

Ordeal 0 Part3

🕒 Created	@October 17, 2022 9:22 PM
🏷️ Tags	

概要

Part 2では実際にデータが保存されている場所、保存方法を学習した。

Part 3では大量のデータを理解できるように、集計の基礎とその方法を学習する。
学習後、様々な観点からデータを集計しデータを理解することができるようになる。

参照

・[出典：KT Channel] 「データを理解する集計の力」

<https://www.youtube.com/watch?v=q11RLh3Z4yY>

#DataSaberBridge

#Tableau

#Ordeal 0

#データ

#集計

集計について

私たちが見るデータは基本大量。

大量のデータを一気に見ても理解できない（＝「データの海に溺れる」とも言う）
ので、「**集計**」することでデータを理解する

集計＝数字を**ある単位・かたまりで集めて計算**する（合計、平均など）

※集計≠合計、合計は集計の手法の1種。

集計で重要なこと＝集計の単位[データの粒度 (Level of Detail, LOD)]を考えること。

集計する際にはデータを

- どの単位で集めるのか
- どのかたまりで集計するのか

を考える必要がある。

例：集計単位を「年」とすれば、各年の売上金額の集計など

集計単位の例

①時間

No.	単位	事例
1	億	地球の歴史
2	秒	動画

②場所

No.	単位	事例
1	国	世界各国のGDP
2	経緯度	人の移動履歴

③人

No.	単位	事例
1	年齢/性別	人口ピラミッド
2	個人	病院の診断結果、カルテ

④モノ

No.	単位	事例
1	カテゴリ	商品カテゴリごとの売上金額

2	製品	各製品の売上金額
---	----	----------

※データの粒度×ストーリーテリング

→最終的に伝えたい内容をもとにデータの粒度を考えて、要点をつないでいく。

(例：2021年以降に注力していく商品開発に向けて、

- 2020年1月から6月の半年間を通じてA社の製品のうちどの商品が一番売上総額が高かったか
- 次に、どの年齢層に一番売上総額が高い商品が売れていたか
- さらに、どの地域に一番売上総額が高い商品が売れていたか

など)

