Data Assessment Process

Original Data site: <https://www.kaggle.com/zynicide/wine-reviews>

Updated data Was provided at: <https://www.kaggle.com/zynicide/wine-reviews/discussion/83970>

Downloaded update file from: <https://drive.google.com/file/d/1CU_HASG-vPw0KANv6_8BRKCM94ardLjv/view>

Redo of the data process – 2021-04-27 22:00

Number of observations: 141,617

Dups: Apparently removed

Added “id” as the name for the 1st column number

Sliced data down:

California: 🡺 39072

All Descriptions Populated: Yes

Designations missing: 18789

Points populated: X

Blanks Prices removed: -186 🡺 38,886

Reduced to Sonoma and Napa valley 🡺 17,945

Title (wine name) blank: 0

Variety blank: 0

Winery Blank: 0

Vintage Blank: 1060

Taster names blank: 15313 🡺 23,573

Saved as Napa-Sonoma\_Wine.csv 17,945

Moved all other data sets out

Reduced the Addresses to Sonoma and Napa

Removed extraneous name

* , LLC – 11
* Winery - 306
* Wines – 174
* Wine – 506
* Vintners – 5
* Vineyards – 283
* Vineyard – 73
* CELLARS – 177
* CELLAR – 7
* CO. – 20
* COMPANY - 64
* ESTATES – 15
* ESTATES – 66
* Family