Artificial Intelligence for Sustainable Societies

Introduction to Data Science

José Braga de Vasconcelos

jose.vasconcelos@ulusofona .pt

Iolanda Velho

iolanda.velho@ulusofona.pt

Data Science Project Report

Measuring User Experience Through Usability Testing to Drive Interface Improvement

Anton Donle

a22407849@alunos.ulht.pt

Ligia Anjos

a22407807@alunos.ulht.pt

Venus Schwidrowski

a22407863@alunos.ulht.pt

Abstract / Project Summary

The report outlines the project's progress, "Measuring User Experience Through Usability Testing to Drive Interface Improvement." Based on user survey data, the project uses design principles to evaluate how different online platforms (e.g., Facebook, Instagram, and Twitter) performed. It leverages tools like Numpy, Pandas, and Matplotlib to gather, compare, and visualize the data to achieve this.

The chosen dataset contains attributes like Name, Age, Gender, Platform, User Experience, and Design Ratings (Likert Scale). Data science techniques were employed to uncover patterns, such as data visualization, normalization, curve fitting, and classification models (Logistic Regression and Support Vector Machines). Findings revealed minimal differences in user ratings across platforms, suggesting similar user experiences. Additionally, average ratings were not significantly influenced by demographic factors like age.

Keywords

Data Science, Design, User Interface, Data visualization.

1. Introduction

Design is crucial in advocating for better user experiences across various projects. Through different specialities and tools, it aims to create objects, applications, websites, and a range of other projects that ensure interactions are as seamless as possible, minimizing barriers and user frustration.

Among all the values related to the design role, one particularly stands out: the concept of being data-driven. Design Thinking is a process that seeks to solve problems through a methodology that considers understanding the issue, exploring it and materializing the best possible solution. They all used data to ensure the team was making the right decisions. This project works in the materialized area, where designers search for insights from the users to improve the work. In this case, five platforms are evaluated using usability tests. (Stickdorn, Hormess, Lawrence, & Schneider, 2018, p. 400)

The goal is to ask users to complete various tasks within the interface and then evaluate how easy or difficult it was to accomplish each task once completed, often using a Likert scale from 1 to 5, (Stickdorn et al., 2018) where 1 is very difficult and 5 is very easy. This study allows designers to understand what improvements to make and create action plans accordingly. This approach is widely used within the digital product design world, so various companies and startups have similarly structured data, facilitating data acquisition.

The project investigates data to understand how different social media were evaluated differently by various age groups and sexes. Moreover, it tries to understand if different trends were visible in the evaluation different social media have received. The results show no significant difference between evaluations from different age groups and sexes. Moreover, the research showed that there is not a strong enough correlation between the ratings a user gave and the social media they evaluated. Both SVM and Linear Regression (Pedregosa et al., 2011) model could not achieve an accuracy more significant than random guessing would have.

Our data analysis process started with data preparation, checking for missing values, visualizing simple features like the age in a histogram and then changing the existing dataset in two ways: Firstly, replacing the names with an anonymized unique ID, and secondly, adding a column of "average" ratings given by the users. We also showed that our data is normally distributed, which makes it suitable for further experimentation.

For this project, the following research questions were investigated: (1) differences between age groups, (2) differences between social media platforms, and (3) correlation between user ratings and social media platforms. The results indicate no significant differences observed across any of the examined dimensions.

2. Datasets description

The Dataset was found in Kaggle, a data science platform. It has 1482 answers with the following features: Name, Age, Gender, Platform, User Experience (qualitative), and Ratings on design (Likert Scale). The dataset was already "cleaned" and no null values were found. There were also no duplicates in our dataset. It seemed that someone had already done data preparation previously, which makes sense as we downloaded it from Kaggle (Rahman, M. A. 2023).

In Figure 1, we can see that the age is relatively evenly distributed amongst the participants.

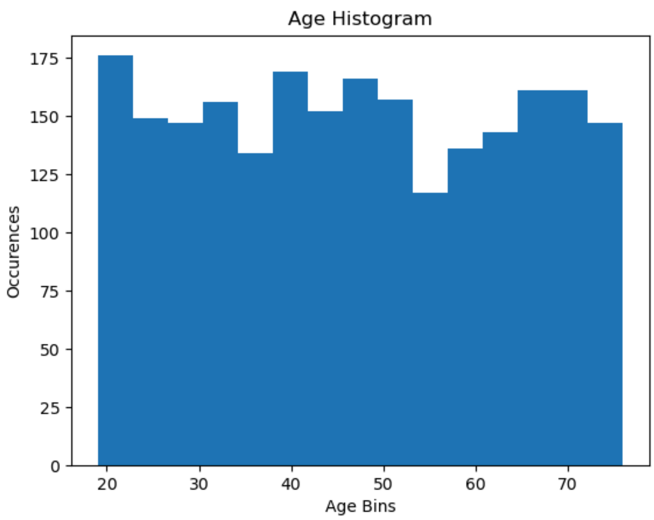

Figure 1: Age distribution in the dataset.

3. Methods and algorithms

In this project different methods were used. They can be broadly grouped into data preparation, data visualization/understanding of the data, curve fitting (linear regression, normal curve estimation) and classification (Logistic Regression and Support Vector

Machine)

3.1 Data preparation

The data was cleaned to determine if missing or duplicated data was in the dataset and to ensure an accurate investigation. For the data on a Likert scale, simple python logic was used to see if all those data points were between 1 and 5. This validation was relevant to identify possible inconsistencies in the data before proceeding with the analyses. Dictionaries were created to help the team to access the data easily.

3.1.1. **Anonymizing Names**

To protect participants' privacy, the initial measure undertaken involved the 'anonymization' of participant's names by replacing them with a unique 5-digit number. A random sample was drawn from a distribution between 10000 and 99999 without replacement. This was done in order to protect privacy while still being able to distinguish them using a unique ID. The column "name" was kept and replaced with the number.

