**Exploratory analysis of Aviation stack dataset**

**Introduction:**

This aviation stack dataset has the details of the future flights that are scheduled and related information like flight type, airline company, age of flight, and other details.

**Data Inspection:**

Dataset size: (102 rows, 15 columns)

Columns :15

Column names: Index(['id', 'fleet\_average\_age', 'airline\_id', 'callsign', 'hub\_code',

'iata\_code', 'icao\_code', 'country\_iso2', 'date\_founded',

'iata\_prefix\_accounting', 'airline\_name', 'country\_name', 'fleet\_size',

'status', 'type'],

dtype='object')

The datatypes of the columns :

Data types:

id object

fleet\_average\_age object

airline\_id float64

callsign object

hub\_code object

iata\_code object

icao\_code object

country\_iso2 object

date\_founded float64

iata\_prefix\_accounting float64

airline\_name object

country\_name object

fleet\_size object

status object

type object

dtype: object

**Missing values:**

id 1

fleet\_average\_age 1

airline\_id 1

callsign 1

hub\_code 2

iata\_code 2

icao\_code 2

country\_iso2 2

date\_founded 5

iata\_prefix\_accounting 5

airline\_name 2

country\_name 3

fleet\_size 3

status 3

type 4

**Summary statistics analysis:**

Summary statistics for 'fleet\_average\_age':

count 101

unique 68

top 5.2

freq 4

**Incorrect values:**

Status distribution:

status

active 95

122 1

historical/administration 1

65 1

58 1

Needs some filtering to these data to make it stable.

**Descriptive Statistics:**

Yearly founded counts:

year\_founded

1907.0 1

1919.0 1

1920.0 1

1923.0 1

1927.0 1

1928.0 1

1931.0 1

1932.0 1

1933.0 3

1934.0 1

1941.0 1

1945.0 3

1946.0 2

1947.0 1

1948.0 1

1949.0 1

1951.0 2

1952.0 1

1957.0 2

1959.0 2

1962.0 2

1965.0 1

1967.0 1

1971.0 2

1972.0 3

**Data visualization:**

Fleet Average Age Distribution:

id fleet\_average\_age airline\_id status type year\_founded

0 13133 10.9 1.0 ... active scheduled 1934.0

1 13134 17.0 2.0 ... active scheduled 1928.0

2 13135 13.8 3.0 ... active scheduled 1931.0

3 13136 12.3 4.0 ... active scheduled 1967.0

4 13137 6.9 5.0 ... active scheduled 2004.0

.. ... ... ... ... ... ... ...

97 13228 4.7 96.0 ... active scheduled 1941.0

98 13229 3.5 97.0 ... active scheduled 2005.0

99 13230 7.7 98.0 ... 58 active 1994.0

100 13231 5.0 99.0 ... active scheduled 1996.0

101 13232 3.7 100.0 ... active scheduled 1993.0

Some incorrect values needs to be cleaned.

A graph of a number of blue bars

Description automatically generated

**Conclusion:**

To sum up, the dataset's exploratory analysis has given rise to a fundamental comprehension of the main characteristics, distributions, and connections within the data.

This dataset which has future scheduled data of flights can be utilized to predict the delay of flights compared with past data records.