# Detecting Trends

A trend is a change in the normal number of complaints we'd expect to see. A trend is defined as

A consistent upward or downward movement, out of the ordinary range. For an upward trend: each count must be equal to or greater than previous count. For a downward trend: each count must be less than or equal to the previous count.

Machine generated alternative text:
Sub_product:Other type of mortgage Company:WELLS FARGO & COMPANY Product:Mortgage 
6 
5 
4 
3 
2 
1 
03 
raw complaint counts 
complaint counts: 7 day average 
avg count 
avg count in trend window 
trend window 
10 
17 
24 
May 
2017 

This plot shows the number of complaints for one issue for the 30 days **before** the end date (2017-05-01). Here the x-axis = day of the month, and the y axis = number of complaints per day for the theme. For the first 19 days of the month there are no complaints for this theme, then starting on the 19th day there are increasing amounts of complaints: a trend.

The last 7 days is highlighted in pale green, this is the "short window" relating to the recent past: most of the time the trends that are happening now are the most important.  The blue line is the "raw" number of complaints per day: it is highly variable as the amount of complaints vary between weekdays and weekends. To make it easier to understand we also plot a green line by taking the average over the previous 7 days for each day. This removes the noise caused by weekdays/weekends. The pale red horizontal dotted line is the average number of complaints in the whole 30-day period. The pale cyan horizontal dotted line is the average number of complaints in the last 7-day period. This clearly shows that the average number of complaints in the last 7 days is approximately three times the number of complaints over the whole month.

**Examples of Trends**

It can be useful to look at real examples of trends to understand some of the issues about discovering these and displaying these:

Machine generated alternative text:
Example Upward Trend 
25 
20 
15 
10 
5 
03 
Jul 
2017 
complaint counts 
trend start date 
complaint counts: 7 day average 
avg count 
avg count in trend window 
trend window 
10 
17 
24 
31 
Date 

This trend is noisy and doesn’t increase in a smooth fashion.

Machine generated alternative text:
25 
20 
15 
10 
5 
Jul 
2017 
Example Upward Trend 
complaint counts 
trend start date 
complaint counts: 7 day average 
avg count 
avg count in trend window 
trend window 
03 
10 
17 
Date 
24 
31 

This trend is short lived and represents a single ‘spike’ in complaints.Machine generated alternative text:
Sub_product:Other type of mortgage Company:WELLS FARGO & COMPANY Product:Mortgage 
6 
5 
4 
3 
2 
1 
03 
raw complaint counts 
complaint counts: 7 day average 
avg count 
avg count in trend window 
trend window 
10 
17 
24 
May 
2017 

This is a trend that overall is increasing but might be showing different scores due to the day of the week: fewer people make complaints at the weekends (see the two low scores in the green part on days 30 and 31)

Machine generated alternative text:
PRODUCT:O&GEO:CA&THEME:Loan servicing, payments, escrow account 
counts 
trend for counts 7 day average 
23 
30 
13 
20 

This is a more complicated trend. The counts are increasing (trending) on the right-hand side but they were also high at the start of the month, and low in the middle. Possible causes:

* This is just a noisy complaint theme and has naturally variability.
* This trend has a monthly periodicity: It might be that people complain at the same time every month – e.g. a few days monthly bills are sent

 It’s necessary to look at the trend over a longer timeframe to find the answer.

## How Trend Detection Works

The first step of the trend detection algorithm groups all the complaints according to their characteristics, counts the daily totals, and then identifies which of these are trending.

All complaints are classified by the Natural Language Classifier and Natural Language Understanding models, and the results are combined with customer data. This gives a number of variables for all complaints: category, process, product, theme, sub-product, age, zip code, state, gender etc. The exact list of fields available depends on the customer data set. The user can determine which of these available fields are of interest for trend detection by updating the config file.

