**BICYCLIST SAFETY**

**Report Highlight**

Unravelling the Factors: Investigating the critical elements elevating the risk for bicyclists. A comprehensive analysis aimed at identifying key reasons behind the heightened probability of death or severe injury.

(Source : PennDOT Open Data)

Prepared By:

**GROUP-6**

**Bharat Modhwadiya**

**Neil Patel**

**Mufaddal Zulfiquar**

**Dhruv Kandhari**

**Rajeshwari Patidar**

# **Executive Summary**

This report presents a comprehensive statistical analysis dedicated to the prediction of incidents involving bicycle fatalities and severe injuries. Leveraging sophisticated analytical tools and methodologies, the study aims to not only quantify the occurrence of such incidents but also delve into the underlying factors that contribute to them. The overarching goal is to enhance our collective understanding of these incidents, ultimately facilitating the development of targeted interventions and safety measures. By examining PennDOT datasets and employing predictive modeling techniques, the report seeks to identify patterns, trends, and key variables associated with bicycle accidents. Advanced predictive modeling techniques will be applied to identify patterns, discern trends, and pinpoint key variables that significantly influence the occurrence and severity of these incidents.

The insights derived from this analysis are poised to serve as a valuable resource for policymakers, urban planners, and safety advocates. By offering a nuanced understanding of the dynamics at play, the report equips decision-makers with actionable intelligence to formulate and implement targeted strategies for accident prevention. Moreover, the findings will contribute to the ongoing dialogue on sustainable urban mobility by highlighting the importance of fostering a safe environment for cyclists—a critical component of modern, eco-friendly transportation systems. In an era where the promotion of cycling as a sustainable and healthy mode of transportation is gaining momentum, prioritizing bicycle safety is paramount. The analytical depth of this report aims not only to identify risk factors but also to empower PennDOT with the knowledge needed to proactively mitigate these risks.

# **Highlights of the Situation**

Bicycle deaths and injury can be caused by numerous factors. The aim would be to determine which factors can be controlled to prevent such causes. Some key highlights derived from data collected over 5 years starting from 2018:

1. Time Series:

A graph of a number of years

Description automatically generated with medium confidence

A graph of injury injuries

Description automatically generated

Observed from time series that 2022 compared to 2021 had less fatalities and severe injuries.

2. Geographic Patterns:

Observed here Philadelphia county has the highest amount (89) of severe injuries/fatalities.

A graph of injuries with numbers

Description automatically generated with medium confidence

An overview of clusters accounting only for deaths.

A map with green circles and a location pin

Description automatically generated

# **Statistical Insights**

**Urban vs. Rural Dynamics:**

In the diverse landscape of Pennsylvania, a stark contrast emerges as we examine the geographical distribution of cyclist fatalities. An overwhelming 94.41% of tragic incidents unfold in **urban** settings, underscoring the unique challenges faced by cyclists navigating city streets. However, the resilience of rural areas is not without exception, with 5.59% of fatalities occurring in these serene landscapes.

Further delving into demographics, our analysis reveals a nuanced gender dimension — approximately **72% of the victims are male**, and 26% are female.

**Unraveling Intersection Dynamics:**

The intersection, a nexus of pathways, proves to be a critical juncture in the cyclist safety landscape. A poignant realization surfaces as we delve into intersect type data — a staggering **51% of cyclist deaths unfold at Mid-Block locations**. Four-Way Intersections follow closely, contributing to 35% of fatalities. Addressing these specific intersection types presents an avenue to curtail the devastating toll, potentially reducing the chances of cyclist fatalities by an impactful 85%. This insight serves as a strategic guide for urban planners and traffic management authorities seeking to prioritize interventions for maximal impact.

\*As per **NTSB safety research report,** Bicycle crashes involving motor vehicles at midblock locations are more likely to result in fatal and serious injuries for the bicyclists. [[1]](#footnote-2)

A graph of different colored squares

Description automatically generated

**Stop Signs as Lifesavers**:

Amidst the bustling intersections and quiet streets, a beacon of safety emerges in the form of stop signs at signals. A compelling revelation unfolds as we observe an 84% reduction in the likelihood of severe injuries or death for cyclists at intersections equipped with stop signs. This finding serves as a clarion call for urban planners and policymakers to accentuate the role of these seemingly simple traffic control measures in safeguarding vulnerable road users.

