Analytics with Unstructured Data

Social Media Data

Commitment Beyond Numbers

PINNACLE
ACTUARIAL RESOURCES, INC.

Roosevelt C. Mosley, Jr., FCAS, MAAA

CAS RPM Seminar
Dallas, TX
March 10, 2015

Social Media Analytics

- Social media – a way of life

- Data

- Data Processing and Analysis

- Analysis Examples

- Analysis Challenges

Social Media – A Way of Life

Commitment Beyond Numbers

PINNACLE
ACTUARIAL RESOURCES, INC.

Social Media – A Way of Life

- 56% of Americans have a profile on a social networking site
- 64% say that social media has an influence on purchasing decisions, with Facebook having the largest influence
- 76% of Twitter users are active users, up from just 47% in 2010
- In June, 2014, YouTube had 11.3 billion video views, Facebook had 12.3 billion
- Facebook has 1.37 billion active worldwide users, Twitter 288 million
- On average 190 million tweets are sent per month

Insurance Social Index

Company	ISI	Audience	Engagement	Response
Nationwide Insurance	83%	72%	83%	90%
USAA	83%	96%	65%	92%
PEMCO Insurance	80%	24%	97%	100%
Esurance	72%	82%	49%	88%
Sun Life Financial Canada	70%	49%	72%	83%
MetLife	67%	88%	42%	79%
Northwestern Mutual	67%	40%	74%	77%
Auto-Owners Insurance	66%	15%	99%	67%
AMICA Mutual	64%	28%	56%	98%
American Family Insurance	64%	69%	50%	75%

Source: Customer Respect Group, "Social Eyes – The Insurers' View of Social Media." February, 2015.

Opportunities for Insurance Companies

- Insurance companies are investing significant resources on a social media presence

- Current and potential customers are voluntarily sharing intimate details of their lives with the world

- Current and potential customers are interacting with companies on a very personal level

- Applications
  - Customer service
  - Customer sentiment
  - Competitive intelligence
 - Broad market trends
  - Claims

Data

Social Media Data

- Twitter data - over 6 million insurance tweets total (January, 2012 to present)
  - 1.6 million GEICO tweets

- Data
  - Content of the tweet
  - Specific tweet recipient
  - Sender of the tweet
  - Language of tweet
  - Where the tweet originated
  - Link to a picture of user
  - Latitude and longitude of the user
  - Date and time of tweet
  - Device/platform

- Advantages of social media data
  - Unfiltered
  - Broad view of non-customer reactions
  - Facilitates more timely analysis of trends
- Sources
  - Third party data aggregators (Hootsuite, GNIP)
  - API
  - Company developers
  - Screen scraping