The trend detection algorithm derives all the combinations of these fields:

1. complaint\_category
2. complaint\_category + process
3. complaint\_category + process + zip code
4. …

This forms a hierarchy of complaint trends: the top level are very generic but more specific details appear as columns are added.

The trend detection algorithm then counts the complaints in each combination for all values in each field:

1. complaint\_category = ‘burdensome request’, …
2. complaint\_category = ‘burdensome request’ + process = ‘Mortgage Application”,
3. complaint\_category = ‘burdensome request’ + process = ‘Mortgage Application” + zip code = 9410
4. …

The complaints are counted for each of these combinations, and evaluated to see if a trend can be found. Complaints are counted over periods of 30, 60 and 90 days. Trend details and trend timelines are viewed in the Surveillance Insights – Complaints Analytics Dashboard:

## Trend Detection Rule:

Given a timeline of complaint counts, it is flagged as trending if:

* There is a general increase or decrease in the last 7 days (short window). See section below for definition of a monotonic increase.
* There is at least one day where the count is unexpectedly large or small given the preceding counts. In technical terms, at least one day has exceeded the expected Poisson distribution score. See section below on Poisson distribution
* There is a clearly defined start and end to the trend within the short window. For the start-date this means there must be at least one day greater than the average count over the long window. For the end-date this means there must be one day with a count greater than the start date (for an increasing trend.)
* The end-count must be greater than the start-count for an increasing trend. The end-count must be less than the start-count for a decreasing trend.

## Trend Risk Score

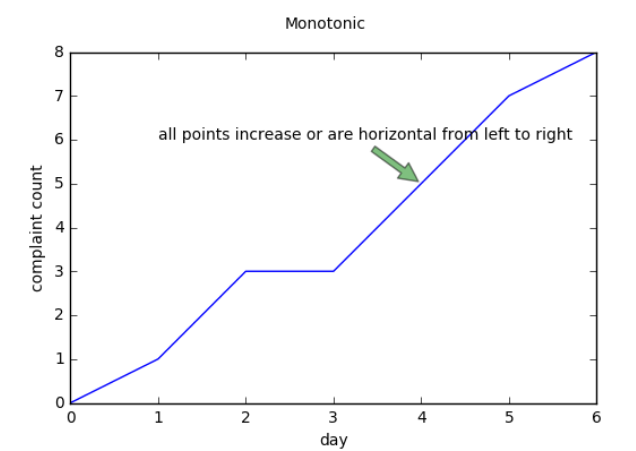
For each timeline that is detected as a trend: the risk score is calculated which expresses how significant each trend is as a percentage score.

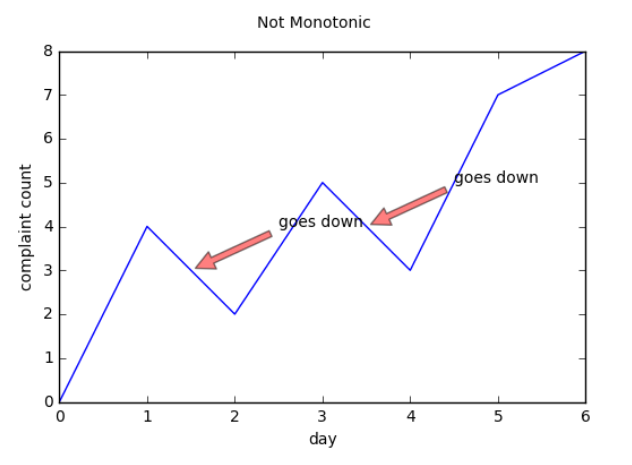
Trends may vary widely in terms of size, speed of growth and other factors. It is difficult to find a “one-size fits all” method for scoring when high-level complaints may have tens or hundreds of thousands of complaints and increase slowly, while a low-level trend may only contain tens of complaints but be increasing rapidly. Early detection of growing trends is a key use case so the trend detection algorithm should highlight both.

