Police Response Prediction

USING SEATTLE POLICE CALL DATA TO PREDICT RESPONSE TIME AND PROBABLE CRIMES

Alex Tresselt
CS 7180 | 12/13/2018

1. INTRO

The challenge I tried to address was using past call data to predict future crimes given only a location and time. This could be useful deciding what areas of the city to visit, or to learn more about the neighborhood they live or work in.
A second challenge was using the same data to predict the Police Department's response time to a call. This could be useful for someone who is calling in to report a crime and would like to know how long they would have to wait before a police officer arrives.

2. DATASET

I used the Call Data dataset that is available on the City of Seattle's Open Data Portal.[1] It is a record of every call for service coming in to the Seattle Police Department's (SPD) Communication Center, through either 911 calls or other means, since July 2009, and is updated every day. The calls are collected and logged by SPD's Computer Aided Dispatch (CAD) system.
The dataset I used for this project was downloaded on November 7th, 2018. It contains 3,713,328 rows and 11 columns.

3. DATA EXPLORATION/UNDERSTANDING

Dataset Structure

	CAD Event Number	Event Clearance Description	Call Type	Priority	Initial Call Type	Final Call Type	Original Time Queued	Arrived Time	Precinct	Sector	Beat
0	2009000287074	ASSISTANCE RENDERED	TELEPHONE OTHER, NOT 911	1	SHOTS - IP/JO - INCLUDES HEARD/NO ASSAULT	--DISTURBANCE - OTHER	08/14/2009 11:12:34 PM	Aug 14 2009 11:13:55:000PM	EAST	GEORGE	G3
1	2009000287075	ASSISTANCE RENDERED	ONVIEW	7	PREMISE CHECK, OFFICER INITIATED ONVIEW ONLY	--PREMISE CHECKS - CRIME PREVENTION	08/14/2009 11:13:49 PM	Aug 14 2009 11:13:50:000PM	NORTH	LINCOLN	L1
2	2009000287078	OTHER REPORT MADE	ONVIEW	7	TRAFFIC STOP - OFFICER INITIATED ONVIEW	--TRAFFIC - MOVING VIOLATION	08/14/2009 11:16:20 PM	Aug 14 2009 11:16:20:000PM	NORTH	LINCOLN	L2
3	2009000287080	NO POLICE ACTION POSSIBLE OR NECESSARY	TELEPHONE OTHER, NOT 911	4	NOISE - DIST, GENERAL (CONST, RESID, BALL PLAY)	--DISTURBANCE - NOISE	08/14/2009 11:17:22 PM	Aug 15 2009 12:16:18:000AM	WEST	DAVID	D2
4	2009000287081	ASSISTANCE RENDERED	911	1	UNKNOWN - ANI/ALI - LANDLINE (INCLUDES OPEN LINE)	--ASSIST PUBLIC - 911 HANG UP, OPEN LINE	08/14/2009 11:17:34 PM	Aug 15 2009 01:40:09:000AM	SOUTH	OCEAN	O2

Figure 1 First 5 rows of the Call Data dataset

Each row represents a call for service. There are 11 columns, describing how the call came in, what crime is being reported, its location, the time reported, the time a police officer arrived, and how the call was resolved.

---

[1] https://data.seattle.gov/Public-Safety/Call-Data/33kz-ixgy

Column Descriptions

Column	Description (taken from the Open Data Portal page)
CAD Event Number	A unique ID assigned to each call
Event Clearance Description	How the call was resolved by the responding police officer
Call Type	How the call was received by the communications center
Priority	Priority of the call, 1-9, with 1 being the most serious
Initial Call Type	The initial classification of the crime being reported
Final Call Type	The final crime classification, made by the responding officer
Original Time Queued	Time the call was logged
Arrived Time	Time the first responding police officer arrived
Precinct	Precinct where the call originated.
Sector	Sector where the call originated. Sectors roll up into Precincts
Beat	Beat where the call originated. Beats roll up into Sectors

Priority Description

Priority

Code	Level	Guideline for Use
1	Immediate	Any incident which poses obvious danger to the life of a citizen or officer. Life threatening crimes in progress such as shootings, stabbings, helping the officer. Major disturbances including those with weapons, serious injury accidents, robbery alarms and prowlers.
2	Urgent	Altercations or actions which if not policed quickly would or could develop into more serious or major proportions; incidents where in there is a threat of violence, injury or damage to property; unknown injury/minor injury accidents; shoplifters in custody.
3	Prompt	Investigations or minor incident type complaints in which response time is not a critical issue.
4	As Available	Mischief or nuisance type complaints such as snow ball throwing, firecrackers, etc. Dispatched after all other higher precedence calls have been assigned.
5	Telephone Reporting	Call events handled by officers assigned to the Internet/Telephone Reporting Unit
6	Secondary Call Taker	Call events handled by call takers assigned in the Communications Section
7	Officer Initiated	Call events initiated by officers for proactive work, such as traffic stops, suspicious stops, premise checks and directed patrol activities

Figure 2 Priority Level Descriptions, from SPD Call for Service Dashboard[2]

Each call is assigned a priority by the Communications Center, based on its severity. The highest is level 1, which indicates an immediate danger to life. The lowest is level 4, which are 'nuisance' complaints that are handled after all higher priority calls have been handled. There are also levels 5-7 and 9, which are used for calls generated for taking reports.