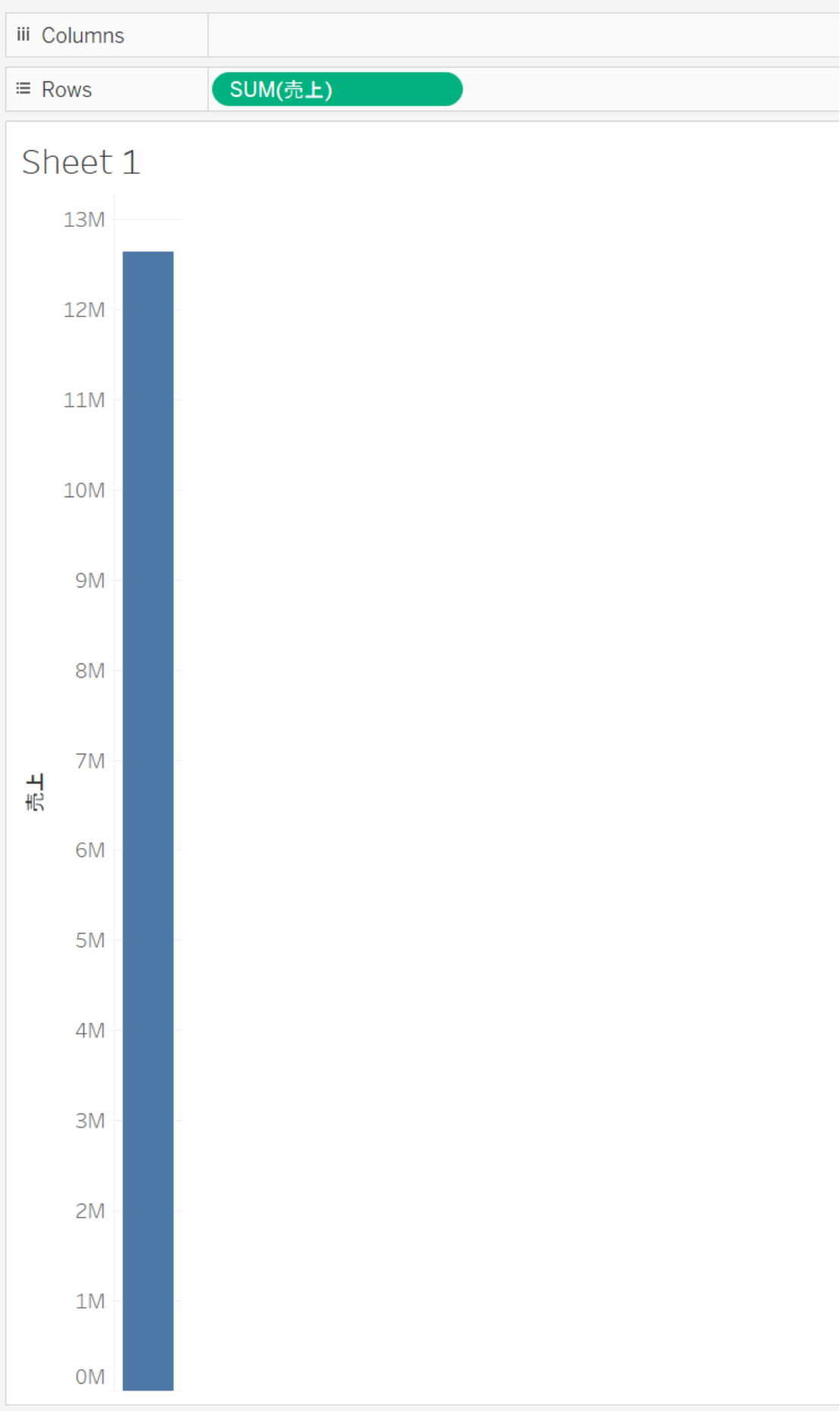
集計における計算方法

集計における基本的な計算方法→合計、平均など

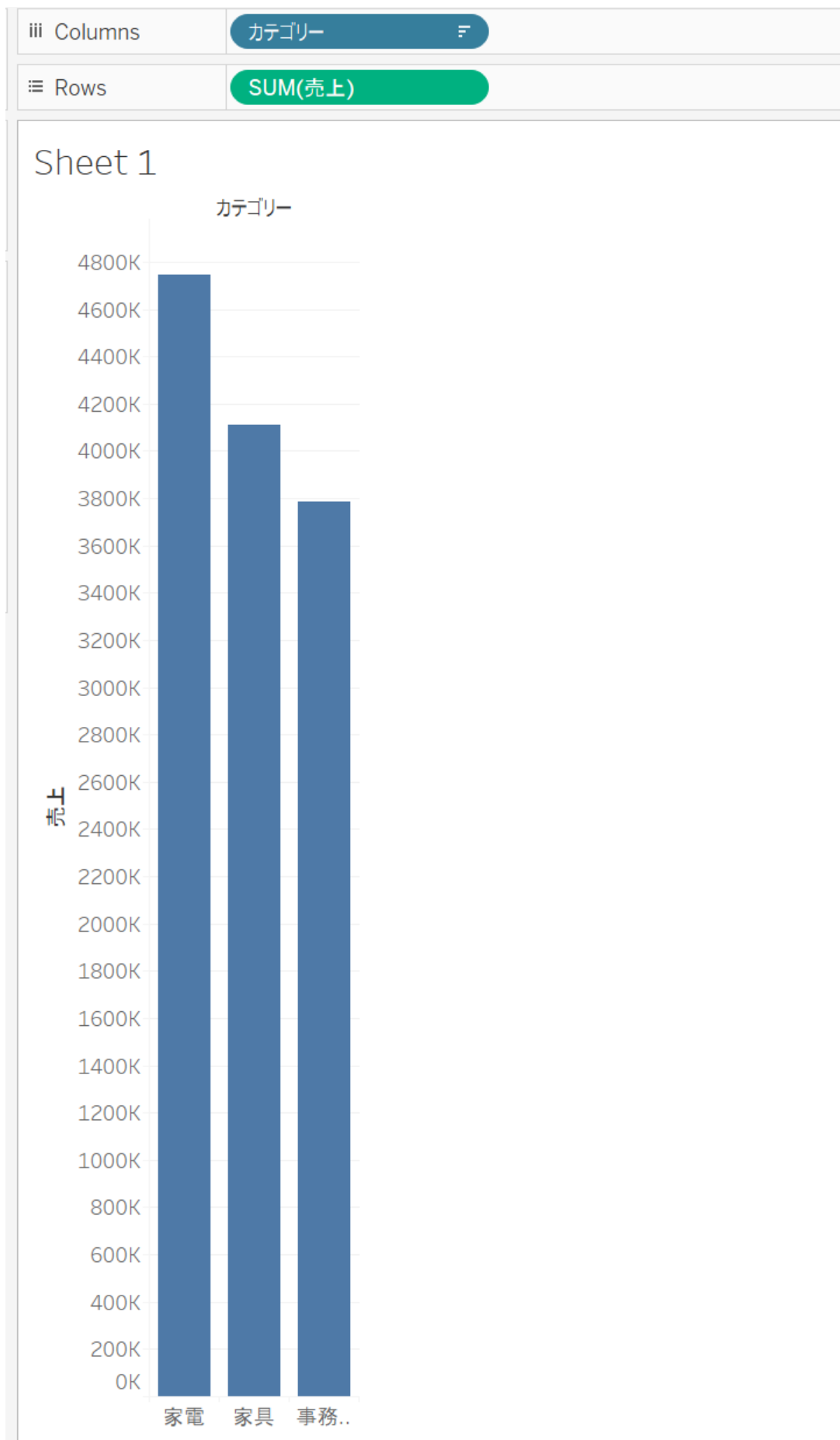
自分や相手が知りたい計算方法を選ぶ。

データの粒度や計算方法により異なる集計結果

■全データの売上合計



■各カテゴリの売上合計



Pages

Filters

Marks

Automatic

Color

Size

Text

Detail

Tooltip

SUM(売上)