The scoring method is designed to allow size and speed of increase and decrease and consistency of increase or decrease to all be considered.

The risk score is calculated from 3 components:

1. The number of complaints since the trend was detected: trends with bigger increases or falls in complaints score higher.
2. The gradient of the increase or decrease: trends which increase or fall more steeply score higher
3. How "monotonic" the fall or increase is. Trends are often noisy i.e. even if there is an overall increase some points are less than the preceding count. A monotonic increase is one where all counts are either greater than or equal to the preceding days complaint count. Trends where the count only goes in one direction score higher because they are more consistent.





A score is created from these three factors between 0 and 85% (the maximum). Some tuning may be required for each data set.

**Risk Score Formula:**

For each trend, these are calculated:

1. Total number of complaints in trend
2. gradient of trend
3. monotone score – does the trend only increase or decrease

Each of these has a weight factor configurable by the user giving the relative importance of each as a fraction adding to 1. For example:

1. complaints weight = 0.5
2. gradient weight = 0.3
3. monotone weight = 0.2

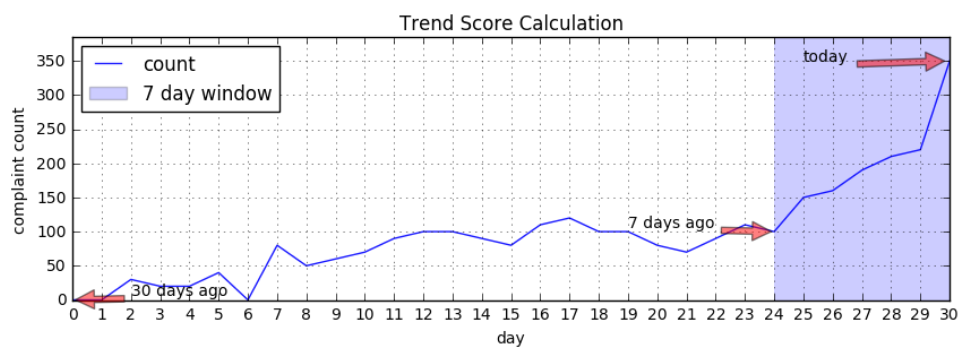
The score is calculated as:

Score = **)**complaints weight \*( )gradient weight\* monotone monotone weight

Where:

* logs are calculated using base 10
* All 3 main terms are raised to the power of the corresponding weight
* The results are multiplied together

**Example Trend Score Calculation**



Here a trend has been identified in the last 7 days:

Number of complaints in last 7 days = 1380 : ( 100 + 150+ 160+ 190+ 210+ 220+ 350)

* Gradient = (final count – initial count)/ 7 days = 350 -100/7 = 35.7
* Monotone score = 100% (all counts from 24th to 30th are greater than preceding day count)

The user has determined these weights:

1. complaints weight = 0.5
2. gradient weight = 0.4
3. monotone weight = 0.1

which means that the total number of complaints is the most important factor, increasing gradient is almost as important, but they don’t care so much if a trend is noisy.

The user has also determined these factors:

* For the number of complaints: all trends will be ranked on a scale between 0 and 10,000:

**Max complaints = 10,000**

* For the gradient: all trends will be ranked on a scale between 0 and 1000. A gradient of 1000 signifies that the fastest growing trend the user has seen or expects to see will grow at a rate of 1000 additional counts per day:

**Max gradient = 1000**

Score = **)**complaints weight \*( )gradient weight\* monotone monotone weight

Score = )**0.5**\*( ) **0.4** \* 1 **0.1**

Score = )**0.5**\*( ) **0.4** \* 1 **0.**1 = 0.7

Score is then scaled to have a maximum of 85%:

**Final Score** = 0.7 \* 85% **= 60%**

### Earliest Trend Detection Date

If a trend is detected, the earliest trend detection date is designated as the first day on which the complaints counts exceeds the long-term average. The earliest trend detected date is shown with a red dot in the UI, see below. If the trend continues increasing over a prolonged period the earliest trend detection date is preserved as the first date on which the count exceeded the long-term average

