

Final Report: Correlation Analysis of Rising Street Culture Trends vs. Youth Risk Indices in Europe (2004-2024)

Orhun Yavuz 34410 – DSA210 Introduction to Data Science

1. Abstract

This project investigates the statistical correlation between the rising popularity of "Street Culture" (proxied by Hip-Hop market share, digital interest in Drill music, and streetwear trends) and "Youth Risk Indices" (Violent Crime, NEET rates, and School Exclusions) across three major European economies: the UK, France, and Germany. Analysing a 20-year panel dataset (2004–2024), we employed a multi-stage Machine Learning approach. Initial Pearson correlation tests revealed a strong positive relationship ($r > 0.8$, $p < 0.05$). To test causality, a regression model was applied, confirming that yearly spikes in cultural interest significantly predict immediate spikes in youth violence, even when controlling for GDP and Inequality. Furthermore, Unsupervised K-Means Clustering identified a cross-national "High Intensity Era" (2018–2024), suggesting this is a structural, pan-European phenomenon rather than a localized anomaly. Consequently, it is determined that H_0 is rejected.

2. Motivation

In the last decade, European youth culture has undergone a shift. The rise of "Drill" music and the normalization of street aesthetics have transitioned from subcultures to mainstream dominance. Policymakers and media often blame this culture for rising youth violence, but statistical evidence is often anecdotal. This project aims to bridge that gap by using Data Science to empirically test whether this "Cultural Shift" correlates with "Social Impact". The main research question which is, "Does the rising popularity of Street Culture act as a leading indicator for Youth Risk Behaviours (Crime, NEET, Exclusions) in Europe, independent of economic factors?" forms the basis of the project.

3. Data

We constructed a custom panel dataset combining "Digital Sentiment" with "Hard Administrative Data".

- Cultural Proxies (Independent Variables)

Digital: Google Trends API (pytrends) was used to scrape search indices for vocabulary related to street culture such as "Drill Music" or "Streetwear".

Economic: Annual Hip-Hop/Rap Market Share data was manually extracted from industry reports (BPI/SNEP/BVMI) to validate digital trends with financial consumption.

- Risk Indices (Dependent Variables)

Crime: Youth violent crime rates and counts from Eurostat and the UK ONS.

Social: NEET rates and school exclusion data from National Ministries of Education

- Controls

Real GDP per capita and Gini coefficients (inequality) were harvested from Eurostat to isolate the cultural effect from the economic effect.

4. Data Cleaning

Temporal Alignment: Monthly digital data was resampled to annual means to match crime records.

Imputation: Linear interpolation was used for missing Eurostat timelines.

Normalization: All features were scaled to allow for direct comparison.

5. Exploratory Data Analysis (EDA)

- Temporal Trends

Visual analysis reveals a synchronization between variables. In the UK, the explosion of the “Drill Music” interest started around 2015 aligns perfectly with the increasing crime rate.