Columns

Rows

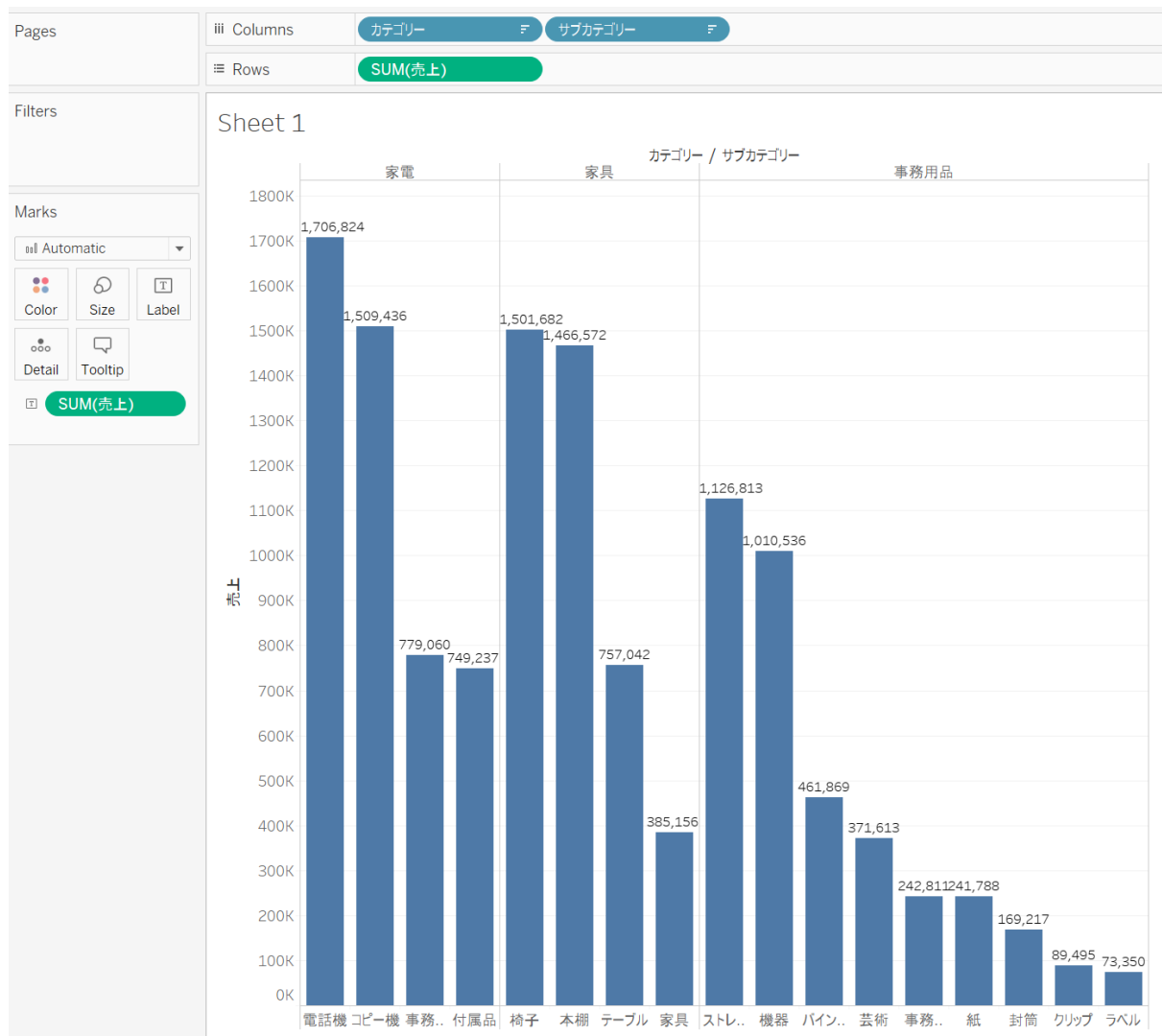
カテゴリー

Sheet 2

カテゴリー

家具	4,110,452
家電	4,744,557
事務用品	3,787,493

■サブカテゴリごとの売上合計



Pages

Filters

Marks

Automatic

Color

Size

Text

Detail

Tooltip

SUM(売上)

Columns

Rows

カテゴリー

サブカテゴリー

Sheet 2

カテゴリー	サブ..	Σ
家電	電話機	1,706,824
	コピー機	1,509,436
	事務機器	779,060
	付属品	749,237
家具	椅子	1,501,682
	本棚	1,466,572
	テーブル	757,042
	家具	385,156
事務用品	ストレージ	1,126,813
	機器	1,010,536
	バインダー	461,869
	芸術	371,613
	事務用品	242,811
	紙	241,788
	封筒	169,217
	クリップ	89,495
	ラベル	73,350

■サブカテゴリごとの売上平均



Pages

Filters

Marks

iii Columns

Rows

カテゴリー

サブカテゴリー

Sheet 2

カテゴリー

サブ..

家電	電話機	508.4
	コピー機	679.0
	事務機器	524.3
	付属品	243.7
家具	椅子	437.3
	本棚	608.3
	テーブル	879.3
	家具	122.1
事務用品	ストレージ	223.2
	機器	580.1
	バインダー	75.1
	芸術	76.4
	事務用品	100.9
	紙	69.2
	封筒	70.9
	クリップ	34.4
	ラベル	28.2

