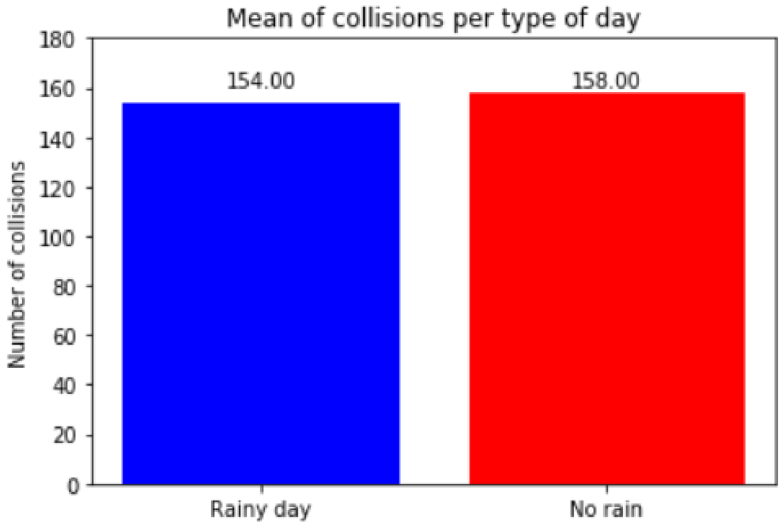
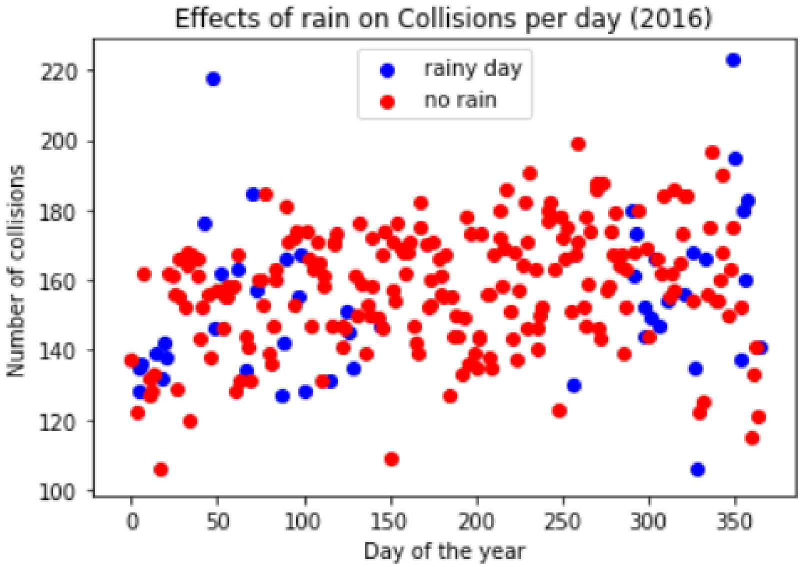
Los Angeles Traffic Collision Study II

Idea From Project 1

Potential cause of increasing in daily collision in LA

- Weather? (Raining vs. Non-Raining)
- Temperature? (Does the heat weaken your ability to drive?)
- Day of the week?



