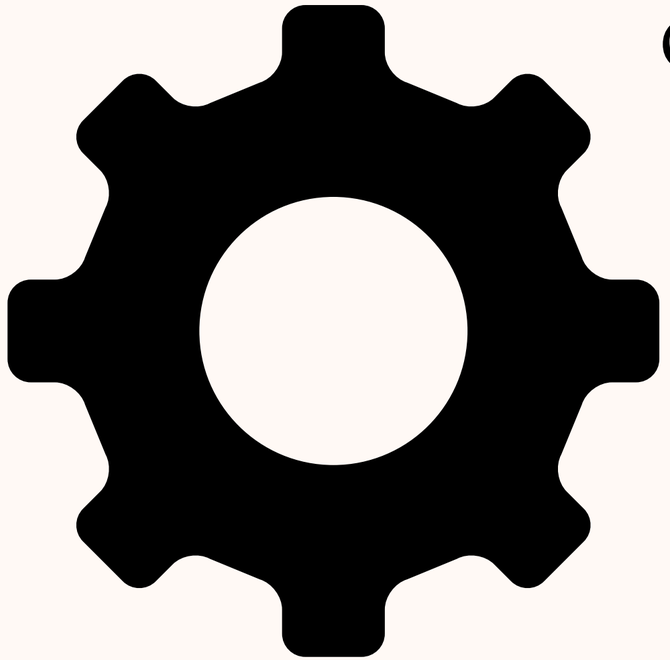
SOMMAIRE

- INTRODUCTION
- OUTILS & METHODOLOGIE
  - Outils
  - Méthodologie
- RESULTATS
- CONCLUSION

INTRODUCTION

Le marché des billets d’avion est caractérisé par une forte volatilité des prix, influencée par de nombreux facteurs tels que la demande, la saisonnalité les standings et les stratégies des compagnies aériennes. Prédire ces variations est un défi majeur tant pour les consommateurs que pour les entreprises du secteur. Ce projet vise à développer un modèle prédictif utilisant des techniques de Big Data et de Machine Learning pour estimer l'évolution des prix des billets d’avion en fonction de divers paramètres (destination, période, historique, etc.). L'objectif est de fournir un outil d’optimisation pour les voyageurs tunisiens et une meilleure stratégie tarifaire pour les compagnies aériennes.

I - OUTILS & METHODOLOGIE



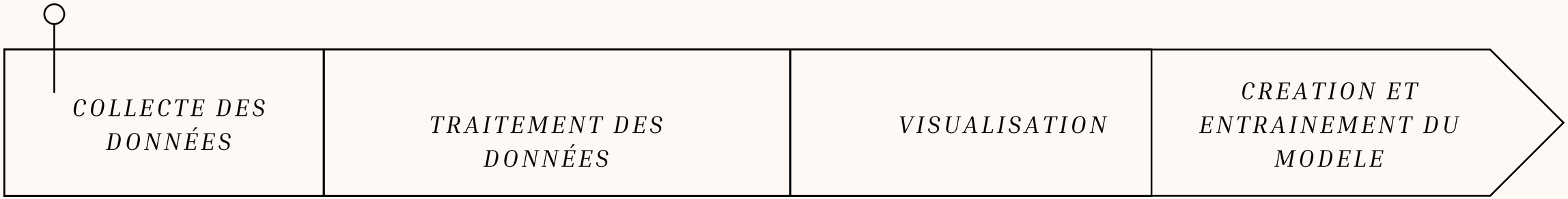
OUTILS

Afin de nous simplifier la tache nous avons travailler dans un container **Docker**, notre Dockerfile nous a permis de configurer l'environnement avec un server Jupyter et Spark. Voici de manière général les autres technologies utilisées.

- PySpark & FindSpark : pour le traitement des données et la création du modèle
- Pandas,Matplolibet Seanborn: pour la visualisation des données
- Chromium & BeautifulSoup : pour la collecte des données

METHODOLOGIE

Source :
Kayak



Collecte des données

Des données sur Kayak ont été collectées avec Playwright et BeautifulSoup pour contourner les captchas et simuler un comportement humain. Ces informations concernent les prix des billets d'avion vers plusieurs destinations, notamment Madrid, Cotonou, Tunis, Bruxelles, Anvers, Ostrava, Luxembourg, N'Djamena, Abidjan, Paris et Alger.

Traitement des données

Dans cette section, nous avons formaté les données, éliminer les doublons, les valeurs nulles et les colonnes inutiles. De plus, nous avons créé de nouvelles colonnes, d'abord pour faciliter la visualisation, puis pour préparer la création du modèle.

