

Final Project Proposal

Kevin Gifford, Mike Gruzynski, Jay Zuniga

W205 Section 3 - Summer 2017

Problem Statement

Anecdotal evidence suggests that many public transit stops in Seattle have a high incidence of crime, including some of those closest to the city's main tourist attractions. Our proposal would create an analytical tool to identify transit stops and routes with the highest incidence of crime in the immediate vicinity and explore if and how crime rates at these locations vary in connections with a number of external factors (season, weather, etc.). This analysis could be useful both to visitors and the city government; visitors could use this analysis to learn which transit locations should be avoided or used with caution, and the City could use the analysis to guide policing efforts.

Data Requirements

Primary data needs and sources for the analysis include the following:

- **Transit Stop and Route Data:** King County Metro (the primary transit provider in Seattle) makes data on its system, including stop locations and transit routes, freely available to developers.
<http://metro.kingcounty.gov/maps/developers-center/>
<http://www.kingcounty.gov/transportation/kcdot/MetroTransit/Developers.aspx>
- **Crime Data:** The City of Seattle publishes Seattle Police Department 911 call response data, including date/time, latitude/longitude location, a description of the incident, and the amount of time officers spent at the scene. The dataset includes records from 2008 to the present and is updated approximately every 4 hours.
<https://data.seattle.gov/Public-Safety/Seattle-Police-Department-911-Incident-Response/mzrk-e8qt>
- **Weather/Climate Data:** Seattle Airport (Sea-Tac or KSEA) publishes historical data on weather, including high/low temperature, rainfall, visibility, along with weather events and weather related daily data. This data set will be used as a secondary resource for the problem statement. The data set goes as far back as 1937, however 2008 to present data will be stored to pair up with crime data set timeline.
https://www.wunderground.com/history/airport/KSEA/2017/7/17/DailyHistory.html?req_city=KSEA&req_state=WA&req_statename=Washington&reqdb.zip=98158&reqdb.magic=4&reqdb.wmo=99999

- **Food Banks Data** - The City of Seattle publishes the [location of local food banks](#). This data set will be used to enhance our location data to see if there's any correlation with the location of existing food banks to crime incidents.
- **Cultural Spaces and LandMarks**- The City of Seattle publishes location of [cultural spaces](#) which includes museums, galleries and theaters as well as [landmarks](#). We will use this data to pinpoint some of the tourist destinations around town.

Data Cleaning and Transformation

Because the published datasets include a large number of records and information about topics that are not relevant to our particular analysis, we anticipate the following cleaning and transformation processing:

- The 911 response data must be scrubbed to remove incomplete records and false alarms. Additionally, specific crime descriptions must be aggregated into broader categories to be useful for our analysis.
- The King County Metro transit data include routes and facilities for the agency's entire service area, so the first cleaning operation will be to narrow the dataset to facilities and routes in Seattle only.
- The Weather/climate data contains many columns we are not primarily interested in, so cleaning operation will be used in order to narrow the dataset down to key weather information. In addition, there are some code-in values that require additional look up and possibility of missing data.
- Food Banks, Landmarks and Cultural Spaces - will be cleaned to ensure addresses and lat/long can be joined with main data sets, 911 response data and transit. We might also have to add tourist spots which are not considered as 'cultural' to ensure we have the full picture of tourist destinations.

Data Processing and Calculations

The primary processing activity will be to calculate the number of crimes reported within a specific distance of each transit stop (one-quarter or one-half mile) and normalize by volume of transit traffic or ridership. This spatial analysis component will be conducted through use of a Python spatial analysis library, such as cartopy, geopy, or pyqgis (or some combination thereof). Some additional spatial analysis may also be necessary to link transit stops and routes and make it possible to query and summarize the crime data by route, as well as stop.

