



Building Recommendations with Unsupervised Learning Techniques

Clustering Single Malt Scotches by Flavour Profile

Chris Journeyay

Machine Learning Coursework 3

Contents

Table of Figures.....	2
Table of Tables	2
Project Overview.....	3
Data Set Overview.....	3
Data types	4
Data Set Exploration	4
Raw Data	4
Data Pre-processing	4
Data Visualisation	5
Principle Component Analysis	6
Clusterability Analysis	8
Hopkins Statistic.....	8
Dissimilarity Matrix	8
Summary of Data Exploration and Clusterability Analysis.....	10
Approach to finding the optimal cluster solution.....	10
Clustering with Partitioning Algorithms.....	11
K-means clustering.....	11
Partitioning Around Medoids (k-medoids) Clustering	13
Hierarchical Clustering.....	15
Agglomerative Hierarchical Clustering.....	15
Divisive Hierarchical Clustering.....	18
Soft Clustering.....	21
Fuzzy Clustering	21
Summary of Models.....	24
Model Selection	24
Recommendation Generation	24
Endnotes	27

Table of Figures

Figure 1 - Density Plot of Whisky Features	5
Figure 2 - Correlation Plot of Whisky Features	6
Figure 3 - Variables vs Principle Components	7
Figure 4 - Individual distilleries vs Principle Components	8
Figure 5 - Whisky Dissimilarity Matrix - Spearman Distance	9
Figure 6 - Whisky Dissimilarity Matrix - Kendall Distance	9
Figure 7 - Optimal Clusters for K-means using average silhouette width	11
Figure 8 - K-means average silhouette plot	12
Figure 9 - K-means clusters plot	13
Figure 10 - K-medoid optimal clusters plot	14
Figure 11 - K-medoid silhouette plot	14
Figure 12 - K-medoids cluster plot	15
Figure 13 - Agglomerative Hierarchical Cluster Dendrogram	16
Figure 14 - Agglomerative Hierarchical Cluster Dendrogram with Clusters	16
Figure 15 - Agglomerative Hierarchical Cluster Silhouette plot	17
Figure 16 - Agglomerative Hierarchical Cluster plot	18
Figure 17 - Divisive Hierarchical Cluster Dendrogram	19
Figure 18 - Divisive Hierarchical Cluster Dendrogram with clusters	19
Figure 19 - Divisive Hierarchical Cluster silhouette plot	20
Figure 20 - Divisive Hierarchical Cluster plot	21
Figure 21 - Fuzzy clustering silhouette plot	22
Figure 22 - Fuzzy clustering cluster plot	23

Table of Tables

Table 1 - Whisky Flavour Profile Features	3
Table 2 - Whisky flavour profile Likert values	4
Table 3 - Sample of Whisky Data Set	4
Table 4 - Principle Component Eigenvalues for Whisky dataset	7
Table 5 - Average Silhouette Width for each cluster using k-means	11
Table 6 - Cluster size of each cluster using k-means	11
Table 7 - K-medoid average silhouette width	14
Table 8 - Agglomerative Hierarchical Cluster Average Silhouette Widths	16
Table 9 - Divisive Hierarchical Cluster average silhouette widths	19
Table 10 - Fuzzy clustering average silhouette widths	21
Table 11 - Fuzzy clustering Dunn's coefficients	22
Table 12 - Summary of model performance	24
Table 13 - Sample recommendation validation	25
Table 14 - Complete set of recommendations	25

Project Overview

The world of single malt scotch whisky is diverse and complex. Scotland's 80+ distilleries produce an array of single malts of incredible variety of styles and age statements and finding a way of exploring this complex landscape is difficult. Using regional designation (i.e.: Speyside) is an option but there can be profound differences even within regions.

This project aims to simplify that by using unsupervised machine learning to cluster single malts by flavour profiles. The goal of this project is to use those clusters to generate recommendations for single malts to the casual drinker who has a known or set of known single malts that they enjoy.

First the optimal clustering algorithm will be selected for the flavour profile data. Once the clusters are complete then recommendations can be generated. Recommendations will be for the most similar single malts as well as a palate challenger – a single malt within the same cluster but that is most unlike the reference malt.

This recommendation system is a great way of giving drinkers a path for exploring the different single malts available.

Data Set Overview

The data set used for this analysis is a set of flavour profiles for single malts produced by distilleries in Scotland. The data was sourced from Kaggleⁱ and is based on a set of tasting notes found in the book *Whisky Classified* by David Wishart.ⁱⁱ

The data set contains 86 observations of single malts by distillery name. Single malts between 10-15 years old are profiled, one for each distillery. Where those age statements are not available, the best-selling single malt from that distillery was selected instead. This represents a sample of the most commonly available single malts from each distillery and the most commonly available choices in restaurants and bars.

The flavour profile is made up of 12 features

Table 1 - Whisky Flavour Profile Features

Feature	Explanation
Body	Weight or fullness of the malt and often referred to as “mouthfeel” – from light through to full bodied.
Sweetness	Presence of sugar in the overall process - Similar terminology used for wine, from dry through to sweet
Smoky	Influence of peat in the distillation process – taste of peat, moss, soil, bonfires, or smoked food
Medicinal	Influence from seashore maturation – taste of iodine, seaweed, anti-septic
Tobacco	Influence of both still shape and maturation - taste of old books, musty, old leather, garden sheds
Honey	Influence of both still shape and maturation barrels - taste of beeswax, mead, butterscotch, caramel
Spicy	Influence of the oak used in maturation barrels - taste of cedar, cinnamon, pepper, woody
Winey	Influence of the original use of the cask before whisky maturation - taste of port, sherry, madeira
Nutty	Influence of the oak used in the maturation barrels - Taste of almonds, walnuts, chocolate, butter
Malty	Influence of distillation – taste of biscuits, cereal, grain

Fruity	Influence of both still shape and maturation – taste of raisins, strawberries, citrus, fruitcake
Floral	Influence of both still shape and maturation – fragrant, lavender, heather, mint, grassy

Data types

Each feature has a value on a Likert scale between 0-4 that represents the intensity of each element in the overall flavour.

Table 2 - Whisky flavour profile Likert values

Value	Intensity
0	Not Present
1	Low Hints
2	Medium Notes
3	Definite Notes
4	Pronounced

This Likert scale is built around tasting notes from numerous publications that all use different terminology and is simplified to accommodate those differences. The data could be considered a meta-study of those publications. Additional information on the collation of the data set can be found on the original author's website. ⁱⁱⁱ

One important note is that these Likert values will be considered *non-parametric*. The goal of this work is to understand how similar or dissimilar each whisky is from each other, not from some theoretical mean. This implications for the data processing;

- the data will not be scaled for any of this analysis
- the choice of distance measure will focus on non-parametric measures

Data Set Exploration

Raw Data

The following is a sample of the whisky data set.

Table 3 - Sample of Whisky Data Set

	Body	Sweetness	Smoky	Medicinal	Tobacco	Honey	Spicy	Winey	Nutty	Malty	Fruity	Floral
Aberfeldy	2	2	2	0	0	2	1	2	2	2	2	2
Aberlour	3	3	1	0	0	4	3	2	2	3	3	2
AnCnoc	1	3	2	0	0	2	0	0	2	2	3	2
Ardbeg	4	1	4	4	0	0	2	0	1	2	1	0
Ardmore	2	2	2	0	0	1	1	1	2	3	1	1
ArranIsleof	2	3	1	1	0	1	1	1	0	1	1	2
Auchentoshan	0	2	0	0	0	1	1	0	2	2	3	3
Auchroisk	2	3	1	0	0	2	1	2	2	2	2	1
Aultmore	2	2	1	0	0	1	0	0	2	2	2	2
Balblair	2	3	2	1	0	0	2	0	2	1	2	1

Data Pre-processing

The data set required no pre-processing however some minor changes were made to correct spelling of certain distilleries. An additional column was added to include the complete name of the malt including the relevant age statement.

Data Visualisation

A simple density plot provided an overview of the characteristics of the single malts under consideration.

