Nic's Cleaning and ML Perspective

Modeling by description_length

qualityPrediction.ipynb | ranking_words.csv

Code

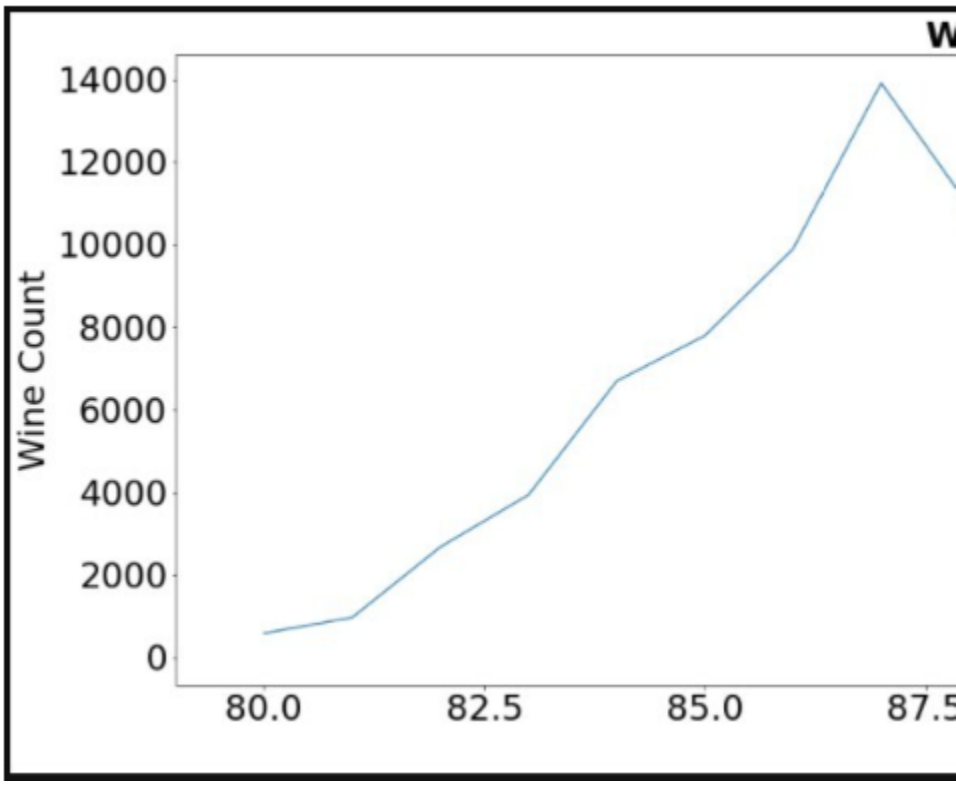
Sentiment Analysis of Wine Data

- Can we use Machine Learning to predict the quality of wine based on
- Dataset: https://www.kaggle.com/zynicide/wine-reviews

