MICHIGAN STATE UNIVERSITY

Project Plan

Twitter Trending Effects on Amazon Sellers

The Capstone Experience

Team Amazon

Michael Chouinard
Erin Hoffman
Caitlin McDonald
Paul Stanos
Justin Vrooman

Department of Computer Science and Engineering
Michigan State University

Fall 2014

From Students…
…to Professionals

Project Overview

- Retrieve Twitter tweets
- Process tweets
  - Parsed for brand/product
  - Scored
- Store tweets
  - Original text
  - Brand/product
  - Scores
- Visualize data
  - Graphs, charts, and tables

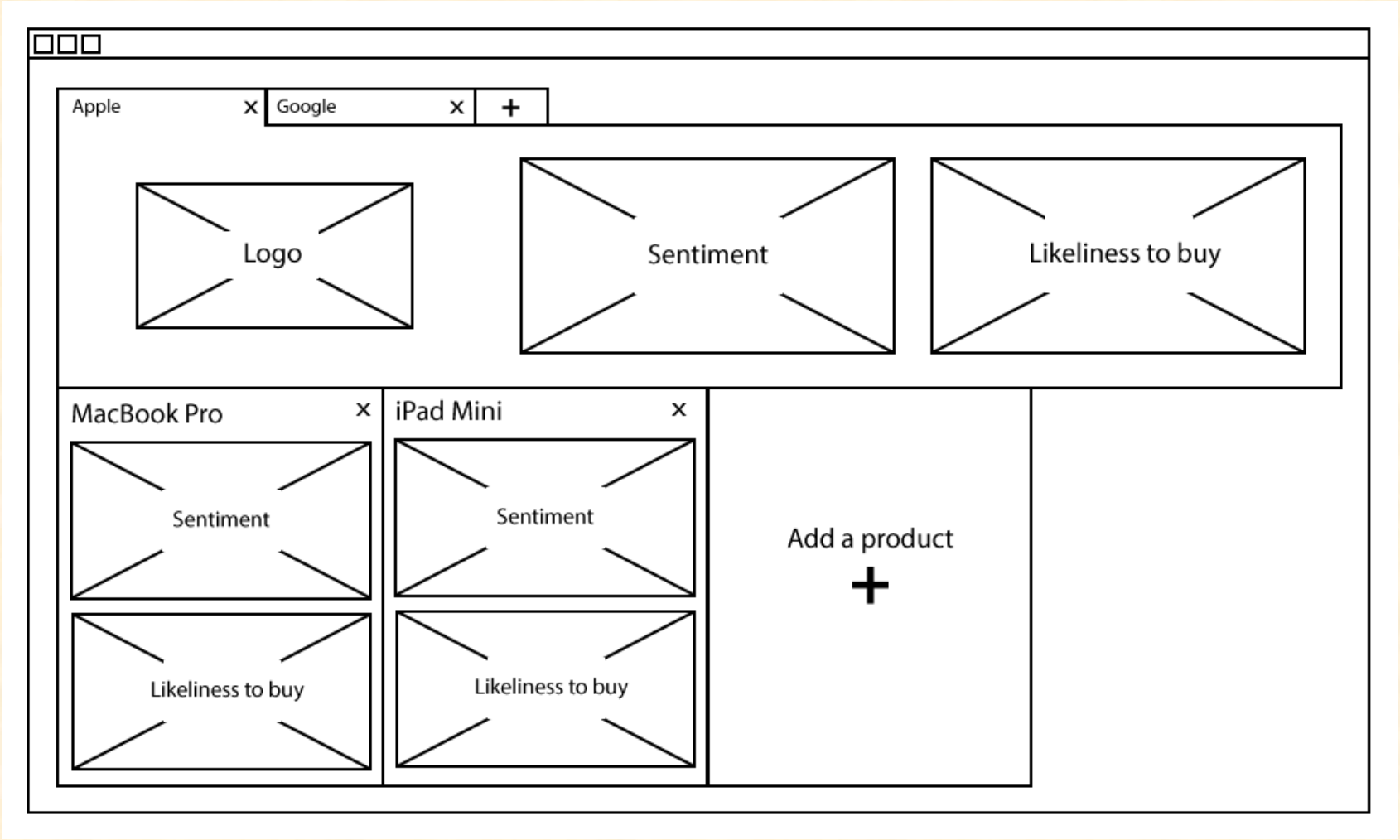
Functional Specifications

- Consume Twitter stream with Amazon Kinesis
- Score tweets using natural language processing techniques
  - Sentiment analysis
  - Purchase likeliness
- Store tweets with scores using Amazon Redshift
  - Web app draws from this stored data
- Provide mobile-friendly web app to display results
  - Multiple views: graphs, charts, and tables
  - Filtering data

