

Structure of Input Vectors

Time-Dependent, Cabin-Specific Information

M Cabin

HELD_BUS_PsJs	HELD_PsJs
HELD_BUS_PsJs_HIST	HELD_PsJs_HIST
HELD_BUS_PsJs_vLW	HELD_PsJs_vLW
HELD_BUS_PsJs_vLW_HIST	HELD_PsJs_vLW_HIST
HELD_BUS_PsJs_vL4W	HELD_PsJs_vL4W
HELD_BUS_PsJs_vL4W_HIST	HELD_PsJs_vL4W_HIST
HELD_BUS_PsJs_vFW_HIST	HELD_PsJs_vFW_HIST

14 variables

C Cabin

HELD_BUS_PsJs	HELD_PsJs
HELD_BUS_PsJs_HIST	HELD_PsJs_HIST
HELD_BUS_PsJs_vLW	HELD_PsJs_vLW
HELD_BUS_PsJs_vLW_HIST	HELD_PsJs_vLW_HIST
HELD_BUS_PsJs_vL4W	HELD_PsJs_vL4W
HELD_BUS_PsJs_vL4W_HIST	HELD_PsJs_vL4W_HIST
HELD_BUS_PsJs_vFW_HIST	HELD_PsJs_vFW_HIST

14 variables

* In principle, these can depend on time (capacity can change leading up to flight day), but are (for a given flight) constant in more than 95% of cases. As such, they are treated as constant (time-independent) variables.

† This variable is a categorical string and has been one-hot-encoded into 5 variables, with one of them being dropped so as to not have a redundant variable (i.e. `drop_first = True` in Pandas). As such, `MACRO_GROUP_NM` is 4 binary variables (`macro_group_nm1`, `macro_group_nm2`, etc).

Constant, 'Summary' Information

DOW_NO
YEAR
MONTH
DAY
PSJs_LY_c
PSJs_LY_m
CAPACITY_c*
CAPACITY_m*
MACRO_GROUP_NM[†]



12 variables

* This is the overall lowest ticket price at that particular snapshot of any competitor, for the equivalent flight leaving on the same day (see my competitor query).

† This is the overall lowest ticket price at that particular snapshot for BA (again, see my query).

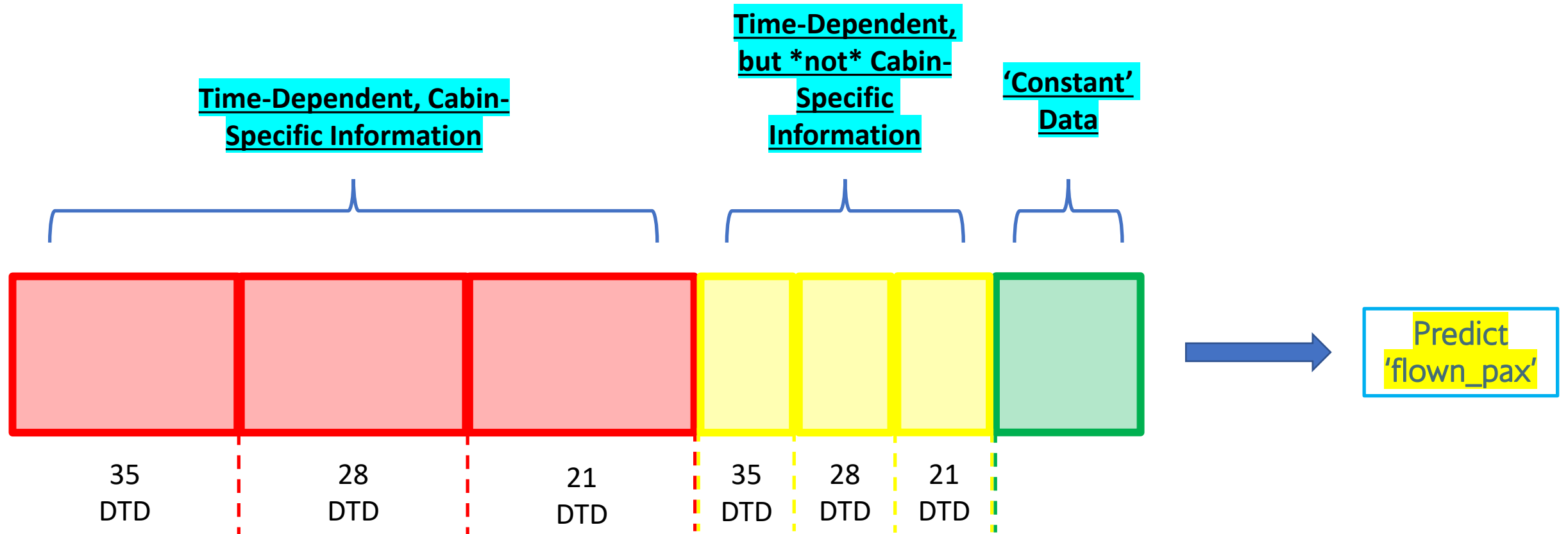
‡ This is the number of available seats for this flight at that particular snapshot, as queried from the SEATS_AVL_QTY table (see Toby's query). This number is often zero and can be negative (depends on whether the booking system is opening / closing tickets for this flight at that time).

