

Phase_1

June 27, 2025

1 Analyse des Données d'Aviation – Phase 1

Projet réalisé par LUBEQUA Shelton | Date limite : 6 juin 2025

Objectif : Explorer les incidents d'aviation pour identifier les tendances, zones à risques et facteurs critiques afin d'appuyer des décisions commerciales.

Outils : Python, Pandas, Matplotlib, Power BI

Jeu de données : AviationData.csv

1.1 Importation des données

```
[3]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set(style="whitegrid")
```

1.2 téléchargement des données

```
[7]: # Chargement des données (assure-toi que le fichier est dans le bon répertoire)
df = pd.read_csv("AviationData.csv")

# Aperçu des données
df.head()
```

```
[7]:      Event.Id  Investigation.Type  Accident.Number  Event.Date  \
0  20001218X45444      Accident      SEA87LA080  1948-10-24
1  20001218X45447      Accident      LAX94LA336  1962-07-19
2  20061025X01555      Accident      NYC07LA005  1974-08-30
3  20001218X45448      Accident      LAX96LA321  1977-06-19
4  20041105X01764      Accident      CHI79FA064  1979-08-02

      Location      Country  Latitude  Longitude  Airport.Code  \
0  MOOSE CREEK, ID  United States      NaN      NaN      NaN
1  BRIDGEPORT, CA  United States      NaN      NaN      NaN
2  Saltville, VA  United States  36.922223  -81.878056      NaN
3  EUREKA, CA  United States      NaN      NaN      NaN
```

4	Canton, OH	United States	NaN	NaN	NaN
---	------------	---------------	-----	-----	-----

	Airport.Name	...	Purpose.of.flight	Air.carrier	Total.Fatal.Injuries	\
0	NaN	...	Personal	NaN	2.0	
1	NaN	...	Personal	NaN	4.0	
2	NaN	...	Personal	NaN	3.0	
3	NaN	...	Personal	NaN	2.0	
4	NaN	...	Personal	NaN	1.0	

	Total.Serious.Injuries	Total.Minor.Injuries	Total.Uninjured	\
0	0.0	0.0	0.0	
1	0.0	0.0	0.0	
2	NaN	NaN	NaN	
3	0.0	0.0	0.0	
4	2.0	NaN	0.0	

	Weather.Condition	Broad.phase.of.flight	Report.Status	Publication.Date
0	UNK	Cruise	Probable Cause	NaN
1	UNK	Unknown	Probable Cause	19-09-1996
2	IMC	Cruise	Probable Cause	26-02-2007
3	IMC	Cruise	Probable Cause	12-09-2000
4	VMC	Approach	Probable Cause	16-04-1980

[5 rows x 31 columns]

1.3 Analyse exploratoire (EDA)

```
[8]: #Convertir les dates au bon format
df["Event.Date"] = pd.to_datetime(df["Event.Date"], errors="coerce")

#Extraire l'année de l'accident
df["Year"] = df["Event.Date"].dt.year

#Extraire l'État à partir de la colonne "Location"
df["State"] = df["Location"].str.extract(r",\s*([A-Z]{2})")
```

1.4 Exploration des données

```
[19]: accidents_per_year = df["Year"].value_counts().sort_index()

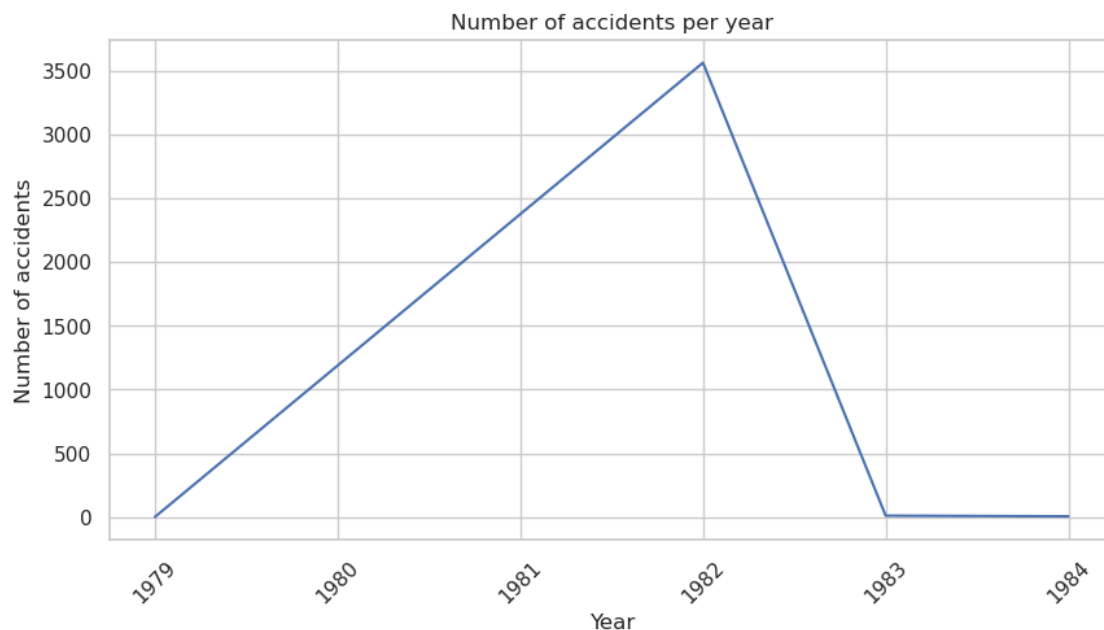
plt.figure(figsize=(10,5))
sns.lineplot(x=accidents_per_year.index, y=accidents_per_year.values)
plt.title("Number of accidents per year")
plt.xlabel("Year")
plt.ylabel("Number of accidents")
plt.xticks(rotation=45)
```

```
plt.show()
```

