Section Detection

Isaiah Valenica - ijvalencia@ucdavis.edu
Will Schmidt - wpschmidt@ucdavis.edu
Christopher Ton - chrton@ucdavis.edu

Github: https://github.com/chriztopherton/ucd_datafestF19

GOAL: Identify Sections of Catalog Based on Location Coordinates

METHOD:
K-MEANS Clustering
- Scatter plots of subgroups (conf. Level > 60 (?))
Create dataframe of all word sizes
Copy URL to clipboard

GERMAN RHINES AND MOSELLES

1973 and 1975 are a happy pair of vintages for lovers of fine German white wines. The 1973s – available right now at our shop – are light, elegant wines blessed with much fruit and flavor. The 1975s – which can be preordered now at substantial savings, with delivery in mid–October – are great, rich wines, the equals of the now legendary '71s. While these fabulous '75s are maturing to their full glory, the '73s are ready to offer much pleasure and value. All are estate–bottled.

No.	1973 RHINE WINES	Bottle	Case
755	JOHANNISBERGER ERNTEBRINGER, Winzergenossenschaft Dry, soft, an appealing casual wine.	2.59	29.55
704	WACHENHEIMER BISCHOFSGARTEN, Dr. Bürklin–Wolf Good balance and bouquet – from one of the Pfalz's most distinguished growers.	2.99	34.10
792	RUPPERTSBERGER GEISBÖHL KABINETT, Dr. Bürklin–Wolf Charming and attractive. Very nice with a number of light foods.	2.99	34.10
702	RÜDESHEIMER BURGWEG, Winzergenossenschaft Rüdesheim lies at the western edge of the Rheingau, making a number of good wines. Burgweg is one of the best. Full–bodied.	2.99	34.10
852	RAUENTHALER WÜLFEN KABINETT, Winzergenossenschaft Rauenthal is especially known for its sweeter wines, but this relatively dry Kabinett is a beauty.	2.99	34.10
718	DEIDESHEIMER LANGENMORGEN SPÄTLESE, Dr. Bürklin–Wolf Top quality, 'late picking' wine from the eminent grower, Dr. Bürklin–Wolf.	4.99	56.90

