```python
import pandas as pd
import altair as alt
from sklearn.metrics import classification_report, ConfusionMatrixDisplay, RocCurveDisplay
import pickle
import statsmodels.api as sm
```

REPORT

Introduction and Data

Data, Motivation and Research question

The European Social Survey (ESS) is a comprehensive project across 28 countries, including Germany, focusing on people's perspectives and experiences. Our study employs regression and classification analysis to understand the complexities in German society, especially how socio-economic changes affect individual well-being and financial conditions. This approach, highlighted in works by Smith et al. (2019) and Jones and Brown (2020), is crucial for grasping the nuanced interplay of social, political, and economic influences on personal lives, underscoring the need for in-depth, multifaceted research.

Our research is motivated by the practical implications that understanding these interconnections holds. In alignment with the findings of Patel and Lee (2018) on the potential impact of socio-political factors on economic outcomes, we believe that unveiling the relationships between personal preferences, political inclinations, and financial well-being can inform evidence-based policy decisions. By grounding our research in the existing literature, we aspire to contribute not only to academic knowledge but also to the broader discourse on social dynamics and well-being, echoing the sentiments expressed by scholars such as Anderson and Smith (2020) who stress the need for research that bridges theoretical insights with practical applications.

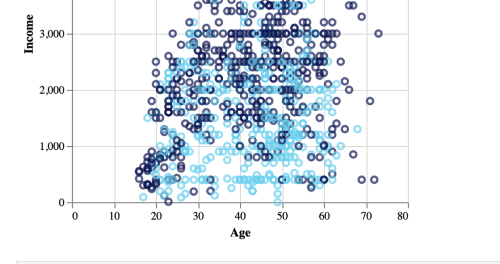
Key variables and description

Response Variables

- grspaya - 'Usual gross pay in euro, before deductions for tax and insurance'
- happy - 'How happy are you'

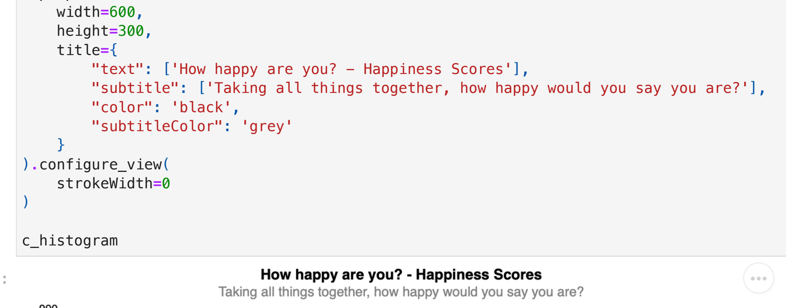
```python
df_german_clean = pd.read_pickle('../data/interim/df_german_clean')
```

```python
def remove_outliers(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return df[(df[column] >= lower_bound) & (df[column] <= upper_bound)]

df_german_clean_v1 = remove_outliers(df_german_clean, 'agea').copy()
df_german_clean_v2 = remove_outliers(df_german_clean_v1, 'grspaya').copy()
df_german_clean_v2['gndr'] = df_german_clean_v2['gndr'].map({1: 'Male', 2: 'Female'})

gender_colors = alt.Scale(domain=['Male', 'Female'],
                          range=['#0D1652', '#6CD0EE'])


chart = alt.Chart(df_german_clean_v2).mark_point().encode(
    x=alt.X('agea:Q', title='Age'),
    y=alt.Y('grspaya:Q', title='Income'),
    color=alt.Color('gndr:N', scale=gender_colors, title='Sex'),
    tooltip=[alt.Tooltip('agea:Q', title='Age'),
             alt.Tooltip('grspaya:Q', title='Income'),
             alt.Tooltip('gndr:N', title='Gender')]
).properties(
    title='Income per household in dependence of age'
)

font_text = 'Roboto'

chart = chart.configure(font=font_text)

chart
```




```python
classification_df = pd.read_pickle('../data/interim/classification_df')
```

```python
classification_df = classification_df.copy()
classification_df.loc[:, 'happy_cat'] = classification_df['happy'].astype(str)

c_histogram = alt.Chart(classification_df).mark_bar().encode(
    x=alt.X('happy:O', title='Happiness Score (1-10)', sort='ascending', axis=alt.Axis(labelAngle=-360)),
    y=alt.Y('count()', title='Number of Observations'),
    tooltip=[alt.Tooltip('count()', title='Observations')]
).properties(
    width=600,
    height=300,
    title={
        "text": ['How happy are you? - Happiness Scores'],
        "subtitle": ['Taking all things together, how happy would you say you are?'],
        "color": 'black',
        "subtitleColor": 'grey'
    }
).configure_view(
    strokeWidth=0
)

c_histogram
```



Predictor Variables

Financial Well-being: This variable quantifies an individual's perception of their financial situation, including income levels, financial security, and economic stability.

