Statistical & Machine Learning Analysis of Aircraft Survival Across Global Airlines

Kaitao Liao

My Motivation

Airlines face billions in aircraft costs — buying, maintaining, and retiring aircraft is one of their biggest financial decisions.

Maintenance & Safety — depend on knowing when aircraft are near the end of service life.

Fuel efficiency & Sustainability — push airlines to retire older, less efficient planes sooner.

Predictive insights from statistics & machine learning — help airlines, investors, and regulators make smarter decisions.

Research Question

Main Question:

What statistical and machine learning patterns explain differences in aircraft survival and performance across global airlines?

Sub-Questions:

1. **Aircraft Age** – How does age affect the probability of retirement across airlines?
2. **Aircraft Models** – Do certain models (e.g., A320, B737) remain active longer than others?
3. **Airline Practices** – Are there statistical differences in aircraft retirement strategies among airlines(e.g., legacy carriers vs low-cost carriers) ?
4. **Prediction** – Can machine learning methods accurately classify whether an aircraft is Active or Retired by using airline, age, and model as predictors?

Dataset

Source: aircraft_data.csv

Variables used:

- **First Flight** → converted into aircraft Age (years)
- **Airline**
- **Model** (e.g., A320, B737, A380)
- **Status** (Active / Retired / Stored)

Data preparation:

- Removed missing/invalid records
- Filtered ages between 0–80 years
- Target variable:
  - Active = 1
  - Non-active = 0

	model	msn	ln	ac_type	airline	first_flight	registration
1	Airbus A220	50001	NA	CS100	Bombardier	16.09.2013	C-FBCS
2	Airbus A220	50002	NA	CS100	Bombardier	03.01.2014	C-GWYD
5	Airbus A220	50005	NA	CS100	Bombardier	04.07.2013	C-GWXZ
6	Airbus A220	50006	NA	CS100	Bombardier	30.10.2015	C-FFCO
7	Airbus A220	50010	NA	CS100	Swiss International Air Lines	16.06.2016	HB-JBA
8	Airbus A220	50011	NA	CS100	Swiss International Air Lines	22.07.2016	HB-JBB
9	Airbus A220	50012	NA	CS100	Swiss International Air Lines	24.09.2016	HB-JBC
10	Airbus A220	50013	NA	CS100	Swiss International Air Lines	06.11.2016	HB-JBD
11	Airbus A220	50014	NA	CS100	Swiss International Air Lines	06.12.2016	HB-JBE
12	Airbus A220	50015	NA	CS100	Swiss International Air Lines	09.02.2017	HB-JBF
13	Airbus A220	50016	NA	CS100	Swiss International Air Lines	26.02.2017	HB-JBG

msn	ln	ac_type	airline	first_flight	registration	status
55115	NA	220-371	JetBlue Airways	15.07.2021	N3062J	Active
55116	NA	220-371	Air Austral	09.04.2021	F-OMER	Active
55117	NA	220-371	Air Canada	maj.21	C-GTZS	Active
55118	NA	220-371	Air Canada	10.05.2021	C-GTZU	Active
55119	NA	220-371	Air Canada	18.05.2021	C-GUAC	Active
55120	NA	220-371	JetBlue Airways	04.08.2021	N3065J	Active
55121	NA	220-371	Air Baltic	10.06.2021	YL-ABA	Active
55124	NA	220-371	JetBlue Airways	17.09.2021	N3077J	Active
55125	NA	220-371	Air Austral	27.06.2021	F-OTER	Active
55126	NA	220-371	Swiss International Air Linesoperated byAir Baltic	16.07.2021	YL-ABB	Active
55127	NA	220-371	Air Baltic	29.07.2021	YL-ABC	Active