Tweets per Month – All Companies

Number of Tweets Per Month
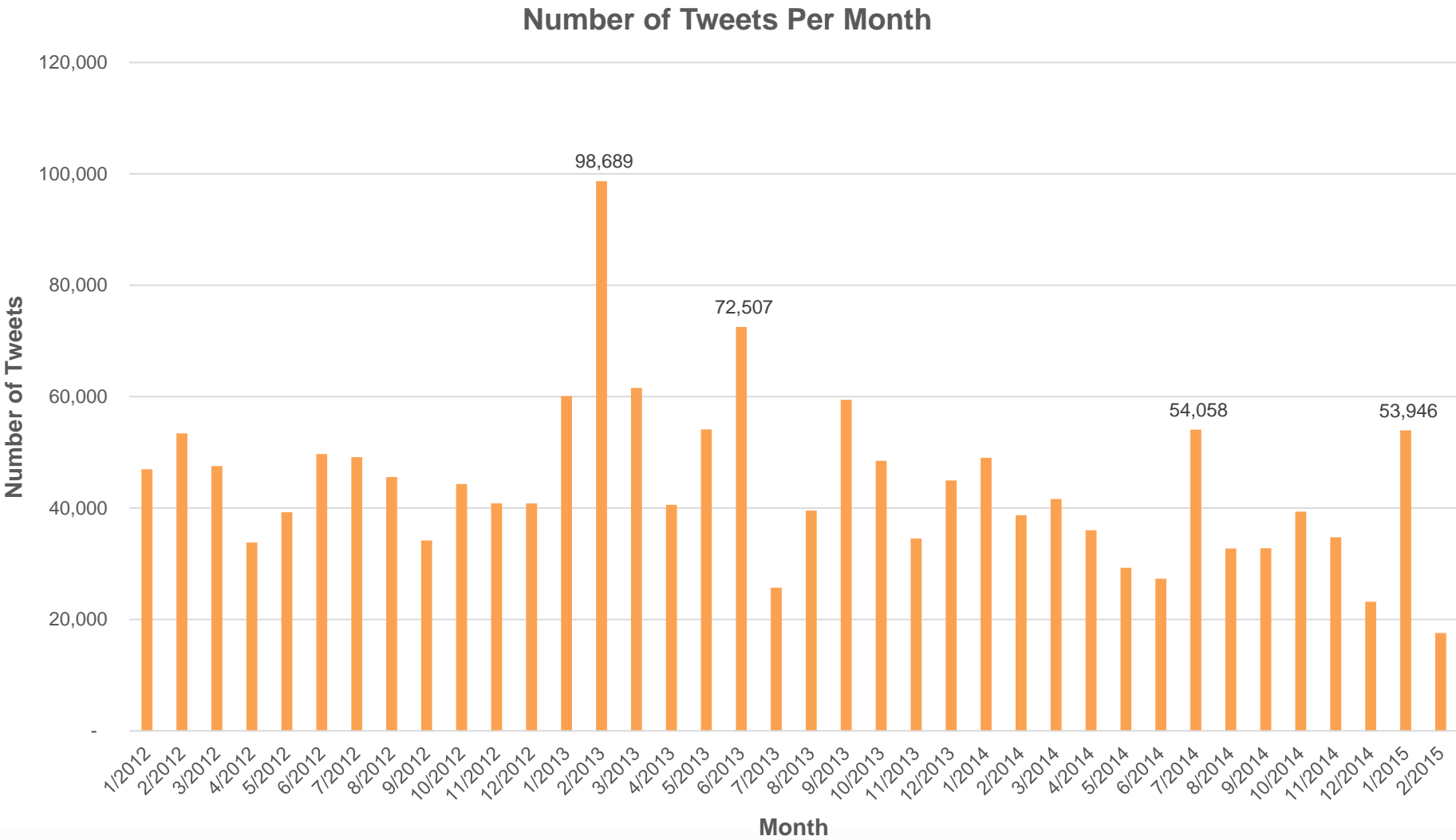
Tweets per Day – February 2013

Number of Tweets per Day

Data Processing & Analysis

Commitment Beyond Numbers

PINNACLE
ACTUARIAL RESOURCES, INC.

Data Processing Steps

- Remove punctuation and symbols (retain @ and #)

- Parse the tweet (35 words worked for Twitter – will need many more for other sources)

Tweet ID	User	Tweet	Word1	Word2	...	Word35
1	@mosley	Text of tweet	W1	W2	…	W35

- Change table structures from tweets in rows to tweets in columns – keep indicator of order

Tweet ID	Word Order	Word
1	1	Word1
1	2	Word2
…	…	…
1	35	Word35

- Correct spelling errors

- Add word indicators

Identifying Themes - Clustering/Segmentation

- Unsupervised classification technique
- Groups data into set of discrete clusters or contiguous groups of cases
- Performs disjoint cluster analysis on the basis of distances computed from one or more quantitative input variables and cluster seeds
- Data points are grouped based on the distances from the seed values
- Objects in each cluster tend to be similar, objects in different clusters tend to be dissimilar

