



# Halloween:

Deadliest day of the year  
for child pedestrians.

**Preventing child fatality by predicting accidents  
in the Washington DC Metropolitan area.**

# Project Summary



Pedestrian fatalities are reported to be 43% higher on Halloween night. DC Public Charter School Board is concerned with this, and would like to set preventative measures to keep children and parents safe. As a result, we built a prediction model to identify areas with high probability of accidents based on real-time weather data and real-time metro data.

## Project Goals

Provide the DC School Board educational resources on implementing real-time monitoring systems that can provide live updates and alerts to parents, guardians, and children about potential risks in proximity.

## Data Sets

### Historical:

- DC Weather Data, 2010-2023
- DC Crash Data, 2000-2020

### Real-time:

- DC Weather Data (API)
- DC Metro Data (API)

## Limitations

- Age and Location Variability - Risk levels may vary since not all areas experience the same safety concerns.
- Safe Route Planning is currently unavailable.

## Future Extensions

Create mobile applications that provide real-time safety information and alerts as well as suggest safer routes to parents and children during Halloween weekend.

## Methodology

1

Explore data.

2

Principle Component Analysis.

3

Machine Learning on historical data.

4

Machine Learning Prediction Modeling.

5

Check Model Performance.

6

Final Thoughts & Recommendations.

# The problem at hand



## Halloween is the deadliest day of the year for child pedestrians

Pedestrians ages 0-17 fatally struck by motor vehicles, total by the day of year for 2000 through 2020.