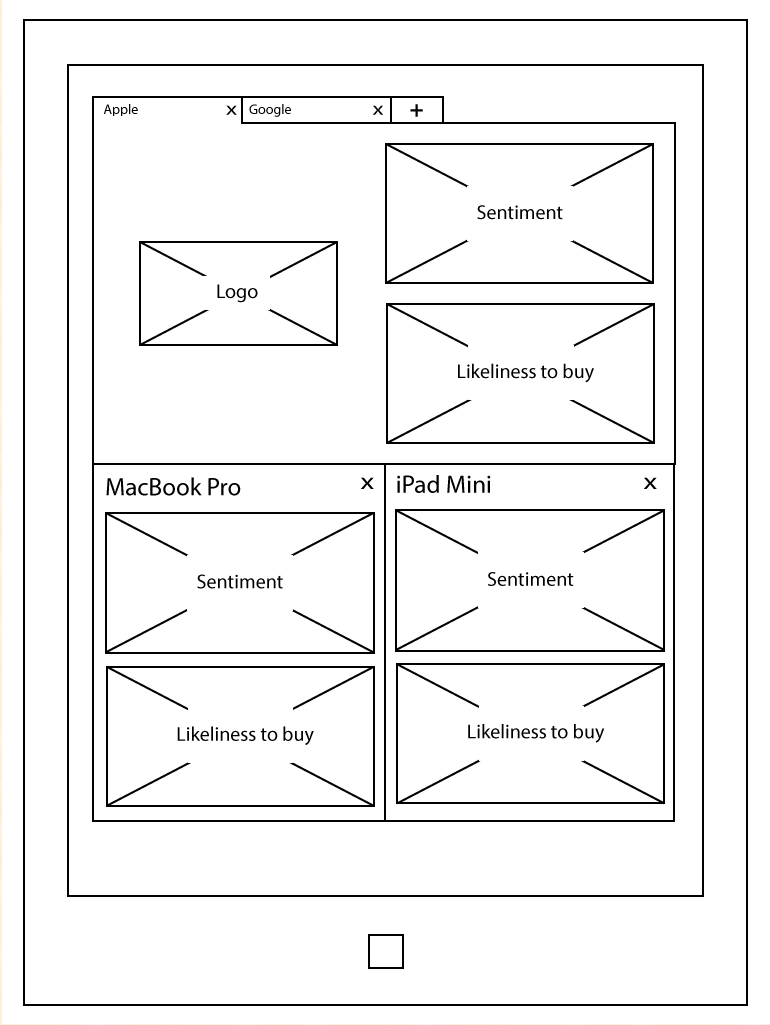
Design Specifications

- Users can filter data by brand/product and time
  - Analyze sentiment over time
- Users can layer multiple data sets on top of each other
  - Analyze sentiment vs. likeliness to purchase
- Web interface designed to be mobile-friendly
  - Desktop, tablet, and phone compatible

