

TEAM 5

Detect Change points and Events

BIA-660-A PROJECT - Yelp Dataset Challenge

12/06/2016

Table of Contents

- Problem Statement
- Algorithm - Change Points
- Algorithm - Events
- Code Demo
- Output and Visualization

Problem Statement

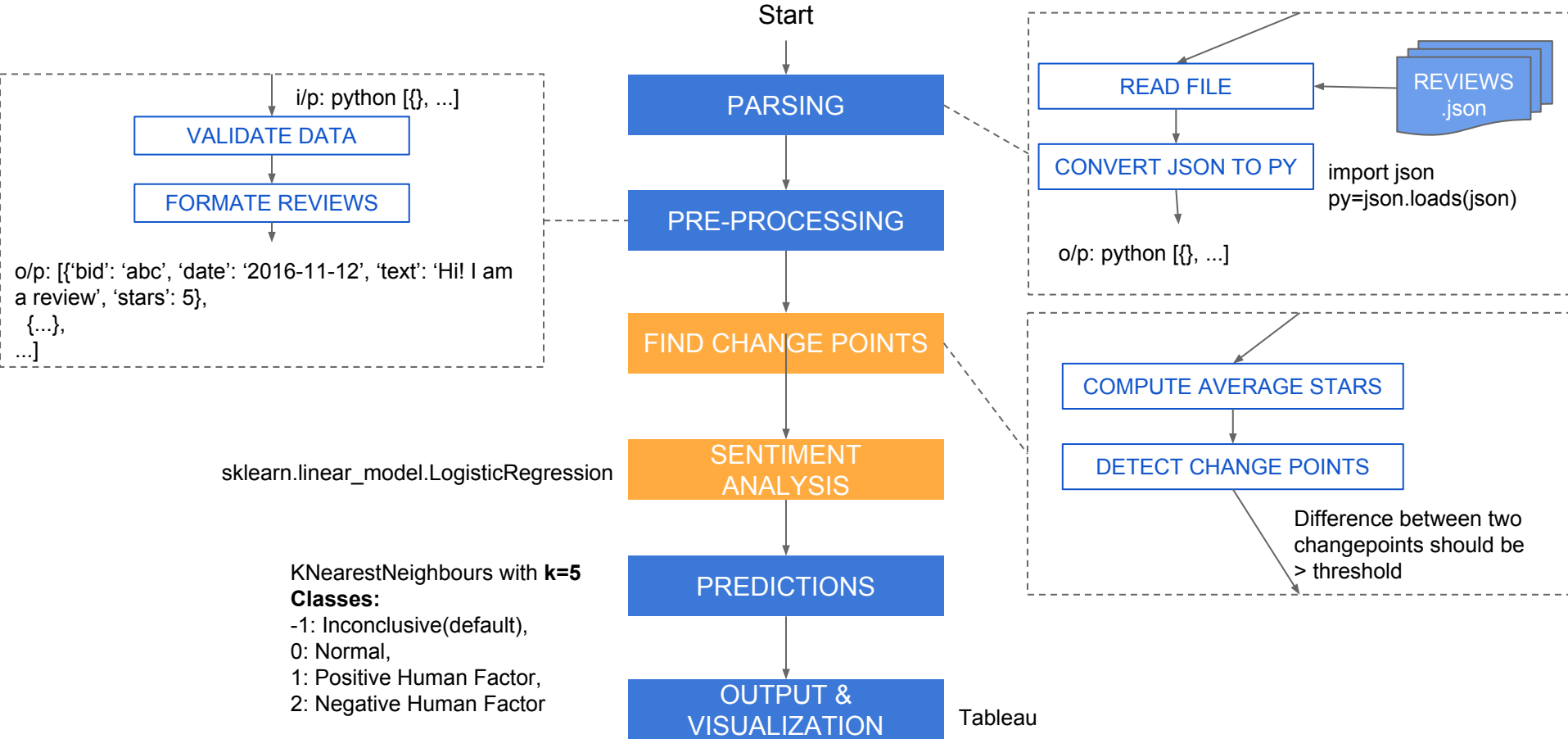
1. Changepoints

Can you detect when things change suddenly (i.e. a business coming under new management)?

