<div align="center">

University of Chicago
Booth School of Business
41000: Business Statistics, Autumn 2021: Final Project.
Due: Dec 10 @ midnight on Canvas

Project: Predicting the Quality of Bordeaux Wines

This can be done as a group project

</div>

Many wines become better with age and their prices increase to reflect the higher value. Wines from some years mature into very good even excellent wines while others prove to be mediocre. There are significant profits to be made from identifying high quality vintages early when prices are low and before everyone realizes their quality. For example, early buyer's of the 1961 Bordeaux vintage made substantial profits as 1961 matured into one of the finest vintages in decades. Hoping for a similar result, many buyers purchase young wines before their quality is known.

Your project is to develop a statistical model to predict which vintages are likely to mature into high quality, highly priced wines. The data set bordeaux.txt and bordeauxp.txt on the course webpage contain market values and explanatory variables for red Bordeaux wines from different vintages or years.

The factors measured are:

variable	description
price	wine price index
sum	summer temperature (degrees C)
har	harvest rain, rain during the months August and September
sep	September temperature
win	winter rain, rain during the months October to March
age	vintage

The prices are for mature wines for the years up to 1980. These prices fully reflect the quality of these vintages. The prices are relative to the baseline of 1961 which is normalized to a value of 100. The data ranges from 1952-1991 and there are no prices from 1980 onwards as the latter wines have not matured. There is an extremely active market, however, for the young wines after they are bottled. People are willing to buy the young wines from the 1980's hoping that they will have matured into classic wines like the 1961 vintage.

Weather for the Bordeaux region is one of the main determinants of the quality of wine. Your data set includes age and the following variables to help you predict prices: Summer temperature, Harvest rain, September temperature and Winter rain.

Harvest rainfall is important because if it rains too much during the harvest season the wines will be too watery or too diluted. The better vintages have dry harvest periods and are said to be more concentrated. Summer temperature is also important because the hotter weather is necessary for the grapes to fully ripen. Riper, sweeter fruit produces a better quality wine. Experts closely follow the September temperature as they think that this is a good indicator of whether the harvest will be good or not. Winter rainfall is important because wetter weather is good for the grape vines early in the growing season. In France, climate change is already affecting the country's winemaking industry. French vintners say heat, drought and erratic weather are altering the landscape and their centuries-old way of working (web link).

There are, of course, a number of experts who claim they can subjectively assess whether a young wine will turn into a great vintage based on their opinions but without statistical analysis. These experts predicted that the 1986 vintage would be among the finest of the decade (report, Latour price graph). As young wines they were sold at similar, and sometimes higher prices to initial buyers than the wines of the other vintages of the past decade.

Using your statistical model for the data you have, predict the relative prices for wines in the 1980's. Based upon your analysis write a paper describing your statistical model and the role of the various factors in

predicting prices. Be sure to describe the use and outcome from any statistical tests you perform. Use your analysis to examine whether your statistical analysis provides and different predictions from the so-called experts.

Bordeaux Wine in 2017, 2018 and 2019

The most recent years 2017, 2018 and 2019 appear to be quite exceptional. You'll need the following input variable

For 2017, the winter rain, summer temperature, September temperature and harvest rain variables are 521, 19, 19, and 86.2, respectively.

For 2018, the winter rain, summer temperature, September temperature and harvest rain variables are 593, 20, 22 and 84.9, respectively.

For 2019, the winter rain, summer temperature, September temperature and harvest rain variables are 550, 21, 19 and 110, respectively.

Predict prices for 2017, 2018 and 2019.

Keep in mind the data analytic tools we've covered and address the following questions:

(a) Train a multiple regression model on the dataset from 1952-1980.

(b) Are any wines outliers or influential in your analysis?

(c) Use your model on the out-of-sample dataset and predict prices for 1981-1991.

(d) Is 1986 a superior vintage as the experts seemed to think?

(e) Is September temperature statistically more relevant than summer temperature?

(f) 1983 and 1989, 1990 are known to be great vintages in hindsight. Does your model agree with this statement and how do they compare to 1961?

Your paper should be between $5 - 10$ pages including exhibits. You will be graded on presentation, writing and analysis.