3.1.2. **Feature Engineering**

According to the correlation matrix (Figure 2), some features have a small correlation with others (either positive or negative). Mostly this stays between -0.01 and 0.01. For further experiments and an easier insight into the data and the rating users gave, a column was added that contained the average rating a user gave. We only used the 17 features that were evaluated on a Likert scale (1-5) for this. This column is called 'avg'.

	Age	Color Scheme	Visual Hierarchy	Typography	Images and Multimedia	Layout	Mobile Responsiveness	CTA (Call to Action) Buttons	Forms and Input Fields	Feedback and Error Messages	Loading Speed	Personalization	Accessibility	Animation and Transitions	Scrolling_Behavior	Gestures and Touch Controls
Age	1.000000	-0.072052	0.115494	-0.013218	0.001590	-0.002016	-0.024842	-0.019818	-0.067698	0.036173	-0.004936	-0.042122	-0.058782	0.047702	-0.016939	0.022131
Color Scheme	-0.072052	1.000000	0.041483	-0.001928	0.088402	0.035608	0.049323	-0.000166	-0.056013	0.027489	0.018706	0.019207	0.015929	0.018878	-0.049969	-0.028058
Visual Hierarchy	0.115494	0.041483	1.000000	0.005805	0.041046	0.056024	0.031366	-0.002422	-0.041295	0.026820	-0.014754	0.027454	0.025974	-0.005651	-0.010954	0.013296
Typography	-0.013218	-0.001928	0.005805	1.000000	-0.010972	0.021637	0.009615	-0.022117	-0.053230	0.042775	0.007485	0.014736	0.021219	0.025093	0.037836	-0.044661
Images and Multimedia	0.001590	0.088402	0.041046	-0.010972	1.000000	0.026768	0.029732	0.037834	0.037069	-0.033277	0.003191	-0.020954	-0.007296	-0.042262	-0.006653	-0.079309
Layout	-0.002016	0.035608	0.056024	0.021637	0.026768	1.000000	0.009058	0.030536	-0.004363	0.096735	0.046454	-0.096379	0.033398	0.002677	0.072914	0.052784
Mobile Responsiveness	-0.024842	0.049323	0.031366	0.009615	0.029732	0.009058	1.000000	-0.004462	-0.026264	0.022437	0.044211	0.016545	-0.018787	0.033838	0.015624	-0.006693
CTA (Call to Action) Buttons	-0.019818	-0.000166	-0.002422	-0.022117	0.037834	0.030536	-0.004462	1.000000	0.117483	-0.026514	-0.037515	-0.050553	-0.055498	-0.006489	-0.002387	0.036113
Forms and Input Fields	-0.067698	-0.056013	-0.041295	-0.053230	0.037069	-0.004363	-0.026264	0.117483	1.000000	-0.033679	-0.056167	0.001199	0.014434	0.066536	0.044515	0.039495
Feedback and Error Messages	0.036173	0.027489	0.026820	0.042775	-0.033277	0.096735	0.022437	-0.026514	-0.033679	1.000000	0.048343	0.045062	0.098807	-0.034288	0.015507	-0.058794
Loading Speed	-0.004936	0.018706	-0.014754	0.007485	0.003191	0.046454	0.044211	-0.037515	-0.056167	0.048343	1.000000	-0.069109	-0.038181	-0.064543	0.018856	0.044116
Personalization	-0.042122	0.019207	0.027454	0.014736	-0.020954	-0.096379	0.016545	-0.050553	0.001199	0.045062	-0.069109	1.000000	-0.048693	-0.031258	-0.018019	-0.066220
Accessibility	-0.058782	0.015929	0.025974	0.021219	-0.007296	0.033398	-0.018787	-0.055498	0.014434	0.098807	-0.038181	-0.048693	1.000000	-0.075885	0.080637	-0.017914
Animation and Transitions	0.047702	0.018878	-0.005651	0.025093	-0.042262	0.002677	0.033838	-0.006489	0.066536	-0.034288	-0.064543	-0.031258	-0.075885	1.000000	0.032887	0.066757
Scrolling_Behavior	-0.016939	-0.049969	-0.010954	0.037836	-0.006653	0.072914	0.015624	-0.002387	0.044515	0.015507	0.018856	-0.018019	0.080637	0.032887	1.000000	0.005815
Gestures and Touch Controls	0.022131	-0.028058	0.013296	-0.044661	-0.079309	0.052784	-0.006693	0.036113	0.039495	-0.058794	0.044116	-0.066220	-0.017914	0.066757	0.005815	1.000000
Search Functionality	-0.045049	0.044369	0.030249	0.003797	-0.071549	0.032595	-0.079171	0.081504	0.005795	-0.004476	-0.041061	-0.048968	-0.051060	-0.008985	0.006083	0.117959
Social_Media_Integration	0.053130	-0.025676	0.109005	-0.035134	-0.039041	-0.067551	0.042966	-0.052126	-0.005759	0.018040	-0.006530	0.041832	-0.017889	0.060699	-0.006045	0.020293

Figure 2: correlation matrix

3.2. Understanding the Data

3.2.1. Qualitative measure: User Experience by social media

To analyse the relationship between the type of qualitative reviews (e.g., "Confusing") and their frequency, a bar chart was created. The X-axis represents various social media platforms, while the Y-axis displays the occurrence of each qualitative term (e.g., "Engaging") for each platform. This visualization facilitates easy comparison of how frequently specific terms are associated with each platform.

The analysis only takes into consideration the 'User_experience' columns within the dataset, which contains qualitative reviews. The occurrences of each review type were counted, and users were grouped based on the social media platform they evaluated. The frequency of each specific review type was then calculated for each platform.

The value_counts() function was applied to this column with the parameter normalize=True to normalize the data for each social media platform. Finally, a graph was plotted with an appropriate title, labels for the X and Y axes, and a legend to ensure clarity.

