Data Mining Edmonton's 311 Service Requests

Kalvin Eng • November 30, 2017

https://goo.gl/cJ9SHN

Objective:

Better understand 311 requests made in the City of Edmonton by finding what items frequently occur together when reporting an issue

Overview

Data

- Schema, description, size
- Attributes
- Exploration

Method

- Frequent itemset mining
- Association rules mining

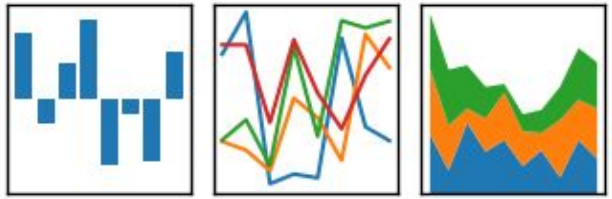
Preliminary Results

- Frequent itemsets

Data Preprocessing & Exploration

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

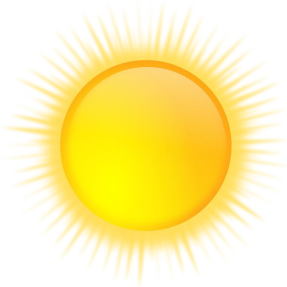


+ a b | e a u®

matplotlib

4 Datasets

311 Requests



Edmonton's Weather

Council Meetings

Alberta's Holidays

Merged by date

311 Data: Schema, Description, Size

Schema

1. Reference Number
2. Date Created
3. Date Closed
4. **Request Status**
5. **Status Detail**
6. **Service Category**
7. **Service Code**
8. **Business Unit**
9. **Neighbourhood**
10. **Community League**
11. **Ward**
12. Address
13. Lat
14. Long
15. Location
16. **Ticket Source**
17. **Calendar Year**
18. Count
19. Posse_Number
20. Transit_Ref_Number

Description

Edmonton's public 311 service requests from January 1, 2013 until November 17, 2017

Size

- Before cleaning: 294 097 rows x 20 columns
- After cleaning: 291 516 rows x 10 columns
  - Attributes kept are **bold** and underlined