METHODOLOGIE

Visualisation des donnés

Nous avons utiliser matpolib pour générer des graphes qui nous ont permis d’aboutir a des conclusions tel que les périodes ou les tickets sont les plus elevées,la compagnie qui occupe la plus grande part de marché et bien d’autre

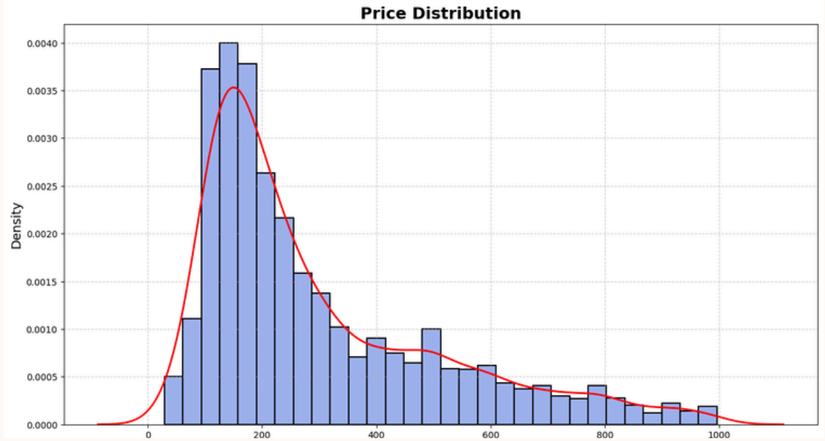
Création et entrainement du modèle

Pour développer et entraîner le modèle, nous avons employé l'algorithme de machine learning Random Forest Regressor, qui a permis de diminuer l'erreur approximative d'environ 78.Le fonctionnement de ce algo est basé sur la création d'une multitude d'arbres de décision lors de la formation. Chaque arbre est construit à partir d'un sous-échantillon aléatoire des données, ce qui permet de capturer une variété de modèles potentiels. Lors de la prédiction, le modèle combine les résultats de tous les arbres pour produire une estimation plus précise et robuste. Cela réduit le risque de surapprentissage et améliore la généralisation du modèle aux nouvelles données. En utilisant cette approche, nous avons pu obtenir des prédictions fiables et cohérentes, même dans des contextes où les relations entre les variables sont complexes et non linéaires.Ensuite nous avons créer une application Flask pour tester efficacement le modèle

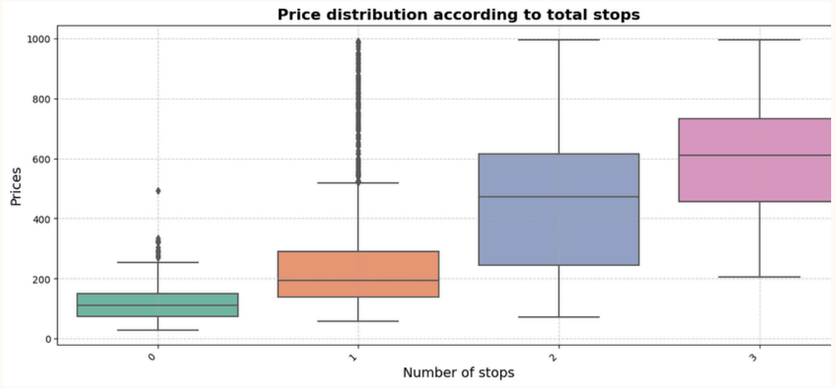
RESULTATS

Après analyse des données, il en ressort certaines conclusions :

- Distribution des prix

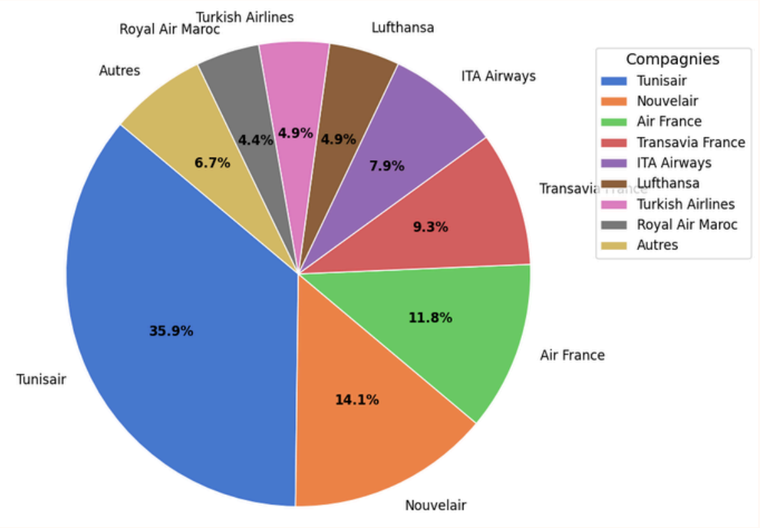


- Le prix est souvent proportionnel au nombre de stops sans doute en raison des ressources mis en oeuvre

