Scotch Whisky Characteristics Analysis

Robert Brown

Northcentral University

TIM-6520: Inferential Statistics and Predictive Analytics

Dr. William Tribbey

October 22, 2023

Scotch Whisky Characteristics Analysis

Introduction

The whisky industry presents a multifaceted arena characterized by a storied tradition and distinctive nuances that set each distillery apart. To traverse this elaborate domain, a succession of data-centric inquiries has unfolded, harnessing the precision of statistical analysis. These efforts have utilized several hypotheses testing methodologies, including the Chi-Square test, Bartlett's test, Levene's test, and ANOVA, to dissect eighty-six various distillery characteristic datasets retrieved from Kaggle.com (Ando, 2018). The principal goal is to demystify the elements that shape whisky characteristics and furnish robust predictive models pertinent to the industry's practical needs.

The targeted audience for this analysis comprises key decision-makers in the Scotch whisky industry. This includes master distillers focused on product quality, marketers aiming for optimal branding strategies, and industry analysts seeking investment opportunities. Agents from the services industry, such as fine dining and hospitality sectors, are part of this audience.

This diverse group of stakeholders embodies the intricate web of relationships within the spirits industry and illustrates the far-reaching impact of data-driven decision-making. Their endorsement of the analytical findings is critical, as it not only validates the identified correlations but also influences strategic decisions that reverberate through the supply chain, branding, and consumer engagement. The ensuing hypotheses, therefore, are not merely academic exercises but pivotal in shaping industry practices and customer experiences. This sets the stage for a rigorous examination of the data, paving the way for a deeper understanding of the nuanced interplay between whisky characteristics and market dynamics.

Hypothesis

These hypotheses serve as the foundation for subsequent statistical testing and analysis. The study posits the following hypotheses to guide the inquiry:

* *H01*: There is no significant correlation between whisky characteristics and the geographic locations of the distilleries. This hypothesis suggests that the taste profile of a whisky is independent of its place of origin.
* *Ha1*: A significant correlation exists between whisky characteristics and the geographic locations of the distilleries, suggesting a regional influence on the taste profile.
* *H02*: There is no significant difference in whisky traits across different clusters of distilleries. This indicates that clustering does not reveal distinct groupings based on taste profiles.
* *Ha2*: There is a significant difference in whisky traits across different clusters of distilleries, indicating that distinct taste profiles can be associated with specific clusters.
* *H03*: Whisky characteristics cannot be significantly predicted based on the available dataset. It assumes the dataset lacks the predictive power to determine specific whisky characteristics.
* *Ha3*: Whisky characteristics can be significantly predicted based on the available dataset, asserting the predictive capacity of the dataset in determining whisky traits.

These hypotheses will be tested using statistical methods to discern the validity of the assumed correlations, differences, and predictability, thus informing the analytical direction of the research (Fisher, 1918).

Analysis Methodologies

This study's methodology adopted a multi-faceted, data-driven approach to dissect the intricacies inherent in whisky characteristics. The primary dataset, sourced from Kaggle (Ando, 2018), included 12 distinct attributes, such as body, sweetness, and smokiness.

K-means clustering, known for its simplicity and efficiency (McQueen, 1967), was employed to segment whiskies based on their taste profiles. Determining the optimal number of clusters was achieved using the Elbow Method, resulting in the selection of five distinct groups.

To correct for the non-normal distribution of certain variables within the whisky dataset, z-score normalization was implemented. This standardization technique is pivotal in inferential statistics to neutralize the effect of differing variances and bring each attribute to a common scale with a mean of zero and a standard deviation of one (Glass, Peckham, & Sanders, 1972). Such normalization ensures that the subsequent inferential tests, like t-tests and ANOVA, are not biased by the scale of the data.

The study then proceeded with hypothesis testing, employing both parametric and non-parametric statistical methods. The parametric t-tests were applied to compare group means. Nonparametric tests such as Spearman's Rank correlation and ANOVA were used to explore the relationships and differences between the various whisky attributes (Spearman, 1904; Fisher, 1918). This blend of statistical tests provided a thorough examination of the dataset, enhancing the interpretation of the data and supporting the study's conclusions.

Testing and Output Performances

The exploratory phase of the analysis was underpinned by K-means clustering, which provided an intuitive grouping based on whisky characteristics. K-means clustering was initially chosen due to its simplicity and efficiency (McQueen, 1967). This algorithm effectively segmented whiskies based on multiple taste profiles. The Elbow Method (see Figure 1) was utilized to ascertain the optimal number of clusters, settling on five as the most appropriate.