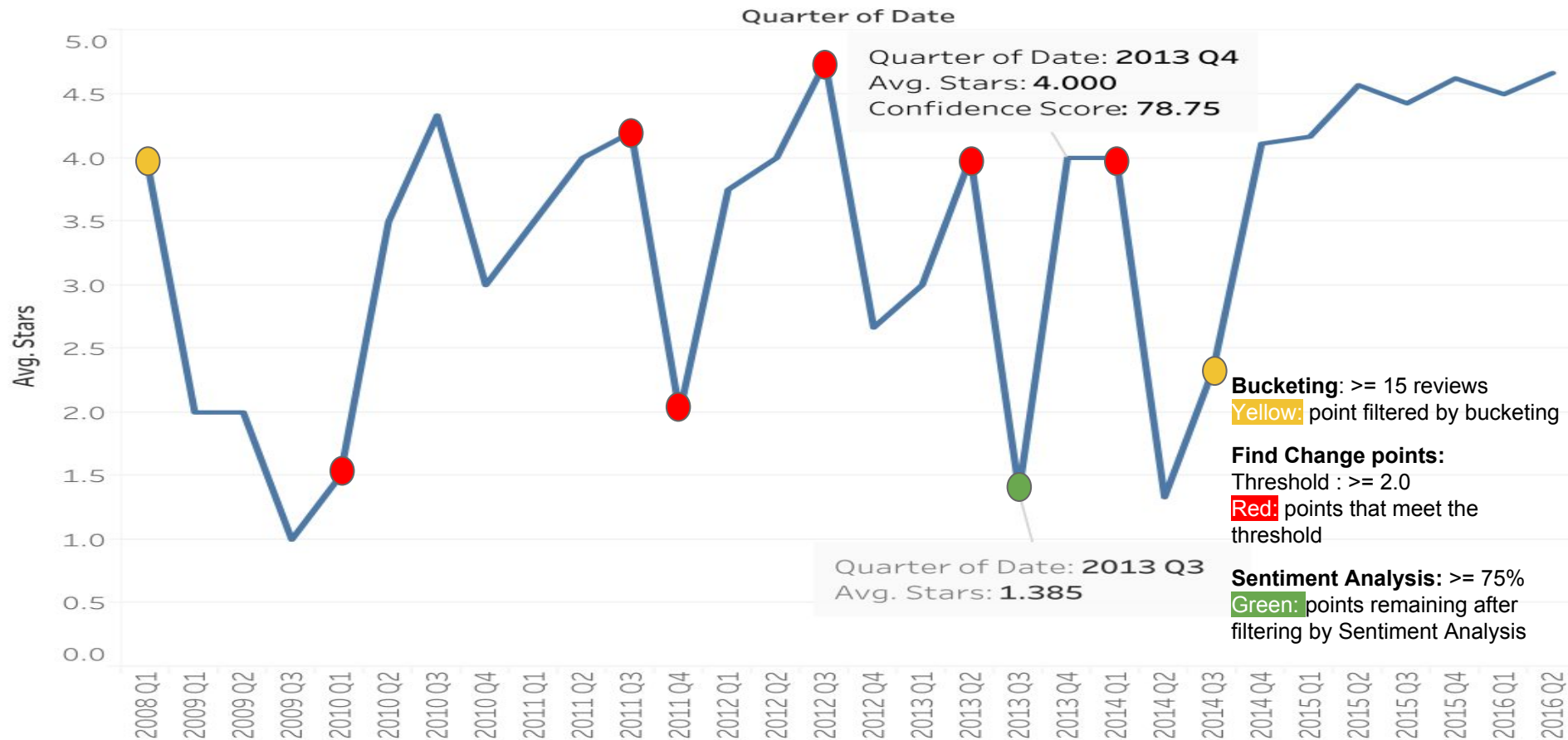
2. Events

Can you see when a city starts going nuts over cronuts?