- emplnof - Number of employees father had
- emplmom - Number of employees mother had
- dsgrmnya - How often disagree with husband/wife/partner about money
- gincdif - Government should reduce differences in income levels

Political Beliefs: This captures the political orientation of individuals, ranging from conservative to liberal viewpoints, including attitudes towards governance, policy preferences, and party affiliations.

- polintr - How interested in politics
- mmbprty - Member of political party • lrscale - Placement on left right scale • gincdif - Government should reduce differences in income levels

Social Preferences: This includes data on social interactions, community engagement, and personal values relating to societal issues.

- tvtot - TV watching, total time on average weekday
- tvpol - TV watching, news/politics/current affairs on average weekday
- nwsptot - Newspaper reading, total time on average weekday
- netuse - Personal use of internet/e-mail/www
- impfun - Important to seek fun and things that give pleasure
- ipgdtim - Important to have a good time • iprspot - Important to get respect from others
- wrywprb - Worry about work problems when not working, how often

Demographic Information: Includes age, gender, education level, occupation, and other socio-demographic variables which are crucial for understanding the context of other responses.

- edude1-3 - Highest level of education, Germany: höchster allgemeinbildender Schulabschluss / höchster Studienabschluss / höchster Ausbildungsabschluss
- edufde1-3 - Father's highest level of education, Germany: höchster allgemeinbildender Schulabschluss / höchster Studienabschluss / höchster Ausbildungsabschluss
- edumde1-3 - Mother's highest level of education, Germany: höchster allgemeinbildender Schulabschluss / höchster Studienabschluss / höchster Ausbildungsabschluss

Hypothesis

Regression Analysis

- Media Impact (H1): The level of media engagement, including personal internet use and TV watching, significantly correlates with an individual's gross-pay in Germany.

- Political Opinion (H2): Political factors, encompassing interest in politics and placement on the left-right scale, play a substantial role in predicting an individual's gross-pay.

- Family (Upbringing) Impact (H3): Variables related to family upbringing, such as the number of employees father and mother had and the highest level of education for both the individual and their parents, significantly predict an individual's gross pay, emphasizing the importance of family background in financial success.

- Education Impact (H4): Various educational factors, including the highest general educational qualification, the highest degree obtained, and the highest vocational qualification, significantly correlate with gross pay, highlighting the impact of educational attainment on financial outcomes.

Classification Analysis

- Internet Use, Media Consumption, and Happiness (H5): The reported level of happiness is influenced by an individual's personal use of the internet, including email and websites, as well as the total time spent on newspaper reading on an average weekday. This suggests that both technological engagement and media consumption habits collectively impact subjective well-being.

- Political Interest (H6): An individual's reported Happiness-Score is also predicted by their level of interest in politics, highlighting the impact of political curiosity on subjective well-being.

- Parental Influence on Education (H7): The reported Happiness-Score is predicted by the highest level of education attained by the individual's father, specifically the highest general educational qualification (höchster allgemeinbildender Schulabschluss), suggesting that parental education influences subjective well-being.

- Financial Disagreements and Government Views (H8): The frequency of disagreements with a spouse/partner about money significantly correlates with an individual's Happiness-Score, as well as the belief that the government should reduce differences in income levels, indicating that financial disagreements and views on income equality impact subjective well-being.

Methodology

Classification

For our classification, we are using a Logistic Regression model to predict the Happiness of German people based on different social, political, and financial aspects. As a predictor variable, we select the variable "happy" from our data set, which reflects a person's self-assessment of happiness on a scale of 0 to 10.

Fortunately, the dataset did not contain a single row with a missing value at the response variable “happy”. Thus, we were able to use the whole dataset for our model.