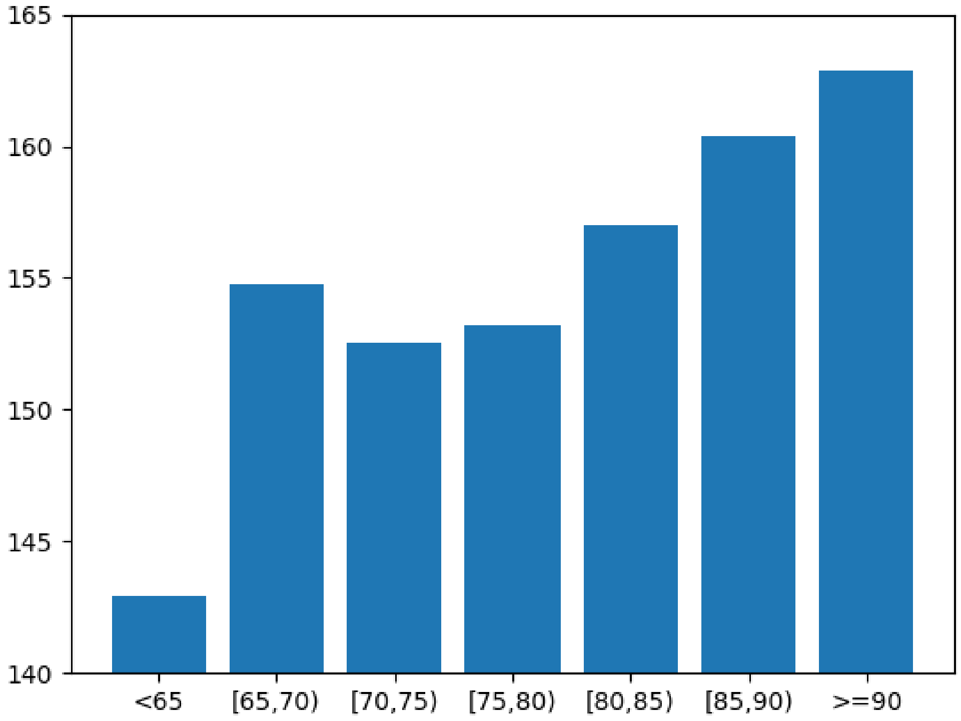


Conclusion of Project 1: Weather

- Due to hug gap between number of rainy day and non-rainy day, statistically, average daily collision number is not affected much by rain or no rain.

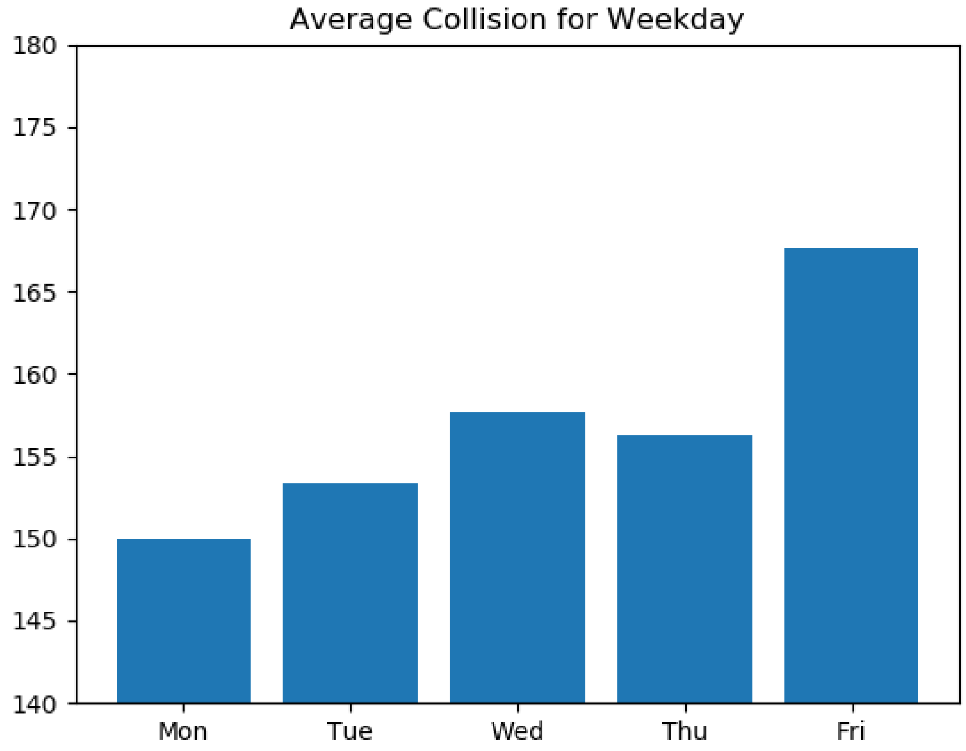
- Further thinking:
  - Intensity of raining
  - Time/Duration of raining

Conclusion of Project 1: Temperature



- The average number of daily collisions trend to higher with higher temperature.

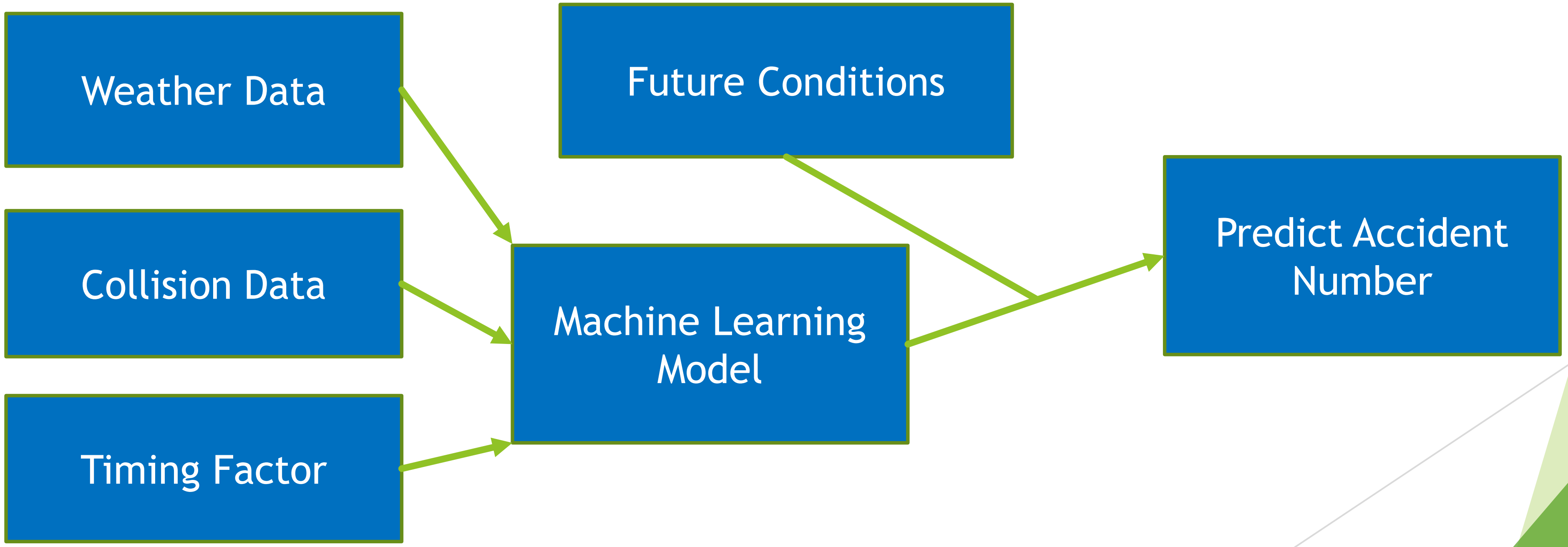
Conclusion of Project 1: Day of the week