For my models, I trimmed the dataset to only include calls with a priority of 3 or higher.

[2] https://www.seattle.gov/police/information-and-data/calls-for-service-dashboard (Methodology tab)

Locations

For privacy reasons, the location where the call originated is given with a “beat” instead of a street address. A beat is the area that an individual police officer will patrol. SPD divides the city into five Precincts, based on geographical location: North, West, East, South, and Southwest. Each Precinct is further divided into several Sectors, and then each Sector contains several Beats. Each Beat is given a letter, matching its Sector, and a number. For example, O3 is a beat located in Ocean Sector, which is in the South Precinct. [3]

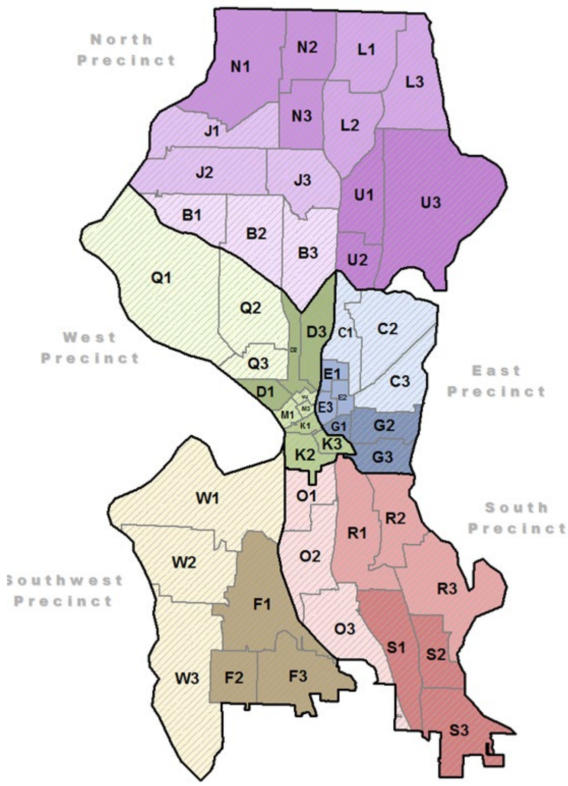

Figure 3 Seattle Police Beat Map[4]

[3] https://www.seattle.gov/police/about-us/about-policing/precinct-and-patrol-boundaries
[4] https://www.seattle.gov/Images/Departments/Police/aboutUs/precinctmap.png

Most Frequently Reported Crimes

```
Final Call Type
--PREMISE CHECKS - CRIME PREVENTION                    319630
--SUSPICIOUS CIRCUM. - SUSPICIOUS PERSON               308718
--DISTURBANCE - OTHER                                  263116
--TRAFFIC - MOVING VIOLATION                           209521
--TRAFFIC - PARKING VIOL (EXCEPT ABANDONED CAR)        190080
--TRAFFIC - MV COLLISION INVESTIGATION                 167921
--ASSIST PUBLIC - OTHER (NON-SPECIFIED)                132945
--SUSPICIOUS CIRCUM. - SUSPICIOUS VEHICLE               93011
--THEFT - CAR PROWL                                     82681
--PROWLER - TRESPASS                                    82520
```

Figure 4 Top 10 Most Reported Crimes, 2009-2018

The ten most frequently reported crimes and their counts.

