# Categorical features

Saturday\_night\_bool, search\_adult\_count, search\_room\_count, search\_children: Make EXP features (for click bool and booking bool)

# New features

Price\_difference\_history = prop\_log\_historical\_price - log(price\_usd) OK

Price\_difference\_user = |visitor\_hist\_adr\_usd – price\_usd|OK

Hotel\_quality\_book = booking(prop\_id)/counting(prop\_id) (idea people second place) OK

Hotel\_quality\_click = click(prop\_id)/counting(prop\_id) (idea people second place) OK

# Handing missing values:

Visitor\_hist\_starrating: remove (95% missing) and create new feature: starrating\_diff = |visitor\_hist\_starrating – prop\_starrating| OK ?

Prop\_review\_score: replace missing values with lowest star rating OK

Prop\_location\_score2: replace by worst case location score OK

srch\_query\_affinity\_score: fill with worst case solution (approach people 2nd place) OK

orig\_destination\_distance = try to replace by average distance of people with same srch\_destination\_id and same visitor\_location\_country\_id. For the remaining NA’s avg? OK

Competitor values: create 1 competitor from 8 by averaging between competitors. Comp\_rate\_total, comp\_rate\_inv, comp\_rate\_percent\_diff. Or replace every NaN with zero?

Gross\_booking\_usd: remove, not an interesting feature to use in the predictions (not in test set) OK