## Sentiment and Topic Modelling US Consumer Financial Protection Bureau

# Documentation

This web application displays the sentiment and topic analysis of complaints received by the **US Consumer Financial Protection Bureau**.

The panel on the left lets you filter the details on whether the customer was *compensated* before or after the complaint was lodged and the type of *product* the complaint was about.

**NOTE:** The filters apply to all the tabs.

The tabs area display the various views pertaining to the complaints.

Contents

[Filters **Error! Bookmark not defined.**](#_Toc493610095)

[Sentiment Analysis 3](#_Toc493610096)

[Topic Modeling 5](#_Toc493610097)

[Topic Modeling 7](#_Toc493610098)

## User Interface

### Filters

The filter panel consists of an area to filter on the data set details and the section to update the topic analysis panels.

The filter on whether a consumer was compensated and by specific product will change as soon as the selected label has changed its value.

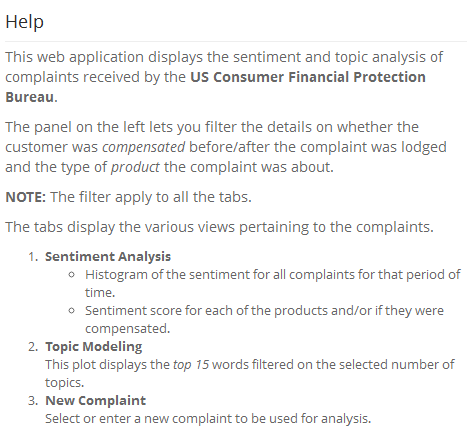
The “Both/All” buttons will select all the values for that filter and the “None” label will deselect all the values.

If the compensation filter is empty (all labels are unselected) then the complaints will only be filtered by products.

If all the products are unselected, then the complaints will only be filtered by the compensation status of the complaint.

To change the number of topics the complaints will be divided into, first select the value (2-5) then click on the “Update” button.

Note that the number of topics filter will only show on tabs that the filter influences, meaning it will not be visible on the “Sentiment Analysis” tab because the filter is not used.