Furthermore, the feedback was divided into two categories positive_categories = ["User-Friendly", "Engaging", "Efficient", "Clear and concise", "Well-structured", "Intuitive", "Adequate"] and negative_categories = ["Confusing", "Inconsistent Navigation", "Limited Menu Options"] creating two graphs.

3.2.2. Mean evaluation by website and category

A new data frame (platform_means_df) was created to store the means of each platform for each evaluated item. The items were in a list called evaluation_columns to allow access only to the quantitative items. Thus, it was possible to determine which website performed lowest in each evaluation column.

Each platform was also evaluated separately to clearly compare its average scores and analyse its performance in each of the evaluated categories. The performance was plotted by generating a bar chart using the average scores for each evaluation feature. To enhance the visualization, a horizontal line representing the average score of the platform was added and, because the results are usually close to each other, The y-axis limits were adjusted to focus on a specific range of values, from 3.5 to 4.5 to improve the clarity of the data visualization.

3.2.3. Average Ratings

To understand the average ratings given overall and to see whether those follow a certain trend, a histogram was plotted counting the frequency of scores that all platforms received. Furthermore, to see if there is any obvious difference in average scores of different social media, we will make normalised versions of this histogram for each of the five social media platforms separately.

3.2.4. Average Rating by Age

Age groups were created by defining bins in intervals of five years, these age ranges were set as follows: 19–24, 25–29, 30–34, 35–39, 40–49, 50–54, 55–59, 60–64, 65–69, and 70–75. This resulted in ten bins of nearly equal size, with the exception of the last bin, which is slightly larger.

Participants were then grouped according to these bins, and the average rating provided by each age group was calculated. The results were visualised by plotting the age groups on the X-axis and their corresponding average ratings on the Y-axis, enabling a clear comparison of the data.

3.3 Curve Fitting: Fitting a normal distribution to average ratings

To understand whether our data follows a normal distribution, we will estimate the parameters and plot a normal distribution over the histogram of occurrences of average ratings.

Fitting a normal distribution will help us to understand the data better by being able to describe it with two parameters: mean and standard deviation.

Furthermore, a linear regression will be used to compare whether the data follows a linear trend (even distribution), or a normal distribution.

If data is normally distributed, it can be easier to train a Machine Learning model on it and furthermore, statistical tests assume a normal distribution of data (Altman, D. G., & Bland, J. M. (1995)). These statistical tests are not very relevant to the research, but it is good practice in a data science report.

3.4. Classification

To further analyse whether the numerical features on a Likert scale are indicative of which social media was evaluated, classification will be performed. Based on the numerical features: 17 values on a Likert scale, and the participants' ages, the classification algorithm will predict which platform has been evaluated. There are 5 different platforms, so for the algorithm to outperform random guessing, it should be correct in more than 20% of its guesses. This approach is called multiclass classification.

Two different algorithms will be used: A Support Vector Machine (Smola, A. J., &

Schölkopf, B. (2004)) and a logistic regression model. The logistic regression model will be using a multinomial, which means ”the loss minimised is the multinomial loss fit across the entire probability distribution”(Smola, A. J., & Schölkopf, B. (2004), Pedregosa et al., 2011). The alternative that could have been used is one versus rest (ovr) evaluation.

The dataset is randomly split into 80% training data and 20% test data. For both experiments this split will stay the same.

For the SVM we will test the performance for different model parameters C. The values C= [1,5,10,25,50,100,150] will be tested, and it will be investigated if the classification model performs better with a certain value for C.

For the logistic regression model training data samples of the size s = [25,50,75,100,200,300,500,1000,1800] were used iteratively, and model performance is plotted for each of them.

For analysing the performance will be compared to a baseline of 1/5 (20%). This is how accurate random guessing would perform guessing one of five classes.

4. Project analysis and results

4.1. Understanding the data

The research with users was well done, it took into account different ages and genders, it evaluated users on different platforms with little difference in quantity between the samples. It's curious to think about how the scores are very close, even on different platforms. This could lead to the understanding that they are well-established platforms in the market and that they are the interface references.

4.1.1. Average Ratings

Plotting the average ratings in a histogram (Figure 3) might suggest that the data is normally distributed, which is to be expected in ordinal data from a Likert scale (Stickdorn et al., 2018). In section 4.2. of the results we will further investigate this. Plotting the average rating distributions by social media (Figure 4) shows that most social media received a fairly similar pattern of average ratings, with Facebook being slightly different with an overall lower distribution but one higher spike. However, a further analysis would be necessary to fully understand if there is a significant difference or if this follows a trend.

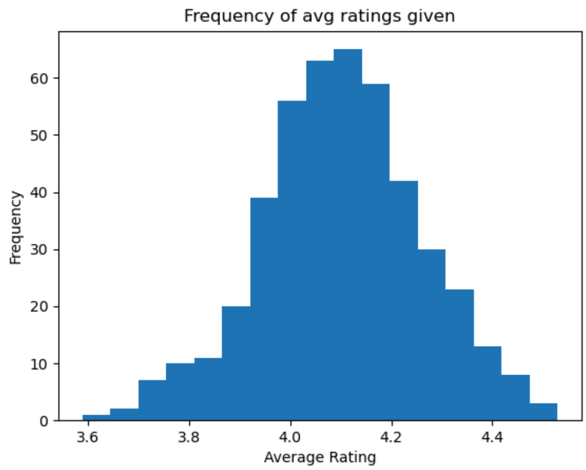

Figure 3: average ratings considering all the platforms.