Source

Edmonton Open Data

311 Data: Current Attributes

Request Status

Open or closed

Status Detail

Additional information about request status

E.g. duplicate request, vehicle gone upon arrival, citizen complied

Service Category

Type of activity or service request

E.g. pothole, snow & ice maintenance, drainage maintenance

Service Code

General description of service request

E.g. road, manhole cover, pothole, dead animal

Business Unit

The business area that is responsible for the service

E.g. roadway operations, drainage operations, traffic engineering

Ticket Source

Where request was initiated

E.g. telephone, mobile app, email, web form

311 Data: New Attributes

Days to Resolution

Amount of days it takes to resolve a request

Month Created

Month the request was created

Day of Week Created

Day of week request created

E.g. Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday

Month Closed

Month the request was closed

Day of Week Closed

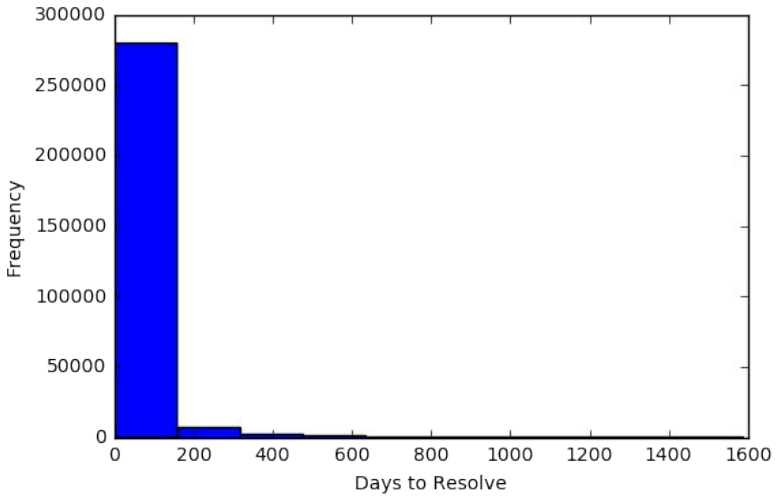
Day of week request closed

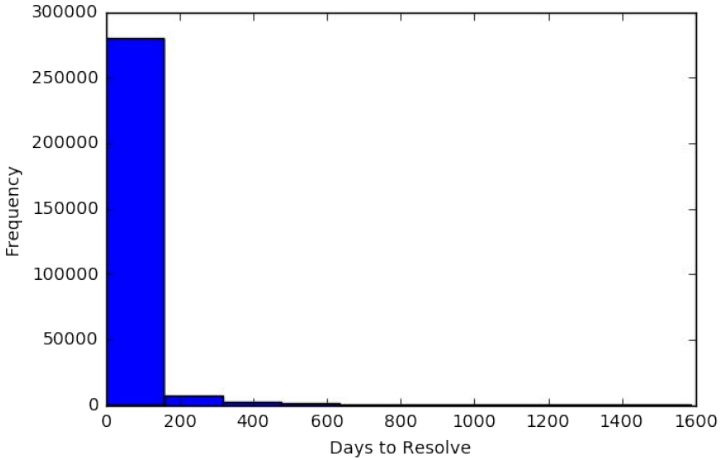
311 Data: Discretized Attributes

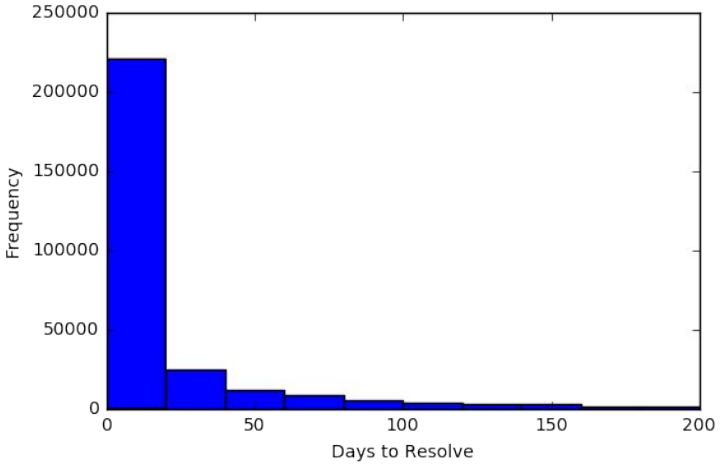
Days to Resolution

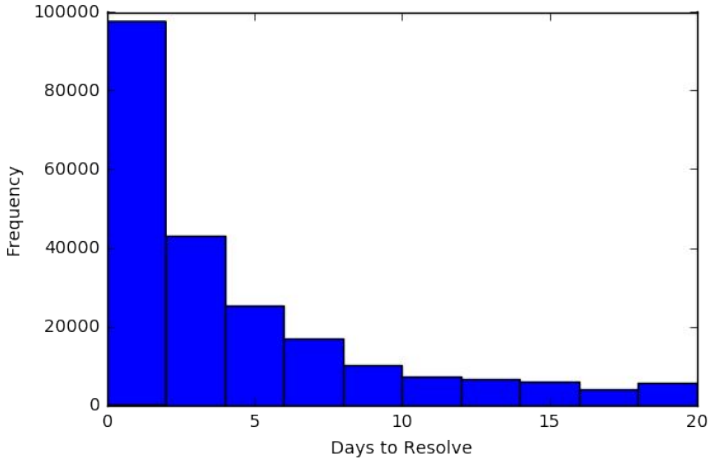
Discretized into 18 values:

- within {<2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 40, 60, 80+} days
- Based on the visualization of frequencies (next slide)



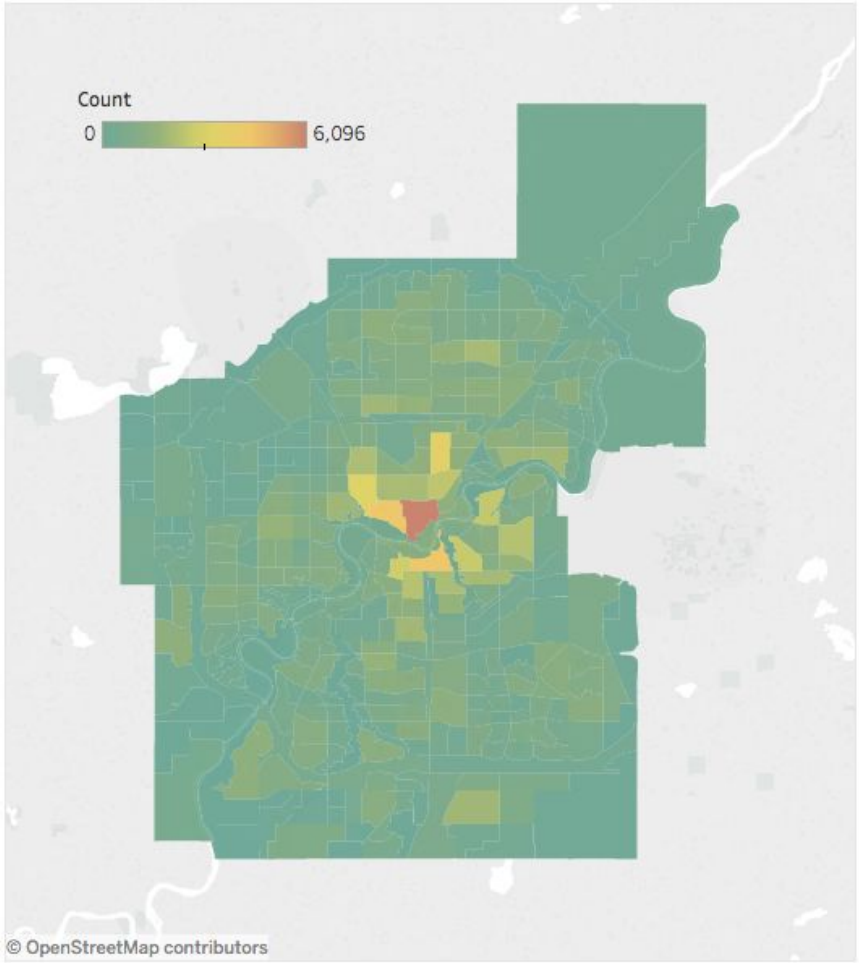






- Days to resolve are mostly < 200.
- Within the <200, it is usually resolved within first 20 days.
- Within the first 20 days, requests are usually resolved within first 2 days.