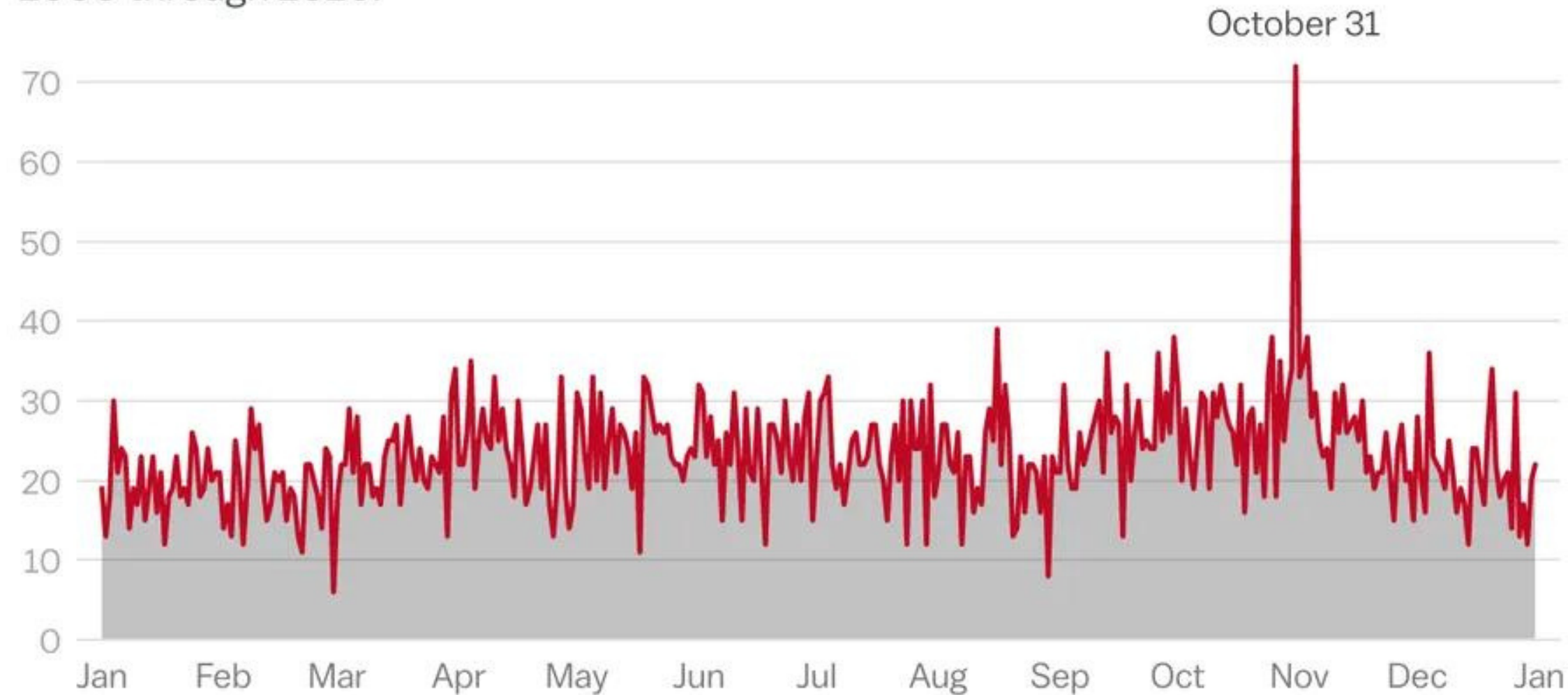


Chart: Muizz Akhtar and Youyou Zhou/Vox • Source: National Highway Traffic Safety Administration

**Vox**

[vox.com](https://www.vox.com)



Parents, guardians, and children will be out and about trick-or-treating during Halloweekend.

**DC Public Charter School Board** is concerned with the spike in child and parent pedestrian fatality the past few years, and would like to set preventative measures to keep children and parents safe.

**OUR  
GOAL**

We will build a prediction model to identify districts with high probability of accidents based on real-time weather data and real-time metro data.