Time-Dependent, but *not* Cabin-Specific Information

MIN_COMP_PRICE*
MIN_BA_PRICE†
SEATS_AVL‡

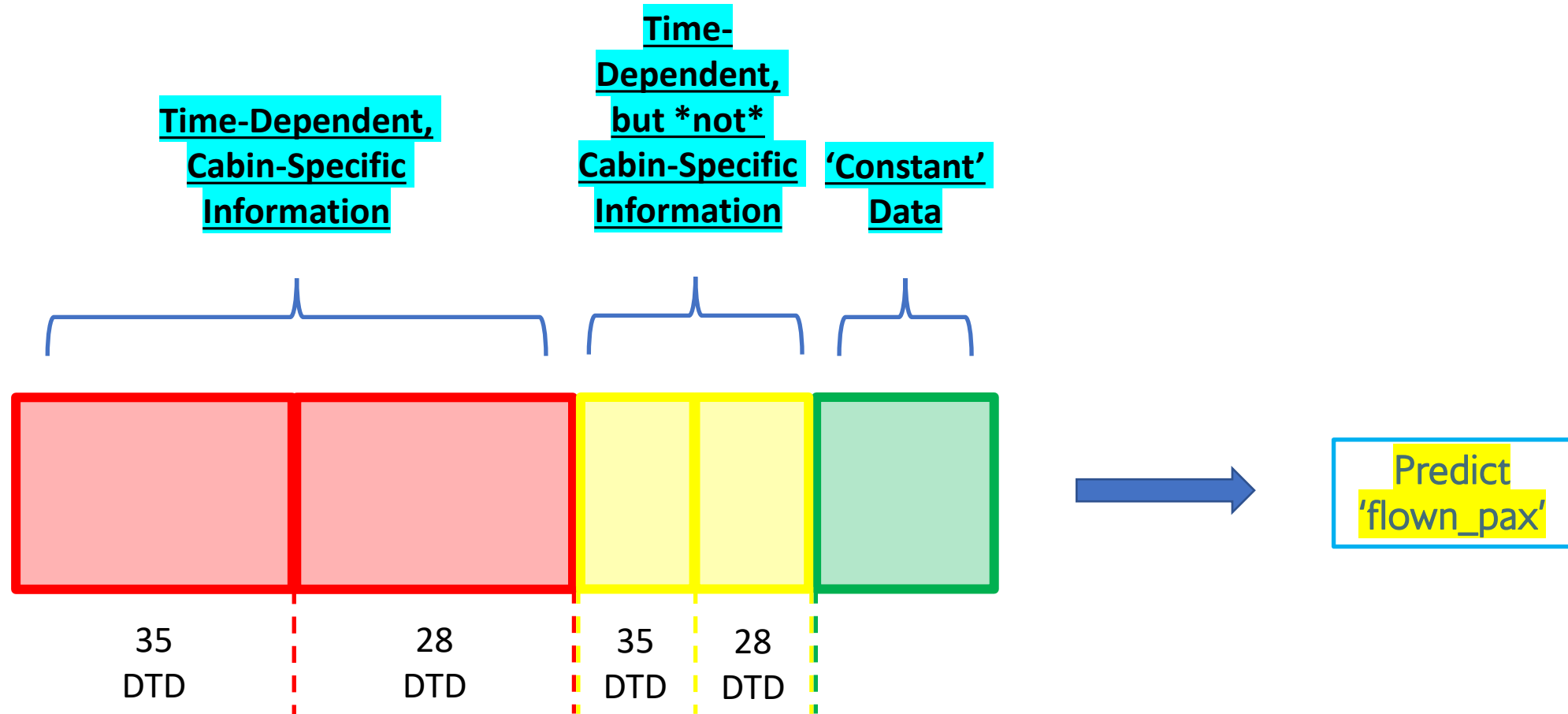
} 3 variables

Example: Predict Passengers Flown at 21 DTD



(in this case, a total of $(14+14) \times 3 + 3 \times 3 + 12 = 105$ variables are fed to the model)

Example: Predict Passengers Flown at 28 DTD



(in this case, a total of $(14+14) \times 2 + 3 \times 2 + 12 = 74$ variables are fed to the model)