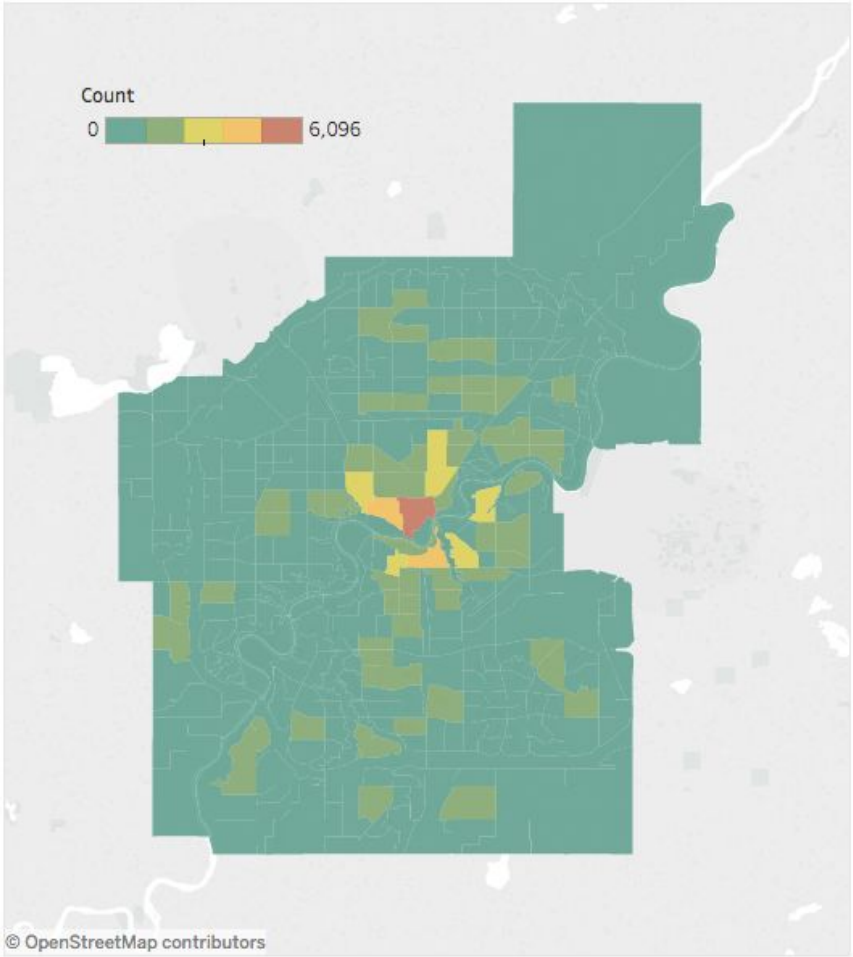
Reports Per Neighbourhood (2013-2017)



https://goo.gl/jERYYa

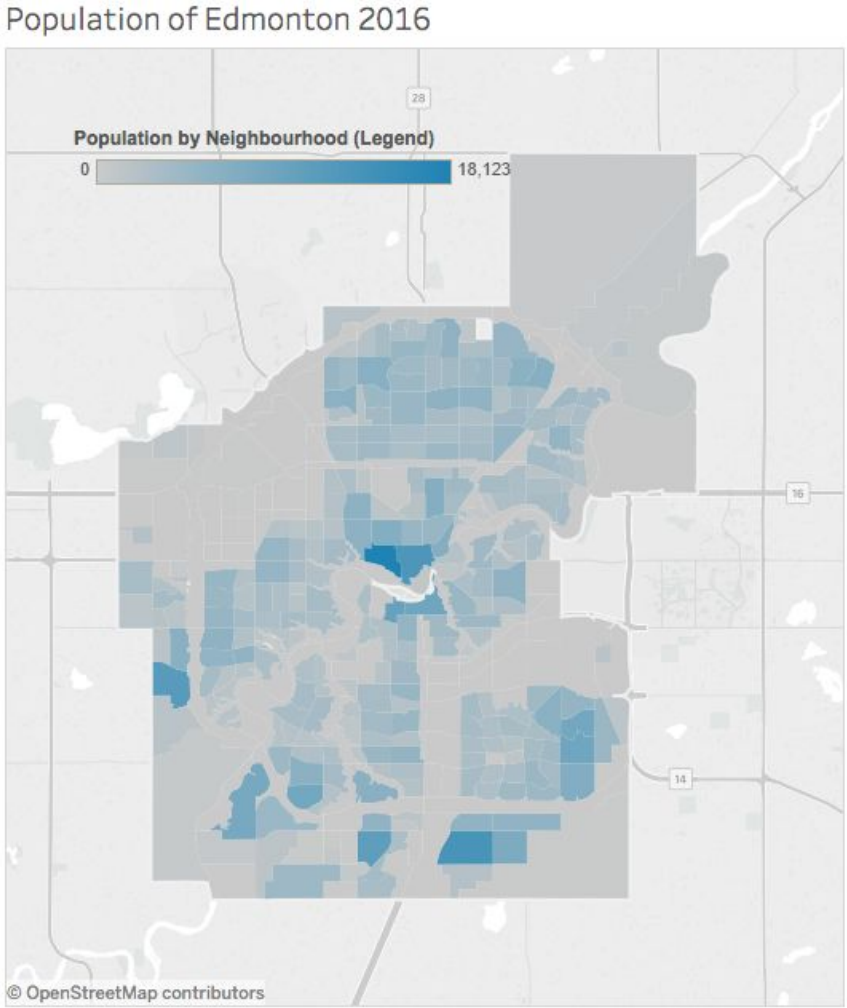
- Requests appear all over Edmonton
- Are there neighbourhoods where they more frequently appear than others?

