Evidence of Reduced Racism Through Sports: The Case of Iñaki Williams

Undergraduate Thesis

Students: Tomás Buscaglia y Martiniano López Fontanelle

Mentor: Osama Khalid

July 31, 2025

Abstract

This thesis examines whether the sustained visibility of Iñaki Williams, the first Black player to become a central figure at Athletic Club de Bilbao, has influenced racial attitudes in the Basque Country. Building on the Parasocial Contact Hypothesis, which posits that mediated exposure to minority figures can reduce prejudice, we employ both traditional crime data and novel social media metrics to assess changes in discrimination following Williams's rise to prominence in early 2016. Using a Difference-in-Differences framework and a Synthetic Control Method, we document a 32–34% reduction in monthly hate crimes and a 65% drop in annual hate-crime counts in the Basque Country relative to appropriate counterfactuals. Complementing these "hard" outcomes, we collect and classify over 2.4 million tweets using a state-of-the-art Spanish-language hate-speech model. Our findings reveal consistent post-2016 declines in hateful speech (–1.7%), aggression (–2.2%), and targeted insults (–1.4%) among Athletic supporters compared to synthetic controls. These results suggest that Williams's symbolic role disrupted entrenched narratives of ethnic exclusivity and catalyzed a modest recalibration of local norms around race and belonging. The thesis contributes to a growing literature on the social impact of minority representation in elite sports and highlights the value of combining official statistics with high-frequency online data to trace shifts in public sentiment.

Contents

1 Introduction 4

2 Literature Review 5
- 2.1 Comparative Cases: Salah and Giannis . . . 5
- 2.2 Context: Racism and Xenophobia in Spain and the Basque Country . . . 6
- 2.3 The Parasocial Contact Hypothesis . . . 6
  - 2.3.1 Application to Iñaki Williams . . . 7

3 Hate Crimes 8
- 3.1 Data Description . . . 9
- 3.2 Monthly Data . . . 10
- 3.3 Annual Data . . . 12
- 3.4 Hate Crime Conclusions . . . 15

4 Twitter Data 15
- 4.1 Scraping . . . 16
- 4.2 Data Description . . . 16
- 4.3 Tweet Classification . . . 17
  - 4.3.1 Preprocessing . . . 17
  - 4.3.2 Classification . . . 17
- 4.4 Results . . . 17
  - 4.4.1 Hateful Speech Probability . . . 18
  - 4.4.2 Aggressive Speech Probability . . . 19

5 Conclusion 20

6 Bibliography 22

7 Appendix 24
- 7.1 Targeted Speech Probability . . . 24

1 Introduction

On February 19, 2015, a promising 20-year-old talent from Athletic Bilbao's academy made his international debut in the UEFA Europa League. Within just eight minutes of play, he achieved a historic milestone by becoming the first player of African descent to score a goal for Athletic Bilbao in the club's 117-year history. In 2016 Williams cemented his place in the team and began a record-setting streak of consecutive appearances in La Liga, eventually reaching 251 matches played without interruption, a testament to both his consistency and growing symbolic visibility.

But why is Iñaki Williams the first player of African descent to play for Athletic Bilbao? The club adheres to a strict "no-outsider" policy, allowing only players born in the Basque Country to join. Williams' eligibility is rooted in his parents' remarkable journey. María and Félix, refugees fleeing Ghana in search of asylum, endured a long and challenging journey before arriving in Spain. Following complications with authorities and a period of detention, María gave birth to Iñaki in Bilbao, a turn of fate that became a profound blessing for the Williams family and, ultimately, for Athletic Bilbao.

In his early years at Athletic Bilbao, Iñaki Williams quickly established himself as a key player, known for his speed, versatility, and durability. Since his debut, he has played over 400 matches for the club, scoring more than 80 goals and providing 20 assists. His contributions were instrumental in Athletic Bilbao's victory in the 2020–2021 Supercopa de España, further solidifying his legacy within the team.

Williams is not only recognized for his football skills but also for his distinctive identity as a player of African descent, a rarity in the football landscape of the region. Media coverage, particularly that which highlights his African heritage, underscores the cultural impact he has had on the Basque Country. Throughout his career, he has been an outspoken advocate for immigrants and Black minorities, using his platform to promote inclusivity and social awareness. Aquino (2017) provides a compelling analysis of his significance in relation to Basque national identity, stating, "Williams's presence on Athletic does not guarantee that African immigrants' lives will improve in any material way; it does suggest, however, that a Black Basque body has become thinkable for the first time. The phenotypic boundaries of Basqueness have widened for good." But has Iñaki had a real impact on the Basque Country?

The idea that exposure to Williams may have contributed to reducing racism is grounded in the parasocial contact hypothesis, Schiappa et al. (2005), which suggests that engagement with celebrities or prominent figures from minority groups can help diminish prejudice toward that group as a whole. This hypothesis has been supported by real-world studies demonstrating causal effects. Notably, Ala' Alrababa'h's (2021) research examined the impact of Mohamed Salah on reducing Islamophobia in Liverpool, while subsequent work by Bosley (2023) explored a similar phenomenon in Milwaukee, analyzing the influence of Giannis Antetokounmpo on racial attitudes. Building on this line of research, we seek to determine whether a comparable effect can be observed in the case of Iñaki Williams, within the distinct cultural and social context of the Basque Country.

To comprehensively assess racism and xenophobia, we incorporated multiple measures following Ala' Alrababa'h's methodology, combining both direct crime data and social network analysis. This approach allows us to capture two distinct dimensions of discrimination: hate crime data reflect the actions of the most overtly racist individuals who engage in criminal behavior, while Twitter data provide insights into broader societal attitudes and the more subtle, everyday expressions of racism. Given that hate crimes, though highly damaging, are relatively rare and extreme events, relying solely on them would offer a limited perspective. By including social media data, we aim to understand how Williams' signing may have influenced more routine and widespread forms of racist behavior among Athletic Bilbao fans.

2 Literature Review

2.1 Comparative Cases: Salah and Giannis

Recent research has turned to the role of celebrity figures in shaping societal attitudes, particularly regarding prejudice and intergroup relations. Two notable studies, Bosley's (2023) analysis of Giannis Antetokounmpo's impact on racial and immigrant bias in Wisconsin, and Alrababa'h et al.'s (2021) investigation into Mohamed Salah's influence on Islamophobic behavior in the UK, offer compelling, complementary evidence for the parasocial contact hypothesis. Drawing on Allport's (1954) foundational Contact Hypothesis, which posits that interpersonal interaction across group lines under the right conditions can reduce prejudice, these studies extend the logic to mediated forms of contact. Specifically, the Parasocial Contact Hypothesis (Schiappa et al., 2005) argues that repeated, one-sided exposure to outgroup members, particularly through media or celebrity fandom, can produce similar reductions in bias, especially when the celebrity's group identity is salient, the portrayal is positive, and the exposure is consistent.

In both cases, the researchers select athletes who embody characteristics associated with historically stigmatized identities, Giannis Antetokounmpo as a Black immigrant in the United States and Mohamed Salah as a visibly Muslim footballer in the United Kingdom. Both studies examine how the athletes' personal traits and public personas influence societal attitudes beyond the sports arena. Crucially, both athletes fulfill the theoretical conditions that make parasocial contact effective: they are not only extremely successful in their respective sports, but also consistently portrayed in the media in a favorable light, and they visibly represent their minority identity. Salah, for instance, often performs sujood (Islamic prayer) after scoring goals and is openly devout, while Giannis is known for publicly discussing his experiences as an immigrant and Black man in the U.S., and has actively participated in the Black Lives Matter movement.

Methodologically, the two studies take different but complementary approaches to identifying the causal effects of celebrity exposure. Alrababa'h et al. employ a multi-method strategy that includes a synthetic control analysis of hate crime data in Merseyside (Liverpool's home region), a large-scale computational text analysis of over 15 million tweets from UK-based soccer fans, and a survey experiment conducted among over 8,000 Liverpool FC fans. Their results demonstrate a significant 16% reduction in hate crimes and a halving of anti-Muslim tweets among Liverpool fans following Salah's arrival. The survey experiment further confirms that priming respondents with Salah's Muslim identity increased perceived compatibility between Islam and British values, suggesting a clear generalization of positive feelings from the individual to the group, precisely the mechanism theorized by parasocial contact. Bosley's analysis of Giannis's effect in Wisconsin adopts a similar synthetic control framework, using hate crime statistics and Google Trends data to evaluate changes in anti-Black and anti-immigrant sentiment before and after Giannis's rise to prominence with the Milwaukee Bucks. While the effects are somewhat less definitive than in the Salah study, the findings nonetheless point to reductions in both hate crimes and racially charged internet searches, particularly during periods of peak fan engagement, such as the Bucks' championship run in 2021.

