBA et la fonctionnalité des UserForms. Cette première esquisse se focalisait sur l'essentiel : une interface utilisateur basique intégrée à Excel, permettant de saisir des données de réservation et de générer une sortie indicative.

Pour la page d'accueil, on avait décidé de faire un UserForm qui s'ouvre dès l'ouverture du fichier Booksure_app.xlsm dans lequel on présentait brièvement l'objectif de l'application. Nous avons mis deux premiers boutons, un bouton « Commencer » qui permettait de passer au formulaire et à la saisie des données et un bouton « Quitter » pour faciliter l'utilisateur souhaitant quitter l'application.

Pour le formulaire, on a décidé de diviser les questions en quatre sections et sur deux pages pour le rendre moins grossier. On a commencé par mettre des questions pour la quasi-totalité des variables que nous avons à disposition dans la base de données pour la prédiction, puis par la suite filtrer uniquement celles qu'on aura sélectionnées. Sur la première page, nous avons deux boutons, le premier pour retourner à la page d'accueil et le second pour passer à la page suivante du questionnaire. Sur la seconde page, nous avons également deux boutons, le premier pour retourner à la page précédente du questionnaire et le second pour lancer la prédiction.

Pour la page de résultat de la prédiction, on avait uniquement configuré différents boutons. Le premier était pour retourner directement à la page d'accueil, un second pour retourner au début du questionnaire et ainsi recommencer une nouvelle prédiction avec de nouvelles données et le dernier permettait de quitter l'application. Ce dernier bouton fermait définitivement l'application et on ne pouvait pas la réouvrir étant donné que nous faisons afficher la page d'accueil sur un UserForm uniquement à l'ouverture du fichier. Dans les prochaines versions, nous allons rectifier cela en ajoutant un bouton sur une feuille Excel.

2. La confection du design de l'application

Après la réalisation de la première esquisse, l'étape suivante a consisté à concevoir le design de l'application (cf. [Version 1](#)). L'objectif était de créer une interface utilisateur intuitive,

claire et agréable. Nous avons d'abord réfléchi à l'ergonomie et à la structuration de l'interface. Et nous avons donc fait le choix de ne plus avoir la page d'accueil qui s'ouvre sous forme de UserForm dès que nous ouvrons le fichier, mais plutôt d'avoir la page d'accueil affichée sur une feuille Excel. Ainsi, à la fin de la prédiction, lorsque nous cliquons sur le bouton permettant de quitter, on se retrouve directement sur une feuille sur laquelle se trouve la page d'accueil. Le problème de fermeture définitive de l'application a donc été rectifié. Ce choix a été fait de telle sorte que par la suite, d'autres feuilles Excel nous permettent d'intégrer d'autres informations.

Concernant l'aspect visuel, un choix de couleurs a été effectué pour garantir à la fois lisibilité et esthétique. Nous avons choisi le bleu marine comme fond de l'application, car c'est une couleur souvent utilisée pour son côté sérieux et conventionnel, cela transmet ainsi un sentiment de confiance et de professionnalisme. De plus, cette couleur sombre permet de réduire la fatigue visuelle et de mettre en valeur les autres éléments de l'interface grâce à un contraste. En complément, le rose fuchsia a été choisi pour les boutons et les éléments sur lesquels on veut attirer l'œil de l'utilisateur en raison de son dynamisme et de sa visibilité. Cette couleur vive attire donc naturellement l'attention, facilitant ainsi l'identification des zones interactives de l'application et les informations importantes. L'association du bleu marine et du rose fuchsia crée une combinaison moderne et audacieuse, reflétant à la fois la fiabilité de l'outil et son approche innovante.

Ce design a été développé de manière itérative, avec plusieurs ajustements réalisés au fil des tests et des retours. Le but était de trouver un équilibre entre esthétique et fonctionnalité, en s'assurant que l'interface soit à la fois attrayante et facile à utiliser.

3. La gestion des oublis et des erreurs de saisies et réinitialisation des choix

Après avoir réalisé le design, la priorité dans le développement de l'application a été d'assurer une gestion efficace des oublis et des erreurs de saisie. Cela permet non seulement de garantir la fiabilité des données collectées, mais aussi d'améliorer l'expérience utilisateur en réduisant les frustrations liées à des erreurs de manipulation.

Nous avons commencé par intégrer des contrôles pour les champs obligatoires. Lorsque l'utilisateur tente de soumettre un formulaire sans avoir rempli toutes les informations requises, un message de rappel apparaît pour l'informer des champs manquants. Ce message précise clairement quel champ il a oublié de compléter. Sur la Figure 18, on montre un exemple de message d'avertissement dans lequel on lui demande de remplir un des champs qu'il a oubliés. Ce message s'affiche jusqu'à ce que tous les champs soient bien renseignés.

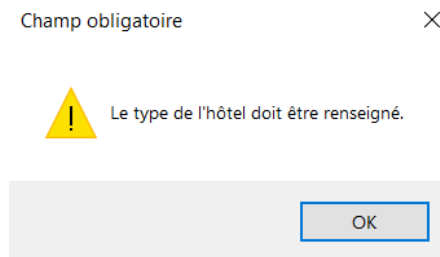


Figure 18 : Message d'avertissement de champ obligatoire oublié

La deuxième étape a été la gestion des erreurs de format saisies par l'utilisateur. Afin de prévenir les erreurs de saisie, nous avons mis en place plusieurs contrôles et validations automatiques. Pour ce qui est des variables pour lesquelles les modalités sont restreintes, nous avons mis des listes déroulantes avec comme option « fmStyleDropDownList ». Avec cette option proposée dans VBA, l'utilisateur ne pourra pas saisir de texte dans la liste déroulante et devra se limiter à choisir parmi les éléments prédéfinis. Pour les autres données à saisir, ce sont soit des données numériques positives pouvant contenir des décimales, soit des entiers positifs également. Pour ce faire, nous avons mis en place des contrôles vérifiant que la saisie est bien un nombre positif et, le cas échéant, que c'est bien un entier (donnée numérique ne contenant pas de virgule). Si un utilisateur saisit une donnée d'un format incorrect, un message d'erreur clair lui indique le problème (cf. Figures 19 et 20) et donne un exemple de saisie ayant le bon format, lui permettant ainsi de corriger rapidement sa saisie.