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer, Tfi
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, classification_repo
```

Clean Data

- Read CSV for starting data
- drop nulls and duplicates from description and points columns
- drop columns down to description and points



Description length vs points

- Add description length column

[7]:
```
# Add description length column
description_points_df = description_points_df.assign(description_leng
description_points_df.head()
```

[7]:

	description	points	description_length
25	Yields were down in 2015, but intensity is up,...	94	215
29	This standout Rocks District wine brings earth...	94	333
60	Concentrated, ripe blackberry and cassis aroma...	91	242
61	Moorooduc's estate Pinot Noir is a solid value...	91	315
62	Smoky aromas of fresh-cut wood blend with berr...	91	229

K.I.S.S - simplify the points distributio

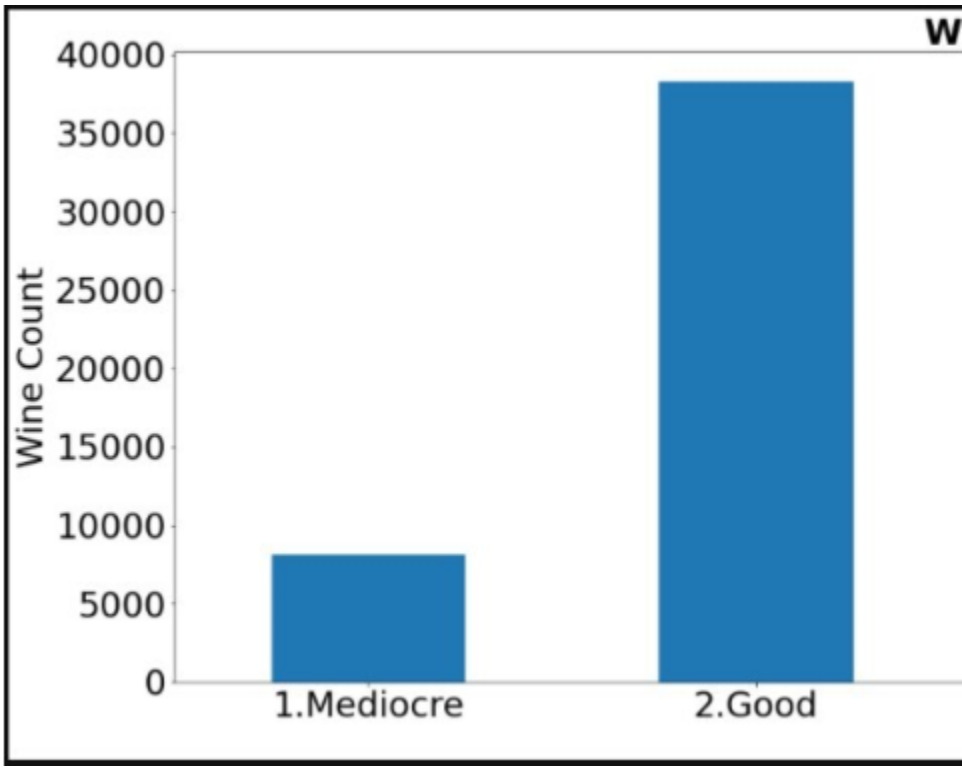
- Keep it simple stupid
- 1 -> Points 80 to 84 (Mediocre)
- 2 -> Points 84 to 88 (Good)
- 3 -> Points 88 to 92 (Very good)
- 4 -> Points 92 to 96 (Outstanding)
- 5 -> Points 96 to 100 (Classic)

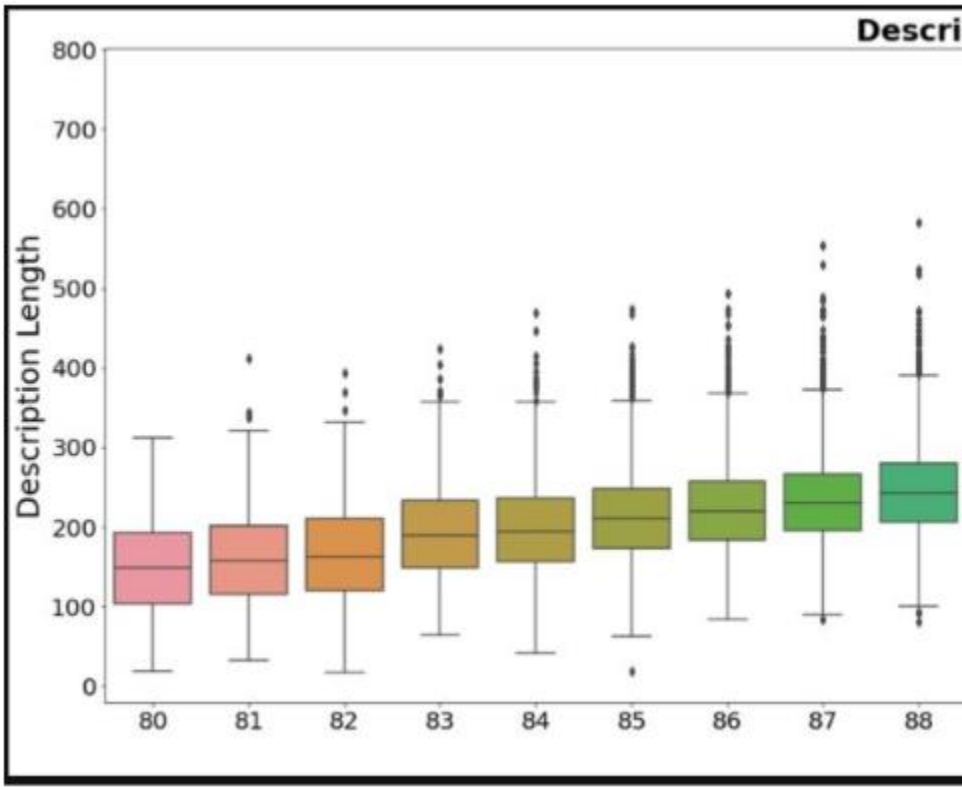
[9]:
```
# simplify the points data
def getQuality(points):
  if points < 84:
    return '1.Mediocre'
  elif points >= 84 and points < 88:
    return '2.Good'
  elif points >= 88 and points < 92:
    return '3.Very good'
  elif points >= 92 and points < 96:
    return '4.Outstanding'
  else:
    return '5.Classic'

# Add kiss_points column
description_points_df = description_points_df.assign(kiss_points = de
description_points_df.head()
```

[9]:

	description	points	description_length	k
25	Yields were down in 2015, but intensity is up,...	94	215	4.Ou
29	This standout Rocks District wine brings earth...	94	333	4.Ou
60	Concentrated, ripe blackberry and cassis aroma...	91	242	3.
61	Moorooduc's estate Pinot Noir is a solid value...	91	315	3.
62	Smoky aromas of fresh-cut wood blend with berr...	91	229	3.





	precision	recall	f1-score	support
1.Mediocre	1.00	0.95	0.97	763
2.Good	0.96	0.99	0.98	3830
3.Very good	0.98	0.97	0.97	3418
4.Outstanding	1.00	0.95	0.98	1135
5.Classic	1.00	0.96	0.98	94
accuracy			0.98	9240
macro avg	0.99	0.96	0.98	9240
weighted avg	0.98	0.98	0.98	9240