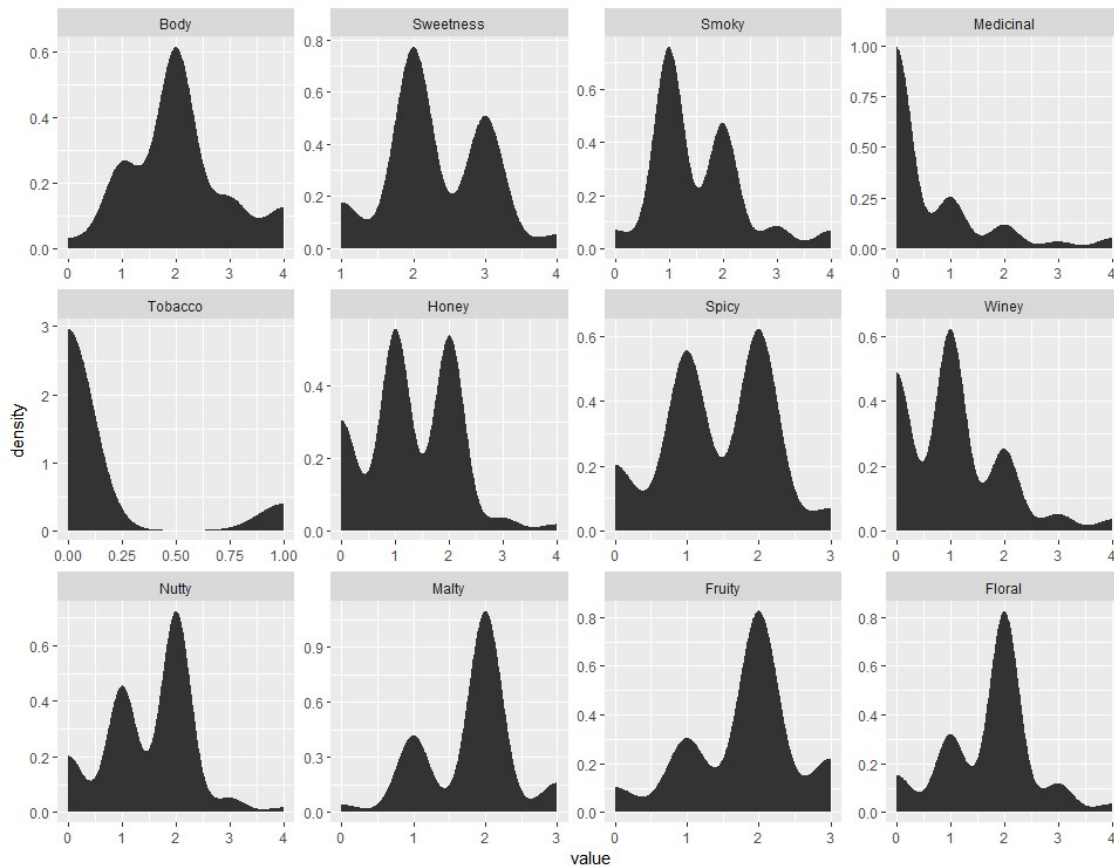


Figure 1 - Density Plot of Whisky Features

Low hints and Medium notes of flavour are most frequent amongst the 12 features. Medicinal and Tobacco flavour occur the least. Sweeter flavours are more prevalent.

There is low variability within each flavour group with a high proportion of the single malts clustered around very close values.

A correlation plot shows how the flavours work together.

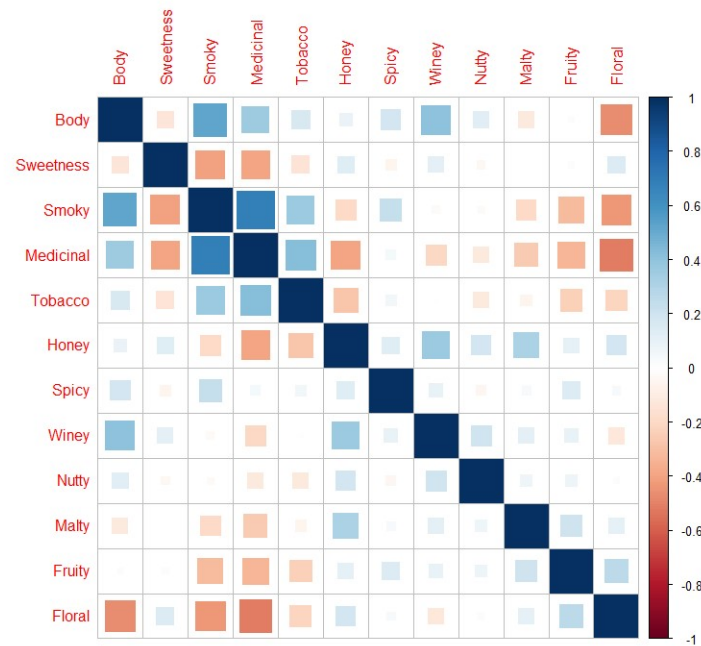


Figure 2 - Correlation Plot of Whisky Features

Some relationships are easily distinguished.

- Smoky and Medicinal flavours are highly correlated and are the antithesis of Floral and fruity flavours
- Both are positively correlated to Body, as are Winey flavours

The remaining relationships between the flavours are less clear. The remaining relationships are not strong, either positively or negatively.

This only highlights the challenges for this work that key differences between the single malts will require additional work to uncover.

Principle Component Analysis

As there appears to be no real patterns that clearly separate the data set, a principle component analysis is run to look at feature extraction. The following table displays the Eigenvalues from that analysis.

Table 4 - Principle Component Eigenvalues for Whisky dataset

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	2.33128029	30.1109794	30.11098
Dim.2	1.48790511	19.2178865	49.32887
Dim.3	0.74017815	9.5601927	58.88906
Dim.4	0.63876410	8.2503219	67.13938
Dim.5	0.55983472	7.2308645	74.37024
Dim.6	0.46394222	5.9923101	80.36256
Dim.7	0.39548319	5.1080886	85.47064
Dim.8	0.35514396	4.5870642	90.05771
Dim.9	0.27083293	3.4980971	93.55580
Dim.10	0.24757748	3.1977281	96.75353
Dim.11	0.17787006	2.2973821	99.05092
Dim.12	0.07348093	0.9490848	100.00000

It takes 7 principle components to explain more than 85% of the variance between the single malts. Once plotted against the first two principle components, the influence of the variables closely reflects the relationships seen in the correlation plot.

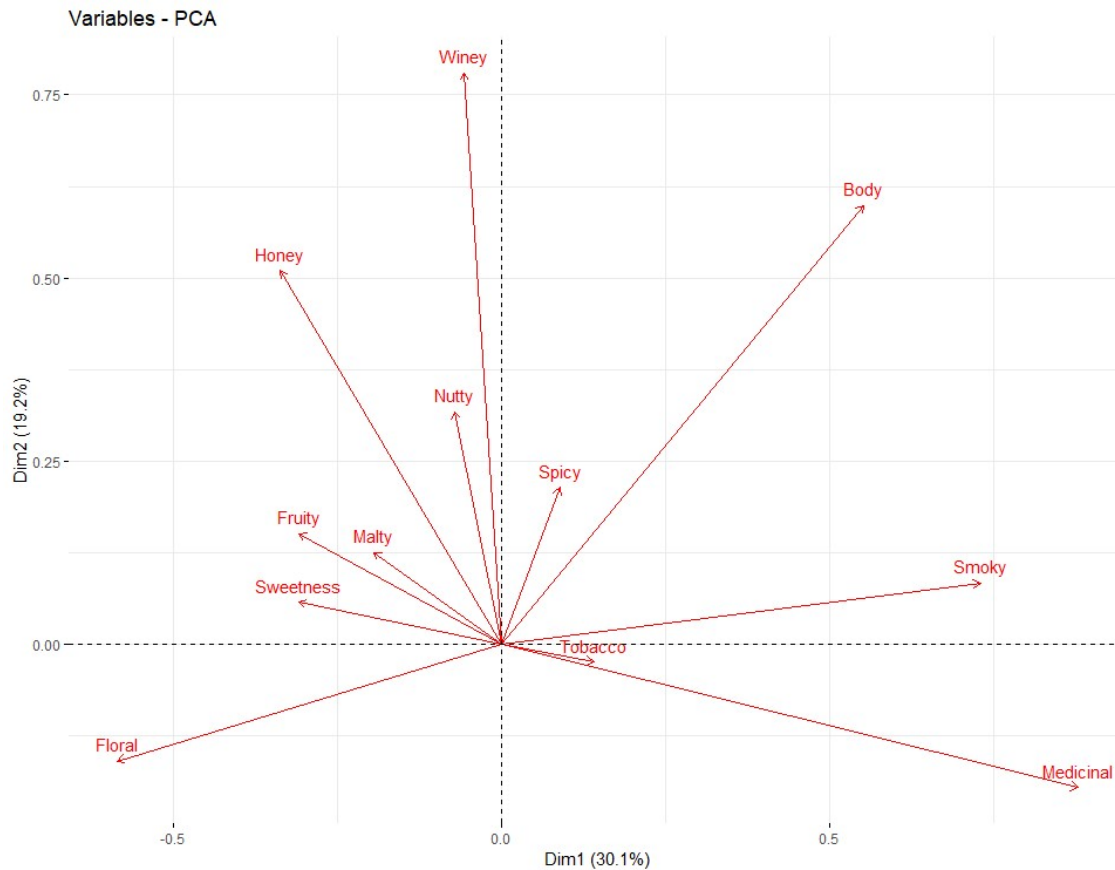


Figure 3 - Variables vs Principle Components

With the principle analysis complete, the data can be plotted against the first two principle components.

