ASTRO-ANALYTICA

IST 615 – Group 1

- Anjali Chintam
- Daniel Jeski
- Gabe Tiger
- Manan Meghani
- Ryan Ondocin
- Sharvil Turbadkar

Agenda

- Product Idea and Relevance
- Cloud Architecture
- Cost and Revenue Models
- Stakeholders and Competitors
- Scope and Limitations
- Future Goals and Expectations
- Demo

Product Idea and Relevance

Product Idea

- Sentiment Analysis on Twitter data on Elections 2020
- The proliferation of social media in the recent past has provided end users a powerful platform to voice their opinions
- Much can be drawn regarding how the election will play out by looking at the opinions expressed through Twitter

Product Relevance

What is Sentiment Analysis?

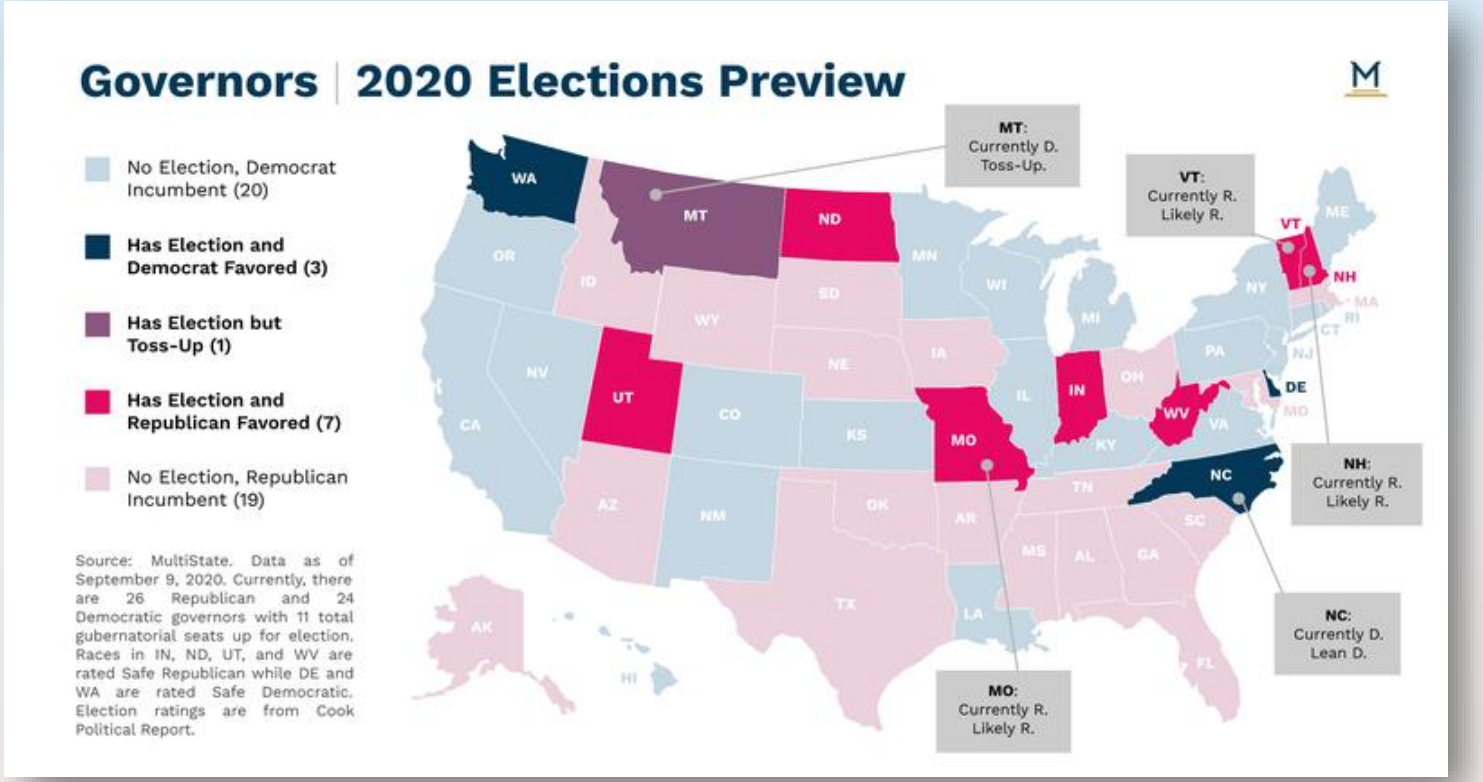
- Contextual mining of text
- Identifies and extracts subjective information in source material
- Allow businesses understand the social sentiment of their brand, product or service
- Monitors the online conversations

How Does it Help?