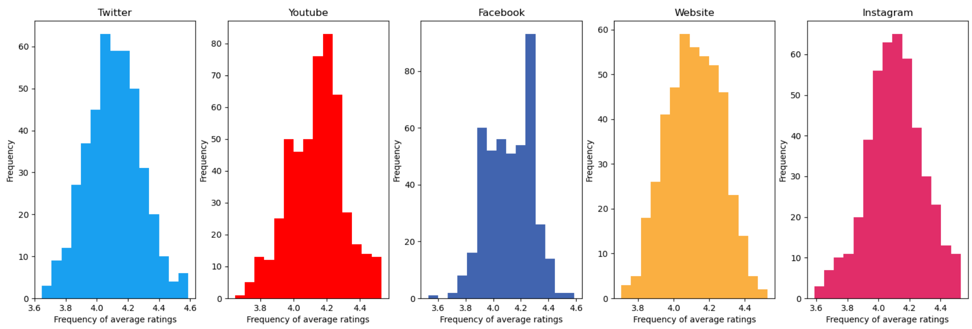

Figure 4: Average ratings to each platform by social media.

4.1.2. Qualitative measure: User Experience by social media

Overall, there is minimal variation in the frequency of reviews across the different review categories, as can be seen in Figure 5. The highest frequency, just under 0.14, is observed in the "Confusing" category, while the lowest, 0.6, is in the "Adequate" category. Despite this, the difference between these categories is notable. The minimum frequency for "Confusing" reviews is slightly below 0.12, whereas the maximum for "Adequate" is 0.8. It is also worth highlighting that other social media platforms in the "Adequate" category scored significantly lower, averaging around 0.6.

Among the 10 categories, Twitter recorded the highest scores in "Confusing" (0.14) and "Efficient" (0.10). However, it scored the lowest in the "User-Friendly" category, with a considerable drop to 0.8. Facebook and Websites did not show significantly higher or lower scores compared to other platforms in any category. In contrast, Instagram stood out in the "Adequate" category, scoring just above 0.8, compared to 0.6 for other platforms.

It is worth considering that Instagram's higher score in the "Adequate" category might be influenced by factors such as a larger user base. However, this is speculative, as the user base sizes for each platform were not researched in this study, and we normalized the data.

While slight variations exist between categories, none of the differences seem to be statistically significant. Categories such as "Confusing," "Clear and Concise," and "Inconsistent Navigation" received the most ratings overall. This suggests that users are particularly attentive to how intuitive and user-friendly a social media platform is. Further investigation would be necessary to confirm these findings.

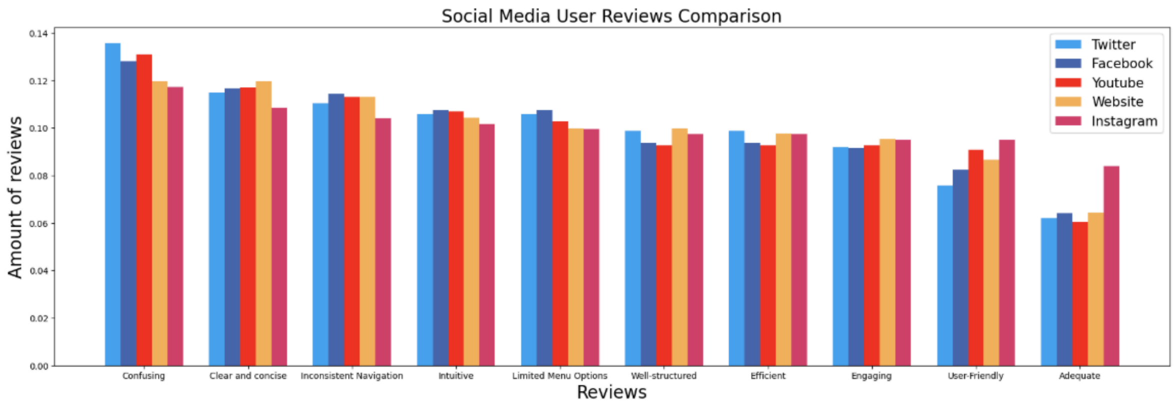

Figure 5: Comparison of the social media reviews.

According to Figure 5 Twitter has higher bars and this is why it may appear to stand out with the most positive reviews, but when looking carefully, it can be seen that Twitter was the platform that received the most negative comments as well (Figure 7).

Figure 6 and 7 give further insight into this. In Figure 7 only the negative reviews are shown and Twitter is perceived as more confusing than the other media, yet on the other negative aspects it is rather average. In Figure 6 it can be seen that there is no clear winner, yet in some categories certain media stand out. Twitter tends to have less positive reviews overall and is perceived as the least user-friendly and adequate, which suggests that it received the fewest positive evaluations.