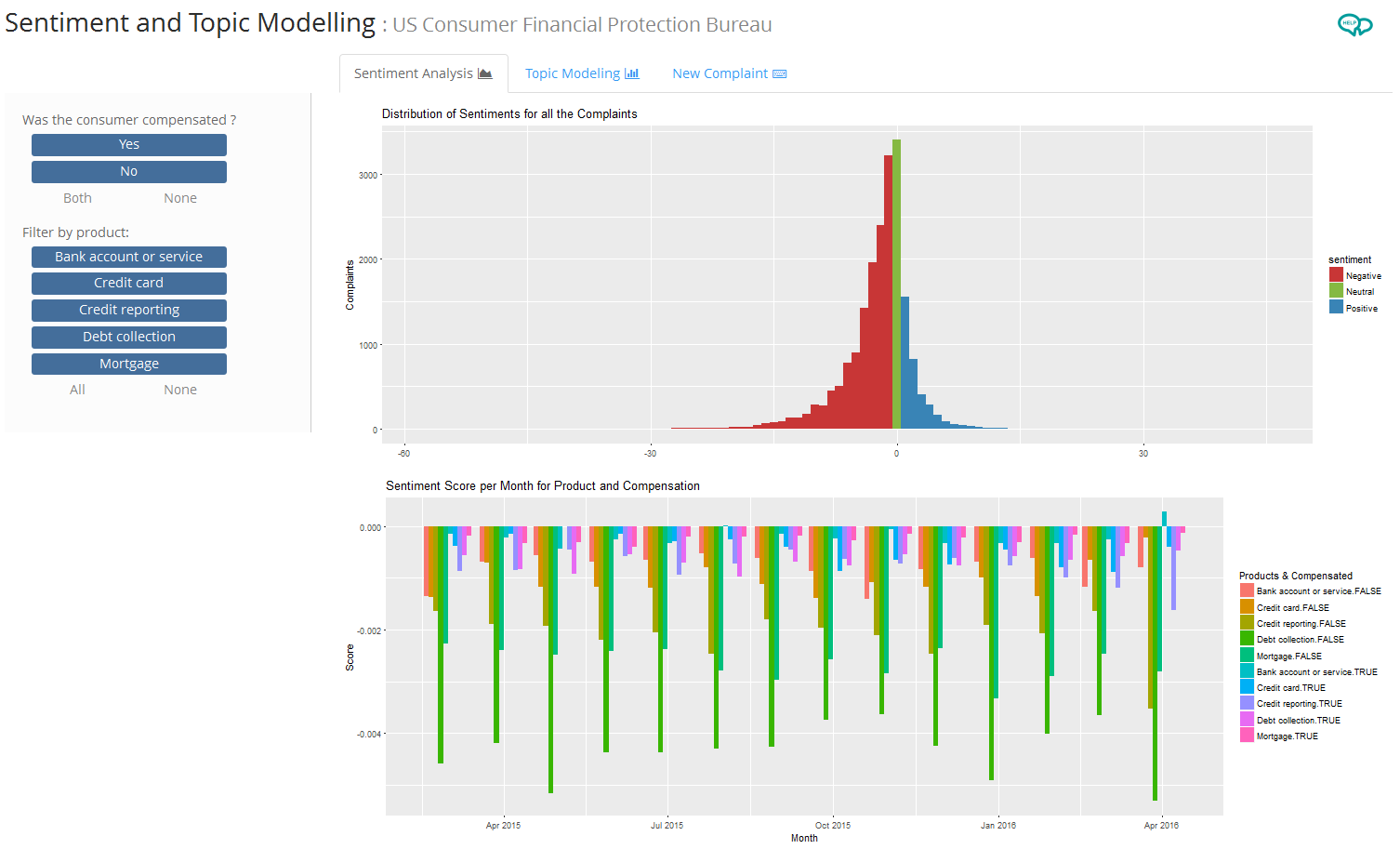
### Help

The help overlay gives a quick overview of the page, the filters and tabs.

The overlay is accessed by clicking on the help icon.