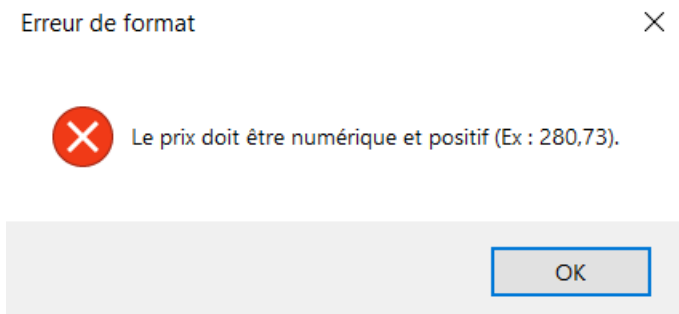


Figure 19 : Message d'erreur de format de la donnée saisie pour les valeurs numériques

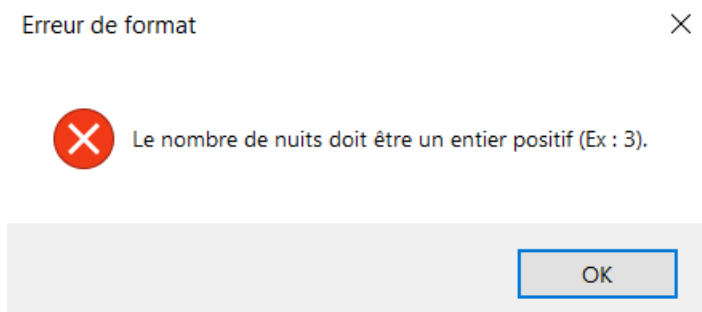


Figure 20 : Message d'erreur de format de la donnée saisie pour les entiers

Le bouton permettant de réinitialiser les choix entrés par l'utilisateur a également été ajouté à cette étape. Ce bouton nous paraissait être relativement important pour que

l'application soit plus fluide et efficace. Afin d'éviter les actions accidentelles, un message de confirmation a été intégré. Lorsqu'un utilisateur clique sur le bouton de réinitialisation, une boîte de dialogue s'affiche à l'écran (cf. Figure 21) pour demander à l'utilisateur s'il est sûr de vouloir supprimer toutes ces saisies. Il peut alors confirmer ou annuler cette action, ce qui permet d'améliorer l'expérience utilisateur.

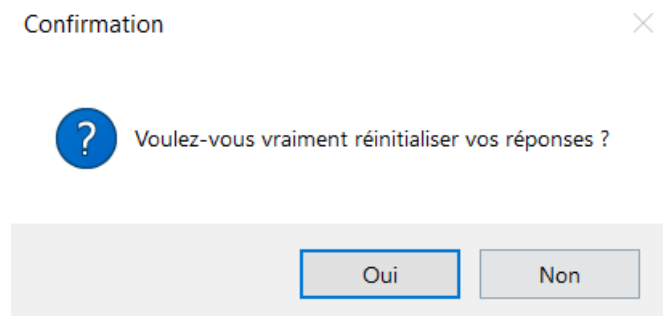


Figure 21 : Message de confirmation de réinitialisation des choix

4. Un formulaire concis et minimaliste

Lors de la conception de l'application, un soin particulier a été apporté à la création d'un formulaire à la fois concis et minimaliste. L'objectif était de maximiser l'efficacité et l'expérience utilisateur tout en réduisant la complexité.

Les questions du formulaire ont été reformulées et limitées aux informations strictement nécessaires à l'utilisateur pour répondre. Ce choix garantit un remplissage efficace et précis des données en évitant de submerger l'utilisateur avec des informations inutiles. Chaque sous-partie du questionnaire est encadrée et clairement identifiée, ce qui rend le formulaire plus structuré et facile à parcourir.

L'interface est désormais plus épurée et améliore le confort visuel de l'utilisateur (cf. Figures 13 et 14).

5. Une page d'information « À propos »

Lors du Sprint 4, nous avons décidé de créer une page contenant diverses informations accessibles depuis la page d'accueil via le bouton « À propos » (cf. Figure 10). Cette page est sur une feuille Excel invisible à l'ouverture de l'application et est rendue visible lorsque l'utilisateur clique sur le bouton prévu pour ouvrir la page. Sur cette page, nous avons mis un bouton « Retour » permettant de revenir à la page d'accueil. Lorsque l'utilisateur utilise ce bouton, la feuille contenant la page « À propos » est rendue à nouveau invisible. Nous avons

fait le choix de rendre cette feuille invisible lorsqu'elle est inactive pour éviter le doublon d'option d'accès à la page et donc de surcharger l'utilisateur d'informations inutiles.

Sur cette page (cf. Figure 11 et 12), nous avons décidé de mettre des informations sur le contexte de création de l'application, de lister les collaborateurs ayant participé à la création de cette application ainsi que leur rôle et leur contact. Enfin, nous avons intégré des informations pour permettre à l'utilisateur de mieux comprendre le fonctionnement de la prédiction, s'il le souhaite. Ces informations comprennent les différentes variables utilisées ainsi qu'une brève explication du modèle prédictif choisi.

6. Une page résultat complète

C'est lors du dernier sprint que nous avons peaufiner la dernière page des résultats (cf. Figures 16 et 17). Dans cette page, on retrouve quatre sections clairement distinctes et encadrées. Chaque section a été pensée pour répondre à des besoins spécifiques.

La section résultat affiche clairement, avec un code couleur, le statut prédit par le modèle. On affiche également avec le même code couleur le pourcentage de risque d'annulation de la réservation. Afficher le pourcentage de risque d'annulation est essentiel pour fournir une vision nuancée et probabiliste de la prédiction. Contrairement à une simple classification binaire (« annulation » ou « confirmation »), le pourcentage de risque reflète l'incertitude inhérente au modèle et permet d'éviter une interprétation trop rigide des résultats.

La section « Recommandations » a été intégrée pour transformer les prédictions en actions concrètes. En proposant des conseils spécifiques adaptés au statut de la réservation, comme envoyer un rappel ou des offres personnalisées au client ou surbooker les réservations, elle aide le gérant à réduire le risque d'annulation ou à maximiser l'utilisation des ressources de l'hôtel. Cette approche proactive ajoute une valeur opérationnelle directe à l'application.

L'affichage de la fiabilité du modèle dans la section « Confiance » est essentiel pour apporter de la transparence et renforcer la confiance des utilisateurs dans les prédictions fournies. La fiabilité permet au gérant de l'hôtel d'évaluer dans quelle mesure la prédiction peut être prise en compte dans sa prise de décision. En incluant cette information, on souligne que les résultats ne sont pas absolus, mais basés sur une probabilité calculée par le modèle. Cela aide à éviter une dépendance aveugle à l'outil, surtout dans les cas où l'incertitude est élevée.