Association Analysis

- Background in market basket analysis
- Identification of items that occur together in the same record
- Produces event occurrence as well as confidence interval around the occurrence likelihood
- Can lead to sequence analysis as well, which considers timing and ordering of events

$$\text{Support} = \frac{\text{Transactions that contain items A \& B}}{\text{All transactions}}$$

$$\text{Confidence} = \frac{\text{Transactions that contain items A \& B}}{\text{Transactions that contain item A}}$$

$$\text{Expected Confidence} = \frac{\text{Transactions that contain item B}}{\text{All transactions}}$$

Transactions = tweets
Items = words

Analytics Model Training

- Random sample of tweets are scored for a target variable
- Model can be developed for the random sample to predict the target
- Model scoring can be applied to new tweets to determine predicted classification and

Analysis Examples

PINNACLE
ACTUARIAL RESOURCES, INC.
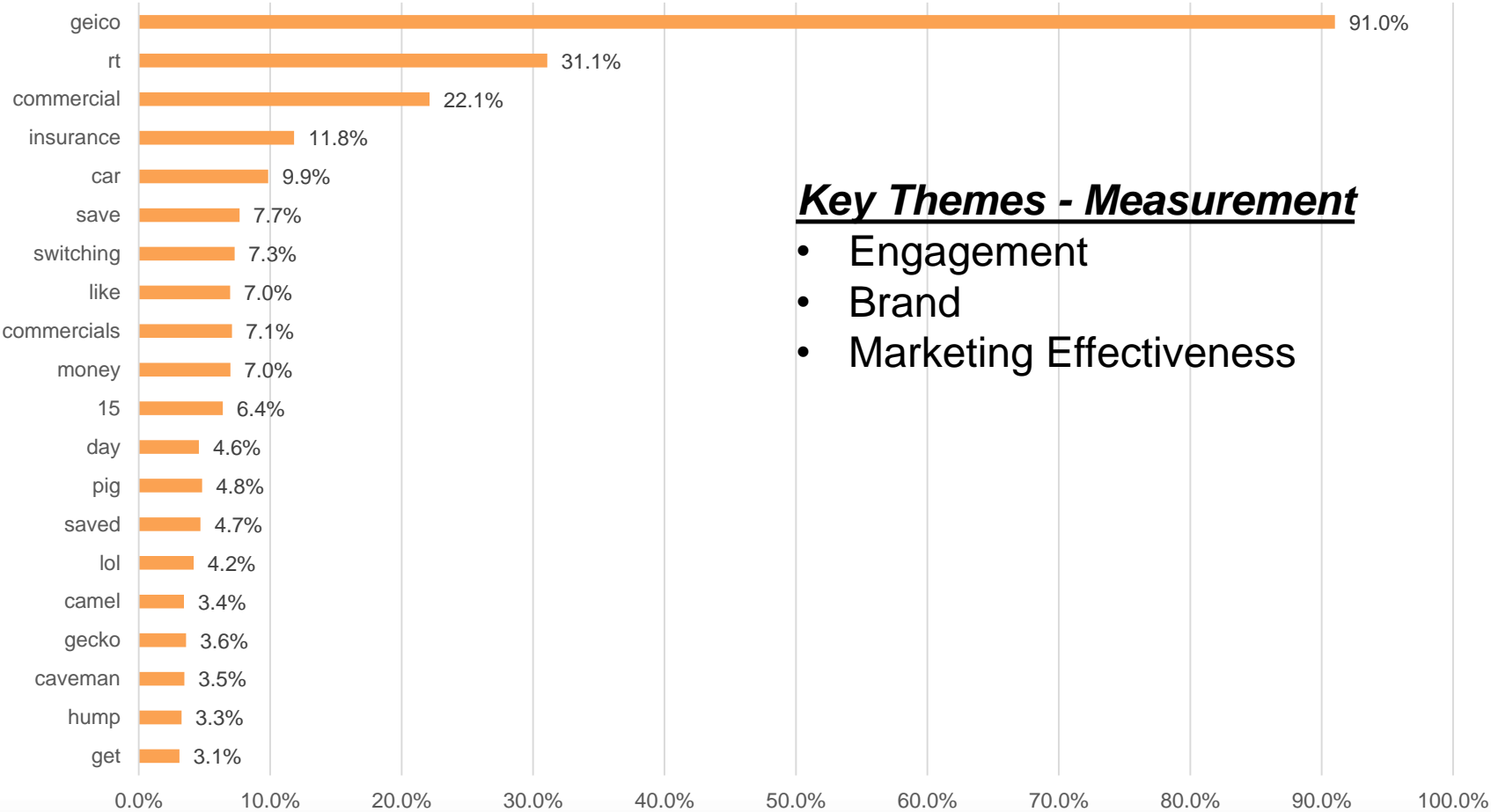
Top 10 Keywords