- Helps data analysts within large enterprises:
  - Gauge public opinion
  - Conduct nuanced market research
  - Monitor brand and product reputation
  - Understand customer experiences

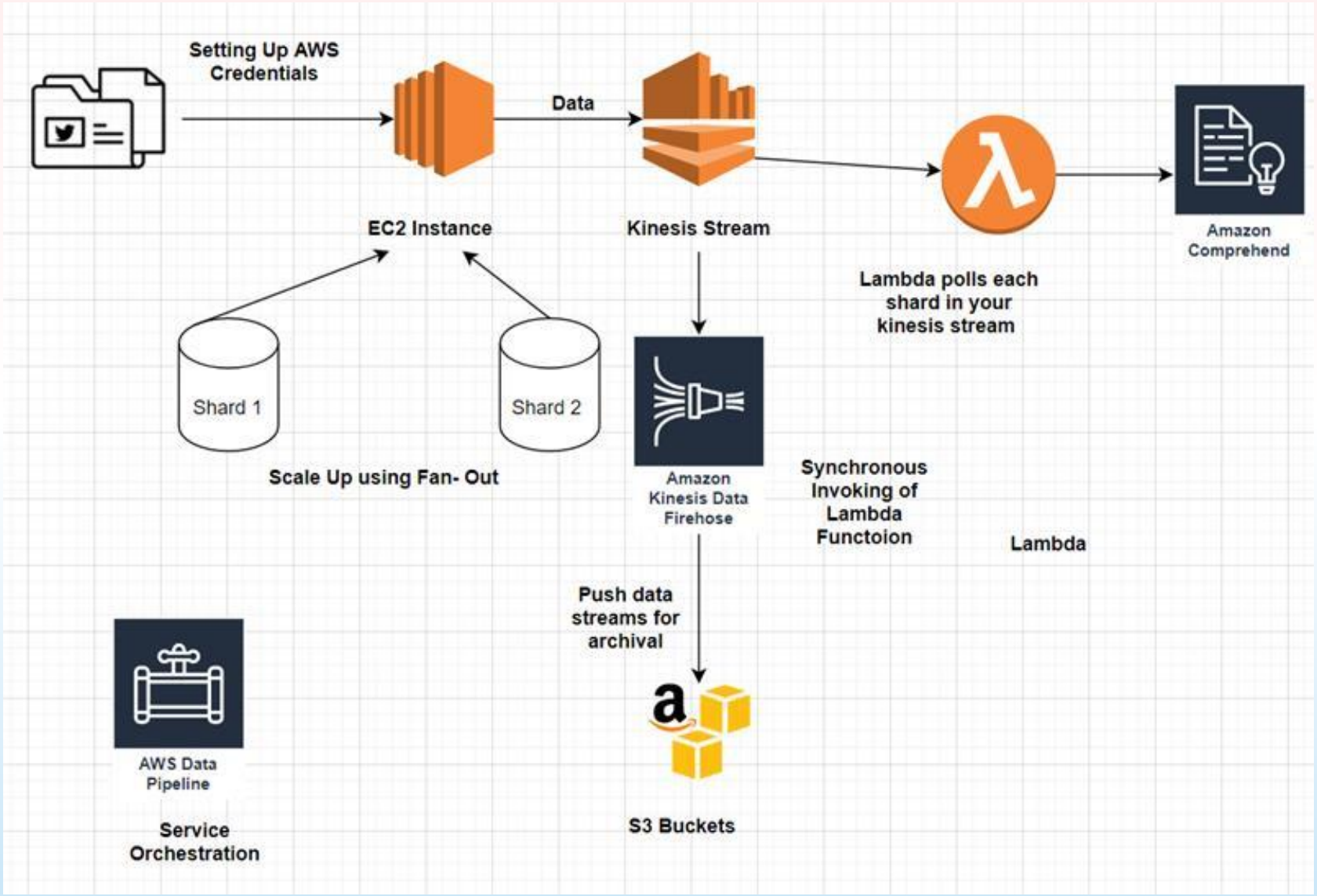
Product Relevance

- Creation of an AI dashboard predicting the 2020 U.S. presidential election by state
- To provide analysis of sentiment of people in particular regions towards a political party
- Broader use case could be any election and increasing granularity (County, City/Town, voting district, etc.)
- Even broader vision to expand into other industries
  - Marketing
  - Research
  - User behavior
  - Primarily by geographical location



Cloud Architecture

Cloud Architecture (Original)



ML Architecture

Cloud Architecture (Updated)

- Simplified architecture
- Cost reduction
- Development overhead reduction and simplicity

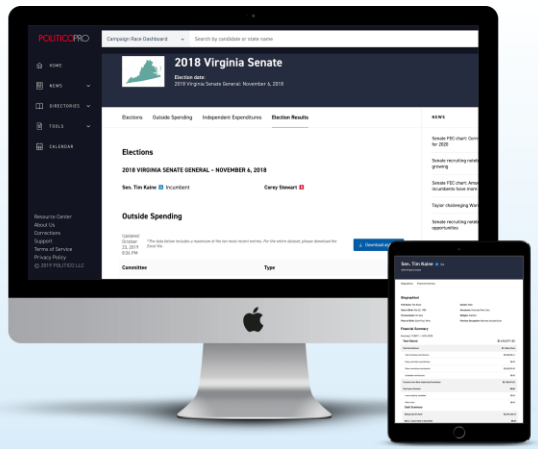
Web Hosting
User Management
Code Execution

EC2

Web Interface Communicates with EC2

User Web Access

User Seeks Data



EC2 Runs Pickle and Stores Data

Pickle Hosting and Data Storage

amazon S3

Cost and Revenue Model

Cost Model – Infrastructure (Original)

- Cost is based on initial implementation
- Cost will scale as company and available services grows
- Information provided by Amazon Web Services; cost for US EAST (N. Virginia) region

Component	Upfront	Description	Monthly	Annual