Reports Per Neighbourhood (2013-2017) Quintiles



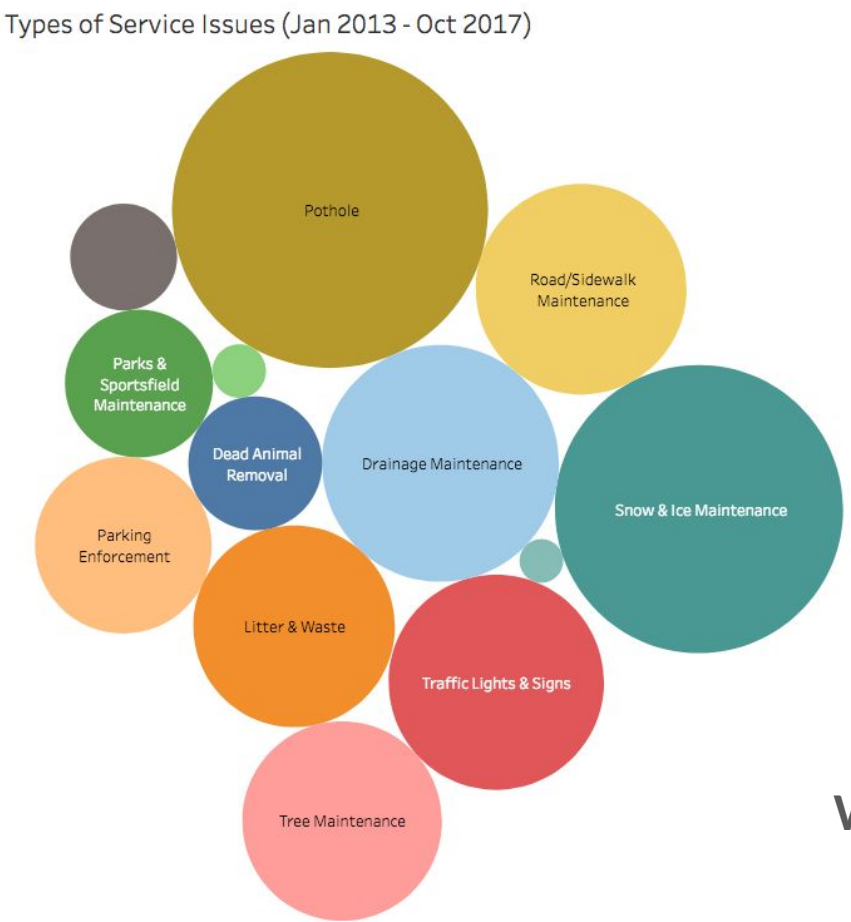
https://goo.gl/YmGF3v

- With quintiles, it is easier to visualize that requests occur more around central neighbourhoods and other pockets around central area
- Why are there more reports from these areas than others?



https://goo.gl/Rg2wB7

- In general, more populous areas have more issues reported
- What are the most frequent kinds of requests?



Service Category
- Dead Animal Removal
- Drainage Maintenance
- Litter & Waste
- Parking Enforcement
- Parks & Sportsfield Maintenance
- Pest Management
- Pothole
- Road/Sidewalk Maintenance
- Snow & Ice Maintenance
- Structure Maintenance
- Traffic Lights & Signs
- Tree Maintenance
- Vandalism/Graffiti

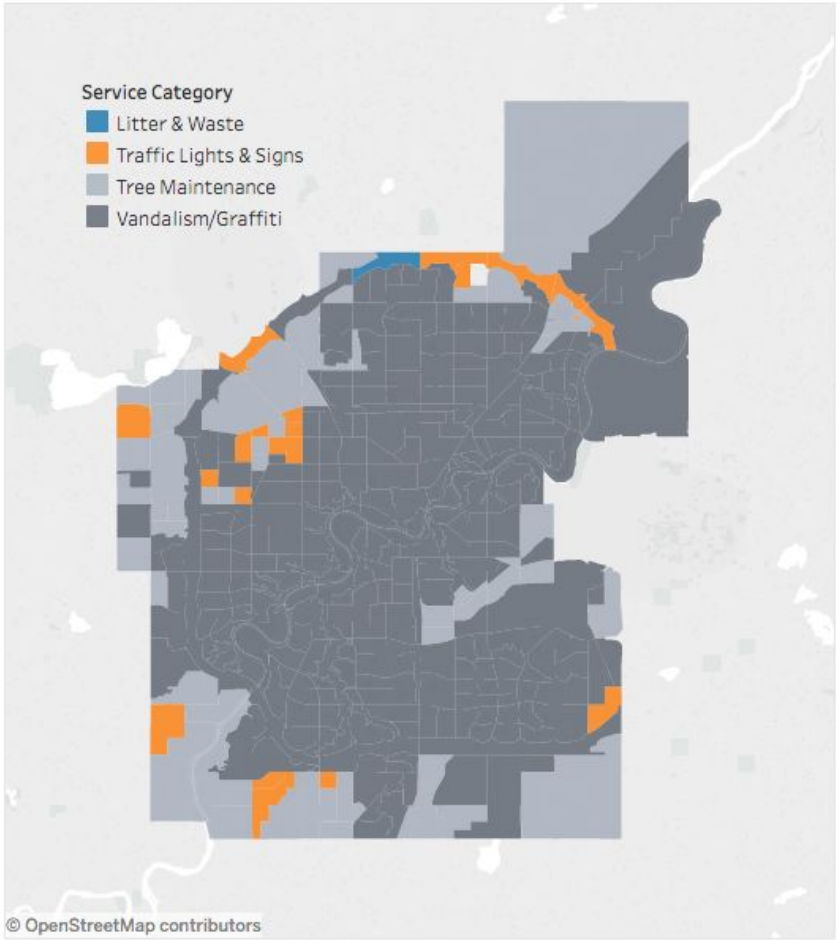
https://goo.gl/L5vMBS

Top 3 Service Requests:

1. Pothole
2. Snow & Ice Maintenance
3. Drainage

What kind of requests occur most frequently in neighbourhoods?