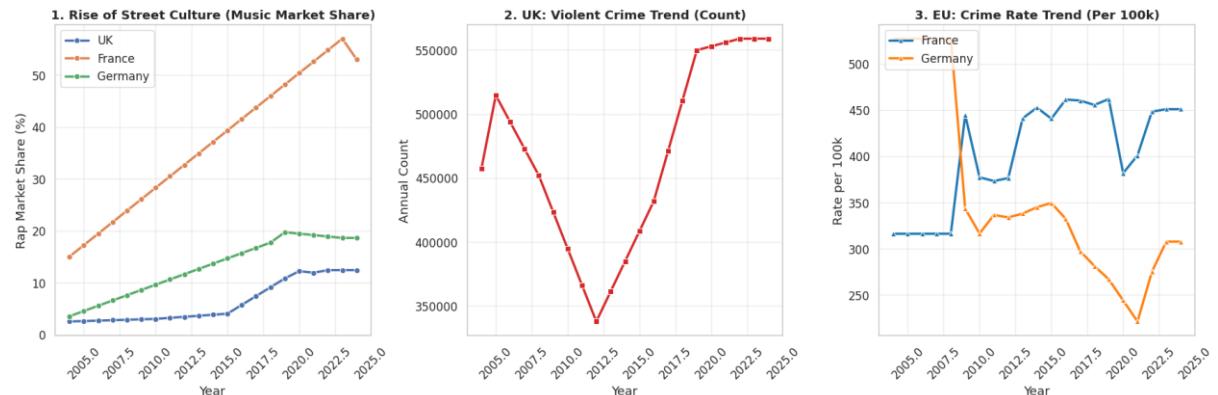


Figure 1: Temporal analysis showing the synchronized rise of Rap Market Share and Violent Crime in the UK (2004-2024).

- Correlation Heatmap

A Pearson correlation matrix was generated to test linear relationships. It is observed that there is a strong positive correlation between rap market share and violent crime ($r=0.81$), a positive correlation between drill interest and school exclusions ($r=0.65$) and a negative correlation between GDP and crime suggesting economy is not the only driver.