S3	None	0.023*2GB	$0.046	$0.28
Kinesis Stream	None	0.015*24*31*4Shards	$44.64	$535.68
Kinesis Firehose	None	20 records/seconds *5KB record Size=500/1048576 *86400 sec/day for a month	$8.24	$98.88
Lambda	None	Executing 5 times a month* 128MB/1024 *Price per 100 MS for 128mb	None	None
Comprehend-Topic Modelling	None	[150-50 MB processed $ flat rate processing]=100* MB billed	$0.40	$4.80
Comprehend - Sentimental Analysis	None	100,0000 election tweets on election data scraped *4 units per request*0.001 Price/unit	$40.00	$480.00
Amazon Data Pipeline	None		$1.20	$14.40
Twitter Premium	None	Scrape tweets that are not truncated and the tweets are immutable	$149.00	$1,788.00
IAM	None		N/A	N/A
Total			**$243.53**	**$2,922.04**

Cost Model – Infrastructure (Updated)

- Simplified architecture
- Removed major cost pain points:
  - Kinesis Stream
  - Kinesis Firehose
  - Comprehend
- Will allow more efficient maintenance costs to allow more freemium service availability

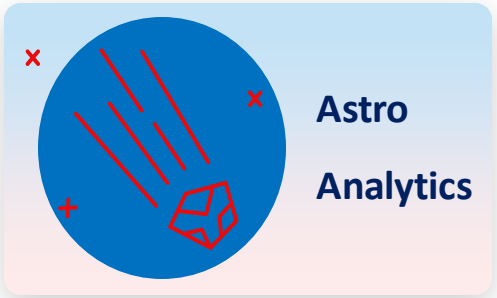
Component	Upfront	Description	Monthly	Annual
S3	None	$0.023/GB Anticipate 2GB per service	$0.046	$0.28
EC2	None	Run and manage web hosting, script and query date Anticipate Linux with SQL Web	$53.50	$642.00
Twitter Premium	None	Scrape tweets that are not truncated and the tweets are immutable	$149.00	$1,788.00
IAM	None		N/A	N/A
Total			**$202.96**	**$2,430.28**