Final Product

The final analysis product would consist of a series of summary tables highlighting those locations/routes with the highest current incidence of crime overall, as well as the following:

- Transit locations with highest year-over-year increases in crime rates;
- Locations/routes with highest rates for specific types of crimes, such as property thefts, assaults, drug use, or public intoxication;

- Give application users secondary information on transit based crime data based on input from weather scenarios
- Figure out if proximity to local food banks has any impact on crime and should be a consideration for transit routes.
- Pinpoint which tourist destinations have the most risk in terms of transit routes exposed to high crime.

APPENDIX CRIME DATA:

CATEGORIES WITH THEFT

- [2] AUTO THEFT
- [4] BURGLARY - UNOCCUPIED STRUCTURE ON RESIDENTIAL PROPERTY
- [9] FORGERY, BAD CHECKS
- [12] SHOPLIFT
- [16] TRESPASS
- [17] PROPERTY - FOUND (FOLLOW UP TO SPD CASE)
- [20] THEFT - CAR PROWL
- [22] BURGLARY - RESIDENTIAL, UNOCCUPIED
- [25] STRONG ARM ROBBERY
- [26] BURGLARY - COMMERCIAL
- [27] THEFT - MISCELLANEOUS
- [28] THEFT - AUTO ACCESSORIES
- [30] PROPERTY DESTRUCTION
- [31] ALARMS - RESIDENTIAL BURGLARY (FALSE)
- [33] ALARMS - COMMERCIAL BURGLARY (FALSE)
- [42] FRAUD (INCLUDING IDENTITY THEFT)
- [50] ARMED ROBBERY
- [51] ALARMS - VEHICLE (FALSE)
- [52] PROPERTY - MISSING
- [54] ALARMS - RESIDENTIAL PANIC (FALSE)
- [53] BURGLARY - RESIDENTIAL, OCCUPIED
- [55] STRONG ARM ROBBERY
- [56] BURGLARY - COMMERCIAL
- [57] ALARMS - COMMERCIAL PANIC (FALSE)
- [58] TRESPASS
- [61] BICYCLE THEFT
- [62] PROPERTY - FOUND (NON SPD GO#)
- [65] TRESPASS - PARKS EXCLUSION
- [66] LICENSE PLATE THEFT OR LOSS
- [67] AUTO THEFT AND RECOVERY
- [117] ALACAD - RESIDENTIAL BURGLARY (FALSE)
- [118] ALACAD - RESIDENTIAL PANIC (FALSE)
- [119] ALACAD - COMMERCIAL BURGLARY (FALSE)
- [120] ALACAD - VEHICLE (FALSE)
- [121] ALACAD - COMMERCIAL PANIC (FALSE)

LIGHT CRIME

[1] HARASSMENT, THREATS
[3] FIGHT DISTURBANCE
[6] MISCHIEF, NUISANCE COMPLAINTS
[8] DISTURBANCE, OTHER
[19] RECKLESS ENDANGERMENT, LITTERING, PARKS CODE VIOLATIONS
[35] HAZARDS
[36] NOISE DISTURBANCE
[39] MISDEMEANOR WARRANT SERVICE
[44] JUVENILE DISTURBANCE
[46] NOISE DISTURBANCE, RESIDENTIAL
[60] PEDESTRIAN VIOLATION
[63] ANIMALS - INJURED, DEAD, DANGEROUS
[70] PARKS EXCLUSION
[72] ANIMAL NOISE, STRAYS, BITES
[73] RECKLESS BURNING
[82] NOISE DISTURBANCE, RESIDENTIAL
[83] PARKS EXCLUSION
[95] PURSUIT
[108] CRISIS COMPLAINT - GENERAL
[110] CROWD MANAGEMENT (Stand by only)
[115] DEMONSTRATION MANAGEMENT (Control tactics used)
[116] CRISIS COMPLAINT - PICK-UP OR TRANSPORT

