WORKING WITH DATASET

Understanding The Dataset

Date	10-11-2022
Team ID	PNT2022TMID29101
Project Name	Project – AIRLINES DATA ANALYTICS FOR AVAITION INDUSTRY
Maximum Marks	2 Marks

The main dataset has been sourced from Kaggle. Additional data with Aiport metadata has also been fetched from Partow (The Global Airport Database). Only relevant information will be provided for the latter due to low importance in the report.

The main dataset i.e. Airline Cancellation/Delay was fetched in the form of multiple CSV files each representing data for each year. Overall, it was nearly 7 GB in size with nearly 68 million rows, comprising of the following fields/columns:

Name	Description	Type(Format)	Example
FL_DATE	Date of the flight	DATE (yy/mm/dd)	2009-05-02
OP_CARRIER	Airline Identifier	STRING	9E
OP_CARRIER_FL_NUM	Flight Number	INTEGER	2216
ORIGIN	Starting Airport Code (IATA Code)	STRING	MLI
DEST	Destination Airport Code (IATA Code)	STRING	MEM
CRS_DEP_TIME	Planned Departure Time	INTEGER	600
DEP_TIME	Actual Departure Time	FLOAT	603.0
DEP_DELAY	Total Delay on Departure in minutes	FLOAT	3.0
TAXI_OUT	The time duration elapsed between departure from the origin airport gate and wheels off	FLOAT	14.0
WHEELS_OFF	The time point that the aircraft's wheels leave the ground	FLOAT	617.0
WHEELS_ON	The time point that the aircraft's wheels touch on the ground	FLOAT	757.0
TAXI_IN	The time duration elapsed between wheels-on and gate arrival at the destination airport	FLOAT	8.0
CRS_ARR_TIME	Planned arrival time	INTEGER	732
ARR_TIME	Actual Arrival Time	FLOAT	805.0
ARR_DELAY	Total Delay on Arrival in minutes	FLOAT	33.0
CANCELLED	Flight Cancelled (1 = cancelled)	FLOAT	0.0
CANCELLATION_CODE	Reason for Cancellation of flight: A - Airline/Carrier; B - Weather; C - National Air System; D - Security	STRING	D
DIVERTED	Aircraft landed on airport that out of schedule	FLOAT	0.0
CRS_ELAPSED_TIME	Planned time amount needed for the flight trip	FLOAT	92.0
ACTUAL_ELAPSED_TIME	AIR_TIME+TAXI_IN+TAXI_OUT	FLOAT	122.0
AIR_TIME	The time duration between wheels_off and wheels_on time	FLOAT	100.0
DISTANCE	Distance between two airports	FLOAT	442.0
CARRIER_DELAY	Delay caused by the airline in minutes	FLOAT	0.0
WEATHER_DELAY	Delay caused by weather	FLOAT	0.0
NAS_DELAY	Delay caused by air system	FLOAT	33.0
SECURITY_DELAY	Delay caused by security	FLOAT	0.0
LATE_AIRCRAFT_DELAY	Delay caused by aircraft reaching late	STRING	0.0