A crucial point of comparison among the cases lies in how the minority identity of each athlete is made visible and socially salient. Mohamed Salah's Muslim identity is not only prominent but continually reinforced, through ritualistic gestures such as sujood prayer after goals, religious references in fan chants, and media portrayals that emphasize his culture. This consistent visibility offers a clear and direct focal point for audiences to reevaluate preconceived notions about Islam. In contrast, Giannis Antetokounmpo's identities, as a Black African immigrant and formerly undocumented youth, are deeply meaningful but less performatively emphasized in his public image. While Giannis's backstory is widely known and frequently cited in biographical features, it is not always foregrounded in the same ritualized or symbolic ways that characterize Salah's presence. Nevertheless, Giannis's capacity to unite a racially segregated Milwaukee through shared support for the Bucks, particularly during emotionally charged moments like the Black Lives Matter protests, suggests that his impact operated through dual mechanisms: parasocial identification with his persona and real-world intergroup contact among fans who gathered collectively in public spaces. Whereas Alrababa'h et al. (2021) conceptualize parasocial contact as the primary mechanism of attitudinal change, Bosley (2023) notes that the civic rituals of fandom, embodied in spaces like Milwaukee's "Deer District", may have created opportunities for face-to-face intergroup encounters that complemented or amplified the parasocial effect. This distinction between mediated and embodied contact becomes essential when considering how audience engagement and identity salience interact to shape prejudice reduction.

Collectively, these studies highlight the growing scholarly recognition of celebrity figures as agents of social change, particularly in the domain of intergroup relations and prejudice reduction. Both Bosley (2023) and Alrababa'h et al. (2021) provide compelling empirical support for the Parasocial Contact Hypothesis (Schiappa et al., 2005), extending Allport's (1954) original theory of intergroup contact into the mediated realm of modern sports fandom. Despite differing in geographic and cultural contexts, the two studies converge on a core set of mechanisms: repeated exposure to a minority figure in a positive and salient light can lead to measurable improvements in social attitudes and behaviors, including reductions in hate crimes and discriminatory discourse.

2.2 Context: Racism and Xenophobia in Spain and the Basque Country

The broader social and political context in which Iñaki Williams' career unfolds is essential to understanding the potential effects of parasocial contact in Spain. In recent years, racism and xenophobia have become prominent topics of public discourse in Spain, driven in part by rising levels of immigration.

In the Basque Country, where Williams plays for Athletic Club de Bilbao, the dynamics of race intersect with a long-standing ethno-regional nationalism. Athletic Club's unique recruitment policy, limited to players with Basque heritage or developed in the region, has historically aligned the club with the preservation of Basque identity. While this policy has fostered a strong sense of local pride and cultural continuity, it has also drawn scrutiny for its lack of racial diversity. Williams, born in Bilbao to Ghanaian parents and raised in a Basque-speaking environment, challenges the essentialist narratives that have long defined Basque identity. His presence as the first Black player to score for the club in its 117-year history challenges traditional views of who can be considered "authentically" Basque and invites a broader societal reflection on race, belonging, and identity.

The potential for Williams to influence racial attitudes in the Basque Country and Spain more broadly must be situated within this cultural landscape. The region remains relatively homogeneous in racial terms, and encounters with Black individuals are often mediated through sports or television, rather than interpersonal relationships. This structural isolation heightens the relevance of parasocial exposure through public figures like Williams, whose repeated media presence and regional integration offer a rare opportunity for mediated intergroup contact.

2.3 The Parasocial Contact Hypothesis

The Parasocial Contact Hypothesis offers a compelling theoretical lens through which to understand the potential impact of Iñaki Williams on racial attitudes in Spain. Rooted in Allport's (1954) classic

Contact Hypothesis, the theory proposes that intergroup contact, when positive, frequent, and endorsed by social norms, can reduce prejudice through mechanisms such as increased empathy, anxiety reduction, and the re-categorization of group boundaries (Pettigrew & Tropp, 2006). However, in many societies, structural segregation and cultural isolation limit opportunities for direct interpersonal contact with outgroups. Recognizing this limitation, Schiappa, Gregg, and Hewes (2005) introduced the Parasocial Contact Hypothesis, suggesting that repeated exposure to minority group members via mass media can produce similar prejudice-reducing effects. This mediated form of contact becomes especially significant in environments where minority visibility is low and stereotypes dominate public discourse.

Empirical support for the parasocial contact hypothesis has been found across multiple domains, including attitudes toward racial, religious, and sexual minorities. For instance, Ramasubramanian (2015) demonstrated that positive media portrayals of Black celebrities could reduce implicit bias among viewers. Similarly, Park (2012) found that exposure to minority characters on television increased acceptance of outgroup members in real-life social contexts. As previously mentioned, the study of Mohamed Salah by Alrababa'h et al. (2021) represents one of the most rigorous tests of this theory in a naturalistic setting.

In applying this framework to Iñaki Williams, several questions arise. Does his visibility as a Black player in a traditionally white, ethnically exclusive team provide the conditions for effective parasocial contact? Are his personal qualities, humility, regional fluency in Basque, and anti-racist advocacy sufficient to generate widespread admiration and empathy? And critically, is his racial identity made salient in ways that allow audiences to generalize their attitudes toward Black individuals more broadly, or is it compartmentalized as an exceptional case? The literature underscores the importance of repeated, emotionally resonant exposure, combined with a clear group identity, for parasocial contact to generate attitudinal shifts (Schiappa et al., 2005; Pettigrew & Tropp, 2006; Park, 2012). Williams' consistent presence in La Liga, his symbolic status within Athletic Club, and his vocal stance against racism provide a promising foundation.

2.3.1 Application to Iñaki Williams

For parasocial contact to generate generalized tolerance, the minority identity of the public figure must be readily visible and cognitively linked to the broader group (Allport, 1954; Schiappa et al., 2005). In the case of Iñaki Williams, his Black identity is unambiguously salient within the context of Spanish and particularly Basque football. Athletic Club de Bilbao has long been associated with an ethno nationalist project of Basque cultural preservation, exemplified by its strict *cantera* policy that historically limited recruitment to players of Basque origin or upbringing. One of the club's mottos is *"Con cantera y aficion no hace falta importacion"* that translates to "With homegrown talent and loyal fans there is no need for importing". As such, the image of a Black footballer representing this institution disrupts longstanding racialized conceptions of Basqueness that have tended to exclude people of African descent.

Williams' racial identity is consistently highlighted in media coverage, often in tandem with narratives about his integration into Basque culture, his fluency in Euskera, and his family's migration story from Ghana via refugee routes. Unlike Muslim identity in Salah's case which was made salient through prayer rituals, personal naming, and media framing, Williams' blackness is visually unavoidable and frequently referenced when issues of race in Spanish football arise. His status as the first Black goalscorer in the club's 117-year history further anchors his identity in the public consciousness as not just a sportsman, but a symbol of racial change within an ethnically coded institution. Thus, the condition of identity salience is robustly met, and arguably even amplified by the regionalist and racially homogenous backdrop against which his career unfolds.

The Parasocial Contact Hypothesis also presupposes that the public figure in question is portrayed in a consistently favorable light (Paolini, Harwood & Rubin, 2010). Williams' public persona, as constructed through both traditional and social media, is overwhelmingly positive. He is frequently described as humble, loyal, and principled, qualities that resonate deeply within the cultural ethos of both Athletic Club and the Basque public. His decision to remain at Athletic Bilbao despite transfer opportunities from wealthier clubs is framed as a gesture of loyalty and rootedness, enhancing his status not just as a talented player but as a moral figure aligned with community values.

Importantly, Williams has also been portrayed as an articulate and courageous advocate against racism. His calm but firm condemnation of racist abuse, especially following an incident during a match against Espanyol in 2020, was widely covered in Spanish and international media. He has consistently framed his responses within a discourse of progress and solidarity, avoiding confrontation while urging institutional reform. This strategic positioning reinforces a narrative in which Williams is not only a victim of discrimination but a dignified agent for positive change. Such portrayals have the potential to evoke admiration, empathy, and ultimately, the rehumanization of Black individuals in a context where racial othering is common (Dovidio et al., 2017).

The third condition for effective parasocial contact is the sustained and repeated exposure of audiences to the minority figure (Schiappa et al., 2005; Park, 2012). In the case of Iñaki Williams, his continuous and high-profile presence within Athletic Club de Bilbao grants him a unique degree of visibility not only in Spanish football, but also within Basque society. Since becoming a regular starter in 2016, Williams has become a mainstay in the club's lineup, famously holding the La Liga record for consecutive appearances (251 matches between 2016 and 2023). His uninterrupted presence ensures that Basque fans, who maintain one of the most passionate and locally embedded football cultures in Europe, have engaged with his image consistently over the course of nearly a decade. This repeated, local exposure allows for familiarity to develop at a scale that is both personal and collective. In this sense, Williams is not just "seen", he is a continuous presence in the cultural fabric of the Basque Country.

Thus, within the specific sociocultural environment of the Basque Country, Williams fulfills the condition of consistent exposure in a particularly potent way. His integration into one of the region's most revered institutions ensures that his presence is not only visible but meaningful. This immersion offers fertile ground for parasocial identification to occur, making it possible for attitudes toward racial minorities to shift gradually through the mechanisms of familiarity, admiration, and everyday engagement.

