"The Perfect Pour"

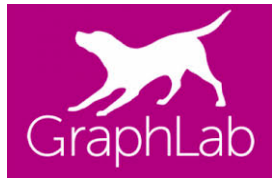
Using Machine Learning to Build a Wine Recommender

Bob Mickus

Workflow

Data Sources → **Data Collection** → **Data Munging** → **Data Modeling**

reg[ular]
expr[essio]n



Exploratory Data Analysis

Wine Data:

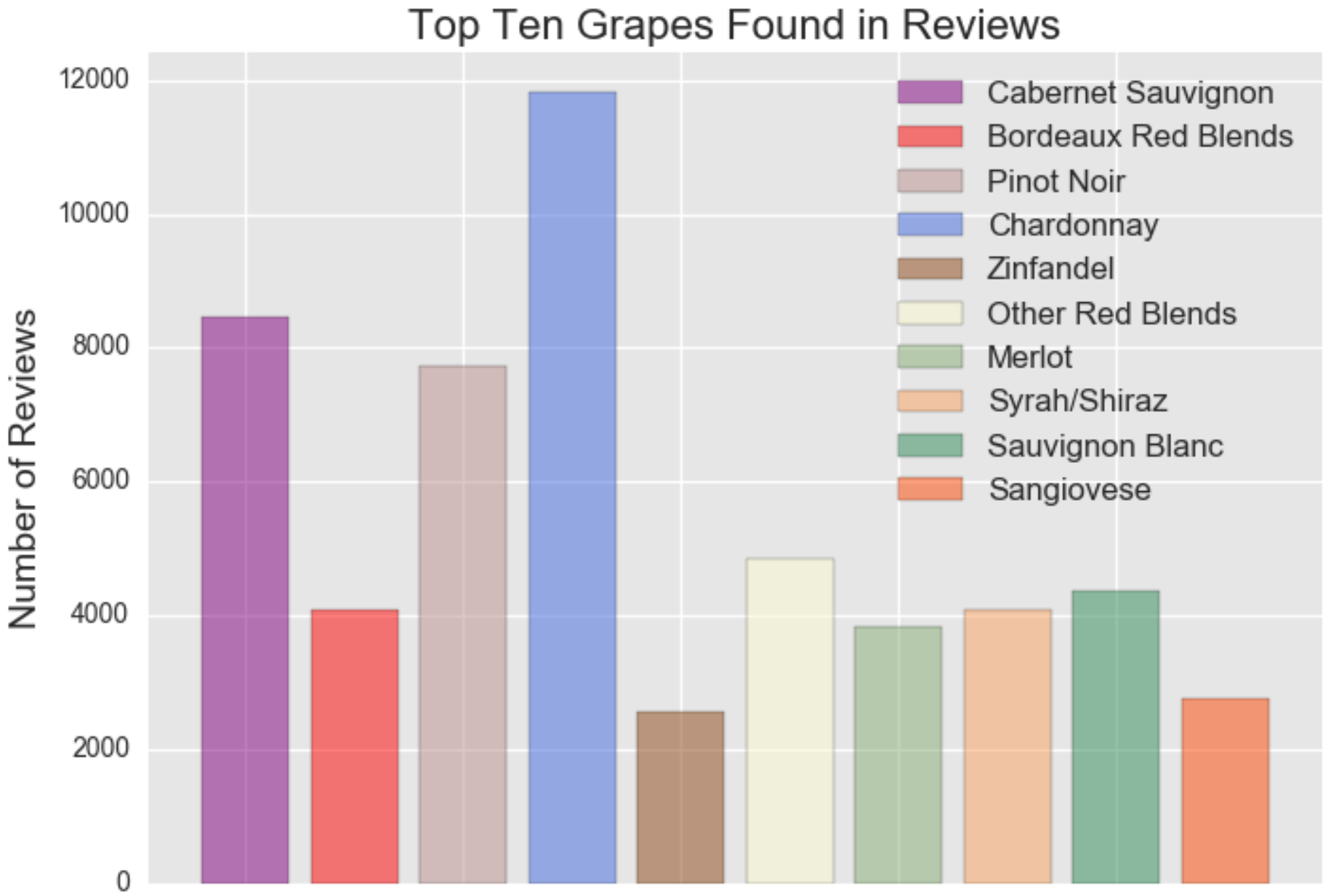
Num Wines:
78,076

Price Range:
$1.98 – $12,429

Wine Type:
Red: 61%
White: 35%

Grape Varietals:
164
Chard: 15%, Cab: 10%

Regions:
109
Napa: 10%



Exploratory Data Analysis

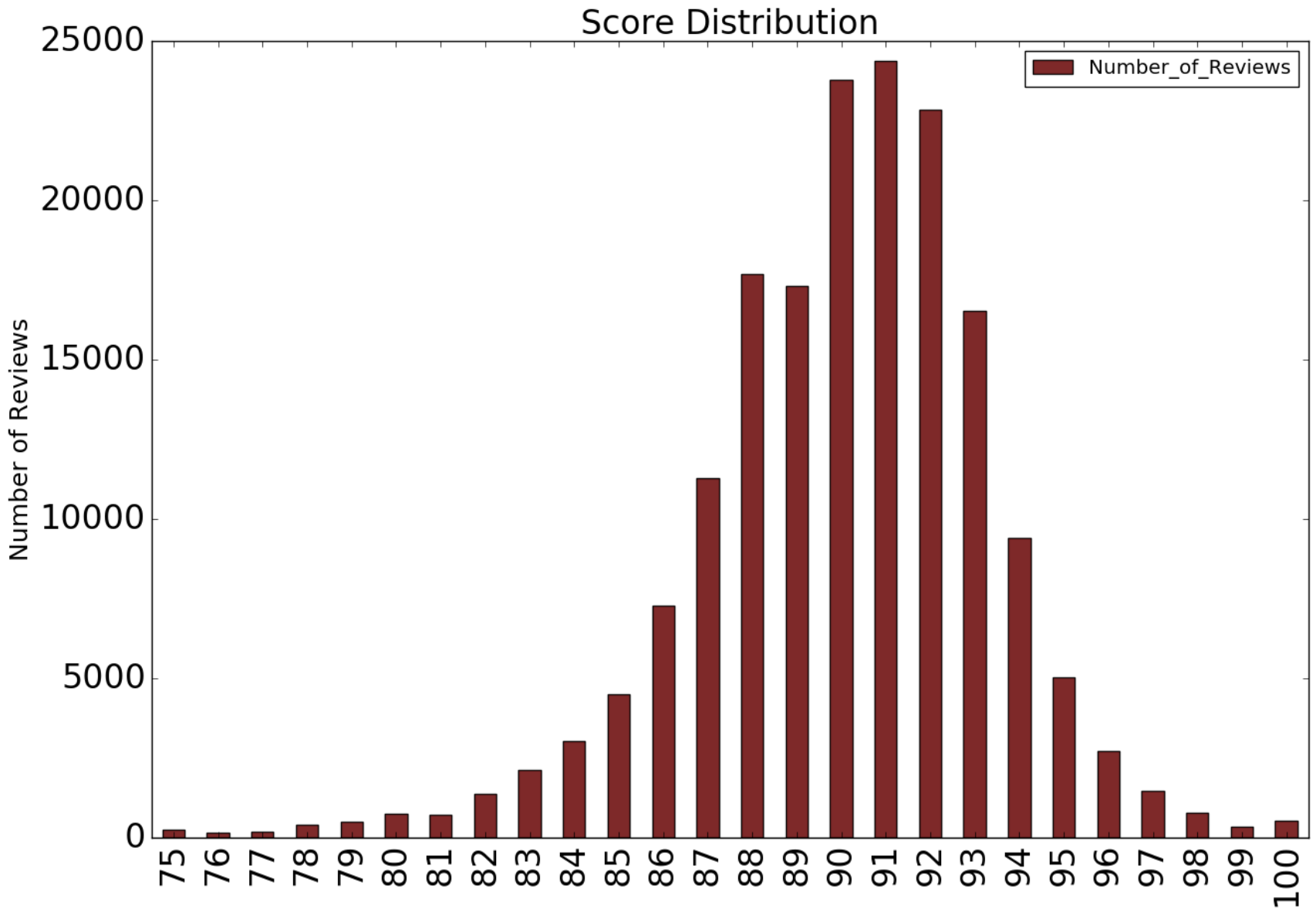
Review Data:

