

---

Is this Scotch?



# A Little Domain Knowledge

---



# Whiskey

## Malt Whiskey

Irish Whiskey

Asian Whiskies

Scottish Whisky = "Scotch"

## Grain Whiskey

Canadian Whiskey

American Whiskey

Bourbon





# Can I Identify Where a Whiskey Comes from?



---

## The Data...

Downloaded as CSV from [Whiskeyanalysis.com](http://Whiskeyanalysis.com)

1. 1611 observations, each representing a different whiskey.
2. 8 different features
  - a. Review average, std. Deviation, number of reviews, type, class, flavor, super flavor, country

## Post EDA

1. 1344 observations
2. 3 different features
  - a. Type, Class (reclassified), Flavor





# Flavor Profile



# Features – Flavor

- A – Full-bodied, sweet, pronounced sherry – with fruity, honey and spicy notes
  - B – Full-bodied, sweet, pronounced sherry – with fruity, floral and malty notes, some honey and spicy notes may be evident
  - C – Full-bodied, sweet, pronounced sherry – with fruity, floral, nutty, and spicy notes, some smoky notes may be evident
  - E – Medium-bodied, medium-sweet – with fruity, honey, malty and winey notes, some smoky and spicy notes may be evident
  - F – Full-bodied, sweet and malty – with fruity, spicy, and smoky notes
  - G – Light-bodied, sweet, apéritif-style – with honey, floral, fruity and spicy notes, but rarely any smoky notes
  - H – Very light-bodied, sweet, apéritif-style – with malty, fruity and floral notes
  - I – Medium-bodied, medium-sweet, quite smoky – with some medicinal notes and spicy, fruity and nutty notes
  - J – Full-bodied, dry, very smoky, pungent – with medicinal notes and some spicy, malty and fruity notes
- 



---

## Other Features

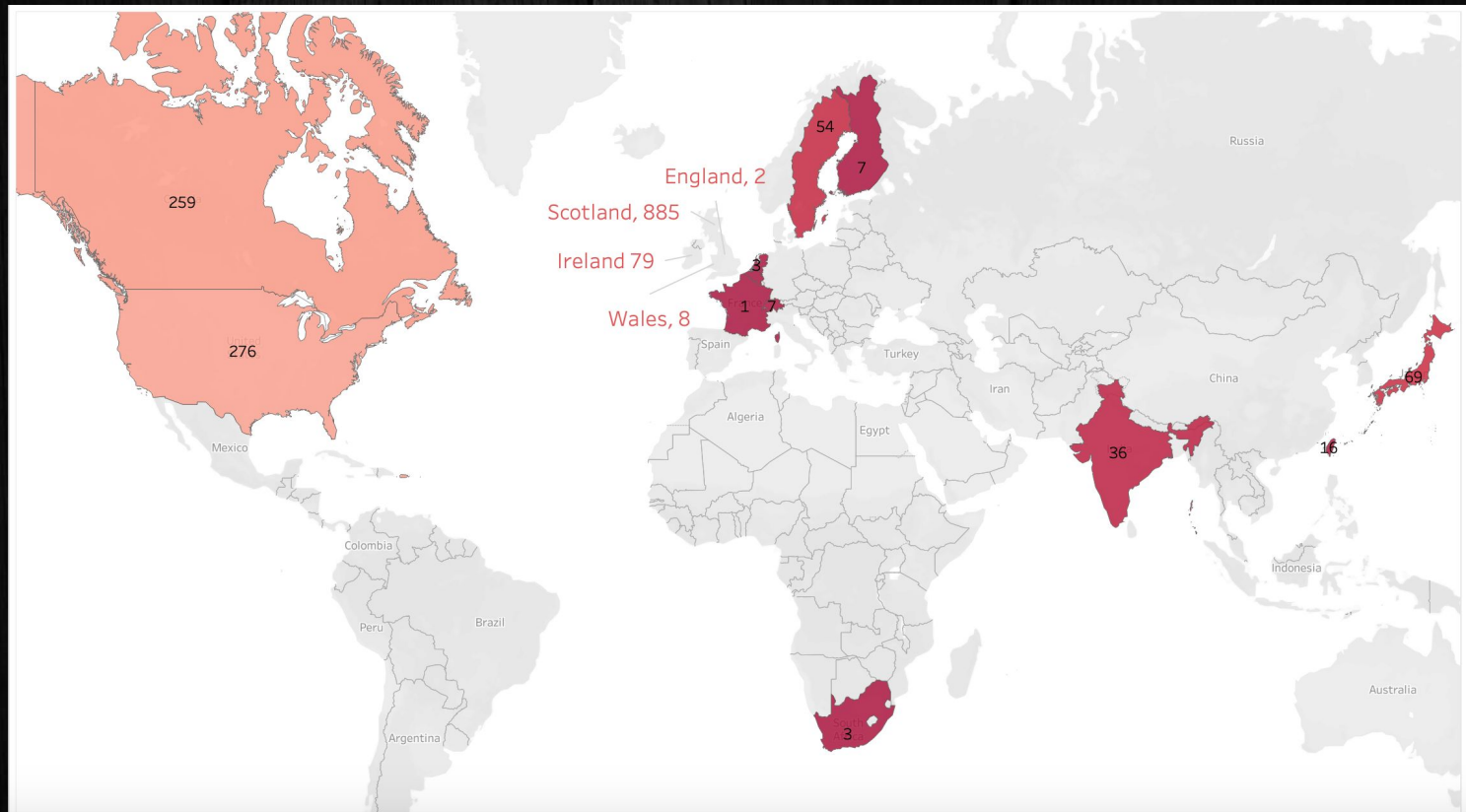
Class - Single-Malt like, Blend, Rye-like, Bourbon-like