Sub-Question 1

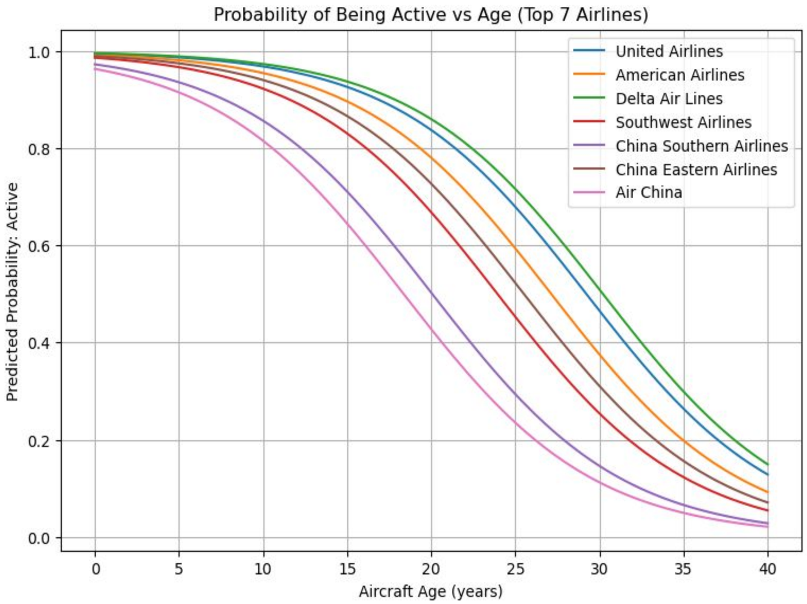
- **Objective**: Estimate how aircraft age influences the probability of being active across different airlines.

- **Method**: **Logistic regression** with predictors: **age and airline.**

- **Findings**:

1. Probability of being active decreases with increasing aircraft age.

2. Clear differences between airlines in how long aircraft tend to remain active.



Sub-Question 2

Objective: To test whether different aircraft models (Airbus vs Boeing families) show different active lifespans depending on design era and usage.

Method: **Kaplan–Meier survival analysis** with **log-rank tests** comparing model pairs.

Pairs chosen based on **design role & era**:

- **A320 vs B737** → Narrowbody aircraft.
- **A330 vs B777** → 1990s widebodies aircraft.
- **A380 vs B747** → Very large, 4-engine aircraft.
- **A350 vs B787** → New-generation widebodies.

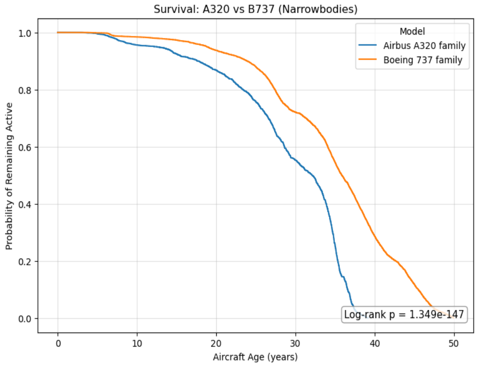
Findings & Factors:

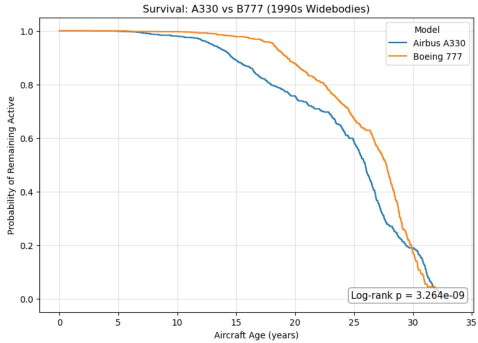
1) **A320 vs B737:**
   a) Both dominate the narrowbody market, but B737 has longer survival.
   b) Factors: introduced earlier (1960s), multiple redesigns (737MAX), very high global demand.
   c) A320 still strong but with slightly younger retirement ages in dataset.