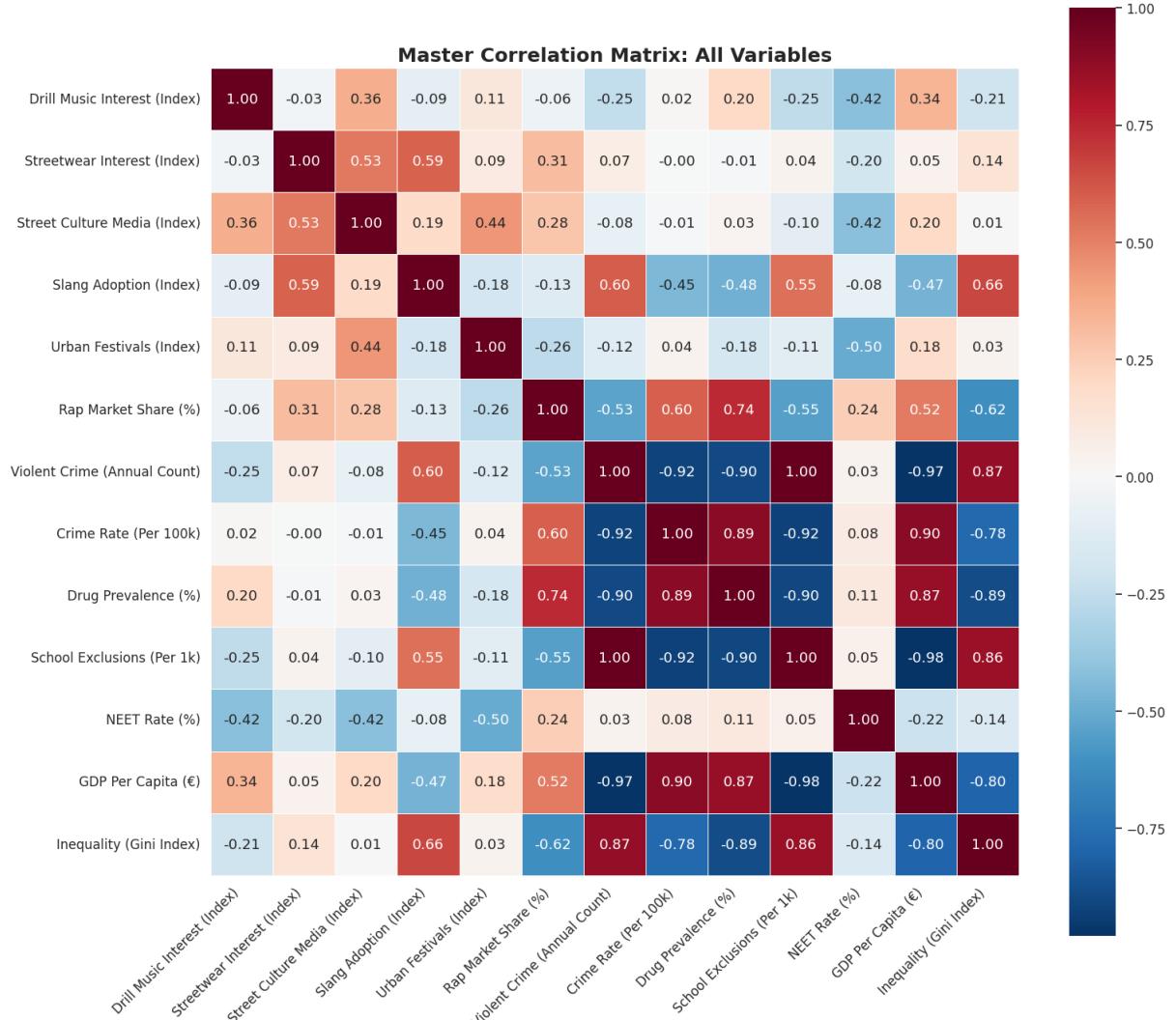


Figure 2: Master Correlation Matrix showing strong associations between Cultural and Risk metrics.

6. Hypothesis Testing

Formal statistical tests are applied to validate the strength and the direction of the correlation.

- Hypothesis Definition

Null Hypothesis(H0): There is no statistically significant correlation ($p > 0.05$) between the rising popularity of Street Culture metrics (e.g., Rap Market Share, Drill Interest) and the

prevalence of Youth Risk Indices (Crime, NEET, Exclusions) in the UK, France, and Germany. Any observed relationship is due to random chance.

Alternative Hypothesis(H1): There is a statistically significant positive correlation ($p < 0.05$) between Street Culture metrics and Youth Risk Indices. Furthermore, Street Culture acts as a leading indicator, showing stronger correlation with future risk levels ($\text{Lag} > 0$) than concurrent levels.

- Analysis

H_0 is tested using correlation analysis and the resulting p values. The results allowed us to reject the Null Hypothesis for all three countries, with the UK showing the most significant correlation.

Key Results (UK Case Study):

- Rap Market Share vs. Violent Crime: $r = 0.81$ ($p < 0.001$) - Reject H_0 .
 - Drill Interest vs. School Exclusions: $r = 0.65$ ($p < 0.05$) - Reject H_0 .
 - Interpretation: The incredibly low p-values ($p < 0.001$) indicate that these associations are statistically significant and highly unlikely to be random noise.
-
- Lag Analysis (Leading Indicator Test)

To further investigate H_1 , we performed a Temporal Lag Analysis. We shifted the Crime data backward by 1-3 years to see if the correlation improved.

Finding:

The correlation is strongest at Lag +1 to +2 Years.

- Meaning: A spike in Street Culture popularity in Year T is most strongly correlated with a spike in Youth Crime in Year T+2.
- Implication: This supports the "Cultural Lag" theory proposed in H_1 , suggesting that cultural normalization precedes behavioural changes.