Machine generated alternative text:
4.6 
O 
3.6 
3.0 
o 
2.6 
2.0 
1.0 
Jul og 
Jul 16 
Jul 23 
Dates 
Jul 30 
Aug 06 

### Trending 7 days and 30 days Calculation

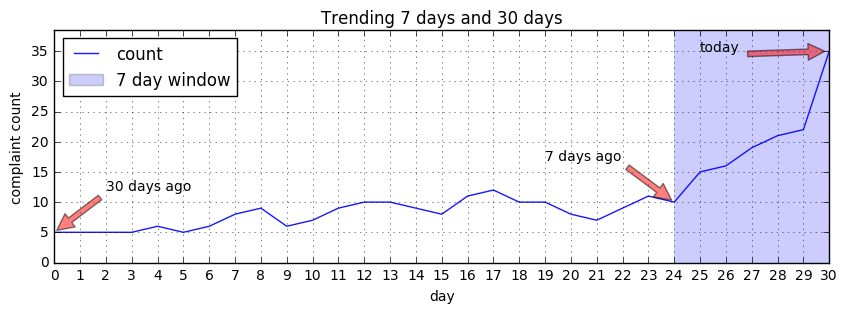
This describes the method for calculating the trending 7 days and trending 30 days scores on the complaints UI. These are displayed in the UI:

Machine generated alternative text:
TRENDING LAST 7 DAYS 
+0.0% 
TRENDING LAST 30 DAYS 
-50.0% 

This calculates how much the trend has increased over a period as a percentage of the original amount:

(increase or decrease / original amount) as a percent

**Example**



**Trend over the last 7 days:**

amount 'today' = 35

original amount (7 days ago) = 10

increase over the last 7 days = 35 - 10 = 25

**Trending last 7 days = 25/10 = 250%**

“a 200% increase”

**Trend over the last 30 days:**

amount 'today' = 35

original amount (30 days ago) = 5

increase over the last 30 days = 35 - 5 = 30

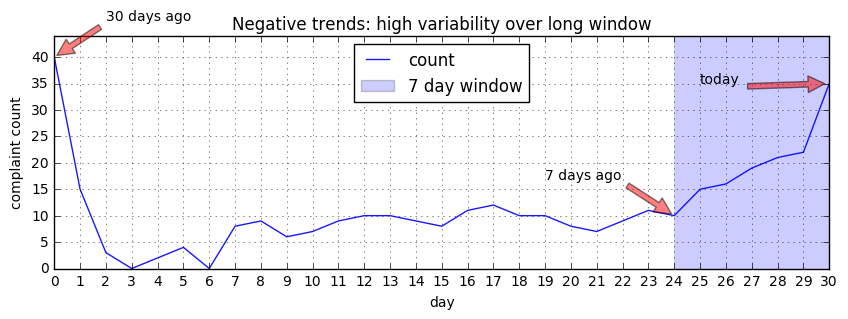
**Trending last 30 days = 30/5 = 600%**

“a 600 % increase

**Negative Scores**

It’s possible for the trend detection to find an increasing trend and still have negative or decreasing trend scores. There are two examples below.