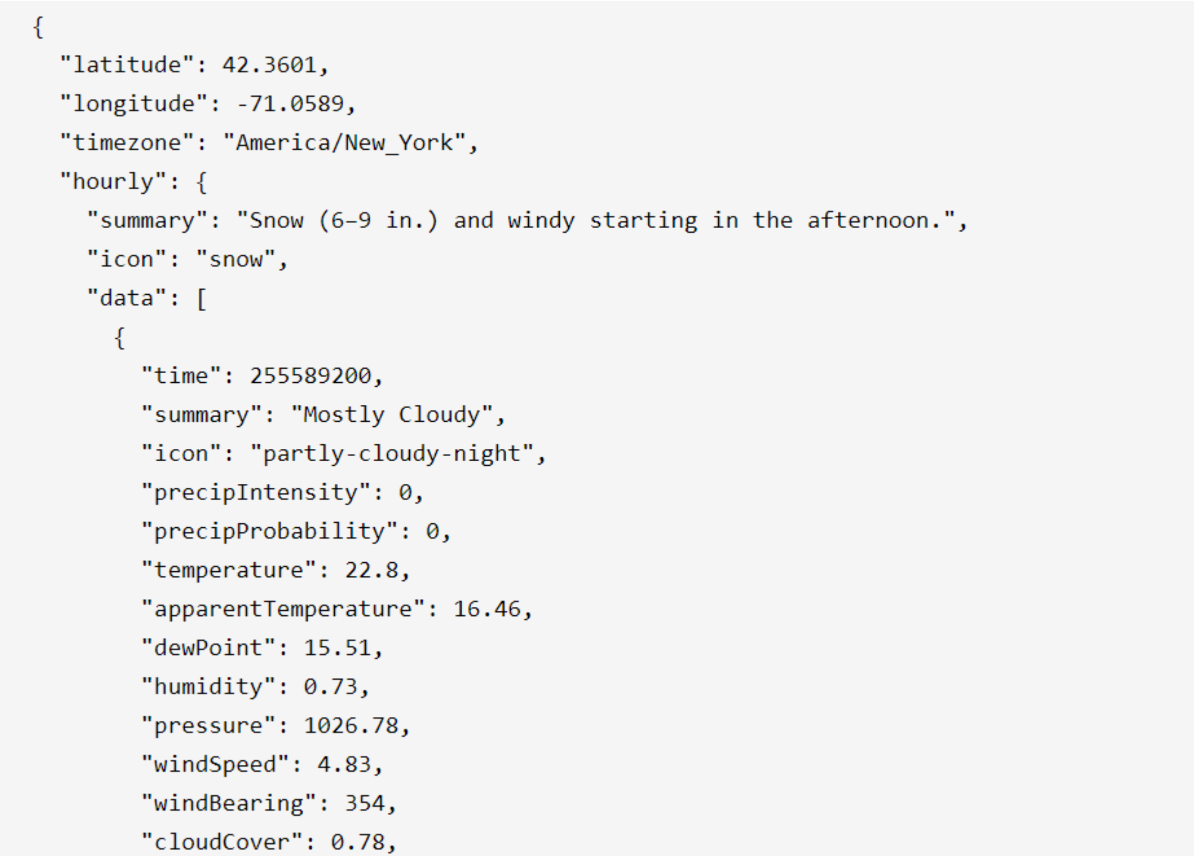
- On average, collisions on Friday is significantly higher than other weekdays.