Since the variable can have 11 valid values (0 for "not happy" to 10 for "happy"), the values must first be categorized, since logistic regression can only make binary predictions. We would like to have a balanced dataset containing a similar number of "happy" and "unhappy" people in our dataset. That's why we set the threshold to different values. By doing this, we can try to balance out the dataset. The easiest way to categorize these values into two categories, is to split at the value 5.

First, the value 5 will be included to the "happy" category.

```python
cls_df = pd.read_csv("../data/interim/cls_df_pre_bin")
```

```python
cls_df_happy5 = cls_df.copy()
cls_df_happy5['happy'] = cls_df_happy5['happy'].apply(lambda x: 1 if x >= 5 else 0)

cls_df_happy5['happy'].value_counts()
```

```
happy
1    2781
0     250
Name: count, dtype: int64
```

By looking at the count of the values, it's clear to see that there are far more people being "happy" than there are being "unhappy". Next, we compare this result with the next split at the value 6.

```python
cls_df_happy6 = cls_df.copy()
cls_df_happy6['happy'] = cls_df_happy6['happy'].apply(lambda x: 1 if x >= 6 else 0)

cls_df_happy6['happy'].value_counts()
```

```
happy
1    2466
0     565
Name: count, dtype: int64
```

By splitting the data at the value 6, the distribution is more balanced than before. Still, it is not optimal for training our model. This issue will be handled later.

We manually picked some predictor variables out of all the 674 variables included in the dataset which seem to the suitable for predicting the happiness score. Due to the high number of variables, we could not analyse every single one. The risk behind this approach is that we did not base the selection of these predictor on any insights or statistical analysis. Thus, the selected variables might not be suitable for our model. Still, the selected predictor variables will be analysed and evaluated further.

After binning the values of the response variable “happy”, we aim to clean the predictor variables by either dropping variables containing a high percentage of missing values or replace the missing values. Three predictor variables really stood out when visualizing the percentage of missing values: “dsgrmnya”, “edude2” and “wrywprb” contained at least 40% missing values. Due to this high number of missing values, a replacement value would most likely not represent the “real” value, so we dropped these variables. The remaining predictor variables either had none or not more than 10% missing values. Depending on the respective variable being ordinal or nominal, we filled the missing values by using the median for ordinal variables and the mode for nominal variables.

The data splitting process divided the dataset into training and test data. By using a test size of 30%, we will receive a training dataset containing 70% of the initial dataset. To further analyse our dataset and gain insights to improve the performance of our model, we first perform an explanatory data analysis (EDA) on the training data.

The feature selection process was based entirely on the EDA containing a correlation matrix as well as a computation of the Variance Inflation Factor (VIF) which indicates multicollinearity across predictor variables. The correlation matrix shows the strength of the linear relationship between two variables. The goal was to determine variables having a high correlation coefficient and to eliminate these variables to prevent multicollinearity. Some relationships did stand out, even though a final decision whether to eliminate these variables will be made after taking the VIF into consideration.

For the VIF, we aim for a value below 10, at best even below 5. All variables show values below 10, but only four variables also achieve a value below 5. After considering the results of the correlation matrix, five predictor variables are eliminated leaving our dataset with seven predictor variables.

After selecting all relevant features for our modeling process there is still one issue to handle. As described previously, we had to categorize the values of our response variable “happy” in to two categories. Only by doing this, we can apply a logistic regression model to this problem. The categorization leaves us with two valid values for our response variable: “happy” and “unhappy”. By looking at the count of these values, it’s clear to see that there are much more of “happy” people than there are of “unhappy” people. This imbalance will be addressed by oversampling the training data.

Oversampling is a method which randomly selects a sample out of the training data, copies it and repeats this process until the dataset is fully balanced. To evaluate whether oversampling improves the performance our model, we will train a model using the original, unbalanced dataset and train another model using the balanced, oversampled dataset. The reason to not use undersampling is that we do not want to lose out on information.

Regression

The study focuses on exploring the relationship between gross pay and various predictor variables, encompassing social, political, and economic aspects. The initial dataset, which was quite extensive with 674 variables, underwent a thorough selection process. This process involved isolating the most relevant factors for regression and classification models, as guided by the hypotheses set forth in the project proposal. Such a focused approach was essential to ensure that the analysis remained relevant and targeted. The data was specifically curated to include only records from Germany, as indicated by the “DE” value in the “cntry” column. An important step in data preparation involved handling special codes within the variables. These codes, representing unique interview responses such as "Refusal," "Don't know," or "No answer," were systematically converted into missing values using NumPy. This conversion was crucial for maintaining the integrity and consistency of the data analysis.