Algorithm - Change Points



Example for a Change Point



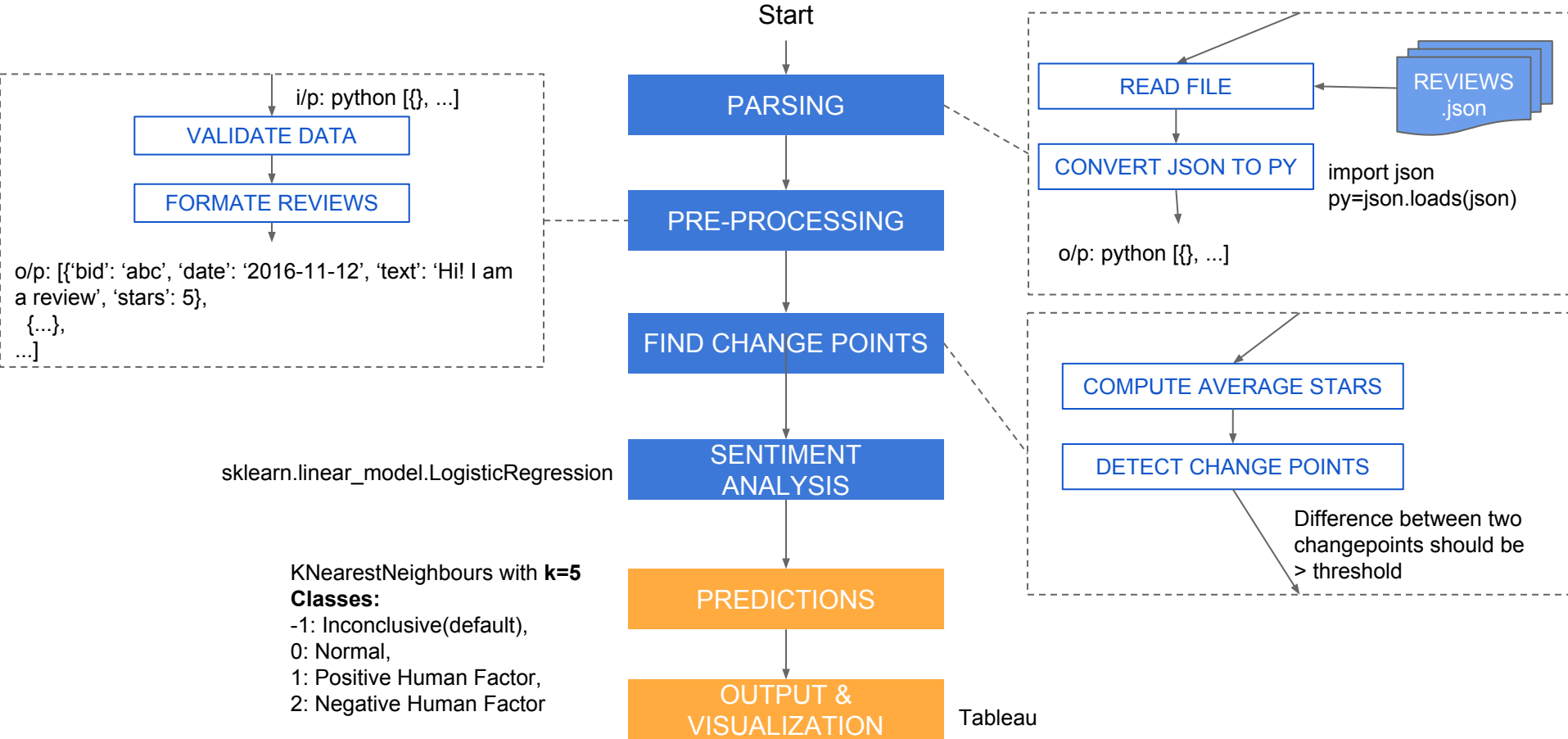
The trend of average of Stars for Date Quarter. The data is filtered on Business Id, which keeps OBpMvu5B9fY-KEbOuxLtFQ.