**Increasing Trends with Negative trending 7 days and trending 30 days scores**

****

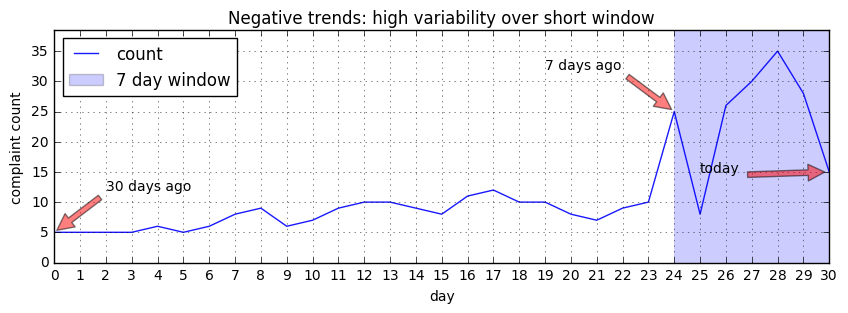
**Trend over the last 7 days:**

amount 'today' = 35  
original amount (7 days ago) = 10  
increase over the last 7 days = 35 - 10 = 25  
Trending last 7 days = 25/10 = 250%

**Trend over the last 30 days:**

amount 'today' = 35  
original amount (30 days ago) = 40  
increase over the last 30 days = 35 - 40 = -5

Trending last 30 days = -5/40 = -12.50%

****

**Trend over the last 7 days:**

amount 'today' = 15  
original amount (7 days ago) = 25  
increase over the last 7 days = 15 - 25 = -10  
Trending last 7 days = -10/25 = -40%

**Trend over the last 30 days:**

amount 'today' = 15  
original amount (30 days ago) = 5  
increase over the last 30 days = 15 - 5 = 10

Trending last 30 days = 10/5 = 200%

# Trend Statistics:

Trends may vary widely in terms of size, speed of growth and other factors. It is recognised that it is difficult to find a “one-size fits all” method. For example, some of the high-level complaints have tens or hundreds of thousands of complaints and may increase slowly, while a low-level trend may only contain tens of complaints. Early detection of growing trends is a key use case so the trend detection algorithm should highlight both. Given this, a number of statistics for each trend are calculated so that there is flexibility in updated the trend detection method in the future if required:

|  |  |
| --- | --- |
| Statistic | Description |
| count | total number of complaints in the long window |
| count in trend window | total number of complaints in the short window |
| mean count | average number of complaints per day for the long window (this includes the days in the short window) |
| mean count in trend window | average number of complaints per day for the short window |
| mean ratio | (mean count in the short window)/ (mean count in the long window) |
| trendinglast7days | increase in % over the last 7 days using the raw counts, calculated as:  (count on the last day - count on 7th day before this)/ count on 7th day before this |
| trendinglast30days | increase in % over the last 30 days using the using the raw counts, calculated as:  (count on the last day - count on 30th day before this)/ count on 30th day before this |
| Trend type id | Type of trend: 'increasing'=1, ’decreasing'=2, 'no trend'=3 |
| Mk trend | True if there is a trend (i.e. if trend type id = 1 or 2) |
| Mk score p-value | The confidence score (p-value) that the Mann-Kendall algorithm has discovered a trend:  Score < .05 = low, Score < 0.1 = high, Score < 0.001 = very high confidence |
| poisson\_scores\_max | The highest Poisson score in the trend window |
| poisson\_above\_thresh\_count\_inc | how many days in the short window had a Poisson score greater than the Poisson threshold for an increasing trend |
| poisson\_above\_thresh\_count\_dec | how many days in the short window had a Poisson score greater than the Poisson threshold for a decreasing trend |

# Mann-Kendall score

The purpose of the Mann-Kendall score is to statistically assess if there is a monotonic upward or downward trend of the variable of interest over time. A monotonic upward (or downward) trend means that the variable consistently increases (or decreases) through time, but the trend may or may not be linear.

# Point-by-point Poisson Model

The Poisson distribution model is a statistical model used to evaluate how unusual a point in a time-series is when compared to previous points. It is used in the trend analysis model to help identify sudden increases (or decreases) in the number of complaints seen that may help indicate the presence of a trend.

The Poisson distribution describes the probability of observing a count of some quantity, when many sources have individually low probabilities of contributing to the count. The point-by-point Poisson model uses the previous point in a time series to define the expectation for the current point, and gives us the unlikeliness of a count of complaints given the previous count.