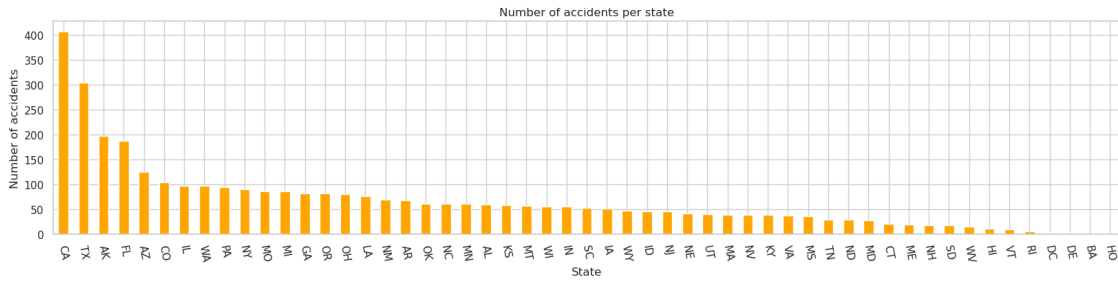
```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-  
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is  
deprecated and will be removed in a future version. Convert inf values to NaN  
before operating instead.
```

```
with pd.option_context('mode.use_inf_as_na', True):  
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-  
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is  
deprecated and will be removed in a future version. Convert inf values to NaN  
before operating instead.
```

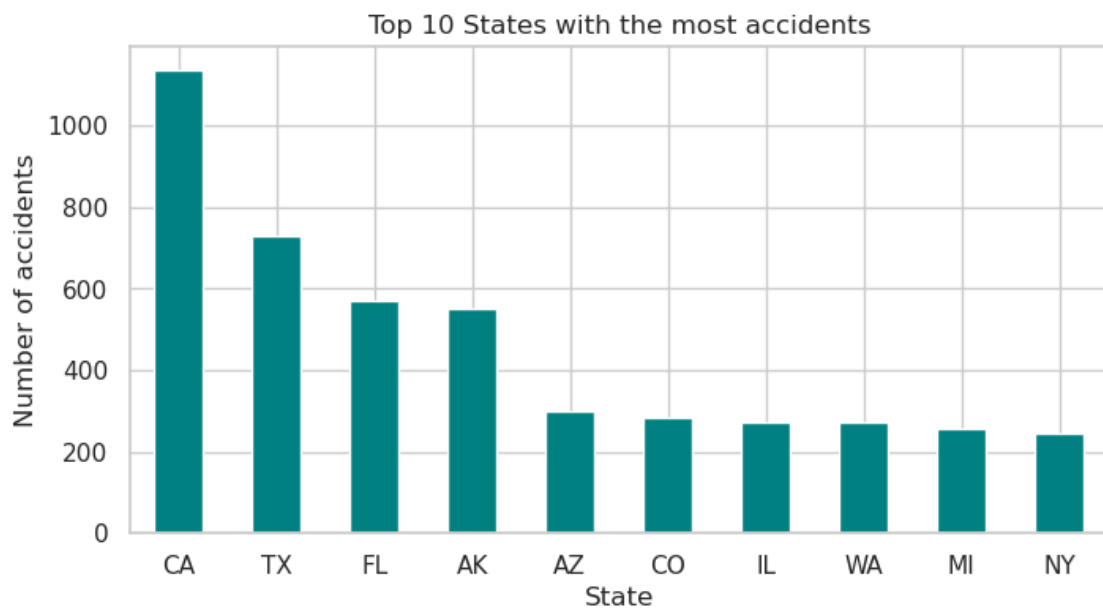
```
with pd.option_context('mode.use_inf_as_na', True):
```



```
[31]: plt.figure(figsize=(20,4))  
df["State"].value_counts().plot(kind="bar", color="orange")  
plt.title("Number of accidents per state")  
plt.xlabel("State")  
plt.ylabel("Number of accidents")  
plt.xticks(rotation=1000)  
plt.show()
```