Automatic

Color

Size

Text

Detail

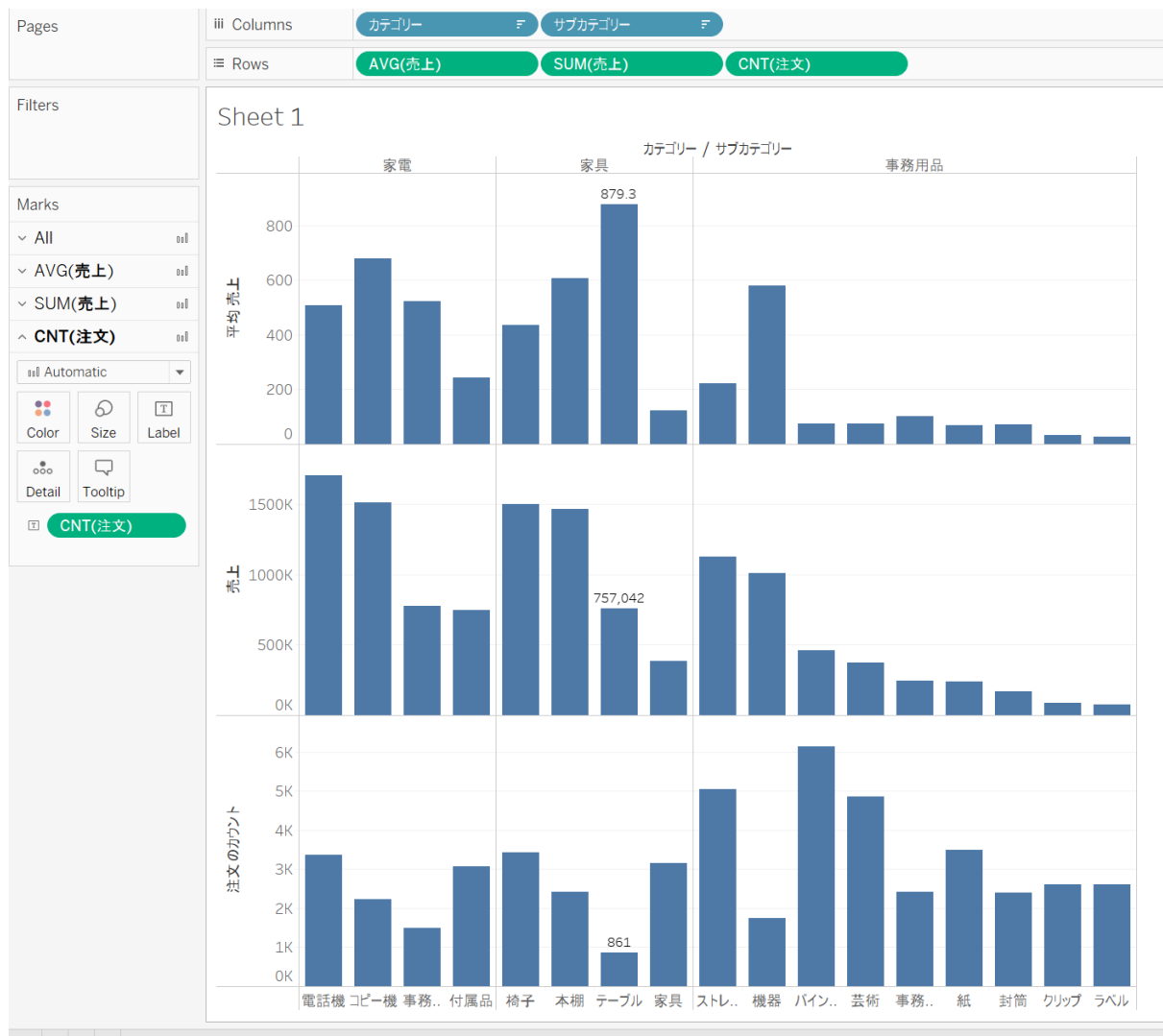
Tooltip

AVG(売上)

