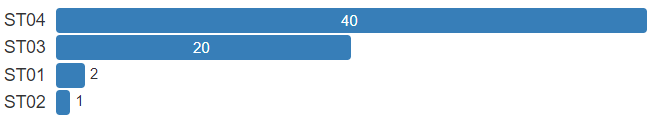
**CITIES**: 6 variáveis (1 numérica e 6 categóricas) e 63 observações, sem valores em falta

- store\_id: alfanumérico com identificação da loja, 63 valores distintos

- storetype\_id: alfanumérico com identificação do tipo de loja, 4 valores distintos



- store\_size: numérico (real) com o tamanho da loja, 32 valores distintos



- city\_id\_old: alfanumérico com identificação da cidade, 19 valores distintos



- country\_id: texto com nome do país, apenas 1 valor (Turquia)

- city\_code: texto com nome das cidades, 19 valores distintos

**PRODUCT**: 10 variáveis (3 numéricas e 7 categóricas) e 699 observações, com 100 valores em falta (1,4%)

- product\_id: alfanumérico com identificação do produto, 699 valores distintos

- product\_length: numérico (real) com comprimento do produto, 123 valores distintos e 18 valores em falta (2,6%) e 1 valor “zero” (0,1%)

- product\_depth: numérico (real) com profundidade do produto, 163 valores distintos e 16 valores em falta (2,3%) e 1 valor “zero” (0,1%)

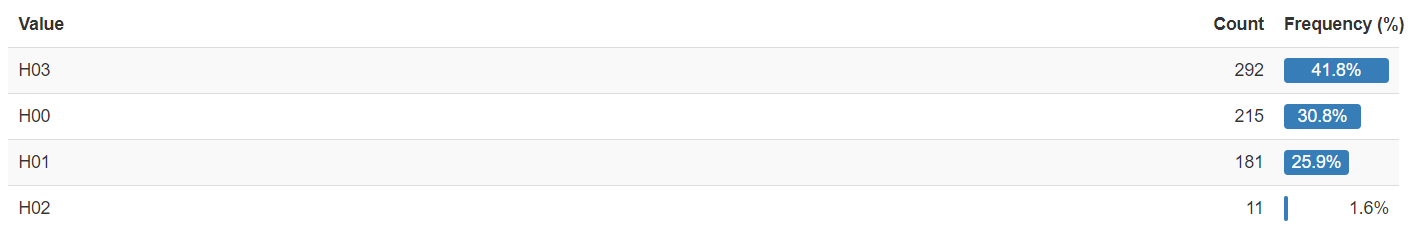
- product\_width: numérico (real) com largura do produto, 144 valores distintos e 16 valores em falta (2,3%) e 1 valor “zero” (0,1%)

- cluster\_id: categórico com indicação de uma segmentação/cluster, 10 valores distintos e 50 valores em falta (7,2%)

Com os dados que temos ainda não consegui perceber como foi feita esta segmentação nem nenhum padrão, sendo que o “cluster\_1” é o mais reduzido e tem 4 produtos (apenas destas hierarquias H000003 e H000004)



- hierarchy1\_id: alfanumérico com identificação da hierarquia de nível “1” (mais abrangente), 4 valores distintos, sem valores em falta



- hierarchy2\_id: alfanumérico com identificação da hierarquia de nível “2”, 18 valores distintos, sem valores em falta

- hierarchy3\_id: alfanumérico com identificação da hierarquia de nível “3”, 79 valores distintos, sem valores em falta

- hierarchy4\_id: alfanumérico com identificação da hierarquia de nível “4”, 168 valores distintos, sem valores em falta

- hierarchy5\_id: alfanumérico com identificação da hierarquia de nível “5”, 373 valores distintos, sem valores em falta

**SALES**: 14 variáveis (6 numéricas e 8 categóricas) e 8.886.058 observações, com 35.271.795 valores em falta (28,4%), mas sem duplicados

- Unnamed: numérico (real) com id sequencial das vendas, 100% valores distintos, sem valores em falta

- store\_id: alfanumérico com identificação da loja, 63 valores distintos e sem valores em falta

- product\_id: alfanumérico com identificação do produto, 615 valores distintos

- date: datas compreendidas entre o dia 02-01-2017 e 31-10-2019

- sales: numérico (real) com as quantidades vendidas (mínimo de 0 e máximo de 43301), com 302296 valores em falta (3,4%) e 7048907 zeros (79,3%)



- revenue: numérico (real) com a receita da venda, 302296 valores em falta (3,4% - igual a ‘sales’) e 7049979 zeros (79,3% - 1072 zeros a mais que ‘sales’)

- stock: numérico (real) com o stock do produto, 302296 valores em falta (3,4% - igual a ‘sales’ e ‘revenue’) e 66086 zeros (0,7%)

- price: numérico (real) com o preço do produto, 606 valores distintos (ou seja, existem produtos com o mesmo preço), 91381 valores em falta (1972 não apresentam valores em ‘sales’, ‘revenue’ e ‘stock’ | 69035 com ‘sales’=0 e ‘revenue’=0 | 69057 com ‘revenue’=0 | 22 com ‘revenue’=0 mas ‘sales’>0 de produtos, lojas e datas diferentes)

- promo\_type\_1: alfanumérico com tipo promoção, 17 valores distintos, sem valores em falta



- promo\_bin\_1: categorias com característica da promoção, 5 valores distintos e 7653515 valores em falta



- promo\_type\_2: alfanumérico com tipo promoção, 4 valores distintos, sem valores em falta



- promo\_bin\_2: categorias com característica da promoção, 3 valores distintos e 8873337 valores em falta (99,9%)



- promo\_discount\_2: numérico (real) com valor do desconto aplicado, 6 valores distintos (entre 16 e 50) e 8873337 valores em falta (igual a ‘promo\_bin\_2’)

- promo\_discount\_type\_2: alfanumérico com tipo desconto, 4 valores distintos e 8873337 valores em falta (igual a ‘promo\_bin\_2’ e a ‘promo\_discount\_2’)