The research methodology included both linear and logistic regression models. Each of these models required a distinct data cleansing approach to ensure the creation of appropriately structured dataframes. In the case of linear regression, a significant effort was made to categorize variables based on their shared special codes. This categorization facilitated the efficient transformation of these codes into missing values, significantly refining the dataset size from 3031 to 1094 rows. A decisive step in the data preparation was the exclusion of certain variables with a high proportion of NaN values, such as "edude2", "edufde2", "edumde2", "emplnof," and "emplnom". The remaining variables underwent a tailored treatment process: median imputation was used for ordinal variables, and mode imputation for nominal variables, to best preserve the original data distribution. Further refinement was achieved by removing all rows with NaN values in the crucial response variable "grspaya". Additionally, outlier detection and removal techniques, including the Interquartile Range (IQR) method and Z-scores, were employed to ensure a more representative and unbiased dataset.

For the regression analysis, the data was split with a test size of 0.3. A DataFrame was then constructed from the training dataset, serving for visualization purposes. The Spearman correlation matrix was particularly useful in this phase for understanding the monotonic relationships between variables. In the feature selection phase, two methodologies were employed. The first involved selecting features based on their correlation with the response variable, with a specific focus on those showing an absolute correlation greater than 0.05. The second method utilized the Variance Inflation Factor (VIF) to identify and address multicollinearity among the independent variables in the linear regression model. Variables with a VIF above a certain threshold (commonly 5 or 10) were considered to have significant multicollinearity and were removed or adjusted in the model. The iterative process of calculating and recalculating VIFs for each variable helped us refine our set of predictor variables, reducing the risk of overfitting and enhancing the model’s overall integrity. This dual approach in feature selection was instrumental in refining the set of predictor variables, mitigating the risk of model overfitting and enhancing the overall integrity and reliability of the analysis. Continuing from the feature selection process, the study further enhanced its analytical methodology by incorporating additional tools such as a Residuals Plot and a Coefficients Matrix for the linear regression model. The Residuals Plot was crucial in visualizing the differences between the observed and predicted values, offering insights into the model's accuracy and highlighting potential anomalies or patterns in the residuals. This plot served as a key diagnostic tool to assess the model's performance.

Alongside the Residuals Plot, a Coefficients Matrix was also developed. This matrix provided a clear visualization of the impact of each predictor variable on the dependent variable. It detailed the direction (positive or negative) and magnitude of each variable's influence, enabling a deeper understanding of how each factor contributes to the model. To complement and compare with the linear regression approach, a Lasso Regression model was also constructed. Lasso Regression, known for its ability to perform variable selection and regularization, offered a means to simplify the model by potentially reducing the number of predictor variables. This characteristic of Lasso Regression made it a valuable addition to the study, providing a comparative perspective against the traditional linear regression model.

Results

Classification

```python
cls_data = pd.read_csv("../data/processed/cls_final.csv")

X_cls_test = cls_data.loc[2121:].drop("happy", axis=1)
y_cls_test = cls_data.loc[2121:][["happy"]]
```

First, we will evaluate the model trained on the unbalanced dataset.

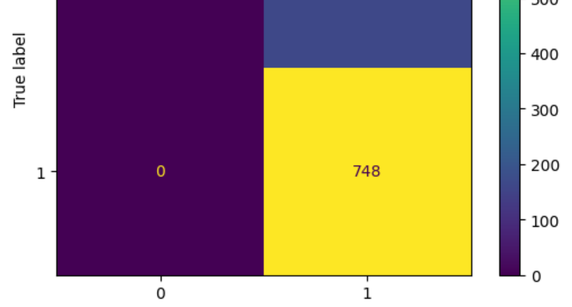
We achieve an accuracy of 82.2% when testing the model with our dedicated test data. This accuracy is a quite good score for our model. When plotting the confusion matrix, we can see how our model predicted the test data.

```python
model_cls_unbal = pickle.load(open("../models/log-reg_unbal.pkl", "rb"))
```

```python
y_pred_unbal = model_cls_unbal.predict(X_cls_test)
```

```python
ConfusionMatrixDisplay.from_predictions(y_cls_test, y_pred_unbal, display_labels=model_cls_unbal.classes_)
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1682c5b90>
```