Calls by Month

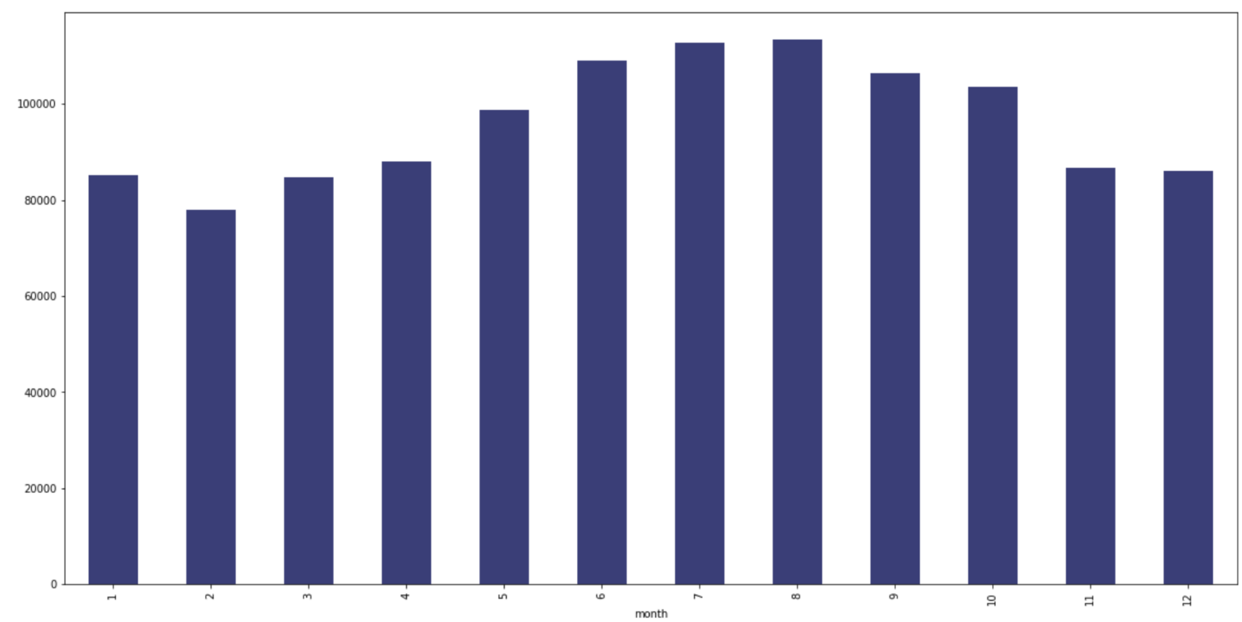

Figure 5 Calls by Month, 2009-2018

Grouping the calls by month reveals a seasonal pattern, with an increase during the summer months, and a decrease in winter. The lowest amount of calls occur in February, and the most in August.

Calls by Day of Week

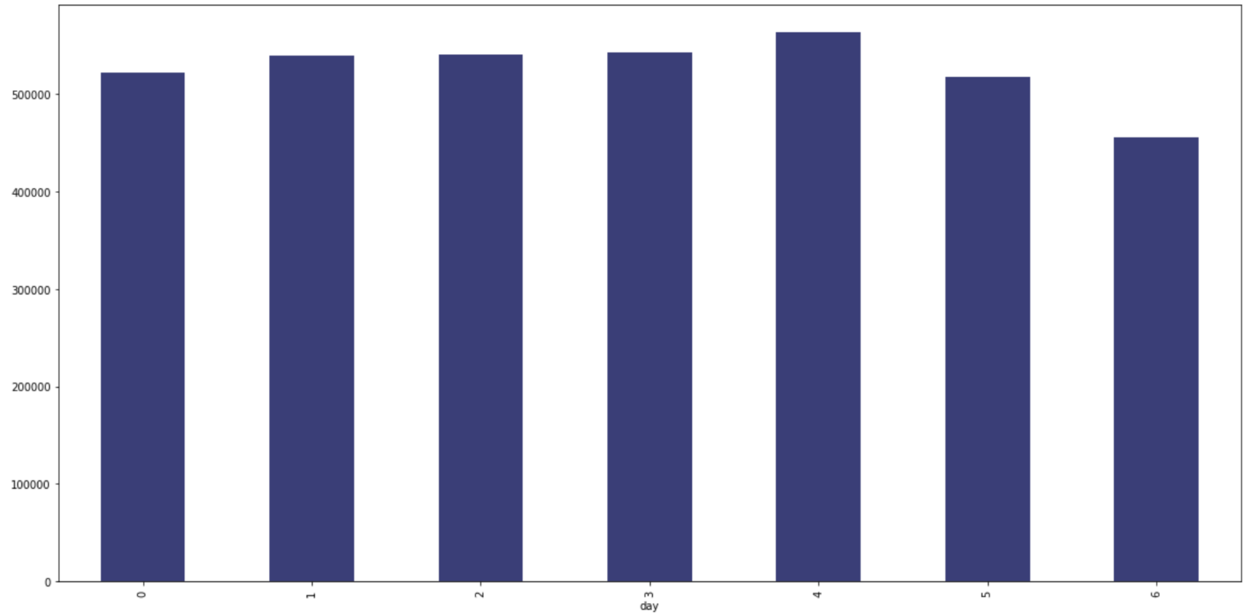

Figure 6 Calls by day of week, 2009-2018

Grouping the calls by day of week (0 = Monday, 6 = Sunday) reveals a slight increase on Fridays, then a drop over the weekend with Sundays having the lowest number of calls.

