Engineering Big Data Systems Final Project Report

Riddhi Kakadiya

Nuid 001811354

Data Set : Wine Reviews, 130k wine reviews with variety, location, winery, price and description

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

Map Reduce Analysis:

1. Number of reviews given by each reviewer.
2. Distinct titles of the wines.
3. Distinct countries that the wines were produced.
4. Average rating of wines per variety.
5. Average price of wines per variety.
6. Wineries each country has.

Analysis using Pig:

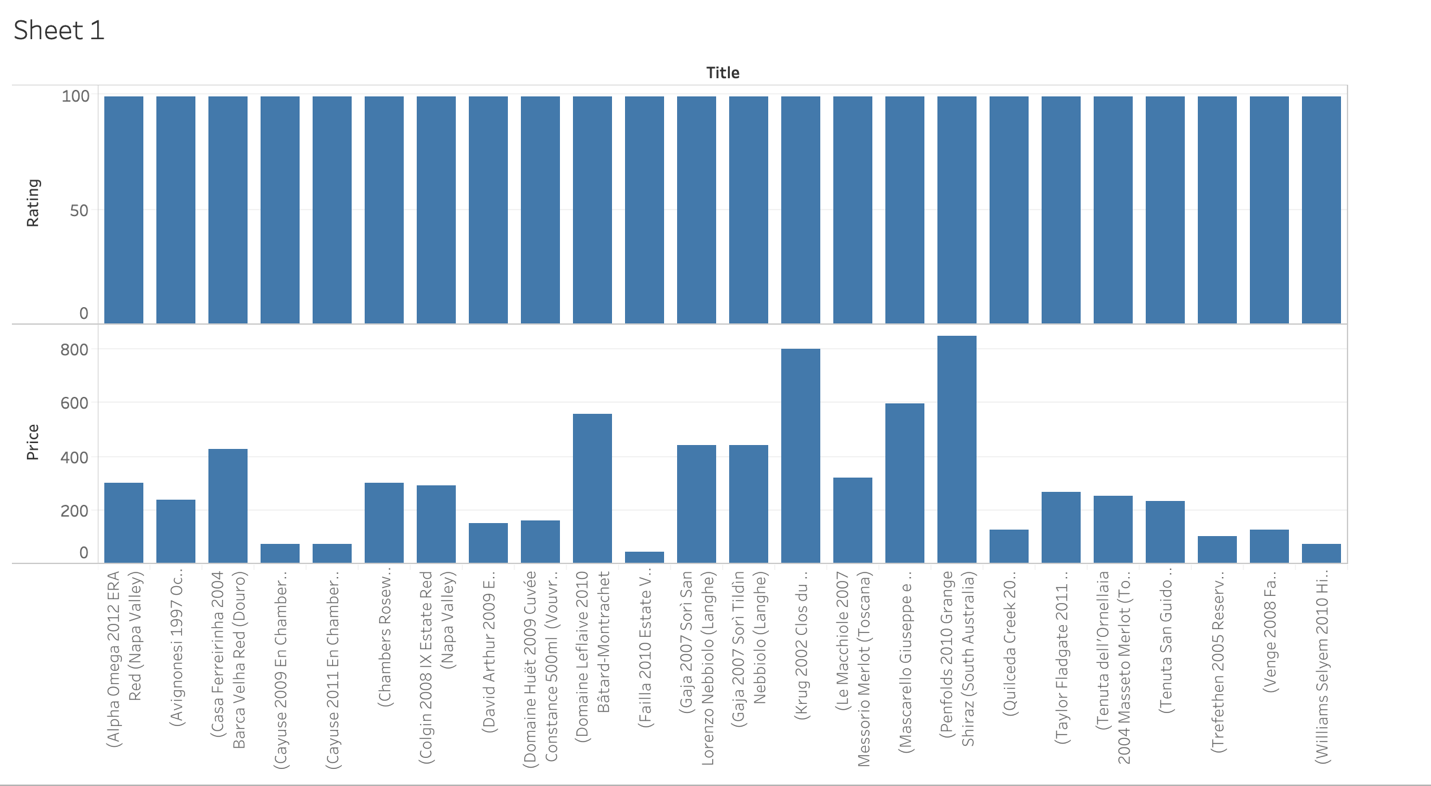
1. 30 cheapest wines.