Sample Results

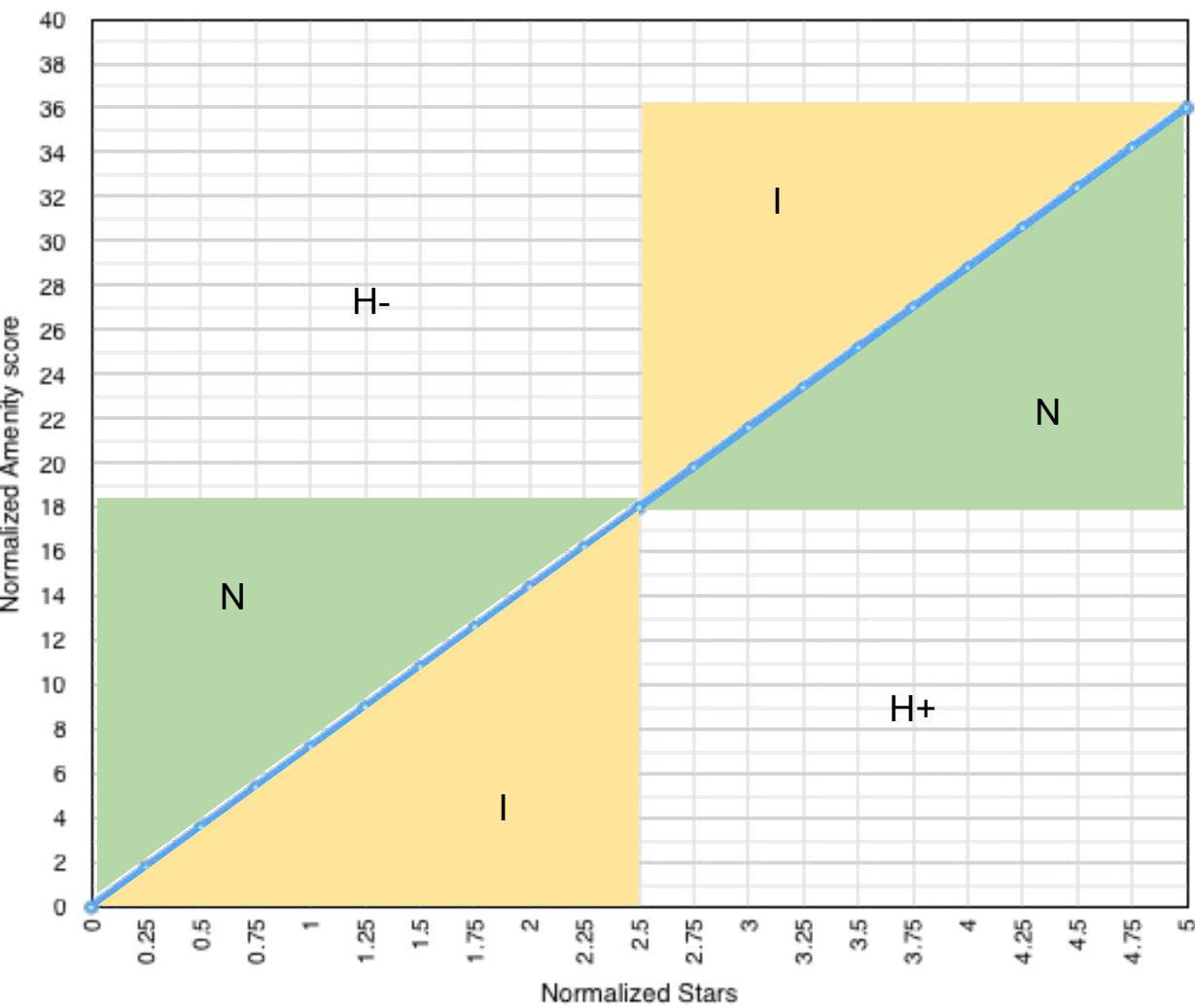
First 5 Change Points

Business Id	Bucket	Slope	Confidence Score
4_yqGLhfQJ-mRyH7i2-KNQ	2016-Q2	2.35	75.00
DfNoB509sMpMbu-qvsgvrA	2016-Q2	2.00	76.15
zDGhPMIxpCjKWVyXROjUzw	2016-Q2	2.24	84.71
KHtaNwoAG76qK6BIYGYJvQ	2016-Q2	-2.18	82.50
cTFq_1sAnQUb9eMZO6FlvQ	2016-Q2	-3.20	81.82