1973 MOSELLE WINES
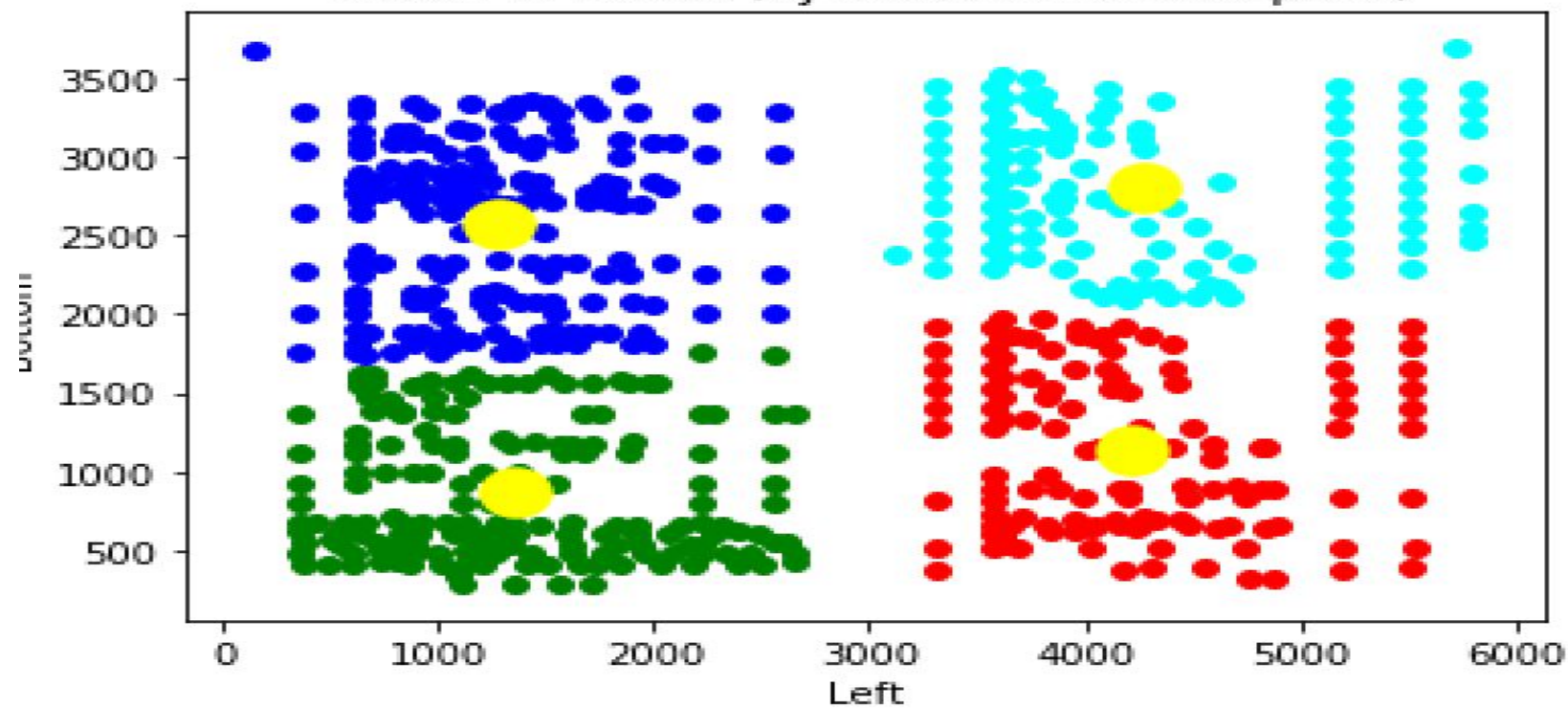
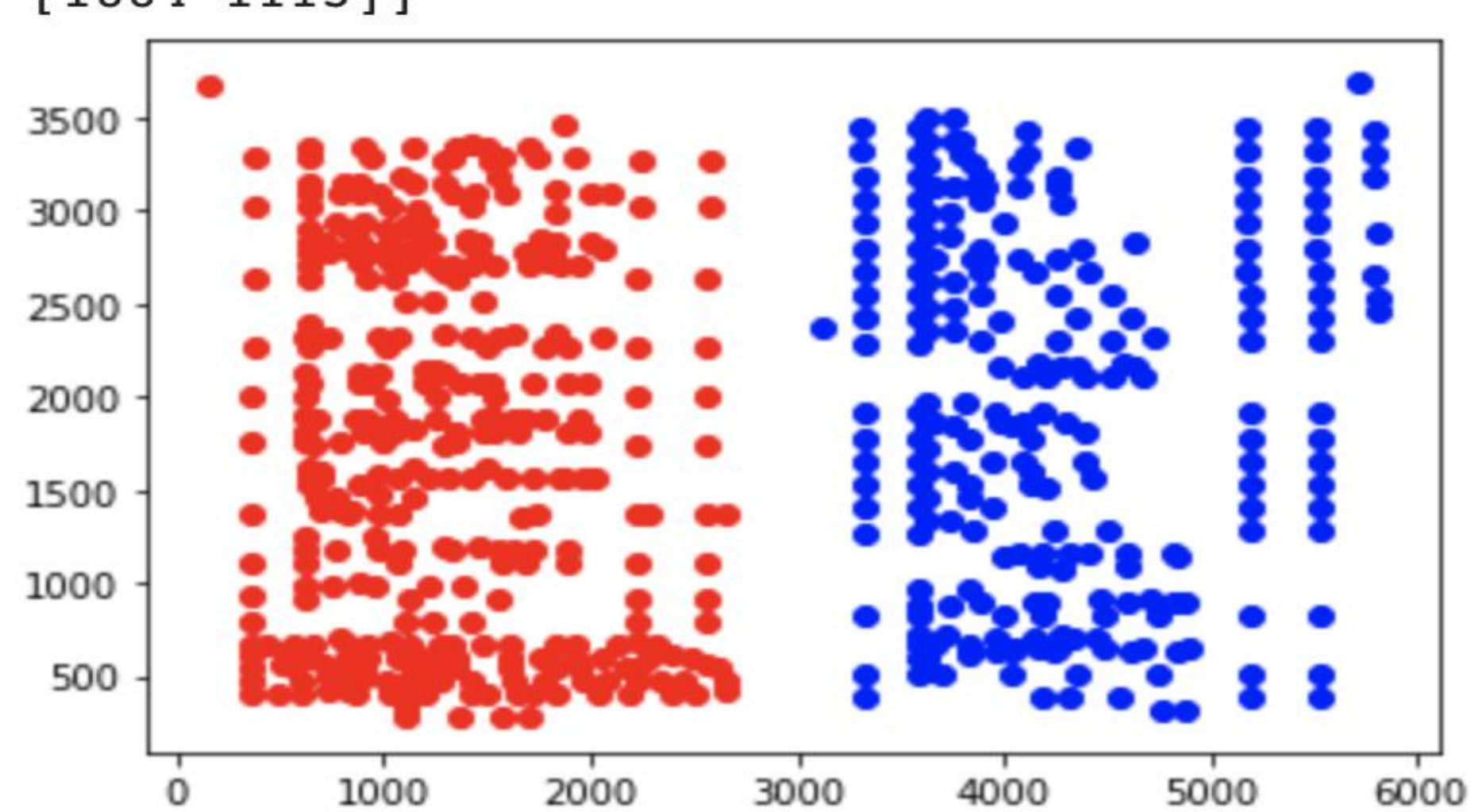| No.		Bottle	Case
625	PIESPORTER MICHELSBERG, Dunweg Piesporter wines are some of the most attractive in the Moselle, and this one from the Michelsberg vineyard is no exception. Typical, light. Also available in half–bottles at $1.44 the bottle, $32.55 the case.	2.59	29.55
675	BERNKASTELER KURFÜRSTLAY, Dunweg Shows all the characteristics that make Bernkastel wines the most famous of Moselles. Good distinction.	2.99	34.10
618	OCKFENER GEISBERGER, Friedrich Wilhelm Gymnasium Sprightly, refreshing example of an elegant Saar.	2.99	34.10