Figure 1

Whisky Elbow Plot

A graph of a number of clusters

Description automatically generated

Note. Elbow Plot was created using the ‘matplotlib’ library and Kmeans from the ‘scikit-learn’ library in Python 3.12 (Hunter, 2007; Pedregosa et al., 2011).

Figure 2 shows that this clustering information provided an initial snapshot of the characteristics and their relation to the regional distilleries before a z-score normalization.

Figure

Clustered Distillery Data Summary

A close-up of a map

Description automatically generated

Note. This early run figure, prior to a z-score normalization, illustrates the results of a K-means clustering algorithm applied to various whisky distilleries and categorized by flavor profile attributes. Each color represents a different cluster, suggesting a grouping of distilleries with similar characteristics. The graphic was created using the ‘scikit-learn’ library in Python 3.12 (Pedregosa et al., 2011).

To manage the non-normal distribution of certain variables within the whisky dataset, z-score normalization was utilized to standardize the data, thereby enabling more effective manipulation (Walker, 2020). This preprocessing step is critical when applying clustering algorithms like K-means clustering, assuming all features have the same scale (Jain et al., 1999).

The z-score normalization process adjusts the data to have a mean of 0 and a standard deviation of 1, thus harmonizing the variance across different attributes (Massaron & Boschetti, 2016). The subsequent analysis leveraged parametric (e.g., t-tests) and non-parametric tests (e.g., Spearman's Rank, ANOVA), accommodating the unique properties of whisky characteristics. The combination of these methodologies, underpinned by the preparatory normalization, provided a comprehensive and robust analysis. The K-means clustering identified distinct groups within the data, as Figure 3 illustrates, demonstrating the practical utility of these preparatory steps in revealing patterns and insights.

Figure

Simplified K-means Cluster After Z-score Normalization

A diagram of different colored triangles

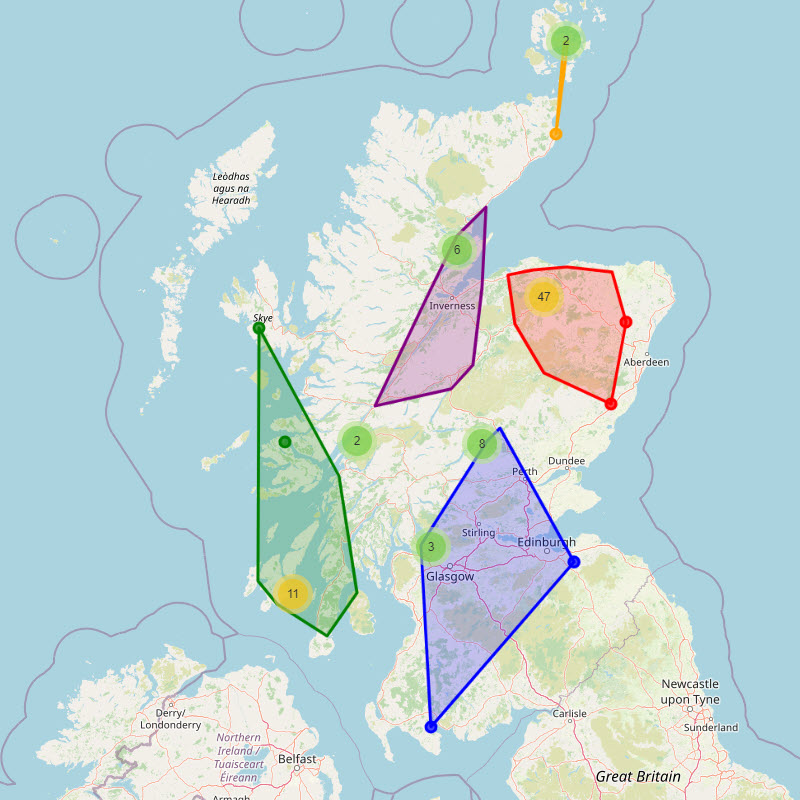
Description automatically generated

Note. This normalization of the cluster points through z-score standardization clarifies the data, offering a more refined view of the Distillery characteristics. It ensures that despite the transformation for scale uniformity, the inherent relationships between the original clustering points remain intact. This analytical step is facilitated by the ‘scikit-learn**’** Python library (Pedregosa et al., 2011).

