IST 718:
Big Data Analytics

Traffic Accidents in Relation to Major Professional Sporting Events:

An Analysis of Four Southeastern Cities

Holden Herrell, Karima Hunter, and Vaibhav Mehta

TOPICS

Background

Obtaining the Data

Scrubbing the Data

Exploring the Data

Modeling the Data

iNterpreting the Data

Background

Business Question

Should Charlotte, NC approve the relocation of the Tampa Bay Rays to their city, giving them a third major professional sports teams/stadiums (MLB, NBA, NFL)?

Concern

Will adding a third team negatively impact traffic accidents?

Study:

Examine cities with three professional sports teams to understand if there are increased traffic accidents on gamedays.

Obtaining the Data

Traffic Accidents

Source: Kaggle

Description: 3M traffic accidents from 2016 to 2019. **Key Fields:**

- Accident Location
- Accident Date
- Severity
- Traffic Conditions
- Weather Conditions

Sporting Schedule

Source: Various*

Description: Contains sporting event data for MLB, NBA, NFL.

Key Fields:

- Game Location
- Game Date
- Attendance

*baseball-reference.com
*pro-football-reference.com
*basketball-reference.com

Scrubbing the Data

Traffic Accidents

- Trim data:
  - Records from 2017-2019
  - Relabel neighboring citys to city of interest
  - Include only records cities neighboring Atlanta, Dallas, Houston, and Miami
  - Remove records with NAs
  - Removed:

    ID, Description, County, Country, Airport_Code, TMC, Source, Street, Timezone, Astronomical_Twilight, Civil_Twilight, and Nautical_Twighlight
- Rename Sunrise_Sunset to Day_or_Night

Scrubbing the Data

Sporting Datasets

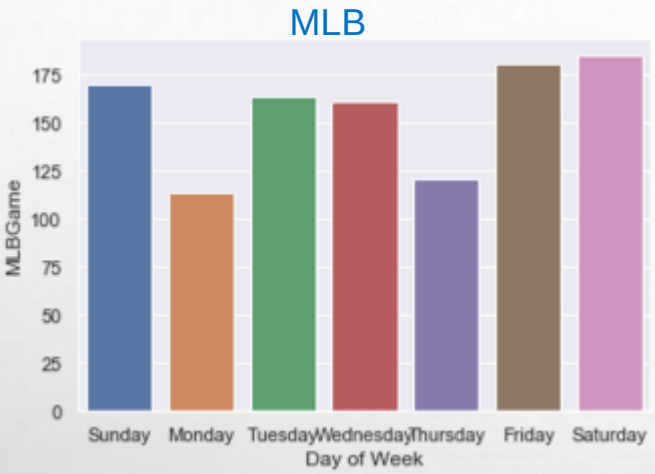
- Trim data:
  - Records from 2017-2019
  - Trim data to only concerned cities
- Isolate dataset to *GameDay, HomeTeam, and Attendance*
- Create Flag for MLBGame, NBAGame, and NFLGame

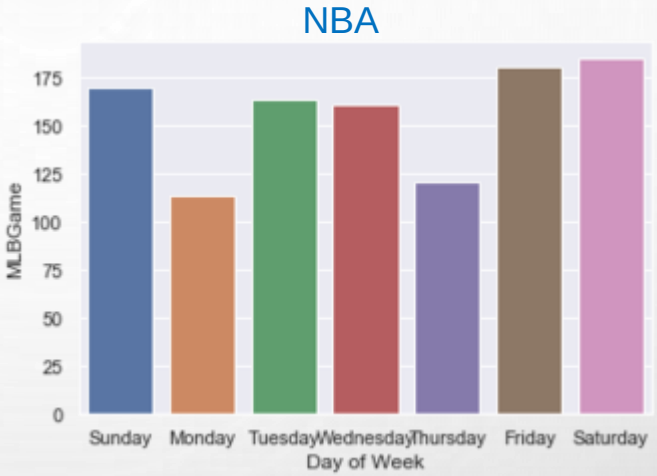
Combining Datasets:

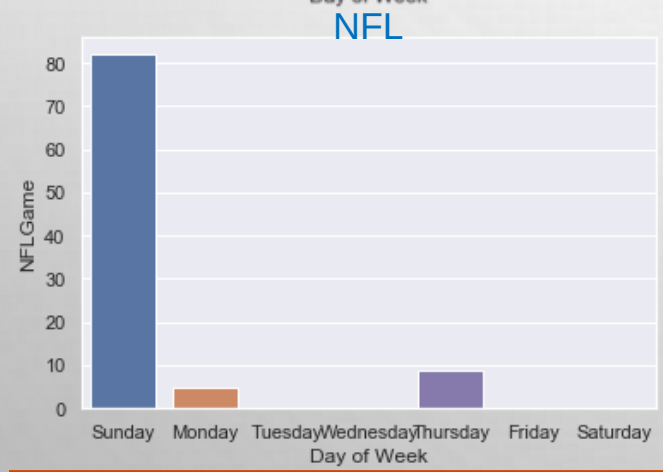
- Formatted all dates to ‘yyyy-mm-dd’
- Added Day of Week field based on date
- Created unique key combination of City, State, and Date to join sports data to accident data