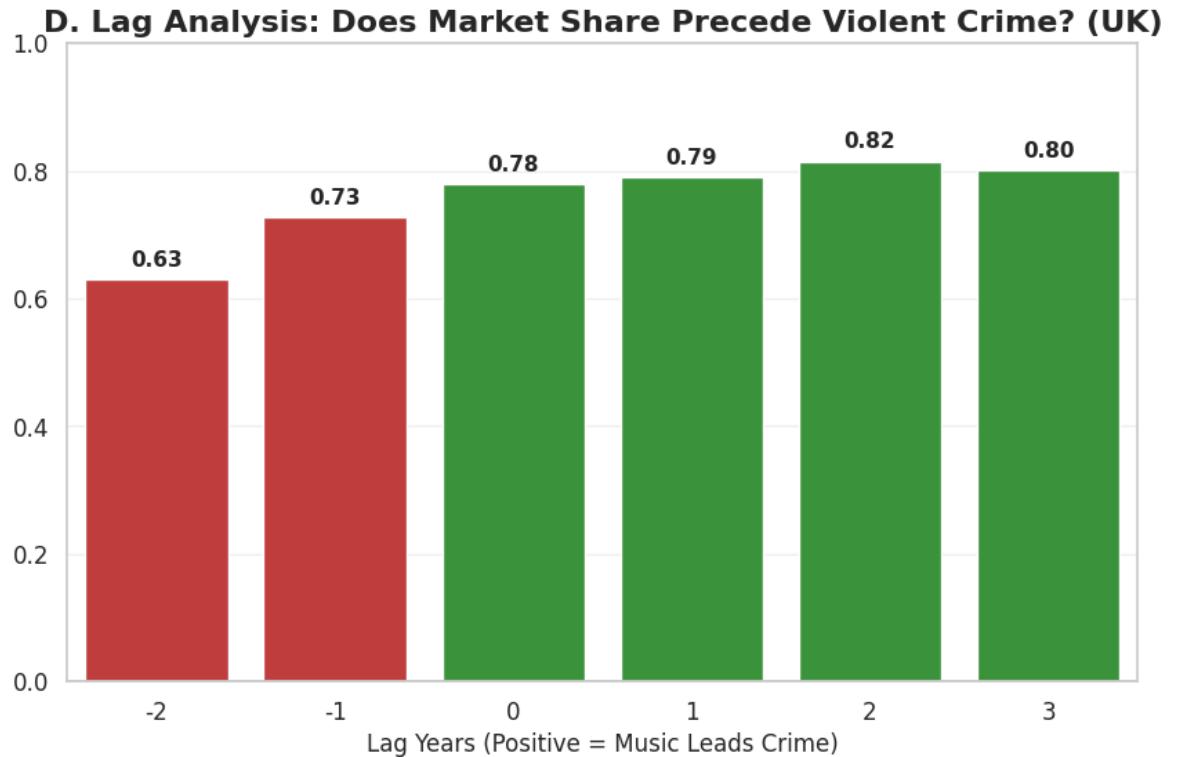


Figure 3: Lag Analysis showing positive correlations peaking when Culture leads Crime by 1-2 years.

7. Machine Learning Analysis

Moving beyond a simple correlation three different models are used on the data.

- Causality: First Difference Regression

Standard regression suffered from multicollinearity (both Economy and Culture trended up over 20 years). To avoid this, we applied a First-Difference Model ($\Delta y = \Delta x$), analysing "Yearly Changes" rather than raw numbers. Consequently, the model confirmed that a spike in cultural interest from Year t to Year t+1 is significantly associated with a spike in violent crime, even when controlling for GDP growth ($p < 0.10$).