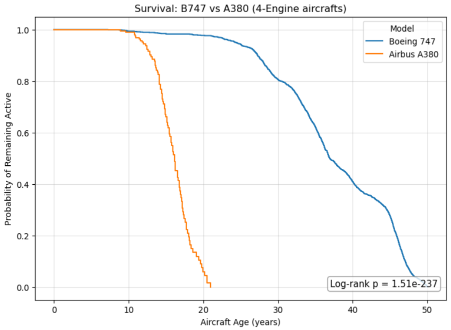
2) **A330 vs B777:**
   a) Lifespans broadly similar since both entered in the 1990s as efficient twin-engine widebodies.
   b) The 777 is more popular for serving long-haul missions (often intercontinental) and has faced slower replacement.

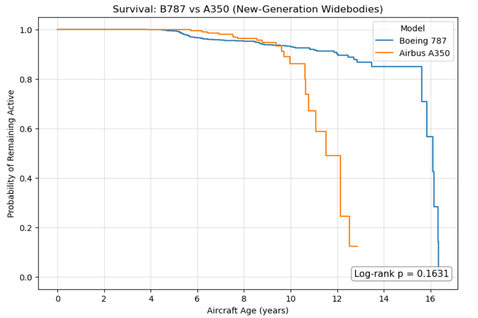
3) **B747 vs A380:**
   a) B747 continues in service longer, especially as freighters.
   b) A380 retires earlier due to high fuel costs, limited airport compatibility, and declining demand post-2010.
   c) Example: many airlines retired A380s during COVID-19, while 747 freighters stayed essential.

4) **B787 vs A350:**
   a) Both are next-generation fuel-efficient designs using composites.
   b) Still relatively young in service so survival curves overlap, no significant difference yet.
   c) Too early to judge long-term differences until more retirements occur.









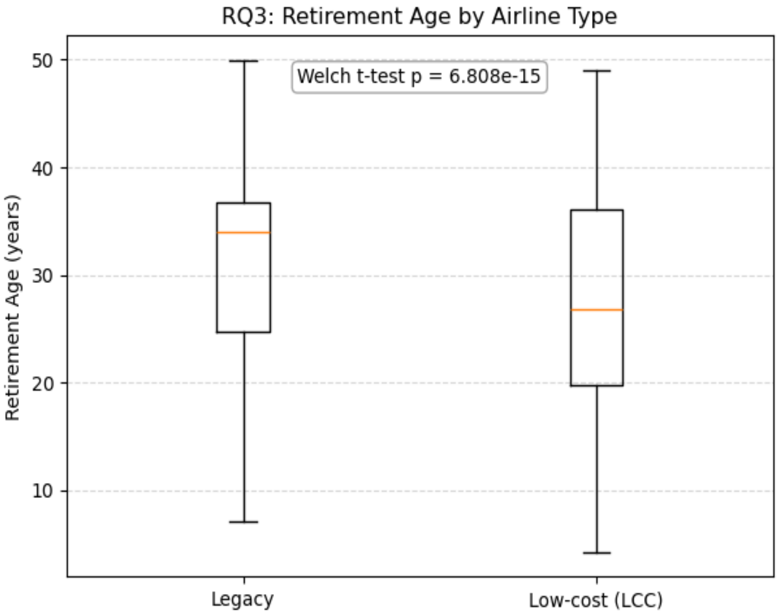
Sub-Question 3

Objective: Test whether legacy airlines retire aircraft later than low-cost carriers (LCCs).

Method: Statistical testing with **Welch's t-test** and **Mann–Whitney U** on retirement ages.

Findings:

1. Legacy carriers retire aircraft at significantly older ages compared to LCCs (supported by extremely small p-values).
2. LCCs show lower median retirement ages, reflecting their tendency to keep younger, more fuel-efficient aircraft.
3. LCC retirement ages are more spread out (some very early retirements), while legacy carriers are more consistent.



Sub-Question 4

Objective: Predict whether an aircraft is **Active** or **Retired** using features such as **age, airline, and model**.