■テーブルの売上合計・平均・注文件数

テーブルの売上合計は小さいのに売上平均が高い

→ テーブルの注文件数が少ない。



Pages	iii Columns																																																																																																																															
	Rows	<div> <div>カテゴリー</div> <div>サブカテゴリー</div> </div>																																																																																																																														
Filters	<div>Sheet 2</div> <div>カテゴリー サブ..</div> <table> <tr> <td>家電</td><td>電話機</td><td>508.4</td></tr> <tr> <td></td><td></td><td>1,706,824</td></tr> <tr> <td></td><td></td><td>3,357</td></tr> <tr> <td></td><td>コピー機</td><td>679.0</td></tr> <tr> <td></td><td></td><td>1,509,436</td></tr> <tr> <td></td><td></td><td>2,223</td></tr> <tr> <td></td><td>事務機器</td><td>524.3</td></tr> <tr> <td></td><td></td><td>779,060</td></tr> <tr> <td></td><td></td><td>1,486</td></tr> <tr> <td></td><td>付属品</td><td>243.7</td></tr> <tr> <td></td><td></td><td>749,237</td></tr> <tr> <td></td><td></td><td>3,075</td></tr> <tr> <td>家具</td><td>椅子</td><td>437.3</td></tr> <tr> <td></td><td></td><td>1,501,682</td></tr> <tr> <td></td><td></td><td>3,434</td></tr> <tr> <td></td><td>本棚</td><td>608.3</td></tr> <tr> <td></td><td></td><td>1,466,572</td></tr> <tr> <td></td><td></td><td>2,411</td></tr> <tr> <td></td><td>テーブル</td><td>879.3</td></tr> <tr> <td></td><td></td><td>757,042</td></tr> <tr> <td></td><td></td><td>861</td></tr> <tr> <td></td><td>家具</td><td>122.1</td></tr> <tr> <td></td><td></td><td>385,156</td></tr> <tr> <td></td><td></td><td>3,154</td></tr> <tr> <td>事務用品</td><td>ストレージ</td><td>223.2</td></tr> <tr> <td></td><td></td><td>1,126,813</td></tr> <tr> <td></td><td></td><td>5,049</td></tr> <tr> <td></td><td>機器</td><td>580.1</td></tr> <tr> <td></td><td></td><td>1,010,536</td></tr> <tr> <td></td><td></td><td>1,742</td></tr> <tr> <td></td><td>バインダー</td><td>75.1</td></tr> <tr> <td></td><td></td><td>461,869</td></tr> <tr> <td></td><td></td><td>6,146</td></tr> <tr> <td></td><td>芸術</td><td>76.4</td></tr> <tr> <td></td><td></td><td>371,613</td></tr> <tr> <td></td><td></td><td>4,864</td></tr> <tr> <td></td><td>事務用品</td><td>100.9</td></tr> <tr> <td></td><td></td><td>242,811</td></tr> <tr> <td></td><td></td><td>2,407</td></tr> <tr> <td></td><td>紙</td><td>69.2</td></tr> <tr> <td></td><td></td><td>241,788</td></tr> <tr> <td></td><td></td><td>3,492</td></tr> </table>		家電	電話機	508.4			1,706,824			3,357		コピー機	679.0			1,509,436			2,223		事務機器	524.3			779,060			1,486		付属品	243.7			749,237			3,075	家具	椅子	437.3			1,501,682			3,434		本棚	608.3			1,466,572			2,411		テーブル	879.3			757,042			861		家具	122.1			385,156			3,154	事務用品	ストレージ	223.2			1,126,813			5,049		機器	580.1			1,010,536			1,742		バインダー	75.1			461,869			6,146		芸術	76.4			371,613			4,864		事務用品	100.9			242,811			2,407		紙	69.2			241,788			3,492
家電	電話機	508.4																																																																																																																														
		1,706,824																																																																																																																														
		3,357																																																																																																																														
	コピー機	679.0																																																																																																																														
		1,509,436																																																																																																																														
		2,223																																																																																																																														
	事務機器	524.3																																																																																																																														
		779,060																																																																																																																														
		1,486																																																																																																																														
	付属品	243.7																																																																																																																														
		749,237																																																																																																																														
		3,075																																																																																																																														
家具	椅子	437.3																																																																																																																														
		1,501,682																																																																																																																														
		3,434																																																																																																																														
	本棚	608.3																																																																																																																														
		1,466,572																																																																																																																														
		2,411																																																																																																																														
	テーブル	879.3																																																																																																																														
		757,042																																																																																																																														
		861																																																																																																																														
	家具	122.1																																																																																																																														
		385,156																																																																																																																														
		3,154																																																																																																																														
事務用品	ストレージ	223.2																																																																																																																														
		1,126,813																																																																																																																														
		5,049																																																																																																																														
	機器	580.1																																																																																																																														
		1,010,536																																																																																																																														
		1,742																																																																																																																														
	バインダー	75.1																																																																																																																														
		461,869																																																																																																																														
		6,146																																																																																																																														
	芸術	76.4																																																																																																																														
		371,613																																																																																																																														
		4,864																																																																																																																														
	事務用品	100.9																																																																																																																														
		242,811																																																																																																																														
		2,407																																																																																																																														
	紙	69.2																																																																																																																														
		241,788																																																																																																																														
		3,492																																																																																																																														