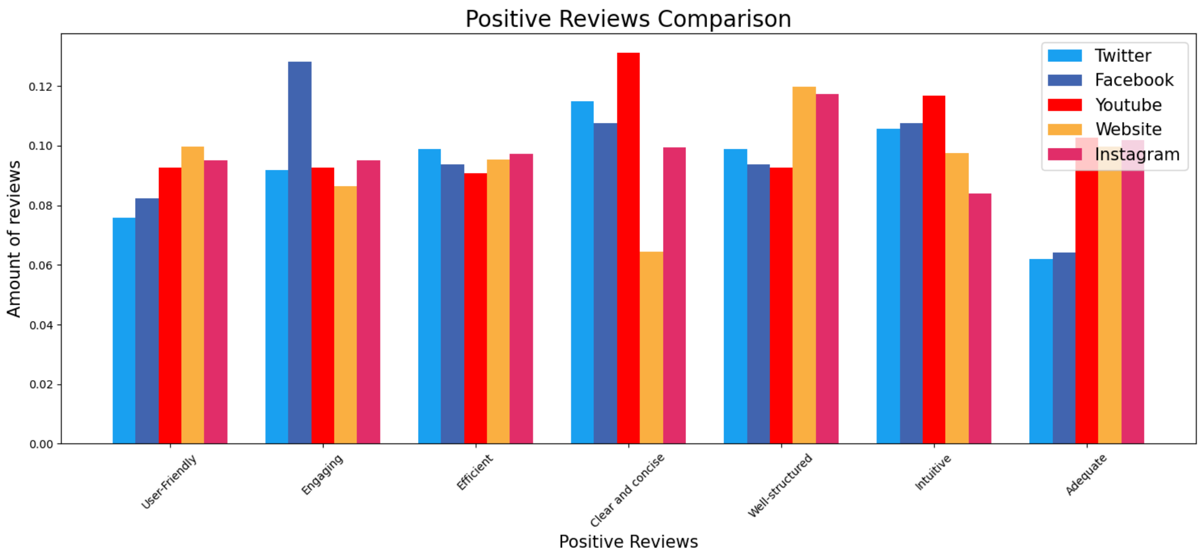

Figure 6: Solely positive reviews

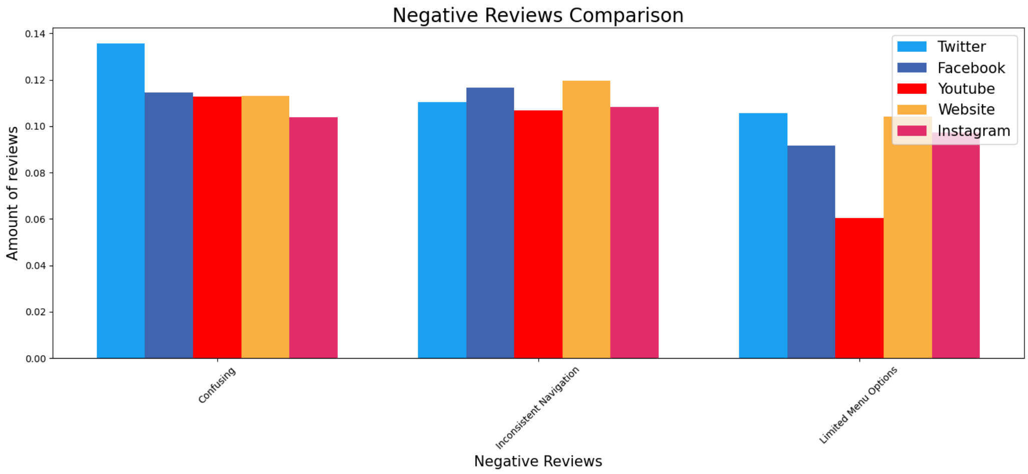

Figure 7: Solely negative reviews

4.1.3. Mean evaluation by website and category

Using the dictionary created in data preparation, it was possible to find out how many users evaluated each website: 435 users evaluated Twitter, 496 users evaluated Youtube, 437 evaluated Facebook, 451 users evaluated another Website and 452 users evaluated Instagram in this research.

This made it possible to discover which platforms performed better or worse in each of the quantitative columns evaluated:

Category	Worst Media	Mean
Color Scheme	Instagram	4.06
Visual Hierarchy: - Mean:	Twitter	4.06
Typography	Facebook	3.89
Images and Multimedia	Twitter	4.04
Layout	Facebook	4.20
Mobile Responsiveness	Website	4.20
CTA (Call to Action) Buttons	Youtube	4.15
Forms and Input Fields	Facebook	4.07
Feedback and Error Messages	Facebook	: 4.06
Loading Speed:	Website	4.07
Personalization	Twitter	4.07
Accessibility	Facebook	4.06
Animation and Transitions	Instagram	4.07

Scrolling_Behavior	Twitter	4.09
Gestures and Touch Controls	Instagram	4.06
Search Functionality	Website	4.07
Social_Media_Integration	Youtube	4.04

Table 1: Worst average scores per category

Mobile responsiveness and Layout had the highest means and typography was consistently the lowest category when comparing all the platforms but overall the averages showed little variation.

The line in the bar charts (Figure 8-12) shows the average value of the platform. Twitter (Figure 8) has 10 features below the average while YouTube (Figure 10) has only 6.

