Data-set variables investigation

Attempt to identify significant relationship between comp_pct and other variables

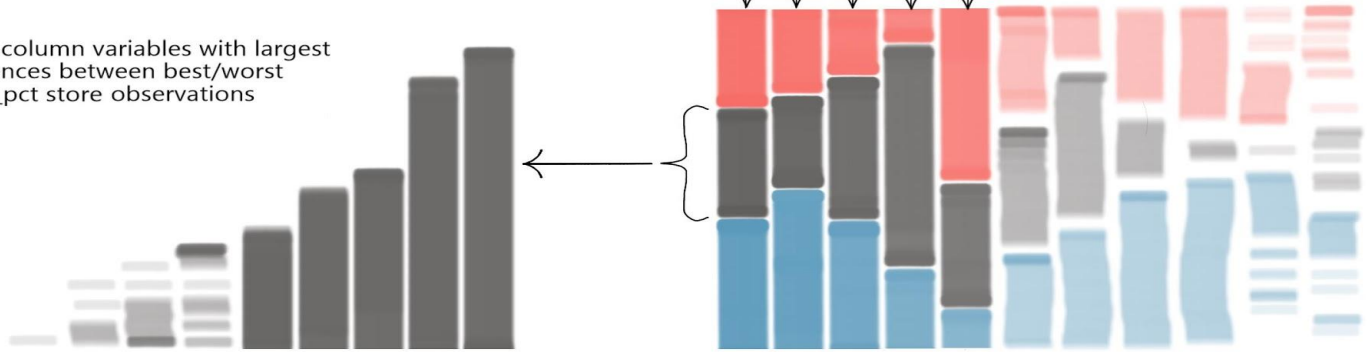
Airport
Outlet

Group and Aggregate by mean: (Mall)

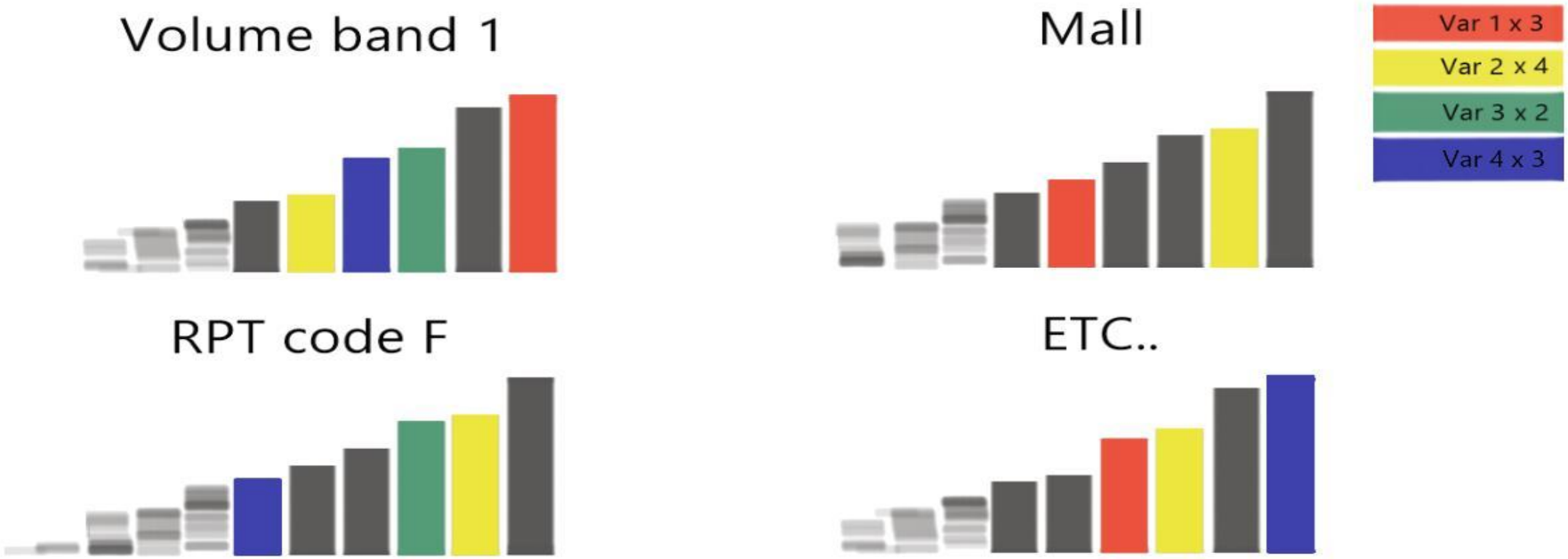
The goal of this approach is to identify column variables with the starkest differences between top and bot comp_pct observations
For this we group the dataset by different classes, rpt codes, and volume bands. We put the column valiables through multiple filters, so to speak, hoping to locate variables that persist throughout each grouping

Column variables

Select column variables with largest differences between best/worst comp_pct store observations



Obtain frequency of these variables across different store group types



Filtering the variables with this process, we selected the top 10 ranked based on frequency.

(Detailed tables(excel format) available)

CLASS GROUPINGS(7)

multi_units: 6/7

multi_value: 5/7

trans_cnt_exchg_in: 4/7

ft_tenure_days: 4/7

cp_total_email_trans: 3/7

cp_total_trans: 3/7

am_tenure_days: 3/7

cp_valid_emails: 3/7

cp_valid_mailing_address: 3/7

accessory_shoe_trees_units: 3/7

VOLUME BAND GROUPINGS(6)

mark_down_amt_ty: 4/6

special_order_amt: 4/6

trans_cnt_exchg_in: 3/6

ft_tenure_days: 3/6

cp_total_email_trans: 3/6

accessory_shoe_trees_value: 3/6

cp_valid_emails: 3/6

cp_valid_mailing_address: 3/6

accessory_shoe_trees_units: 3/6

cp_total_trans: 3/6

RPT CODE GROUPINGS(3)

multi_units: 3/3

trans_cnt_exchg_in: 2/3

cp_valid_emails: 2/3

cp_total_trans: 2/3

trans_cnt_exchg_out: 2/3

cp_total_email_trans: 2/3

special_order_amt: 2/3

cp_valid_mailing_address: 2/3

accessory_shoe_care_value: 2/3

multi_value: 2/3

PERSISTED ACROSS CLASSES, VOLUME BANDS, AND RPT CODES AT LEAST ONCE

trans_cnt_exchg_in

cp_total_email_trans

cp_total_trans

cp_valid_emails

cp_valid_mailing_address