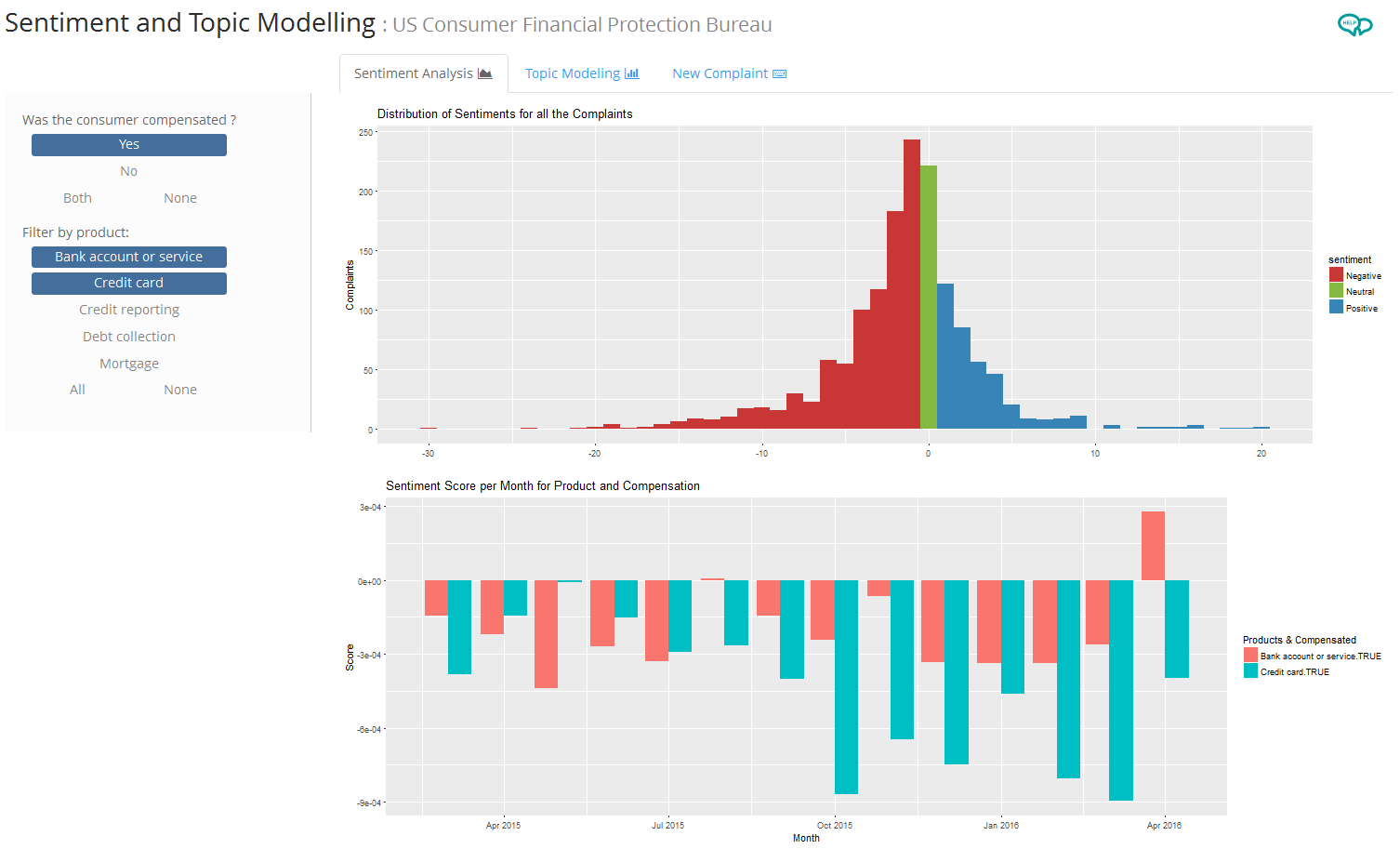
## Sentiment Analysis



This selection of filters display the complete dataset with both values for compensation and all the products.

The histogram show at the top displays the number of complaints based on their sentiment.