3 Hate Crimes

We begin our analysis with what can be described as a "hard test" of the parasocial contact hypothesis: the examination of hate-crime rates in the Basque Country surrounding the rise of Iñaki Williams. If his visibility and symbolic role within Athletic Club de Bilbao truly shifted public animus toward black immigrants then we should observe a discernible drop in hate crimes motivated by racism and xenophobia in the region. Hate crimes are typically committed by individuals on the ideological margins rather than by "everyday" citizens. Indeed, empirical work finds that perpetrators of ideologically motivated violence disproportionately exhibit traits associated with radicalization, prior violent offenses and social isolation, distinguishing them from the general population. (Schuurman & Carthy, 2025).

Moreover, hate crimes tend to be public or semi-public acts (e.g., vandalism of religious sites, assaults in public spaces) making offenders acutely sensitive to social norms and community sanction. Thus, a sustained reduction in hate-motivated incidents would imply not only a shift in extremists' private attitudes but also a recalibration of what is publicly permissible. It is also important to recognize that, unlike instantaneous expressions of prejudice on social media, hate crimes typically require planning, mobilization, and in many cases coordination.

Criminological research shows that many hate crimes are retaliatory in nature, often arising as delayed responses to perceived provocations or triggering events. As King and Sutton (2013) document, hate incidents tend to increase in the aftermath of salient events that generate intergroup grievances, suggesting that bias-motivated violence commonly emerges after a period of grievance accumulation and rationalization. Consequently, any influence that Williams may exert on public attitudes is unlikely to translate immediately into reductions in hate crime. Instead, we expect a lagged effect as social norms evolve and the recalibration of what is deemed publicly permissible behavior gradually diffuses through society. Our empirical framework therefore allows for a temporal delay between Williams's rise to prominence and observable changes in hate crime rates.

3.1 Data Description

To assess the impact of Iñaki Williams' presence on real-world hate crimes, we sought granular, time-series data from two primary sources. Our first approach involved contacting numerous police departments and government agencies at the local, provincial, and national levels across Spain. We specifically requested monthly hate crime statistics at the provincial level, covering the period from 2013-2014 onwards. While this effort yielded limited success, the Basque Country Police Department provided us with a report detailing the monthly number of reported hate crimes for three regions within the Basque Country. Other provinces either did not respond to our requests or provided only annual data, which was insufficient for our analysis.

Our second data source was the *Observatorio Español del Racismo y la Xenofobia*[1], a body within the *Ministerio de Inclusión, Seguridad Social y Migraciones* of the Spanish government. This source provided two types of relevant data: (1) monthly aggregate hate crime statistics for the entire country of Spain, and (2) annual hate crime statistics disaggregated by province.

After all, we are left with two complementary datasets. The first is a monthly dataset that provides disaggregated hate crime statistics for the Basque Country alongside corresponding national-level data for Spain. This dataset spans from January 2014 to December 2019. The monthly granularity of this data enables us to analyze temporal trends and identify deviations in the Basque Country relative to national patterns over time.

The second dataset is annual and covers the same time frame 2014 to 2019 but provides data disaggregated at the regional and provincial levels across all of Spain's autonomous communities. This dataset allows for a cross-sectional comparison of hate crime trends between the Basque Country and other regions over the same period.

The trade-off, therefore, lies in the level of geographic and temporal aggregation. The monthly comparison provides fine-grained temporal data but a broader geographic scope (Basque Country vs. Spain). The annual comparison offers more geographically specific data (provincial level) but at a wider temporal scale (annual).

In both datasets, we operationalize hate crimes using the official definition provided by Spanish authorities. Specifically, we include all reported "delitos de odio" (hate crimes), defined as: any criminal offense, whether against individuals or property, where the victim, location, or target is selected due to their real or perceived association with a protected group. These protected characteristics include,

[1] Observatorio Español del Racismo y la Xenophobia.(OBERAXE) consulted at https://www.inclusion.gob.es/web/oberaxe

but are not limited to, race, national or ethnic origin, language, skin color, religion, age, physical or mental disability, and sexual orientation. This definition, issued by Spain's Ministry of the Interior and used by national observatories such as OBERAXE, ensures consistent and rigorous measurement of hate-motivated incidents across regions and time periods.

Additionally, for our analysis to be properly identified, we needed to ensure that no major immigration trends or events disproportionately affected the Basque Country compared to Spain as a whole or other regions. To address this, we examined immigration data and found no evidence of significant regional discrepancies that could confound our results.

To test our hypothesis, we implement two distinct methodological approaches corresponding to the two datasets previously described. Both rely on an event-study framework that exploits the timing of Iñaki Williams' incorporation into Athletic Club de Bilbao and leverage the structure of each dataset to produce consistent estimates of the potential effect.

We designate January 2016 as our treatment period reflecting the period where Iñaki Williams solidified his position as a starter in the first team and began receiving heightened media and public attention within the Basque Country. In this period Williams delivered 5 goals across 8 matches underscoring his newfound status as an indispensable first-team figure. To further justify establishing this period as the treatment we analyzed the total number of tweets mentioning Iñaki from a selection of Athletic Club fan accounts. A clear spike in mentions is observed during this goalscoring streak. Given both statistics, we denote this month as the treatment period.

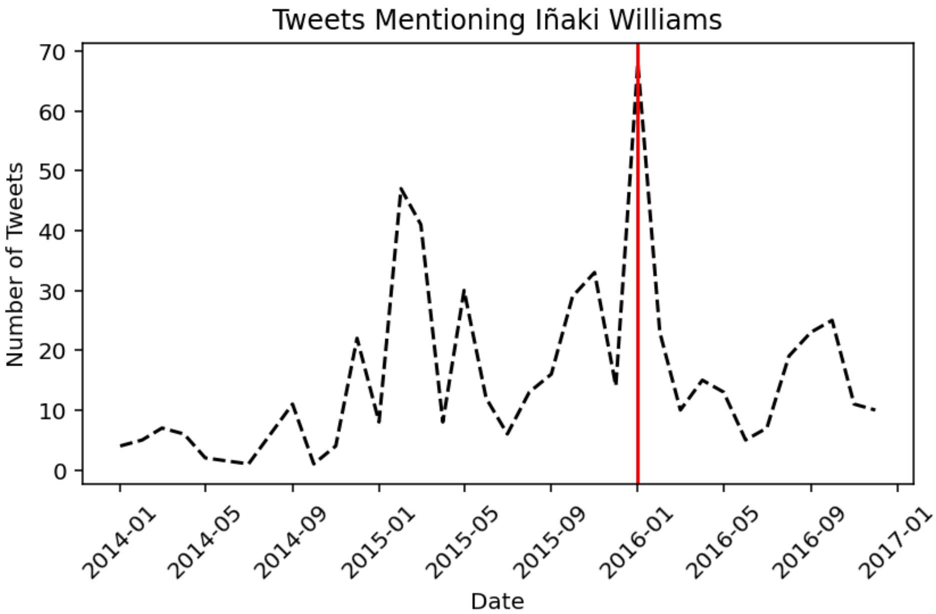

Figure 1: Tweets mentioning Iñaki Williams. *Source: Author's elaboration based on collected tweets.*

3.2 Monthly Data

We begin our empirical analysis by implementing a Difference-in-Differences (Ashenfelter & Card, 1984) model using our monthly panel dataset to evaluate whether the incorporation of Iñaki Williams into Athletic Club de Bilbao corresponds with a reduction in hate crimes in the Basque Country, relative to trends observed in the rest of Spain. The model exploits temporal variation before and after Williams' incorporation as a starter, with the Basque Country serving as the treated unit and Spanish national data as the control group.

The regression specification includes month fixed effects to account for national seasonal shocks and time-varying factors common across regions. Our main outcome variable is the monthly rate of hate crimes per 100,000 inhabitants, constructed as follows: the total number of reported hate crimes in a given month divided by the regional population, multiplied by 100,000. The dataset comprises 72 monthly observations for both the treatment and control units, covering the period from January 2014 to December 2019. The dependent variable ranges from a minimum of 0.00 to a maximum of 1.05 hate crimes per 100,000 residents, with a mean of 0.3324 and a standard deviation of 0.1949. We apply a five month moving average to our graph to mitigate the high temporal variability that is characteristic of crime data.