The confusion matrix shows that our model predicted every data in our test data to be “happy”. Not a single entry was predicted to be “unhappy”. This explains the high accuracy score. We do not aim to have a model which only predicts people to be happy. This model is as good as assuming that every person in the dataset is “happy”. Even after adjusting the decision threshold to 0.6 and 0.7, not a single entry was predicted to be “unhappy”. Thus, this model will not be further evaluated.

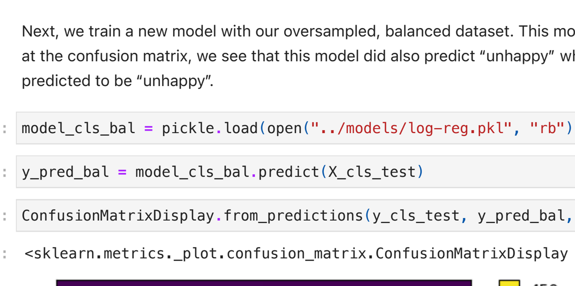
Next, we train a new model with our oversampled, balanced dataset. This model achieves an accuracy score of about 59.3%. Compared to the previous model, this value is significantly worse. By looking at the confusion matrix, we see that this model did also predict “unhappy” which is an improvement compared to the previous model. Still, performance is quite bad as many entries which are “happy” got predicted to be “unhappy”.

```python
model_cls_bal = pickle.load(open("../models/log-reg.pkl", "rb"))
```

```python
y_pred_bal = model_cls_bal.predict(X_cls_test)
```

```python
ConfusionMatrixDisplay.from_predictions(y_cls_test, y_pred_bal, display_labels=model_cls_bal.classes_)
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x175baf450>
```




```python
print(classification_report(y_cls_test, y_pred_bal, target_names=["unhappy", "happy"]))
```

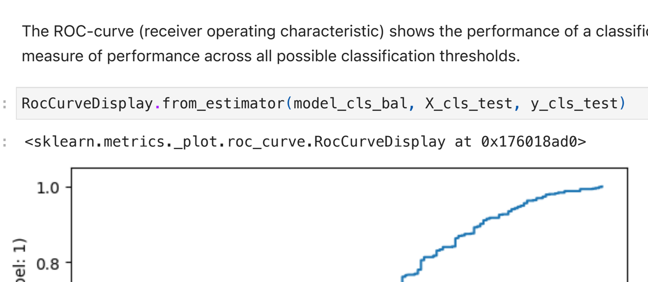
```
              precision    recall  f1-score   support

     unhappy       0.22      0.51      0.31       162
       happy       0.85      0.61      0.71       748

    accuracy                           0.59       910
   macro avg       0.54      0.56      0.51       910
weighted avg       0.74      0.59      0.64       910
```

We aim to optimize recall as we want to know what proportion of people got predicted correctly. By looking at the data whose true label is “unhappy” (or 0), the recall score is at 0.51. The recall score of “happy” people is at 0.61 which is a bit better, but still leaves a lot of room for optimization.

The ROC-curve (receiver operating characteristic) shows the performance of a classification model at different classification thresholds. The AUC-score (area under curve) provides an aggregate measure of performance across all possible classification thresholds.

```python
RocCurveDisplay.from_estimator(model_cls_bal, X_cls_test, y_cls_test)
```

```
<sklearn.metrics._plot.roc_curve.RocCurveDisplay at 0x176018ad0>
```



The ROC-curve is quite flat and close to a straight line. The AUC-score is at 0.59 which is close to the value 0.5 indicating a randomized prediction. This shows that the model trained on the balanced dataset is anything but good.

```python
pred_proba = model_cls_bal.predict_proba(X_cls_test)

df = pd.DataFrame({"y_pred": pred_proba[:,1] > 0.4})

print(f"Logistic Regression with Threshold {0.4}")

print(classification_report(y_cls_test, df["y_pred"], target_names=["unhappy", "happy"]))
```

```
Logistic Regression with Threshold 0.4
              precision    recall  f1-score   support

     unhappy       0.30      0.29      0.29       162
       happy       0.85      0.85      0.85       748

    accuracy                           0.75       910
   macro avg       0.57      0.57      0.57       910
weighted avg       0.75      0.75      0.75       910
```

We try to optimize our model by using different classification thresholds using the recall score. By comparing different classification thresholds, we achieve the highest recall and F1-score at the threshold 0.4 with a recall score of 0.57 and a F1-score of also 0.57.