GEICO Keywords

Percentage of Tweets

Keyword	Percentage
geico	91.0%
rt	31.1%
commercial	22.1%
insurance	11.8%
car	9.9%
save	7.7%
switching	7.3%
like	7.0%
commercials	7.1%
money	7.0%
15	6.4%
day	4.6%
pig	4.8%
saved	4.7%
lol	4.2%
camel	3.4%
gecko	3.6%
caveman	3.5%
hump	3.3%
get	3.1%

Key Themes - Measurement
- Engagement
- Brand
- Marketing Effectiveness

Advertising "Focus Groups"
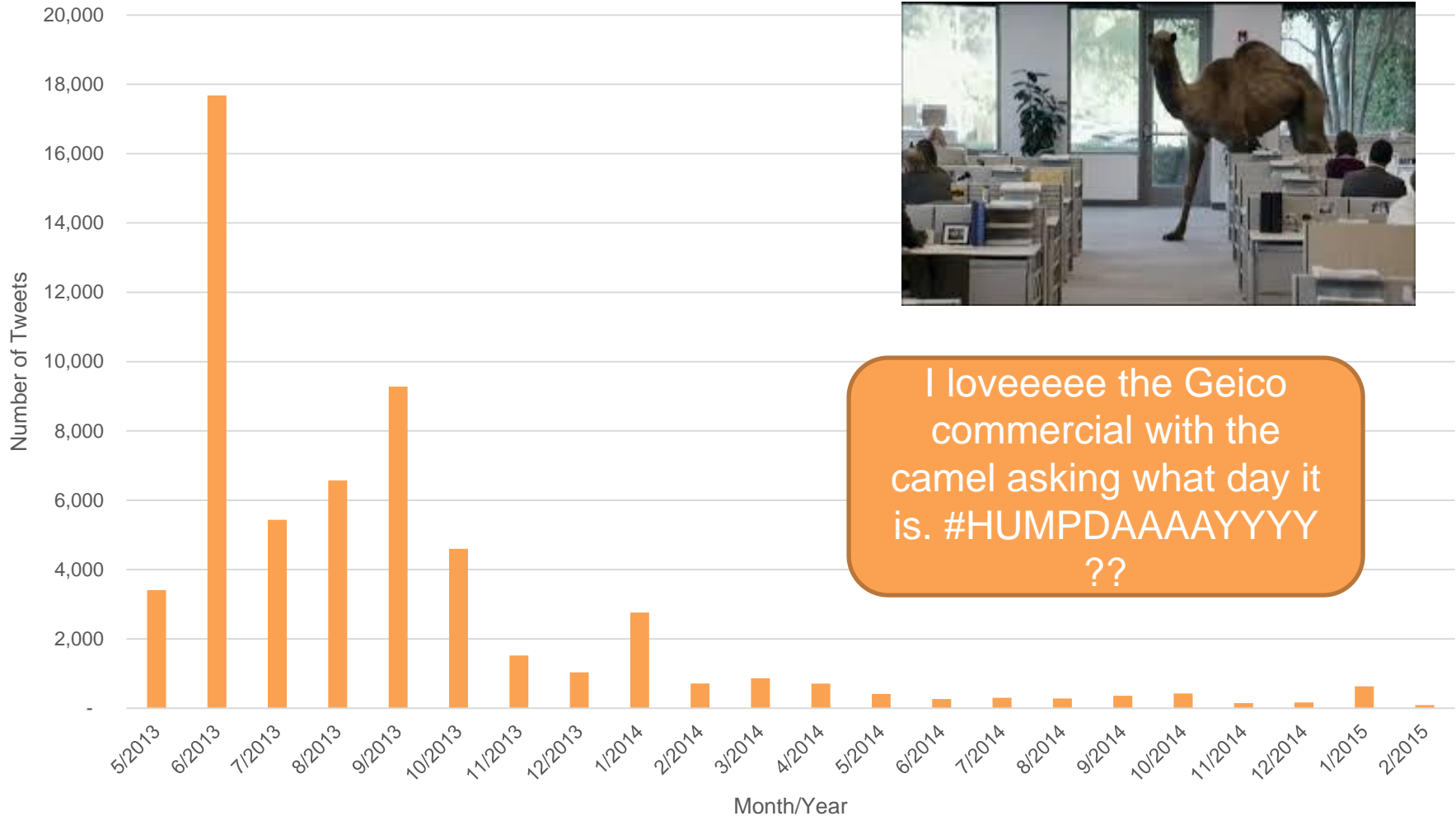
Camel – Sentiment Chart

