# INFO 7250 Project Process Record

## 211121

Data Fields Explanation

FL\_DATE:

date of the flight

OP\_CARRIER / OP\_UNIQUE\_CARRIER:

flight operation company

OP\_CARRIER\_FL\_NUM:

flight number, maybe remains the same if it’s a round trip in the same day

ORIGIN:

origin of the flight

DEST:

destination of the flight

CRS\_DEP\_TIME:

computer reservation system departure time

only 2019 doesn’t have this column

DEP\_TIME:

actual departure time

different meaning for 2020 (?)

**DEP\_DELAY:**

DEP\_TIME – CRS\_DEP\_TIME

Actual departure time – computed departure time

TAXI\_OUT:

The Taxi Out phase includes a lot of things happening in the cockpit and the cabin. The flight attendants are making sure the cabin is ready for takeoff and conducting the emergency brief. The pilots are starting the engines and requesting permission to taxi.

Boarding to WHEELS\_OFF

WHEELS\_OFF:

To depart or travel away (from someone or something) on wheels or on a wheeled vehicle or apparatus.

Time that wheel off the runaway (departure)

WHEELS\_ON:

Time that wheel on the runaway (arrival)

TAXI\_IN:

WHEELS\_ON to getting to the gate

CRS\_ARR\_TIME:

computer reservation system arrival time

ARR\_TIME:

Actual arrival time

**ARR\_DELAY:**

Arrival delay

CANCELLED,

CANCELLATION\_CODE,

DIVERTED,CRS\_ELAPSED\_TIME,

ACTUAL\_ELAPSED\_TIME,

AIR\_TIME,DISTANCE,

CARRIER\_DELAY,

WEATHER\_DELAY,

NAS\_DELAY,

SECURITY\_DELAY,

LATE\_AIRCRAFT\_DELAY,

Unnamed: 27

How to perform data cleaning

Just dump the null value

For value in different format, CRS\_DEP\_TIME: 2019.csn doesn’t have this, skip 2019.csv

I may discard 2019.csv and 2020.csv from the data analysis

For missing value in dep\_delay, I use 0.0 as default value.

MapReduce Task Explanation:

ADA1:

Average departure delay time based on carriers

ADA2:

Sum departure delay time based on carriers

ADA3:

Max departure delay time based on carriers

ADA4:

Min departure delay time based on carriers

ADA5:

Average departure delay time based on origin

ADA6:

Sum departure delay time based on origin

ADA7:

Max departure delay time based on origin

ADA8:

Min departure delay time based on origin

ADA9:

Average arrival delay time based on carriers

ADA10:

Sum arrival delay time based on carriers

ADA11:

Max arrival delay time based on carriers

ADA12:

Min arrival delay time based on carriers

ADA13:

Average arrival delay time based on origin

ADA14:

Sum arrival delay time based on origin

ADA15:

Max arrival delay time based on origin

ADA16:

Min arrival delay time based on origin

ADA17:

Average departure delay time based on carriers and flight route(origin to destination)

ADA18:

Average arrival delay time based on carriers and flight route(origin to destination)

ADA19:

Max departure delay time based on carriers and flight route(origin to destination)

ADA20:

Max arrival delay time based on carriers and flight route(origin to destination)

ADA21:

min departure delay time based on carriers and flight route(origin to destination)

ADA20:

min arrival delay time based on carriers and flight route(origin to destination)

ADA23:

Average departure delay time based on carriers and month

ADA24:

Max departure delay time based on carriers and month

ADA25:

Min departure delay time based on carriers and month

ADA26:

Average arrival delay time based on carriers and month

ADA27:

Max arrival delay time based on carriers and month

ADA28:

Min arrival delay time based on carriers and month

ADA29:

Average departure delay time based on route and month

ADA30:

Max departure delay time based on route and month

ADA31:

Min departure delay time based on route and month

ADA32:

Average arrival delay time based on route and month

ADA33:

Max arrival delay time based on route and month

ADA34:

Min arrival delay time based on route and month