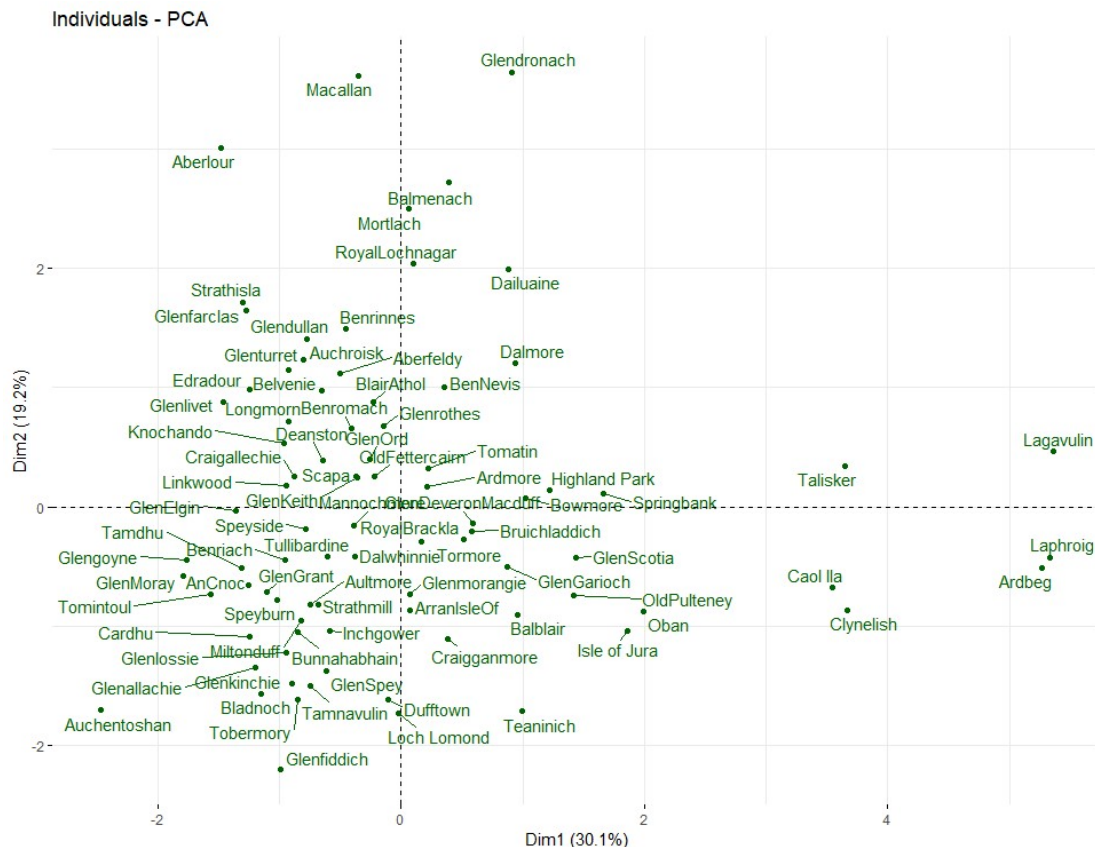


Figure 4 - Individual distilleries vs Principle Components

There are some signs of clusters around the smoky and medicinal elements as well as the winey element, but the rest of the data shows no real clustering against the first two principle components.

A clusterability analysis can better inform the approach.

Clusterability Analysis

Hopkins Statistic

In order to determine whether clustering is possible the Hopkins Statistic (H) for the data set is calculated.

The Hopkins value for the data set is 0.38.

The factoextra package in R used to calculate the Hopkins Statistic returns a value of 1-H so the closer the result is to zero, the more clustered the data. A value of 0.5 would indicate that the values are random however we are somewhere in the middle. There appears to be weak clustering to the data set and this will have an impact on the approach to the analysis.

Dissimilarity Matrix

Dissimilarity matrices are run against the data set using different distance measures. As the data is treated as non-parametric then the distance measures are derived using both the Spearman and Kendall correlation distances. Both matrices are then plotted and analysed.

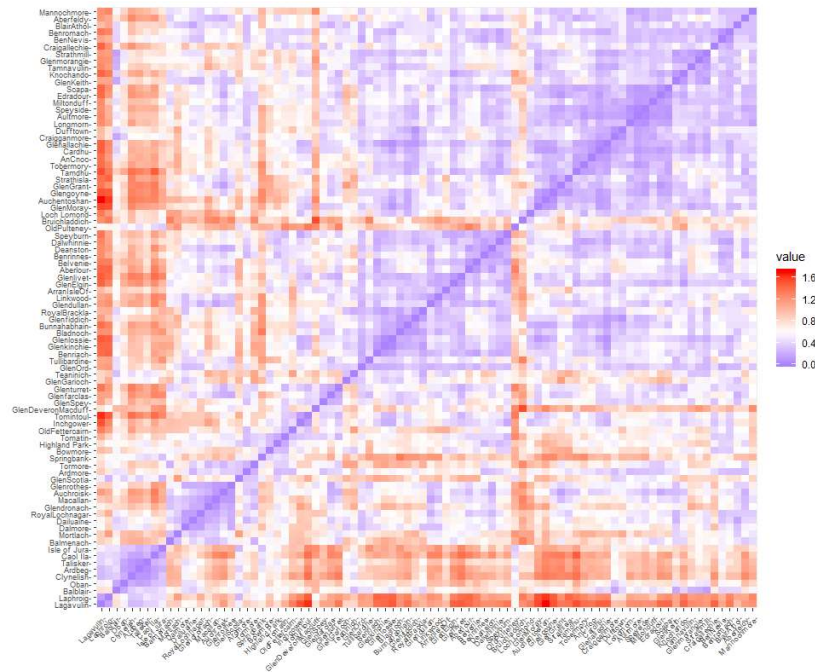


Figure 5 - Whisky Dissimilarity Matrix - Spearman Distance

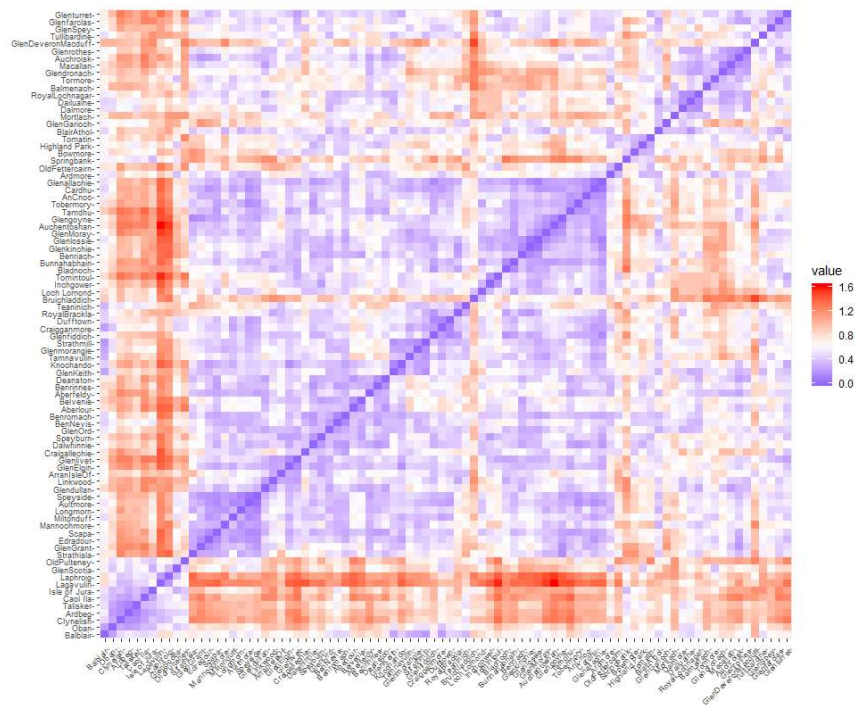


Figure 6 - Whisky Dissimilarity Matrix - Kendall Distance

Both plots demonstrate a clear tendency for the data to cluster. Visually, the Kendall distance measure provides clearer clusters. This is the distance measure that will be chosen for the data set.

Summary of Data Exploration and Clusterability Analysis

Finding clusters will be challenging for this data set. The higher dimensionality of principle components and a higher Hopkins statistic are indicators of that challenge. However, clusters are evident in the dissimilarity correlation plots and this validates the non-parametric approach to handling the data however it also highlights the challenge of dealing with outliers.

The approach to developing clusters is shaped by the nature of the data and the nature of the outcome we are looking to achieve. This is flavour profile data for whiskies and the goal is to recommend drinks of similar flavour profiles so at the core this data is subjective. Our recommendations do not have to be 100% - we are simply looking for the best approximation of flavour profiles for recommendations. This grants some degree of freedom to the approach enabling an approach that requires a balance between degree of similarity and usefulness to the end consumer. There will be no perfect solution to this, simply a best efforts approach to finding meaningful distinctions and then using them to derive recommendations.

Approach to finding the optimal cluster solution

The approach to clustering the data involves investigating multiple unsupervised learning approaches against the data set. This involved partitioning and hierarchical methods as well as newer soft methods. The complete list is as follows;

- Partitioning algorithms
 - K-means clustering
 - K-medoids clustering
- Hierarchical Clustering
 - Agglomerative hierarchical clustering
 - Divisive hierarchical clustering
- Soft Method
 - Fuzzy clustering

For each method the optimal number of clusters (k) is selected with the optimal number chosen by comparing average silhouette widths as a guideline and avoiding negative values as much as possible. The following guidelines are used for interpreting silhouette width;

- Observations with a large silhouette width are very well clustered
- Smaller silhouette widths (around 0) means that the observation lies between two clusters
- Observations with a negative S_i are probably placed in the wrong cluster

Each method is then compared, and the best approach is then chosen based on the best balance between the number of clusters, cluster size and silhouette widths. The recommendations are then derived from the chosen approach.

Clustering with Partitioning Algorithms

K-means clustering

In order to perform k-means clustering finding the optimal k value is the first step. As the silhouette width is our key statistic then finding k that maximises that value is key.

The Nbclust function in R is used to run multiple scenarios and the average silhouette width is plotted against each value of k.

