

# Analyse du Dataset : Global EV Sales (2010–2024)

Rapport Complet d'Analyse Exploratoire des Données



NOUHAILA EN-NAIDI 22006162 Groupe 2 Finance

Date : December 6, 2025

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Contexte du Projet . . . . .	2
1.2	Objectifs de l'Analyse . . . . .	2
<b>2</b>	<b>Le Dataset</b>	<b>3</b>
2.1	Source et Provenance . . . . .	3
2.2	Types de Problèmes Possibles . . . . .	3
2.2.1	Problème 1 : Régression . . . . .	3
2.2.2	Problème 2 : Séries Temporelles . . . . .	3
2.2.3	Problème 3 : Clustering . . . . .	3
2.3	Structure Générale . . . . .	4
2.4	Dictionnaire des Variables . . . . .	4
2.5	Glossaire . . . . .	4
<b>3</b>	<b>Chargement et Exploration Initiale</b>	<b>5</b>
3.1	Installation des Dépendances . . . . .	5
3.2	Importation des Bibliothèques . . . . .	5
3.3	Chargement du Dataset . . . . .	5
3.4	Exploration Initiale . . . . .	5
<b>4</b>	<b>Nettoyage des Données</b>	<b>7</b>
4.1	Analyse de la Qualité . . . . .	7
4.2	Code de Nettoyage . . . . .	7
<b>5</b>	<b>Analyse Exploratoire des Données</b>	<b>8</b>
5.1	Évolution Mondiale des Ventes . . . . .	8
<b>6</b>	<b>Conclusions</b>	<b>9</b>

# Chapter 1

## Introduction

### 1.1 Contexte du Projet

Ce projet analyse l'évolution mondiale des ventes de véhicules électriques (EV) entre 2010 et 2024. Les objectifs principaux sont :

- Étudier la croissance du marché mondial des véhicules électriques.
- Analyser les tendances par région et par pays.
- Comparer les types de motorisation (BEV vs PHEV).
- Comprendre les facteurs influençant l'adoption des EV.

### 1.2 Objectifs de l'Analyse

- **Analyse descriptive** : comprendre la structure et l'évolution du marché.
- **Analyse comparative** : identifier les régions leaders.
- **Préparation des données** : nettoyage, normalisation et structuration.
- **Insights stratégiques** basés sur les tendances.

# Chapter 2

## Le Dataset

### 2.1 Source et Provenance

Élément	Détail
Plateforme	Kaggle
Nom	Global EV Sales 2010–2024
Auteur	Patrick L. Ford
Lien	Kaggle Dataset
Période	2010–2024
Thématique	Mobilité électrique
Licence	Open Data

### 2.2 Types de Problèmes Possibles

#### 2.2.1 Problème 1 : Régression

- Type : régression supervisée
- Cible : `EV_Sales`
- Algorithmes : Régression Linéaire, Random Forest, XGBoost, Gradient Boosting

#### 2.2.2 Problème 2 : Séries Temporelles

- Prévision 2025–2030
- Méthodes : ARIMA, SARIMA, Prophet, LSTM

#### 2.2.3 Problème 3 : Clustering

- Segmentation des pays selon leurs patterns d'adoption
- Algorithmes : K-means, DBSCAN, Clustering hiérarchique

## 2.3 Structure Générale

Élément	Valeur estimée
Nombre de lignes	200–500
Nombre de colonnes	6–8
Taille mémoire	< 1 MB
Période	2010–2024
Fréquence	Annuelle
Granularité	Pays/Année/Type
Unité	Véhicules vendus

## 2.4 Dictionnaire des Variables

Variable	Rôle	Type	Description	Exemple	Manquant
year	Feature	Numérique	Année	2010	Non
country	Feature	Catég.	Pays	China	Non
region	Feature	Catég.	Région	Europe	Possible
ev_sales	Target	Numérique	Ventes EV	1 500 000	Possible
type	Feature	Catég.	Motorisation	BEV/PHEV	Non
market_share	Feature	Numérique	Part de marché	0.18	Possible
ev_stock	Feature	Numérique	Parc EV	10 000 000	Possible

## 2.5 Glossaire

- **BEV** : Battery Electric Vehicle (100% électrique)
- **PHEV** : Plug-in Hybrid Electric Vehicle (hybride rechargeable)

# Chapter 3

## Chargement et Exploration Initiale

### 3.1 Installation des Dépendances

```
!pip install kagglehub[pandas-datasets]
!pip install pandas numpy matplotlib seaborn scikit-learn
```

### 3.2 Importation des Bibliothèques

```
import kagglehub
from kagglehub import KaggleDatasetAdapter
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
```

### 3.3 Chargement du Dataset

```
df = kagglehub.load_dataset(
    KaggleDatasetAdapter.PANDAS,
    "patricklford/global-ev-sales-2010-2024",
    ""
)
df.head(10)
```

### 3.4 Exploration Initiale

- Dimensions du dataset

- Types de données
- Valeurs manquantes
- Statistiques descriptives

# Chapter 4

## Nettoyage des Données

### 4.1 Analyse de la Qualité

- Détection des valeurs manquantes
- Doublons
- Valeurs aberrantes
- Cohérence des années

### 4.2 Code de Nettoyage

```
df_clean = df.copy()
df_clean = df_clean.drop_duplicates()
df_clean.columns = df_clean.columns.str.lower().str.replace(" ", "_")
df_clean = df_clean.fillna(df_clean.median(numeric_only=True))
```

# Chapter 5

## Analyse Exploratoire des Données

### 5.1 Évolution Mondiale des Ventes

```
yearly_sales = df_clean.groupby("year")["ev_sales"].sum()
plt.plot(yearly_sales)
plt.title("Evolution des ventes EV (2010-2024)")
plt.show()
```

# Chapter 6

## Conclusions

- Forte croissance du marché des véhicules électriques.
- La Chine domine largement le marché mondial.
- Les BEV surpassent progressivement les PHEV.
- Base solide pour des modèles ML/Time Series.