Pour finir, la section « Facteurs principaux », qui donne les facteurs principaux qui influencent le résultat de la prédiction, vise à rendre le modèle plus interprétable et à offrir

des clés de compréhension au gérant. En listant les variables ayant eu le plus d'impact sur le résultat, comme le délai de réservation, le tarif ou la durée du séjour, cette section montre pourquoi une réservation est perçue comme risquée.

7. Dernières améliorations de l'expérience utilisateur

Lors de ce dernier sprint, nous avons aussi apporté quelques dernières modifications pour améliorer l'expérience de l'utilisateur.

Après avoir demandé des avis de personnes extérieurs au projet, quelques suggestions nous ont été proposés. Il y a notamment eu comme proposition l'ajout d'une page récapitulative des données saisies par l'utilisateur (cf. Figure 15). Cette idée est fondée sur le fait que notre formulaire est décomposé sur deux pages et cela permet à l'utilisateur de confirmer les choix qu'il a pu faire. Cela améliore l'expérience utilisateur et enlève toute frustration liée à des erreurs de saisie qu'il aurait pu faire. En effet, lorsqu'il lance la prédiction il est sûr de lui et cela lui évite de devoir tout recommencer dans le cas où il ne l'était pas. Sur cette page, nous retrouvons donc un bouton « Corriger » lui permettant de retourner au début du formulaire dans l'éventualité où il s'est trompé. Le bouton « Lancer la prédiction » a donc été déplacé sur cette page et ne se trouve plus sur la deuxième page du formulaire. Il y a toujours un bouton « Quitter » permettant de quitter complètement l'application.

La seconde amélioration a été l'ajout d'un bouton d'information pour la variable décrivant le type de clients (cf. Figure 22). Lorsqu'on clique sur ce bouton, un message qui décrit les modalités de cette variable s'affiche à l'écran (cf. Figure 23). Nous avons fait ce choix car nous estimions que les modalités de cette variable n'étaient pas très parlantes. Ainsi, cet ajout permet à l'utilisateur de mieux comprendre la variable et de choisir la bonne modalité en toute confiance, sans se demandant s'il se trompe ou non.

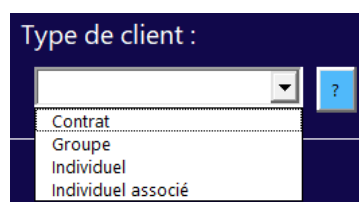


Figure 22 : Bouton d'information et modalités pour la variable décrivant le type de client



La variable 'Type de client' décrit la catégorie du client :

- Contrat : Client dont la réservation est liée à un contrat ou un allotement (ex : entreprises).
- Groupe : Client faisant partie d'un groupe (ex : conférences, mariages).
- Individuel : Client ayant réservé indépendamment, sans être lié à un groupe ou contrat.
- Individuel associé : Client ayant réservé individuellement mais lié à d'autres réservations individuelles (ex : amis voyageant ensemble).

OK

Figure 23 : Message d'informations sur les modalités de la variable décrivant le type de client

Nous avons également amélioré la gestion des erreurs de saisie de l'utilisateur. Désormais, lorsque l'utilisateur tape un résultat non attendu, la police du texte saisi est immédiatement affichée en rouge (cf. Figure 24). Cela permet d'attirer immédiatement l'attention sur les erreurs et facilite leur correction. Cette mise en évidence visuelle rend l'expérience utilisateur plus intuitive en guidant rapidement l'utilisateur vers les champs nécessitant une modification. Cela améliore l'efficacité du formulaire.

Historique et préférences du client :

Nombre de réservations non annulées :	Nombre de réservations annulées :
<input type="text" value="Bonjour"/>	<input type="text" value="1,5"/>
Nombre de demandes spéciales :	
<input type="text" value="-2"/>	
Combien de jours avant son arrivée le client a-t-il réservé son séjour ?	
<input type="text" value="18"/>	

Figure 24 : Exemple illustrant la gestion des erreurs de saisie (texte en rouge)

La dernière amélioration que nous avons apportée est une page de chargement (cf. Figure 25). Cette page avertit l'utilisateur. Cette page permet de gérer l'attente entre l'entrée des données et l'affichage des résultats de manière transparente et professionnelle. Elle informe l'utilisateur que le système est en train de traiter sa demande, ce qui réduit les risques de confusion ou d'impatience.

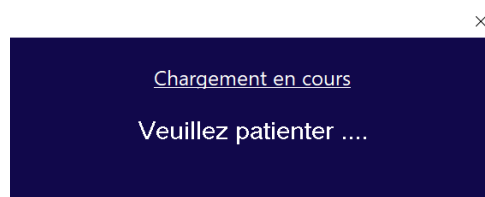


Figure 25 : Page de chargement

VI. Environnement de développement

A. Langage de programmation

Python, langage polyvalent et accessible, a joué un rôle central dans le développement de notre application. Grâce à sa syntaxe claire et intuitive, il a permis une manipulation fluide des données via des bibliothèques comme Pandas et NumPy, et une création efficace de modèles prédictifs avec Scikit-learn. En particulier, notre modèle final de prédiction a été implémenté à l'aide d'un Random Forest, choisi pour sa robustesse et sa capacité à gérer des données complexes sans nécessiter un prétraitement intensif.

La synergie entre Python et VBA a également été un atout majeur du projet. Alors que VBA, intégré à Excel, a permis de concevoir une interface utilisateur intuitive, Python nous a quant à lui permis d'utiliser sa puissance analytique impressionnante. Nous avons ainsi développé un pipeline VBA (Python permettant d'associer l'affichage d'une interface graphique avec le travail algorithmique de Python).

De cette façon, nous avons pu envoyer des données brutes qui ont été enregistrées via Excel, transférées vers un fichier .csv, analysées par des scripts Python et renvoyées à Excel après analyse pour affichage. Cette approche garantit une expérience utilisateur fluide tout en maintenant des capacités analytiques avancées.

L'un des défis rencontrés a été la gestion dynamique des chemins d'accès, car le projet devait fonctionner indépendamment de son emplacement sur les machines des utilisateurs. Cela a été résolu grâce à une automatisation complète des chemins pour Python et pour VBA.

La question que vous devez certainement vous poser à présent est la suivante : « En quoi ces éléments sont-ils pertinents vis-à-vis de leur démarche ? ». En effet, cela peut paraître obsolète, mais tout ce travail nous amène vers un sujet essentiel : La sécurisation des données. Dans le monde actuel, où les individus mal intentionnés exploitent souvent les informations sensibles à des fins personnelles, il est crucial d'assurer une protection optimale des données. C'est pourquoi il n'est pas possible ici de récupérer une quelconque information relative à une réservation ou à un client, puisque celles-ci ainsi que les fichiers associés sont directement supprimés à la suite de l'affichage.