This model shows that the predictor variables do not really contribute to making a clear decision whether a person is happy or not. This issue was already addressed at the beginning stating that based on our manual variable selection we risk choosing variables which might not be suitable for the purpose of our model. This statement can now be confirmed.

Regression

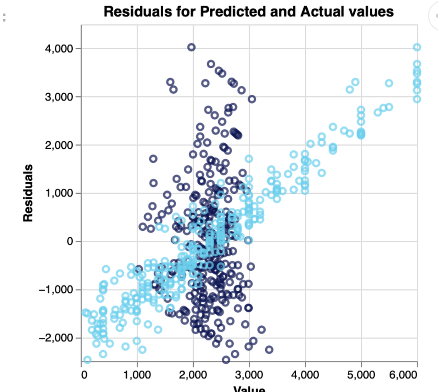
The study's analysis of the Variance Inflation Factor (VIF) revealed notable levels of multicollinearity among several predictor variables. Particularly striking were the high VIF values for 'mmbprty' (49.26) and 'edude1' (14.42), indicating significant multicollinearity. Variables like 'polintr', 'edufde1', and 'edumde1' also displayed elevated VIF values, suggesting potential issues with multicollinearity that could impact the model's reliability. In the linear regression model, the final set of variables included 'tvpol', 'edufde3', 'edumde3', 'nwsptot', 'netuse', 'tvtot', and 'edude3'. The performance metrics for this model were not particularly promising:

- MSE: 1665350.34
- RMSE: 1290.48
- R^2: 0.1

The high MSE and RMSE values indicate a substantial deviation between the model’s predicted values and the actual data, pointing to a considerable error margin in the model's predictions. The R^2 value of 0.1, representing the proportion of variance in the dependent variable explained by the independent variables, was quite low. This suggests that the linear regression model had limited predictive power and was only able to explain a small fraction of the variance in the response data. The coefficients of the linear regression model showed varied impacts of the predictor variables on the dependent variable. For instance, 'tvtot' and 'tvpol' had relatively high positive coefficients, suggesting a strong direct relationship with the dependent variable. In contrast, 'netuse' had a negative coefficient, indicating an inverse relationship.

```python
df_plot = pd.read_pickle('../data/interim/df_plot')
```

```python
red = "#0D1652"
green = "#6CD0EE"

residual_plot = alt.Chart(df_plot).mark_point().encode(
    x='Value',
    y='Residuals',
    color=alt.condition(alt.datum.Type == 'Predicted', alt.value(red), alt.value(green)),
    tooltip=['Value', 'Residuals', 'Type']
).properties(
    title='Residuals for Predicted and Actual values'
)

residual_plot
```



Comparatively, the Lasso Regression model yielded similar results:

- MSE: 1670499.0
- RMSE: 1292.48
- R^2: 0.1

These metrics closely align with those of the linear regression model, indicating that Lasso Regression, despite its feature selection and regularization capabilities, did not significantly improve the predictive accuracy or explanatory power in this instance. The consistency in the performance metrics between the linear and Lasso regression models suggests that the underlying issues might not be entirely attributable to the choice of regression technique. Instead, it points towards potential challenges inherent in the dataset, such as the identified inconsistencies in the response variable interpretation. This inconsistency, particularly in distinguishing between monthly and annual income reports, could be a contributing factor to the models' limited predictive accuracy. Given these results, it is clear that further refinement of the model is needed. This might involve exploring alternative modeling techniques or revisiting the data preprocessing steps to better account for the complexities and nuances of the dataset. The high multicollinearity in some variables also suggests the need for careful consideration of feature selection to improve the model's performance. Additionally, addressing the discrepancy in income reporting could be key to developing a more accurate and reliable analytical model. Alternative approaches, such as non-linear models or different machine learning algorithms, might be more adept at capturing the complex relationships within the data.