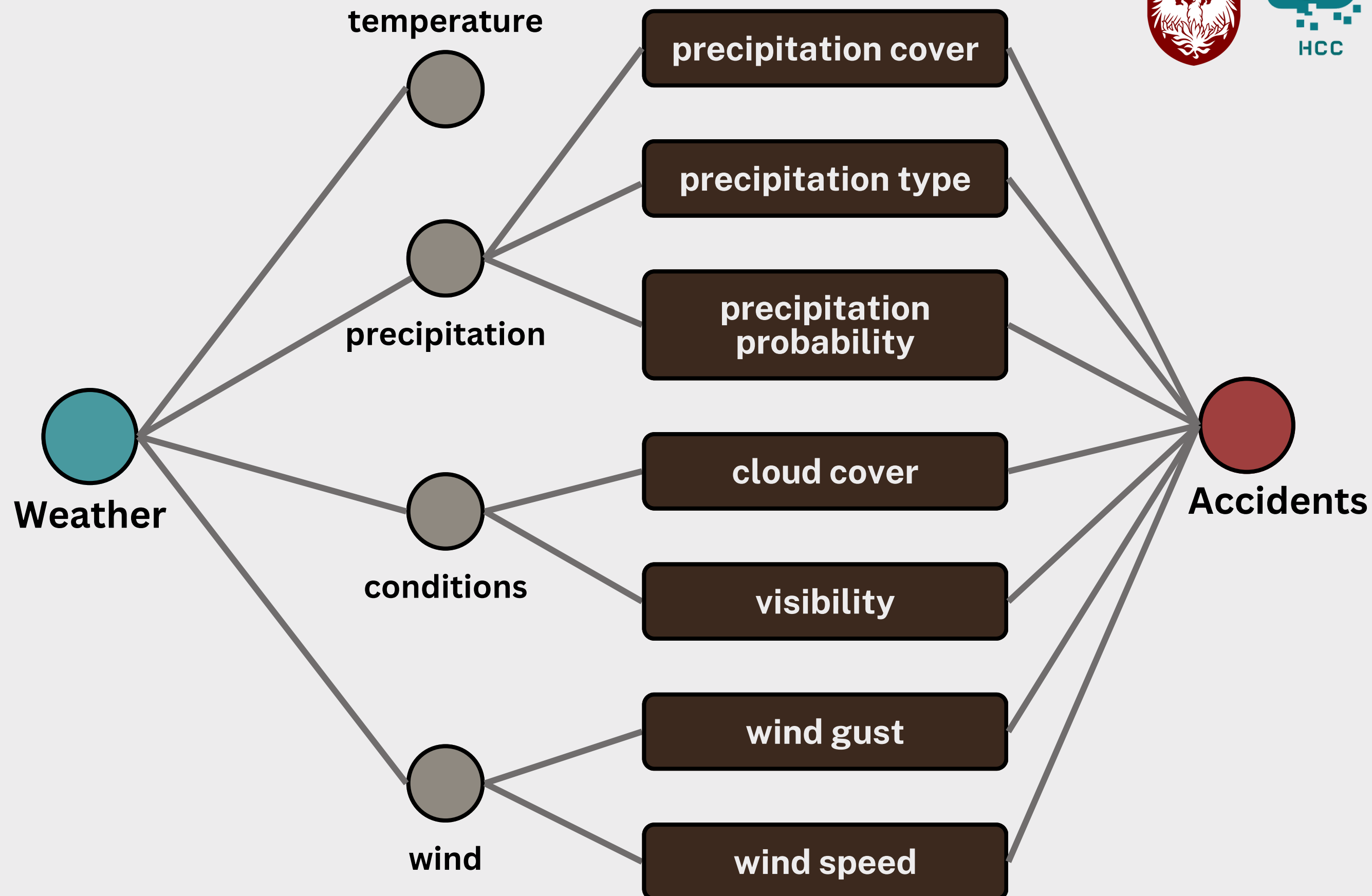
## HISTORICAL DATA SOURCES

DC's October's Weather  
Data from 2010 - 2023

[weatherapi.com](https://weatherapi.com)

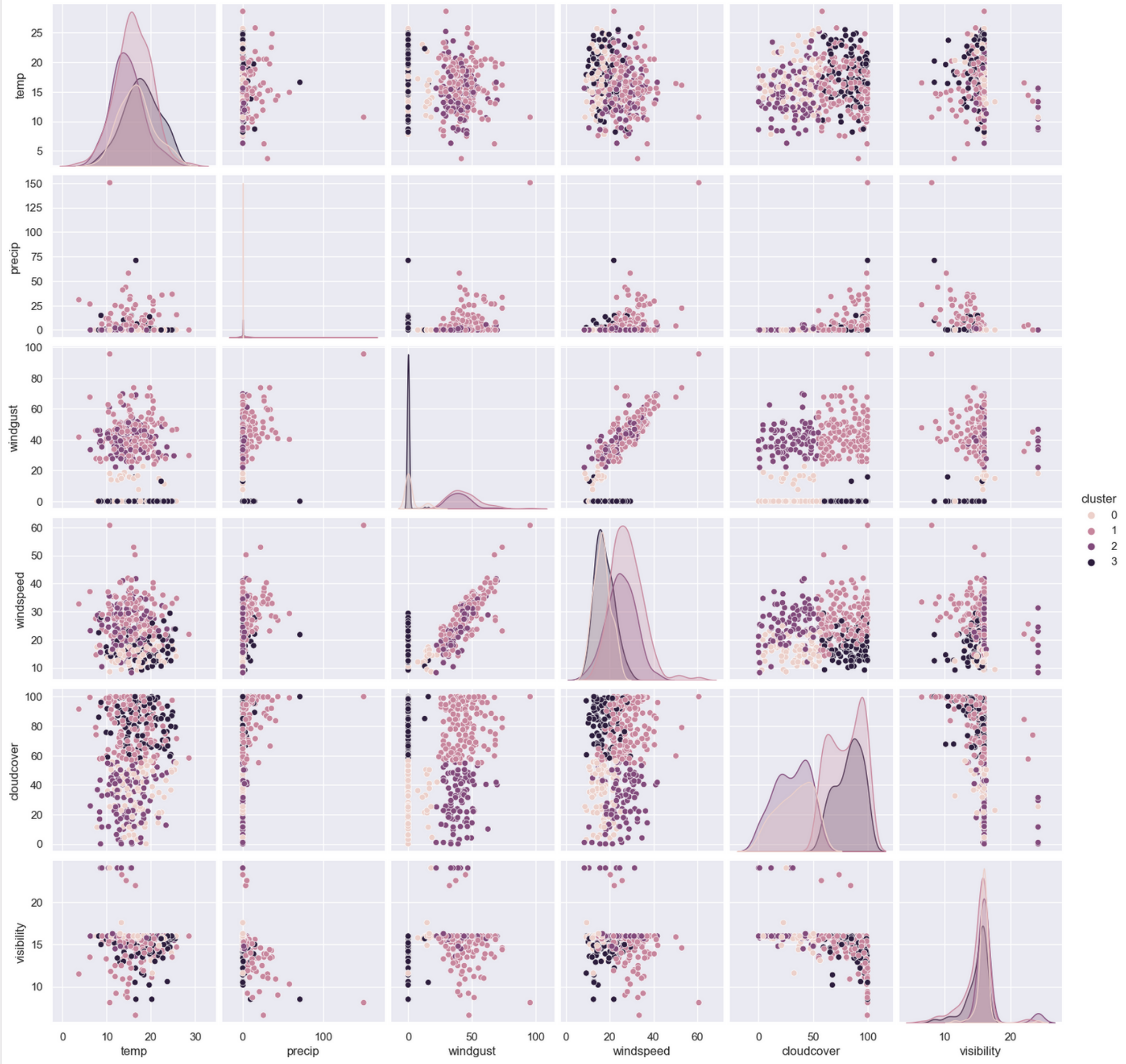
DC's October Crash Data  
from 2000 - 2020