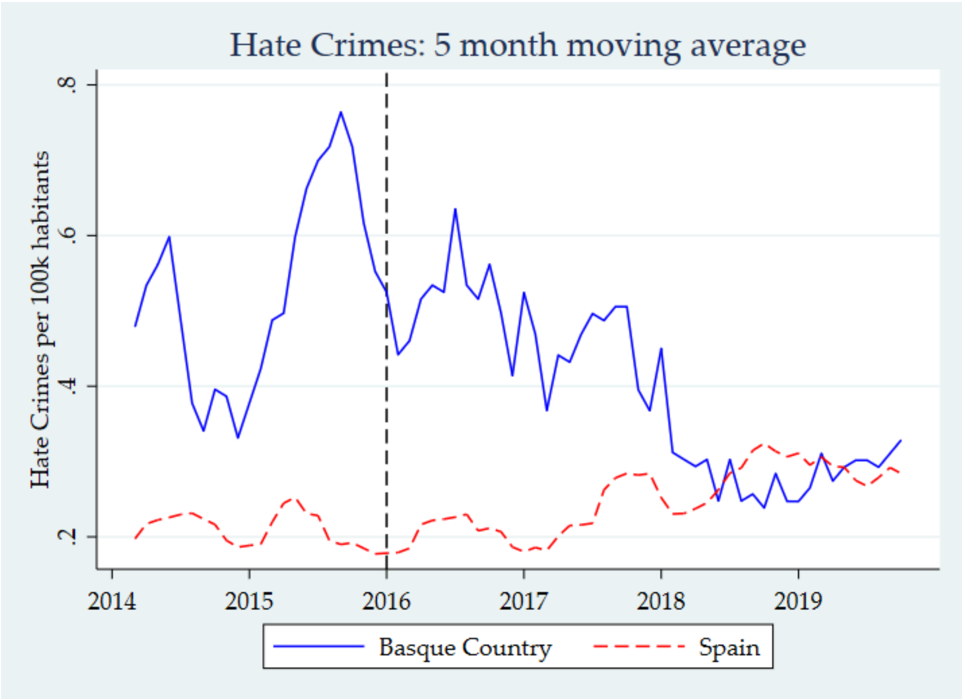

Figure 2: Hate crime evolution. *Source: Author's elaboration based on data from OBERAXE*

As illustrated in Figure 2, the trend in hate crimes in the Basque Country exhibits a pronounced downward shift following Iñaki Williams' incorporation into Athletic Club de Bilbao. Historically, the Basque Country has recorded one of the highest rates of racism motivated hate crimes among Spain's autonomous communities (Lozano Martín et al., 2023). Rates per 100,000 habitants in the region were consistently elevated relative to the national average until approximately 2016, at which point the gap began to narrow. One plausible explanation is that the Basque Country has traditionally been a relatively insular region, with limited integration of immigrant populations and a strong cultural emphasis on longstanding social cohesion. In contrast, the national rate for Spain has remained comparatively low and stable throughout the period, only beginning to increase around 2018.

The key explanatory variable in our analysis is the interaction term between the treatment group (Basque Country) and the post-treatment period (after Williams became a regular starter). This variable captures the causal effect of Williams' emergence on hate crime rates under standard DiD assumptions.

The model is specified as follows:

$$\text{HateCrime}_{it} = \beta_0 + \beta_1 \cdot \text{treated} + \beta_2 \cdot \text{post\_treatment} + \beta_3 \cdot \text{interaction} + \sum_i \beta_i \cdot \text{month\_fe}_i + \varepsilon_{it} \quad (1)$$

Regression Terms

β_0**:** Baseline level of hate crimes in the control group prior to treatment.
β_1**:** Average difference in hate crimes between the treated and control groups before treatment.
β_2**:** Post-treatment change in the control group.
β_3**:** The Difference-in-Differences effect (the causal effect of the treatment).
$\sum_i \beta_i \cdot \textbf{month\_fe}_i$**:** Month fixed effects to control for time shocks common to all regions.
ε**:** The error term.

This specification allows us to isolate the differential change in hate crimes in the Basque Country that coincides with the timing of Williams' ascent to prominence, while controlling for both regional differences and national temporal trends. The identification assumption is that, if there had been no treatment, the Basque Country would have followed a parallel trend to the rest of Spain in the evolution of hate crimes.

Statistical inference in the context of a single treated unit poses well-known challenges, particularly due to the risk of serial correlation and biased standard errors. As highlighted by Bertrand, Duflo, and Mullainathan (2004), conventional inference methods can severely understate standard errors in Difference-in-Differences settings. To address this, we apply Newey–West standard errors, which adjust for heteroskedasticity and autocorrelation in the error structure (Newey & West, 1986).

Table 1: Effect of Treatment on Hate Crimes

	(1) Baseline Model	(2) Controlling for Employment Level
Interaction	-0.165**	-0.157*
Standard Error	(0.0803)	(0.0807)

Our results indicate that the main variable of interest has a negative and statistically significant effect on hate crimes per 100,000 inhabitants. Specifically, we estimate a reduction of 0.165 incidents, which represents a substantial decline given that the overall mean for the period 2014–2019 is 0.428. When compared to the pre-treatment mean of 0.51, this corresponds to an approximate 32% decrease in the incidence of hate crimes. The effect is statistically significant at the 5% level with a p-value of 0,016.

We extend our baseline Difference-in-Differences model by estimating a second specification, Regression (2), that includes a control variable. Specifically we include the unemployment rate, a variable that is key to understanding whether the effect we are capturing can be attributed to Williams, or to a rising unemployment rate (Raphael and Winter-Ebmer, 2001). The inclusion of this covariate results in a modest attenuation of the estimated treatment effect. The coefficient remains statistically significant at the 10% level.

3.3 Annual Data

Our second empirical strategy relies on a yearly panel dataset spanning from 2014 to 2019, which includes observations for all provinces and autonomous communities in Spain. Based on this data structure, we employ the Synthetic Control Method to estimate how hate crime rates in the Basque Country evolved following the incorporation of Iñaki Williams into Athletic Club relative to what would have happened had he not joined the team.

The Synthetic Control Method, introduced by Abadie and Gardeazabal (2003) and further developed by Abadie, Diamond, and Hainmueller (2010), offers a compelling alternative to Difference-in-Differences by constructing a weighted combination of synthetic units that closely replicates the treated unit's pre-intervention trajectory. This approach is especially well-suited for case studies with a single treated unit, such as the Basque Country in our analysis, and provides a transparent and statistically grounded estimate of the treatment effect through comparison with its synthetic counterpart.

However, the application of the synthetic control algorithm to this dataset posed specific challenges. The historically high rate of hate crimes per 100,000 inhabitants in the Basque Country significantly limited the pool of comparable regions. In all iterations of the model, the algorithm selected only one other region, typically the one with the second-highest hate crime rate, as a viable match in the pre-treatment period. This lack of donor pool diversity weakened the credibility of the synthetic counterfactual. Given this constraint, we opted to implement the synthetic control using the absolute number of hate crimes, rather than normalizing the outcome by population size.

To estimate the treatment effect for the treated unit, we calculate the difference between the observed outcome in the post-treatment period and the corresponding counterfactual outcome generated by the synthetic control. This yields one treatment effect estimate for each of the T post-treatment periods. As a summary measure, we also compute the average treatment effect across all post-treatment periods, providing a simple summary of the treatment effect.

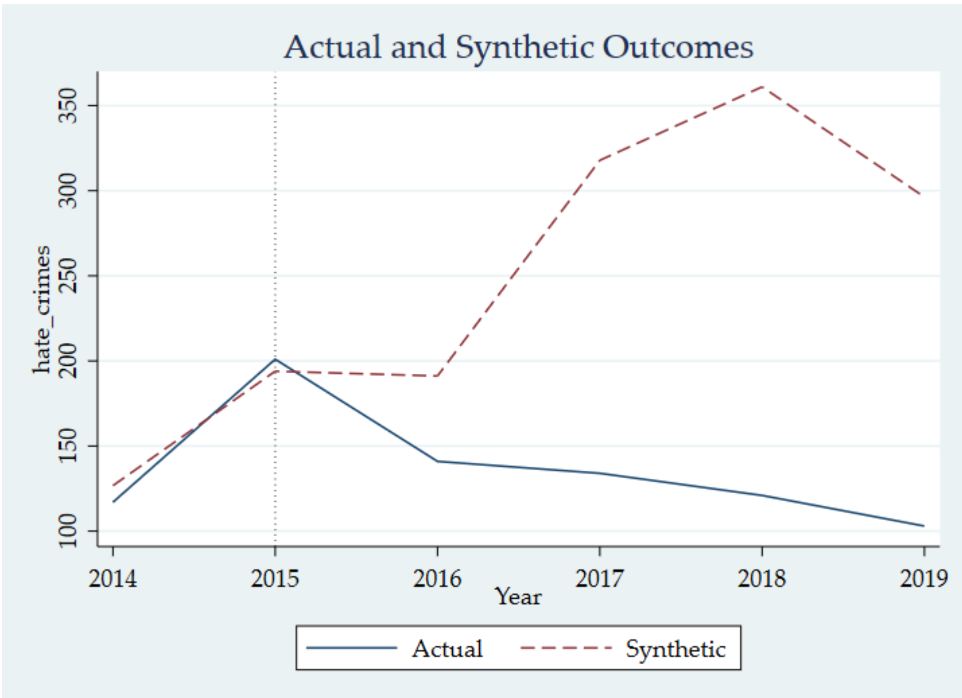

Figure 3: Synthetic Control vs. Treated Unit. *Source: Author's elaboration based on data from OBERAXE*

Figure 3 shows our synthetic control comparison. After matching the pretreatment trend we can see that the actual values in the basque country go downwards after 2016 while the counterfactual synthetic unit increases consistently. We find an average treatment effect of -166 hate crimes when comparing the two. For the last period we observe that our synthetic unit has 296 hate crimes while the actual outcome is approximately 103.That is a 65% drop between the actual outcome and the counterfactual.