Marks

Automatic

Color

Size

Text

Detail

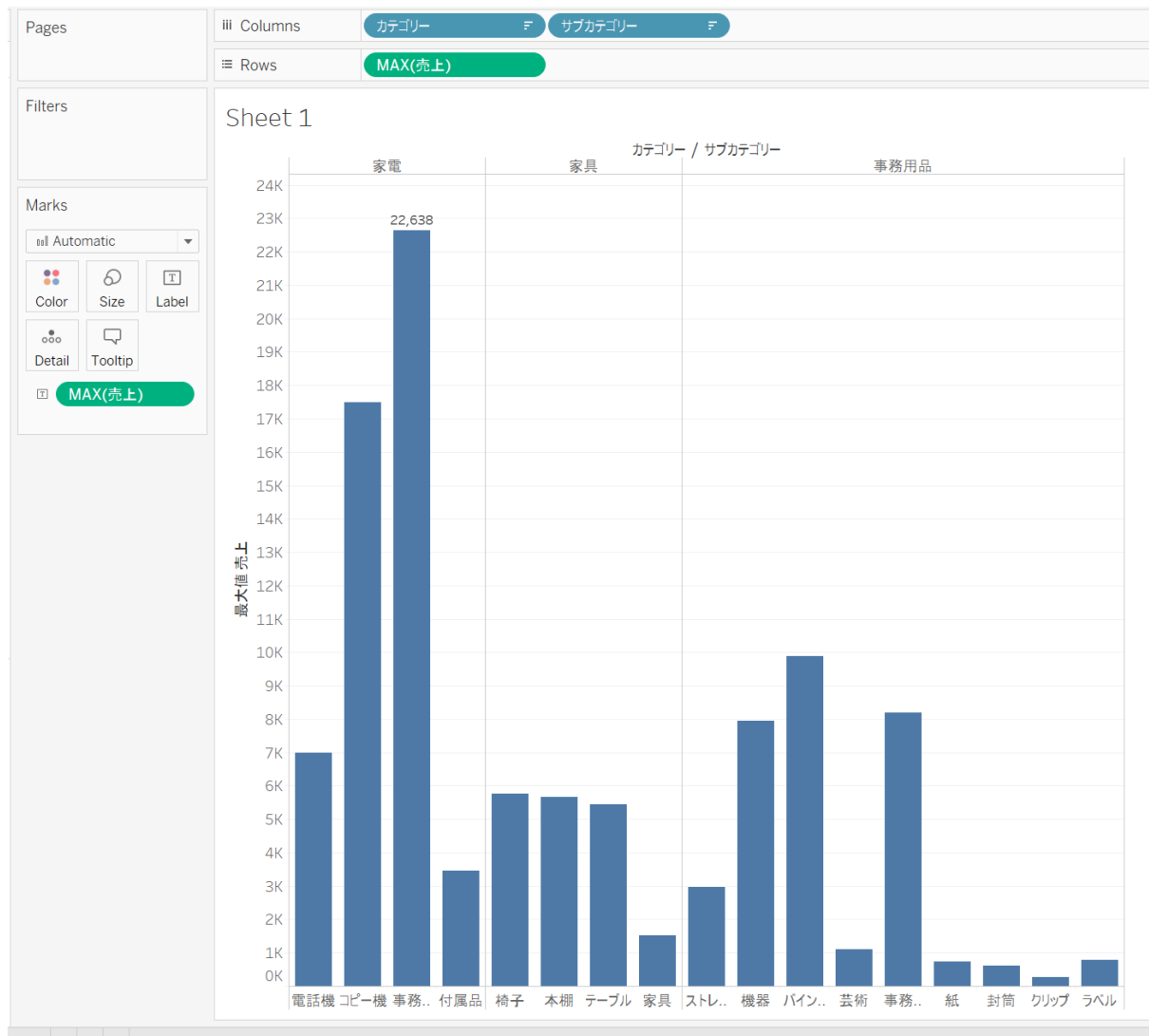
Tooltip

AVG(売上)

SUM(売上)

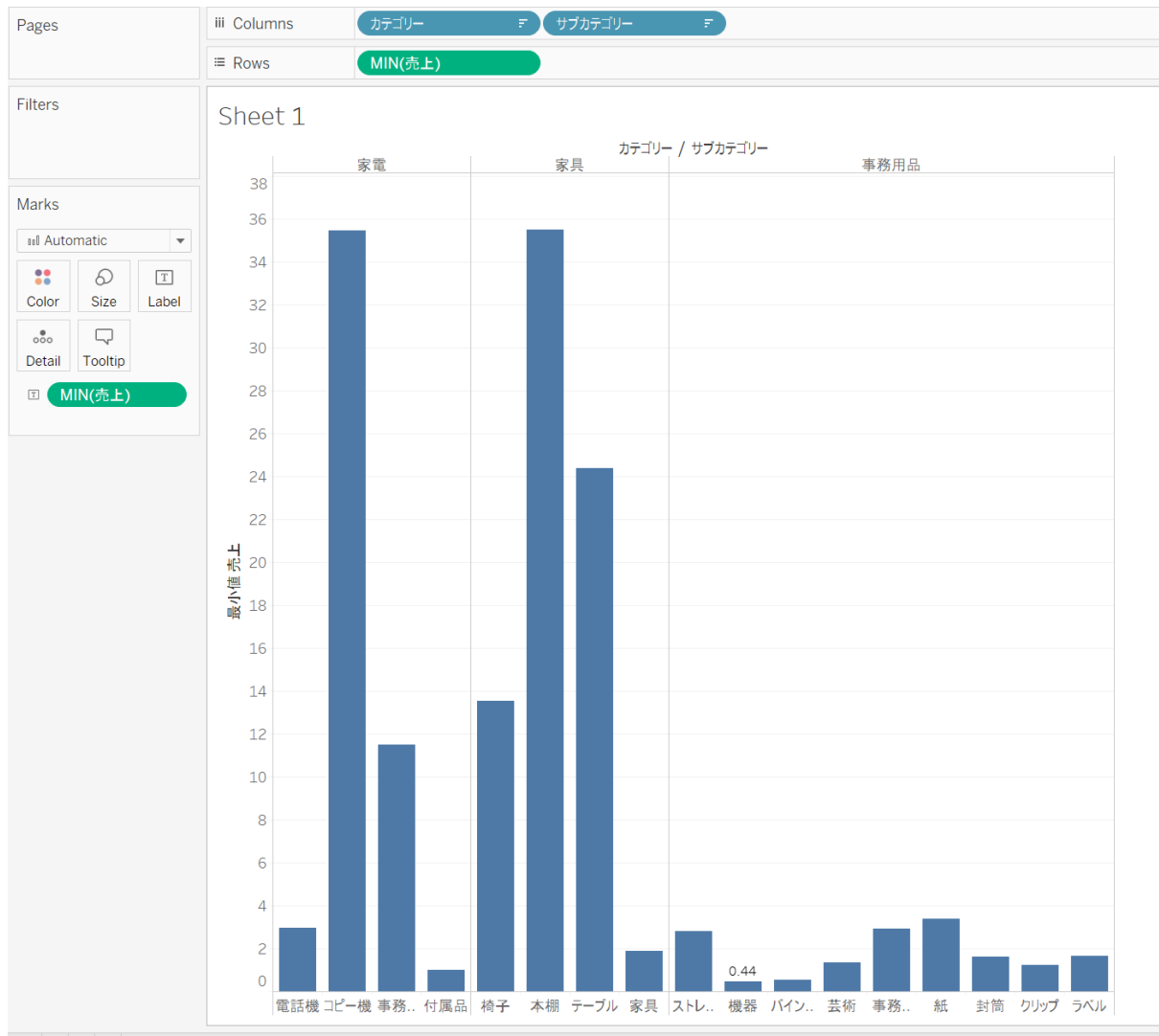
CNT(注文)