**Signalized Intersections and the 30% Conundrum**:

At the heart of intersection safety lies the presence or absence of signals. A sobering revelation encapsulates our findings - 30% of cyclist fatalities transpire at signalized intersections. The converse of this statistic is equally compelling, with an alarming **70% of deaths occurring in the absence of signalization**. This calls for a nuanced approach to infrastructure planning, shedding light on the critical role signals play in averting cyclist fatalities.

\*As per **National Association of City Transportation Officials (NACTO) (2019) Report,** Page 4 states:  "Unsignalized intersections are particularly dangerous for bicyclists, accounting for 65% of bicyclist fatalities in all crashes.

Page 15 mentions: "The absence of traffic signals can increase the risk of conflicts and crashes for all road users, including bicyclists.[[2]](#footnote-3)

A graph of a number of people

Description automatically generated with medium confidence

(Crash at Signalized Intersection: 0 = No , 1= Yes)

**The absence of stop signs and signals emerges as a crucible for cyclist fatalities for Mid-Block and Four-way Intersection.** The confluence of these factors paints a vivid picture of high-risk areas, beckoning for immediate attention and strategic interventions. This synthesis of insights lays the groundwork for a focused and impactful roadmap to enhance cyclist safety across Pennsylvania.

**Additional Insights:**

In our pursuit of understanding factors contributing to cyclist fatalities, a comprehensive exploration dispels common misconceptions. **Speeding, alcohol and drug involvement, mobile phone usage, tailgating, aggressive driving, and illumination, while important facets of road safety, do not emerge as primary catalysts for severe injuries or death in the cyclist demographic**. This nuanced understanding allows us to refine our focus, directing resources towards the most impactful interventions tailored to the unique challenges faced by cyclists on Pennsylvania's roads.

A graph showing a graph of a bicycle crash

Description automatically generated with medium confidence

Illumination- revealing that the majority of severe crashes and fatalities unfold during the day-evening transitional period, particularly between 4 to 6 pm, underscoring the imperative for enhanced visibility and safety measures during these critical hours.

**Data Pre-Processing:**

The data is captured statewide from years 2018-2022 and merged. Once merged:

1. Filter Bicycle variable > 0 to see only bicycle data
2. Remove duplicates
3. Drop variables with Zero variance for both numerical and categorical. Including categorical with high cardinality.
4. Identify which variables have 70% or more Null values. In which case the feature was dropped.
5. Identify which variables are highly correlated to avoid multicollinearity
6. Implement VIF (Variable Inflation Factor) to test for multicollinearity, if VIF value is above 10 the feature is dropped.
7. Missing Value Imputation for these values:

A black text on a white background

Description automatically generated

Numerical values used median and mode was used for categorical values.

1. “Actual\_Dispatch” time for police to arrive on scene was created by taking Arrival\_TM and subtracting it from DISPATCH TIM
2. Create feature called Severity\_Binary by filtering out rows in MAX\_SEVERITY\_LEVEL with 0 or 9. Convert value of 1 (Fatal), and 2 (Severe Injury) to value of 1 showing fatal/serious injury. The other values will be converted to 0.

0--- >32428, 1---> 4593

A graph of data distribution

Description automatically generated with medium confidence

# **ML MODELS**

## Logistic Regression:

A binomial logistic regression is created to predict which features heavily influence fatality and severe injury. Numerous models were created in order to determine probability of a fatality/severe injury occurring. In order to address the imbalance issue, class weight parameter was added. Two models were created in order to balance precision and recall and another to optimize precision .

**Model 1: Balanced Precision and Recall**

A number of numbers in a row

Description automatically generated with medium confidence

Model 1 is optimized to have a balanced precision and recall metric. While maintaining an accuracy of 80% for prediction.

Minimizing Consequences of False Positives and False Negatives:

* False Positives (FP): Predicting a bicycle injury when there isn't one. By optimizing precision, you are focusing on minimizing the instances where the model incorrectly predicts a positive outcome when it should have been negative (FP).
* False Negatives (FN): Failing to predict a bicycle injury when there is one. By optimizing recall, you are focusing on minimizing the instances where the model fails to predict a positive outcome when it should have (FN).
* Balancing precision and recall helps in minimizing both types of errors. False positives may lead to unnecessary interventions or resources being deployed, while false negatives may result in missed opportunities for preventing or addressing actual injuries.

**Model 2: Precision Optimization**

A screenshot of a number

Description automatically generated

In the second logistic model precision is the metric optimized.