Exploring the Data

MLB and NBA have most of their games played on the Weekdays



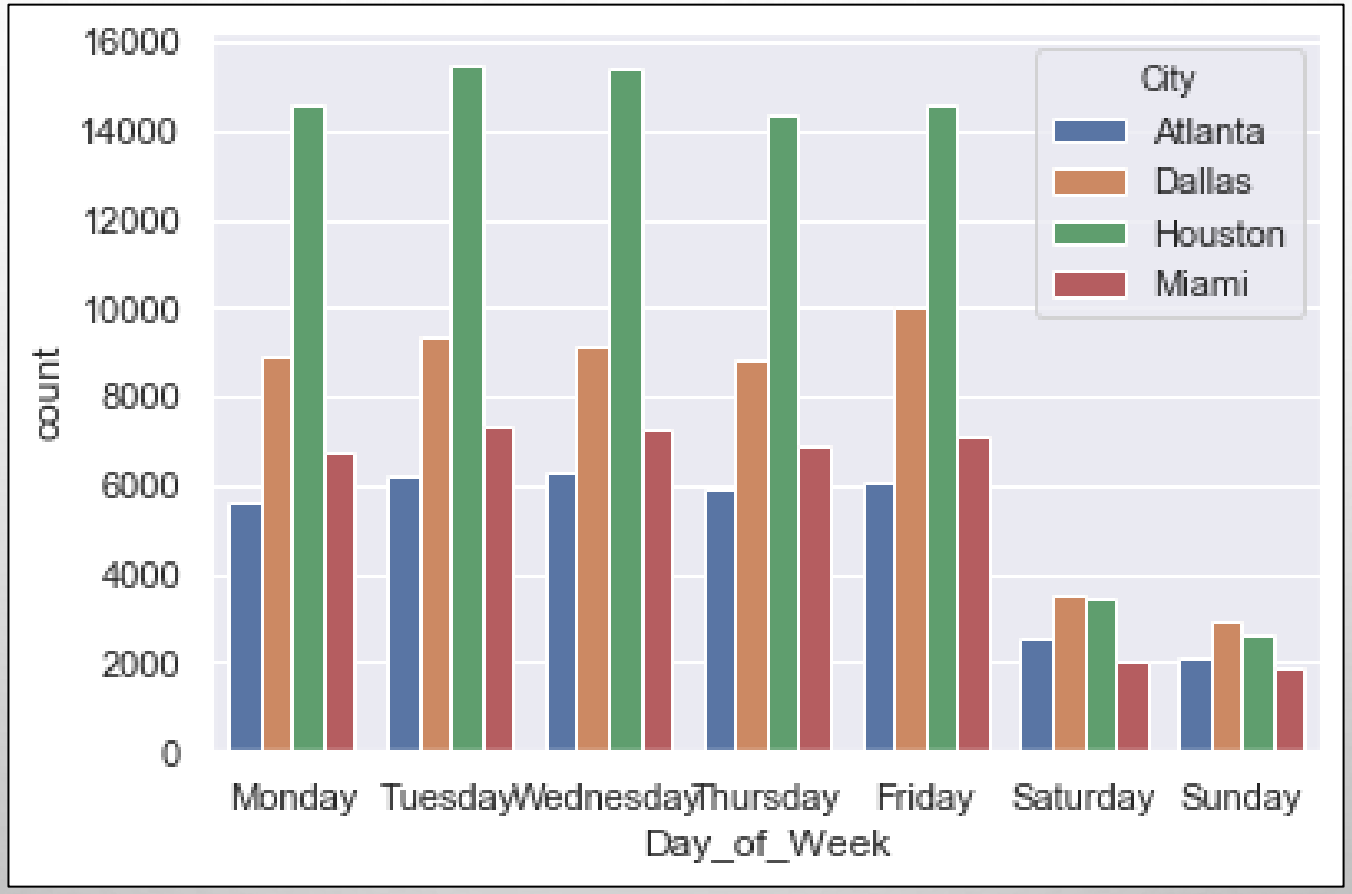




City	MLB	NBA	NFL
Atlanta	Braves	Hawks	Falcons
Dallas	Rangers	Mavericks	Cowboys
Houston	Astros	Rockets	Texans
Miami	Marlins	Heat	Dolphins