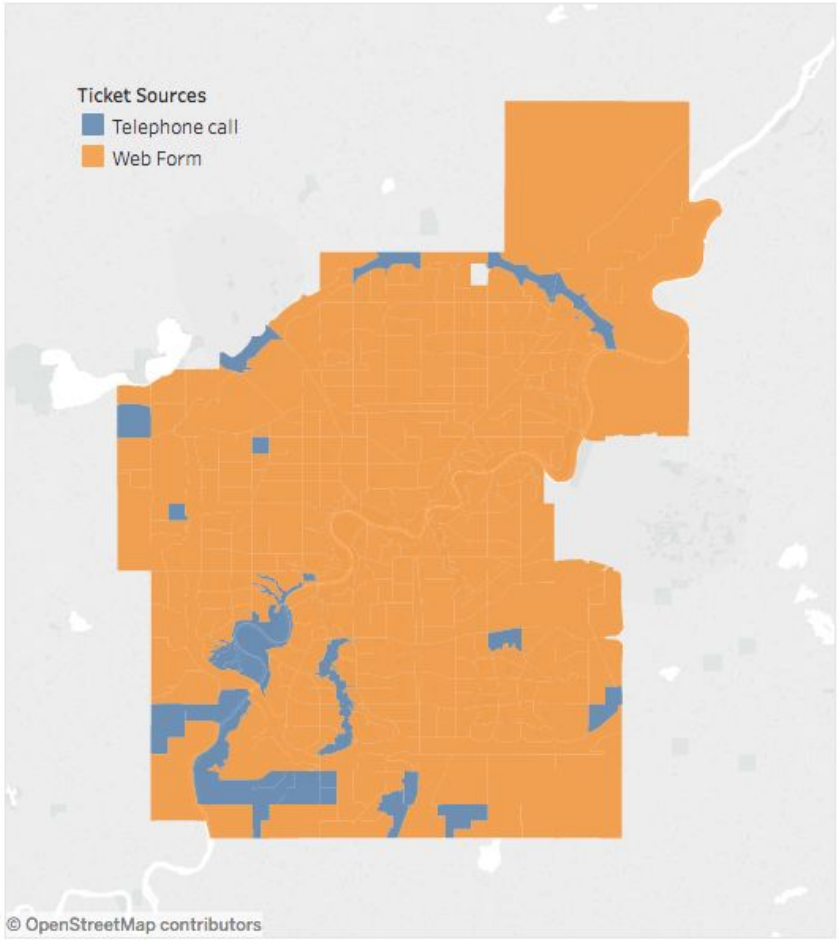
Most Commonly Reported Issues in Neighbourhoods



https://goo.gl/t5rXGQ

- The number of request types per neighbourhood is different than the most frequent overall
- What are some factors that occur most frequently for issues in a neighbourhood?

Most Common Source of 311 Requests



https://goo.gl/kY4YPG

- Most common sources of requests are telephone call and web form
- What kinds of sources occur most frequent with different request types?

Weather Data: Schema, Description, Size

Schema (after cleaning)

1. date
2. avg_hourly_temperature
3. min_windchill
4. avg_relative_humidity
5. avg_dew_point
6. max_wind_gust
7. wind_gust_dir_10s
8. avg_health_index
9. Heatdegdays
10. cooldegdays
11. growdegdays_5
12. growdegdays_10
13. precipitation
14. sunlight

Description

Edmonton's weather data from December 31, 2012 until November 16, 2017

Size

- Before cleaning: 1782 rows x 59 columns
- After cleaning: 1765 rows x 14 columns

Source

Canada

weatherstats.ca based on Environment and Climate Change Canada data

Weather Data: Attributes

avg_hourly_temperature

Average temperature of each day from 12 AM until 12 AM next day

min_windchill

Lowest wind chill value for the day (how cold the weather feels)

avg_relative_humidity

How hot weather feels on the day

avg_dew_point

Temperature to which air must be cooled to become saturated with water vapor; used to determine comfort

(max_wind_gust, wind_gust_dir_10s)

Maximum speed of wind and if it exceeds 29 km/h then the direction of the wind is also recorded

Weather Data: Attributes

avg_health_index

Index for the quality of air

(heatdegdays, cooldegdays)

Value used to estimate the heating and cooling requirements of buildings

(growdegdays_5, growdegdays_10)

Value used to determine the grow time in days for crops that grow at 5 and 10 degrees

precipitation

Amount of precipitation received

sunlight

Duration of sunlight for the day

Weather Data: Quantized Attributes

avg_hourly_temperature

Below freezing: {yes, no}

min_windchill

Risk of frostbite:
{low, moderate, high, very high, severe, extreme}

avg_relative_humidity

Comfortable: {yes, no_low, no_high}

avg_dew_point

Perception of comfortableness:
{severe, extreme, quite, somewhat, ok, comfortable, very, dry}

(max_wind_gust, wind_gust_dir_10s)

Windy: {yes, yes with direction, strong winds}

Direction: {N, NNE, NE, ENE, E, ESE, SE, SSE, S, SSW, SW, WSW, W, WNW, NW, NNW}

Weather Data: Quantized Attributes

avg_health_index

Air quality risk: {low, moderate, high, very high}

(heatdegdays, cooldegdays)

Heating day: {yes, no}
Cooling day: {yes, no}

(growdegdays_5, growdegdays_10)