Cette sécurisation des données est l'un des piliers fondamentaux de notre travail. Sans cette précaution, une application commerciale telle que celle que nous avons développée ne pourrait être mise en ligne en toute confiance.

B. Répot GIT

Lien : [Booksure l'application de gestion de réservations](#)

Toutes les explications relatives à notre projet sont disponibles dans le « Readme.md » de notre répot GIT. Le lien est disponible ci-dessus.

VII. Limites et ouvertures

A. Limites de l'application

Contexte spécifique des données

Les données utilisées pour entraîner et tester l'application proviennent exclusivement d'hôtels situés au Portugal. En conséquence, le modèle est spécifiquement adapté aux comportements de réservation et d'annulation dans ce pays, ce qui pourrait limiter son efficacité dans d'autres contextes géographiques (par exemple, des pays avec des comportements de consommation différents ou des saisons touristiques décalées).

Hypothèses simplificatrices

Certaines variables explicatives n'ont pas été incluses ou simplifiées pour des raisons de disponibilité des données ou de modélisation (par exemple, l'impact des événements locaux ou la météo).

Limites techniques

L'application actuelle ne prend pas en charge les réservations en temps réel, mais s'appuie sur un traitement des données déjà collectées. Les performances du modèle peuvent diminuer si les comportements des clients évoluent (par exemple, à la suite de la pandémie ou de nouvelles habitudes de réservation).

B. Ouvertures possibles

Adaptation à d'autres contextes

Enrichir la base de données avec des réservations provenant d'hôtels situés dans d'autres régions ou pays pour rendre le modèle plus général.

Amélioration des fonctionnalités

Intégrer l'application au logiciel de gestion des hôtels (PMS - Property Management System), pour qu'elle puisse directement proposer des recommandations basées sur les données des réservations en cours.

Ajouter une fonctionnalité interactive permettant à l'utilisateur hôtelier de visualiser les prédictions et d'ajuster certains paramètres (par exemple, annulations prévues par canal ou par période).

Développement d'un tableau de bord prédictif

Proposer un tableau de bord dédié à la prédiction et à l'analyse des annulations, intégrant des indicateurs visuels comme des graphiques ou des alertes.

VIII. Collaboration et sprints suivant la méthode AGILE

Ce projet est mené par un groupe de cinq personnes, chacun occupant un rôle spécifique au sein de l'équipe pour assurer une répartition claire des responsabilités. La collaboration repose sur l'utilisation de Git, qui facilite le suivi et le partage des contributions. Les tâches sont attribuées soit lors des réunions d'équipe, soit directement via des tickets assignés sur la plateforme Git, garantissant une organisation efficace et structurée.

Répartition des rôles selon les forces de chacun :

- Product Owner : Baya BENOUREDJA
- Front interface / Scrum Master : Lucile SAILLANT
- Data Engineer / Scrum Master : Lucas COLARD
- Data Analyst : Clara RAKOTOARISOA
- Data Scientist : Rodolphe GABA

Les présentations des avancées de notre travail ont été organisées en quatre sprints, chacun d'une durée de 15 jours sauf exception. Une ou deux réunions par sprint sont organisées, soit en présentiel, soit en distanciel, pour s'organiser et faire le point sur les avancées et réalisations de chacun. Chaque sprint a permis de définir des objectifs précis, ainsi que des critères d'achèvement (DoD - Definition of Done) pour assurer la qualité et la complétude des tâches. Voici un résumé des réalisations de chaque sprint.

A. SPRINT 1

Objectifs :

- Conception d'un produit analytics en réponse à un besoin client identifié.
- Construction de la roadmap produit découpant les étapes en différents lots de valeur.
- Construction d'un premier jeu de données permettant de traiter la problématique.
- Validation de la prise en main du setup de développeur.

DOD :

- Des maquettes d'interface sont construites et permettent de comprendre le parcours de l'utilisateur au sein de l'application.
- Une roadmap produit détaillée par compétences est construite & découpée dans le temps en différents lots de valeur.
- La variable d'intérêt & les variables explicatives sont présentées de façon explicite & documentées au sein du rapport.
- Un premier script python exécutable est disponible au sein du projet GitHub validant l'autonomie de l'équipe dans l'utilisation des outils de développement.

Résumé du sprint :

Au départ, nous avons choisi de travailler sur la prédiction de la probabilité d'acceptation ou de refus d'un prêt bancaire. Cependant, après un premier sprint et une présentation de ce sujet, nous avons décidé de changer de direction en raison de la difficulté à trouver une base de données fiable. Nous avons alors opté pour la prédiction des annulations de réservations d'hôtel. Cette réorientation nous a conduits à réadapter ce qui était réutilisable du premier sprint et à ajuster les tâches pour aligner notre travail sur le nouveau sujet, tout en veillant à réorganiser le planning du sprint suivant afin d'éviter tout retard.

La raison pour laquelle nous avons choisi ce sujet est qu'il s'inscrit dans un secteur en pleine mutation, offrant des opportunités intéressantes pour développer des solutions prédictives.

Pour mener ce projet, nous avons utilisé un jeu de données issu de l'article de recherche de Nuno Antonio et al., trouvé sur Kaggle. Nous avons structuré notre travail à travers une roadmap organisée en lots de valeur avec une planification chronologique, en adoptant une approche agile pour garantir un développement efficace et progressif. Le parcours utilisateur de l'application a été défini grâce à des maquettes, précisant les principales entrées, sorties et fonctionnalités clés, afin d'assurer une expérience fluide et intuitive. Par ailleurs, nous avons pris en main le setup technique en créant un dépôt Git pour la collaboration, en élaborant un formulaire VBA avec une page d'accueil pour gérer les données, et en développant un script Python basique pour tester les premières fonctionnalités.



Figure 26 : Maquettes de l'interface

B. SPRINT 2

Objectifs :

- Procéder au nettoyage de la base de données tout en identifiant les variables d'intérêt et les variables explicatives.
- Élaborer un modèle initial de régression linéaire.

- Intégrer la protection des données personnelles dans la stratégie produit.
Implémentation d'un pipeline VBA/Python communiquant dans les deux sens.

DOD :

- Une base de données propre, variables d'intérêt identifiées et variables explicatives sélectionnées.
- Un modèle basique de régression linéaire qui permet de prédire la variable d'intérêt en fonction de plusieurs variables explicatives.
- Processus de collecte et de gestion des données sécurisés en faisant en sorte que les réponses ne soient pas sauvegardées mais seulement utilisées pour la prédiction.