2.Top 25 rated wines that are have least prices.

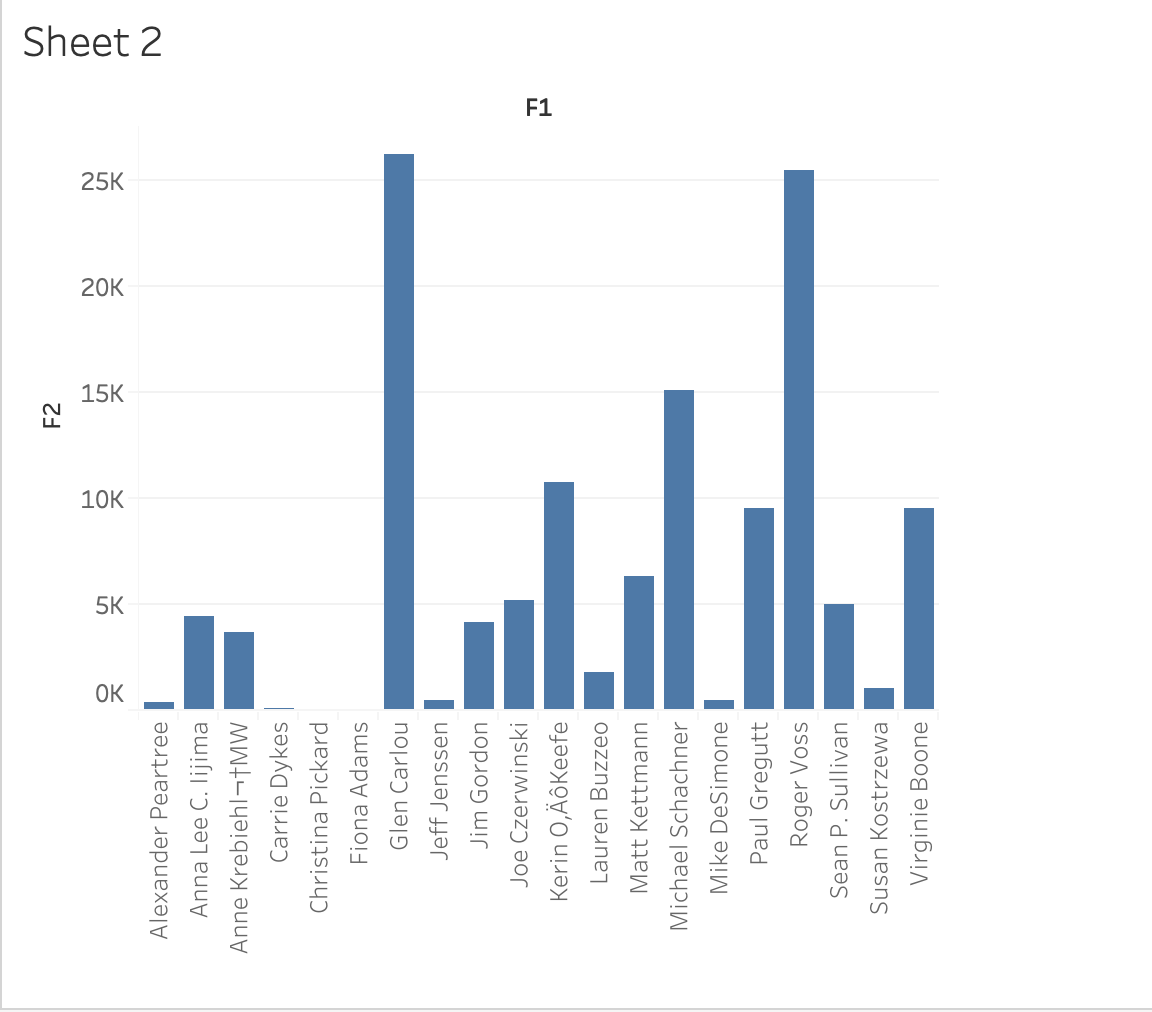
3.Cheapest 25 rated wines that are have highest ratings.