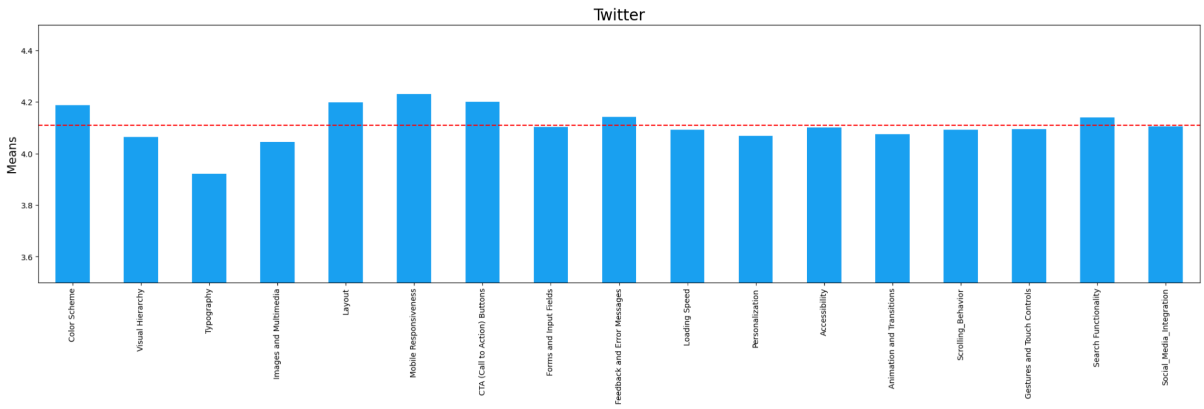

Figure 8: Twitter's means for every quantitative evaluation column

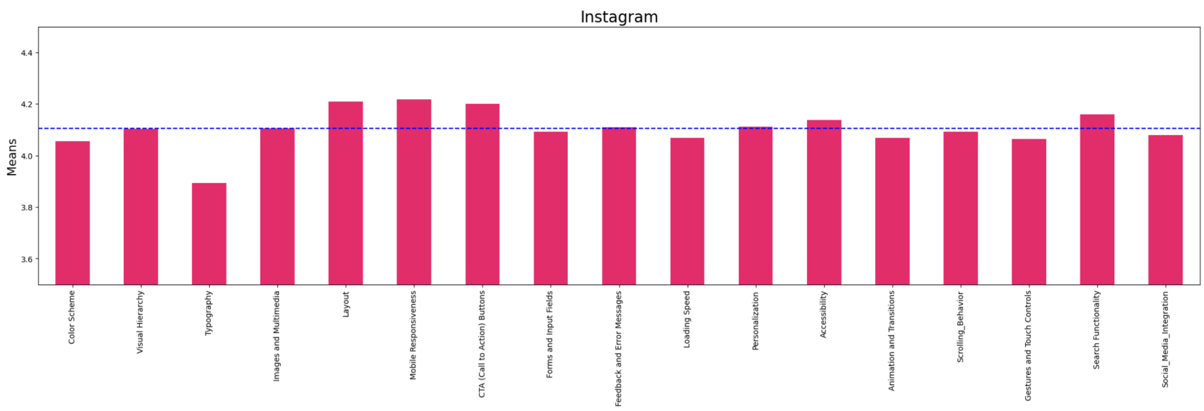

Figure 9: Instagram's means for every quantitative evaluation column

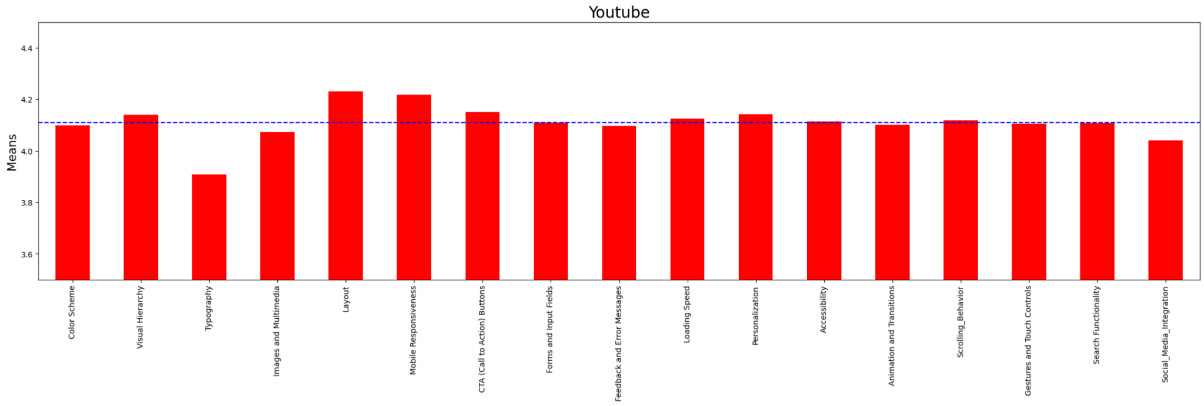

Figure 10: YouTube's means for every quantitative evaluation column

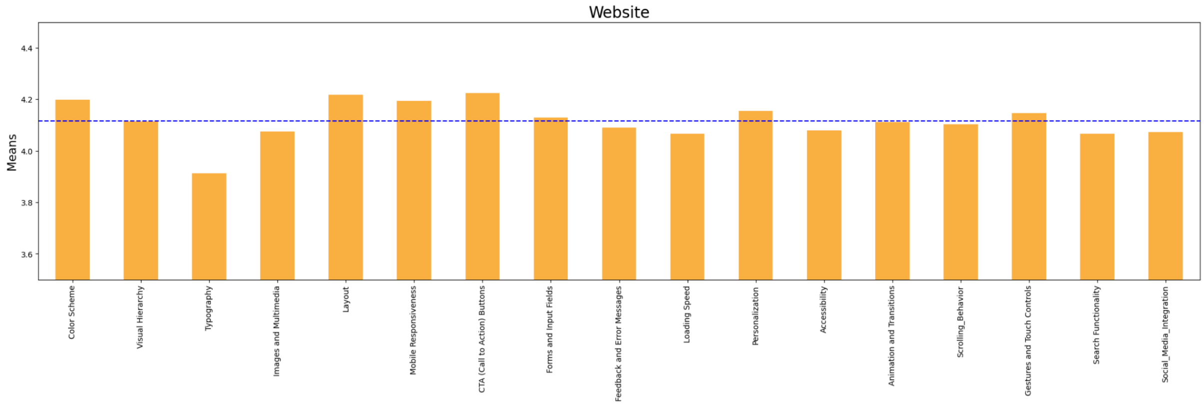

Figure 11: Website's means for every quantitative evaluation column

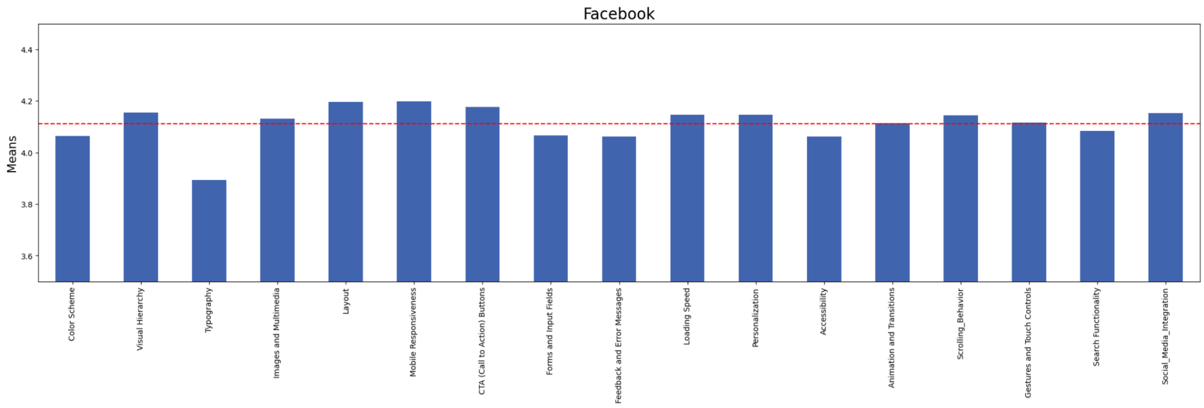

Figure 12: Facebook's means for every quantitative evaluation column

4.1.4. Average Rating by Age

Figure 13 shows that there is minimal variation in average ratings across different age groups, showing no significant differences in the data. While smaller intervals (e.g., 1.5, 1.1, 1.2, 1.3) might reveal slight variations, any patterns would likely be too marginal to provide meaningful insights. As a result, we decided not to pursue further analysis, as it would not yield significant findings regarding differences in average ratings among age groups.

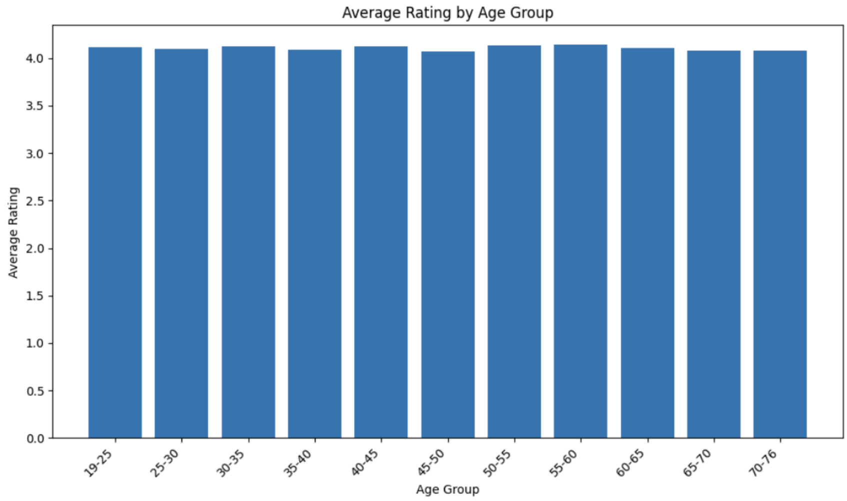

Figure 13: Average rating by age group

4.2. Curve Fitting

A normal distribution could be fitted almost perfectly to the average ratings histogram. The parameters of the estimated normal distribution are $\mu = 4.1$ and $\sigma = 0.16$. That means that the mean or expected value is 4.1 with a standard deviation of 0.16. The standard deviation is fairly low, which means our data tends to be clustered around the mean. The plotted normal distribution can be seen in Figure 14.

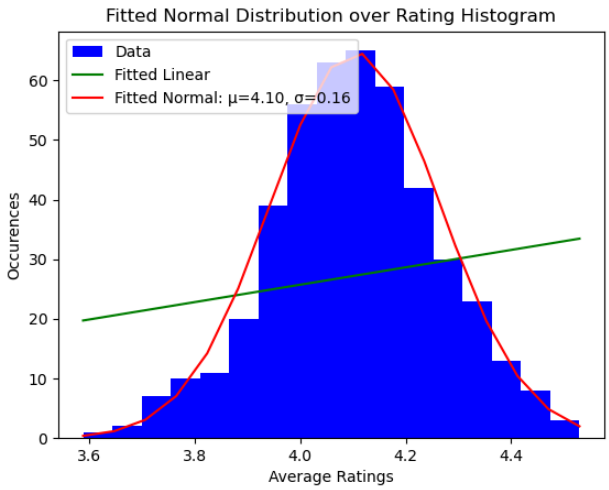

Figure 14: Fitted Normal distribution over rating histogram

Since the normal distribution fits our data very precisely, it means our data is normally distributed and hence it would be suited for performing statistical testing of hypotheses. For reference, a linear regression algorithm has been used and plotted through the data, but it is obvious that it could not be fitted to the histogram of the distribution of ratings well. You could say it is underfitting and not describing the complexity of the data, which makes sense because it is a normal distribution. If the difference was not obvious at first sight, it would be possible to perform a linear least square or another algorithm to see which of the lines fits the data better.

