

# Documentation de l'Analyse des Données de Vols

Hamza Elhaiki

April 1, 2024

## 1 Introduction

Ce document présente la documentation de l'analyse des données de vols collectées à partir du scraping de sources en ligne. L'objectif de cette analyse est d'extraire des insights utiles à partir des données collectées.

## 2 Nettoyage et Prétraitement des Données

### 2.1 Conversion des Prix

Les prix des vols sont convertis en nombres flottants en supprimant les symboles non numériques, les espaces et les symboles de devise.

### 2.2 Traitement de la Durée

La durée des vols est convertie en minutes en extrayant les heures et les minutes de la chaîne de caractères.

### 2.3 Imputation des Valeurs Manquantes

Les valeurs manquantes dans les prix des vols sont imputées en utilisant la moyenne des prix pour chaque aéroport de destination.

## 3 Analyse Statistique

### 3.1 Statistiques par Destination

Les statistiques des prix et des durées de vol sont calculées pour chaque aéroport de destination, y compris le prix maximal, le prix minimal, le prix moyen, la durée maximale, la durée minimale, la durée médiane et l'écart type de la durée.

## **4 Clustering des Données**

### **4.1 K-means Clustering**

Les données sont regroupées en clusters à l'aide de l'algorithme K-means. Les caractéristiques sélectionnées pour le clustering sont la durée du vol et le prix du billet.

## **5 Visualisation**

Les clusters obtenus à partir du clustering K-means sont visualisés dans un graphique 2D, avec la durée du vol en minutes sur l'axe des x et le prix du billet en euros sur l'axe des y.

## **6 Conclusion**

Ce document fournit un aperçu du processus d'analyse des données de vols. Pour plus de détails sur les méthodes utilisées et les résultats obtenus, veuillez consulter le code source.