TODAY'S MODEL REVIEW
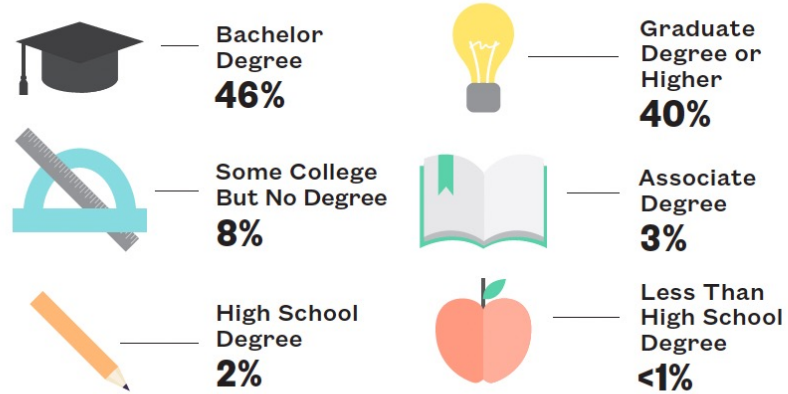
- SXSW Demographics
- Business Objective
- Tweet Word Cloud
- Sentiment Model Solution
- Next Steps

insight
CONSULTING

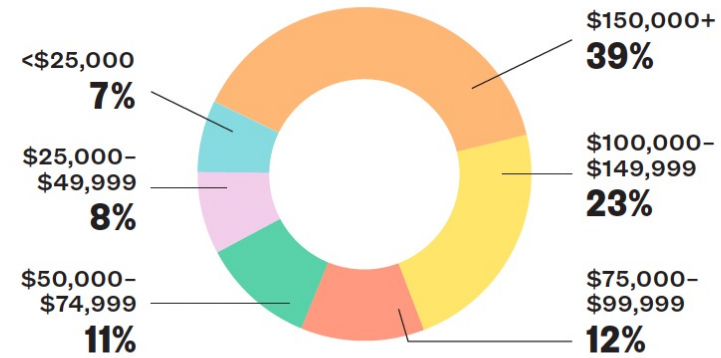# SXSW DEMOGRAPHICS SKEW COLLEGE-EDUCATED, UPPER INCOME & TRENDY

EDUCATION

Bachelor Degree **46%**

Graduate Degree or Higher **40%**

Some College But No Degree **8%**

Associate Degree **3%**

High School Degree **2%**

Less Than High School Degree **<1%**

HOUSEHOLD INCOME

ALL CONFERENCE REGISTRANTS

<$25,000 **7%**

$150,000+ **39%**

$25,000–$49,999 **8%**

$100,000–$149,999 **23%**

$50,000–$74,999 **11%**

$75,000–$99,999 **12%**

HOUSING SITUATION

Owned or Being Purchased **52%**

Rented **46%**

Other **2%**

SXSW = Desirable Market

4 SXSW DEMOGRAPHICS 2019

Source: https://explore.sxsw.com/hubfs/2019-SXSW-CONFERENCE+FESTIVALS-Demographics.pdf
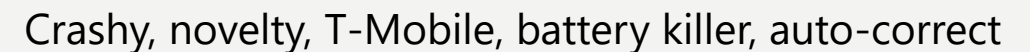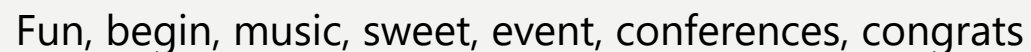
PREDICTING CONSUMER SENTIMENT