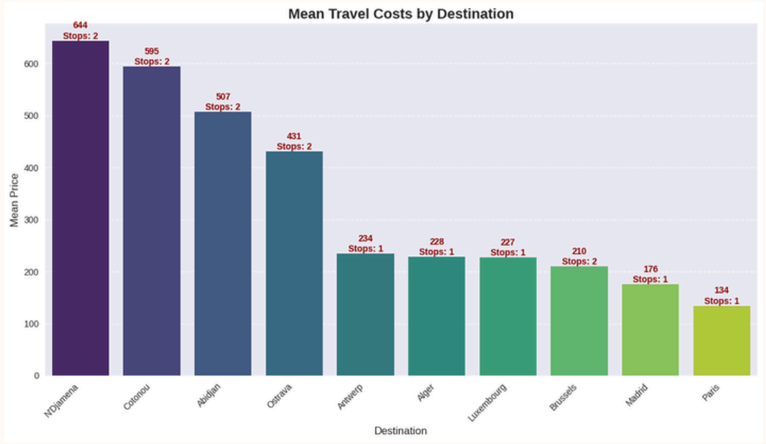


RESULTATS

- Tunisair,NouvelAir,FranceAir vont partir des compagnies les plus utilisés par les tunisiens lors des vols

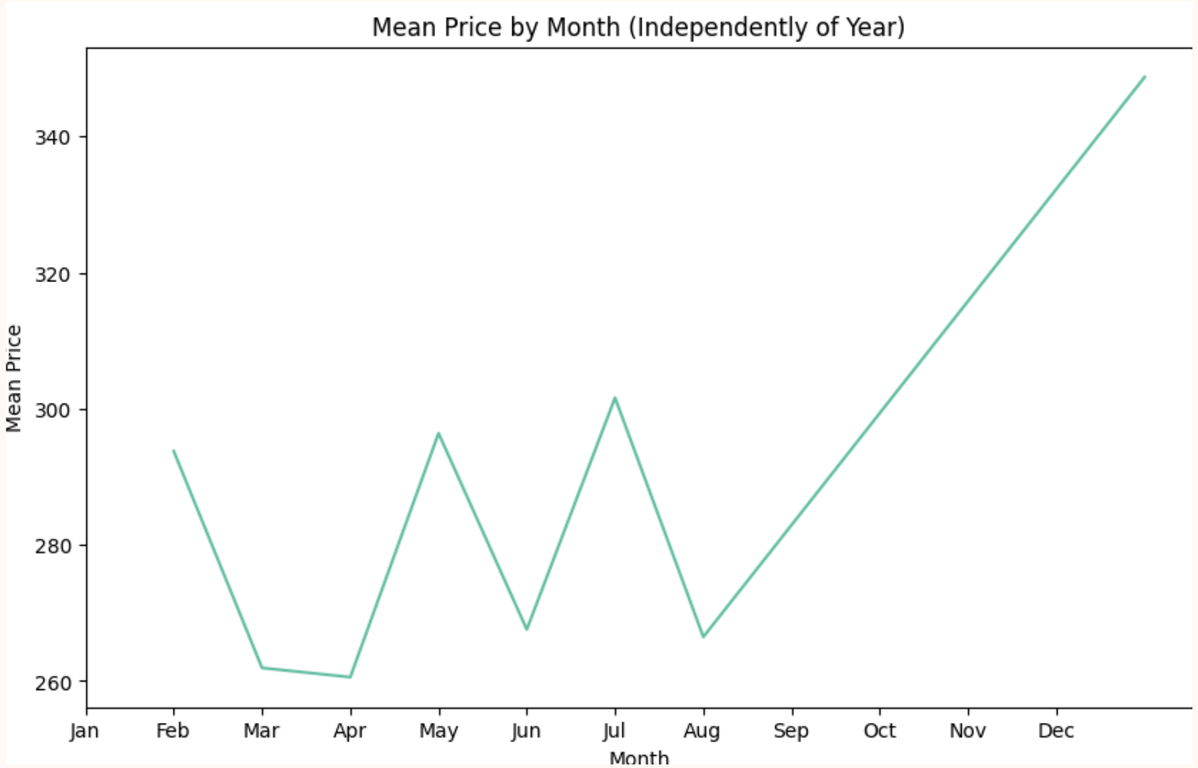


- Les vols vers l'Afrique subsaharienne ont des prix plus elevées comparé a ceux vers l'Europe. Ces vols pour la pluspart ont des escales.