Proposal

- Is it possible to use the available data to generate a model to predict how many collisions will occur?



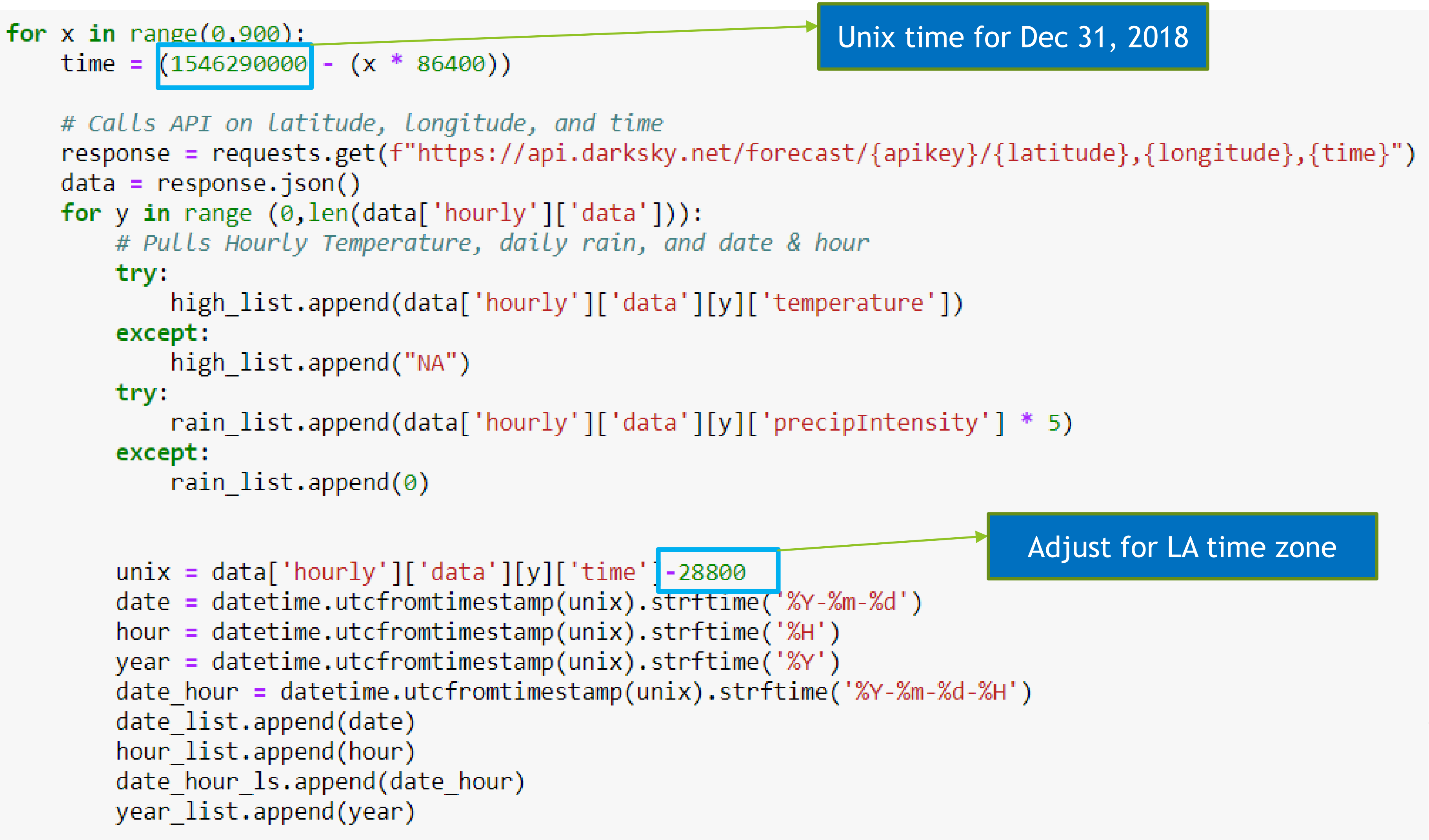
Data Collection: Weather Data



```
{
  "latitude": 42.3601,
  "longitude": -71.0589,
  "timezone": "America/New_York",
  "hourly": {
    "summary": "Snow (6–9 in.) and windy starting in the afternoon.",
    "icon": "snow",
    "data": [
      {
        "time": 255589200,
        "summary": "Mostly Cloudy",
        "icon": "partly-cloudy-night",
        "precipIntensity": 0,
        "precipProbability": 0,
        "temperature": 22.8,
        "apparentTemperature": 16.46,
        "dewPoint": 15.51,
        "humidity": 0.73,
        "pressure": 1026.78,
        "windSpeed": 4.83,
        "windBearing": 354,
        "cloudCover": 0.78,
```