4.3. Classification: SVC & Multi-class Logistic Regression

Since the distribution of average ratings is slightly different for each social media and there is some correlation between the different values, we wanted to understand if this is enough difference to classify the social media platform that has been evaluated based on the numerical features.

However, the prediction of social media based on ratings given by the user failed. Neither the SVM with different parameters for C (Figure 15), nor the logistic regression model (Figure 16) could achieve an accuracy larger than random guessing.

We can see in Figure 15, that the SVM only is slightly better than random guessing with certain values of C. This suggests that our classification model was not able to learn anything from the dataset, however using values for $C > 25$ seems more appropriate for the dataset.

The logistic regression model when evaluated on the test data (Figure 16) is not significantly more accurate than random guessing. We can observe that for small training sizes, the model overfits and achieves an accuracy of almost 1 on the training data, while only getting an accuracy of 0.2 on the test data. As the training size goes up, the performance evaluated on the training data drops, which is to be expected, however, the performance evaluated on the test data never improves.

Since both algorithms were not able to learn anything from the training data that generalized to the testing data, this suggests that either the correlation between the ratings and the media that were evaluated was not strong enough, or that there is not enough training data to identify the correlation.

For comparison, it was attempted to predict the platform that was evaluated based only on the average rating and the participant's age, which provided marginally improved results, but still not enough to be statistically significant. This has not been included in the report but can be seen in the Jupyter Notebook.