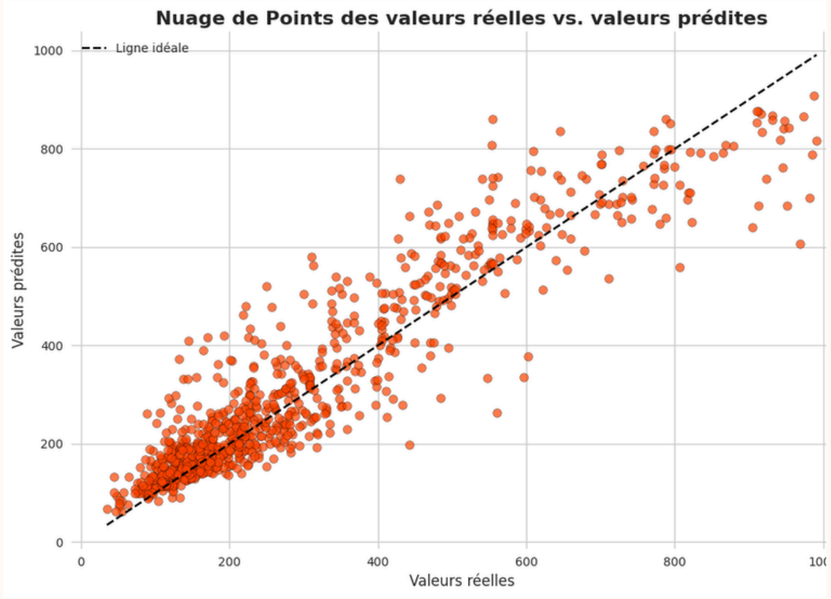
RESULTATS

- En Mai, Juilet et Aout il a aussi une augmentation du prix des billets d’avions en raison de la forte demande touristique pendant ces mois. Les vacances scolaires et le climat estival attirent un grand nombre de voyageurs désireux de profiter de destinations ensoleillées. Cela entraîne une hausse de la fréquentation des aéroports et une pression sur les compagnies aériennes qui ajustent leurs tarifs en conséquence. Pour les voyageurs, il peut être judicieux de réserver à l'avance pour bénéficier de prix plus avantageux ou de considérer des destinations moins prisées pour éviter cette hausse saisonnière des prix.