Camel Sentiment Chart
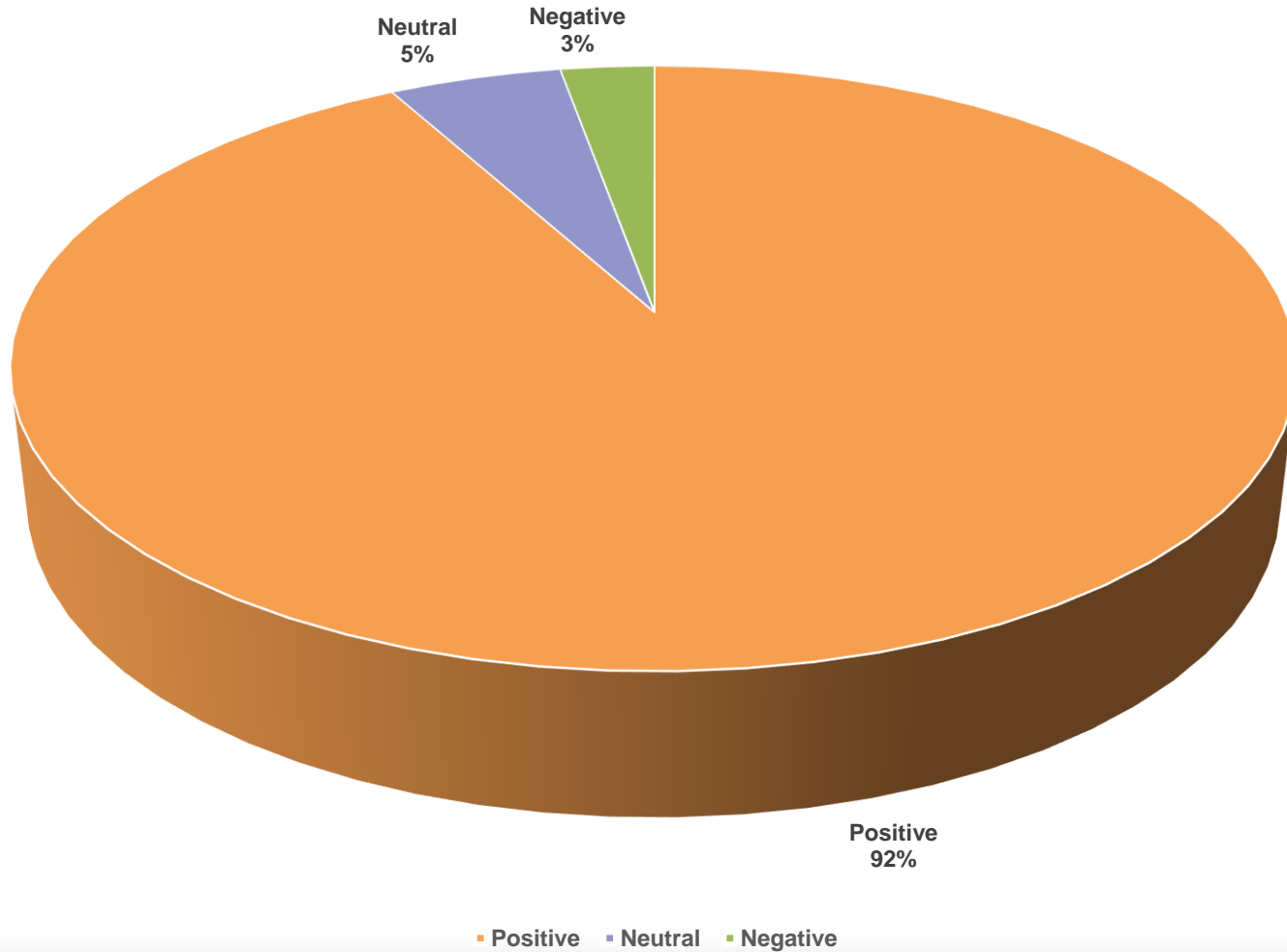