Visualization tool used: Tableau

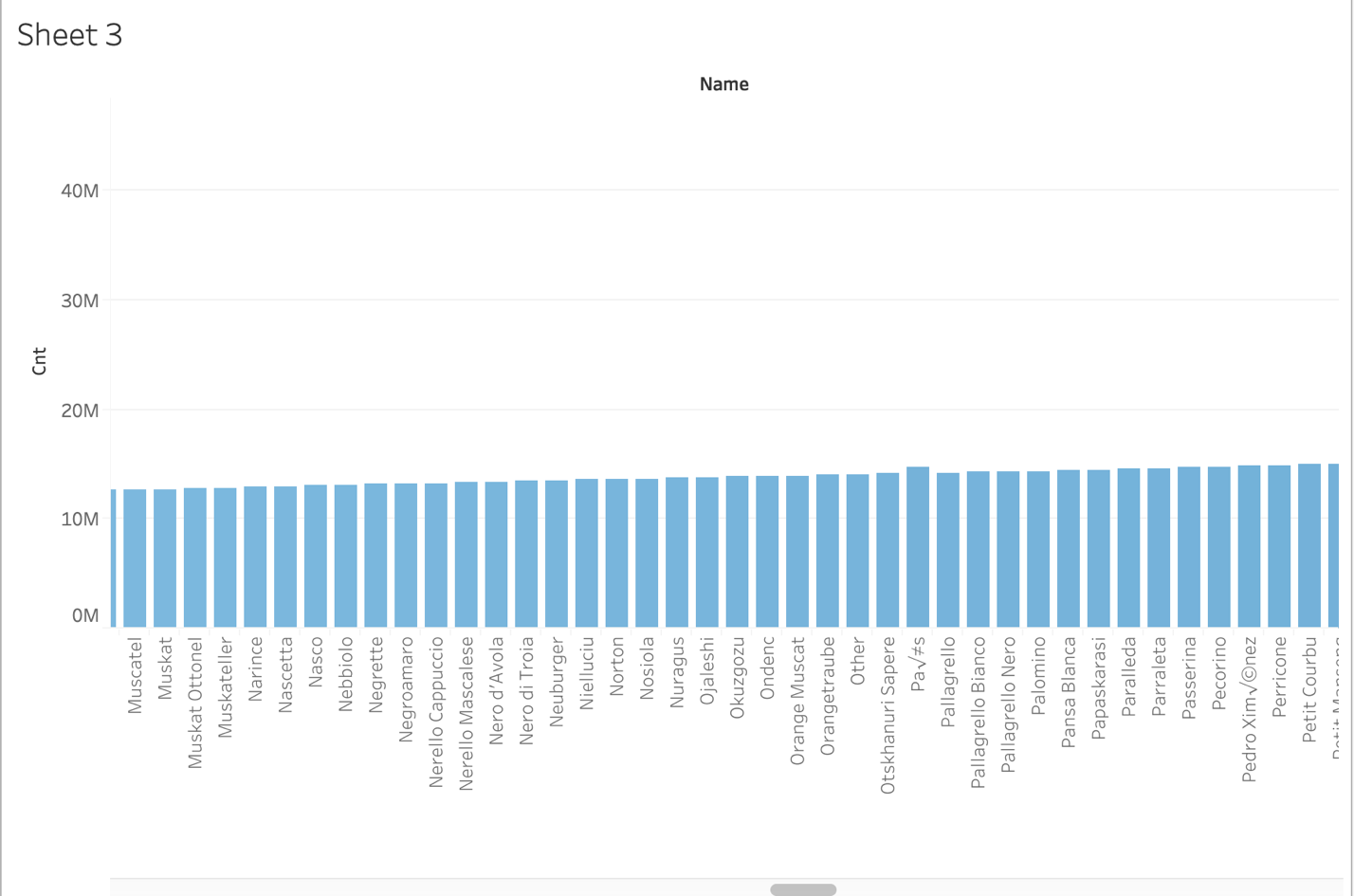
1. Top rated wines and their prices



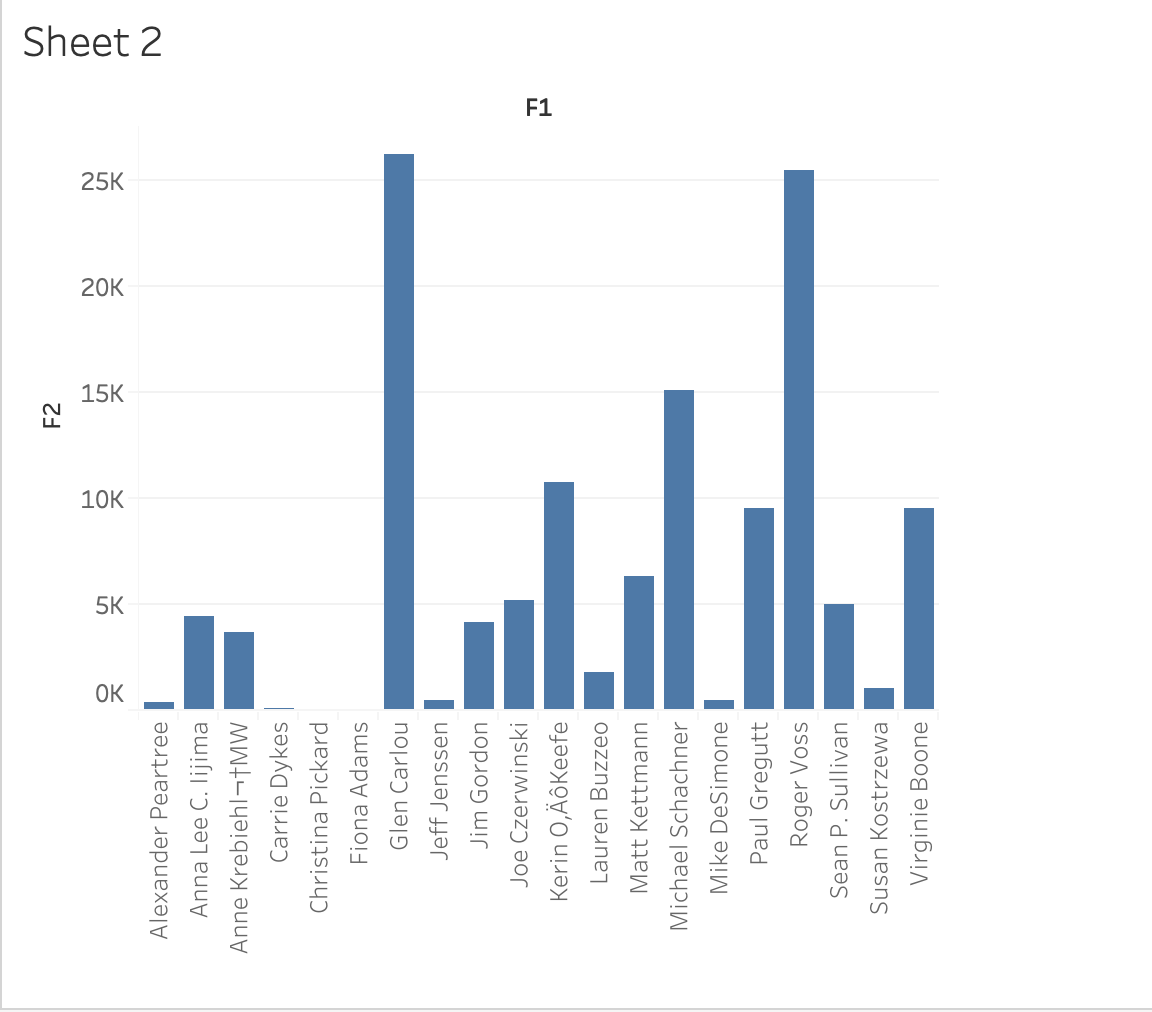
1. Reviews given by each reviewer



3.Average ratings of each wine



1. Reviewers and their reviews



1. Average prices of wines