Business Objective: A smartphone consortium led by Apple and Google want to prove via a 2-week proof-of-concept whether Twitter data can be used to detect sentiment with the highly desirable consumers at SXSW.
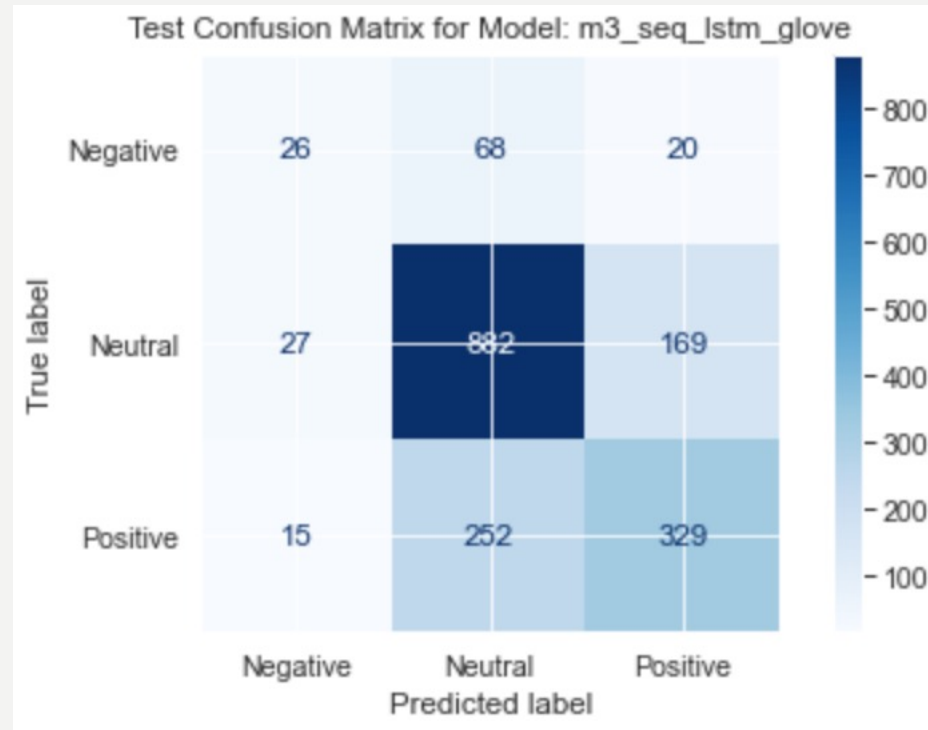
Requirements

1. Develop predictive model to classify tweet sentiment

2. Explain performance and any limitations for positive and negative model predictions

3. Suggest potential solutions to evolve beyond a POC model

SOME WORDS WERE LINKED TO ONLY A POSITIVE OR NEGATIVE EMOTION

Fun, begin, music, sweet, event, conferences, congrats

Crashy, novelty, T-Mobile, battery killer, auto-correct

PERFORMANCE OF OUR BEST (POC) MODEL IS NOT SUFFICIENT

Test Confusion Matrix for Model: m3_seq_lstm_glove

Overall Accuracy
- Model predicted correctly 69% of time
- Best performance on neutral tweets

Negative Tweets
- When predicted, right only 38% of time
- Able to predict only 23% of true negatives

Positive Tweets
- When predicted, right 64% of time
- Able to predict 55% of true negatives

Conclusion: Proof-of-Concept effort shows promise but falls short of usability

PHASE 1 IMPROVEMENT IDEAS

We recommend investing in a Phase 1 project to improve predictive accuracy but addressing these shortcomings in the POC model:

- **Data** – Greatly increase tweet sample size from 9K (only 6% were negative)

- **Use Case** – Work backwards from specific use case goals (actions) to identify tweet patterns

- **Sentiment Scoring** – In addition to human labeling, automate word sentiment scoring

- **Improve Models** – Deeper exploration of parameter tuning in predictive models

Questions?

"In God we trust, all others bring data." – Edward Deming

Insight Consulting Group is a boutique consulting firm specializing in analytics, business intelligence, strategy, data visualization, and data warehousing for Fortune 500 client accounts.

Visit www.insightconsultinggrp.com for more information.

Please contact us at:

Chuck Utterback, Principal 770.298.3169
chuck.utterback@insightconsultinggrp.com

https://github.com/cutterback
https://www.linkedin.com/in/chuckutterback/

THANKS