Résumé du sprint :

Pour le traitement des données, nous avons effectué un nettoyage approfondi, incluant la gestion des valeurs manquantes, des incohérences, des doublons, et des valeurs aberrantes, ainsi que la transformation de variables quantitatives en binaires. Une analyse exploratoire a permis d'examiner les principales caractéristiques des données via des graphiques, des statistiques descriptives, et l'étude des corrélations, afin de détecter tendances et anomalies. Nous avons ensuite développé un modèle de régression logistique de base, défini les variables clés et évalué les résultats obtenus. Enfin, l'intégration entre VBA et Python a été mise en place, avec des tests et une sécurisation garantissant la suppression des données après chaque prédiction. Parmi les faits marquants, nous avons également dû intégrer une pause pédagogique de 15 jours, ce qui a impacté la planification globale du projet.

C. SPRINT 3

Objectifs :

- Une interface utilisateur intuitive permettant d'afficher les résultats des estimations de réservation, avec une présentation claire des données prédictives.
- Gérer les saisies erronées du formulaire lorsque les entrées ne respectent pas le format attendu.
Effectuer une analyse univariée et bivariée pour identifier les corrélations et les tendances entre les variables, afin d'affiner la sélection des données.
- Réaliser une sélection de variables pour retenir celles ayant le plus d'impact sur l'issue des réservations, puis mettre à jour le modèle de prédiction en intégrant ces variables pour améliorer la précision.

DOD :

- Une interface qui affiche les résultats de réservation de manière lisible, avec des données d'estimation qui se chargent et se mettent à jour correctement en fonction des nouvelles prédictions.

- Des messages d'erreur s'affichent en cas de saisie manquante ou de format incorrect, guidant l'utilisateur pour corriger les entrées.
- Les analyses univariées et bivariées des variables pertinentes sont effectuées, les corrélations identifiées sont synthétisées, et les données préparées et formatées.
- La sélection des variables, accompagnée de la documentation des critères, est réalisée et le modèle de prédiction est mis à jour avec ces variables.

Résumé du sprint :

Durant cette quinzaine, nous avons réorganisé les scripts Python en suivant les bonnes pratiques de codage et commencé la rédaction du README. L'analyse exploratoire a été approfondie, avec un comparatif des données avant et après nettoyage, en tenant compte de l'équilibre de la base. Le modèle prédictif a été amélioré. Enfin, dans l'interface utilisateur, nous avons géré les entrées erronées avec des messages d'erreur et intégré les résultats du modèle avec un affichage clair des prédictions.

D. SPRINT 4

Objectifs :

- Rendre le formulaire de l'application plus concis et l'adapter pour qu'il tienne sur une seule page, tout en améliorant la clarté des prédictions et l'expérience utilisateur.
- Développer un modèle Random Forest, comparer ses performances au modèle de régression logistique à l'aide de métriques comme la précision.

DOD :

- Le formulaire est réduit à une seule page, les informations sont présentées de manière claire, testé par les utilisateurs, et validé pour une amélioration de l'expérience.
- Le modèle Random Forest est développé, comparé au précédent avec des métriques clés et validé si performances supérieures.

Résumé du sprint :

Pour ce sprint, nous avons bénéficié d'un délai légèrement prolongé de 18 jours. Nous avons consacré une part importante de ce temps à la rédaction du rapport, à la mise à jour de toutes les tâches accomplies, ainsi qu'à l'optimisation de l'application. Nous avons également travaillé sur d'autres modèles prédictifs, tels que le Random Forest et XGBoost, ainsi que sur deux autres modèles, que nous avons comparés au modèle initialement conçu afin de sélectionner le modèle final. En ce qui concerne l'application, le formulaire a été amélioré, avec des informations présentées de manière claire, testé par les utilisateurs et validé pour améliorer leur expérience. Par ailleurs, nous avons réécrit le README en anglais. De plus, nous avons réalisé les tests unitaires manquants pour garantir la robustesse et la fiabilité de notre application.

E. SPRINT 5

Objectifs :

- Finaliser le formulaire en y incluant uniquement les entrées pertinentes et mettre à jour la documentation du projet.
- Améliorer la sortie des prédictions en ajoutant des analyses supplémentaires et des recommandations.

DOD :

- Le formulaire doit être mis à jour avec uniquement les entrées pertinentes, et le bouton "À propos" doit être actualisé avec les détails des variables utilisées et les informations sur le modèle employé..
- La sortie de prédiction doit inclure des intervalles de confiance clairs, des analyses supplémentaires pertinentes présentées sous forme de graphiques ou des recommandations visuellement attractives intégrées dans l'interface.

Résumé du sprint :

Ce sprint marque l'achèvement du projet, avec pour objectif principal la finalisation de l'ensemble du travail. Nous avons amélioré les résultats de prédiction en intégrant des recommandations et des détails sur l'algorithme utilisé, ainsi que le pourcentage de risque d'annulation des réservations offrant ainsi à l'utilisateur la liberté de les interpréter selon son contexte. Le modèle de prédiction a été ajusté et optimisé pour garantir une meilleure performance et une plus grande précision. Le rapport a été révisé et complété pour inclure toutes les informations nécessaires, tandis que le support de présentation a été préparé de manière claire et professionnelle. Le dépôt Git a également été réorganisé, avec un nettoyage et une suppression des éléments superflus pour une structure plus cohérente.

Ce sprint a permis de consolider l'ensemble du travail réalisé, assurant ainsi une application fonctionnelle, optimisée et prête à être utilisée.

IX. Annexes

A. Dictionnaire des variables

Variable	Description	Type	Exemple
Hotel	Indique le type d'hôtel où la réservation a été faite.	Catégorielle	City Hotel
ADR	Tarif moyen journalier, calculé en divisant le total des transactions d'hébergement par le nombre total de nuits.	Décimal	105.5
Adults	Nombre d'adultes pour la réservation.	Entier	2
Agent	ID de l'agence de voyage ayant effectué la réservation.	Entier	240
ArrivalDateDayOfMonth	Jour du mois prévu pour l'arrivée du client.	Entier	9
ArrivalDateMonth	Mois prévu pour l'arrivée du client.	Catégorielle	July
ArrivalDateWeekNumber	Numéro de la semaine pour l'arrivée.	Entier	27
ArrivalDateYear	Année de la date d'arrivée.	Entier	2015
AssignedRoomType	Code du type de chambre assignée	Catégorielle	C
Babies	Nombre de bébés pour la réservation.	Entier	1
BookingChanges	Nombre de modifications faites à la réservation initiale.	Entier	3
Children	Nombre d'enfants pour la réservation.	Entier	0
Company	ID de l'entreprise ayant réservé ou payé la réservation	Entier	51
Country	Pays d'origine de la réservation	Catégorielle	ARG
CustomerType	Catégorie de la réservation	Catégorielle	Contract
DaysInWaitingList	Nombre de jours où la réservation est restée en attente avant confirmation.	Entier	0