With the new K-means cluster analysis (Figure 4), visualized against the latitude and longitude in a Folium map, it is now possible to reflect that data grouping to a geographical map, offering an innovative way for stakeholders to consider distillery characteristics about their geographical location.

Figure

Summary Geographical Representation of K-means Results



Note. The geographical representation of the K-means clustering output onto the actual locations of the distilleries illustrates potential regional patterns in whisky flavor profiles. This adds a dimension of analysis that enhances the comprehensiveness of the data interpretation, providing a tangible connection between the abstract clusters and real-world locations. The figure was created using the ‘Folium’ library in Python 3.12 (Folium Contributors, 2022).

Subsequent to K-means clustering, a Chi-Square test quantitatively assessed associations between whisky characteristics, with significant findings indicating strong dependencies among specific attributes such as 'Body' and 'Smoky' (p-value = 0.001).

The low p-values in Table 1 indicate statistically significant associations between sensory notes, demonstrating how distinct flavors can co-occur or vary independently, thus providing essential insight for crafting targeted product development.

Table

Chi-Square Test P-values

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Traits*** | ***Body*** | ***Sweetness*** | | ***Smoky*** | ***Medicinal*** | ***Tobacco*** | | ***Honey*** | | ***Spicy*** | ***Winey*** | ***Nutty*** | ***Malty*** | ***Fruity*** | ***Floral*** |
| ***Body*** | NA | | 0.612 | 0.001 | 0.001 | 0.651 | 0.069 | | 0.450 | | 0.008 | 0.872 | 0.283 | 0.487 | 0.007 |
| ***Sweetness*** | 0.612 | | NA | 0.005 | 0.002 | 0.261 | 0.233 | | 0.909 | | 0.227 | 0.916 | 0.573 | 0.155 | 0.753 |
| ***Smoky*** | 0.001 | | 0.005 | NA | 0.000 | 0.000 | 0.047 | | 0.051 | | 0.985 | 0.078 | 0.451 | 0.130 | 0.002 |
| ***Medicinal*** | 0.001 | | 0.002 | 0.000 | NA | 0.003 | 0.213 | | 0.332 | | 0.850 | 0.078 | 0.580 | 0.143 | 0.000 |
| ***Tobacco*** | 0.651 | | 0.261 | 0.000 | 0.003 | NA | 0.032 | | 0.655 | | 0.719 | 0.859 | 0.628 | 0.095 | 0.015 |
| ***Honey*** | 0.069 | | 0.233 | 0.047 | 0.213 | 0.032 | NA | | 0.011 | | 0.196 | 0.865 | 0.059 | 0.528 | 0.796 |
| ***Spicy*** | 0.450 | | 0.909 | 0.051 | 0.332 | 0.655 | 0.011 | | NA | | 0.288 | 0.977 | 0.898 | 0.904 | 0.636 |
| ***Winey*** | 0.008 | | 0.227 | 0.985 | 0.850 | 0.719 | 0.196 | | 0.288 | | NA | 0.568 | 0.021 | 0.136 | 0.461 |
| ***Nutty*** | 0.872 | | 0.916 | 0.078 | 0.078 | 0.859 | 0.865 | | 0.977 | | 0.568 | NA | 0.000 | 0.927 | 0.402 |
| ***Malty*** | 0.283 | | 0.573 | 0.451 | 0.580 | 0.628 | 0.059 | | 0.898 | | 0.021 | 0.000 | NA | 0.695 | 0.788 |
| ***Fruity*** | 0.487 | | 0.155 | 0.130 | 0.143 | 0.095 | 0.528 | | 0.904 | | 0.136 | 0.927 | 0.695 | NA | 0.217 |
| ***Floral*** | 0.007 | | 0.753 | 0.002 | 0.000 | 0.015 | 0.796 | | 0.636 | | 0.461 | 0.402 | 0.788 | 0.217 | NA |

*Note.* Created using R language ‘knitr’, and ‘tidyr’ libraries (R Core Team, 2022; Wickam et al., 2023; Wickham & Grolemund, 2017; Xie, 2015).

A Chi-Square statistic of 39.164 for the association between 'Body' and 'Smoky' with a p-value of 0.001 against 16 degrees of freedom indicates a significant relationship beyond chance. Table 2 presents the expected frequencies under the null hypothesis of no association derived from the marginal totals and sample size, reinforcing the observed dependencies.

