**Wine-Rating and Price**

Found this csv file on Kaggle.com and choose it for an excel project as it has a complex data and easy to understand. Original source is <https://www.vivino.com> and represent their vine database.

I want to explore the file to find insights. Create interactive charts to answer my questions.

Steps:

1. Will start by making a copy of the raw data.
2. Create a table for the entire data in a new sheet called Table1. This contains 8 columns and 4458 rows.
3. Now will inspect each column.

* ‘Name’ column: no empty cells ,interesting names as ‘И 2017’ which I assume is Ok. Notice that names include year, will leave as it is for the moment. If wish can easly be romeved with formula:

=TEXTBEFORE(A2,RIGHT(A2,5))

* ‘Country’ column: no empty cells, 30 countries, no cleaning needed =COUNTA(UNIQUE(Table1[Country]))
* ‘Region’ column: all looks fine
* ‘Winery’ column: all looks fine
* ‘Rating’ column: all looks fine. Rating scores from 2.5 to 4.8
* ‘NumberOfRatings’ column: all looks fine. Numbers starting from 25 up to 20293. This is interesting as there are no wines with under 25 ratings.
* ‘Price’ column: all looks fine. Prices start from 3.55 up to 3410.79 with unknown currency.
* ‘Year’ column: starts from 1998 up to 2019 with 8 rows having ‘N.V.’ No empty cells. Will leave as it is for the moment.

1. Check for duplicates.

* Firstly I checked for duplicates with same name and winery to find out there are same wines from same winery produced in different regions.

=CONCAT(Table1[@Name],Table1[@Winery])

=SORT(FILTER(Table1,COUNTIF(J2:J8667,J2:J8667)>1),1)

* Check for all columns

=CONCAT(Table1[@])

=COUNTIF(J2:J8667,J2:J8667)>1

* No duplicate rows

1. Create a new sheet called ‘Explore’
2. Q1 Which countries are available?

=UNIQUE(Table1[Country])

Converted cells to geography to get current statistics and their flags.

1. Q2 How many vineyards per country are available?

=UNIQUE(Table1[Country])

=COUNTA(UNIQUE(FILTER(Table1[Winery],Table1[Country]=M2)))

Then for data validation choose to count Romanian vineyards

=FILTER(Table1[Winery],Table1[Country]=M15) then paste values only to sort for easy count and compare

1. Q3 Which countries have the biggest selection of wines?

Created a pivot table with count of country

1. Q4 Which are the countries with the best average ratings? What are the prices?

Created a pivot table with average of ratings, total ratings, minim price, average price and maxim price

1. Q5 Which are the best rated wines and where they come from?

Created a third pivot table with name, region and country for rows and average of ratings for values.

1. Q6 I would like to see a boxplot chart for prices over the years.

Created new sheet ‘Prices’, inserted a Box and Whisker chart with ‘Price’ for Y values and ‘Year’ for horizontal category.

Sorted table by ‘Year’ column ascending to have same result in chart and make-it more readable.

Added slicers for Country, Year and Price for a dynamic chart.

Added a title, picture and a bit of colour for a nicer visual effect.

1. Q7 Are older wines better?

Created a new sheet called ‘Ratings’

Added the same theme as previous page.

Created a new pivot chart in ‘Explore’ sheet and from there a pivot chart with ‘Average of Rating’ and ‘Sum of Ratings’ on Values, ‘Year’ on Categories and ‘Values’ on Legend.

To answer the question will leave ratings of 4 and above on dark blue and for less than 4 will change in light blue.

Vivino.com offers 8,666 red wines from 2,719 vineyards in 30 countries. Italy and France have the biggest selection with 2,650 (30.6%) and 2,256 (26%) to choose from. Same name from same vineyard but different year are considered distinct.

You can choose red wines from 1988 up to 2019 with one exception 1994. Plus 8 noted with N.V.

Number of ratings start with 25 up to 20,293 and score from 2.5 up to 4.8 Highest rating received by 4 wines: ‘Veneto Alzero Cabernet 2009’ from Italy, ‘L'Ermita Velles Vinyes Priorat 2008’ from Spain, ‘Barolo Riserva Monfortino 2013’ from Italy and ‘Beckstoffer Las Piedras Vineyard Cabernet Sauvignon 2015’ from US.

On average Moldova and Lebanon received best ratings, both with 4.16 To consider that both are with 1 vineyard.

Cheapest is priced for 3.55, named ‘Shiraz’ it is produced in 2018 Spain. Most expensive is: ‘Pomerol 2012’ for 3410.79 followed by ‘Saint-Émilion Grand Cru 2010’ for 1599.95

As seen on ratings chart older years received better ratings scores.