DepositType	Indication sur le dépôt effectué pour garantir la réservation	Catégorielle	No Deposit
DistributionChannel	Canal par lequel la réservation a été effectuée	Catégorielle	Corporate
IsCanceled	Indique si la réservation a été annulée	Binaire	0
IsRepeatedGuest	Indique si le client est récurrent ou non	Binaire	1
LeadTime	Nombre de jours entre la date de réservation et la date d'arrivée	Entier	342
MarketSegment	Type de marché par lequel la réservation a été faite	Catégorielle	Online TA
Meal	Type de repas inclus dans la réservation	Catégorielle	BB
PreviousBookingsNotCanceled	Nombre de réservations précédentes non annulées par le client	Entier	0
PreviousCancellations	Nombre de réservations précédentes annulées par le client	Entier	11
RequiredCardParkingSpaces	Nombre de places de parking réservées par le client	Entier	3
ReservationStatus	Dernier statut de la réservation	Catégorielle	Check-Out
ReservationStatusDate	Date à laquelle le statut actuel de la réservation a été mis à jour.	Date	01/07/2015
ReservedRoomType	Code du type de chambre réservée	Catégorielle	F
StaysInWeekendNights	Nombre de nuits passées ou réservées le week-end	Entier	19
StaysInWeekNights	Nombre de nuits passées ou réservées en semaine	Entier	6
TotalOfSpecialRequests	Nombre total de demandes spéciales du client (ex. : lits jumeaux)	Entier	5

B. Roadmap du projet

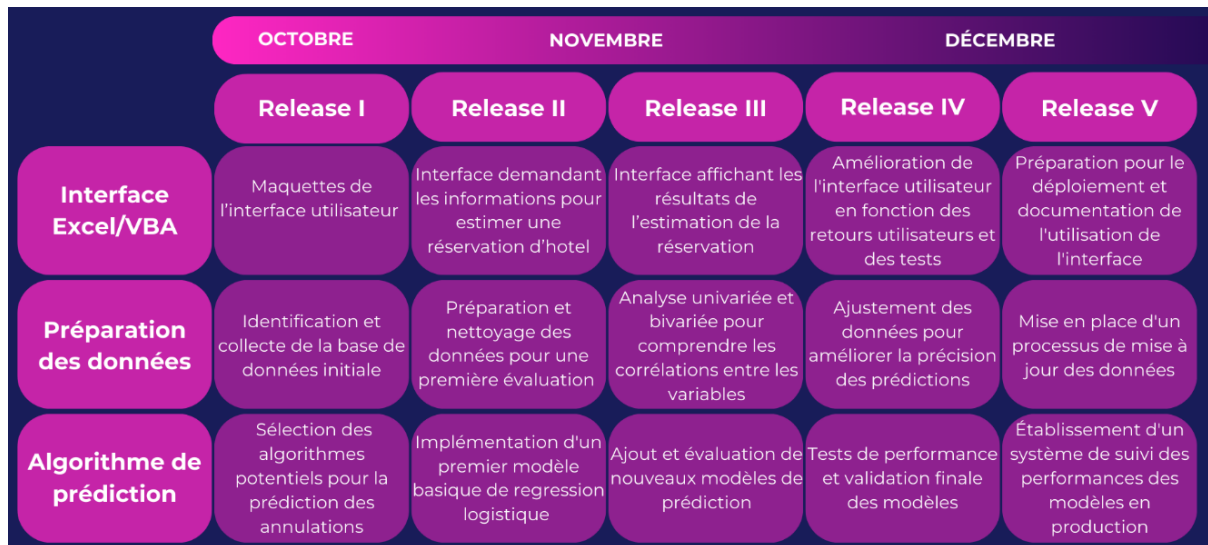


Figure 27 : Roadmap du projet

C. Analyse des corrélations des variables deux à deux

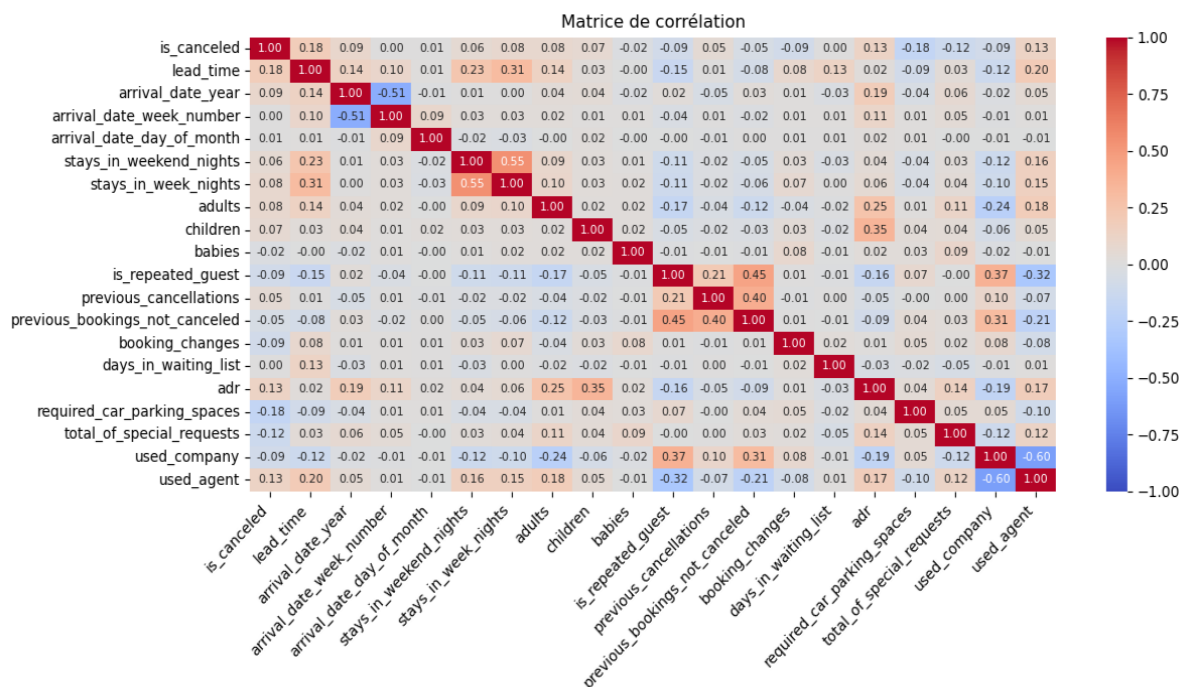


Figure 28 : Matrice de corrélation

D. Performances du modèle de régression logistique

```

Optimization terminated successfully.
Current function value: 0.570124
Iterations 7

Results: Logit
=====
Model:                Logit                Method:                MLE
Dependent Variable:    is_canceled          Pseudo R-squared:      0.032
Date:                 2024-11-04 20:01      AIC:                   99145.2261
No. Observations:      86940                BIC:                   99201.4640
Df Model:              5                    Log-Likelihood:        -49567.
Df Residuals:          86934                LL-Null:               -51211.
Converged:             1.0000                LLR p-value:           0.0000
No. Iterations:        7.0000                Scale:                 1.0000
=====