Num Reviews:
211,452
83% had a rating

Num Unique Users:
334

Num Wines Reviewed:
95,264
61% only had one review

Rating Range:
0 – 100
Mode = 91
Mean = 89.85
Median = 90 σ = 5



Building the Recommender Models

I. Content-Based Filtering:

If you like an item then you will usually also like a “similar” item based on similarity of the items being recommended (rather than learned from the interaction data)

Similarity_type: cosine
Number of wines: 78076
Features: Wine Name, Vintage, Grape Varietal, Region/Location, Price, Avg Critic Rating

Predict on:

Grape	Location	Wine Name	Price	Avg Score
Cab Sauvignon	California	Freakshow Cab - 2014	20.99	90

Recommendations:

Grape	Location	Wine Name	Price	Avg Score
Cab Sauvignon	California	Smoking Loon Cab - 2014	8.99	85
Cab Sauvignon	California	Layer Cake Cab - 2014	14.99	90
Cab Sauvignon	California	Carnivor Cab - 2014	11.99	90

Building the Recommender Models

II. Collaborative Filtering: Item – User Similarity

Build a model to find users who share the same rating patterns with the user whom the prediction is for.

	Wine 1	Wine 2	Wine 3	Wine 4
Bob	91		?	89
Jake	90		92	93
Frank		82		
Kayla	92			90

Algorithm based on:
- -K-nearest neighbors
- -Pearson correlation

Recommendations:

user_id	wine_name	score	rank
bob_mickus	2003 E. Guigal Côte-Rôtie ...	98.5	1
bob_mickus	2007 Maybach Family Vineya...	98.25	2
bob_mickus	1959 Château Lafite Rothsc...	98.0	3
bob_mickus	1989 Domaine Huet Vouvray ...	98.0	4
bob_mickus	2001 Pride Mountain Vineya...	97.3658929312	5
bob_mickus	1976 Joh. Jos. Prüm Wehlen...	97.25	6
bob_mickus	2009 Saxum James Berry Vin...	97.25	7
bob_mickus	1996 Domaine Leflaive Chev...	97.0	8
bob_mickus	1989 E. Guigal Côte-Rôtie ...	97.0	9
bob_mickus	1997 Romano Dal Forno Reci...	96.8333333333	10

"a full-bodied, concentrated wine that has sweet tannin, excellent mid-palate depth and a terrific finish…."

Model Comparisons

RMSE (on predicting 0 – 100 score):

Model 1:
Item – User Similarity 12.4

Model 2:
Default Matrix Factorization 3.8

Model 3:
Hyper-parameters Optimized on MF 2.3

Model 4:
Include Price as a Feature 1.7

Precision (on predicting a good wine)

Model 5:
Logistic Regression Classifier .997

Next Steps...

Next Steps:

- Conduct sentiment analysis on the text comments contained in the reviews and incorporate into the model
- Continue to download more reviews
- Build front-end UX
- Integrate the Perfect Pour with the Matrix
- ”Test” some more wines!

AlchemyAPI

Can I recommend a wine for you?

Bob Mickus

Email: bob@mickus.com
LinkedIn: www.linkedin.com/in/bobmickus
Project Repository: www.github.com/bobmickus/perfectpour