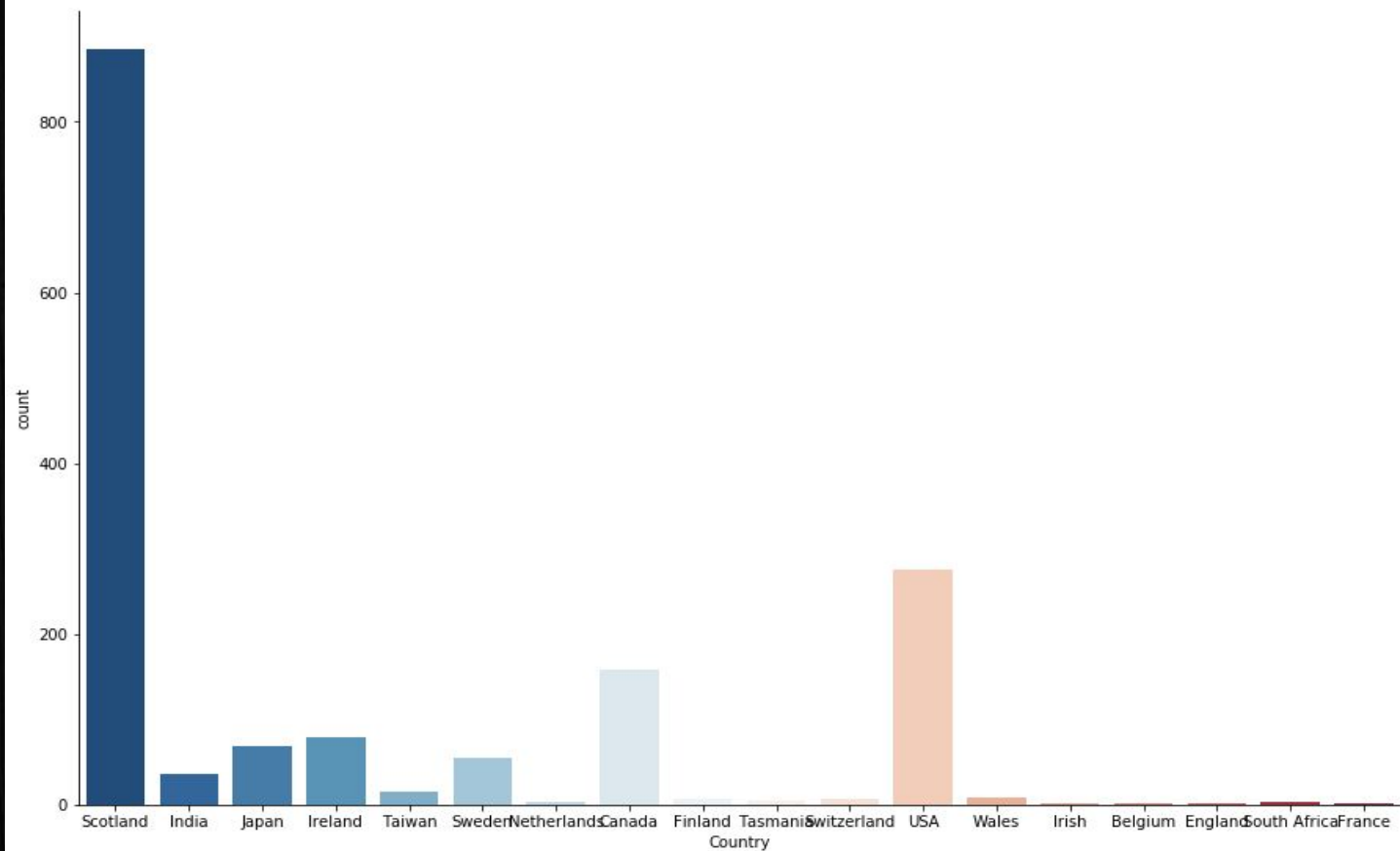
Type - Malt, Blend, Grain, Rye, Wheat, Barley, Bourbon, Flavored