**Table 2**

Summary of Chi-test results of 'Body' and 'Smoky'

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ***Category 1*** | ***Category 2*** | ***Category 3*** | ***Category 4*** | ***Category 5*** |
| **Row 1** | 0.093 | 1.047 | 0.651 | 0.116 | 0.093 |
| **Row 2** | 0.884 | 9.94 | 26.186 | 1.105 | 0.884 |
| **Row 3** | 2.093 | 23.547 | 14.651 | 2.616 | 2.093 |
| **Row 4** | 0.512 | 5.756 | 3.581 | 0.640 | 0.512 |
| **Row 5** | 0.419 | 4.709 | 2.930 | 0.523 | 0.419 |

*Note.* Expected frequencies assume independence under the null hypothesis. The significant Chi-Square values suggest a departure from independence, confirming the associations between categories.

This statistical validation provides empirical evidence for the relationships between variables, guiding data-driven decision-making and preempting bias in the analysis. Bartlett's and Levene's tests verified the homogeneity of variances (Table 3), validating the use of specific parametric tests in the analysis. Bartlett's test for homogeneity of variances yielded a test statistic of 2.747, accompanied by a p-value of 0.601, suggesting the absence of significant variance differences across the groups for the 'Body' characteristic of whiskies.

Table

Summary of Bartlett's and Levene's Statistical Tests

|  |  |  |  |
| --- | --- | --- | --- |
| ***Statistical Measure*** | ***Statistic*** | ***P-value*** | ***Significant*** |
| Bartlett | 2.747 | 0.601 | False |
| Levene | 1.106 | 0.360 | False |

*Note.* Homogeneity of variances verified by Bartlett's and Levene's tests facilitates the use of ANOVA. Data was computed with the SciPy library (Virtanen et al., 2020).

Consequently, this result supports the assumption of variance homogeneity, a prerequisite for applying specific parametric tests like Spearman’s Rank and ANOVA. The z-score normalized values for the 'Body' attribute were consistent across the clusters, reinforcing the robustness of the clustering methodology and justifying subsequent parametric analyses. This finding is pivotal for stakeholders as it confirms the reliability of the data's structure, allowing for more complex, assumption-dependent statistical tests to be confidently utilized in further scrutiny of the whisky characteristics.

The Spearman Rank Correlation analysis offered a comprehensive view of the relationships between the various whisky characteristics, as seen in Table 4.

Table

Summary of Spearman’s Ranking

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Trait*** | ***Body*** | ***Sweetness*** | ***Smoky*** | ***Medicinal*** | ***Tobacco*** | ***Honey*** | ***Spicy*** | ***Winey*** | | ***Nutty*** | | ***Malty*** | | ***Fruity*** | ***Floral*** |  |
| ***Body*** | 1 | -0.128 | 0.452 | 0.191 | 0.167 | 0.1 | 0.202 | | 0.356 | | 0.127 | | -0.102 | 0.017 | -0.419 | | |
| ***Sweetness*** | -0.128 | 1 | -0.371 | -0.316 | -0.131 | 0.108 | -0.042 | | 0.08 | | -0.046 | | -0.022 | 0.044 | 0.14 | | |
| ***Smoky*** | 0.452 | -0.371 | 1 | 0.504 | 0.293 | -0.083 | 0.275 | | -0.017 | | 0.054 | | -0.15 | -0.274 | -0.38 | | |
| ***Medicinal*** | 0.191 | -0.316 | 0.504 | 1 | 0.352 | -0.366 | 0.061 | | -0.246 | | -0.157 | | -0.282 | -0.319 | -0.379 | | |
| ***Tobacco*** | 0.167 | -0.131 | 0.293 | 0.352 | 1 | -0.269 | 0.067 | | 0.048 | | -0.118 | | -0.078 | -0.227 | -0.249 | | |
| ***Honey*** | 0.1 | 0.108 | -0.083 | -0.366 | -0.269 | 1 | 0.103 | | 0.373 | | 0.179 | | 0.283 | 0.067 | 0.15 | | |
| ***Spicy*** | 0.202 | -0.042 | 0.275 | 0.061 | 0.067 | 0.103 | 1 | | 0.066 | | -0.022 | | 0.016 | 0.122 | 0.004 | | |
| ***Winey*** | 0.356 | 0.08 | -0.017 | -0.246 | 0.048 | 0.373 | 0.066 | | 1 | | 0.171 | | 0.153 | 0.077 | -0.114 | | |
| ***Nutty*** | 0.127 | -0.046 | 0.054 | -0.157 | -0.118 | 0.179 | -0.022 | | 0.171 | | 1 | | 0.034 | 0.081 | -0.005 | | |
| ***Malty*** | -0.102 | -0.022 | -0.15 | -0.282 | -0.078 | 0.283 | 0.016 | | 0.153 | | 0.034 | | 1 | 0.209 | 0.11 | | |
| ***Fruity*** | 0.017 | 0.044 | -0.274 | -0.319 | -0.227 | 0.067 | 0.122 | | 0.077 | | 0.081 | | 0.209 | 1 | 0.263 | | |
| ***Floral*** | -0.419 | 0.14 | -0.38 | -0.379 | -0.249 | 0.15 | 0.004 | | -0.114 | | -0.005 | | 0.11 | 0.263 | 1 | | |

