# Understanding and Visualizing Data

**Course Project**

**Instructions:**

This project guides you through the process of applying a data decision framework to an important decision-making situation in your organization or career.

*Once you have completed all three parts of the project, submit this project document and any supporting documents to your instructor for grading. A submit button can be found on the Part Three assignment page. Information about the grading rubric is available on any of the course project assignment pages online.* *Do not hesitate to contact your instructor if you have any questions about the project.*

#### **Part One – Draft a Data-Collection Plan**

In this part of the project, you will identify a situation that requires you to make a decision, identify what data you will use to inform the decision, and draft a plan to collect the data you need.

In order to satisfy this part of the project, you produce a plan that answers the questions listed in the following three steps.

Step 1: **Identify a decision that requires, or would be enhanced by, data analysis.**

**What is the situation?**

My wife and I own a wine shop that has been doing very well and want to expand our wine selection to meet the rising demand for international wines. We have data from last years sales and ratings on the variety of wines we have brought in from California and the ratings given by the professional taster from from many international wineries.

**What are the parameters or options in the decision?**

We only have space in our shop floor for ten more wines in our area we are remodeling and we need to keep our price point below $200 so we will only include data about the wines from International wineries under two hundred dollars.

**Who are the key stakeholders?**

Our stakeholders are myself and my wife and our loyal customers.

**How do you hope or expect data to help illuminate your decision?**

Using the data from known the taster reviews, our customer ratings and the points/$100th we should be able to select the best options for our shop to bring in.

Step 2**: Identify data that will help you better understand the situation.**

A data set based on Wine Enthusiast (winemag.com) combined with our personal sales records

**What are the key performance indicators for your situation?**

Projected Revenue and the Grand total sales.

**What defines the range of cases you will consider?**

Price under $200 USD, ratings over 7.0, country of origin outside US

**What are the variables you will consider?**

Winery, title, price, customer ratings, points, variety, a new feature created by dividing the points by the number of hundreds of dollars

**Is each variable categorical or quantitative?**

Price, points, points/$100th and rating are quantitative rest are categorical

**What purpose does each variable have in informing your decision?**

Region of origin determines where it came from allowing us to select wines by the natural growing conditions of the wine grapes.

Price determines how much it will cost to purchase and our margin for resale

Rating/points the quality/favor of the product

Points/$100th allows us a third a third variable to compare the international wines and known success of wines the known success of wines in our store.

Step 3: Develop a data-gathering plan. (100-250 words)

Where will the data come from?

Is it observational or experimental data?

Who will collect it?

How much data will you need (sample size)?

How will you assure that it is representative of the population?

What steps will you take to mitigate potential bias?

The data will come from a Kaggle dataset we have found based on Wine Mag’s Wine enthusiast section. The data is observational as it is the objective opinion of the wine taster expert. It has been collected into a repository on Kaggle. I will use all 130 K entries and. I will eliminate the bias created by the taster by incorporating a customer rating of all customers who came to our wine tasting event over the last summer.

#### **Part Two – Identify Data Summaries and Visualizations**

In this part of the project, you will identify the statistical summaries and visualizations you believe will best help you make your decision.

In order to complete this part of the project, you need to answer the following questions.

What summary statistics will you use to inform your decision?

We will narrow our search by filtering out all tasting reviews where the price is $200 or more.

We will use the average rating for each variety of wine and then by the average rating for a winery and finally a specific wine.

The average points for a location will be used as well.

We will filter the wines first by being less than $200 then by points per $100th and we will filter for wines included in the highest rating and highest sales.

Are you interested in statistics that are sensitive to or resistant to outliers, and why?

I am interested in statistics that are sensitive to outliers, the more outliers I have the less likely my average and a few other

What visualizations will you use to inform your decision?

Be sure to indicate what variables and what scale you are using for each visualization (e.g. histogram showing frequency of page hits per hour over a 24-hour period)

If you have data related to your project, create at least one of the visualizations you have listed above and include a copy of it here.

**Note:** Though your work will only be seen by those grading the course and will not be used or shared outside the course, you should take care to obscure any information you feel might be of a sensitive or confidential nature.

#### **Part Three – Data and Your Decision**

Part three is the culmination of your project. In this part of the project, you will make a determination about whether the data you have or planned to have is going to be sufficient to make a good decision. If you have data related to your project, you will create an interactive dashboard to help you analyze the data. If not, you will create a mock-up of the dashboard.

In order to complete this part of the project, you need to answer the following questions and present either a working dashboard or a mockup of a dashboard.

Part a: Questions about your data-model-insight framework

What are you attempting to model with your data?

I am attempting to model the best wines for bringing into our shop based on the customer ratings, taster points, points/$100th and overall sales from known wine varieties.

What are the KPIs for the situation you are trying to understand?

Points/$100th, price and points.

What is the relationship between your variables and the KPIs?

The wine varieties from the international wineries and their taster points, price, have a correlation with the prices, customer ratings, taster points and the points/$100th of the shops current wines.

What are the limitations of your model?

We are limited to comparing only the wine varieties and wineries that the taster has sampled. We also are limited since we have no customer rating data that overlaps both the shop data and the international data.

Do you feel your model, as defined, is “good enough” to inform your decision? Why or why not? I do think the model we have created is capable of providing comparable wines to those our customers already have indicated they like.

Part b: Your Project Dashboard

If you have data for your project, you may be able to put together a dashboard in Excel that is a good working model. If so, include an Excel workbook as part of your project submission.

As an alternative, you may turn in a mockup that indicates what elements you would like your dashboard to include.

In either case, your dashboard or mockup should include

* a readout of KPI’s, clearly labeled as such
* one or more visualizations
* summary statistics, as needed

Indicate how the elements of the dashboard are connected. In a mockup, this could be arrows drawn between different elements on the dashboard. For a dashboard with actual data, you might include this information as text.