Calls by Hour

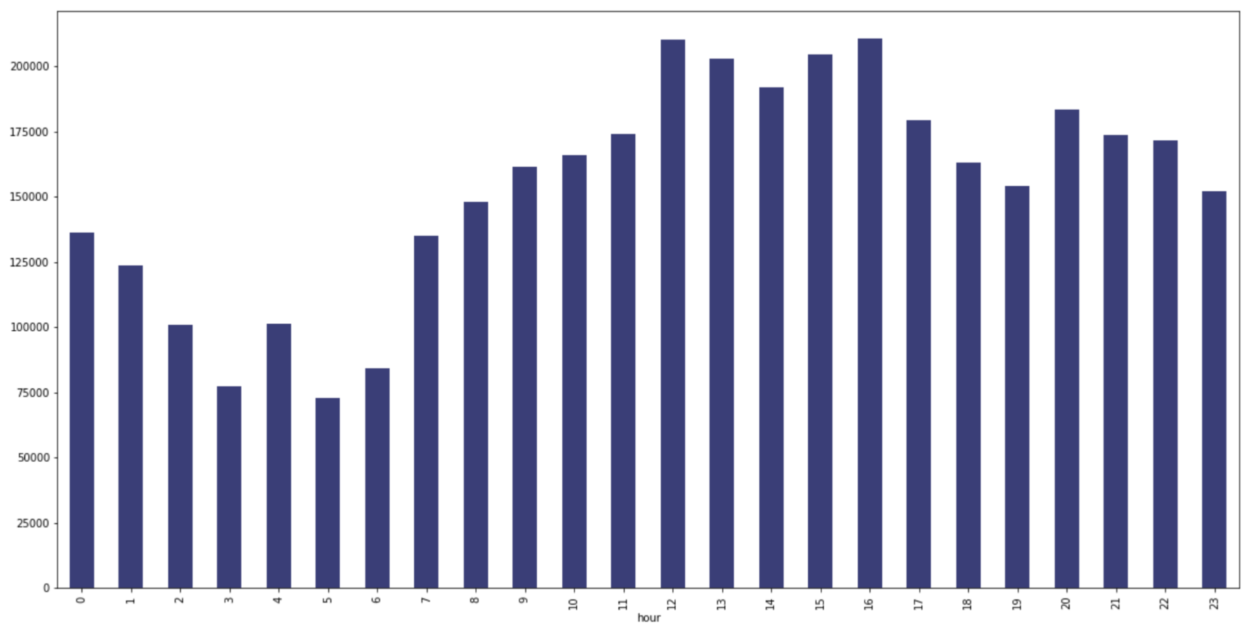

Figure 7 Calls grouped by hour, 2009-2018

Grouping the calls by hour also reveals a pattern. The calls are most frequent from noon to early afternoon, and are least frequent during early morning.

Calls by Minute

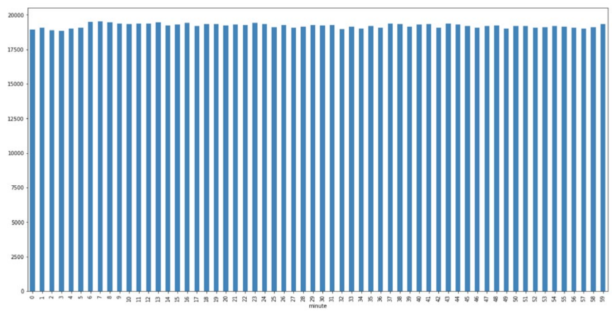

Figure 8 Calls by minute, 2009-2018 using a bar chart

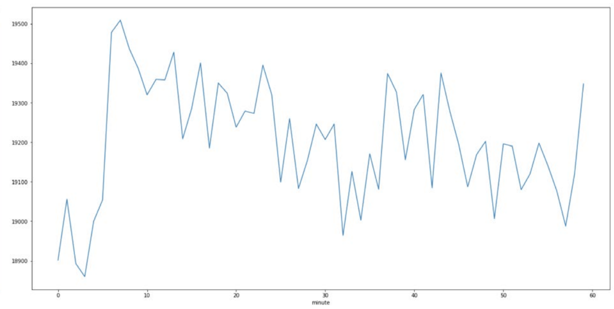

Figure 9 Calls by minute, 2009 - 2018 using a line plot

Grouping the calls by minute shows a slight decrease in calls during the first five minutes of the hour, but overall very similar across the hour. However, using a line plot instead of a bar chart appears to show much more variability, even though there is only a 3% difference between the lowest and highest point on the plot.