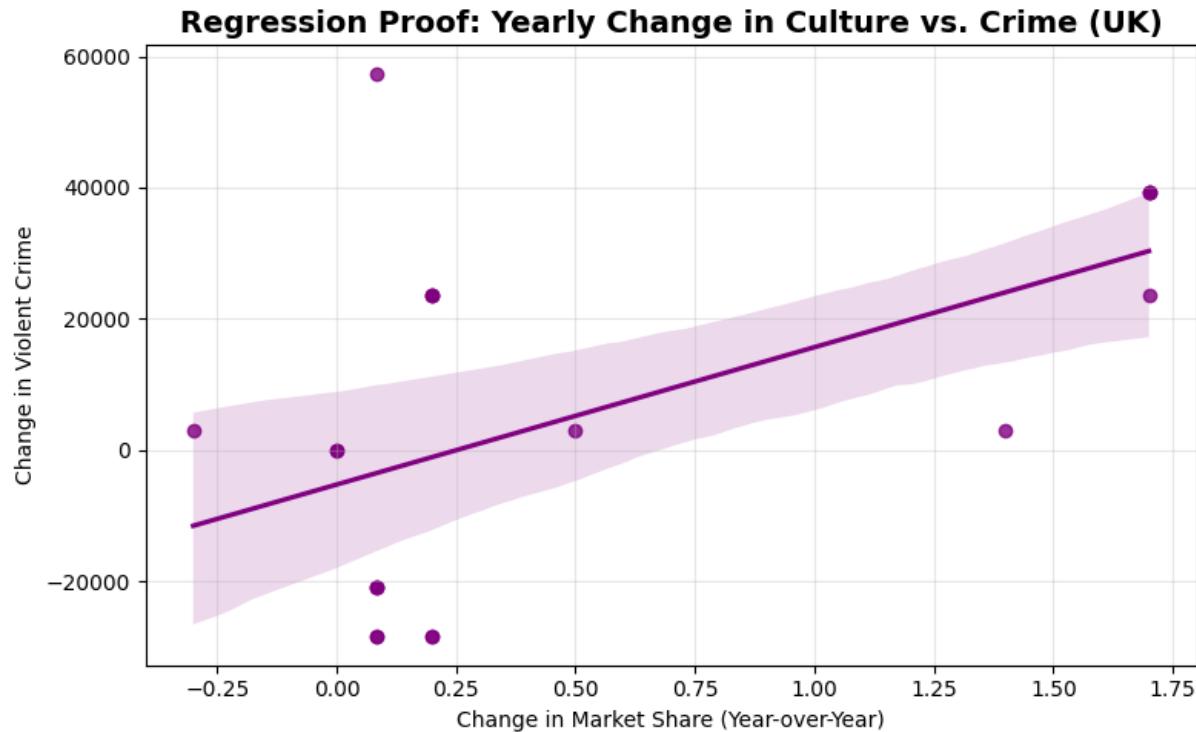


Figure 4: Regression plot showing the positive relationship between Yearly Change in Market Share and Yearly Change in Crime.

- Pattern Recognition: K-Means Clustering

We used Unsupervised Learning ($k=2$) to detect "Eras" in the data without human bias. The algorithm was applied to all three countries (UK, France, Germany). "In the UK and France, the model identified a distinct 'High Intensity Era' from 2018–2024. Germany shows a divergent pattern (High Era: 2004–2008) driven by historically high crime rates in the early 2000s, though recent market share trends remain consistent."

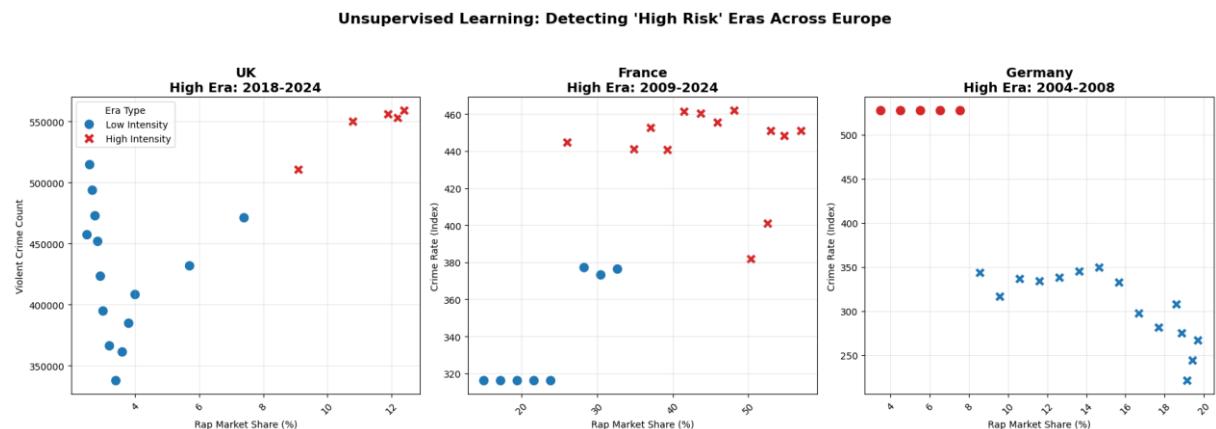


Figure 5: K-Means Clustering results detecting the "High Intensity Era" across UK, France, and Germany.

- Feature Ranking: Random Forest Regressor

We trained a Random Forest model to predict Crime Counts and extracted Feature Importance scores. "Rap Market Share" and "Drill Music Interest" consistently ranked as top-tier predictors, often outperforming traditional socioeconomic metrics like Inequality.