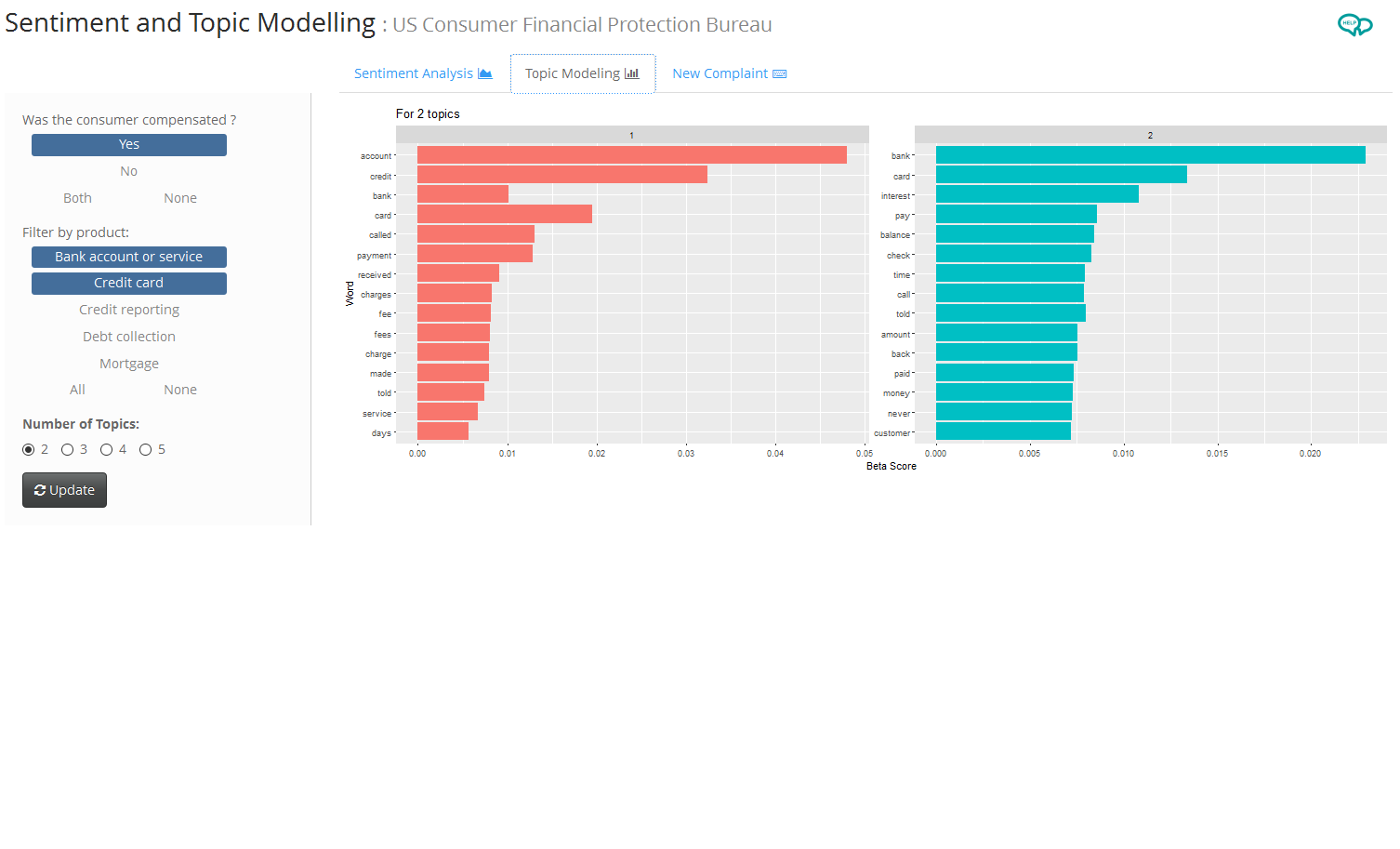
The bottom plot displays the sentiment score for the product with the compensation value; and grouped by month. The score is calculated on the overall difference of sentiment and the number of words in the complaint.