Average Response Time by Year

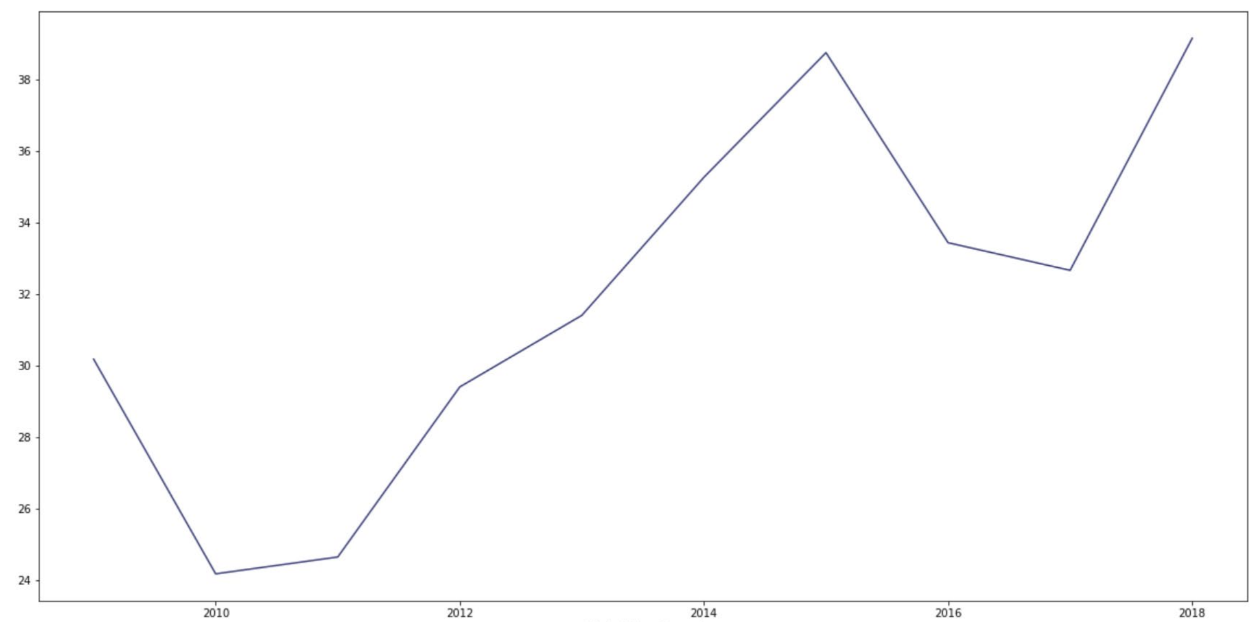

Figure 10 Average Response Times in Minutes from 2009-2018

Graphing the average response time shows an overall increase in the response time, although there was a decrease from 2015-2017. Data from 2009 and 2018 are incomplete, so may not represent an accurate trend. Because the averages change every year, I decided to train my model using only calls from one year, 2017.

Response Time by Precinct

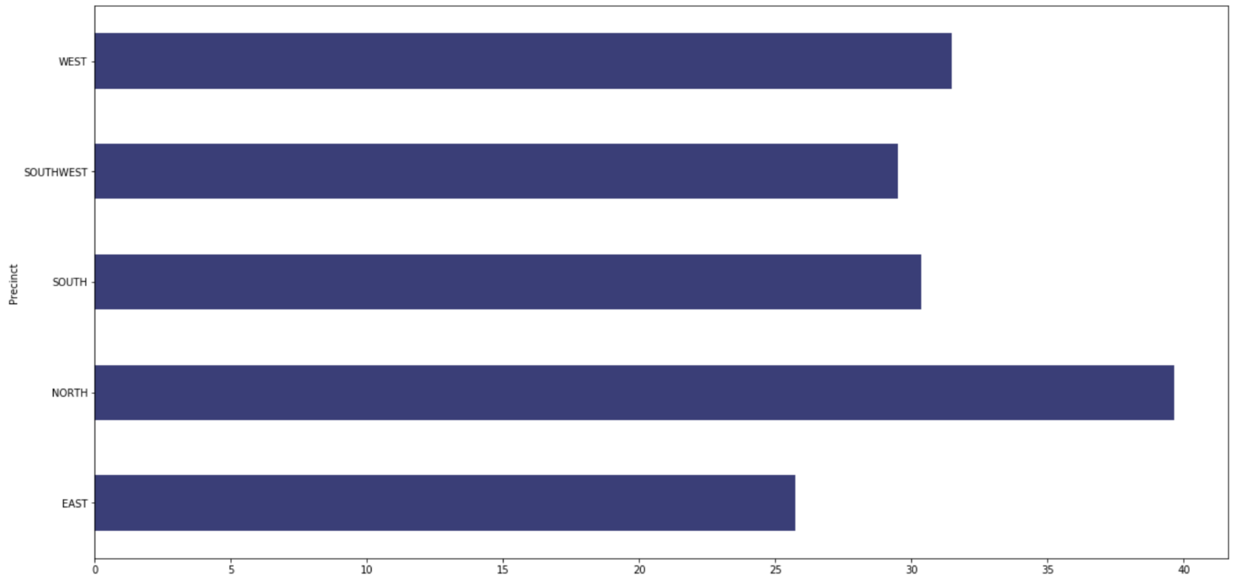

Figure 11 Average Response Time by Precinct, 2017

Grouping average response time by precinct shows that the North precinct has the highest average, almost 40 minutes, while the East precinct has the lowest, with just over 25.