Exploring the Data

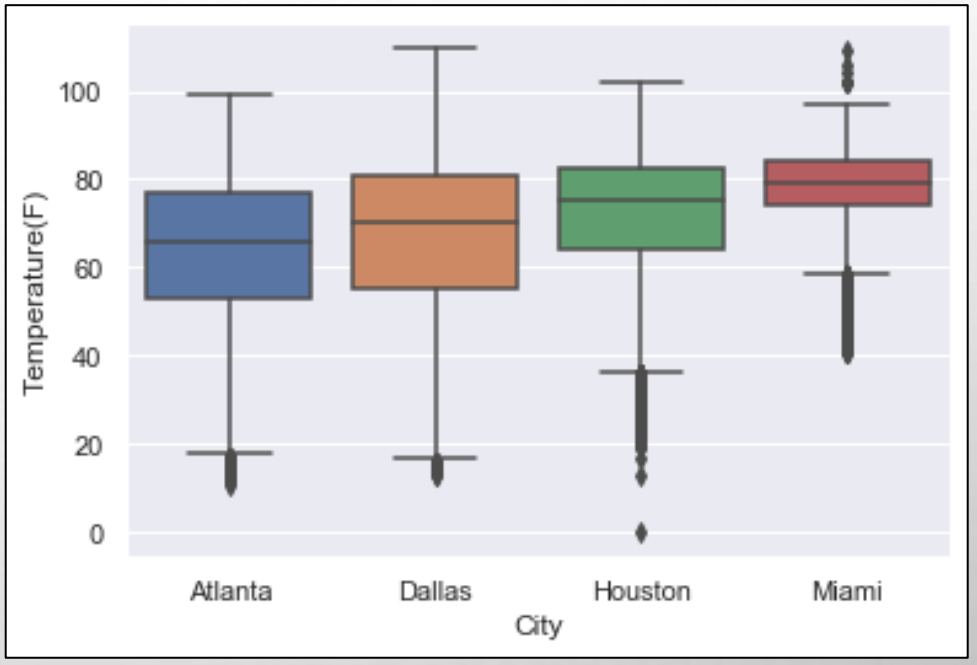
Most Traffic Accidents appear to occur on weekdays



Exploring the Data

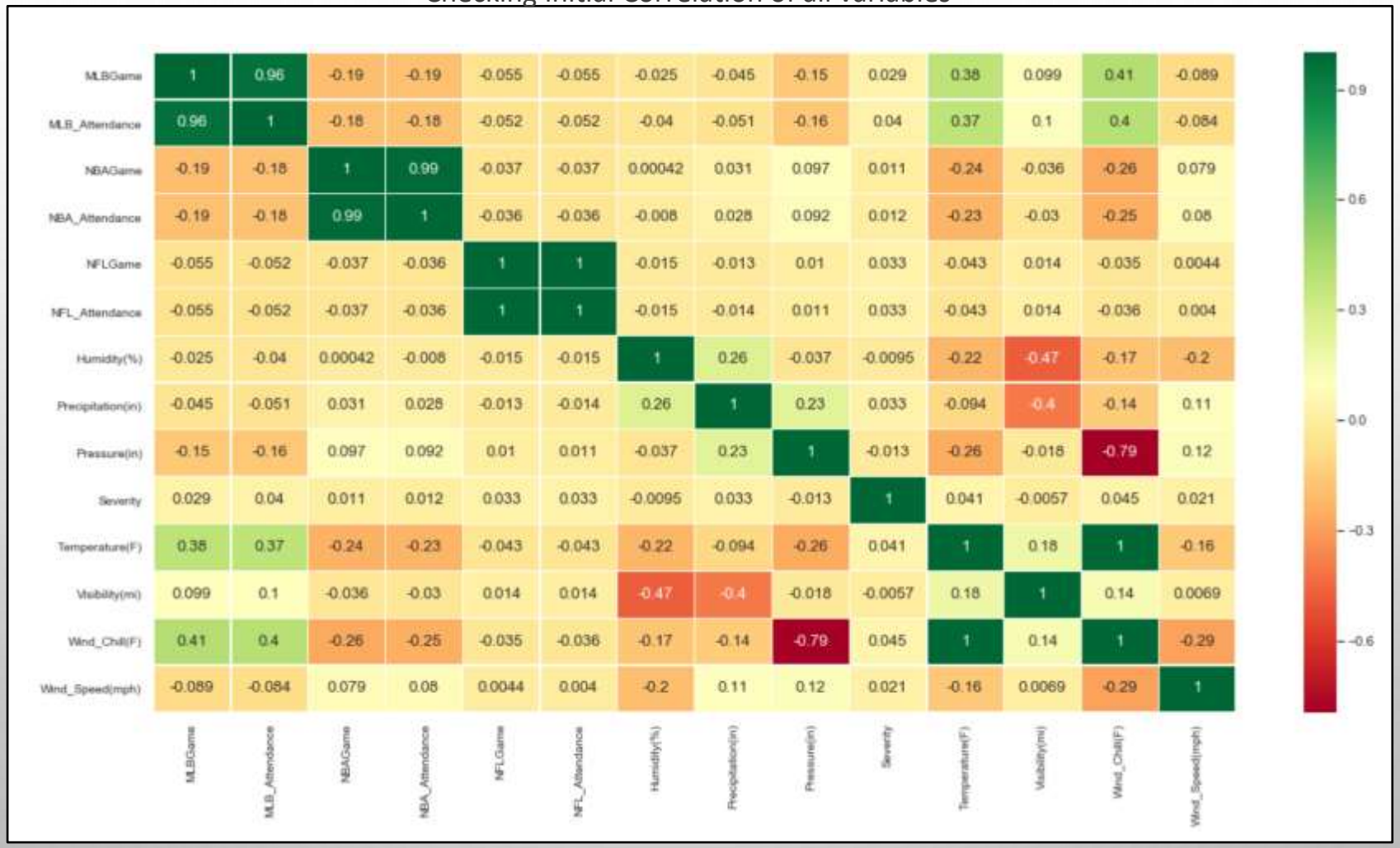
Most accidents occur during clearer weather