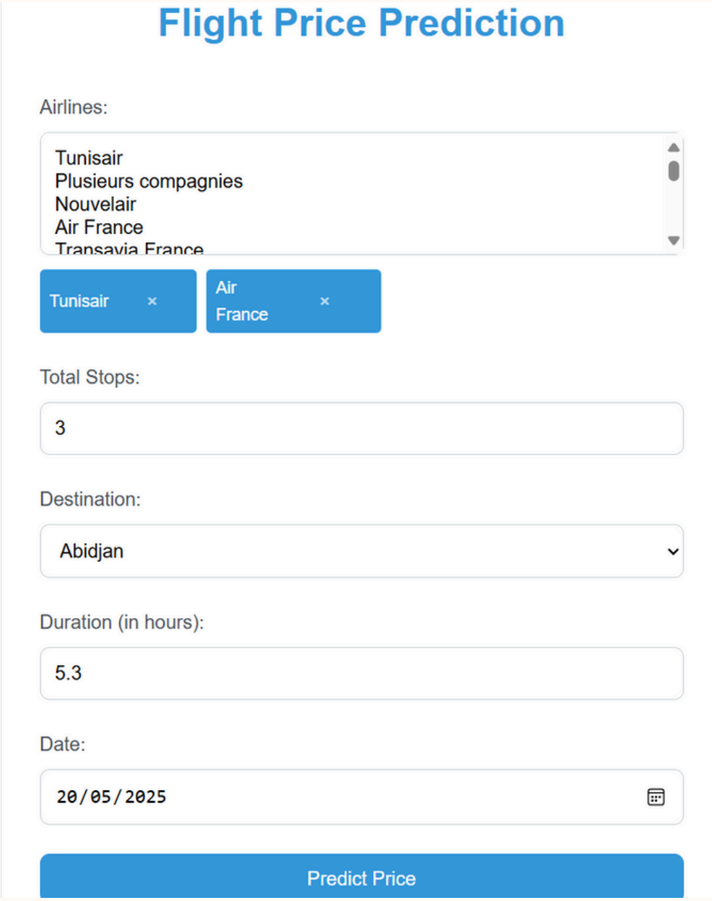


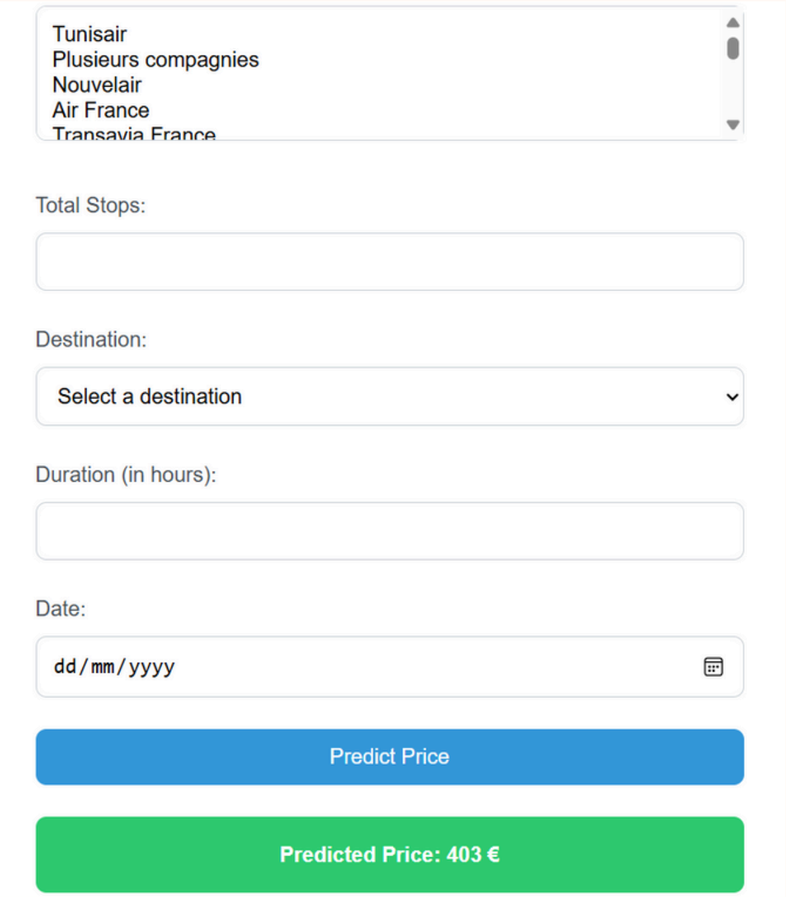
RESULTATS

Ainsi, nous avons élaborer le modèle qui nous permet d'avoir une marge d'erreur de 78 euros et une compréhension des données de 86 %.



RESULTATS


Flight Price Prediction
Airlines:
Tunisair
Plusieurs compagnies
Nouvelair
Air France
Transavia France
Tunisair ×
Air France ×
Total Stops:
3
Destination:
Abidjan
Duration (in hours):
5.3
Date:
20/05/2025
Predict Price


Tunisair
Plusieurs compagnies
Nouvelair
Air France
Transavia France
Total Stops:
Destination:
Select a destination
Duration (in hours):
Date:
dd/mm/yyyy
Predict Price
Predicted Price: 403 €

CONCLUSION

En conclusion, les tarifs des billets d'avion varient selon plusieurs paramètres comme l'agence, les escales et la durée des vols. L'analyse des prix en classe standard révèle des tendances sur les agences et le comportement des Tunisiens, mais les données sont limitées à cause des contraintes de scraping. Pour améliorer la précision des modèles, il est nécessaire d'utiliser un programme capable d'intégrer divers paramètres. Il serait également utile d'étudier l'impact des saisons touristiques et les politiques tarifaires des compagnies aériennes. Une collaboration avec des agences de voyage et une plateforme interactive pour les consommateurs pourraient enrichir l'étude et affiner les modèles prédictifs.

MERCI POUR VOTRE ÉCOUTE

Des questions ?