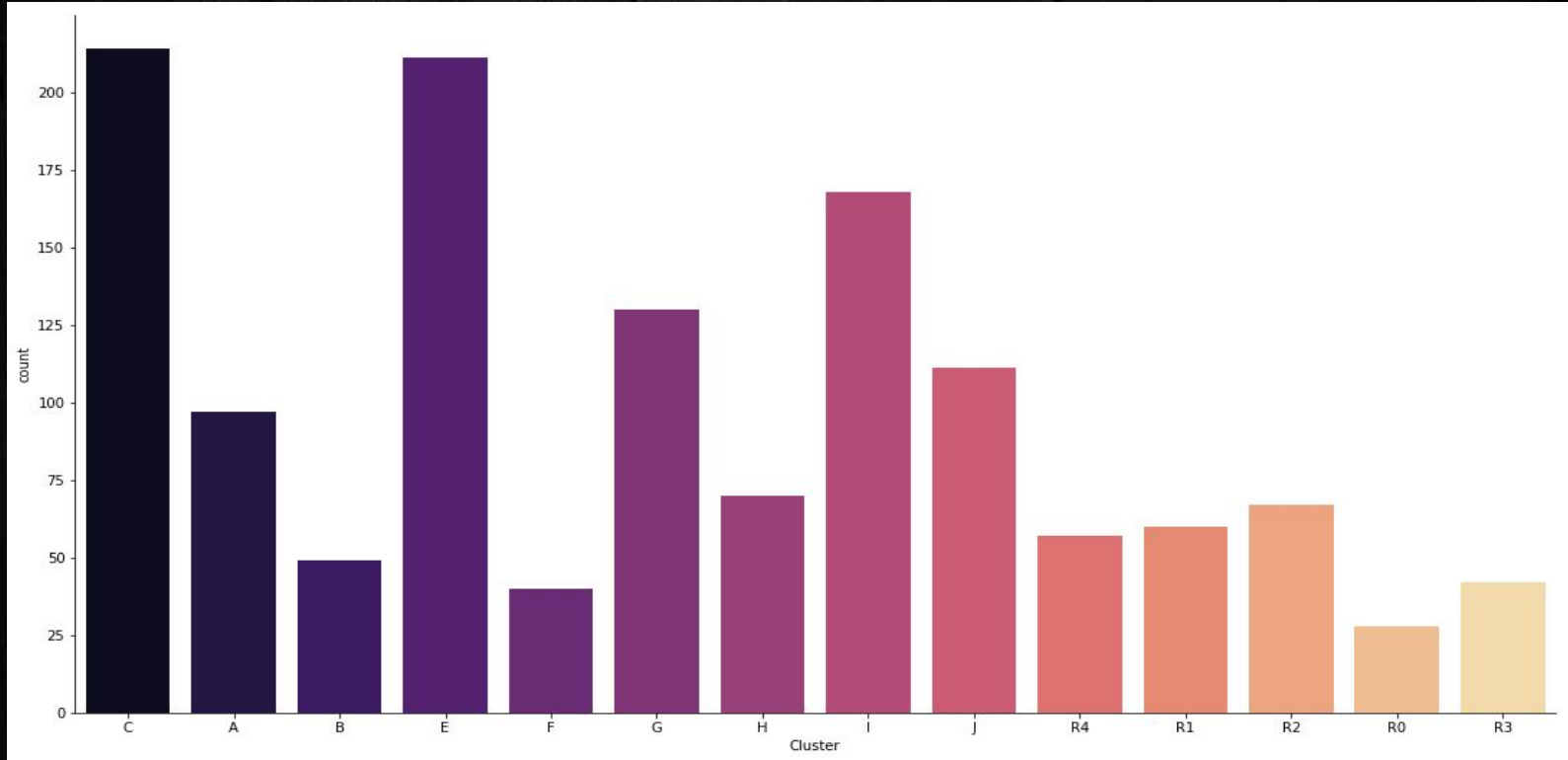


# Maps

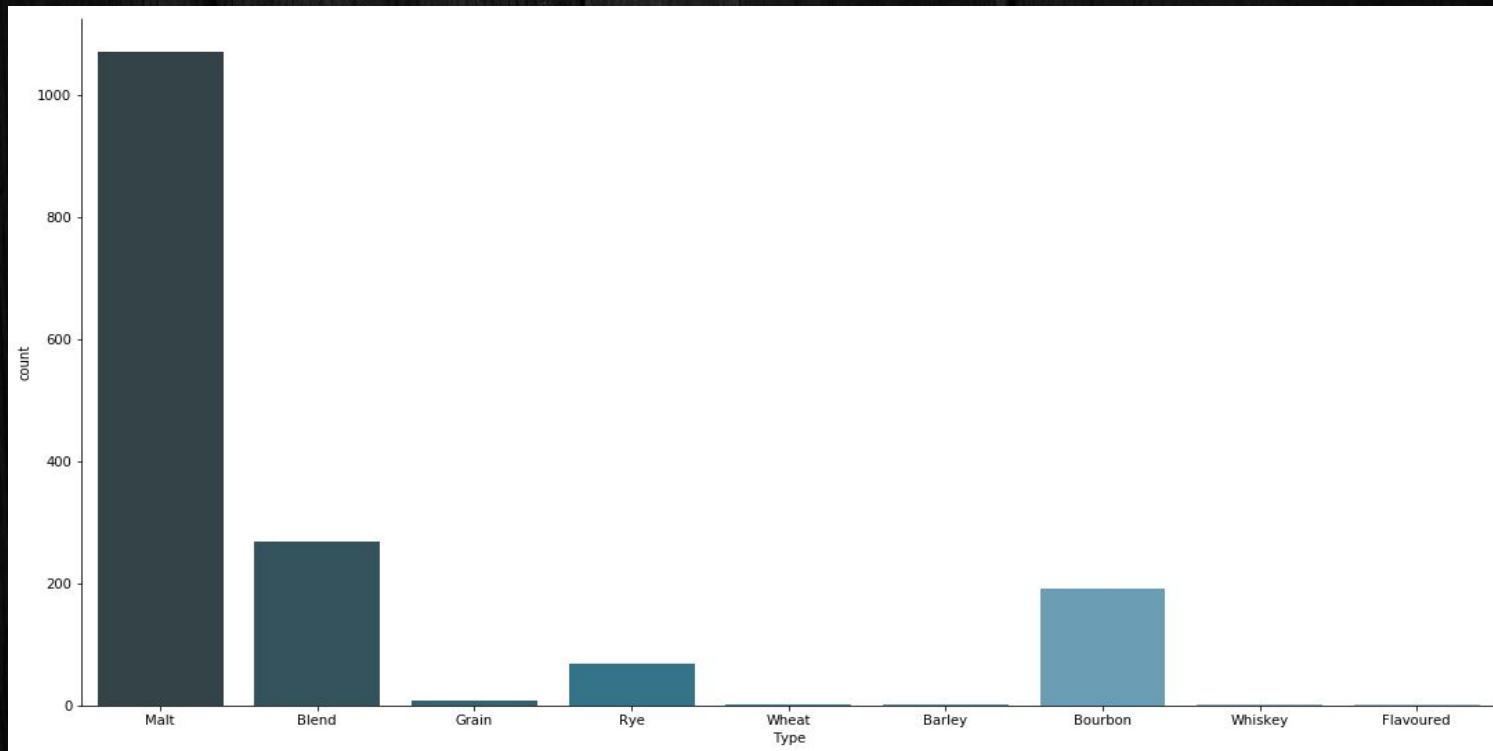




# Flavors/Clusters



# Type







# Start Your Models...



---

# Our Baseline Model

DummyClassifier – Strategy: Uniform

Precision Score: 15%

Recall Score: 23%

F1 Score: 12%

