Air Data & HPG Data

Memo: air_gist, hpg_gist, air_hpg_gist

Tables

Table info

	air			hpg		others		
	air_visit	air_reserve	air_store_info	hpg_reserve	hpg_store_info	date_info	store_id_relation	sample_submission
shape	(252108, 3)	(92378, 4)	(829, 5)	(1355053, 4)	(4690, 5)	(517, 3)	(150, 2)	(32019, 2)
key	(air_store_id, visit_date)	each reservation	store_id	(air_store_id, visit_date)	store_id	calendar_date	air_store_id	(air_store_id, visit_date)
meaning	252108 days of visits	92378 days of reservations	829 restaurants	1355053 days of reservations	4690 stores	517 dates	150 stores both in air and hpg	32019 days of visits
process	--	group by (air_store_id, visit_date)	--	group by (hpg_store_id, visit_date)	--	--	--	--

There is no null values in any of these tables.

Merge Tables
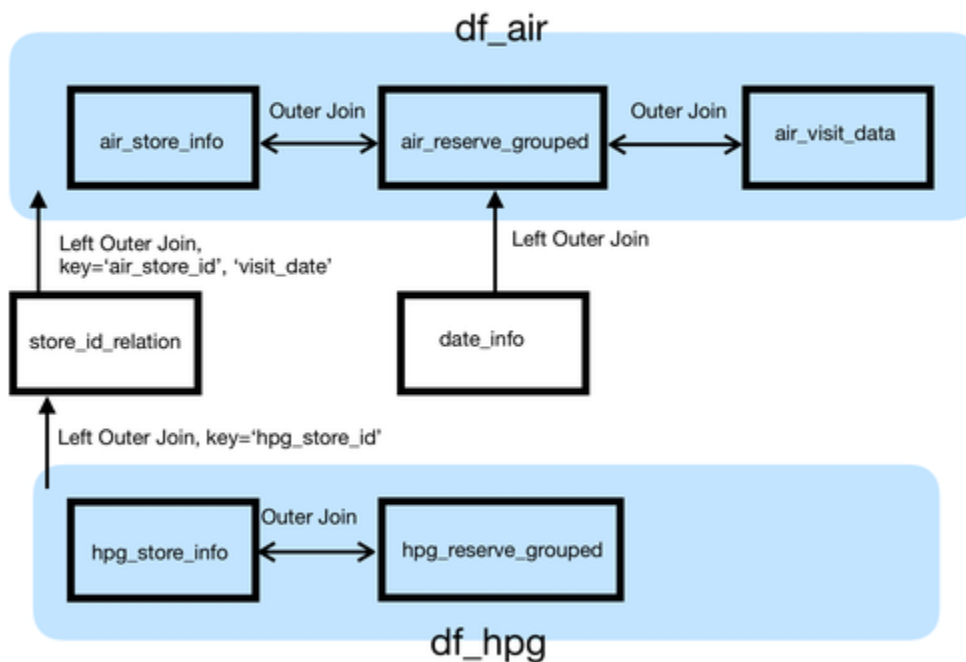
Dataframes

Dataframe info

	df_air	df_hpg	df_store_id_relation	df_submission
shape	(254389, 10)	(1355053, 7)	(150, 2)	(32019, 2)
#stores	829	13325	150	821

- There are 821 stores in submission data, all from df_air
- 150 out of 829 stores exist in both df and air data
- 8 stores disappearing in df_submission from df_air all have null reserve_visitors for all visit_dates

Next Step

1. Explore feature distribution of stores existing in both, and those only in air
2. List feature engineering possibilities
3. Track visitor variations through dates