Algorithm - Change Points



Stars per Amenities



Legend

N: Neutral Scenario - Everything is working as it should

I: Inconclusive - Cannot determine due to less data

H+: Positive Scenario - Humans are doing GOOD job

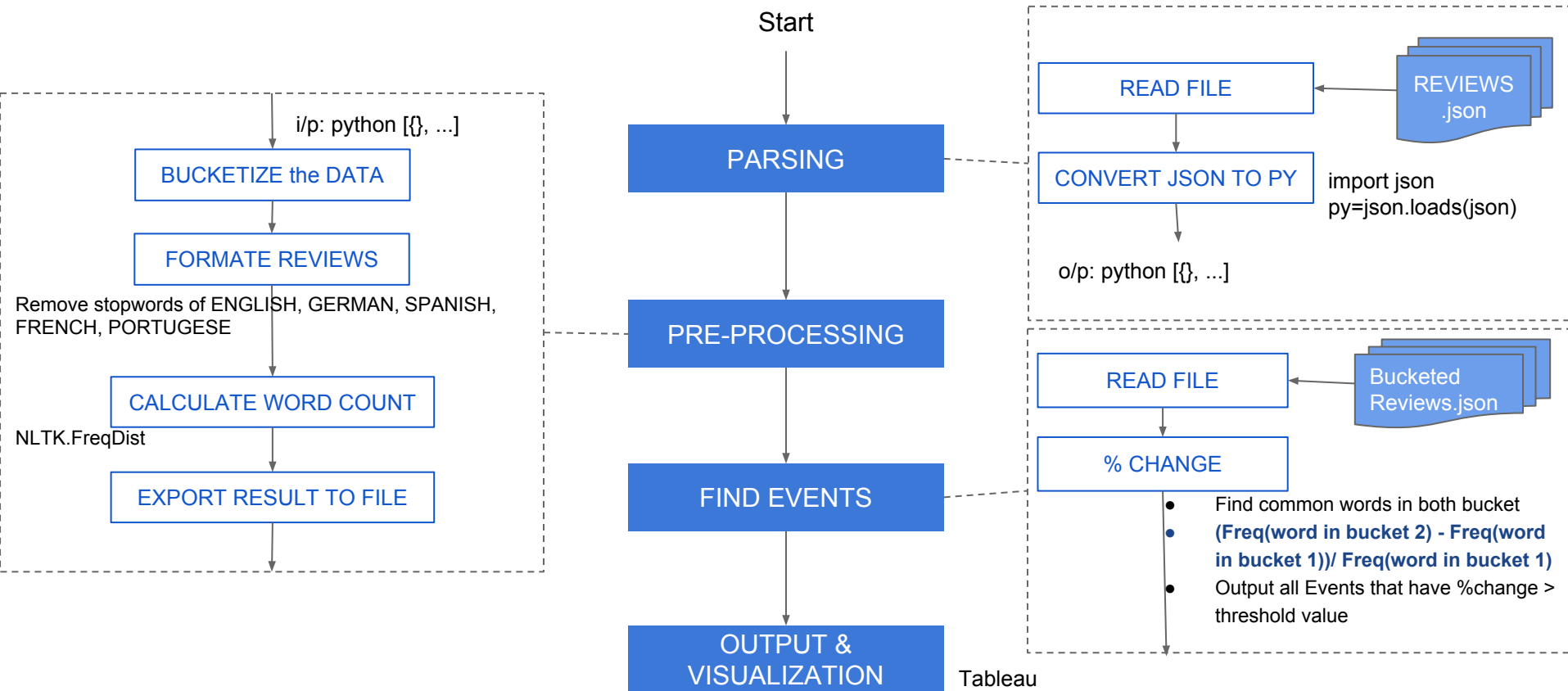
H-: Negative Scenario - Humans are doing BAD job