Method: Random Forest classification with feature importance analysis and **ROC evaluation**.

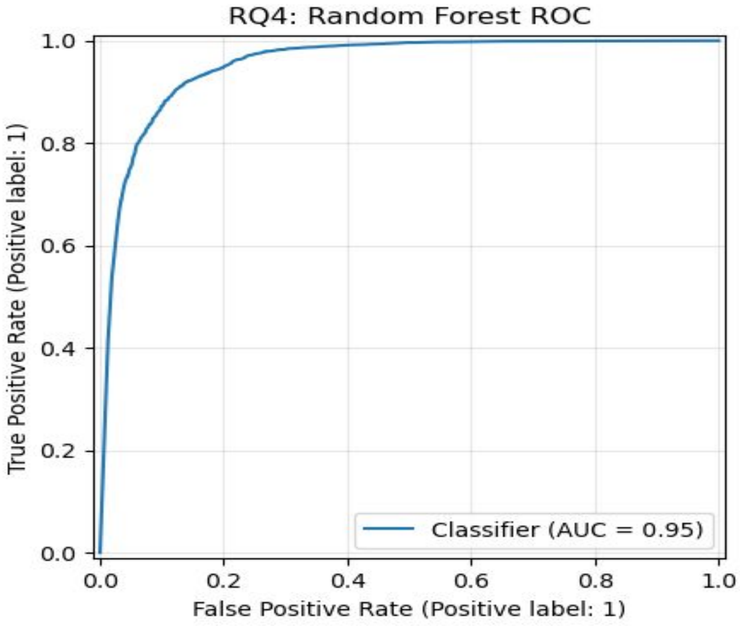
Findings:

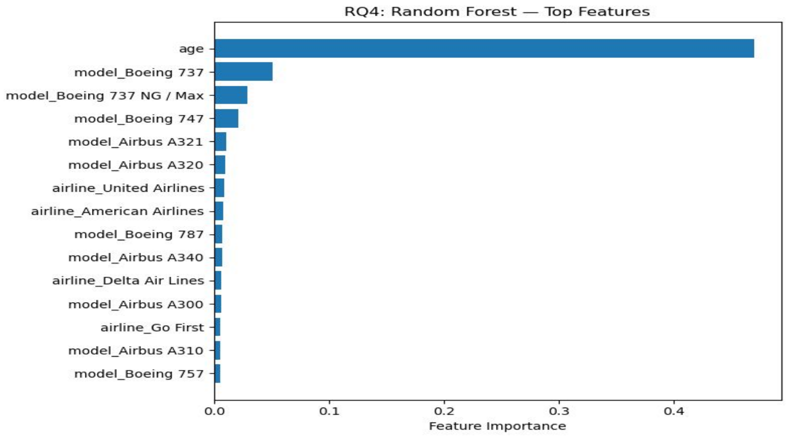
1. The model achieves **high predictive performance** (AUC ≈ 0.95), showing strong ability to distinguish active vs retired aircraft.

2. **Aircraft age** is by far the most important predictor, dominating other features.

3. Certain models (e.g., **Boeing 737 family, Boeing 747**) and airlines also contribute, but to a much smaller degree.

4. Indicates that **retirement is primarily age-driven**, while airline and model exert secondary effects.





Key Takeaway & Conclusion

1. The study shows that aircraft survival is primarily **age-driven**, with older planes facing higher probabilities of retirement after about 25–30 years.

2. **Model design** also plays a role, as newer-generation aircraft (e.g., Boeing 787, Airbus A350) show different survival curves compared to older widebodies (e.g., Boeing 777, Airbus A330).

3. **Airline strategies**: legacy carriers often extend aircraft lifespans, while low-cost carriers retire aircraft earlier to prioritize efficiency.

4. Integrating **survival analysis** and **machine learning** not only explains why aircraft retire but also allows reliable prediction, helping airlines plan fleets more efficiently.