continued

No.	1973 MOSELLE WINES	Bottle	Case
860	ZELTINGER HIMMELREICH, Dunweg Lovely, delicious, superb class. Marries well with a light dinner.	2.99	34.10
594	BRAUNEBERGER JUFFER KABINETT, Ferdinand Haag Flowery and feminine. A wonderful wine from one of Germany's great vineyards.	3.49	39.80

THE GREAT 1975 RHINES
(Order now for delivery in mid–October)

No.		Bottle	Case
724	ELTVILLER SONNENBERG RIESLING Q.B.A. Von Simmern	3.69	42.10
728	STEINBERGER KABINETT, Staatsweinguter	3.99	45.50
732	SCHLOSS VOLLRADS GRÜNLACK, Graf Matuschka–Greiffenclau	3.99	45.50
735	RÜDESHEIMER BERG ROTTLAND KABINETT, Staatsweinguter	4.29	48.95
738	SCHLOSS VOLLRADS BLAULACK, Graf Matuschka–Greiffenclau	4.79	54.65
739	RAUENTHALER BAIKEN KABINETT, Schloss Eltz	4.99	56.90

THE GREAT 1975 MOSELLES
(Order now for delivery in mid–October)

No.		Bottle	Case
668	PIESPORTER MICHELSBERG RIESLING Q.B.A., Otto Dunweg	3.29	37.55
671	BERNKASTELER KURFÜRSTLAY RIESLING Q.B.A., Otto Dunweg	3.29	37.55
673	ZELTINGER HIMMELREICH RIESLING Q.B.A., Otto Dunweg	3.29	37.55
674	OCKFENER GEISBERG RIESLING KABINETT, Friedrich Wilhelm Gymnasium	4.29	48.95
678	PIESPORTER GOLDTRÖPFCHEN KABINETT, Von Kesselstaat	4.79	54.65
679	JOSEPHSHOFER KABINETT, Von Kesselstaat	4.79	54.65
683	GRAACHER HIMMELREICH SPÄTLESE, Friedrich Wilhelm Gymnasium	5.49	62.60
684	ZELTINGER HIMMELREICH KABINETT, Friedrich Wilhelm Gymnasium	5.49	62.60
687	SCHARZHOFBERGER KABINETT, Egon Müller	5.99	68.30
688	SCHARZHOFBERGER SPÄTLESE, Egon Müller	7.99	91.10

RESULT:

Cluster of Words (by bottom-left-most point)

Recommendations/Next-Steps

- Create clusters within the clusters to potentially identify subheadings
- Create clusters based on size of words
- Return the top most values in each cluster (create a threshold)
- Look for outliers in size and positioning within clusters (potential header)
- Once you locate header, use semantics to determine topic of cluster