HEAVY CRIME

[32] FELONY WARRANT SERVICE
[34] CASUALTY (NON CRIMINAL/TRAFFIC) - MAN DOWN, SICK PERSONS, INJURED, DOA)
[40] MISSING PERSON
[43] ASSAULTS, OTHER
[77] GANG GRAFFITI
[78] DISTURBANCE, GANG RELATED
[80] DRIVE BY SHOOTING (NO INJURIES)
[81] ASSAULTS, FIREARM INVOLVED
[84] FOUND PERSON
[89] HOMICIDE
[96] ASSAULTS, GANG RELATED
[100] LOST PERSON

POSSIBLE CRIME

[5] SUSPICIOUS VEHICLE
[10] SUSPICIOUS PERSON
[13] MENTAL COMPLAINT
[15] ABANDONED VEHICLE
[45] SUSPICIOUS CIRCUMSTANCES - BUILDING (OPEN DOOR, ETC.)

[48] HARASSMENT, THREATS - BY TELEPHONE, WRITING
[68] PERSON WITH A GUN
[76] PROWLER
[79] PERSON WITH A WEAPON (NOT GUN)
[92] ASSAULTS, GANG RELATED
[94] MENTAL PERSON PICK-UP OR TRANSPORT
[107] AWOL
[113] AWOL

SEX, DRUGS, ALCOHOL
[14] LIQUOR VIOLATION - INTOXICATED PERSON
[37] NARCOTICS, OTHER
[38] DRIVING WHILE UNDER INFLUENCE (DUI)
[41] LIQUOR VIOLATION - ADULT
[47] LEWD CONDUCT
[49] NARCOTICS FOUND, RECOVERED
[59] PROSTITUTION
[64] NARCOTICS, DRUG TRAFFIC LOITERING
[69] NARCOTICS, DRUG TRAFFIC LOITERING
[71] LIQUOR VIOLATION - MINOR
[74] CASUALTY - DRUG RELATED (OVERDOSE, OTHER)
[85] VICE, OTHER
[86] NARCOTICS ACTIVITY REPORT
[87] PORNOGRAPHY
[88] NARCOTICS WARRANT SERVICE
[93] SEX OFFENDER - FAILURE TO REGISTER
[98] LIQUOR VIOLATIONS (BUSINESS)
[101] GAMBLING
[109] MARIJUANA PUBLIC USE (NOT DISPENSARY)
[111] SOAP (STAY OUT OF AREA OF PROSTITUTION) ORDER VIOLATION
[123] DOMESTIC SEX TRAFFICKING, ADULT
[124] DOMESTIC HUMAN TRAFFICKING, ADULT

VEHICLE BASED
[7] MOTORIST ASSIST
[11] ACCIDENT INVESTIGATION
[21] AUTO RECOVERY
[23] PARKING VIOLATION (EXCEPT ABANDONED VEHICLES)
[29] BLOCKING VEHICLE
[99] TRAFFIC CONTROL (SPECIAL EVENTS)
[114] MOTOR VEHICLE COLLISION
[125] TRAFFIC - COMMUNITY TRAFFIC COMPLAINT (CTC)
[126] TRAFFIC - BICYCLE VIOLATION

[127] TRAFFIC - SCHOOL ZONE ENFORCEMENT

MISCELLEANOUS (DELETE)

[24]

[75] HARBOR - WATER EMERGENCIES

[90] HARBOR - VESSEL ABANDONED

[91] HARBOR - DEBRIS, NAVIGATIONAL HAZARDS

[97] HARBOR - BOAT ACCIDENT

[102] HARBOR - MARINE FIRE

[103] HARBOR - ASSIST BOATER (NON EMERGENCY)

[104] HARBOR - CODE VIOLATION

[105] HARBOR - VESSEL THEFT AND RECOVERY

[106] HARBOR - VESSEL THEFT

[112] NULL

[122] HARBOR - BOATING UNDER THE INFLUENCE

[128] HARBOR - VESSEL RECOVERY