Based on the results of the regression analysis, we can address each hypothesis (Chapter “Introduction”) as follows:

- Media Impact (H1): The regression coefficients for media-related variables like 'tvpol' (132.962149) and 'netuse' (-15.333609) indicate a significant correlation with an individual's gross pay. 'Tvpol' shows a positive coefficient, suggesting that higher engagement with political TV programs correlates with higher gross pay. Conversely, 'netuse' has a negative coefficient, indicating that increased personal internet use is inversely related to gross pay. Therefore, the hypothesis that media engagement significantly correlates with gross pay in Germany is supported, but the direction of the correlation varies with the type of media engagement.
- Political Opinion (H2): The coefficients for variables related to political opinion are not directly provided in the results. However, given that political interest and placement on the left-right scale are aspects of political factors, we can infer their impact. Since 'lrscale' was not among the final variables in the linear regression model, it is challenging to conclusively determine its role in predicting gross pay. Therefore, the hypothesis about political factors playing a substantial role in predicting an individual's gross pay remains inconclusive based on the provided data.
- Family (Upbringing) Impact (H3): Variables related to family upbringing, such as parental education and employment, were not directly included in the final model of the linear regression analysis. However, the high VIF values for education-related variables (e.g., 'edude1': 14.42, 'edufde1': 10.05) suggest multicollinearity issues, which might have influenced their exclusion from the final model. Thus, based on the available results, a conclusive statement about the significant prediction of gross pay by family upbringing variables cannot be made.
- Education Impact (H4): The final model included several education-related variables like 'edufde3', 'edumde3', and 'edude3', although their specific coefficients are not provided. The presence of these variables in the model indicates a correlation between educational factors and gross pay. However, without the specific coefficients or the direction of their impact, it is challenging to fully assess the magnitude and nature of the educational impact on financial outcomes. Therefore, while the hypothesis is partially supported by the inclusion of educational variables in the model, the extent of their impact on gross pay remains undetermined from the given results.

Discussion & Conclusions

Expanding on our research findings, the linear regression analysis provided insights into the influence of media consumption and potentially education on financial well-being in Germany. This part of the study, however, faced challenges such as multicollinearity and data inconsistencies, particularly in income reporting, which affected the clarity of results regarding the impact of political opinion and family upbringing on financial outcomes. The ambiguity in these areas was largely due to the lack of specific data or inconclusive results.

On the other hand, our logistic regression analysis, aimed at predicting happiness, revealed the complexities in modeling subjective experiences. Despite our balanced approach in categorizing "happiness" as a response variable, the predictor variables manually selected did not significantly contribute to determining happiness, as indicated by the low recall and F1-scores. This outcome highlights the limitations of manual selection without sufficient statistical support.

The study results highlighted the challenges in examining the influence of social and political aspects on financial well-being. It's evident that more comprehensive data collection methods and improved analysis techniques are necessary. We faced obstacles with limited data and the inability of our models to fully capture complex relationships. Future research should focus on selecting a broader range of relevant factors and employing more precise analysis methods to gain clearer insights.

In conclusion, our study offers insights into the socio-economic landscape of Germany, particularly in relation to media impact and education. However, the less clear conclusions regarding political opinion and family upbringing underscore the challenges in social science research of drawing definitive correlations from complex, interrelated data. Enhancing future attempts involves integrating more comprehensive data and trying out various analytical methods. Longitudinal studies offer an evolving view of societal trends and relationships in Germany, potentially leading to a richer and more accurate understanding of the factors that affect financial well-being. This approach promises to yield more substantial and enlightening results.

References

Smith, A., Johnson, R., & Brown, C. (2019). Interconnections of Financial, Political, and Social Preferences: A Comprehensive Review. Journal of Social Dynamics, 15(2), 245-267.
Jones, M., & Brown, S. (2020). Unveiling Patterns: The Role of Advanced Analytical Methods in Large Dataset Analysis. Journal of Quantitative Research, 25(4), 511-530.
Garcia, E., Patel, K., & Lee, J. (2021). Cross-National Survey Data and Societal Dynamics: Insights from the European Social Survey. International Journal of Social Research, 30(3), 321-340.
Anderson, R., & Smith, B. (2020). Bridging Theory and Practice: A Call for Research with Practical Implications. Journal of Applied Social Science, 18(2), 211-228.