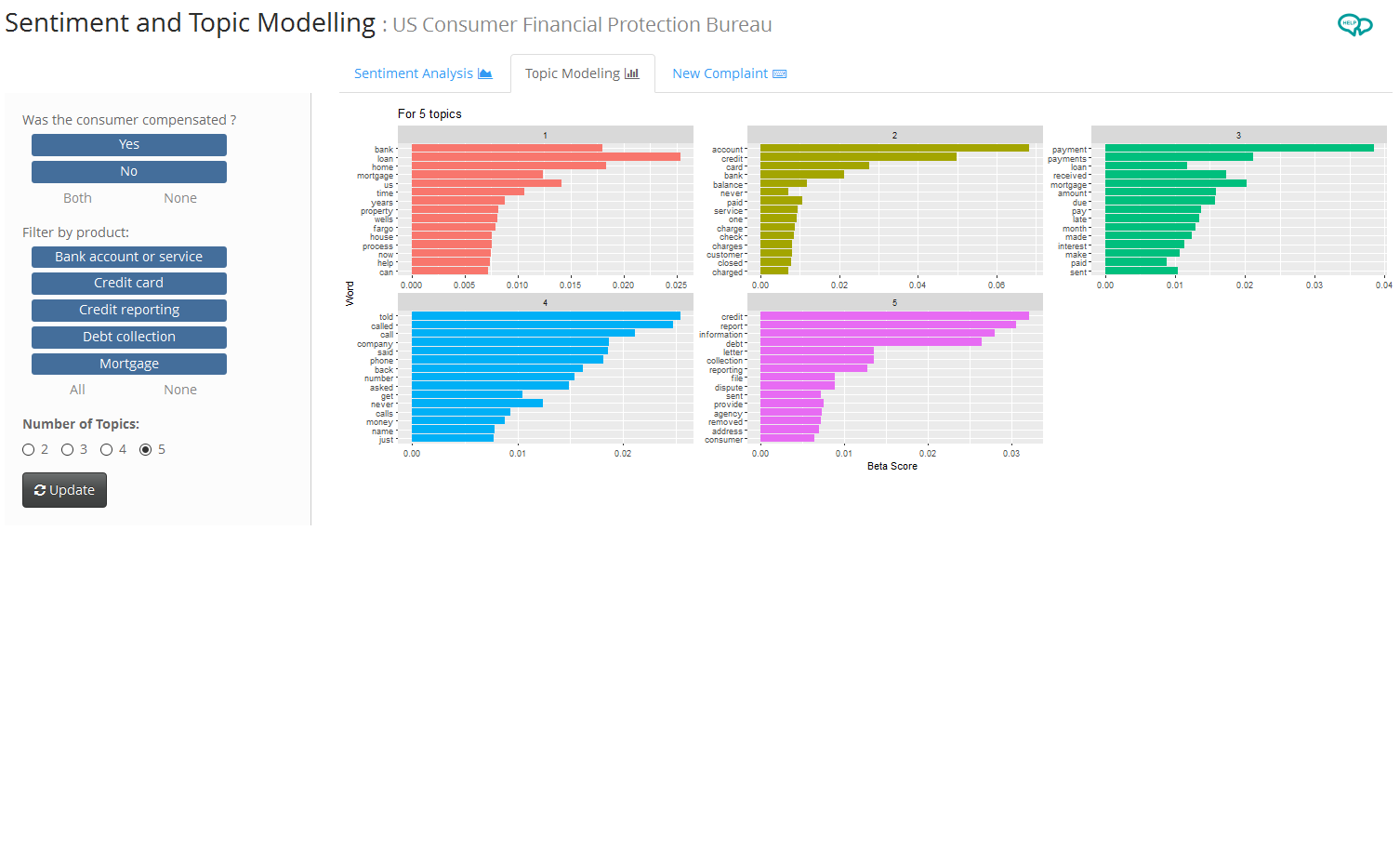
In this selection, the complaints are filtered by only those who received compensation, with the products being either “Bank account or service” or “Credit Card”.

The histogram shifted slightly to include more positive sentiments and in the sentiment score graph there are a couple of months where the score of both product is close to zero (total score / number of words). April 2016 even has a positive value for “Bank account or service”.