Condition	# of Accidents
Clear	47,908
Mostly Cloudy	38,359
Partly Cloudy	27,487
Overcast	27,467
Fair	18,037
Light Rain	8,627
Cloudy	6,237
Rain	1,907
Fog	1,490



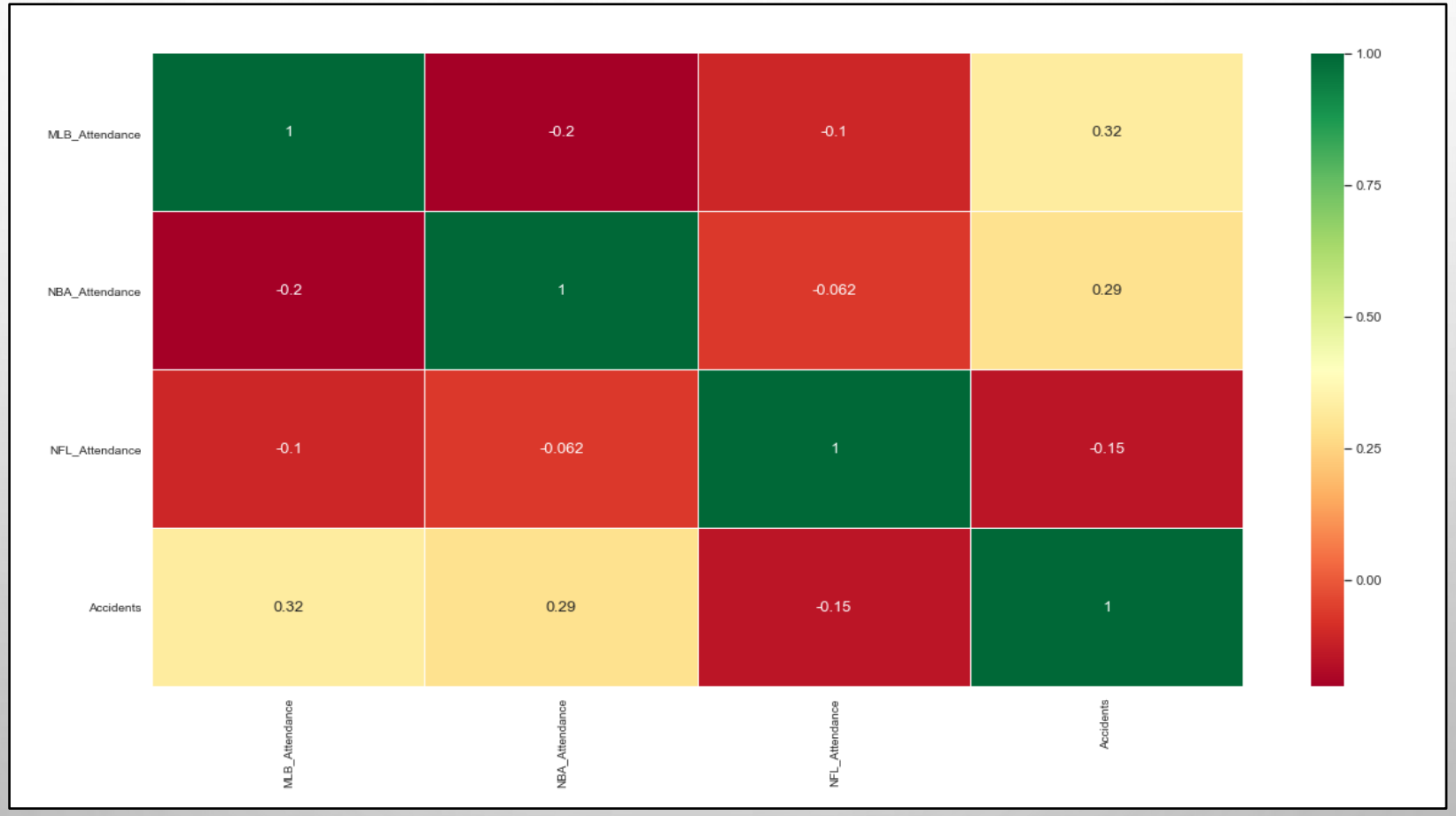
Exploring the Data

Checking Initial Correlation of all variables