- Dark Sky API:
  - Hourly temperature
  - Hourly precipitation Intensity

API call to get weather data



```
for x in range(0,900):
    time = (1546290000 - (x * 86400))

    # Calls API on latitude, longitude, and time
    response = requests.get(f"https://api.darksky.net/forecast/{apikey}/{latitude},{longitude},{time}")
    data = response.json()
    for y in range (0,len(data['hourly']['data'])):
        # Pulls Hourly Temperature, daily rain, and date & hour
        try:
            high_list.append(data['hourly']['data'][y]['temperature'])
        except:
            high_list.append("NA")
        try:
            rain_list.append(data['hourly']['data'][y]['precipIntensity'] * 5)
        except:
            rain_list.append(0)


        unix = data['hourly']['data'][y]['time'] -28800
        date = datetime.utcfromtimestamp(unix).strftime('%Y-%m-%d')
        hour = datetime.utcfromtimestamp(unix).strftime('%H')
        year = datetime.utcfromtimestamp(unix).strftime('%Y')
        date_hour = datetime.utcfromtimestamp(unix).strftime('%Y-%m-%d-%H')
        date_list.append(date)
        hour_list.append(hour)
        date_hour_ls.append(date_hour)
        year_list.append(year)
```

Unix time for Dec 31, 2018

Adjust for LA time zone

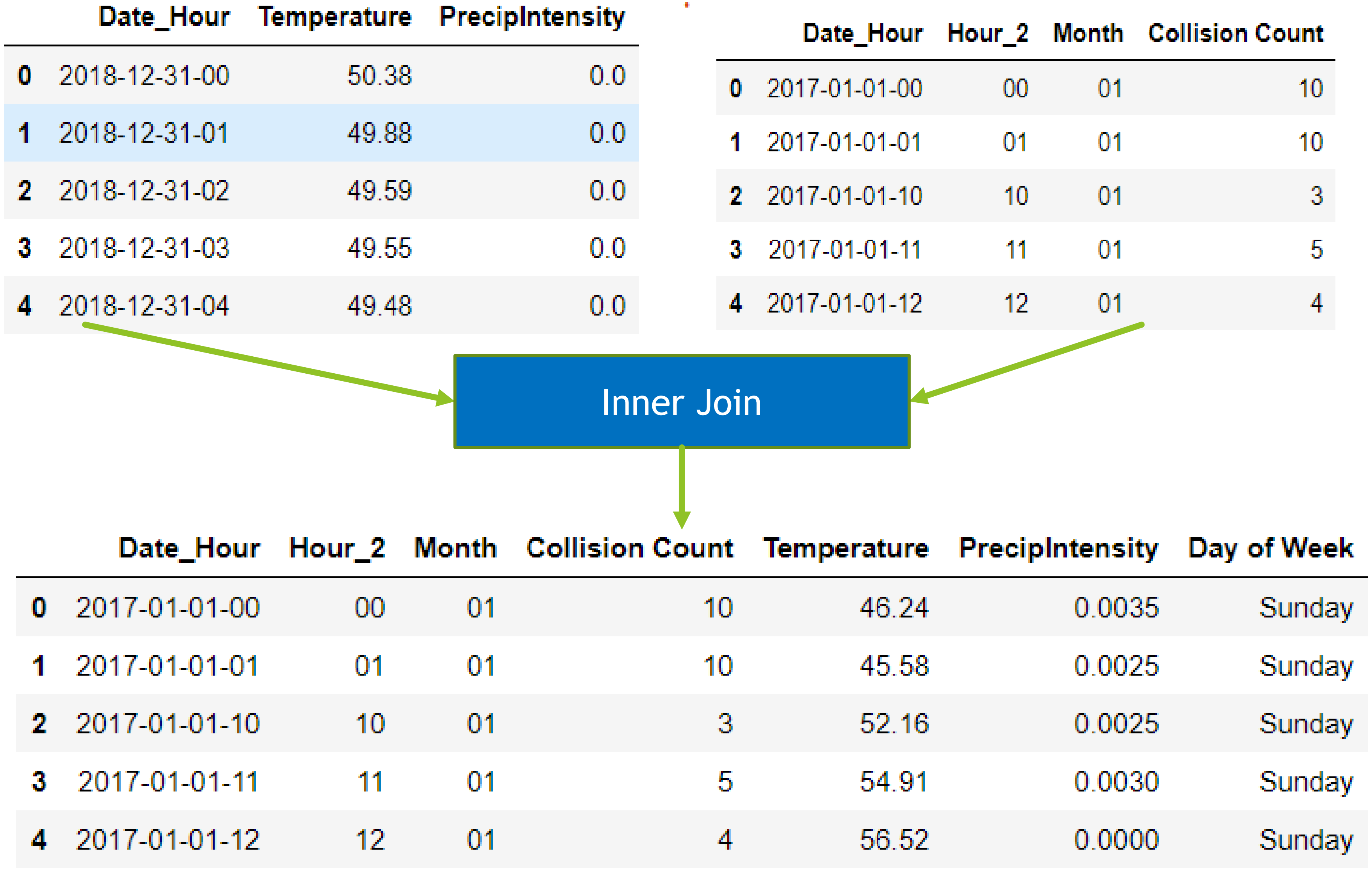
Data Collection: Collision Data

- Source:
  - LOS ANGELES OPEN DATA (https://data.lacity.org/)
  - https://data.lacity.org/A-Safe-City/Traffic-Collision-Data-from-2010-to-Present/d5tf-ez2w