*Note.* The Spearman's Rank Correlation Coefficient is a nonparametric measure of rank correlation that assesses the strength and direction of the association between two ranked variables. It is less prone to errors in distribution than a Pearson Correlation and provides insights into monotonic relationships where variables move together but not necessarily at a constant rate. In Python, the ‘scipy.stats’ library offers a Spearman's Rank Correlation implementation via the ‘spearman’ function (Virtanen et al., 2020).

To decipher these values, where there are strong positive correlations between attributes such as 'Body' and 'Smoky', and 'Medicinal' and 'Smoky', it indicates that whiskies with a fuller body tend to be smokier, and those with smokier profiles often have more medicinal qualities. In contrast, negative correlations were observed between 'Body' and 'Floral', and 'Sweetness' and 'Smoky', suggesting that lighter-bodied whiskies might have more floral notes, and those with sweeter profiles are less likely to be smoky. These correlations provide a nuanced understanding of the intrinsic relationships between the different whisky flavor profiles.

The ANOVA tests conducted on the clustered data revealed significant differences across the whisky clusters for all characteristics, with extreme disparities observed in traits like 'Medicinal', 'Body', and 'Smoky'. These significant F-values and very small P-values suggest that the clusters are distinct in their defining characteristics, as seen in Table 5.

Table

ANOVA Results for Whisky Traits

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Traits*** | ***Sum of Squares*** | ***Degrees of Freedom*** | ***F-value*** | ***p-value*** |
| ***Body*** | 51.16 | 4 | 29.73 | <.00001 |
| ***Sweetness*** | 17.28 | 4 | 5.09 | .00104 |
| ***Smoky*** | 41.26 | 4 | 18.67 | <.00001 |
| ***Medicinal*** | 58.30 | 4 | 42.63 | <.00001 |
| ***Tobacco*** | 29.59 | 4 | 10.62 | <.00001 |
| ***Honey*** | 35.19 | 4 | 14.03 | <.00001 |
| ***Spicy*** | 9.14 | 4 | 2.41 | .056 |
| ***Winey*** | 45.83 | 4 | 23.11 | <.00001 |
| ***Nutty*** | 15.25 | 4 | 4.37 | .00302 |
| ***Malty*** | 21.14 | 4 | 6.60 | .00012 |
| ***Fruity*** | 30.94 | 4 | 11.38 | <.00001 |
| ***Floral*** | 36.56 | 4 | 14.97 | <.00001 |

*Note.* The ANOVA results highlight significant differences in whisky trait means across different clusters, indicating that the clusters can be distinguished based on their sensory attributes. Traits such as 'Medicinal', 'Body', and 'Smoky' displayed high F-values, suggesting substantial inter-cluster variability. These results imply that the clustering methodology effectively separates whiskies into meaningful categories. For conducting ANOVA in Python, the ‘scipy.stats’ module provides the ‘f\_oneway’ function, part of the SciPy library (Virtanen et al., 2020).

For instance, 'Medicinal' qualities showed the highest F-value, indicating a solid differentiation in this trait across the clusters. 'Spicy', although significant, had a higher P-value than others, hinting at a more subtle variation in this characteristic among the different clusters.

These findings underscore the presence of distinct groupings within the whisky varieties, which aligns with the K-means clustering. These results are more than just academic exercises; they carry substantial weight for the stakeholders. This can have practical implications in marketing and product development, and for distilleries, the insights could guide product development. The data offers investors and suppliers a predictive lens into which distilleries or characteristics might be more lucrative. For the services industry, it provides a data-backed method to select premium spirits that align with their customer experience goals.