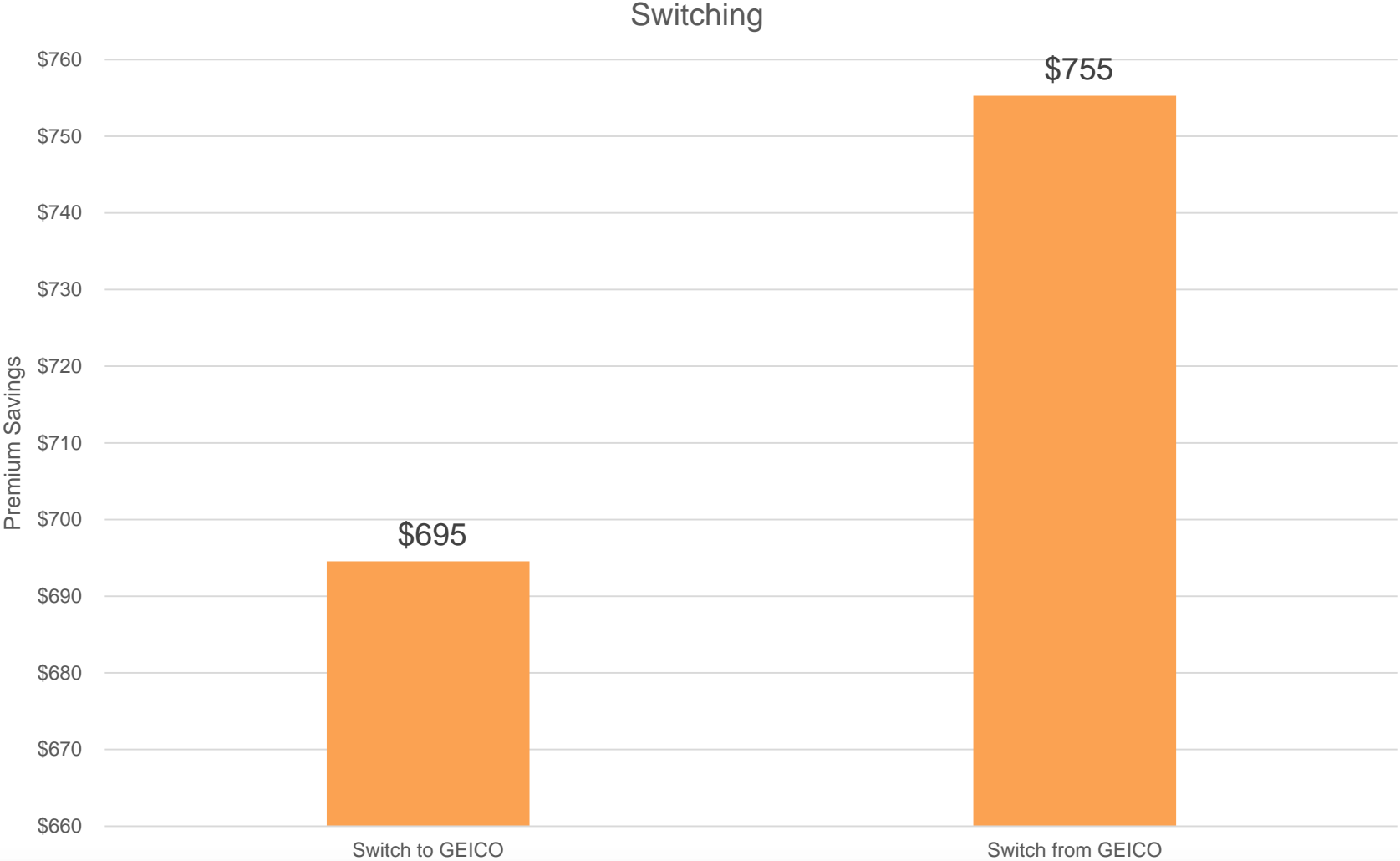
Claims Applications