## Topic Modeling



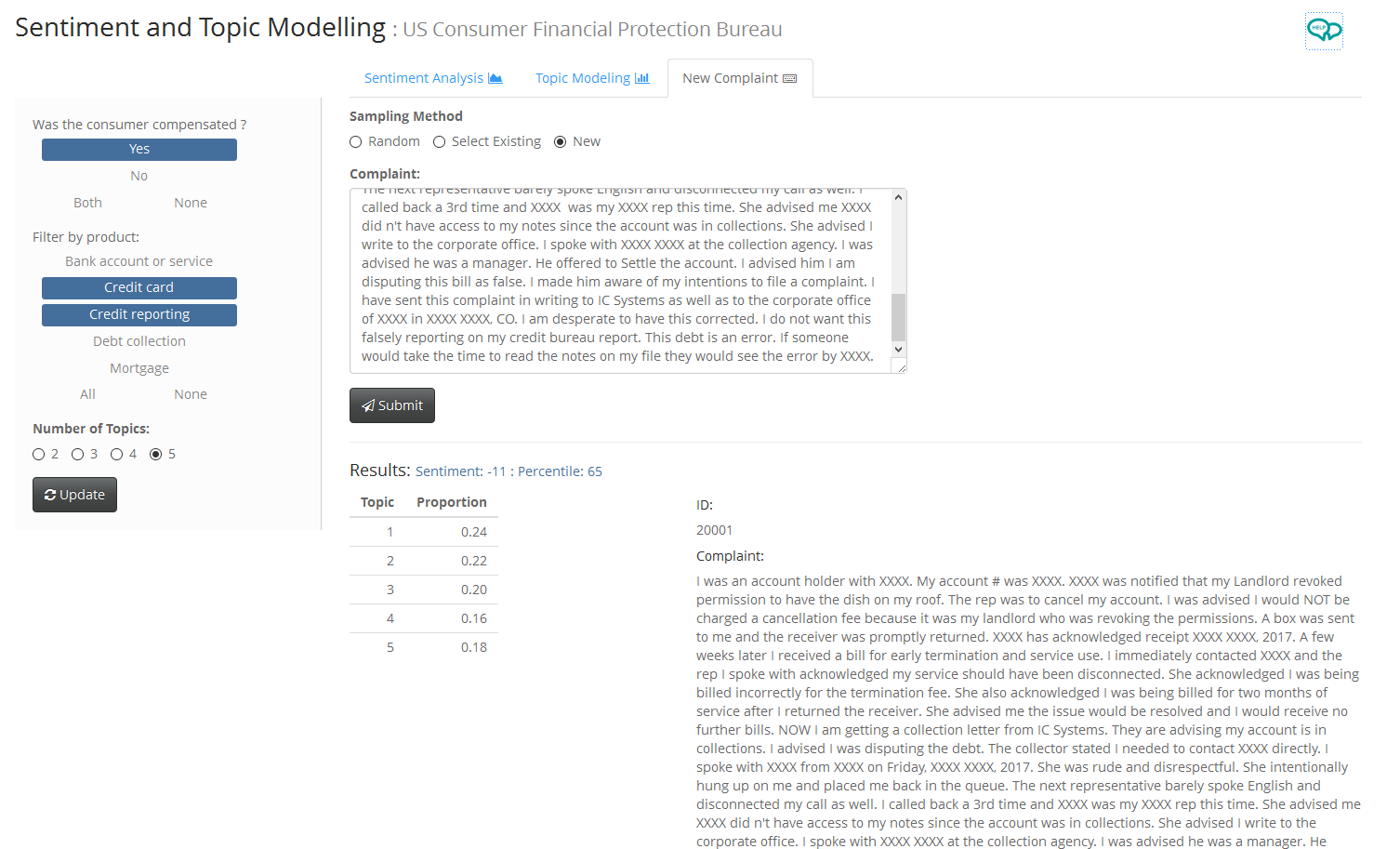
The topic modeling uses the filters as described above and adds a selector for the number of topics to use to evaluate the complaints. In this case the same filters are used as shown on the previous page and it is filtered by 2 topics.



All the values and products are selected and the analysis was done on 5 topics.

## Topic Modeling

|  |  |
| --- | --- |
|  | The “New Complaint” tab allows the user to either select an existing complaint (randomly or by entering the complaint ID) or to enter new text that is then examined using the filters selected and the number of topics.  The result of the selection will display the overall sentiment, the percentile it belongs, the complaint id (20001 if it is new), the complaint text and the topic proportional scores.  The “Random” selector randomly selects an id from the 20 000 complaints and displays the results.  The “Select Existing” allows the user to enter an existing ID to use in the analysis.  The “New” selector allows the user to enter new complaint to use. |
|  |



This screen shows the selection of complaints that where compensated and “Credit card” and “Credit reporting” selected with 5 topics to analyze. The new complaint text that was entered results in a close comparison between topics with topic 1 being the best match. The sentiment score is low at -11 but it falls within a reasonable range as it is similar to other complaints in the group (percentile).