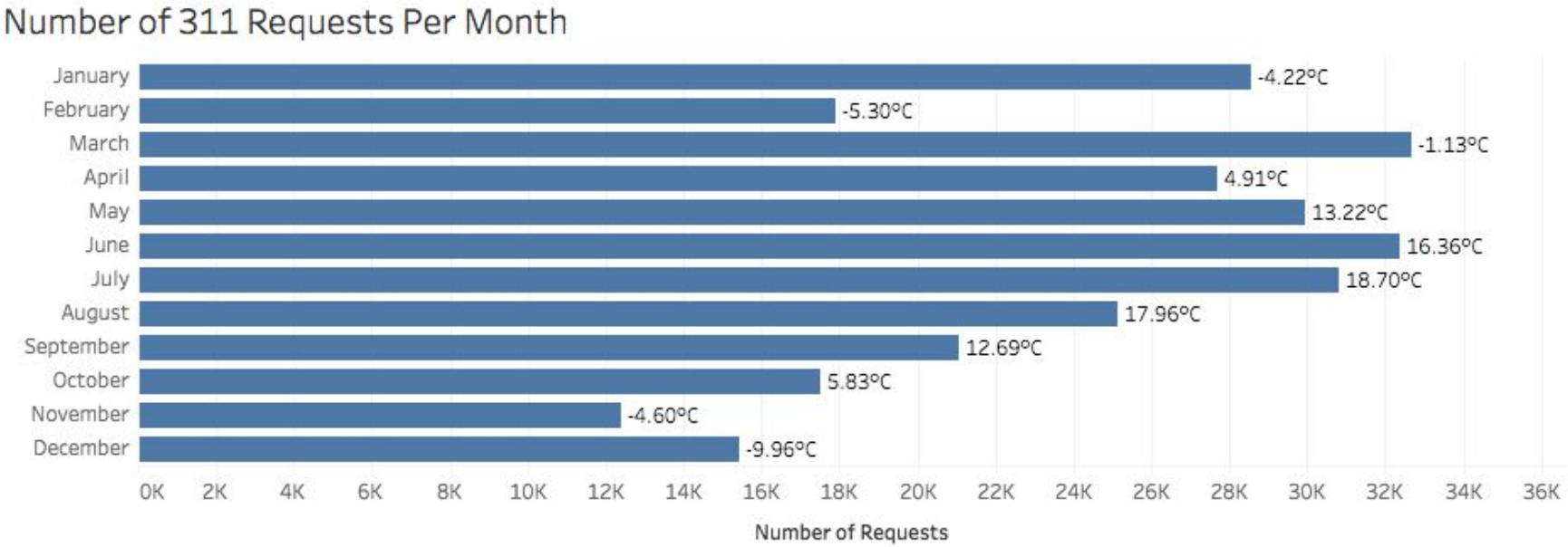
5 degree crops grow: {yes, no}
10 degree crops grow: {yes, no}

precipitation

Any precipitation: {yes, no}

sunlight

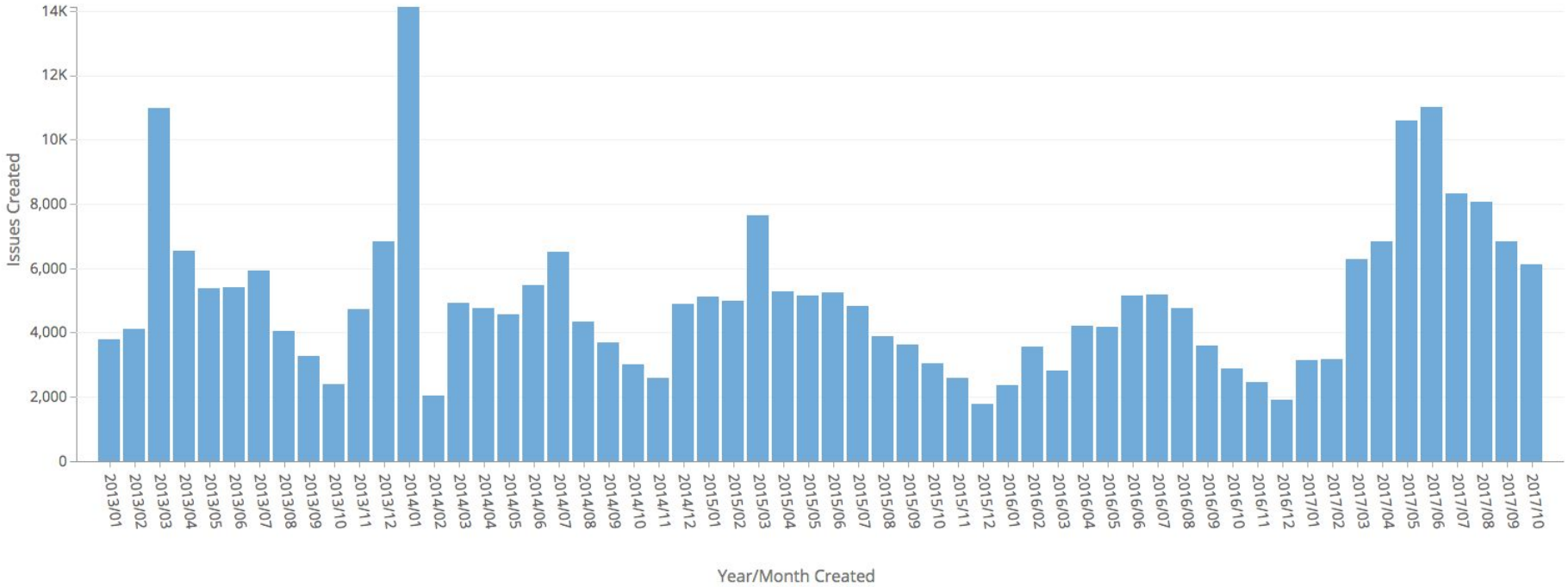
Sunlight Amount:
{ <8, 8, 9, 10 , 11, 12, 13, 14, 15, 16+}