A Message from Mayor Eric Garcetti

We are sharing city data with the public to increase transparency, accountability and customer service and to empower companies, individuals and non-profit organizations with the ability to harness a vast array of useful information to improve life in our city. I hope that this data will help drive innovation and problem solving within the public and private sectors and that Angelenos will use it to more deeply understand and engage with their city. I encourage you to explore data.lacity.org to conduct research, develop apps or simply to poke around.

Data Ready

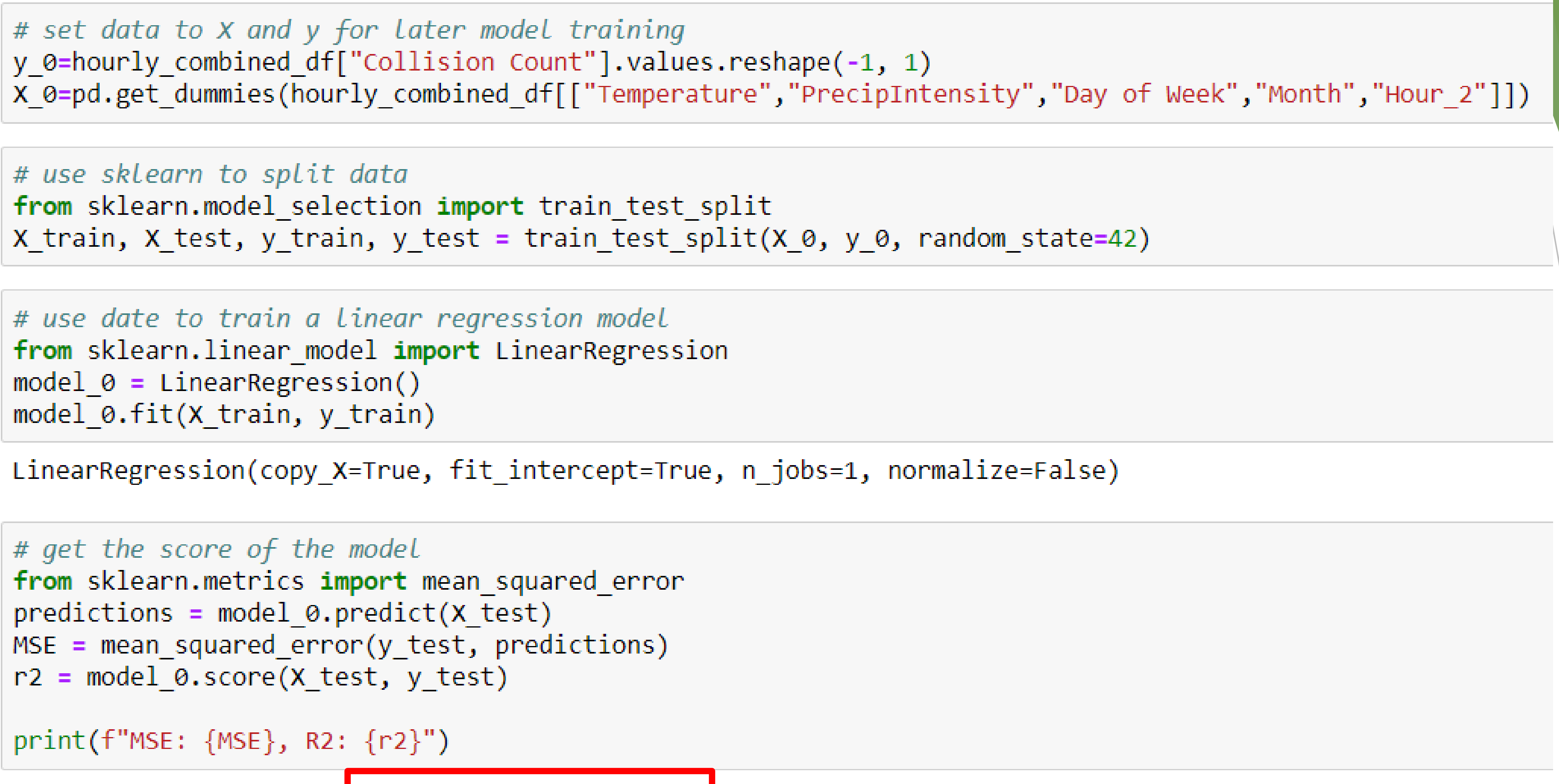


	Date_Hour	Temperature	PrecipIntensity
0	2018-12-31-00	50.38	0.0
1	2018-12-31-01	49.88	0.0
2	2018-12-31-02	49.59	0.0
3	2018-12-31-03	49.55	0.0
4	2018-12-31-04	49.48	0.0