```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
previous_cancellations	1.0606	0.0420	25.2574	0.0000	0.9783	1.1429
booking_changes	-0.4677	0.0153	-30.5756	0.0000	-0.4977	-0.4377
total_of_special_requests	-0.4940	0.0103	-47.8053	0.0000	-0.5143	-0.4737
used_company	-1.6005	0.0491	-32.5908	0.0000	-1.6967	-1.5042
used_agent	-0.4248	0.0105	-40.3671	0.0000	-0.4454	-0.4042
is_repeated_guest	-2.0425	0.0824	-24.7766	0.0000	-2.2041	-1.8809

Figure 29 : Évaluation du modèle de régression logistique pour la prédiction des annulations de réservations hôtelières

E. Evaluation du modèle final : RandomForest

Matrice de confusion

		Predicted	
		Predicted Positive	Predicted Negative
True	Positive class	3451	1296
	Negative class	1117	3636

Métriques d'évaluation du modèle

	precision	recall	f1-score	support
0	0.76	0.73	0.74	4747
1	0.74	0.76	0.75	4753
accuracy			0.75	9500
macro avg	0.75	0.75	0.75	9500
weighted avg	0.75	0.75	0.75	9500

Figure 30 : Indicateurs de performance du modèle

Seuil de décision

Selon le graphique ci-après, le seuil de 0.5 pour la répartition des scores prédits semble convenir et permet un compromis entre la spécificité et la sensibilité du modèle. Ce seuil est logique puisque le modèle de forêt aléatoire fonctionne via un système de vote majoritaire avec les arbres qui le constituent.

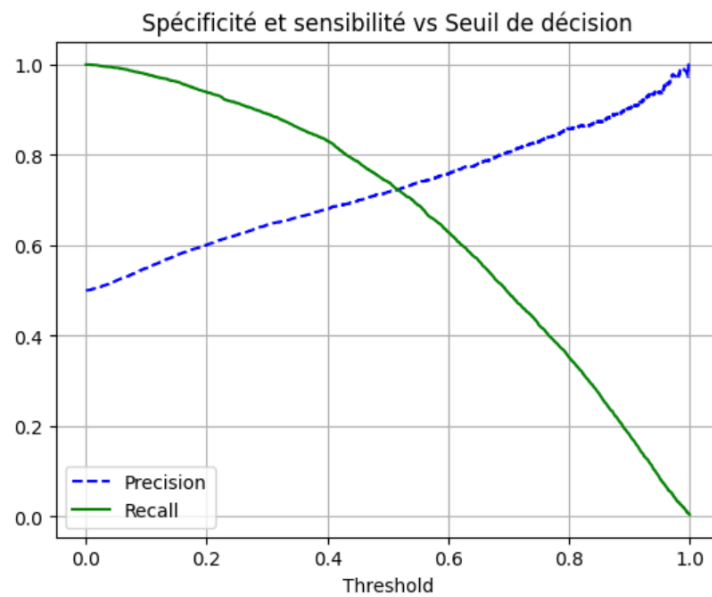


Figure 31 : Seuil de décision permettant un compromis entre spécificité et sensibilité

F. Versions de l'applications

1. Version 0

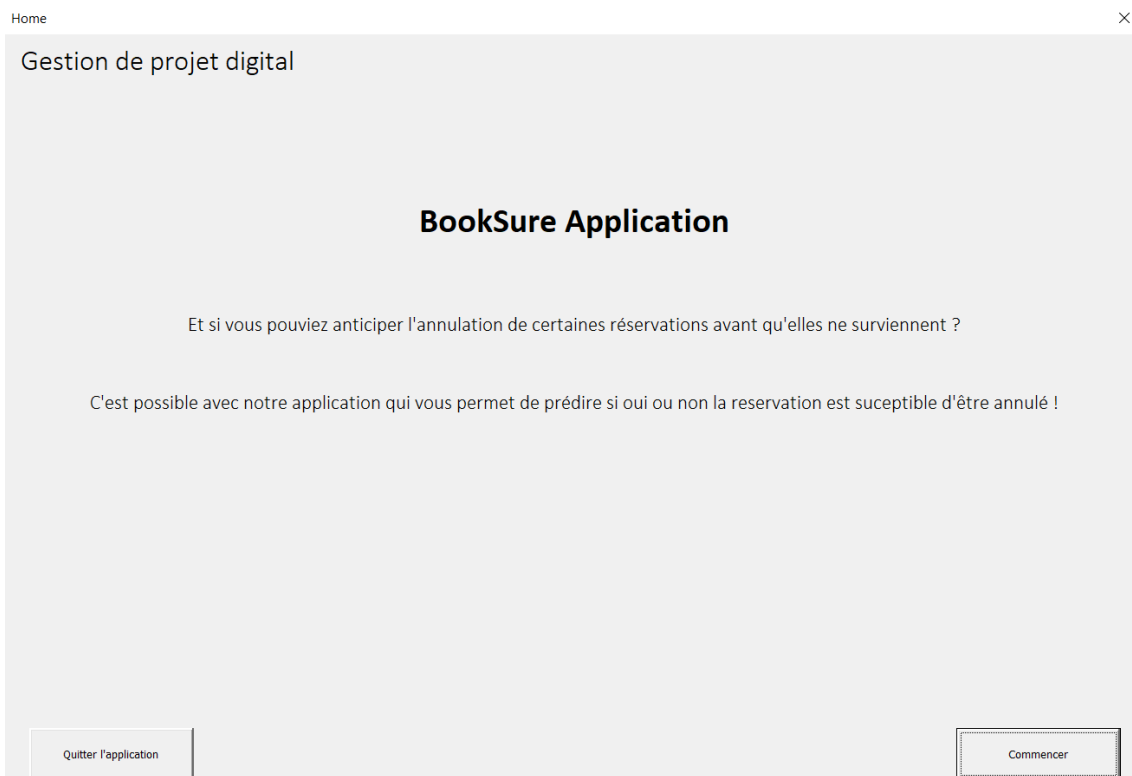


Figure 32 : Page d'accueil de l'application (version 0)

Questionnaire

Informations générales sur la réservation :

Quel est le type de votre hotel ?