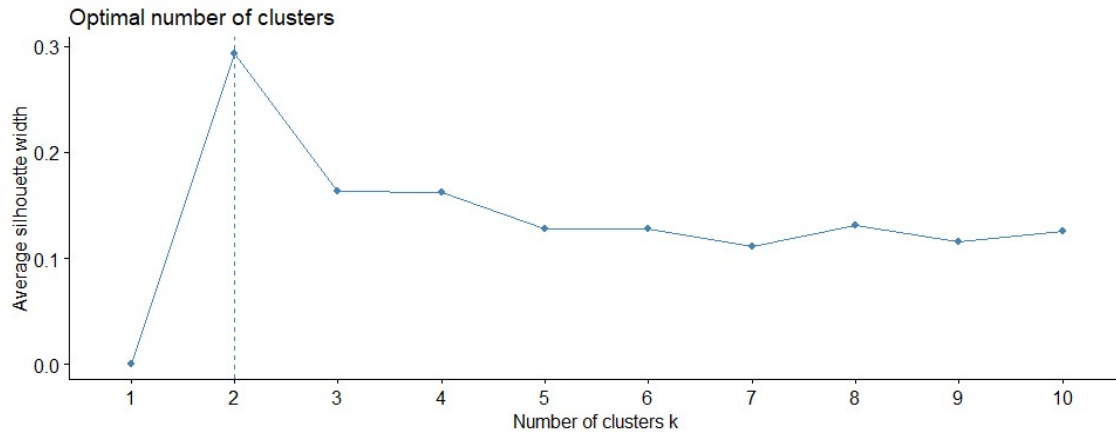


Figure 7 - Optimal Clusters for K-means using average silhouette width

Although 2 provides the best average, the overall numbers need investigation to determine if it is the optimal choice.

The k-means algorithm was run against the Kendall Dissimilarity Matrix using multiple values of k. The average silhouette values are then calculated for each cluster along with the resulting size of each cluster.

Table 5 - Average Silhouette Width for each cluster using k-means

K-value	2	3	4	5	6
Cluster 1	0.25	0.19	0.18	0.13	0.12
Cluster 2	0.04	0.18	0.11	0.08	0.03
Cluster 3		0.03	-0.05	0.00	0.12
Cluster 4			0.12	0.11	0.06
Cluster 5				0.09	-0.04
Cluster 6					0.13

Table 6 - Cluster size of each cluster using k-means

K-value	2	3	4	5	6
Cluster 1	65	56	12	8	9
Cluster 2	21	18	19	8	10
Cluster 3		12	26	14	8
Cluster 4			29	26	15
Cluster 5				30	24
Cluster 6					20

No solution is optimal although k=5 appears to be the best choice. Plotting the silhouette width for the clusters gives the following:



Figure 8 - K-means average silhouette plot

The overall low silhouette width values indicate many data points that are very close to the next adjacent group. The large number of negative values indicates many mis-classified data points; Cluster 3 is particularly problematic as half of the values appear to be mis-classified. Overall this indicates poor performance for this approach.

Plotting the clusters against the first two principle components of the Kendall Distance Matrix gives the following:

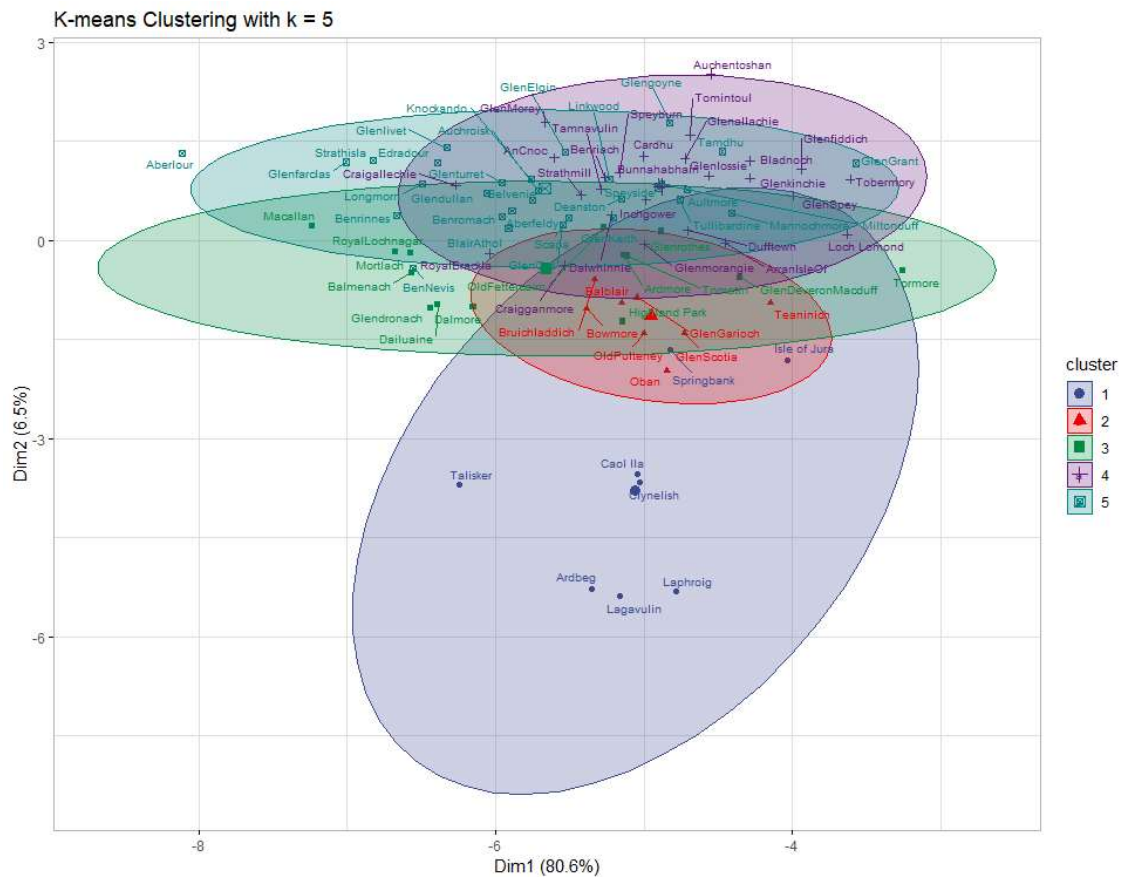


Figure 9 - K-means clusters plot

Over 87.1% of the variance between the single malts can be found within these two principle components, this validates our approach to using this distance matrix approach to finding clusters.

The plot is difficult to interpret as the makeup of the principle components from the distance matrix cannot be determined. The overlapping clusters does explain the silhouette width challenge – the centres of cluster 2-5 are quite close to one another and the data points are all very densely packed. The optimum solution will require a better way of dealing with the nature of the data.

Partitioning Around Medoids (k-medoids) Clustering

Partitioning Around Medoids is the next method run on the data. The choice of medoid over means used in k-means makes this approach less sensitive to noise and outliers which might improve performance.

The optimal number of k is first sought using maximal values of silhouette width as the target.

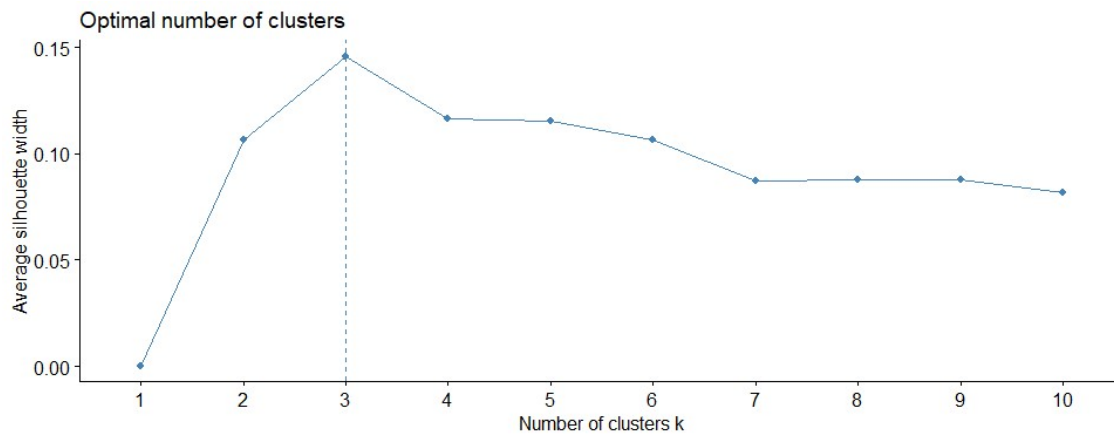


Figure 10 - K-medoid optimal clusters plot

The overall average silhouette width is even less than with k-means. In order to complete the picture, the average silhouette values are then calculated for each cluster using multiple values of k.

Table 7 - K-medoid average silhouette width

K-value	3	4	5	6	7
Cluster 1	0.08	0.02	0.05	0.11	0.09
Cluster 2	0.17	0.13	0.06	0.02	0.03
Cluster 3	0.19	0.19	0.11	0.06	0.06
Cluster 4		0.08	0.20	0.20	0.17
Cluster 5			0.02	0.03	0.07
Cluster 6				0.03	0.01
Cluster 7					0.15

Visual inspection of the plots against does confirm that the best choice for this method is k=3. A plot of the silhouette widths by cluster provides a view of the clusters.



Figure 11 - K-medoid silhouette plot

Once again, several the data points are very close to their neighbouring cluster with many mis-classified points.