To match the pre treatment trend, the synthetic unit is created using the following regions and weights:

Table 2: Synthetic Control Weights by Region

Region	Weight (%)
Barcelona	89%
Asturias	11%

To assess the robustness of our synthetic control results, we implement an in-space placebo test, a standard falsification technique used to determine whether the observed treatment effect for the treated unit is unusually large relative to effects estimated for untreated units (Abadie, Diamond, & Hainmueller, 2010). This approach involves iteratively reassigning the treatment to each control unit (in this case, the other Spanish regions) while maintaining the original treatment date and estimating synthetic controls for these placebo cases. The resulting distribution of placebo effects provides a benchmark against which we can evaluate the magnitude and uniqueness of the treatment effect observed for the Basque Country.

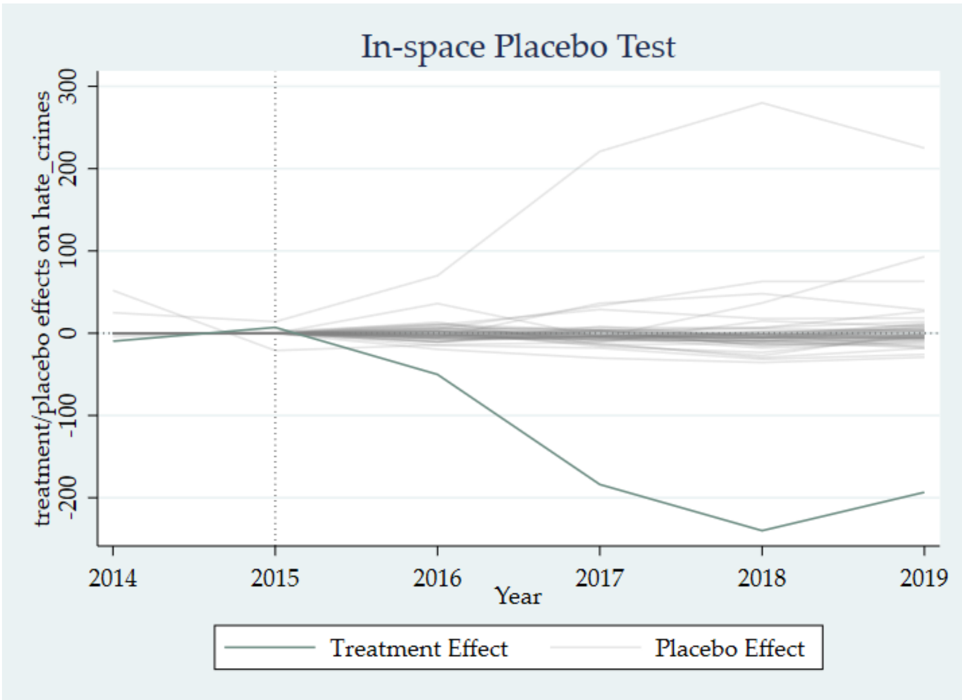

Figure 4: Placebo Tests. *Source: Author's elaboration based on data from OBERAXE*

Figure 4 displays the outcome of this placebo test. As the graph shows, following the treatment year, the treatment effect diverges sharply downward, indicating a notable decline in hate crimes in the Basque Country. In contrast, the placebo effects remain relatively flat or fluctuate upward, and no other region exhibits a comparable or more negative trend. This divergence reinforces the credibility of our findings, suggesting that the reduction in hate crimes is not a spurious result, but rather a treatment-specific effect plausibly associated with Iñaki Williams' growing visibility and symbolic relevance.

As a second robustness check, we conduct an in depth analysis to assess whether the observed decline in hate crimes may be part of a broader regional decrease in overall criminal activity. To do so, we examine the evolution of all other available categories of crime over the same period.

As shown in Figure 5, the decline appears to be specific to hate crimes. Other crime categories either remain relatively stable or exhibit upward trends. This differential pattern strengthens the plausibility of our findings, suggesting that the reduction in hate crimes is not simply driven by general shifts in

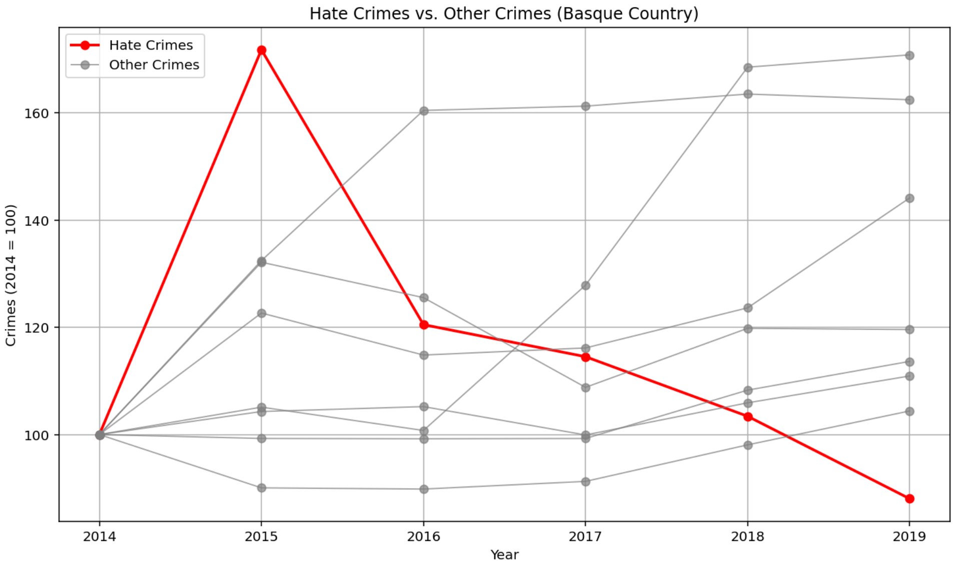

Figure 5: Comparison of Hate Crimes and Other Types of Crimes *Source: Author's elaboration based on data from OBERAXE*

regional crime dynamics.

3.4 Hate Crime Conclusions

In sum, our analysis across both monthly and annual datasets provides consistent, if cautiously interpreted, evidence of what might be termed the "Williams effect." Following Iñaki Williams elevation to regular starter status in early 2016, we observe a substantial downward shift in hate-crime rates, approximately a 34% reduction in monthly incidents per 100,000 inhabitants, and a decrease of 65% total annual offenses in the Basque Country. Although limitations in data granularity temper the strength of causal claims, the convergence of findings from a Difference-in-Differences specification and a synthetic control framework suggests that Williams' unprecedented visibility and symbolic role may have contributed to a broader recalibration of social norms regarding race and xenophobia.

4 Twitter Data

Social media platforms like Twitter serve as a real-time barometer of public sentiment and social preferences, providing high-frequency, large-scale data that complement traditional survey or crime-record statistics, which are often costly, coarse, and subject to reporting lags (Tumasjan et al., 2010). Unlike official hate-crime records, which can take weeks or months to be compiled and published, Twitter data are available almost immediately, allowing researchers to detect shifts in social attitudes with minimal delay (Arcila Calderón et al., 2024).

For example, Müller and Schwarz (2021) show that spikes in anti-refugee social-media posts predict corresponding increases in offline hate crimes, indicating that online turmoil often precedes physical-world violence. Because formal hate-crime data can lag weeks behind real events and under-report many incidents, tracking hate-speech dynamics on Twitter provides a timely and granular proxy for latent social hostility. In the sections that follow, we describe how we harvested, processed, and classified tweets from Athletic Club fan communities and matched control clubs, and how we constructed monthly hate-speech intensity indices to evaluate the influence of Iñaki Williams's presence on online negativity among Athletic Club supporters.

4.1 Scraping

To assemble a comprehensive database of tweets from different fan bases , we implemented a multi-stage scraping protocol designed to balance coverage and compliance with Twitter's rate limits. First, we created ten separate Twitter profiles, each registered via a different Google account, in order to distribute request load and reduce the risk of blocking.

Next, for each of the four clubs (Athletic Club as the treated unit and Real Sociedad, Real Betis and Rayo Vallecano as controls), we performed a day-by-day advanced search over the period from January 1, 2014 to December 31, 2016. On each calendar day, we recorded the URLs of the first five tweets posted by the official club account. We then appended the path segment "/retweets" to each URL and scraped the first ten unique users who had retweeted that post. We treat retweeters as a proxy for genuine fans. After removing duplicate user IDs, we randomly sampled 2,000 accounts per club to form our supporters panels.

For each sampled account, we conducted a month-by-month advanced search across the same three-year window, retrieving the first 20 tweets per account per month. This procedure yielded on average approximately 615,800 tweets per club over the three-year period. To comply with Twitter's enforcement threshold of 50 link-scraping requests before triggering a 15-minute block, our script monitored request counts and automatically paused for 15 minutes when the limit was reached. We rotated through the ten Twitter profiles to maintain continuous scraping without interruption. This approach allowed us not only to respect the rate limit imposed but also to manage the CPU-intensive task effectively and leverage the frontier of available computational power.