[opendata.dc.gov](https://opendata.dc.gov)

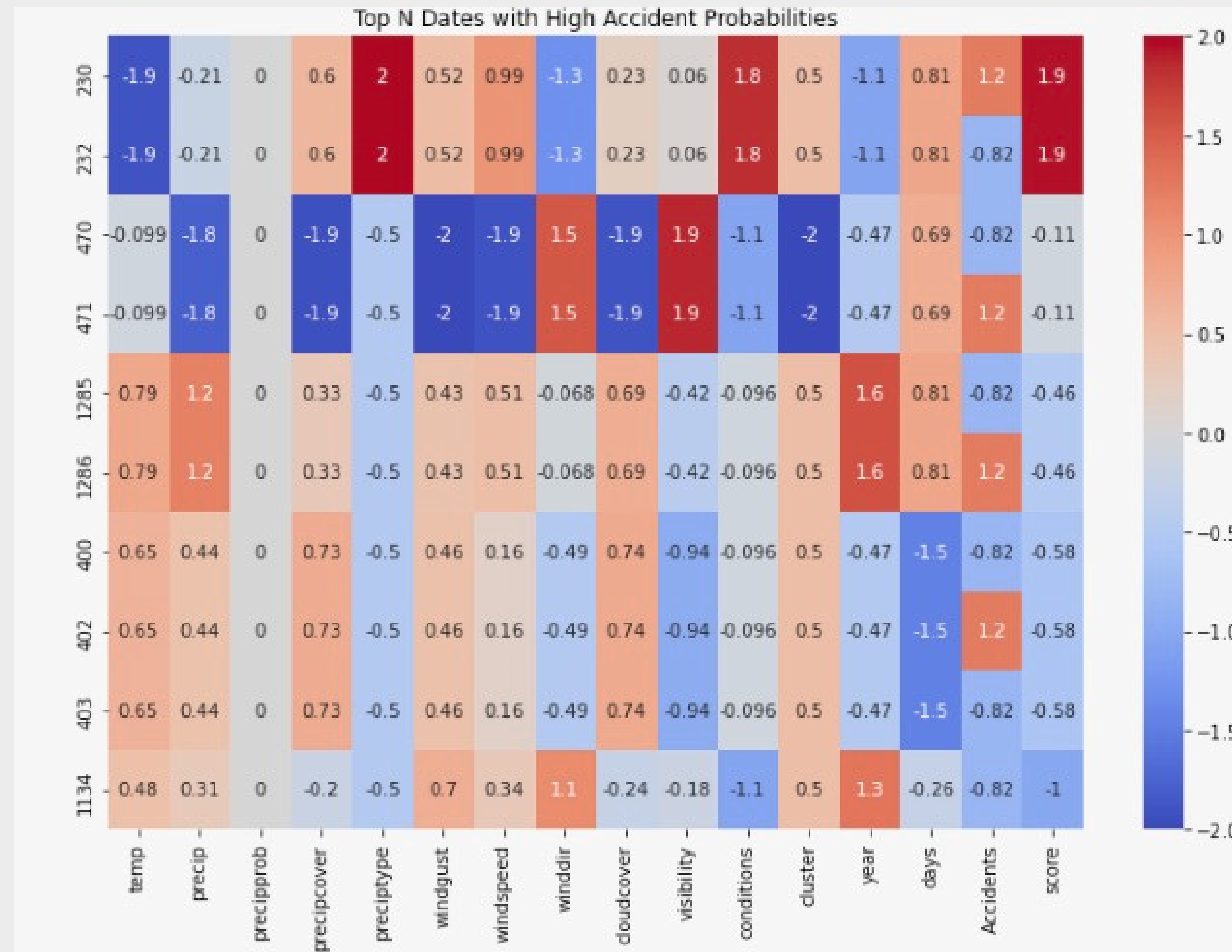




# Principle Component Analysis



# Machine Learning on Historical data



**Purpose:** Highest accident occurrence date

	Feature	Coefficient
0	temp	-0.161196
1	precip	0.158580
2	windspeed	0.176140
3	precipprob	0.000000
4	precipcover	0.148076
5	precipctype	0.015413
6	cloudcover	-0.453543
7	windgust	-0.172364
8	windspeed	0.176140
9	winddir	-0.262211
10	visibility	-0.201199
11	conditions	0.094825