Quel est le prix moyen d'une nuit pour cette réservation ?
Montant : €

Quel est le type de dépôt pour garantir la réservation ?

Informations sur le séjour :

Quel est la date d'arrivée prévue ?

Jour

Mois

Année

Combien de nuits sont réservées en week-end ?

Combien de nuits sont réservées en semaine ?

Quel est le type de chambre réservé ?

Combien de fois la réservation a-t-elle été modifiée ?

Pendant combien de jours la réservation a-t-elle été en liste d'attente avant d'être confirmée ?

[Retour à la page d'accueil](#)

[Suite du questionnaire](#)

Figure 33 : Première page du formulaire de l'application (version 0)

Questionnaire

Informations sur le client :

Depuis quel pays la réservation a-t-elle été effectuée ?

Quel est le type de client de la réservation ?

Combien y a-t-il d'adultes dans la réservation ?

Combien y a-t-il d'enfants dans la réservation ?

Combien y a-t-il de bébés dans la réservation ?

Le client est-il un client récurrent ?
☐ Oui ☐ Non

Historique et préférences du client :

Quel est le nombre de réservations précédentes non annulées par le client ?

Quel est le nombre de réservations précédentes annulées par le client ?

Combien de demandes spéciales a formulé le client ?

Combien de place de parking a réservé le client ?

Combien de jours avant son arrivée le client a-t-il réservé son séjour ?

Quel type de repas a réservé le client ?

[Retour à la page d'accueil](#)

[Page précédente](#)

[Lancer la prédiction](#)

Figure 34 : Deuxième page du formulaire de l'application (version 0)

Quizz

Questionnaire

Informations générales sur la réservation :

Quel est le type de votre hotel ?

Quel est le prix moyen d'une nuit pour cette réservation ?
Montant : €

Quel est le type de dépôt pour garantir la réservation ?

Informations sur le séjour :

Quel est la date d'arrivée prévue ?
Jour Mois Année

Combien de nuits sont réservées en week-end ?

Combien de nuits sont réservées en semaine ?

La chambre réservée par le client est-elle la même que celle qui lui a été assignée ?
☒ Oui ☐ Non

Combien de fois la réservation a-t-elle été modifiée ?

Pendant combien de jours la réservation a-t-elle été en liste d'attente avant d'être confirmée ?

Quitter

Réinitialiser les choix

Suite du questionnaire

Quiz

X

Questionnaire

Informations sur le client :

Depuis quel pays la réservation a-t-elle été effectuée ?

Quel est le type de client de la réservation ?

Combien y a-t-il d'adultes dans la réservation ?

Combien y a-t-il d'enfants dans la réservation ?

Combien y a-t-il de bébés dans la réservation ?

Le client est-il un client récurrent ?

☒ Oui ☐ Non

Historique et préférences du client :

Quel est le nombre de réservations précédentes non annulées par le client ?

Quel est le nombre de réservations précédentes annulées par le client ?

Combien de demandes spéciales a formulé le client ?

Combien de place de parking a réservé le client ?

Combien de jours avant son arrivée le client a-t-il réservé son séjour ?

Quel type de repas a réservé le client ?

Quitter

Page précédente

Réinitialiser les choix

Lancer la prédiction

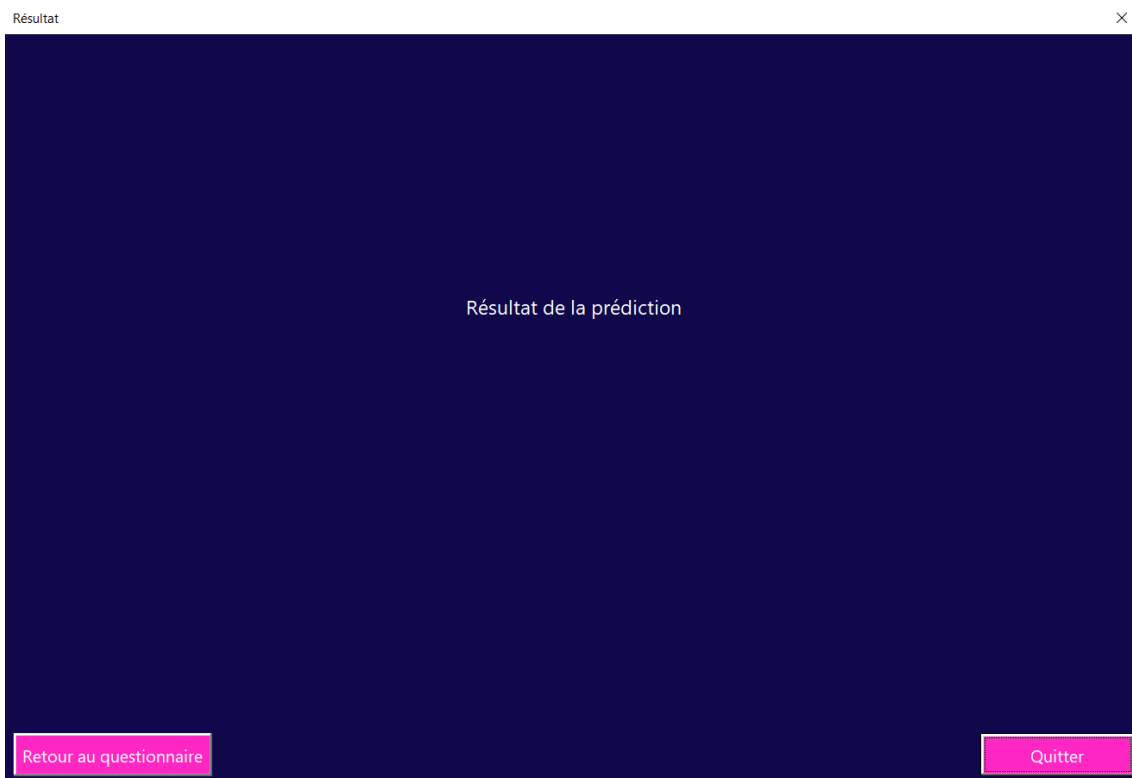


Figure 39 : Page résultat de l'application (version 1)