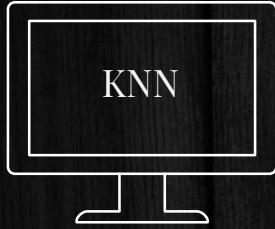
Simple Binary Classifier – KNN – Scotch or Not?

Precision Score: 94%

Recall Score: 94%

F1 Score: 94%

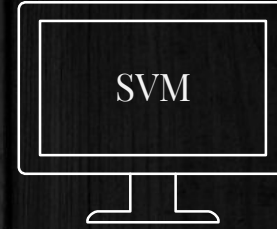




Precision Score: 42%  
Recall Score: 39%  
F1 Score: 39%



Precision Score: 36%  
Recall Score: 38%  
F1 Score: 29%



Precision Score: 40%  
Recall Score: 44%  
F1 Score: 35%



Precision Score: 44%  
Recall Score: 40%  
F1 Score: 35%



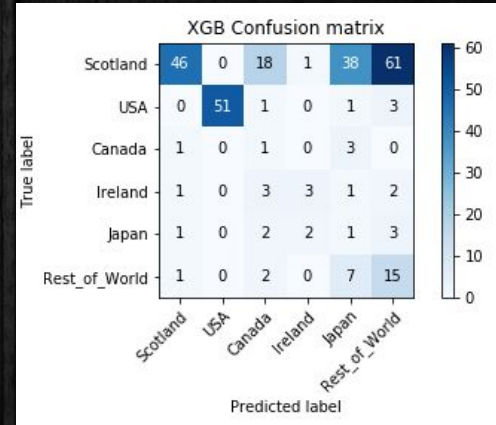
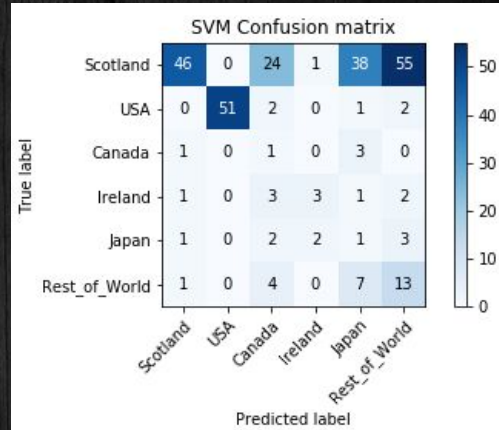
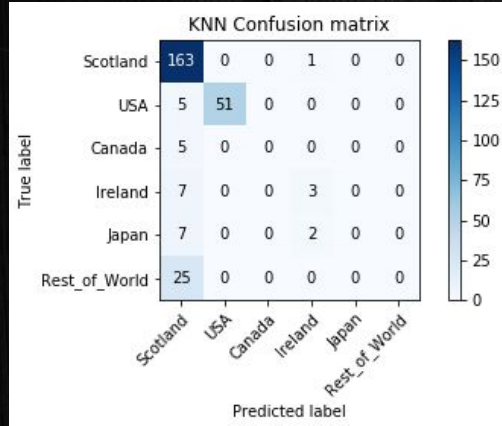
Precision Score: 36%  
Recall Score: 30%  
F1 Score: 25%