Response Time Stats

Priority	count	mean	std	min	25%	50%	75%	max
1	36725.0	8.078502	8.989111	1.0	4.00	6.0	10.00	292.0
2	64058.0	30.953167	43.894056	1.0	7.00	14.0	36.00	927.0
3	32236.0	60.719630	77.417079	1.0	12.00	30.0	79.00	1257.0
4	1264.0	79.829905	98.727463	1.0	15.00	45.0	103.00	848.0
5	635.0	84.968504	90.234267	1.0	17.00	64.0	125.00	1134.0
6	60.0	287.766667	252.483755	1.0	84.75	196.0	455.00	947.0
7	2.0	21.500000	24.748737	4.0	12.75	21.5	30.25	39.0
9	1.0	5.000000	NaN	5.0	5.00	5.0	5.00	5.0

Figure 12 Response Time Statistics for 2017

The priority of the call has a great effect on the response time. A priority one call has an average of 8 minutes response time, while priority 2 is 31 minutes, and priority 3 double that at 60 minutes.

4. DATA PREPROCESSING

Missing Data

Using `.dropna()` to remove missing data dropped 35,435 rows, or about 0.95% of the data.

Date/Time Formats

The Original Time Queued and Arrived Time columns are stored in different date formats, and are stored as strings in the dataset:

Original Time Queued	08/14/2009 11:12:34 PM
Arrived Time	Aug 14 2009 11:13:55:000PM

I used a custom `to_datetime()` format to convert them both to DateTime objects so they would be more useful. I could then easily add more features such as minute, hour, day of week, and month that the call occurred.

Converting to DateTime objects also allowed me to more easily calculate the response time, which is the Arrived Time minus the Original Time Queued. I first calculated the response time in minutes, then rounded to the nearest 10 minutes to make it easier to predict.

Call Types

The call types (both Initial and Final) are stored as strings. They appear to be a group and a subgroup:

```
--TRAFFIC - MOVING VIOLATION
--TRAFFIC - MV COLLISION INVESTIGATION
--TRAFFIC - PARKING VIOL (EXCEPT ABANDONED CAR)
```

These three call types are all in the Traffic group. While most of the call types are stored this way (two dashes, the group, a dash, then the subgroup), not all are. I was unable to write a custom function that would separate all types into their groups.

5. MODELING

Crime Prediction

	Final Call Type	Beat	hour	day	month
72665	--PROWLER - TRESPASS	F2	12	2	2
72894	--PROWLER - TRESPASS	Q1	19	0	4
73541	--THEFT - SHOPLIFT	C1	19	2	4
73777	--ASSAULTS, OTHER	J2	19	2	4
74449	--DISTURBANCE - OTHER	J3	16	0	8

Figure 13 Crime Prediction Model

For crime prediction I created a model with four features, beat, hour, day, and month, to try and use classification to predict Final Call Type. I selected these features because they would be the only features someone would know if they wanted to predict crime in an area they were going to.

Response Time Prediction

	Priority	Initial Call Type	Precinct	Response
72894	2	160	4	40.0
73541	2	140	0	10.0
73777	2	140	1	40.0
74240	3	136	4	10.0
74502	2	140	0	40.0

Figure 14 Response Time Prediction Model

For response time I created a model with three features, priority, call type, and precinct, to try to predict response time.

Training

Both models were trained using a Decision Tree Classifier and a train/test split of 80%/20%.

6. EVALUATION

Classifier Selection

I tested the crime prediction model using several different classifiers. Decision Tree resulted in the highest accuracy, so I decided to use it for both the crime prediction and response time prediction models.

Model	Y_test == Y_pred
Naïve Bayes	0.1055
Linear Regression	0.0
K-Nearest Neighbors	0.0914
Decision Tree	0.1155

Figure 15 Classifier Accuracy

Feature Selection

For the response time model, I ran several tests with different features to find out which ones would result in the highest accuracy. Precision, Call Type, and Precinct resulted in the highest Recall and F-Scores, and only a 0.001 lower Precision score than the highest scoring feature set.

Features	Precision	Recall	F-Score
Call Type	0.270	0.366	0.265
Call Type, Precinct	0.286	0.368	0.269
Call Type, Precinct, Time	0.252	0.246	0.249
Priority	0.203	0.335	0.247
Priority, Precinct	0.227	0.353	0.253
Priority, Call Type, Precinct	0.285	0.369	0.276
Priority, Call Type	0.274	0.363	0.268

Figure 16 Feature Scores

In earlier testing, I had also tried including Sector, Beat, and other time features, but did not record the results. They all had lower scores. I thought that narrowing down the location to the beat level would improve the prediction, but that was not the case.