- Brand evangelists
  - can't say enough wonderful things about @GEICO and how they handled my accident claim! Made a bad event tolerable. Thank you!
  - Thanks @GEICO for making my claim so easy this week. Rough week was made smooth because of your service. Claims adjuster April was great.

- Complaints
  - @GEICO_Service I've been waiting for over a week to hear about a glass claim I made. I need to get it fixed...
  - @Geico made me drive my car damaged for 7 days and now that it's in the shop I can't get a speedy supplement claim review, smh.

- Service
  - @GEICO_Service trying to access my claim on the website, and my claim # is 26 not 16 digits - where do i find the right #?
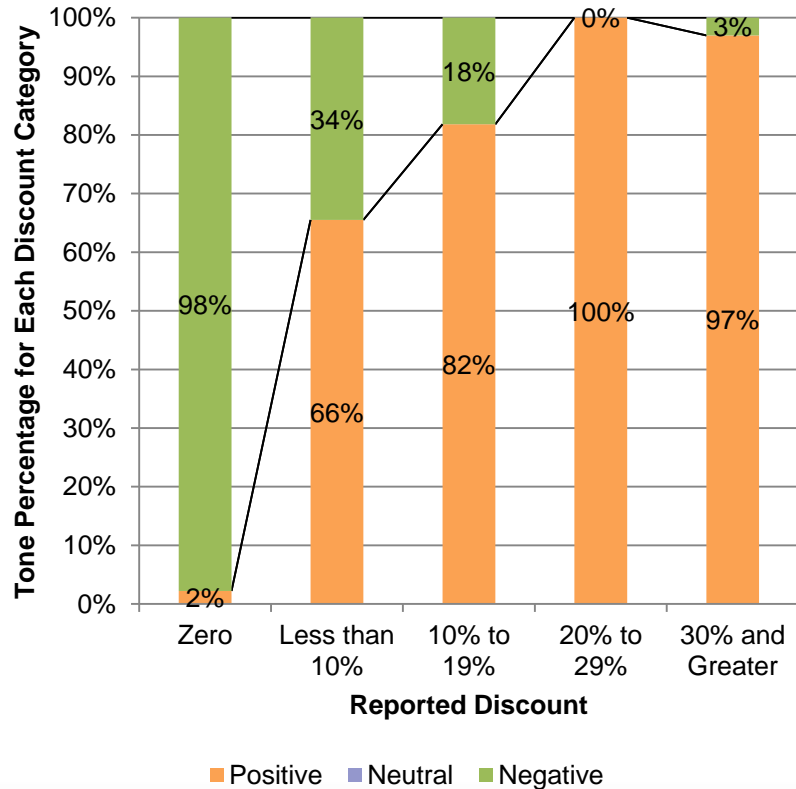
Savings