Plotting the clusters against the principle components for the Kendall distance matrix provides the following:

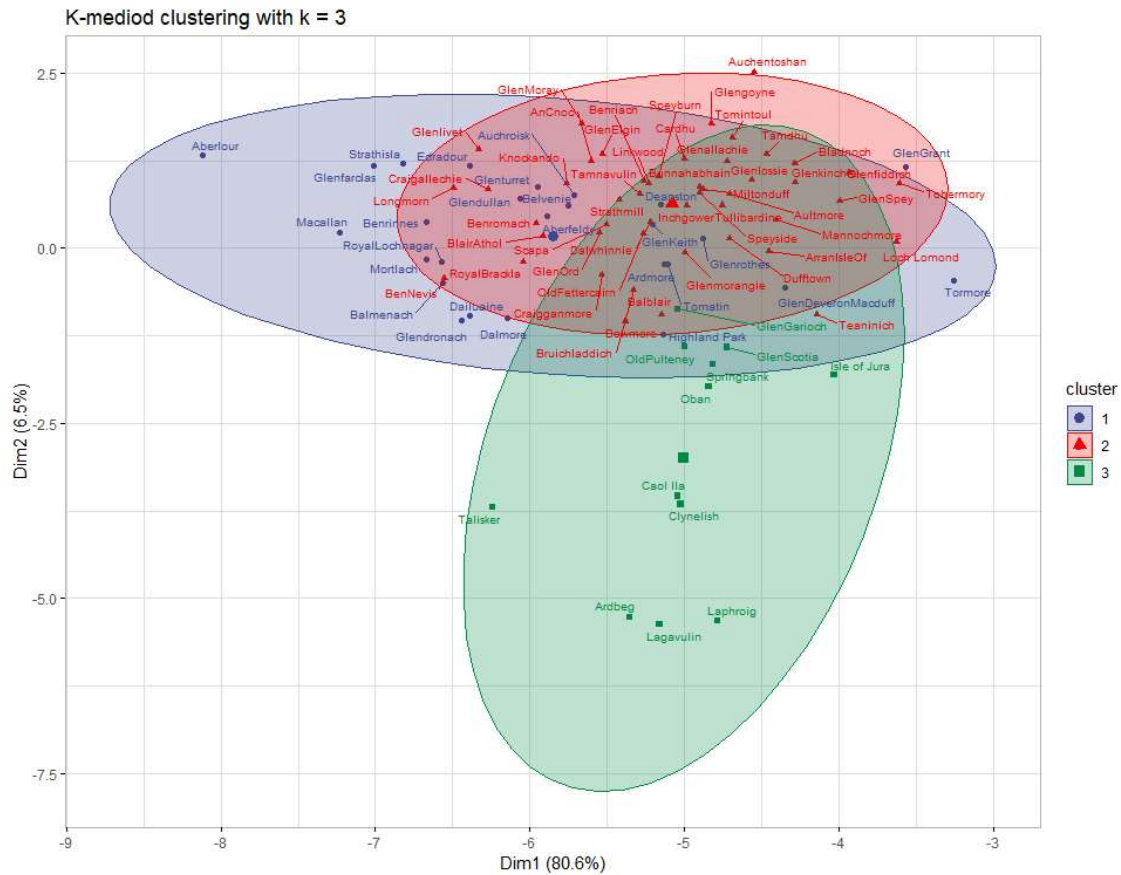


Figure 12 - K-medoids cluster plot

This demonstrates the opposite problem to the k-means approach. The smaller k value has led to more loosely built clusters with weaker relationships.

Hierarchical Clustering

Agglomerative Hierarchical Clustering

This tree-based approach requires a choice of clustering methods and to find the best option we look for the method that provided the highest Agglomerative Coefficient – a measure of the cluster structure that is derived.

There are four methods tested: Average, Single, Complete, Ward with the best choice being the one that provides the highest coefficient. The results are as follows;

average	single	complete	ward
0.7598576	0.4767442	0.8471761	0.9239513

The Ward method is selected.

The Agglomerative Hierarchical Clustering algorithm was run against the Kendall Dissimilarity Matrix using the Ward method to derive the following dendrogram.

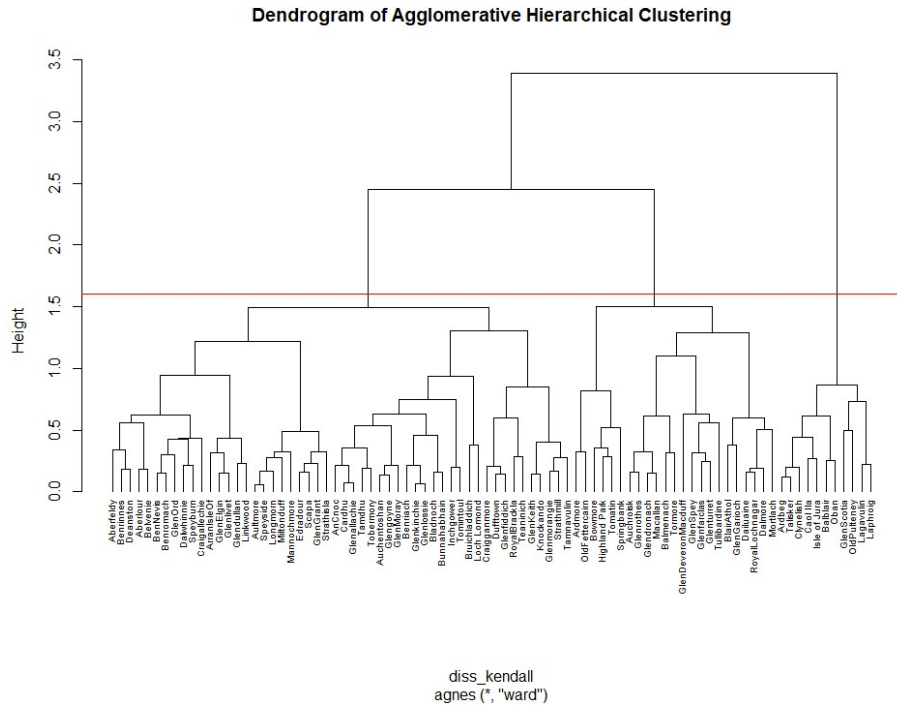


Figure 13 - Agglomerative Hierarchical Cluster Dendrogram

The optimal k for this solution appears to be k = 3 and the rest of the analysis is run with that chosen value.

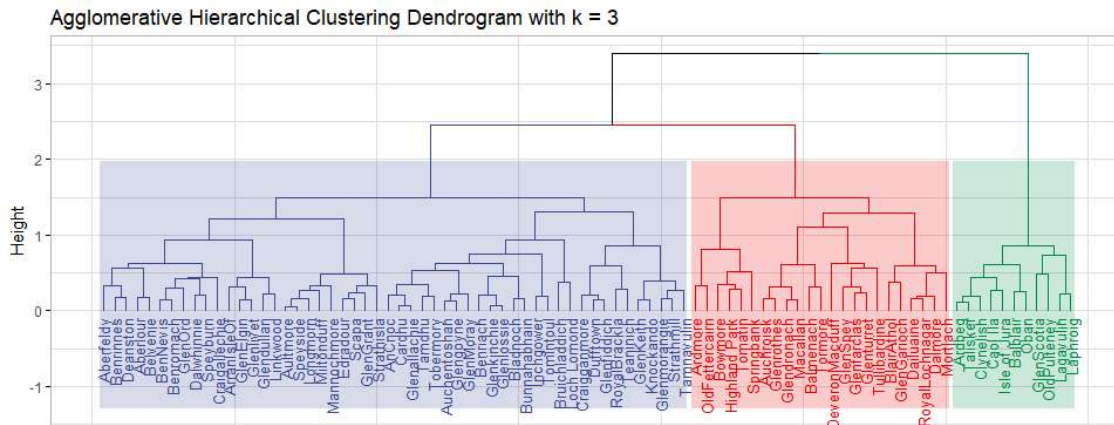


Figure 14 - Agglomerative Hierarchical Cluster Dendrogram with Clusters

This model generated the following average silhouette widths

Table 8 - Agglomerative Hierarchical Cluster Average Silhouette Widths

K-value	3
Cluster 1	0.16
Cluster 2	0.23
Cluster 3	0.05

A plot of the silhouette widths by cluster provides a view of the clusters.

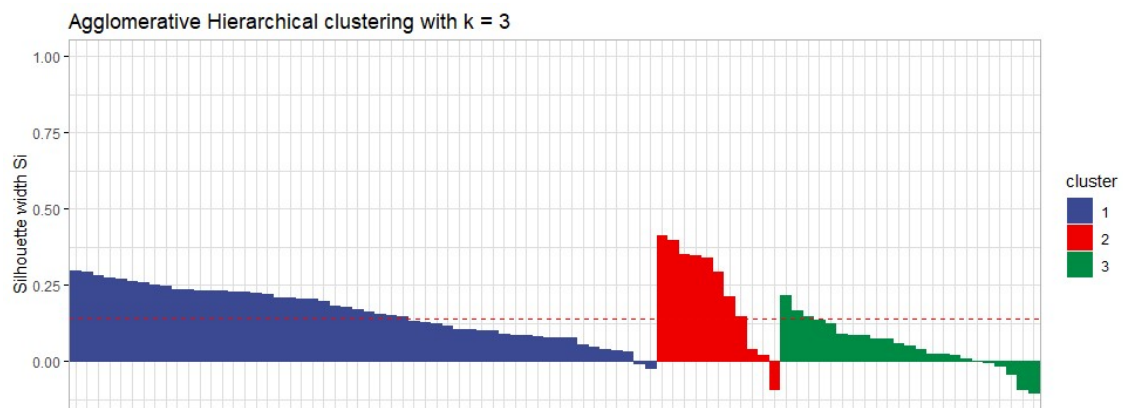


Figure 15 - Agglomerative Hierarchical Cluster Silhouette plot

This result is not dissimilar to the k-medoids approach and the clusters appear almost identical. As a result, both approaches share the same issues with low silhouette width however there are fewer mis-classified points.