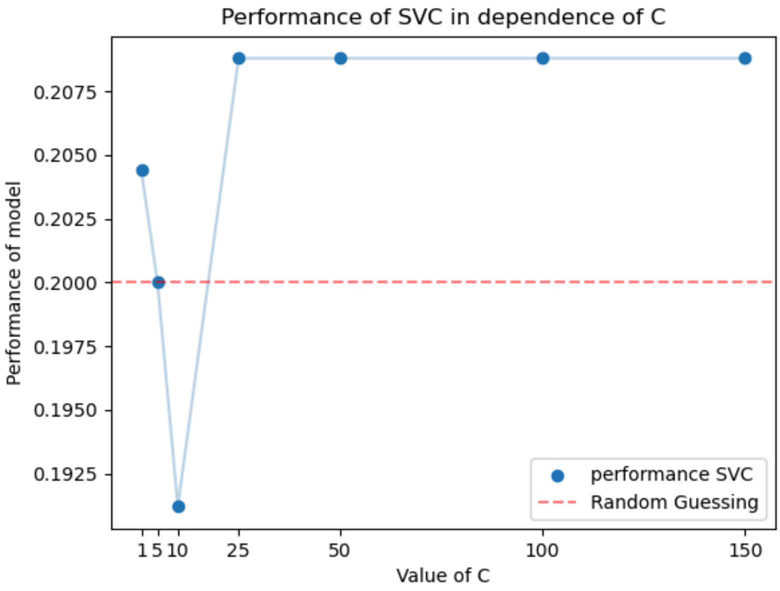

Figure 15: Support Vector Machine Classification

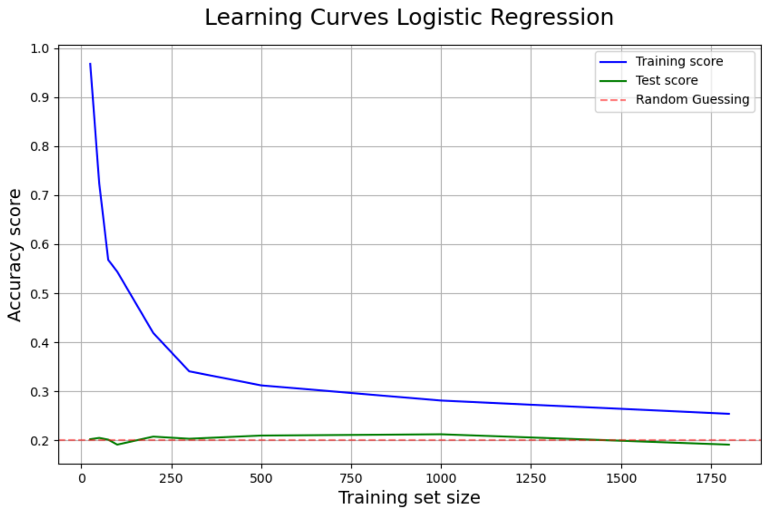

Figure 16: Logistic Regression Classification

Conclusion and future development

For further development of this project, acquiring more data would be crucial. This research indicated that there is not a significant difference between evaluations given by certain user groups or received by specific platforms. To understand and predict a trend whether some platforms are better in some ways than others, a dataset with more detailed and differentiable data would be necessary.

The analysis showed that some platforms are slightly worse than others, and if this trend could be statistically confirmed, it would be possible to tell designers which parts to focus on or improve. More specific analysis could be helpful in identifying exactly which processes of a website / platform need improvement, why and how. Identifying the target market, their needs and cross-referencing whether the platform met these needs or not, and if not how to change them so that they do.

The project's development was also important for learning more about the best way to visualize the data and avoid biasing the user with a large amount of data on the same page as Figure 5 shows. When we combined all the graphs, positive and negative comments were mixed, creating the false impression that Twitter had good comments.

Our research showed that the data was normally distributed. Throughout the project we did not use statistical tests, because they did not seem necessary and in the scope of the project. In the machine learning parts, different random seeds for data sampling and more through hyperparameter testing would have given a more accurate and less random (biased) result, but this wasn't done, because the predictions were too bad or the algorithms couldn't learn enough.

Mario Filho discusses the *survivorship bias* in his book with an example from World War 2: analysts examined returning airplanes and reinforced the areas most hit by bullets. However, later, they realized it was necessary to strengthen the less-hit areas, as airplanes that didn’t return may be more hit in critical, unobserved parts (Filho, 2021). This example raises an important question that needs to be considered in data science projects: are we using the correct data? In the dataset used in this research, we don't know exactly the criteria used to interview users. Therefore, it is essential to consider whether inactive or former users were included in the analysis. Excluding these users might create a distorted view of the overall experience and the analyses could be distorted.

The project was essential for the team to learn and interact with their first database, learn about the process and go through it from start to finish. However, the team does not see possible applications in research and development with this dataset. For the sake of practice, it would have been beneficial to find a dataset that has not been previously cleaned. Furthermore, the features of our dataset were not different enough from each other to make for meaningful results or a successful classification.

Bibliography

Altman, D. G., & Bland, J. M. (1995). Statistics notes: the normal distribution. Bmj, 310(6975), 298.

Filho, M. (2021). Manual prático de data science (Kindle Edition).

Hunter, J. D., Droettboom, M., Caswell, T., Firing, E., & The Matplotlib Development Team. (2003–2023). *Matplotlib documentation*. https://matplotlib.org/stable/

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825–2830. https://scikit-learn.org/stable/

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14, 199-222.

Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018). This is service design doing: Applying service design thinking in the real world. O´Reilly Media.

Rahman, M. A. (2023). *UI UX dataset* [Dataset]. Kaggle.

The NumPy Development Team. (2025). *NumPy documentation*. https://numpy.org/doc/

The pandas development team. (2025). *pandas documentation*. https://pandas.pydata.org/docs/