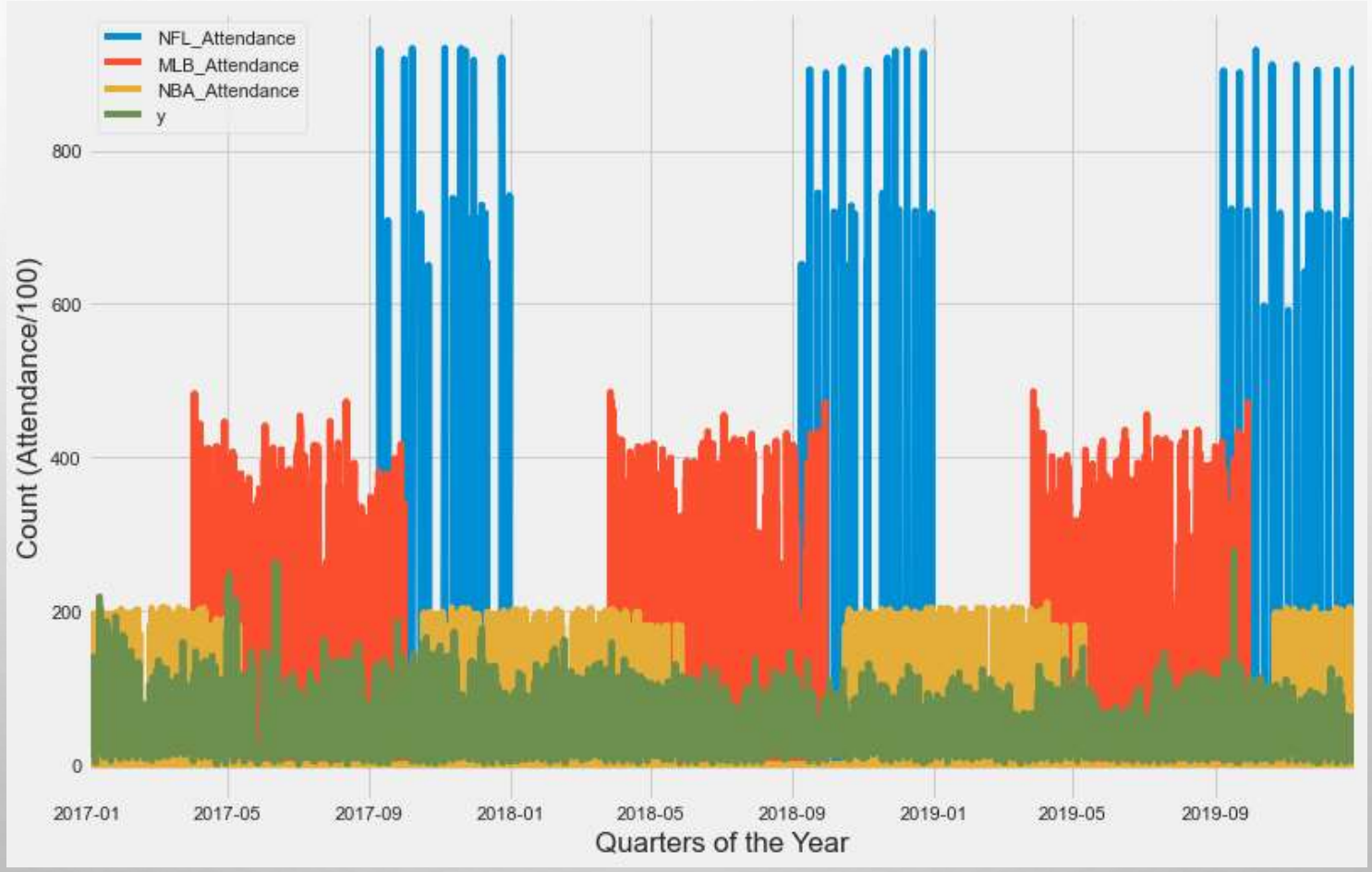
Exploring the Data

Checking Initial Correlation of Sporting Events and Traffic Accidents



Exploring the Data

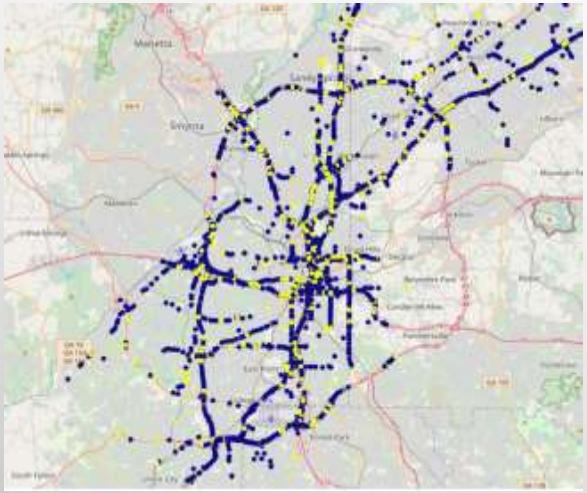
NFL has the greatest level of attendance