	Date_Hour	Hour_2	Month	Collision Count
0	2017-01-01-00	00	01	10
1	2017-01-01-01	01	01	10
2	2017-01-01-10	10	01	3
3	2017-01-01-11	11	01	5
4	2017-01-01-12	12	01	4

Inner Join

	Date_Hour	Hour_2	Month	Collision Count	Temperature	PrecipIntensity	Day of Week
0	2017-01-01-00	00	01	10	46.24	0.0035	Sunday
1	2017-01-01-01	01	01	10	45.58	0.0025	Sunday
2	2017-01-01-10	10	01	3	52.16	0.0025	Sunday
3	2017-01-01-11	11	01	5	54.91	0.0030	Sunday
4	2017-01-01-12	12	01	4	56.52	0.0000	Sunday



```
# set data to X and y for later model training
y_0=hourly_combined_df["Collision Count"].values.reshape(-1, 1)
X_0=pd.get_dummies(hourly_combined_df[["Temperature","PrecipIntensity","Day of Week","Month","Hour_2"]])
```

```
# use sklearn to split data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_0, y_0, random_state=42)
```

```
# use date to train a linear regression model
from sklearn.linear_model import LinearRegression
model_0 = LinearRegression()
model_0.fit(X_train, y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

```
# get the score of the model
from sklearn.metrics import mean_squared_error
predictions = model_0.predict(X_test)
MSE = mean_squared_error(y_test, predictions)
r2 = model_0.score(X_test, y_test)