Plotting the clusters against the first two principle components of the Kendall distance matrix provides the following:

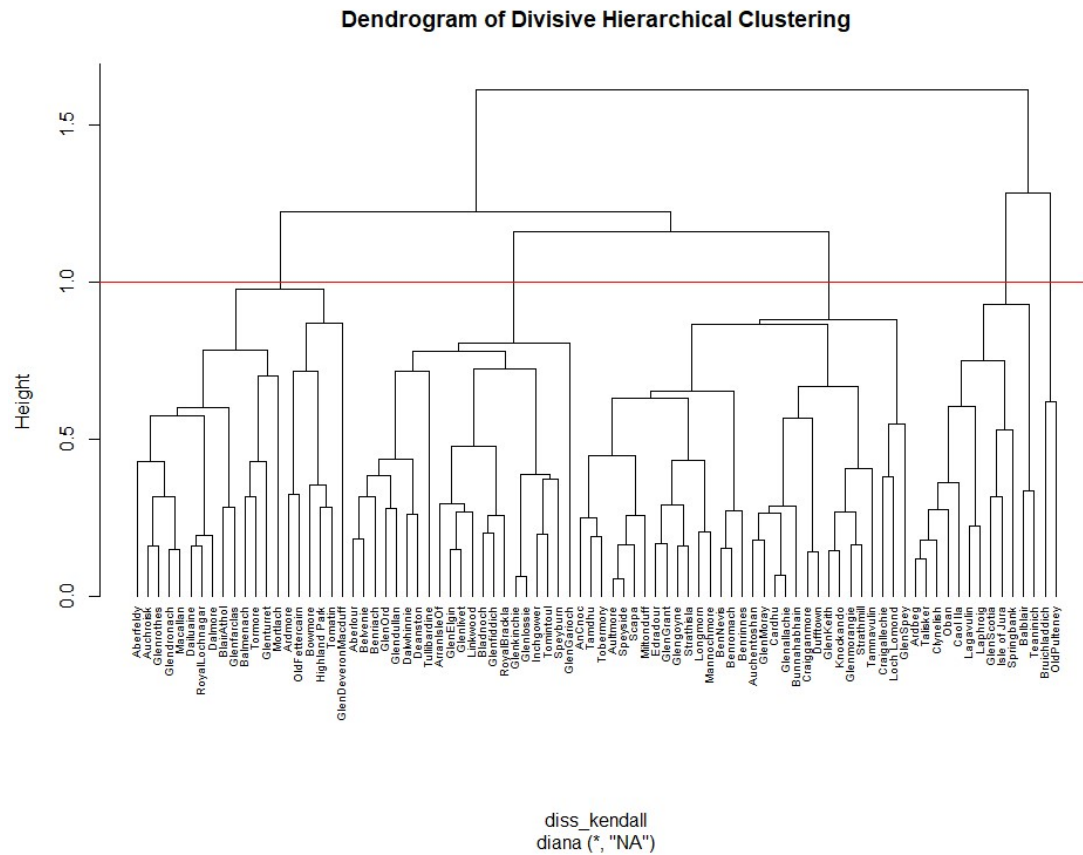


Figure 17 - Divisive Hierarchical Cluster Dendrogram

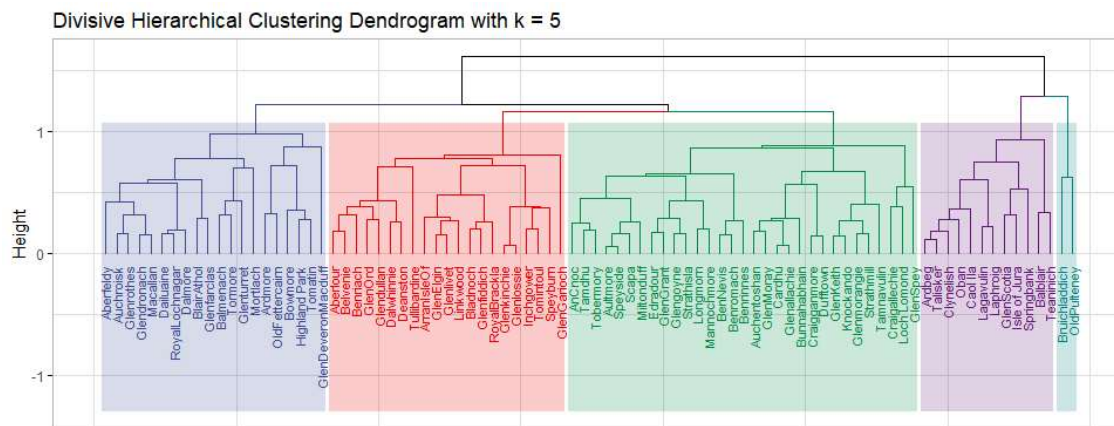


Figure 18 - Divisive Hierarchical Cluster Dendrogram with clusters

This model generated the following silhouette widths

Table 9 - Divisive Hierarchical Cluster average silhouette widths

K-value	5
Cluster 1	0.02
Cluster 2	0.12
Cluster 3	-0.04

Cluster 4	0.00
Cluster 5	0.27

A plot of the silhouette widths by cluster provides a view of the clusters.



Figure 19 - Divisive Hierarchical Cluster silhouette plot

This result does not provide a suitable result. Not only are the average silhouette widths too small moreover most of the data points appear to be misclassified.

Plotting the clusters against the first two principle components of the Kendall distance matrix provides the following:

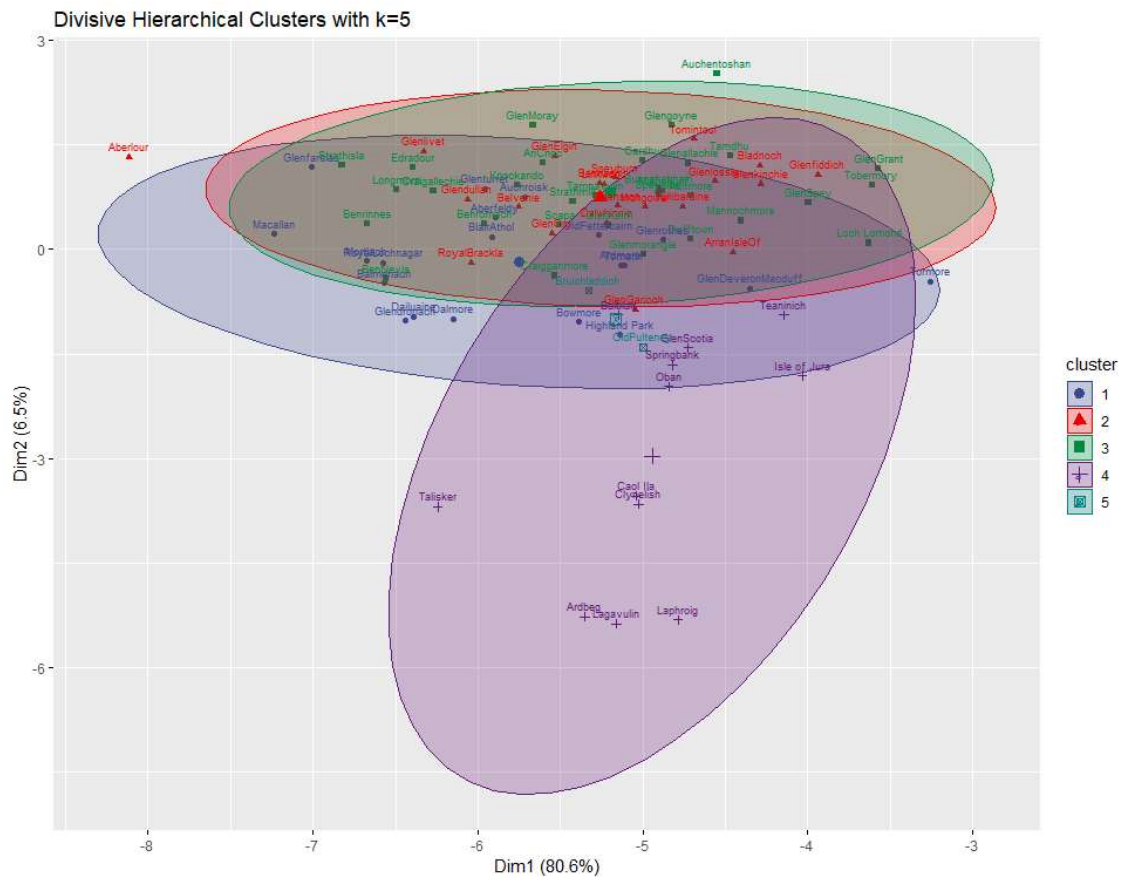


Figure 20 - Divisive Hierarchical Cluster plot

Like the k-means approach, the proximity of the cluster centres to each other leads to create many misclassified points.

Soft Clustering

Fuzzy Clustering

For the final model used is Fuzzy Clustering. This method generalises the partitioning cluster methods allowing observations to belong to more than one cluster. Assignment is based on the probability that an observation belongs to a cluster.