7. SERVICE

For both my models I put together APIs using flask and swagger. Both have health checks to make sure the service is running correctly.

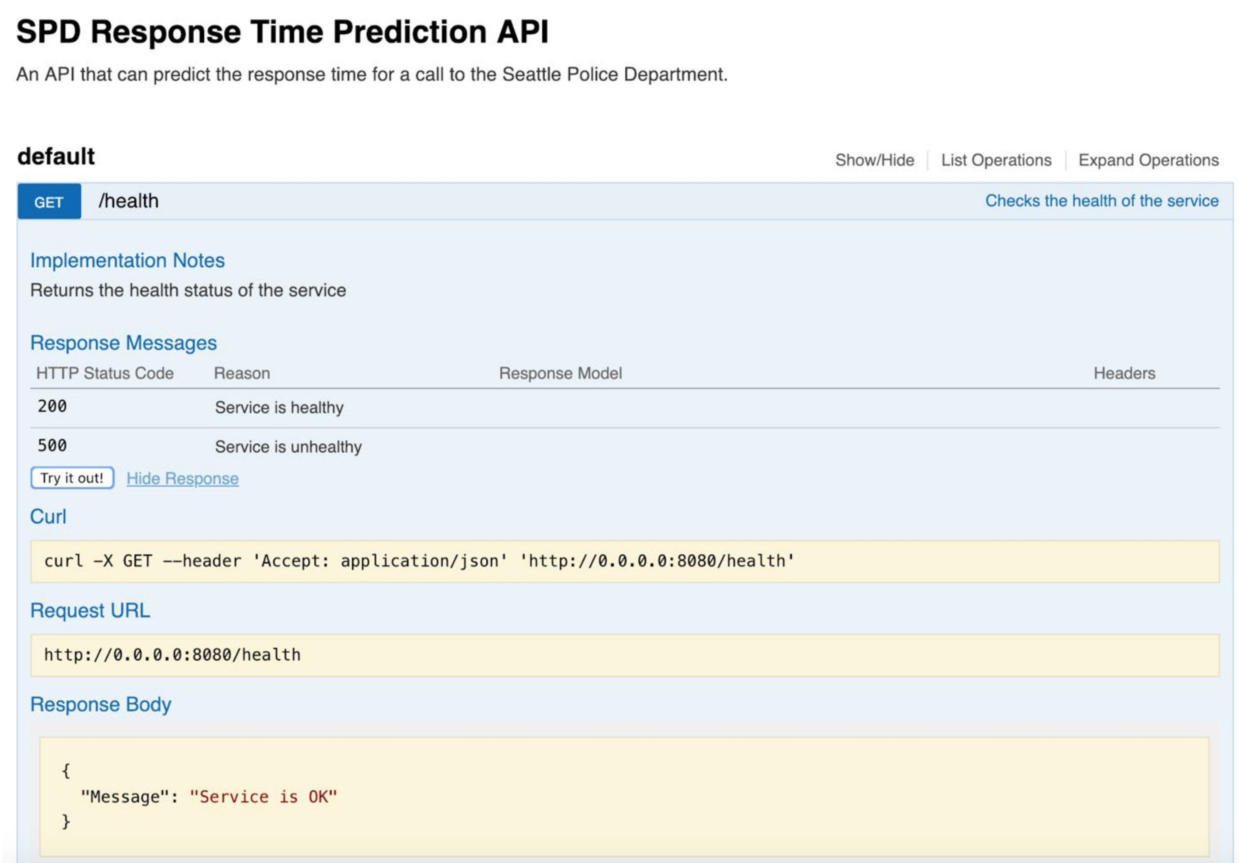

Figure 17 Response Time Prediction API health check

For the Response Time Predictor, users can input the priority of their call, the calltype they are reporting, and the precinct they are calling from. The service will then return the predicted response time.