In order to bypass Twitter's recommendation algorithms and ensure that results reflected strict reverse-chronological order, we installed a Chrome extension[2] that restored the classic "Old Twitter" interface for the entirety of the scraping process.

4.2 Data Description

Our raw dataset comprises 2,463,153 individual tweets scraped from the fan panels of Athletic Club, Real Sociedad, Real Betis and Rayo Vallecano between January 1, 2014 and December 31, 2016. On average, each club contributed roughly 615,800 tweets, each of which retains the full text, a precise UTC timestamp and basic engagement metrics (likes and retweets). To illustrate the nature of the content, consider this fan comment from August 21, 2016:

(@Pablo_Egea)

> *"Lamentables los cánticos hacia Iñaki Williams! ¿Habrá alguna reacción @Tebasjavier?"*
>
> *"The [racists] chants directed at Iñaki Williams are deplorable! Will there be any reaction, @Tebasjavier?"*

This example shows direct references to Williams and hints at potential hostility within the fan base. Subsequent sections will describe how we cleaned, classified, aggregated and smoothed these raw tweets into monthly club-level indices.

[2] Old Twitter Settings Chrome Extension: https://chromewebstore.google.com/detail/old-twitter-layout-2025/jgejdcdoeeabklepnkdbglgccjpdgpmf?pli=1

4.3 Tweet Classification

This section describes how raw tweets were transformed into numerical indicators of hate-speech intensity. We first standardized each message via a dedicated preprocessing routine. We then applied a state-of-the-art Spanish hate-speech model *pysentimiento* [3] to generate continuous scores. Finally, those scores were aggregated into monthly club-level series in preparation for causal analysis.

4.3.1 Preprocessing

Every tweet was passed through the *pysentimiento.preprocessing.preprocess_tweet* function, which was specifically designed to prepare input for the *robertuito-hatespeech* classifier. This routine replaces user mentions and URLs with the tokens @user and @url, collapses sequences of repeated characters to a maximum of three, normalizes laughter patterns (for example "jajajajajaja" becomes "jaja"), and converts emojis into descriptive text labels (for example an angry face emoji becomes :angry_face:). In addition, hashtags are split into individual words by a heuristic segmentation algorithm and all text is lowercased. These steps produce a cleaner, more consistent input for the downstream classifier and remove noise that can impair transformer-based language models.

4.3.2 Classification

After cleaning, each tweet was fed into the *robertuito-hate-speech* analyzer from the *pysentimiento* toolkit. This model reads the entire tweet and assigns three scores: a hate speech probability, which measures explicit derogatory language or slurs directed at a protected group; an aggressive language probability, which captures profanity, threats or a generally hostile tone; and a targeted insults probability, which reflects direct attacks aimed at a specific individual or group. These scores range from zero (very unlikely) to one (very likely). By keeping them as continuous values, we capture how strong or mild the language is instead of forcing a yes-no decision.

Under the hood, RoBERTuito first learned general language patterns by pretraining on a massive corpus of over 500 million Spanish-language tweets, exposing it to slang, emojis, abbreviations and a wide variety of online expressions . It then underwent a second stage of fine-tuning on the SemEval 2019 HatEval dataset, which taught it to distinguish hate speech, aggression and targeting. When processing a tweet, the model splits the text into smaller subword units to handle unknown or misspelled words, uses multiple transformer layers to capture contextual relationships among all tokens, and finally produces three probability scores. Because it was fine-tuned on a benchmark hate-speech corpus, it can accurately recognize messages that combine hateful and aggressive content.

Once each tweet had its three probability scores, we merged these outputs with the tweet metadata (text, timestamp, likes and retweets) and grouped by club and calendar month. We computed the average probability for each category, yielding 36 monthly observations per club.

4.4 Results

In this section we examine the *pysentimiento* hostility measures, hate speech and aggression[4]. For each measure, we first visualize its monthly time series. We then estimate a synthetic control to quantify the causal effect of Iñaki Williams's debut, reporting pre-treatment fit statistics, donor weights, average treatment effects and placebo test results.

[3]For further information consult https://huggingface.co/pysentimiento/robertuito-hate-speech
[4]For Targeted Speech Probability see Appendix section 7.1

4.4.1 Hateful Speech Probability

Figure 6 plots the monthly probability of hate speech for Athletic Club and its three control clubs from January 2014 through December 2016. The vertical red line marks January 2016, when Iñaki Williams embarked on his first multi-game goal streak.

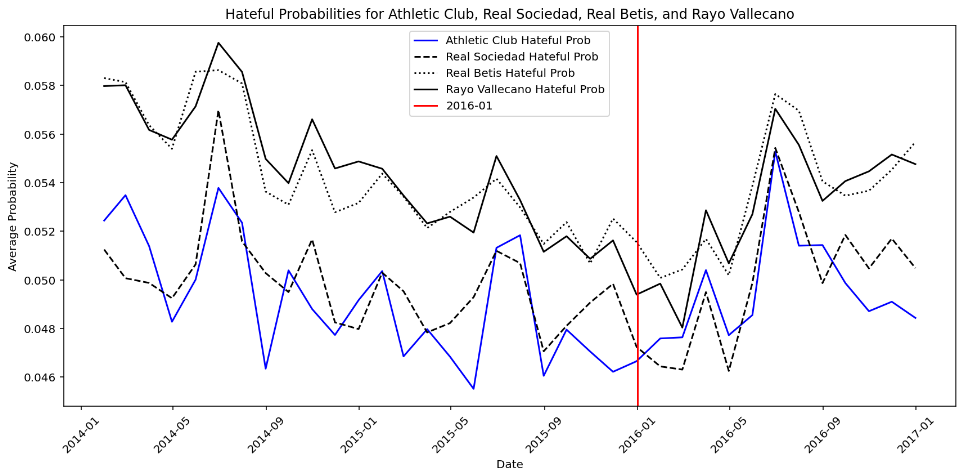

Figure 6: Evolution of Hate Speech Probabilities. *Source: Author's elaboration based on collected tweets.*

In the months before January 2016, Athletic Club's hate speech probability drifts downward, from roughly 0.052 in early 2014 to about 0.0468 by December 2015. Real Sociedad follows a very similar path, declining from approximately 0.051 to 0.0465 over the same span. Real Betis's probability falls from near 0.053 down to about 0.0492, and Rayo Vallecano moves from around 0.052 to 0.0490.

Around April 2016, all four series registered an uptick in hostile language. Athletic Club peaks at approximately 0.053 in late summer 2016 before moderating back to near 0.049 by year-end. Real Sociedad and Real Betis climb even higher, reaching roughly 0.052 and 0.056 respectively, while Rayo Vallecano rises to about 0.055.

Although each fan base experiences a rebound in hate speech following Williams's breakthrough, Athletic Club's increase is more muted compared with the controls. This suggests that while hostile language surged across all communities in early 2016, the magnitude of that surge was relatively smaller among Athletic supporters.

Synthetic Control

Figure 7 compares Athletic Club's monthly hate speech probability with that of its synthetic control from January 2014 through December 2016. In the 24 months before January 2016, the fitted synthetic series closely tracks Athletic's observed values, yielding a root mean squared prediction error of 0.00221. Donor weights assign 85.90% to Real Sociedad and 14.10% percent to Rayo Vallecano, with Real Betis excluded.

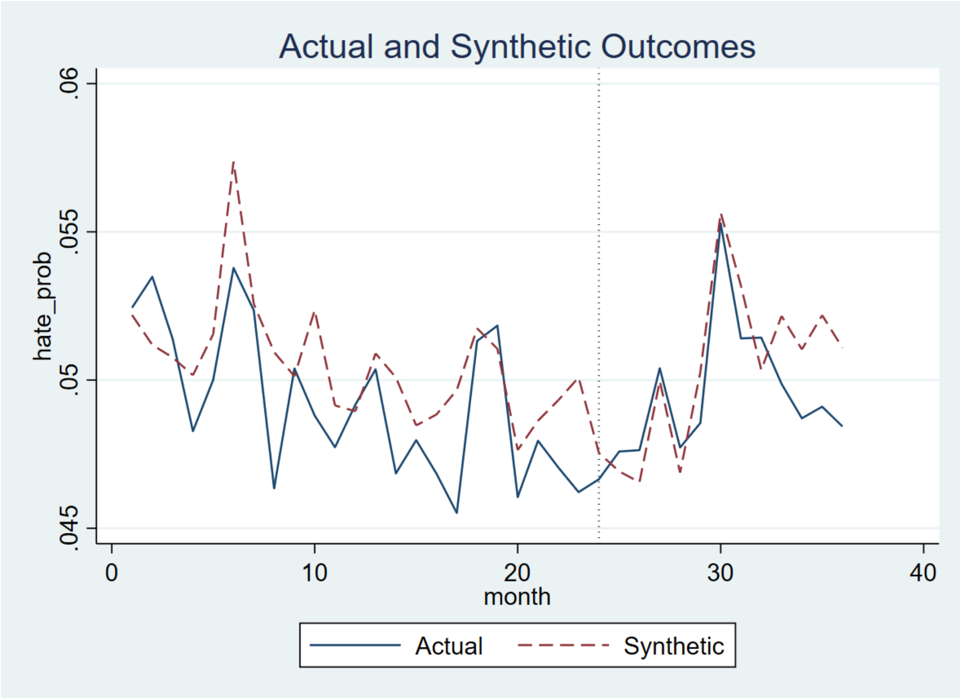

Figure 7: Synthetic Control for Hate Speech Probability. *Source: Author's elaboration based on collected tweets.*

In the year after Williams's breakthrough, the gap between the actual and counterfactual series never exceeds 0.0031 in absolute value. The average treatment effect is -0.0008, indicating a slight reduction in hate speech probability relative to the synthetic control. This reduction represents a decrease in −1.7%. However, by the end of the observation period, the actual series falls approximately 5.3% below its synthetic counterpart.