Cost Model – Staffing & Overhead

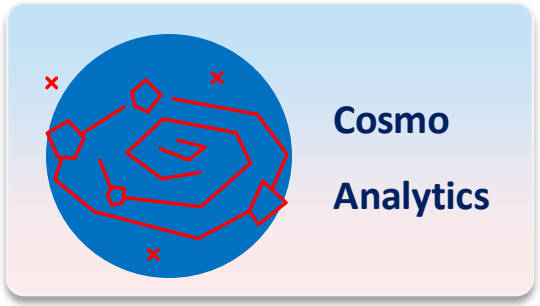
Expense	Description	Number Required	Individual Cost	Estimated Total Cost
Full-Time Staff	Need to employee 6 full-time staff at approx. 50k/yr. plus fringe (30%)	6	$65,000	$390,000
Part-Time Staff	Need to have at least one staff member part-time handling administrative tasks at the office location for walk-in inquiries. 25k/yr.	1	$25,000	$25,000
Employee Workstations	Need to purchase equipment for home workstations for the full-time staff and 1 station for part time. Includes laptop, monitor, keyboard, mouse, headset, etc. ~5k at 4-year cycle	7	$5,000/4	$8,750
Rent	Need to have a small office. Estimate in the Syracuse area are $17/SF/YR at 1, 145 SF	1,145 SF	$98.88	$19,470
Total			**Total/yr.**	**$443,220**

Example based on startup with office location

Revenue Model



- **Freemium model in which some of the advanced services will be paid**
- No consultation necessary
- Data refreshed weekly
- Includes:
  - Analytics dashboard
  - Sentiment analysis by state, candidate, topic
- Intended for based overview and sentiment snapshotting
- Great no-cost data entry



- **Paid development option**
- Consultation required
- Real-time dashboards
- Services includes:
  - Customized interface
  - API available for data integration
  - Access to historical data
  - SSO integration
- Intended for full integration
- Cost varies by scope of implementation

Stakeholders and Competitors

Stakeholders

- Based on the current model:
  - Campaign managers
  - News Agencies
  - Political Pundits
  - Marketing teams
  - Project managers
  - Developers
  - Voters/consumers

- Broader implementation of the current model
  - Specific models for candidates requiring our services like Governors, judges, etc.
  - General public

Competitors

- **Direct - Consulting**
  - Global Strategy Group
  - GMBB Advertising
  - Precision Strategies
  - Storefront Political Media
  - Parkside Group
  - Left Hook
  - SKDKnickerbocker
- **Direct – Free/Data Driven**
  - Lux Election
  - FiveThirtyEight
- **Indirect**
  - COVID-19 Observatory
  - Indiana University
  - iSchool @ SU

Key Differentiators

Our goals to set us apart…

- Focus on building data-driven marketing campaigns
- Develop inclusive strategies for communications, social media, and advertising
- Remove difficult points of entry requiring consultation (“freemium”)
- Expert consultation when necessary
- Flexible solutions tailored to individualize needs and use cases

Scope and Limitations

Scope and Limitations

Scope

- Create customizable dashboards to see sentimental trends for candidates by issue
- View Geographic heatmaps maps containing twitter analytics by state, congressional district or county
- Analyze weekly changes in sentiment over time (time series analysis)

Limitations

- Subjective Nature of sentiment analysis
- Lack of interpretability
- Lack of granularity in Classification(+/-, 0-10)
- Web scraping is limited by API behaviors, 18,000 tweets/ 15 minutes
- We are here to analyze not to influence

Future Goals and Expectations

Future Goals and Expectations

- Refine development strategy
  - Expand freemium services
  - Users will avail freemium service for a month after which we will tailor in their needs and make analysis accordingly
- Increasing robustness
  - Adding services like Kinesis, Elastic Search to expand scope and speed
- Expand social media portfolio
  - Facebook
  - Instagram

Demo

Questions?