■サブカテゴリ内の売上の最大値



Pages	Columns																																																				
	Rows	<div> <div>カテゴリー</div> <div>サブカテゴリー</div> </div>																																																			
Filters	<div> <div>Sheet 2</div> <div> <div>カテゴリー</div> <div>サブカテゴリー</div> </div> <table> <tr> <td>家電</td><td>電話機</td><td>6,999</td></tr> <tr> <td></td><td>コピー機</td><td>17,500</td></tr> <tr> <td></td><td>事務機器</td><td>22,638</td></tr> <tr> <td></td><td>付属品</td><td>3,450</td></tr> <tr> <td>家具</td><td>椅子</td><td>5,760</td></tr> <tr> <td></td><td>本棚</td><td>5,668</td></tr> <tr> <td></td><td>テーブル</td><td>5,451</td></tr> <tr> <td></td><td>家具</td><td>1,519</td></tr> <tr> <td>事務用品</td><td>ストレージ</td><td>2,963</td></tr> <tr> <td></td><td>機器</td><td>7,959</td></tr> <tr> <td></td><td>バインダー</td><td>9,893</td></tr> <tr> <td></td><td>芸術</td><td>1,113</td></tr> <tr> <td></td><td>事務用品</td><td>8,188</td></tr> <tr> <td></td><td>紙</td><td>734</td></tr> <tr> <td></td><td>封筒</td><td>605</td></tr> <tr> <td></td><td>クリップ</td><td>271</td></tr> <tr> <td></td><td>ラベル</td><td>786</td></tr> </table> </div>		家電	電話機	6,999		コピー機	17,500		事務機器	22,638		付属品	3,450	家具	椅子	5,760		本棚	5,668		テーブル	5,451		家具	1,519	事務用品	ストレージ	2,963		機器	7,959		バインダー	9,893		芸術	1,113		事務用品	8,188		紙	734		封筒	605		クリップ	271		ラベル	786
家電	電話機	6,999																																																			
	コピー機	17,500																																																			
	事務機器	22,638																																																			
	付属品	3,450																																																			
家具	椅子	5,760																																																			
	本棚	5,668																																																			
	テーブル	5,451																																																			
	家具	1,519																																																			
事務用品	ストレージ	2,963																																																			
	機器	7,959																																																			
	バインダー	9,893																																																			
	芸術	1,113																																																			
	事務用品	8,188																																																			
	紙	734																																																			
	封筒	605																																																			
	クリップ	271																																																			
	ラベル	786																																																			
Marks	<div> <div>Automatic</div> <div> <div>Color</div> <div>Size</div> <div>Text</div> </div> <div> <div>Detail</div> <div>Tooltip</div> </div> <div>MAX(売上)</div> </div>																																																				

■サブカテゴリ内の売上の最小値



Pages	Columns																																									
Filters	Rows	<div> <div>カテゴリー</div> <div>サブカテゴリー</div> </div>																																								
<div>Marks</div> <div> <div>Automatic</div> <div>Color</div> <div>Size</div> <div>Text</div> <div>Detail</div> <div>Tooltip</div> <div>MIN(売上)</div> </div>	<div>Sheet 2</div> <table> <tr> <th>カテゴリー</th><th>サブ..</th><th></th></tr> <tr> <td rowspan="4">家電</td><td>電話機</td><td>2.97</td></tr> <tr> <td>コピー機</td><td>35.46</td></tr> <tr> <td>事務機器</td><td>11.51</td></tr> <tr> <td>付属品</td><td>0.99</td></tr> <tr> <td rowspan="4">家具</td><td>椅子</td><td>13.55</td></tr> <tr> <td>本棚</td><td>35.49</td></tr> <tr> <td>テーブル</td><td>24.37</td></tr> <tr> <td>家具</td><td>1.89</td></tr> <tr> <td rowspan="9">事務用品</td><td>ストレージ</td><td>2.80</td></tr> <tr> <td>機器</td><td>0.44</td></tr> <tr> <td>バインダー</td><td>0.56</td></tr> <tr> <td>芸術</td><td>1.34</td></tr> <tr> <td>事務用品</td><td>2.92</td></tr> <tr> <td>紙</td><td>3.38</td></tr> <tr> <td>封筒</td><td>1.63</td></tr> <tr> <td>クリップ</td><td>1.24</td></tr> <tr> <td>ラベル</td><td>1.67</td></tr> </table>		カテゴリー	サブ..		家電	電話機	2.97	コピー機	35.46	事務機器	11.51	付属品	0.99	家具	椅子	13.55	本棚	35.49	テーブル	24.37	家具	1.89	事務用品	ストレージ	2.80	機器	0.44	バインダー	0.56	芸術	1.34	事務用品	2.92	紙	3.38	封筒	1.63	クリップ	1.24	ラベル	1.67
カテゴリー	サブ..																																									
家電	電話機	2.97																																								
	コピー機	35.46																																								
	事務機器	11.51																																								
	付属品	0.99																																								
家具	椅子	13.55																																								
	本棚	35.49																																								
	テーブル	24.37																																								
	家具	1.89																																								
事務用品	ストレージ	2.80																																								
	機器	0.44																																								
	バインダー	0.56																																								
	芸術	1.34																																								
	事務用品	2.92																																								
	紙	3.38																																								
	封筒	1.63																																								
	クリップ	1.24																																								
	ラベル	1.67																																								