Highest accident occurrence date: 230

**Index 230** - Saturday October 29, 2011

# ML Prediction Modeling



## DATA SOURCE

Weather data from  
weatherapi.com

Metro data from  
WMATA API

```
X = combined_df[['temp', 'windspeed', 'precip', 'cloudcover']]
y = combined_df['Accidents']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

RandomForestClassifier
RandomForestClassifier(random_state=42)

y_proba = clf.predict_proba(X_test)[: , 1]

y_proba = clf.predict_proba(X_test)

print("Unique classes in y_train:", y_train.unique())
print("Unique classes in y_test:", y_test.unique())

print("Shape of y_proba:", y_proba.shape)

known_classes = y_train.unique()
mask = y_test.isin(known_classes)
y_test_filtered = y_test[mask]
y_proba_filtered = y_proba[mask]

print("Unique classes in y_test_filtered:", y_test_filtered.unique())
print("Number of columns in y_proba_filtered:", y_proba_filtered.shape[1])
```





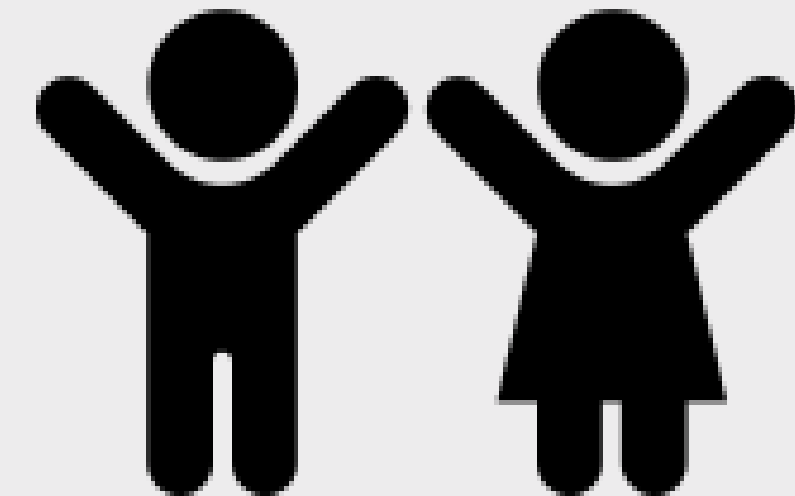
# Modeling Results

```
[42] def predict_proba(temperature, humidity, wind_speed, precipitation, model, scaler):  
    features = [[temperature, humidity, wind_speed, precipitation]]  
    features_scaled = scaler.transform(features)  
    return model.predict_proba(features_scaled)[0][1]  
  
probability = predict_proba(20, 65, 5, 0, clf, scaler)  
print(f'Probability of accident on October 31st, 2023 by historical data and Real time weather and transportation data: {probability:.4f}')
```

Probability of accident on October 31st, 2023 by historical data and Real time weather and transportation data: 0.2683

Probability of an accident happening on October 31st, 2023 based on historical data, real-time weather and transportation data is currently **26.83%**.

# Real-time Monitoring



# Recommendations



The **DC Public Charter School Board** should:

- send out installation instructions for real-time monitoring systems that can provide live updates and alerts to parents, guardians, and children about potential risks in proximity.
- Educate school faculty to raise awareness about Halloween safety:
  - costume visibility (bright, reflective clothing).
  - responsible adult supervision.
  - being attentive to streetlights and real-time alerts.

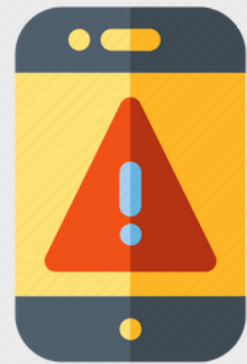
# Limitations



- Age and Location Variability:
  - Risk levels may vary by location and the age of children since not all areas experience the same safety concerns.
- Safe Route Planning is currently unavailable.
  - feature that suggests new routes to safe and well-lit paths for trick-or-treaters so pedestrians can avoid high-risk areas is currently unavailable.



# Proposed Future Extensions



## - Build a Pedestrian Safety App

Create mobile applications that provide real-time safety information and alerts as well as suggest safer routes to parents and children during Halloweekend.



**Thank you!**