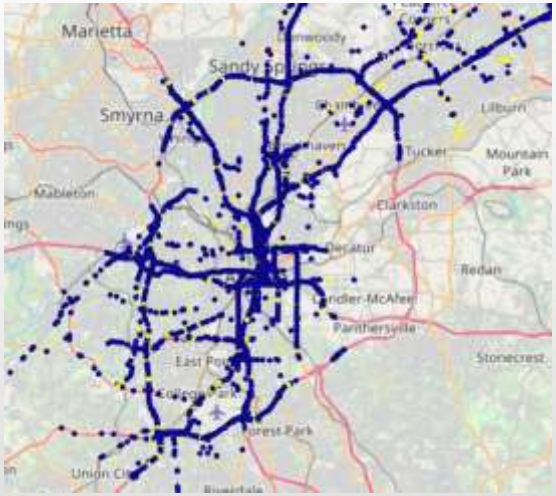
Exploring the Data

Geographic distribution of accidents for Atlanta

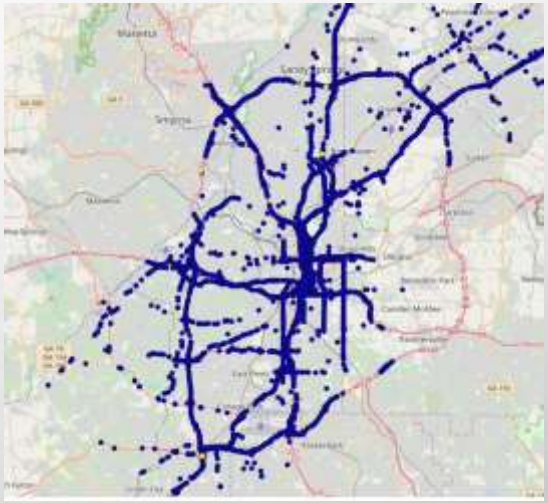
MLB



NBA

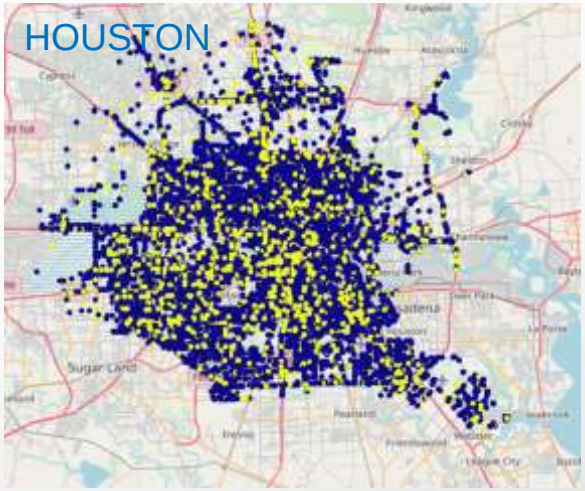


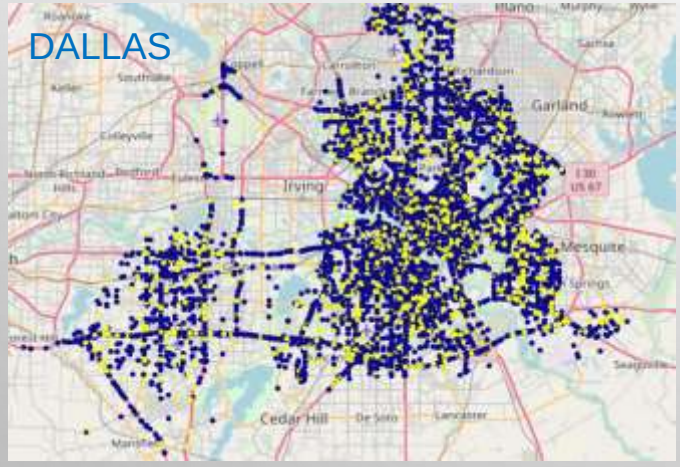
NFL

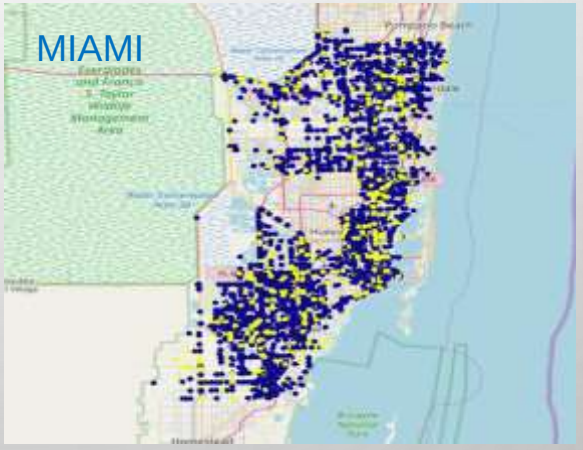


Exploring the Data

Review MLB accidents in our four cities

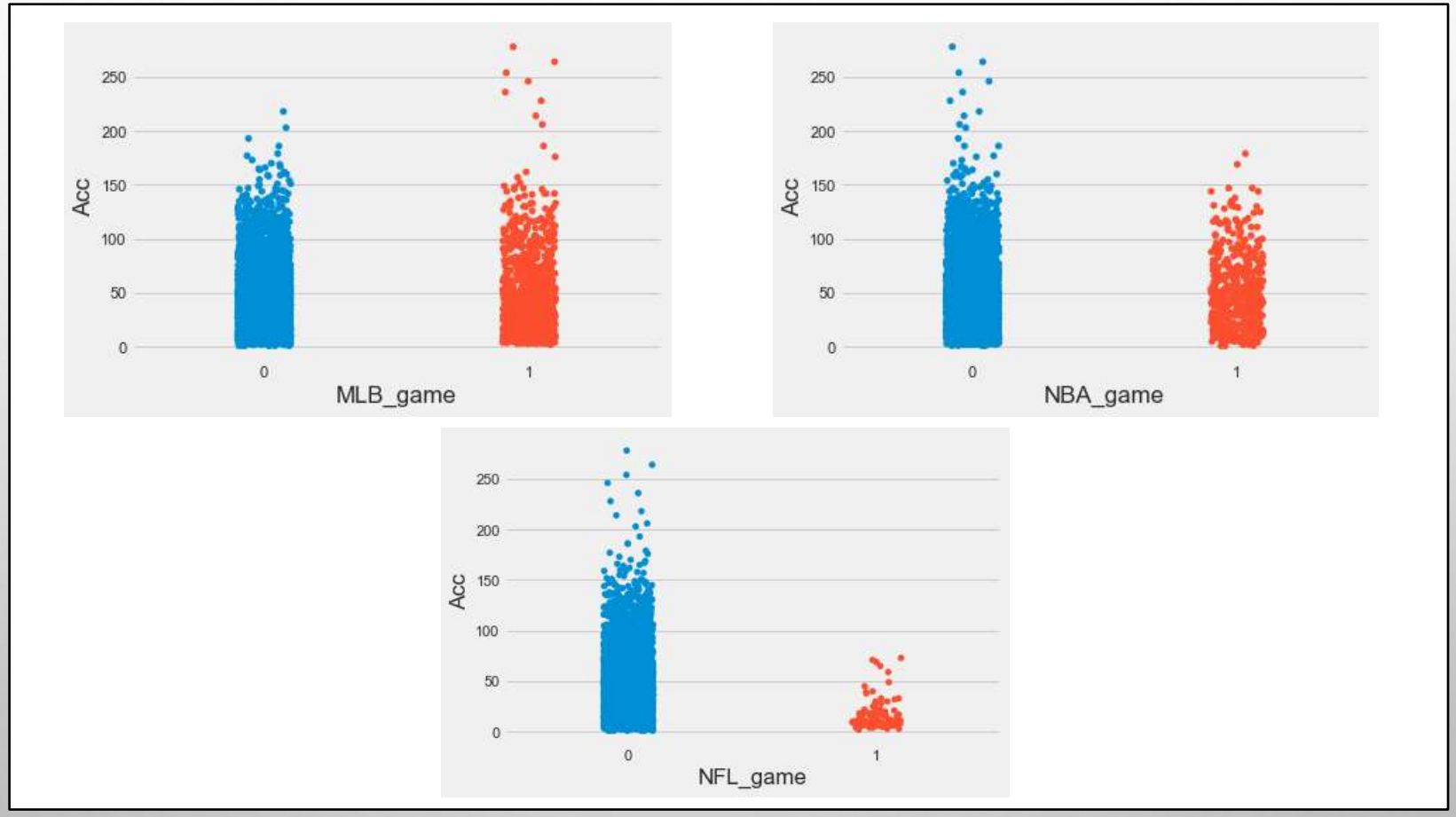






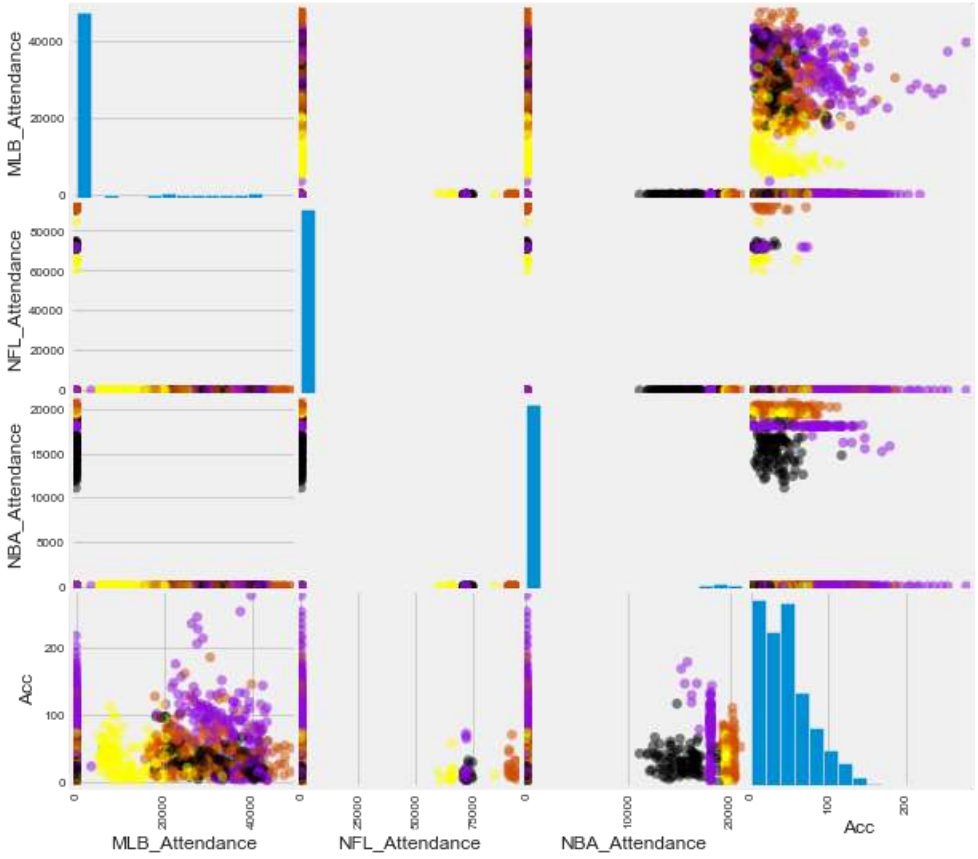
Modeling the Data

Presence or Absence of Sporting Event & Accidents



Modeling the Data

Scatter Plot of Attendance to Accidents by City



Hue = City

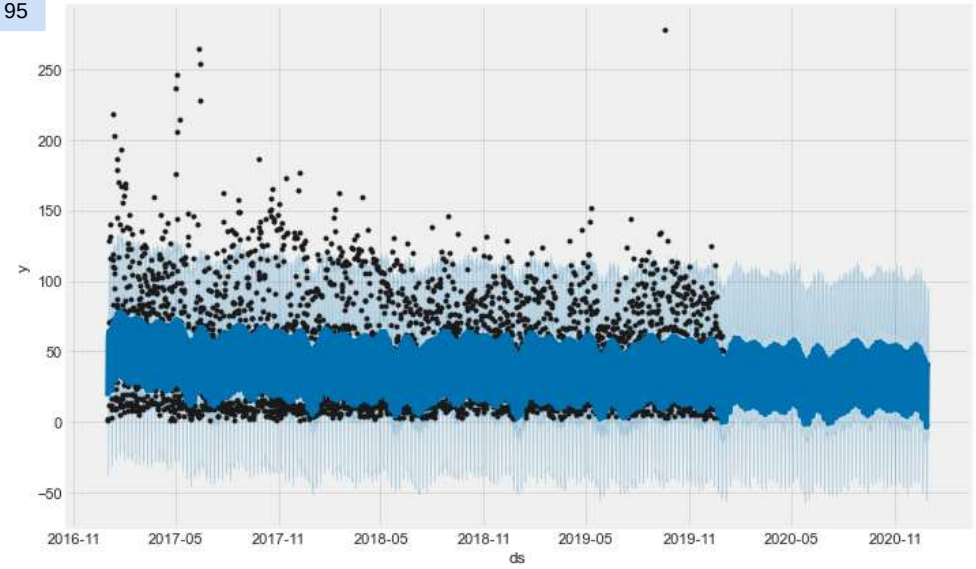
Modeling the Data

Time Series Modeling & Prophet Forecasting

Future Date 2020	Accidents Predicted	Accidents Predicted Lower End	Accidents Predicted Upper End
2020-12-26	0	-53	55
2020-12-27	-4	-55	55
2020-12-28	37	-11	84
2020-12-29	41	-8	94