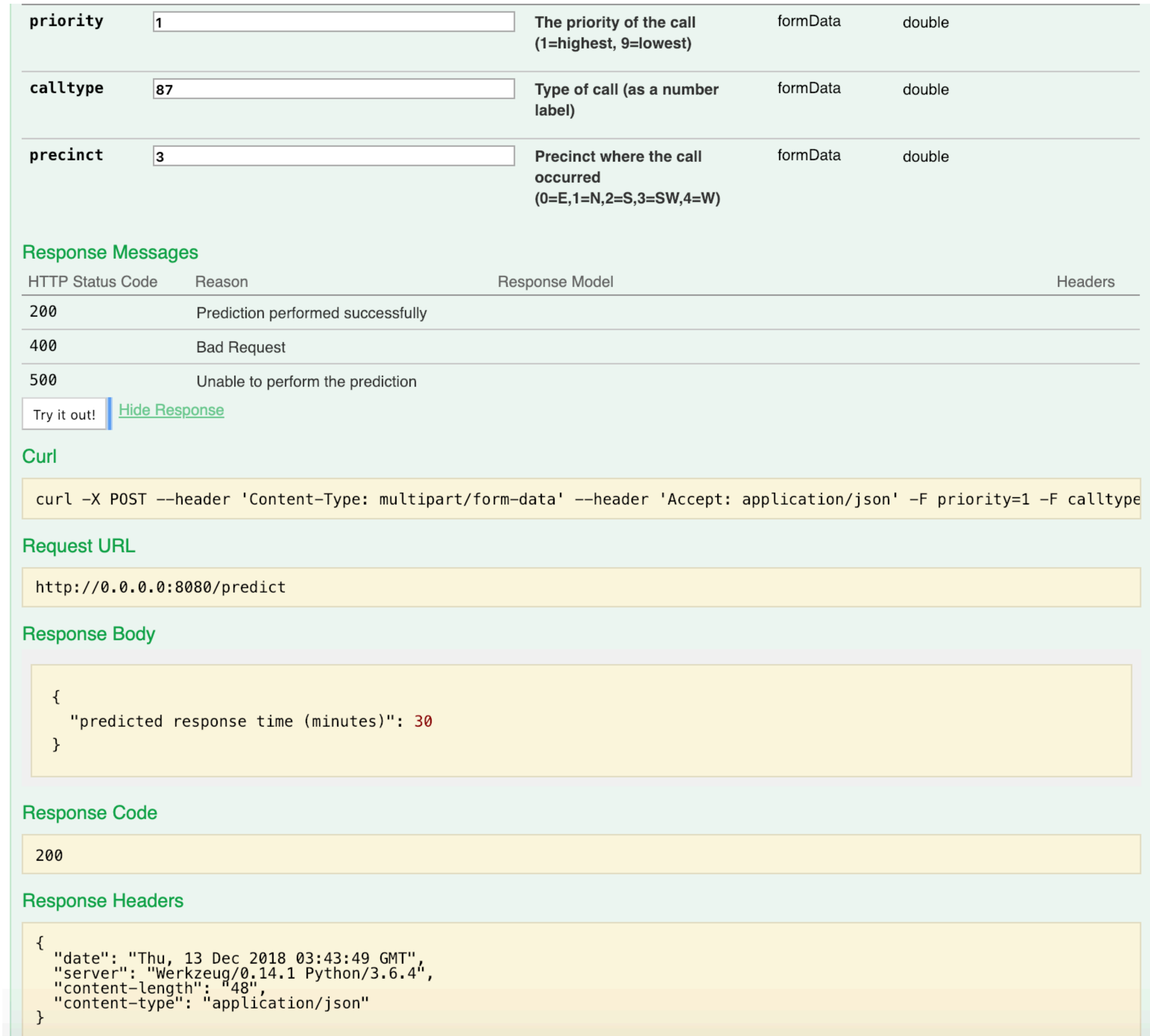

Figure 18 Response Time Prediction API

For the Crime Predictor API, users will enter the location, time, day, and month they are interested in getting a prediction for.

Parameters

Parameter	Value	Description	Parameter Type	Data Type
beat	14	**Location (SPD Beat)**	formData	double
hour	18	**Time (Hour from 0-23)**	formData	double
day	4	**Day (Monday = 0, Sunday = 6)**	formData	double
month	11	**Month (1-12)**	formData	double

Figure 19 Crime Prediction API Fields

And then the service will return the top five most likely crimes to occur at that location and time.

Response Body


```
{
  "predicted crimes": [
    "['--PERSON - FOUND PERSON']",
    "['--PERSON - A.W.O.L.']",
    "['--NARCOTICS - NARS REPORT']",
    "['WARRANT - MISD WARRANT PICKUP']",
    "['--ASSAULTS, OTHER']"
  ]
}
```


Figure 20 Crime Prediction API Response

8. CONCLUSION

Both models resulted in much lower accuracy than I had hoped, so I do not think that they would be very useful currently. If I had more time, I could make further improvements to the model by either finding a better data set, or gaining a much better understanding of this one, such as how calls are classified and prioritized by the communications center, what the different call types and groups mean, the difference between crimes (like THEFT – CAR PROWL and PROWLER – TRESPASS), and staffing levels/how officers respond to calls.

Overall, I think this dataset requires a much greater understanding of Seattle Police procedures and operations than I have to be useful for any kind of prediction. If I was able to improve it, however, I think it would be useful for people to know which areas of the city to avoid during certain times because of high crime, or to save time by knowing how long it will take for an officer to response to their call.

Some other things I could do with more time would be to improve the service. My model currently requires the user to know the number that the model's LabelEncoder has encoded the call type/crime to in order to predict. I would like to have found a way to have the user simply select it from a drop down list. A drop down list would also be better for the beat and precinct fields as well, or even being able to select from a map. I would also have like to put the service on AWS or used Docker.