Precision Score: 37%  
Recall Score: 36%  
F1 Score: 27%

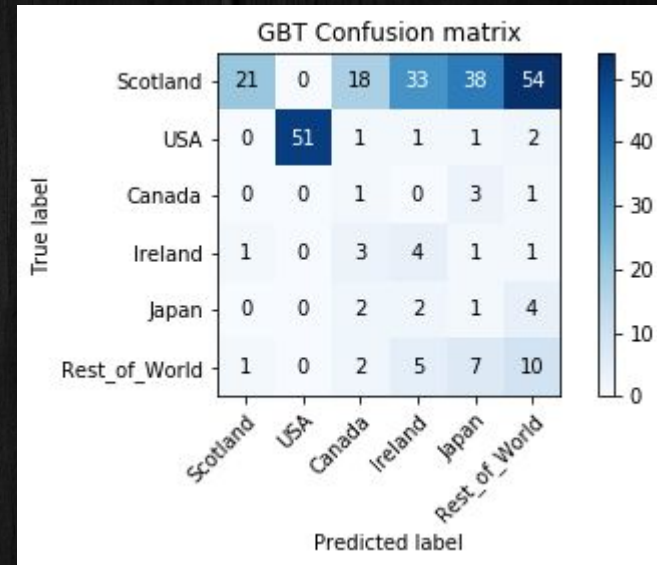
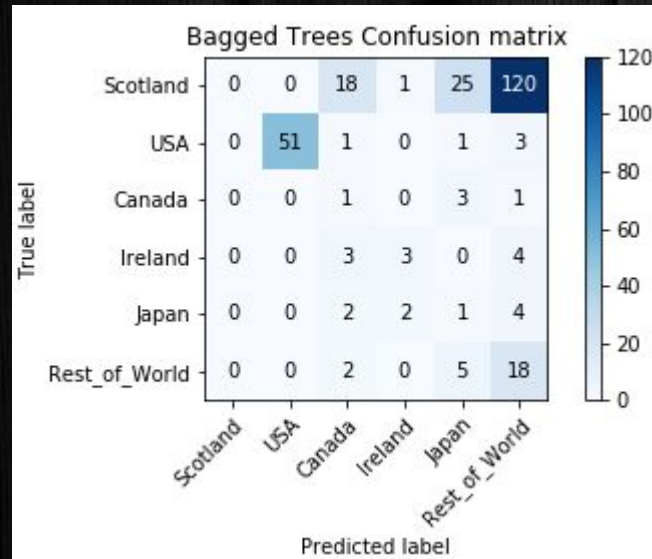


# Confusion Matrices





# Confused Confusion Matrices



---

# Conclusions

## SVM - Ideal Model

- Highest precision and recall score.
- Homoscedastic - errors evenly distributed
- Type 1 errors are desirable over Type 2 errors