Severity of Consequences:

* Predicting a bicycle injury when there isn't one might lead to unnecessary interventions, but the consequences may not be as severe as missing an actual injury.
* False positives may result in inconvenience or additional precautionary measures, but false negatives (missing an actual injury) could have more severe consequences, especially if the injury is fatal or requires immediate medical attention.

## Random Forest:

A random forest model provided higher accuracy by 0.01 compared to model 2 and 0.09 compared to model 1.

A screenshot of a graph

Description automatically generated

Parameters can be adjusted to heavily favour precision or recall. If the consequences of missing a bicycle accident are severe then the model needs to show an increase in recall, even at the cost of lower precision. On the other hand, if the consequences of a false alarm (predicting an accident when there isn't one) needs to be prioritized then precision would need to be optimized.

## Results

From the logistical regression model these are the top 10 features provide insights into the likelihood of a bicycle-related incident based on the odds ratios. Model shows being a Male represents a higher chance of death/injury which reflects the finding from NHTSA (National Highway Traffic Safety Administration)[[3]](#footnote-4).

A screenshot of a computer

Description automatically generated

Ex: A bicycle-related incident increases by approximately 29.12% for each one-unit increase in the ILLUMINATION feature.

Below are the top 10 features that influenced the random forest model. Urban and County features seem to have the most impact on the model predictive power.

A screenshot of a phone

Description automatically generated

# **Limitations:**

1. Data Quality and Completeness:

- The accuracy of the analysis heavily relies on the quality and completeness of the PennDOT datasets. Incomplete or inaccurate data may introduce biases and limit the generalizability of the findings.

2. Imbalanced Data:

- The dataset may suffer from class imbalance, where the occurrences of severe injuries or fatalities are significantly lower than non-severe incidents. This imbalance can affect the performance of predictive models and bias the results.

3. Temporal Scope:

- The study's temporal scope is limited to the years 2018-2022. Changes in infrastructure, policies, or urban development beyond this period may not be fully captured in the analysis, potentially impacting the generalizability of recommendations.

4. Model Complexity:

- The predictive models, while informative, are simplifications of real-world complexities. Factors not considered in the models may influence bicycle-related incidents, and the models' predictions should be interpreted with awareness of these limitations.

# **Conclusions**

In conclusion, this report provides valuable insights into the dynamics of bicycle-related incidents in Pennsylvania, emphasizing the significance of urban-rural disparities, intersection types, and the critical role of traffic control measures. The identification of high-risk areas, such as those lacking stop signs and signals, serves as a foundation for targeted interventions.

The predictive models, though powerful, are not without limitations. Acknowledging the potential biases introduced by data quality, variable selection, and imbalances is essential. Despite these limitations, the models contribute valuable perspectives on the factors influencing incident severity.

Moving forward, the synthesis of data-driven insights and predictive modeling underscores the importance of a holistic approach to cyclist safety. Recommendations stemming from this analysis can guide policymakers, urban planners, and safety advocates in crafting interventions tailored to the unique challenges faced by cyclists. As the promotion of cycling gains momentum, prioritizing safety measures informed by robust analyses becomes imperative for fostering sustainable and secure urban mobility.

# **References:**

1. “Safety Research Report - Bicyclist Safety on US Roadways: Crash Risks and Countermeasures.” *Safety Research Report - Bicyclist Safety on US Roadways: Crash Risks and Countermeasures*, www.ntsb.gov/news/events/Pages/2019-DCA18SS002-BMG.aspx. Accessed 28 Nov. 2023.
2. Ink, Social. “Intersection Treatments.” *National Association of City Transportation Officials*, Nacto, 19 July 2019, nacto.org/publication/urban-bikeway-design-guide/intersection-treatments/.
3. “Bicycle Safety.” *NHTSA - Bicycle Safety* , NHTSA, www.nhtsa.gov/book/countermeasures/countermeasures-work/bicycle-safety. Accessed 28 Nov. 2023.

1. **Bicycle Safety on US Roadways: Crash Risks and Countermeasures** by the National Transportation Safety Board (NTSB) (2019).  [↑](#footnote-ref-2)
2. **The Impact of Intersection Design on Bicyclist Safety” by the National Association of City Transportation Officials (NACTO) (2019)** [↑](#footnote-ref-3)
3. 3. [Bicycle Safety - NHTSA](https://www.nhtsa.gov/book/countermeasures/countermeasures-work/bicycle-safety) [↑](#footnote-ref-4)