Sample Results

First 5 Change Points with Reasons

Business Id	Starts	Amenity Score	Reason
4_yqGLhfQJ-mRyH7i2-KNQ	3.0	2	Due to Positive Human Factors
i0rx72lqWuGOyWadGjRLVg	3.5	10	Due to Positive Human Factors
0CeEgarUoVP2Nq9s1Mn0Aw	3.0	12	Due to Positive Human Factors
JCyuOGUzMrXxJOj0pkoRMg	1.5	1	Inconclusive: due to insufficient data
UCid2Tas6K4SB9sHO9gQ8w	3.0	13	Due to Positive Human Factors

Algorithm - Events



Sample Results

Word	%Change	Description
pokemon	1100.0	Mobile Game
kanoa	350.0	Wireless Earphone Company
andarra	600.0	Transport Company
peanutbutter	400.0	Food
barbocoa	400.0	Caribbean Food
uzbek	300.0	Country ?
mortimer	300.0	?
hasbrown	300.0	Food
asbestos	300.0	Chemical used Construction
wahlburger	300.0	Food Company
macdonald	300.0	Food Company

Bucket: 2016-Q2

DEMO

Run Command:

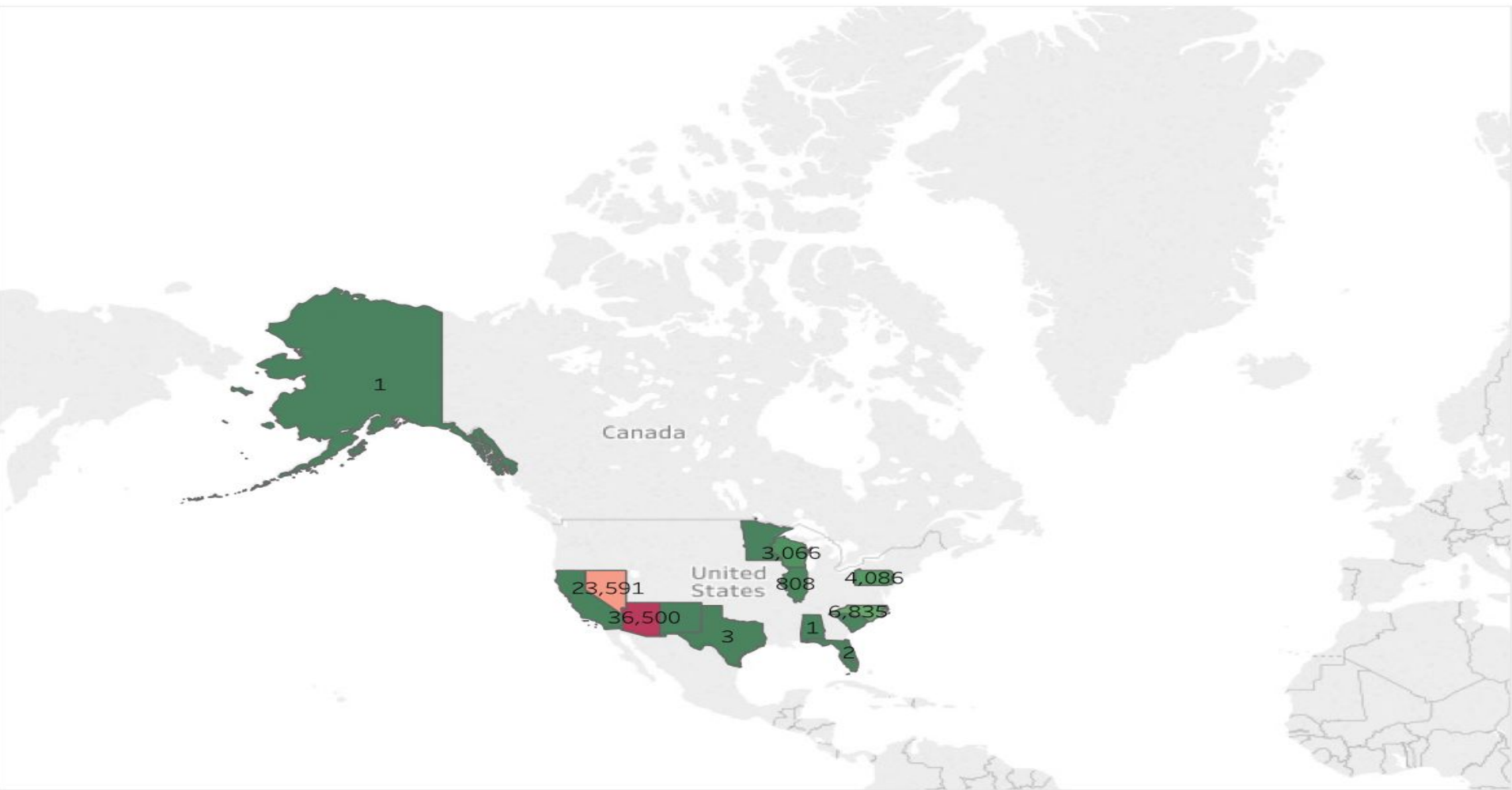
```
python Main.py --bucket quarter --averagestars 2.0 --confidencescore 74.9 --granularity general  
--kneighbours 5 --eventthreshold 300 --dispalyeventforbucket 2016-Q2
```