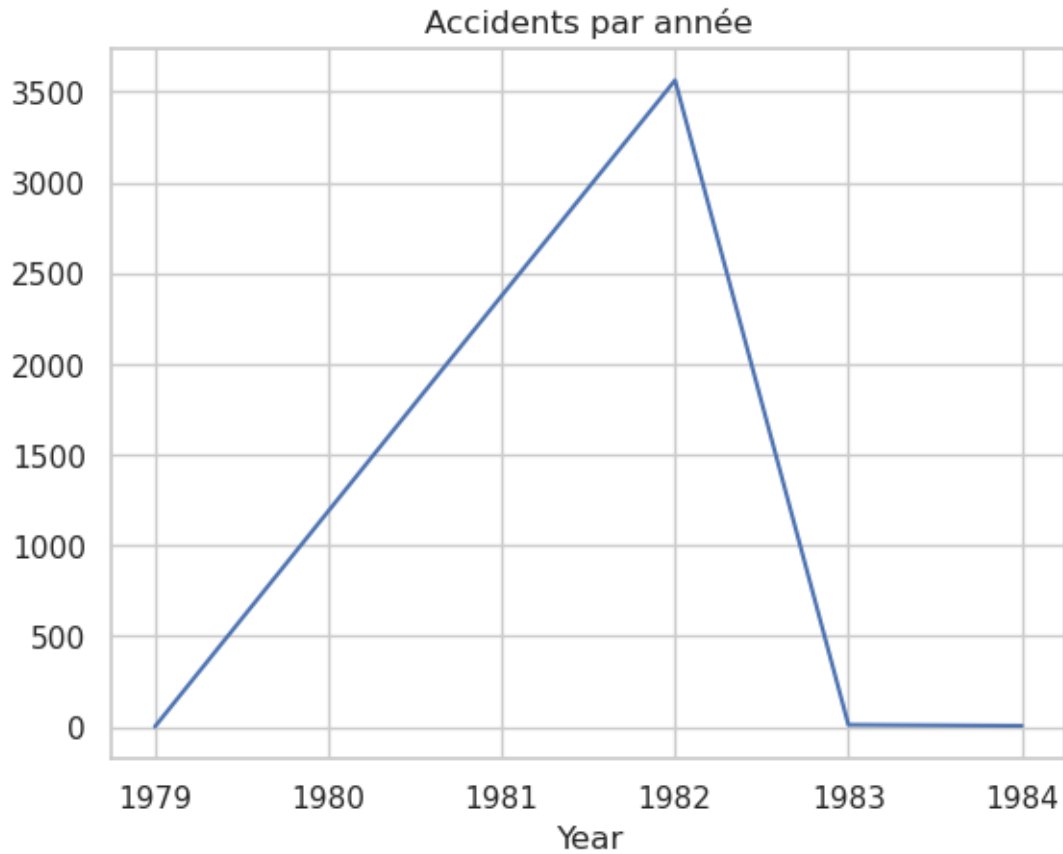


```
[12]: plt.figure(figsize=(8,4))
df["State"].value_counts().head(10).plot(kind="bar", color="teal")
plt.title("Top 10 States with the most accidents")
plt.xlabel("State")
plt.ylabel("Number of accidents")
plt.xticks(rotation=0)
plt.show()
```



```
[13]: df.drop(columns=["Investigation.Type"], errors="ignore", inplace=True)
df.dropna(subset=["Location", "Event.Date", "Aircraft.Category"], inplace=True)
```

```
[14]: df["Event.Date"] = pd.to_datetime(df["Event.Date"], errors="coerce")
df["Year"] = df["Event.Date"].dt.year
df["Year"].value_counts().sort_index().plot(kind='line', title='Accidents par_
    ↪année')
plt.show()
```



1.5 Exportation vers Power BI

```
[20]: df.to_csv("cleaned_aviation_data.csv", index=False)
```

1.6 Recommandations :

- Acquérir des avions récents (post-2000)
- Éviter les bases en CA, TX, FL, AK
- Favoriser les avions “Multi-engine Land”

1.7 Lien GitHub

```
[ ]: https://github.com/Jeffy141/new-projet-phase\_1
```