Analysis of employees retirement age of company

Introduction

This project uses the Python `Polars` package to analyze the quantitative data of company A's former employees, providing a comprehensive analysis of the average age, standard deviation, and other metrics.

The data was sourced from the following GitHub repository: https://github.com/nogibjj/Mini_PJT_3_Polar_ISL/blob/main/HR.csv"

  – From the available variables, I specifically focused on the "Age" at retirement.
  – Calcuated mean, median, standard Deviation and so on

Descriptive Statistics

In [10]:
```python
import polars as pl
import matplotlib.pyplot as plt

# Load dataset using Polars
ppl = pl.read_csv("HR.csv")

# header
ppl.head()
print(ppl.head())
```

shape: (5, 35)

Age	Attrition	BusinessTravel	DailyRate	…	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
i64	str	str	i64		i64	i64	i64	i64
41	Yes	Travel_Rarely	1102	…	6	4	0	5
49	No	Travel_Frequently	279	…	10	7	1	7
37	Yes	Travel_Rarely	1373	…	0	0	0	0
33	No	Travel_Frequently	1392	…	8	7	3	0
27	No	Travel_Rarely	591	…	2	2	2	2

In [11]:
```python
# Summary stat
full_desc = ppl.describe()
print(full_desc)
```

shape: (9, 36)

statistic	Age	Attrition	BusinessTravel	…	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
str	f64	str	str		f64	f64	f64	f64
count	1470.0	1470	1470	…	1470.0	1470.0	1470.0	1470.0
null_count	0.0	0	0	…	0.0	0.0	0.0	0.0
mean	36.92381	null	null	…	7.008163	4.229252	2.187755	4.123129
std	9.135373	null	null	…	6.126525	3.623137	3.22243	3.568136
min	18.0	No	Non-Travel	…	0.0	0.0	0.0	0.0
25%	30.0	null	null	…	3.0	2.0	0.0	2.0
50%	36.0	null	null	…	5.0	3.0	1.0	3.0
75%	43.0	null	null	…	9.0	7.0	3.0	7.0
max	60.0	Yes	Travel_Rarely	…	40.0	18.0	15.0	17.0

Data Visialization

In [2]:
```python
import polars as pl
import matplotlib.pyplot as plt

# Load dataset using Polars
ppl = pl.read_csv("HR.csv")

# Calculate statistics for Age
full_desc = ppl.describe()
age_mean = ppl.select(pl.col("Age").mean()).item()
age_median = ppl.select(pl.col("Age").median()).item()
age_std = ppl.select(pl.col("Age").std()).item()

# Convert Polars dataframe to a list for plotting
ages = ppl.select("Age").to_series().to_list()

# Plot histogram for Age
plt.figure(figsize=(8, 6))
plt.hist(ages, bins=10, edgecolor="black")
plt.title("Average retirement age of company A")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
```