Although this synthetic-control analysis does not reveal a significant shift in hate speech probability, the actual series trends just below its counterfactual in the final several months, suggesting a modest late-season decline in hostile language among Athletic supporters.

4.4.2 Aggressive Speech Probability

As defined by Perez, et. al. (2021), aggressive speech probability is the likelihood of a hateful comment containing aggressive language. Figure 8 plots the monthly probability of aggressive speech among Athletic Club supporters and its three principal control clubs, Real Sociedad, Real Betis and Rayo Vallecano from January 2014 through December 2016. The vertical red line again denotes January 2016, the point at which Iñaki Williams began his sustained first-team appearances.

In the twenty four months preceding January 2016, Athletics Club's aggressive speech probability declined from approximately 0.032 to 0.028. The other three clubs follow a similar downward trend.

Once Williams's scoring run begins, there is a clear uptick in aggressive language across the board. Real Betis and Rayo Vallecano show the sharpest rebounds, each climbing roughly 0.002 points by mid-2016, while Real Sociedad's increase is slightly smaller. Athletic Club's rise is the most muted, peaking at about 0.0295 in September 2016 before drifting back toward 0.0275.

Compared with the hate-speech series, these aggressive-speech trends display less variation in both decline and rebound. That narrower range suggests that, although Athletic fans did become more aggressive

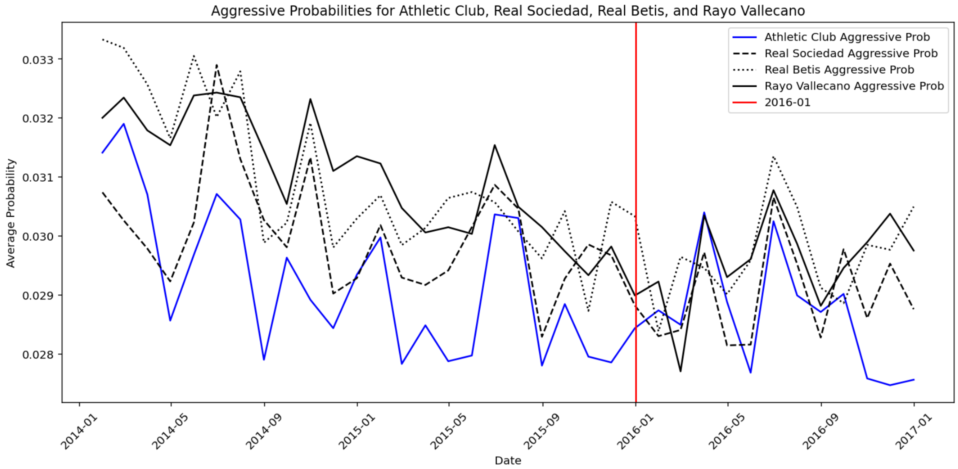

Figure 8: Evolution of Aggressive Speech Probabilities. *Source: Author's elaboration based on collected tweets.*

on Twitter after January 2016, their change was relatively modest compared not only to their own hate-speech response but also to the sharper swings seen in the other three fan bases.

Synthetic Control

The synthetic-control analysis for the aggressive language probability (treatment in January 2016, period 25) shows a tight pre-treatment fit, with a root mean squared prediction error of 0.00143. The synthetic control assigns a 60% weight to Real Sociedad and the remaining 40% to Real Betis, leaving Rayo Vallecano out.

In the post-treatment period (months 25–36), Athletic Club's average treatment effect is –0.0006, reflecting a a slight reduction in aggressive-speech probability relative to the counterfactual. This reduction represents a decrease in −2.2%. Figure 9 illustrates that the actual and synthetic trends track closely throughout the post-treatment period and differ around month 35.

Even though the actual treatment effects found in the synthetic control method are relatively small, we can observe an actual gap around month 35 of around 0,02. That represents a 6,4% decline relative to the counterfactual synthetic unit. Notably, this divergence becomes most pronounced in September 2016, indicating a somewhat delayed manifestation of the "Williams effect".

5 Conclusion

This analysis brings together "hard" crime statistics and "soft" online discourse to advance our understanding of how high-profile minority figures can reshape social norms. By situating Iñaki Williams alongside cases such as Mohamed Salah in Liverpool and Giannis Antetokounmpo in Milwaukee, we have shown that sustained visibility of a Black Basque athlete can produce measurable reductions not only in officially recorded hate crimes but also in the everyday hostility expressed by football fans online.

Our Difference-in-Differences estimates revealed a one-third drop in monthly hate-crime rates in the Basque Country after January 2016, consistent with the 30–40 percent declines documented following Salah's debut in Merseyside, while the Synthetic Control Method uncovered a two-thirds reduction in annual hate-crime counts relative to a weighted Spanish counterfactual. These "hard" outcomes resonate with the Parasocial Contact Hypothesis, suggesting that even indirect exposure can undercut the grievances and rationalizations that King and Sutton (2013) identify as precursors to bias-motivated

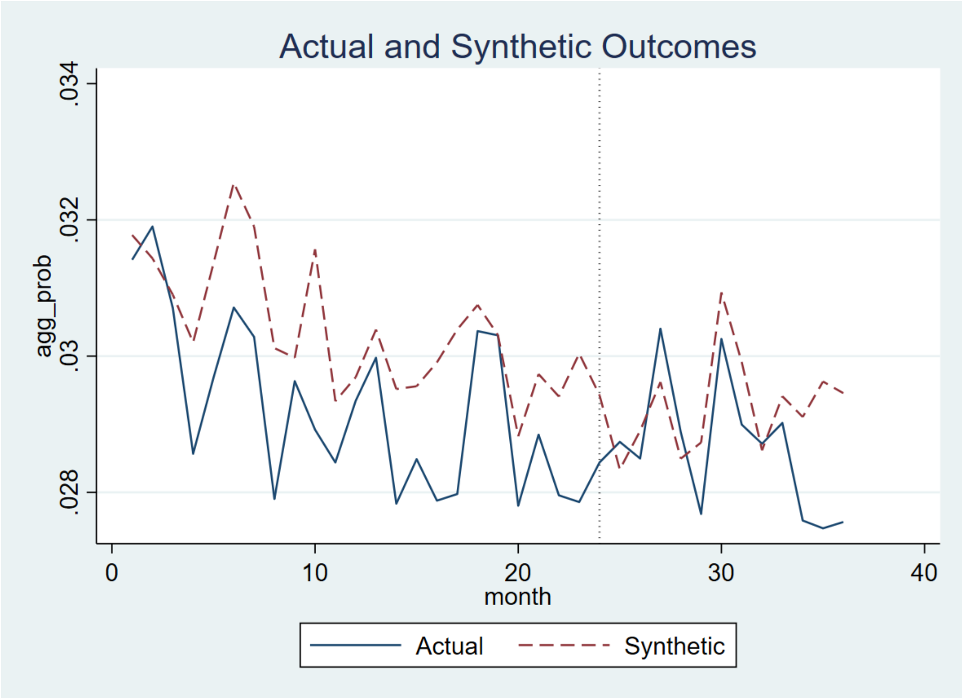

Figure 9: Synthetic Control for Aggressive Speech Probability. *Source: Author's elaboration based on collected tweets.*

violence.

Turning to online speech, our Twitter dataset, over 2.4 million tweets classified with the RoBERTuito model, yielded declines in hate-speech probability (–5.3%), aggression (–6.4%) and targeted insults[5] (–4.9%) against the synthetic counterfactual units. The timing and magnitude of these shifts mirror the "soft" attitude changes observed among Antetokounmpo fans, reinforcing that mediated representation can diffuse new social norms.

Taken together, these findings suggest a cascading effect: Williams's role elevated the symbolic salience of racial diversity in Basque identity, which then rippled through fan conversation and ultimately into fewer hate crimes.

Looking ahead, combining survey-based measures of prejudice with longitudinal social-media experiments could reveal how Williams' influence works, whether by easing intergroup anxiety, strengthening in-group norms, or changing people's sense of what is morally acceptable. Extending this framework to other minority figures in different cultural contexts would also test the generality of our results. For policymakers and club leadership alike, the implication is clear: actively promoting diverse role models may yield dividends in both public sentiment and community safety.

[5]See Appendix 7.1 for further analysis

6 Bibliography

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review, 93*(1), 113–132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*(490), 493–505.