The fuzzy clustering algorithm was run against the Kendall Dissimilarity Matrix for multiple values of k to derive the following average silhouette widths. The specified membership exponent used was 1.2 as it provided the most stable response for the algorithm.

Table 10 - Fuzzy clustering average silhouette widths

K-value	3	4	5	6	7	8
Cluster 1	0.10	0.20	0.33	0.32	0.20	0.17
Cluster 2	0.29	0.08	0.08	0.07	0.29	0.24
Cluster 3	0.38	0.27	0.44	0.39	0.42	0.28
Cluster 4		0.41	0.11	0.10	0.04	0.36
Cluster 5			0.21	0.21	0.17	-0.01
Cluster 6				0.24	0.22	0.16

Custer 7					0.13	0.21
Cluster 8						0.26

To determine the goodness of fit for each value of k, Dunn's partition coefficient and normalised variant is calculated for each one. The optimal k should be chosen on the model that maximises the Normalised values.

Table 11 - Fuzzy clustering Dunn's coefficients

K-value	3	4	5	6	7	8
Dunn's Coefficient	0.7684	0.7071	0.7173	0.7142	0.7188	0.7326
Normalised	0.6526	0.6094	0.6466	0.6570	0.6719	0.6944

In this model a higher normalised Dunn Coefficient did lead to less clear clusters. The number of clusters, k, was chosen to be 6 as it provided the best balance. A plot of the silhouette widths by cluster provides a view of the clusters.

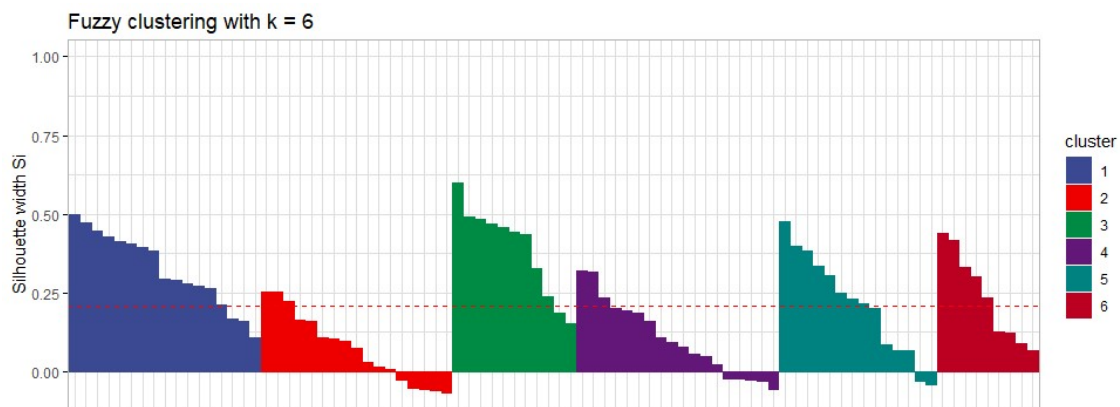


Figure 21 - Fuzzy clustering silhouette plot

This result provides stronger clusters with fewer misclassified observations. Plotting the clusters against the first two principle components of the Kendall distance matrix provides the following:

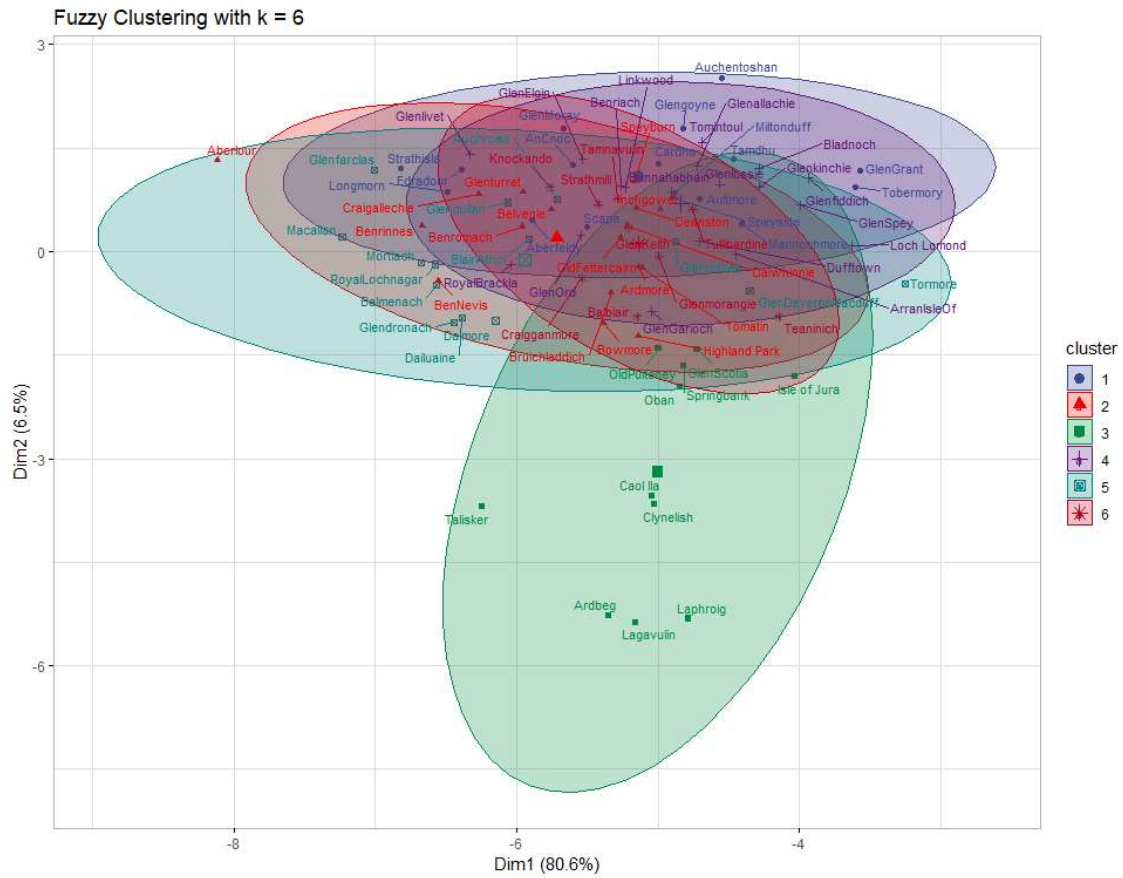


Figure 22 - Fuzzy clustering cluster plot

This provides much more densely packed clusters and deals with the close cluster centres better than other methods.

Summary of Models

The following table summarises the performance of each model

Table 12 - Summary of model performance

Model	K-means	K-medoids	Agglomerative Hierarchical Clustering	Divisive Hierarchical Clustering	Fuzzy Clustering
Chosen k	5	3	3	5	6
Avg. Silhouette Widths					
Cluster 1	0.13	0.08	0.16	0.02	0.32
Cluster 2	0.08	0.17	0.23	0.12	0.07
Cluster 3	0.00	0.19	0.05	-0.04	0.39
Cluster 4	0.11			0.00	0.10
Cluster 5	0.09			0.27	0.21
Cluster 6					0.24
Cluster sizes					
Cluster 1	8	26	52	20	17
Cluster 2	8	48	11	21	17
Cluster 3	14	12	23	31	11
Cluster 4	26			12	18
Cluster 5	30			2	14
Cluster 6					9
Total number of misclassified observations					
	14	13	8	19	12

k-means – sensible sized clusters with very low average silhouette widths for each cluster – sign of weak clustering

k-medoids – better performance but provided 1 very large cluster (>20 observations) with low silhouette width- sign of weak cluster

agglomerative hierarchical cluster – similar performance to k-medoids but far fewer misclassified observations

divisive hierarchical clustering – lead to one very small cluster (2 observations) and a large cluster with an average silhouette width of <0, highest number of misclassified observations

fuzzy clustering – sensible cluster sizes, better highest average silhouette widths (but not perfect) with second lowest misclassified observations

Model Selection

Fuzzy clustering with k = 6 is the chosen algorithm creating the clusters to generate recommendations. This approach provides the balance between strength of clusters, as measured by the average silhouette width, fewer misclassified observations and size of clusters.

Recommendation Generation

Generation of the recommendations is very straightforward. The whiskies are first separated into their clusters using the fuzzy clustering algorithms. For each whisky in each cluster, the first three nearest neighbours, as measured using the Kendall distance matrix, are assigned as the

recommended alternatives. The palate challenger whiskey is selected from the same cluster but using the furthest neighbour as measured by the Kendall distance matrix.

A random sample set of recommendations is collated in order to inspect the quality of the recommendations and by proxy the clustering.

The sample whisky chosen is Strathmill. The process recommends Glenmorangie, Glen Keith and Tamnavulin as the closest whiskies from the cluster, and recommended Teaninich as the palate challenger. All are from cluster 6.

The flavour profiles for those whiskies are as follows:

Table 13 - Sample recommendation validation

Distillery	Body	Sweetness	Smoky	Medicinal	Tobacco	Honey	Spicy	Winey	Nutty	Malty	Fruity	Floral
Strathmill	2	3	1	0	0	0	2	0	2	1	3	2
Glenmorangie	2	2	1	1	0	1	2	0	2	1	2	2
Glen Keith	2	3	1	0	0	1	2	1	2	1	2	1
Tamnavulin	1	3	2	0	0	0	2	0	2	1	2	3
Teaninich	2	2	2	1	0	0	2	0	0	0	2	2

This looks like a sensible set of recommendations; the nearest neighbours are very close in profile to Strathmill whereas the palate challenger is also similar but with stronger medicinal and smoky flavours and without the nutty and malty elements. This indicates a sensible outcome for this work.