To run the program read the README file

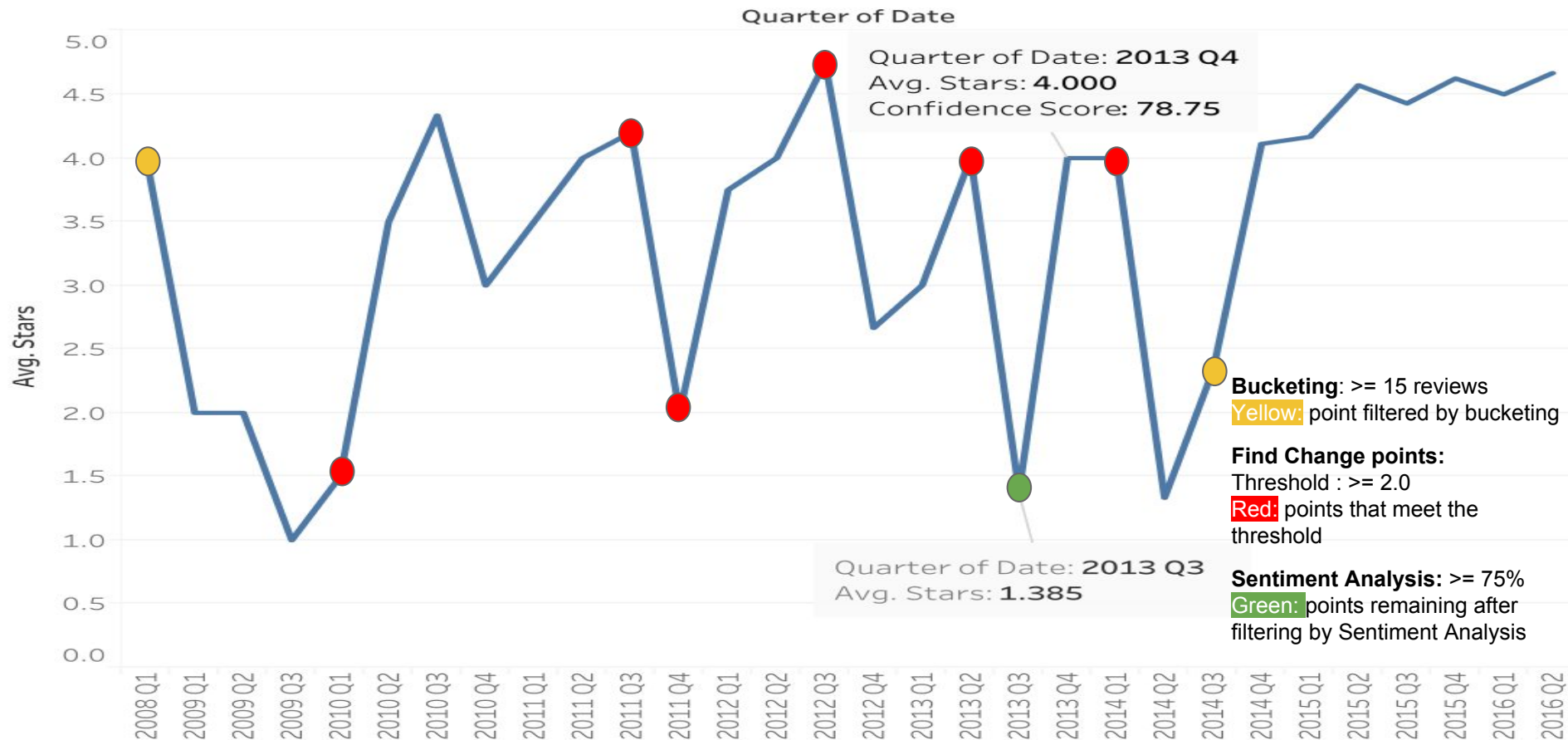
RESULTS & VISUALIZATION

Complete Result in Tabular format can be found under “/Results/” folder

Number of reviews in the USA

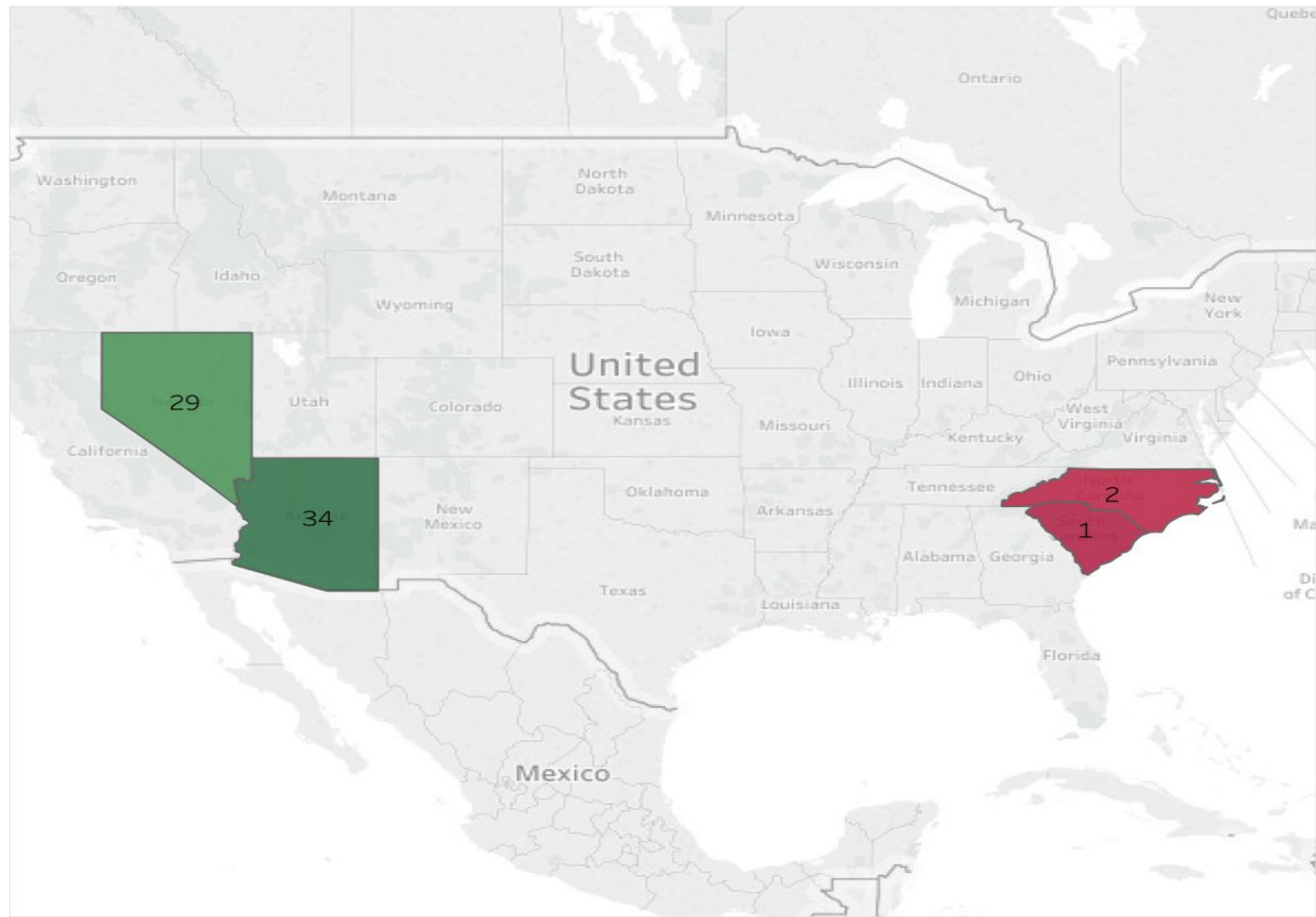


Example for a Change Point




The trend of average of Stars for Date Quarter. The data is filtered on Business Id, which keeps OBpMvu5B9fY-KEbOuxLtFQ.