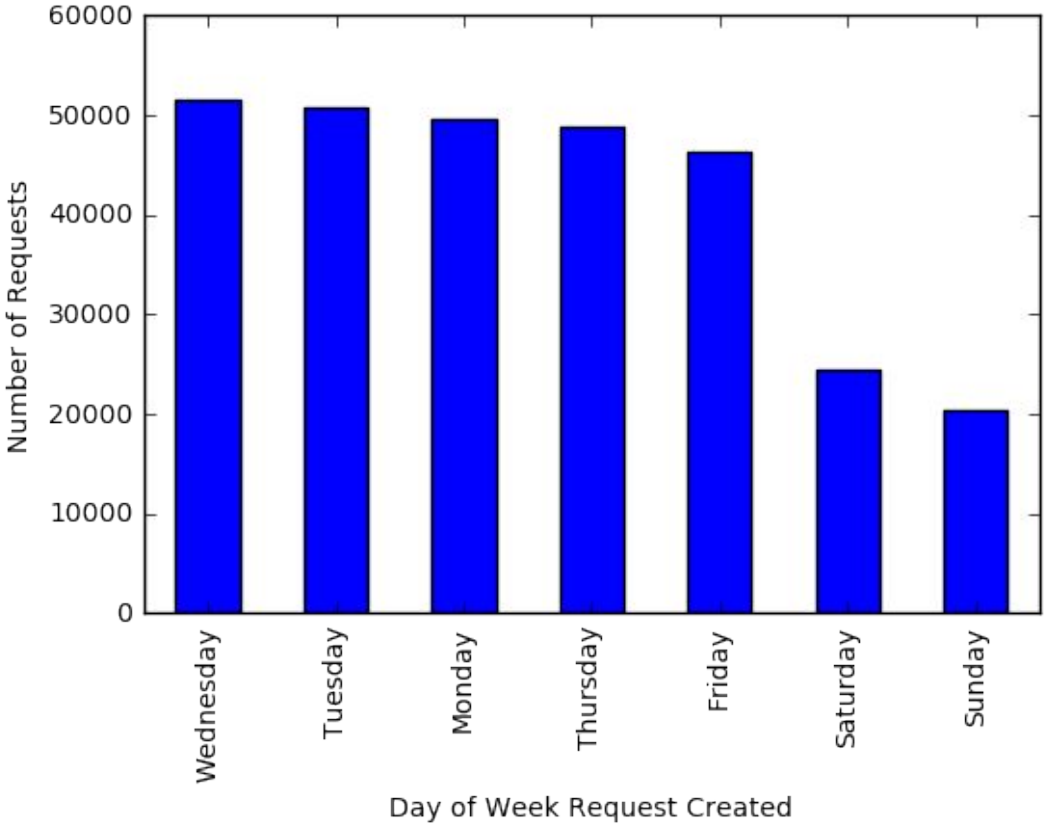
https://goo.gl/JrusDv

- Most requests occur in the warmer months
- What kinds of weather conditions occur most frequently when an issue is reported?

Issues Created over Months & Years



- Issues appear to spike in certain months
- What week of the day are requests most often created?



- Requests appear to be created most frequently during weekdays
- What about other types of days such as holidays or council meeting days?

Meeting Data: Schema, Description, Size

Schema

1. MEETING_ID
2. **MEETING_TYPE**
3. RECORD_TYPE
4. **MEETING_DATE**
5. MEETING_TIME
6. MEETING_LOCATION

Description

City of Edmonton's council and committee meetings planned meetings from 2013 until 2017

Size

- Before cleaning: 719 rows x 6 columns
- After cleaning: 719 rows x 2 columns
  - Attributes kept are **bold** and underlined

Source

Edmonton Open Data

Meeting Data: Attributes

Meeting Type

Types of meetings such as council meetings, hearings and committee meetings

28 unique values are categorized into:

- Council Meetings
- Hearings
- Committee Meetings

Meeting Date

Date meeting has occurred

Holiday Data: Description

Description

Statutory holidays in Alberta from 2013-2017

- Dates generated using algorithms implemented in python-holidays library

Source

https://github.com/ryanss/python-holidays/blob/master/holidays.py

Merged Data: Description, Size

Description

Transactions contain items relating to:

- Issue types
- Location
- Time
- Weather
- Events

Size: Before binarization

343637 rows × 44 columns

Size: After binarization

343637 rows × 946 columns

Merged Data: Schema

1. Request Status
2. Status Detail
3. Service Category
4. Service Code
5. Business Unit
6. Neighbourhood
7. Community League
8. Ward
9. Ticket Source
10. Calendar Year
11. Days to Resolution
12. Month Created
13. Month Closed
14. Day of Week Request Created
15. Day of Week Request Closed
16. holidays_created
17. avg_hourly_temperature_created
18. min_windchill_created
19. avg_relative_humidity_created
20. avg_dew_point_created
21. max_wind_gust_created
22. wind_gust_dir_10s_created
23. avg_health_index_created
24. heatdegdays_created
25. cooldegdays_created
26. growdegdays_5_created
27. growdegdays_10_created
28. precipitation_created
29. Sunlight_created
30. avg_hourly_temperature_closed
31. min_windchill_closed
32. avg_relative_humidity_closed
33. avg_dew_point_closed
34. max_wind_gust_closed
35. wind_gust_dir_10s_closed
36. avg_health_index_closed
37. heatdegdays_closed
38. cooldegdays_closed
39. growdegdays_5_closed
40. growdegdays_10_closed
41. precipitation_closed
42. sunlight_closed
43. MEETING_TYPE_created
44. MEETING_TYPE_closed