Summary of Findings

Throughout the analysis, various statistical techniques unearthed the intricate relationships between different whisky characteristics. K-means clustering initially segmented whiskies into distinct groups, revealing patterns and potential market niches. The Chi-Square test provided evidence of significant associations between sensory notes across all the whisky characteristics. Bartlett's and Levene's tests confirmed the homogeneity of variances, thus validating the application of ANOVA. The ANOVA results highlighted significant differences across clusters for various attributes, highlighting distinct profiles for each cluster.

Implications and Recommendations for Stakeholders

The findings reveal distinct clusters of whisky characteristics that offer actionable insights for industry stakeholders. Distilleries can capitalize on this data by tailoring their production techniques—such as barley selection, maturation time, and barrel characteristics—to enhance the flavors most prevalent in their cluster, aligning their offerings with identified market trends. This targeted approach can inform ingredient selection and processing methods, potentially leading to a more substantial market presence.

Suppliers and distributors should consider these insights to forge strategic partnerships with distilleries whose flavor profiles meet or create new consumer demands. For the service industry, including bars and restaurants, there is an opportunity to refine spirit selections to match the consumer palate better, using data to inform and enhance the overall consumer experience. As stakeholders contemplate the following steps, they are encouraged to investigate further variables like barley types, aging processes, and wood influence on flavors to deepen their understanding of whisky production. This alignment with current consumer trends and detailed flavor profiling is poised to open new avenues for market expansion and reinforce positions in existing markets, underpinning informed business strategies and future research endeavors in the whisky domain.

Conclusion

These data-driven insights offer a 'Go' for stakeholders to refine their operational, production, and marketing tactics. The evidence suggests there are concrete and measurable attributes that can be manipulated to influence whisky profiles, aligning with consumer preferences and creating opportunities for innovation.

The groundwork laid by this research invites a continuous loop of analysis and feedback where industry players can adapt to changing tastes and trends. Future research may involve a deeper dive into consumer preferences, production techniques, and even the effects of terroir on whisky characteristics, ensuring that the industry remains at the forefront of tradition and innovation.

References

Ando, K. (2018). Data analysis of whisky review dataset. Retrieved September 20, 2023, from <https://www.kaggle.com/code/koki25ando/data-analysis-of-whisky-review-dataset>

Fisher, R. A., Sir. (1918). The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh, 52, 399–433. <https://hdl.handle.net/2440/15097>

Folium Contributors. (2022). Folium: python data, Leaflet.js maps [Computer software]. <https://github.com/python-visualization/folium>

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 42(3), 237–288. <https://doi.org/10.3102/00346543042003237>

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. Computing in Science & Engineering, 9(3), 90–95. <https://doi.org/10.1109/mcse.2007.55>

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering. ACM Computing Surveys, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>

Massaron, L., & Boschetti, A. (2016). Regression analysis with python (1st ed.). Packt Publishing.

Pedregosa, F., Varoquaux, G., Gramfort, A., Blondel, M., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., & Perrot, M. (2011). Scikit-learn: Machine learning in python. JMLR.org. <http://jmlr.org/papers/v12/pedregosa11a.html>

R Core Team. (2022). R: A language and environment for statistical computing (R Foundation for Statistical Computing) [R Computer Language]. <https://www.R-project.org/>

Spearman, C. (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72. <https://doi.org/10.2307/1412159>

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A. J., Jones, E., Kern, R., Larson, E.,...Vázquez-Baeza, Y. (2020). Scipy 1.0: Fundamental algorithms for scientific computing in python. Nature Methods, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Walker, M. (2020). Python data cleaning cookbook: Modern techniques and python tools to detect and remove dirty data and extract key insights. Packt Publishing.

Wickham, H., & Grolemund, G. (2017). R for data science: Import, tidy, transform, visualize, and model data (1st ed.). O'Reilly Media. <https://r4ds.had.co.nz/>

Xie, Y. (2015). knitr: A general-purpose package for dynamic report generation in r [Cran - package knitr]. cran.r-project.org. <https://cran.r-project.org/web/packages/knitr/index.html>

Wickam, H., Vaughan, D., & Girlich, M. (2023, January 24). tidyr: Tidy messy data [Cran - package tidyr]. Posit.co. <https://cran.r-project.org/web/packages/tidyr/index.html>