Presentation Guide

Slide 1(Nyarai) – Introduce our Project and group members

Slide 2(Nyarai) – California is very known for its wine and we are curious whether they have the best wines based on this speculation (seen lots of advertisements, commercials, news reports, etc)

* Read hypothesis and explain wine review relative to overall cost, which means best rating for its price or best balance between the two factors we have in judging quality of wines

Slide 3(Nyarai) – Introduce questions on the slide and explain that these are the questions we came up with in order to find evidence to support the criteria of “best”.

* Does California produce the highest quality (professionally rated) wine compared others? (This question helps us use one factor as comparison against other wines as part of the criteria)
* How does California wine reviews compare to the rest of the world when it comes to the variety of wines reviewed? ( This question will explain the diversity of California wines compared to other wines)
* Does spending more get you better wine? ( This question will explore the relationships between price vs reviews and give a bigger perspective on the criteria used in judging wine)

Slide 4(Nyarai) – Tell them we got the data of wine reviews from Kaggle and we narrowed down to these 3 questions to help answer broaden the perspective of our motivation and summary.

Slide 5-7(Chris) – Chris will explain the data cleanup and exploration for 3 minutes

Slide 8(me) – I will explain the breakdown of the total number of reviews and the proportion and that US and California made up out of the total number of reviews and how this particularly made us choose California as our interest.

Slide 9(me) – Explain the two histograms we have and compare review ratings between California and that found globally

Slide 10(me) – Explain differences in variety count and the sheer number of varieties found in CA that is found elsewhere. (Show pie chart in jupyter notebook of top 20 varieties of wine in CA) as a comparison to the pie chart in slide)

Slide 11(me) – Explain the relationship of price vs review and the correlation coefficients found. (Possibly show violin plot in jupyter notebook to broaden perspective)

Slide 12 (me) – Explain we did not find statistical differences based on the chi square test

Slide 13(Audelia) – Discuss findings and what we learned(e.g. most wines fell under $250; California wines were more expensive with only a marginally higher review than other wines; huge proportion from USA; CA is known for zinfandel, but not as well known by rest of world; overall California did not have the best wine vs review rating despite the high # of varieties and wine count)

Slide 14 (Audelia)- Discuss difficulties with dataset(e.g. git branch and merging commits, could not get specific coordinates to do heat map of wineries in case any enthusiasts wanted to go on a winery tour, couldn’t find enough good data sources, dataset was limited to price and review ratings as given metrics)

Final Slide- Thank the audience for listening and ask if they have any questions.