Allport, G. W. (1954). *The Nature of Prejudice* (25th ed.). Cambridge, MA: Perseus Books Publishing, L.L.C.

Alrababa'h, A., Marble, W., Mousa, S., & Siegel, A. A. (2021). Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on Islamophobic behaviors and attitudes. *American Political Science Review, 115*(4), 1111–1128.

Amino Martín, A. M. L., Soto, A. B., & Bautista, F. J. (2023). Study and analysis of hate crimes in Spain (2013–2020). *Revista de Cultura de Paz, 7*, 141–161.

Aquino, M. (2017). A panther among lions: Iñaki Williams, race and Basque identity at Athletic Club de Bilbao. *Studia Iberica et Americana: Journal of Iberian and Latin American Literary and Cultural Studies*, (4), 119–136.

Arcila Calderón, C. A., Sánchez Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., ... & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: The role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities and Social Sciences Communications, 11*, Article 1369.

Ashenfelter, O. C., & Card, D. (1984). Using the longitudinal structure of earnings to estimate the effect of training programs.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics, 119*(1), 249–275.

Bosley, M. (2023). *The Giannis Effect: How Celebrities Impact Prejudice in Their Communities*. Economics Honors Projects, 118.

Carrell, S. E., Hoekstra, M., & West, J. E. (2015). The impact of intergroup contact on racial attitudes and revealed preferences. *National Bureau of Economic Research* (No. w20940).

Dovidio, J. F., Love, A., Schellhaas, F. M., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations, 20*(5), 606–620.

Hopkins, D. J. (2010). Politicized places: Explaining where and when immigrants provoke local opposition. *American Political Science Review, 104*(1), 40–60.

King, R. D., & Sutton, G. M. (2013). High times for hate crimes: Explaining the temporal clustering of hate-motivated offending. *Criminology, 51*(4), 871–894.

Llopis-Goig, R. (2010). Racism and xenophobia in Spanish football: Facts, reactions and policies. *Physical Culture and Sport. Studies and Research, 47*(1), 35–43.

Martín, A. M. L., Soto, A. B., & Bautista, F. J. (2023). Study and analysis of Hate Crimes in Spain (2013–2020). *Revista de Cultura de Paz, 7*, 141–161.

Meleady, R., Seger, C. R., & Vermue, M. (2017). Examining the role of positive and negative intergroup contact and anti-immigrant prejudice in Brexit. *British Journal of Social Psychology, 56*(4), 799–808.

Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association, 19*(4), 2131–2167.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix.

Observatorio Español del Racismo y la Xenofobia (OBERAXE). (n.d.). Informes y estadísticas sobre delitos de odio. *Ministerio de Inclusión, Seguridad Social y Migraciones.* Retrieved from:

`https://www.inclusion.gob.es/oberaxe/es/publicaciones/documentos/`

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin, 36*(12), 1723–1738.

Park, S. Y. (2012). Mediated intergroup contact: Concept explication, synthesis, and application. *Mass Communication and Society, 15*(1), 136–159.

Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., & Martínez, M. V. (2021). pysentimiento: A python toolkit for opinion mining and social NLP tasks. *arXiv preprint arXiv:2106.09462.*

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*(1), 65–85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751.

Ramasubramanian, S. (2015). Using celebrity news stories to effectively reduce racial/ethnic prejudice. *Journal of Social Issues, 71*(1), 123–138.

Raphael, S., & Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *Journal of Law and Economics, 44*(1), 259–283.

Rosenzweig, L. R., & Zhou, Y. Y. (2021). Team and nation: Sports, nationalism, and attitudes toward refugees. *Comparative Political Studies, 54*(12), 2123–2154.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs, 72*(1), 92–115.

Schuurman, B., & Carthy, S. L. (2025). Citizens, extremists, terrorists: Comparing radicalized individuals with the general population. *Terrorism and Political Violence*, 1–20.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–185.

7 Appendix

Athletic Club team photo, January 2017. *Source: https://www.athletic-club.eus/equipos/athletic-club/2016-17/copa/partidos/fc-barcelona-vs-athletic-club/1_5487*

7.1 Targeted Speech Probability

Figure 10 plots the monthly probability of targeted hate speech among Athletic Club and the three control clubs from January 2014 though December 2016.

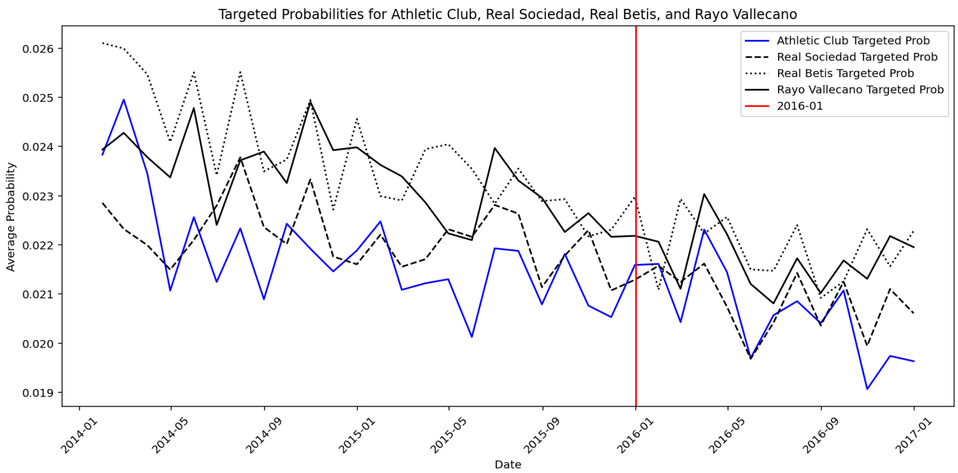

Figure 10: Evolution of Targeted Speech Probabilities. *Source: Author's elaboration based on collected tweets.*

The graph shows us that the targeted probability is much more stable than the aggressive and hateful probabilities. Athletic Club, after an initial drop from 0,024 stabilizes around the 0,022 level. Real Betis shows a similar trend and level while Real Sociedad and Real Betis start from a slightly higher point and show a downward trend.

After January 2016 we see a bit of a downwards trend in all 4 clubs. A clear change of trend can be seen around September 2016, with a significant drop in Athletic Clubs Targeted probability while the others remain at a similar level.

Synthetic Control

In this case the synthetic control algorithm assigns Real sociedad a 98% weight and Rayo Vallecano the remaining 2%, leaving Real Betis out.

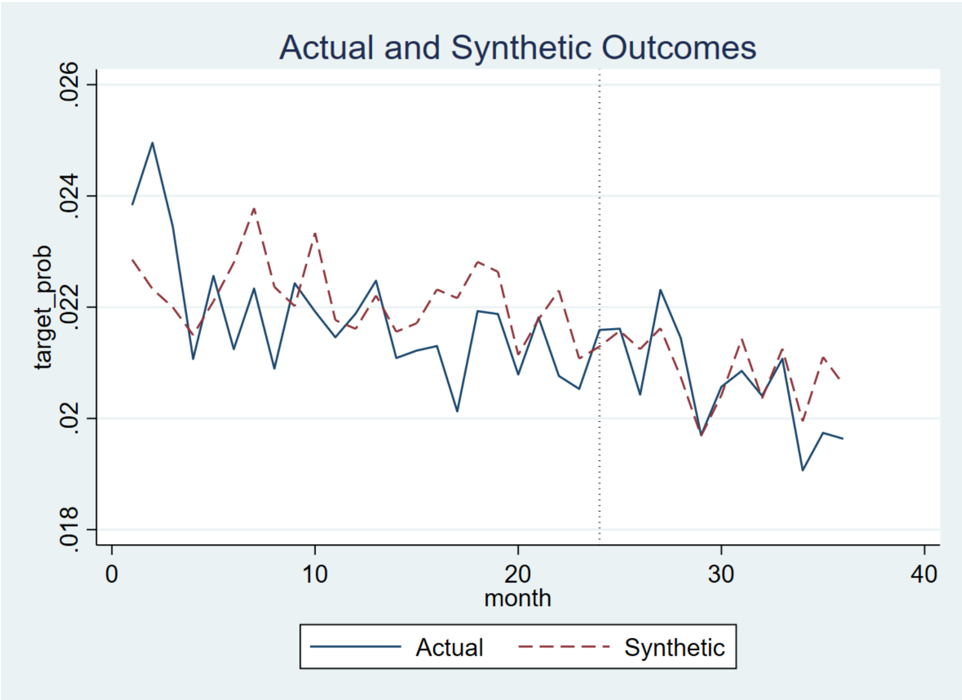

Figure 11: Synthetic Control for Targeted Speech Probability. *Source: Author's elaboration based on collected tweets.*

The average treatment effect for the post-treatment period is -0,0003, representing a decrease in -1.4%. As in previous analysis we can observe a significant gap between the actual and synthetic units towards the end of the series. Around September 2016 we observe a decrease in Athletic´s Club probability relative to the synthetic unit. This represents a reduction of approximately 4,9% relative to the synthetic counterpart.