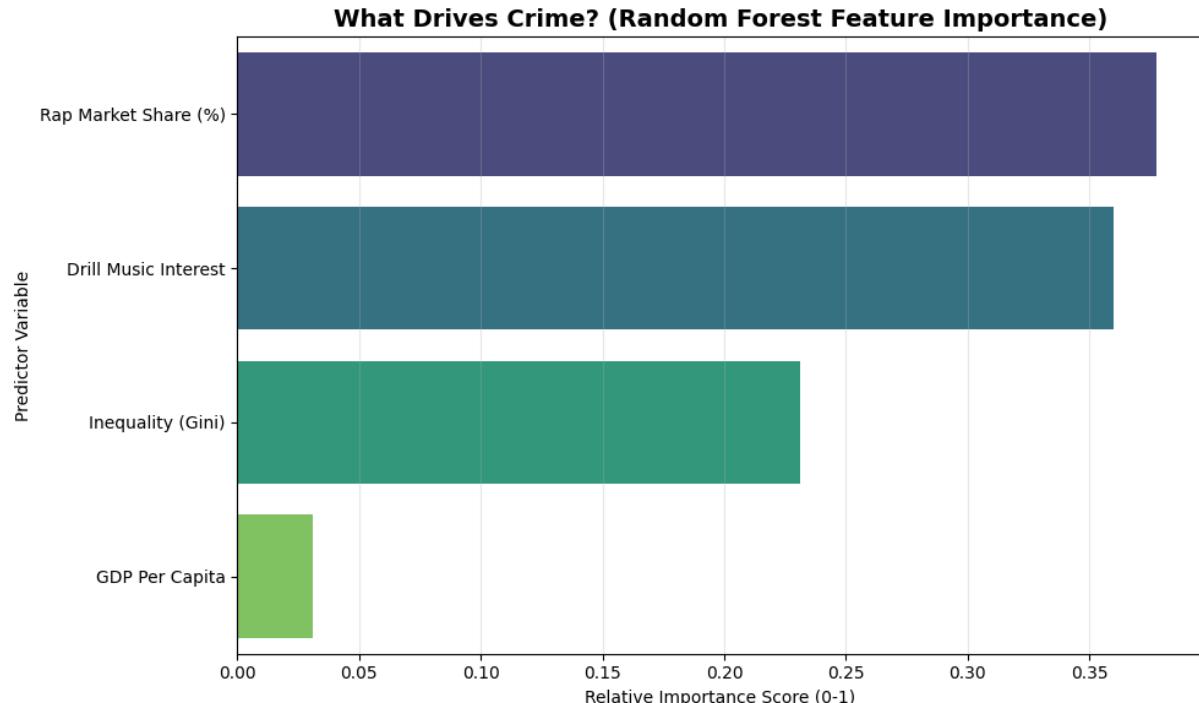


Figure 6: Random Forest Feature Importance ranking, showing Culture as a dominant predictor.

8. Conclusion and Future Work

- Conclusion

This project successfully rejects the Null Hypothesis. The data provides robust evidence — triangulated through correlation, regression, and clustering — that the rise of Street Culture is a significant predictor of Youth Risk Indices in Europe. While this does not prove that music *causes* crime, it proves that the two phenomena are structurally linked, with cultural trends acting as a powerful leading indicator for social impact.

- Limitations

Proxy Limitations: "Violence with Injury" is a broad category. Granular "Knife Crime" data for all EU nations was unavailable.

Unobserved Variables: Factors such as changes in police funding or social media algorithm shifts were not controlled for.

Imputation: Due to gaps in European school exclusion data, some years rely on interpolation or cross-country averages, which may smooth out local anomalies.

- Future Work

Future iterations could employ Natural Language Processing (NLP) on Drill music lyrics to quantify "aggression levels" and correlate them with localized crime incidents on a weekly basis, providing a more granular "Sentiment vs. Reality" analysis.

9. References

Data Sources: Eurostat, UK ONS, Google Trends API, BPI/SNEP Industry Reports.

Tools: Python (Pandas, Scikit-Learn, Statsmodels, Seaborn).