2020-12-30	40	-11	95

- Daily prediction for year 2020
- Accidents rarely exceed 150/day
- Accident prediction < 125/ day

Actual & Predicted Accidents per Day for SE Cities)



Modeling the Data

K Nearest Neighbor Prediction Models

Model	y	x1	x2	x3	x4	Parameters
A	Accidents	NFL Attendance	NBA Attendance	MLB Attendance	City	KNN: 10 Accuracy: 2.07%
Prediction A1	17	1200	0	0	Atlanta	
Prediction A2	32	800	0	12	Miami	
B	Accidents	NFL Game	NBA Game	MLB Game	City	KNN: 6 Accuracy: 29.55%
Prediction B1	Low Risk*	0	1200	800	Miami	
Prediction B2	Medium Low Risk*	700	0	1800	Houston	
C	Accidents	NFL Game	NBA Game	MLB Game	City	KNN: 6 Accuracy: 1.35%
Prediction C1	8	1	0	0	Atlanta	
Prediction C2	16	1	0	1	Houston	
D	City	NFL Attendance	NBA Attendance	MLB Attendance	Accidents	KNN: 12 Accuracy: 47.11%
Prediction D1	Miami	0	900	0	35	
Prediction D2	Atlanta	700	0	1800	20	

*Accident Bins: 0-20 = No Risk; 21-40 = Low Risk; 41-60 = Medium Low Risk; 61-80 = Medium Risk; 81-100= Medium High Risk; 101-150 = High Risk; 150+ = Unacceptable Risk

Interpreting the Data

Interpreting

1. Seasonality of Sporting Events vs. Seasonality of Accident Data

2. No Strong Causal Relationship Between Games and Accidents, weather is more correlated with accidents.

3. Game Type and Day of the Week- MLB and NBA Friday games are at increased risk of accidents

Interpreting the Data

Conclusion

1. Data is supportive of building a stadium within an Urban area, especially if the games are on the weekend.

2. Sporting games played in inclement weather, may increase the risk of accidents in Urban Southeast Cities.

3. Planning committee should use models to predict attendance at all 3 major sports events and risk of accidents to encourage drivers to take public transportation.