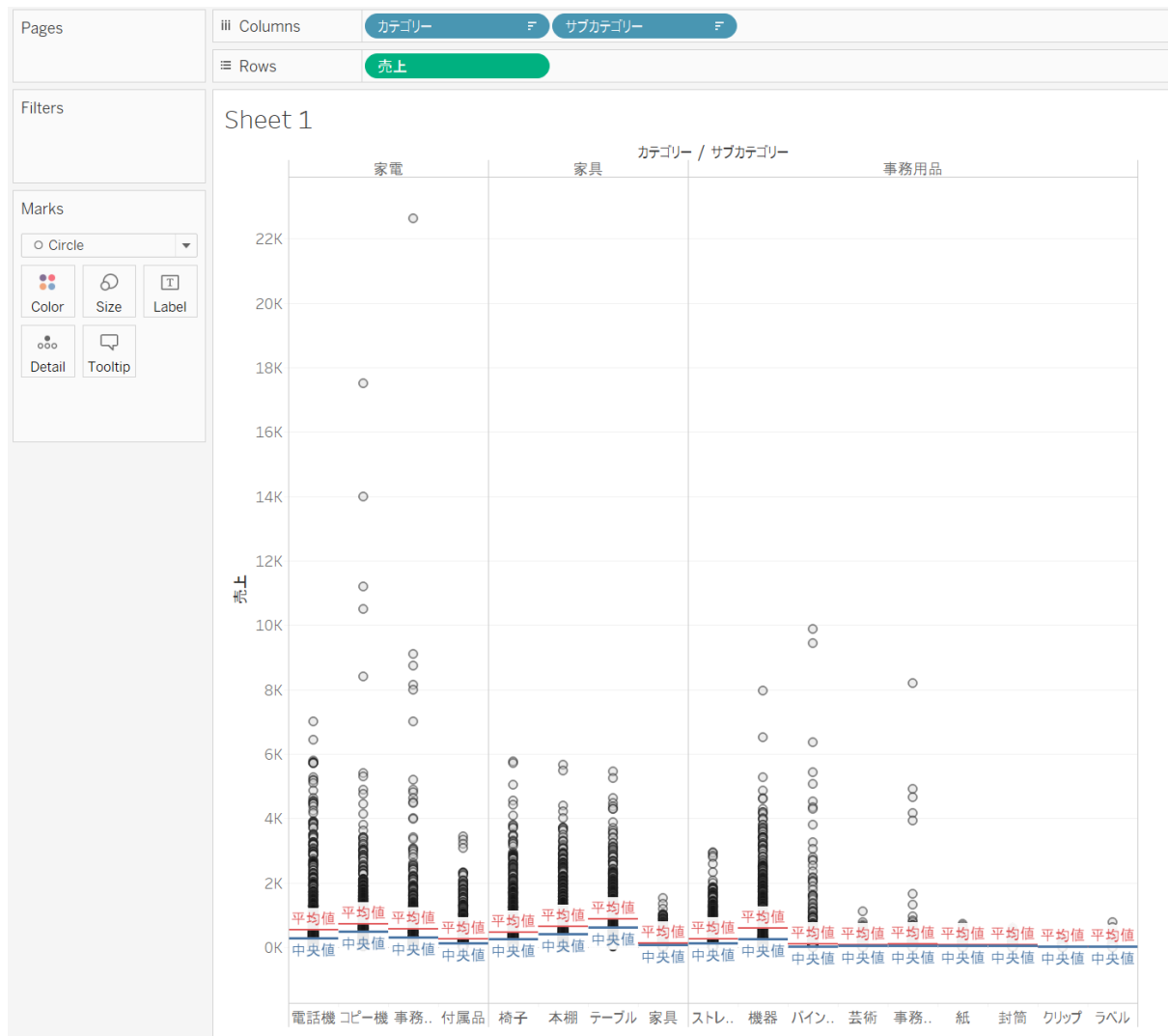
■平均値と中央値

中央値は、平均値に比べて外れ値（データが密集しているところから離れている値）に影響されにくい集計結果が得られる。

集計方法を中央値もしくは平均値を使用するかは

- 自分の見たいもの
- 相手に見せたいもの

に応じて判断する。



まとめ

データを理解する集計には

- データの粒度
- 計算方法
- ストーリーテリング

が必要。

データの粒度と計算方法次第で様々な角度でデータを把握することができる。

また、データを通じて伝えたいものに応じて、データの粒度と計算方法を適切に選んでいくことが重要。

データの粒度 × 計算方法

粒度の範囲を理解し、適切な計算方法を選ぶことで
私たちは様々な角度でデータを俯瞰することができる



データを理解する集計の力

データの粒度
×
ストーリーテリング
×
計算方法