**Data Dictionary**

## Data Background

* The TV airing data is consisted of 13 columns: Media, Estimate, Access, DMA, Market, Station ID, Length, Date Aired, Time Aired, Spot Cost, Spot Type, GRP and Impression (000). The data dictionary is provided below.
* Data range: 2018-12-31 ~ 2019-12-29
* Time zone: EST

## Data Column Definition

**Media**: “TV” only.

**Estimate**: quarter of the year in which the spot was aired. “QQYY” format.

**Access**: the creative code, used to indicate which creative was used in the spot. \*In most cases, there would be multiple versions of TV advertisements shown on TV. In media world, we call each version of these advertisement a unique creative.

**DMA**: either “900” or (blank).

**Market:** indicates the broadcast coverage of the station on which the spot was aired. In general, the larger the coverage, the higher the GRP: National > Regional > Local.

**Station ID**: indicates on which station (hashed) the spot was aired.

**Length**: spot length in seconds.

**Date Aired**: the spot airing date in MM/DD/YYYY format.

**Time Aired**: the spot airing time in HH:MM:SS format.

**Spot Cost**: Spot cost in $. In general, the higher the spot cost, the higher the GRP.

**Spot Type**: Internal system Tracking code (can be ignored).

**GRP**: individual spot’s GRP. If it’s 0, the spot is unrated and this’s where we need to generate projected GRP. otherwise, it’s the actual accurate GRP for the spot (rated).

**Impression (000)**: Number of eyeballs exposed to the spot (in thousand). If the spot is unrated, then the impression is also unknown (shown as 0).

Among all available information, we believe the above highlighted columns are worth considering as dependent variables for the unrated spot GRP estimation project.

In addition, we also consider **DOW** (day of week) and **daypart** as two important time dimensions that should effectively improve the prediction accuracy.

* The DOW can be determined based on **Date Aired** column.
* The daypart can be determined based on **Time Aired (EST)** column, following the below hour break down:
  + Late Night: 12AM-2AM
  + Overnight: 2AM-6AM
  + Morning: 6AM-9AM
  + Daytime: 9AM-4PM
  + Early Fringe: 4PM-7PM
  + Primetime: 7PM-10PM
  + Late Fringe: 10PM-12AM