Switching

$755

$695

Premium Savings

$760
$750
$740
$730
$720
$710
$700
$690
$680
$670
$660

Switch to GEICO   Switch from GEICO
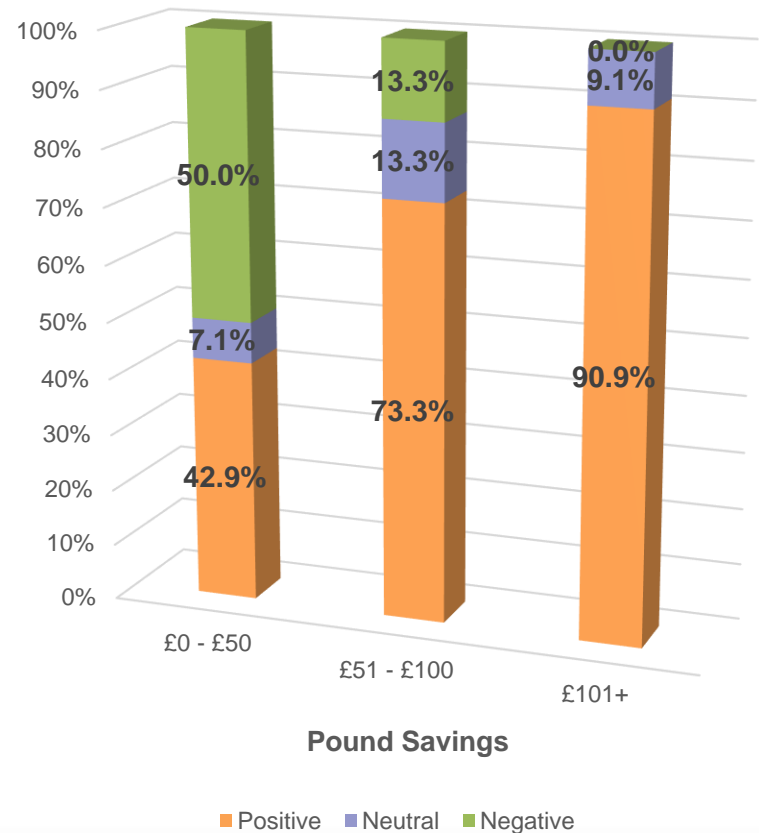
Measuring Customer Sentiment

Tweet

Thought I earned a 30% discount through the @Progressive snapshot, only to learn that policy rates went up the amount of my discount #notcool

Reported Snapshot Discount by Tone

Tone Percentage for Each Discount Category

- Zero: Negative 98%, Positive 2%
- Less than 10%: Negative 34%, Positive 66%
- 10% to 19%: Negative 18%, Positive 82%
- 20% to 29%: Positive 100%, 0%
- 30% and Greater: Negative 3%, Positive 97%

Reported Discount

■ Positive ■ Neutral ■ Negative

Pound Savings by Tone

- £0 - £50: Negative 50.0%, Neutral 7.1%, Positive 42.9%
- £51 - £100: Negative 13.3%, Neutral 13.3%, Positive 73.3%
- £101+: Negative 0.0%, Neutral 9.1%, Positive 90.9%

Pound Savings

■ Positive ■ Neutral ■ Negative

Analysis Challenges

- Accessing and collecting information
- Context
- Relevance
- Influence
- Sentiment
- Raw, unfiltered customer data
- Near real-time analysis needed

Thank You for Your Attention

Roosevelt C. Mosley, Jr., FCAS, MAAA

309-807-2330

rmosley@pinnacleactuaries.com

PINNACLE
ACTUARIAL RESOURCES, INC.