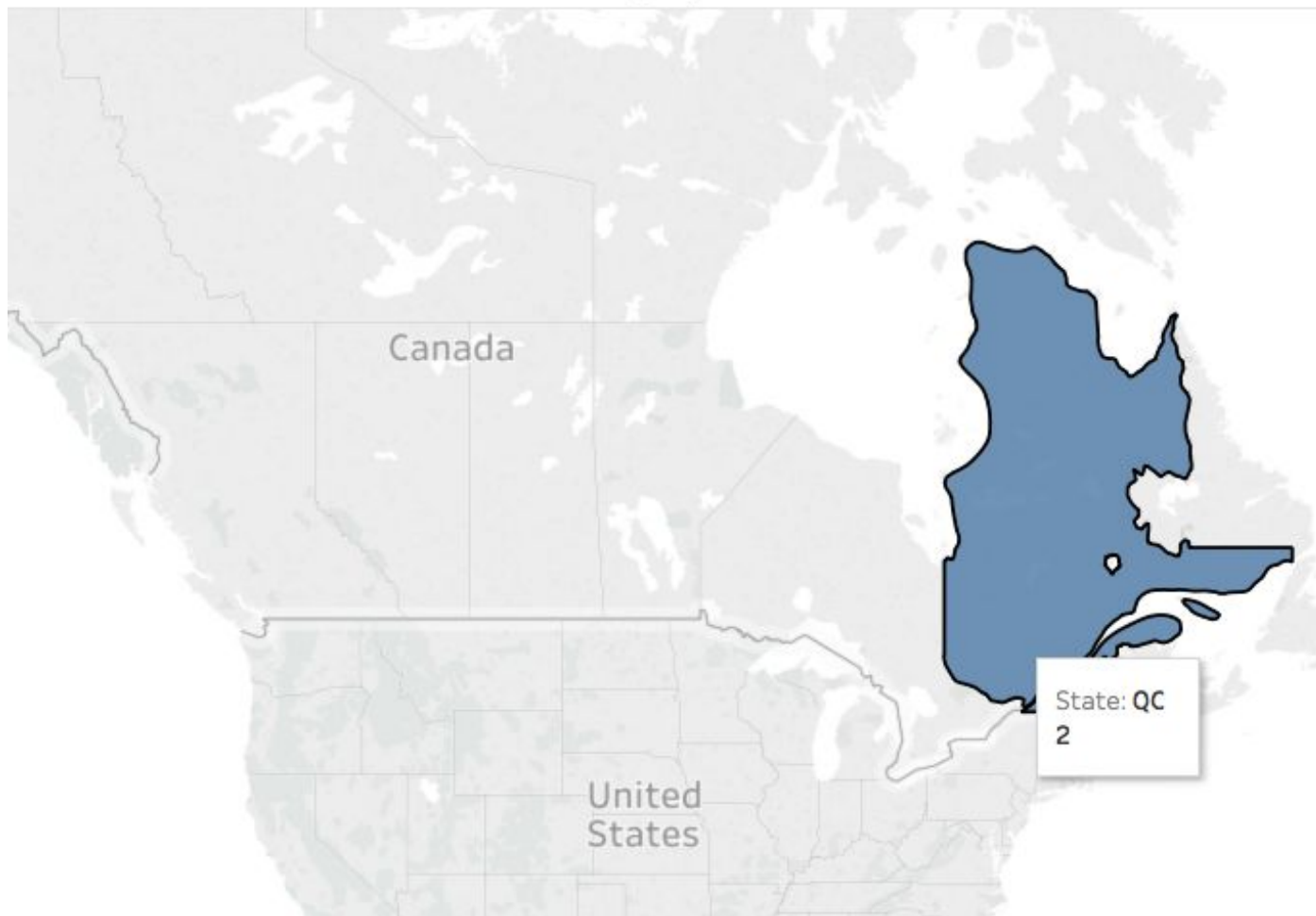
Number of change points in the USA



Change Points with the reasons for the state of Arizona

Business Id	Name	Confidence Score	Reason 
0BpMvu5B9fY-KEbOuxLtFQ	Sushi Ave	78.75	Due to Positive Human Factors
0CeEgarUoVP2Nq9s1Mn0Aw	Mad Dog Saloon	76.15	Due to Positive Human Factors
1oFvqm7eSpJbVpxpPxueaA	Buca di Beppo Italian Restaurant	77.14	Due to Positive Human Factors
00xD3o61AjrGOpEHIGPEwg	Taco Shack	75.00	Due to Positive Human Factors
16VWORTVIsJ6QZt3HTxj3w	Shahrazad Middle Eastern Cuisine	75.00	Due to Positive Human Factors
240mQWFTnbGv-e8GyEMAw	Z'Tejas Southwestern Grill	75.00	Due to Positive Human Factors
-wvryBujtHabx4jCW98eqQ	Juicd Life	84.71	Due to Positive Human Factors
C_Tu6eMyzSJhDwg2XXtnUw	Dre's Hair Salon	75.00	Due to Positive Human Factors
CNxxXXv7NVkoLWjU7t4gwA	Steak 'n Shake	78.75	Due to Positive Human Factors
cTFq_1sAnQUB9eMZ06FlvQ	Prestige Car Wash	81.82	Inconclusivedue to insufficient data
CvufqyRBbS2Ed2wQIAXCMw	Angelina's Pho & Grill Bar	78.75	Due to Positive Human Factors
EtijSxqheu_LHohSvKQfLA	Ken's Nails and Spa	81.82	Due to Positive Human Factors
FSoc19hv_VnUisJ856QZ5w	Firehouse Subs	78.00	Due to Positive Human Factors

Number of change points in Canada



end.