Market Basket Data

Diapers, Beer

Apples, Oranges

Apples, Oranges, Beer

311 Request Data

Pothole, Downtown, Windy

Tree Issue, Downtown

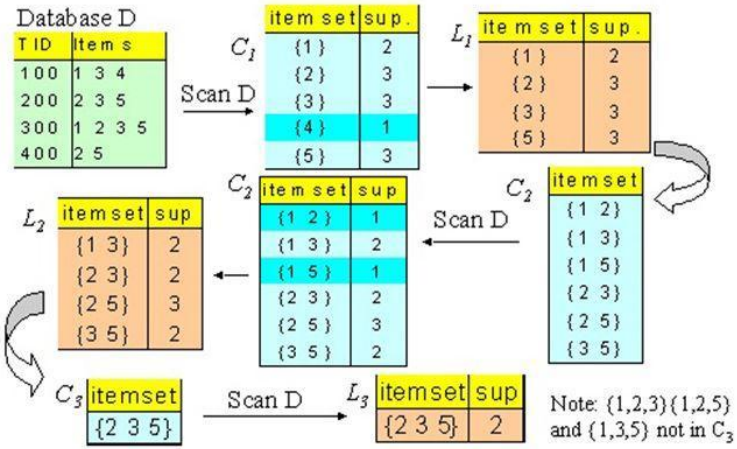
Pothole, Belgravia, Windy

Edmonton's 311 data is analogous to market basket data

There are transactions with different items

Method: **Frequent Itemset Mining**

Finding sets of items that appear in (are related to) many of the same transactions

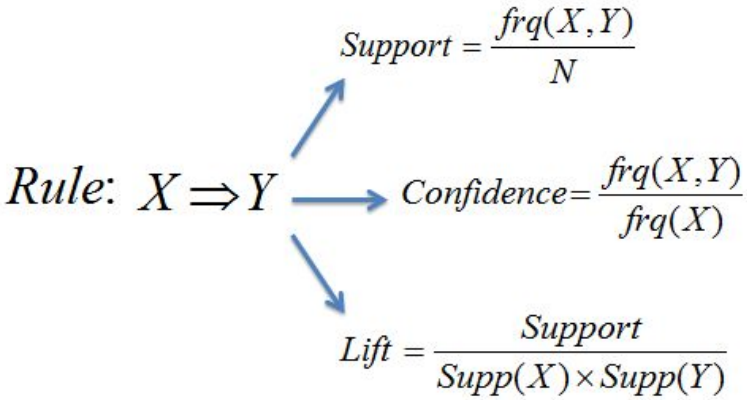


Source: http://auburnbigdata.blogspot.ca/2013/03/apriori-algorithm.html

Method: **Association Rule Mining**

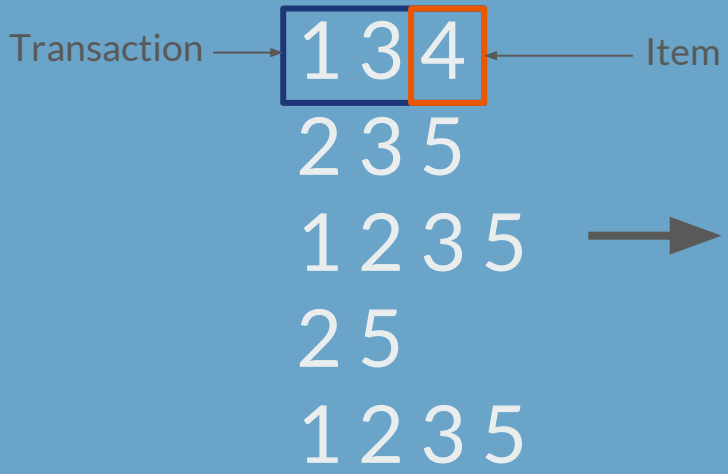
Finding interesting associations among frequent itemsets

Support: how frequent itemset appears
Confidence: how often rule is found to be true
Lift: measure interestingness



Sources: http://chem-eng.utoronto.ca/~datamining/dmc/association_rules.htm
https://en.wikipedia.org/wiki/Association_rule_learning

Preliminary Results



FPGrowth

Transactions count from database : 343637

Min Support: 25%

Max memory usage: 958.90232849121 mb

Frequent itemsets count : 204667

Total time ~ 9419 ms

Preliminary Results: **Frequent Itemsets**

When request is created: {uncomfortable relative humidity (>50%), feels dry, indoor heating is needed}

- Support: 228804 / 343637 = 66.58%
- These conditions occur often in the winter
- Be better prepared to receive requests from the different channels (telephone, mobile app, e-mail) when these conditions occur

Preliminary Results: **Frequent Itemsets**

When a request is created for the Roadway Operations department:
{uncomfortable relative humidity (>50%), feels dry, indoor heating is needed}

- Support: 133803 / 343637 = 38.94 %
- These conditions often occur in the winter
- Roadway Operations should be prepared to deal with more requests in the winter

Preliminary Results: **Frequent Itemsets**

When a request is closed within 2 days: {indoor heating is needed}

- Support: 122247 / 343637 = 35.57%
- Outside conditions are colder
- Since a good portion of requests occur in colder weather, should prepare suitable equipment for crews that need to work in colder weather

Preliminary Results: **Frequent Itemsets**

When a telephone call is made for a request: {moderate risk of frostbite, uncomfortable relative humidity (>50%), feels dry, indoor heating is needed}

- Support: 91849 / 343637 = 26.73 %
- This is a good portion out of the total telephone calls (227 178)
    - ~40% of telephone calls
- Be better prepared to receive telephone calls on these days

Conclusion

- Understand Edmonton's 311 request data using events that occur at certain times as items
- Clean and visually explore the datasets
- Preliminary results demonstrate that some weather conditions occur more frequently with certain types of requests