DV AIRFLY INSIGHTS

**About the dataset:**

* The dataset has **484,551 rows** and **29 columns**.
* There are **null values**: (Dest\_Airport: 1,479 missing),(Org\_Airport: 1,177 missing)
* There are 2 duplicate rows

**KPI’s**

 % Missing values reduced

 Duplicate records removed

 % of dataset retained after cleaning

 Preprocessing time reduced

**Cleaning Process**

1. Import libraries
2. Load dataset
3. Check summary , datatypes , shape
4. Check for null values and replaced with mode value
5. Checked for duplicates and removed them
6. Converting Date format
7. Creating day , month , year , route columns

**Insights**

1. Missing values in Org\_Airport and Dest\_Airport were filled with mode → No nulls left in these key categorical columns.
2. Duplicate rows were detected and removed → dataset integrity improved.
3. New columns created : day,month,hour,route