Once complete, the clusters are reassembled into a single table to deliver the complete set of recommendations. To make it easier for the average drinker to interpret, the names of the distilleries are replaced with the name of the single malts along with their age statement (where available).

Table 14 - Complete set of recommendations

Single Malt	Recommendation 1	Recommendation 2	Recommendation 3	Palate Challenger
Aberfeldy 12 Year	Speyside	Tamdhu	Scapa	Auchentoshan
Abelour 10 Year	Balvenie	Deanston	Benromach	Bruichladdich
AnCnoc 12 Year	Cardhu	Tamdhu	Tobermory	Longmorn
Ardbeg 10 Year	Talisker	Clynelish	Oban	Springbank
Ardmore 12 Year	Benrinnes	Fettercairn	Benromach	Bruichladdich
Isle of Arran 10 Year	Glenlivet	Linkwood	GlenElgin	GlenGarioch
Auchentoshan 10 Year	Glengoyne	GlenMoray	Cardhu	Aberfeldy
Auchrisk 10 Year	Glenrothes	Macallan	Glenfarclas	Mortlach
Aultmore 12 Year	Speyside	Scapa	Miltoduff	GlenMoray
Balblair 10 Year	Strathmill	Craigganmore	GlenKeith	Tullibardine
Balmenach 12 Year	Tormore	Mortlach	Glendronach	GlenDeveronMacduff
Balvenie 10 Year	Aberlour	Speyburn	Deanston	Bruichladdich
Ben Nevis 10 Year	Benromach	Dalwhinnie	Benrinnes	Bruichladdich
Benriach 12 Year	Glenkinchie	Glenlossie	RoyalBrackla	GlenSpey
Benrinnes 15 Year	Deanston	Dalwhinnie	BenNevis	Bruichladdich
Benromach 5 Year	BenNevis	Dalwhinnie	Craigellechie	Bruichladdich
Bladnoch 16 Year	Bunnahabhain	Glenfiddich	Glenkinchie	Tomintoul

Single Malt	Recommendation 1	Recommendation 2	Recommendation 3	Palate Challenger
Blair Athol 12 Year	Dailuaine	Glenfarclas	Dalmore	GlenDeveronMacduff
Bowmore 12 Year	Highland Park	Tomatin	Benromach	Deanston
Bruichladdich 10 Year	Benromach	Inchgower	Fettercairn	Glenturret
Bunnahabhain 12 Year	Bladnoch	Glenallachie	Glenlossie	GlenGarioch
Caol Ila 12 Year	Ardbeg	Clynelish	Isle of Jura	GlenScotia
Cardhu 12 Year	Aultmore	AnCnoc	Edradour	Mannochmore
Clynelish 14 Year	Ardbeg	Talisker	Caol Ila	Springbank
Craigellachie 14 Year	Benromach	BenNevis	Speyburn	Bruichladdich
Cragganmore 12 Year	Balblair	Tamnavulin	Strathmill	Tullibardine
Dailuaine 16 Year	RoyalLochnagar	Dalmore	BlairAthol	GlenDeveronMacduff
Dalmore 12 Year	RoyalLochnagar	Dailuaine	Glenrothes	Tormore
Dalwhinnie 15 Year	Benrinnes	Speyburn	Benromach	Bruichladdich
Deanston 12 Year	Benrinnes	Aberlour	Dalwhinnie	Bruichladdich
Dufftown 15 Year	Glenfiddich	Glenallachie	Bunnahabhain	GlenGarioch
Edradour 10 Year	Scapa	GlenGrant	Speyside	Tobermory
Glen Deveron 10 Year	Ardmore	Tomatin	Benromach	Balvenie
Glen Elgin 12 Year	Glenlossie	Glenfiddich	Glenkinchie	GlenGarioch
Glen Garioch 12 Year	Glenrothes	Glenfarclas	Tormore	RoyalLochnagar
Glen Grant 10 Year	Macallan	Glenrothes	RoyalLochnagar	GlenDeveronMacduff
Glen Keith 10 Year	Macallan	RoyalLochnagar	Auchroisk	GlenDeveronMacduff
Glen Moray 12 Year	Glenlivet	Linkwood	Glenallachie	GlenGarioch
Glen Ord 12 Year	Glenrothes	Auchroisk	Blair Athol	Balmenach
Glen Scotia 14 Years	Dufftown	Glenallachie	Bladnoch	GlenGarioch
Glen Spey 12 Years	Royal Brackla	Glen Ord	Glenkinchie	Dufftown
Glenallachie 12 Year	Auchentoshan	Strathisla	GlenGrant	Miltonduff
Glendroanch 12 Years	Edradour	Glengoyne	Speyside	Tobermory
Glendullan 12 Year	Knockando	Strathmill	Balblair	Teaninich
Glenfarclas 10 Year	Glenlossie	Benriach	Glenallachie	Glen Spey
Glenfiddich 12 Year	Glen Elgin	Linkwood	Bladnoch	GlenGarioch
Glengoyne 10 Year	Glenkinchie	Glenallachie	Benriach	ArranIsleOf
Glenkinchie 10 Year	Strathmill	Knockando	Glen Keith	Tullibardine
Glenlivet 12 Year	Auchentoshan	Cardhu	Glengoyne	Aberfeldy
Glenlossie 10 Year	Benriach	Glenlossie	Glenlivet	Tomintoul
Glenmorangie 10 Year	Auchroisk	Macallan	Glenfarclas	Mortlach
Glenrothes 12 Year	Isle of Jura	Laphroig	Oban	Caol Ila
Glenturret 10 Year	Bladnoch	Glenlivet	Glenallachie	Benriach
Highland Park 12 Year	Tomatin	Speyburn	Inchgower	Bruichladdich
Inchgower 14 Year	Tomatin	Bowmore	Dalwhinnie	Bruichladdich
Isle of Jura 10 Year	Speyburn	Glenturret	Tomatin	Ardmore
Knockando 12 Year	Caol Ila	Ardbeg	GlenScotia	Lagavulin
Lagavulin 16 Year	GlenKeith	Strathmill	Glenmorangie	Teaninich
Laphroig 12 Year	Laphroig	Clynelish	Ardbeg	Springbank
Linkwood 12 Year	Lagavulin	GlenScotia	Clynelish	OldPulteney
Loch Lomond	Glenlivet	GlenElgin	ArranIsleOf	Loch Lomond
Longmorn 15 Year	Bunnahabhain	Glenlossie	Glenallachie	Linkwood

Single Malt	Recommendation 1	Recommendation 2	Recommendation 3	Palate Challenger
Macallan 10 Year	Speyside	Aultmore	Mannochmore	AnCnoc
Mannochmore 12 Year	Glendronach	Auchroisk	Glenrothes	GlenDeveronMacduff
Miltonduff 15 Year	Longmorn	Speyside	Aultmore	Tamdhu
Mortlach 16 Year	Aultmore	Edradour	Speyside	Strathisla
Oban 14 Year	Dailuaine	Balmenach	BlairAthol	GlenDeveronMacduff
Fettercairn 12 Year	Ardbeg	Talisker	Clynelish	Laphroig
Old Pulteney 12 Year	Isle of Jura	Caol Ila	Oban	Springbank
Royal Brackla 10 Year	Glenfiddich	Glenkinchie	Benriach	Loch Lomond
Royal Lochnagar 12 Year	Dailuaine	Dalmore	Macallan	GlenDeveronMacduff
Scapa 14 Year	Aultmore	Edradour	Speyside	GlenMoray
Speyburn 10 Year	Dalwhinnie	Inchgower	Balvenie	Bruichladdich
Speyside 12 Year	Aultmore	Longmorn	Scapa	Auchentoshan
Springbank 10 Year	Isle of Jura	Talisker	GlenScotia	OldPulteney
Strathisla 12 Year	Glengoyne	GlenGrant	Longmorn	Miltonduff
Strathmill 12 Year	Glenmorangie	GlenKeith	Tamnavulin	Teaninich
Talisker 12 Year	Ardbeg	Clynelish	Oban	OldPulteney
Tamdhu 10 Year	Tobermory	AnCnoc	Scapa	Mannochmore
Tamnavulin 12 Year	Strathmill	Craigganmore	Glenmorangie	Tullibardine
Teaninich 10 Year	Balblair	Glenmorangie	Strathmill	Tullibardine
Tobermory 10 Year	Aultmore	Tamdhu	Speyside	GlenGrant
Tomatin 12 Year	Highland Park	Glenturret	Speyburn	Bruichladdich
Tomintoul 10 Year	Glenlivet	Glenlossie	Glenallachie	GlenGarioch
Tormore 12 Year	Balmenach	Auchroisk	Glendronach	Dalmore
Tullibardine 12 Year	GlenKeith	Strathmill	Balblair	Teaninich

Endnotes

ⁱ <https://www.kaggle.com/koki25ando/scotch-whisky-dataset>

ⁱⁱ Wishart, David. *Whisky Classified: Choosing Single Malts by Flavour*. Pavilion Books, London. 2006 Edition

ⁱⁱⁱ <http://www.wishart.org/classification.html>