print(f"MSE: {MSE}, R2: {r2}")
```

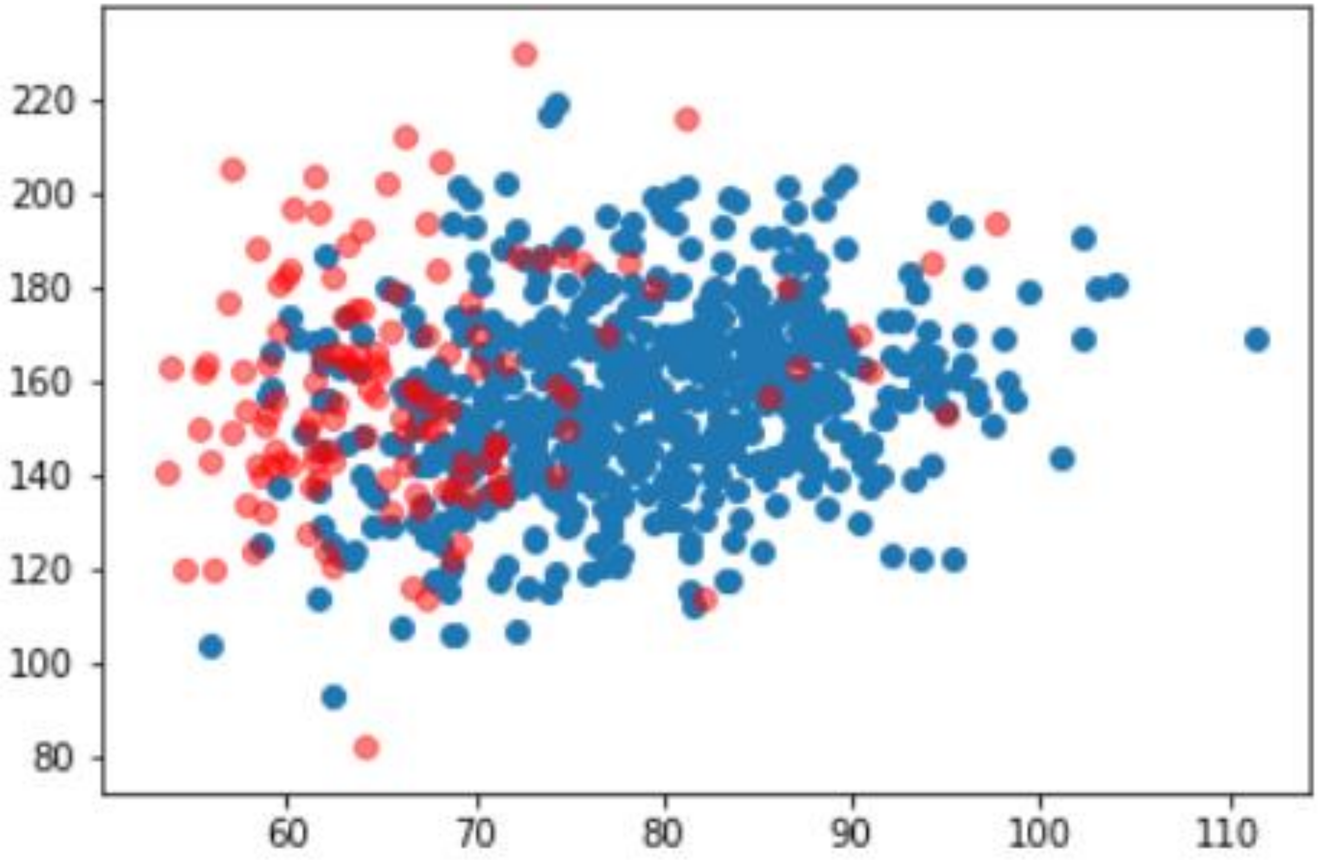
```
MSE: 8.690810300443479, R2: 0.4323312050636351
```

Not Good!!

What’s wrong?

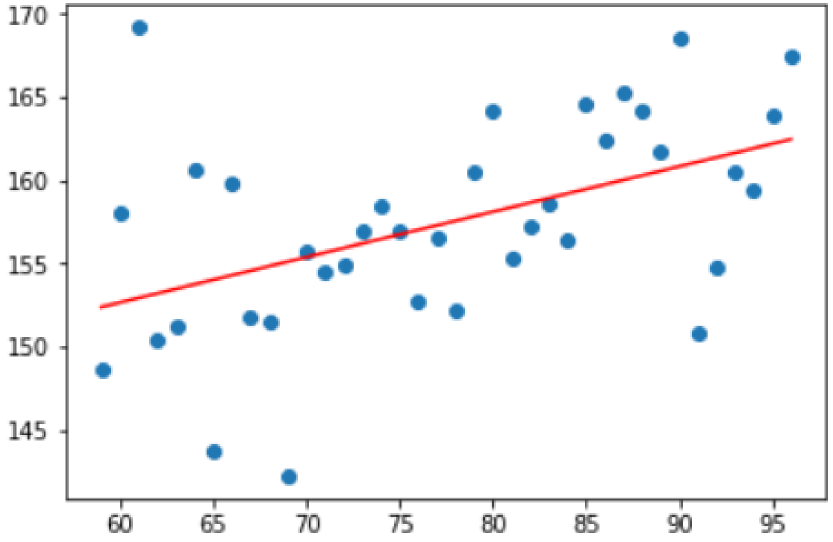
- Not all factors are included:
  - Sunshine
  - Visibility
  - Road condition
  - Wind speed
  - Pedestrian

Other Attempt:
Daily Temperature vs Collision

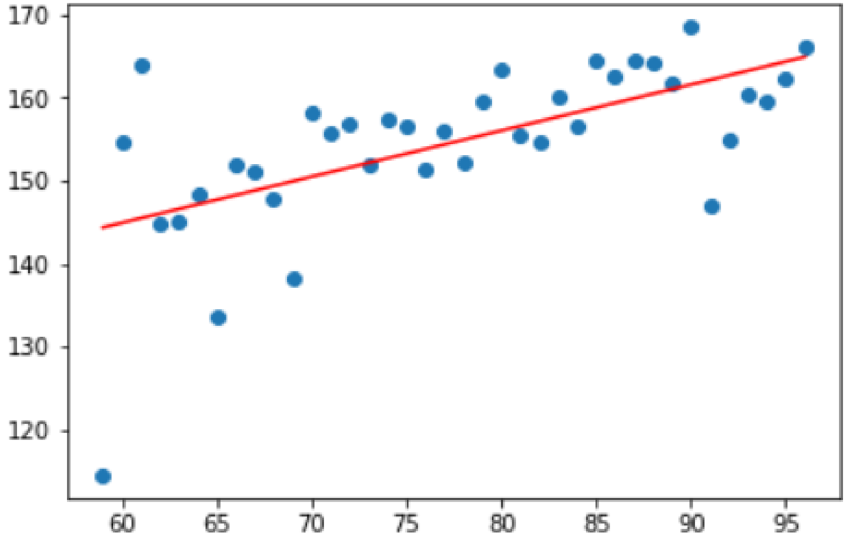


Other Attempt: Temperature vs Average Collision

Including Rainy Day



Excluding Rainy Day



Thank You!

- Project 1 Team:
  - Anthony Hyunh
  - Giselle Chiu
  - Greg Black
- Los Angeles Open Data (https://data.lacity.org/)
- UCI Data Analytic Bootcamp:
  - Chris Shoe
  - Michael
  - Peter Kim