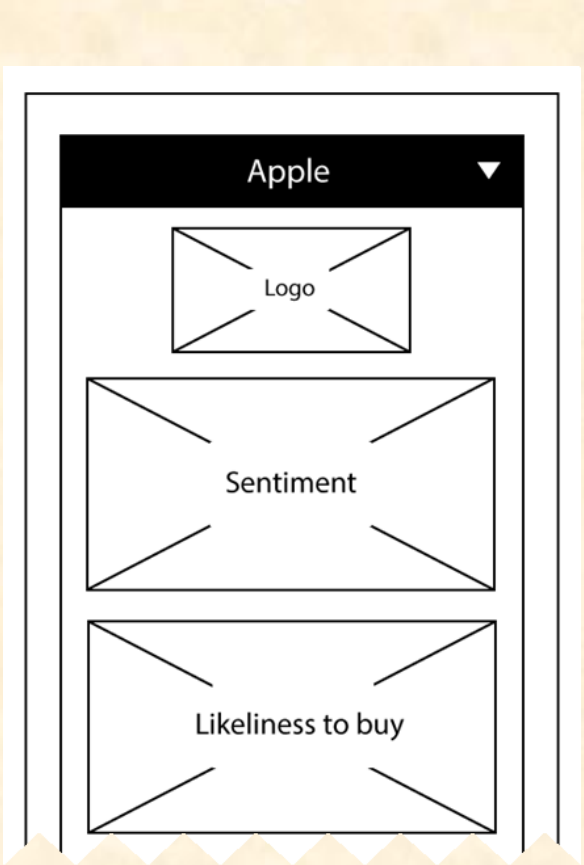
Screen Mockup: Desktop Interface

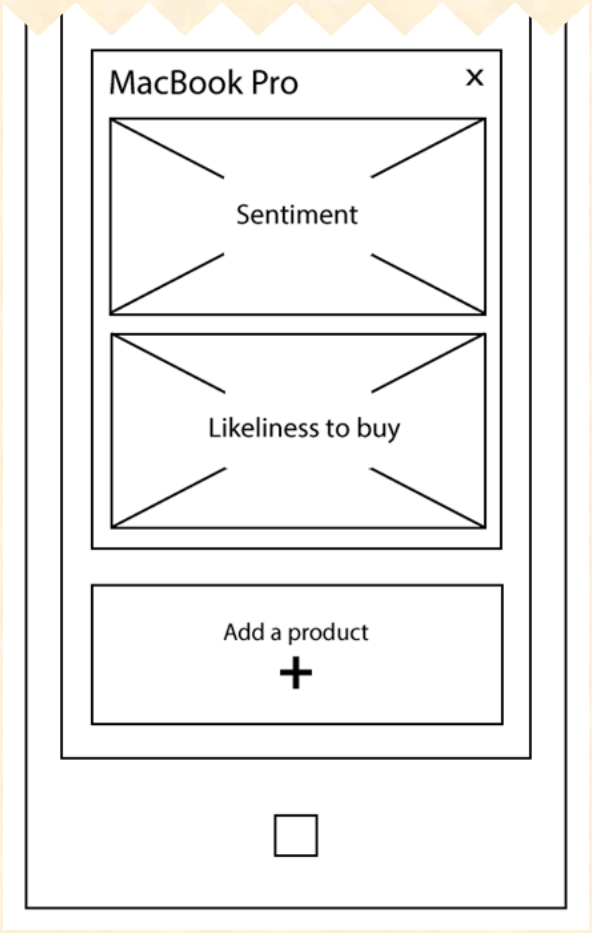


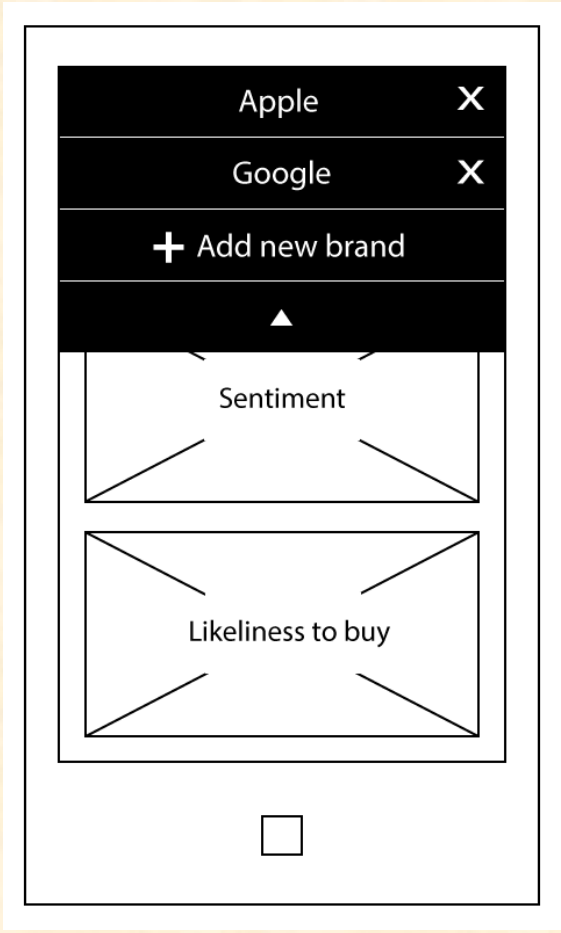
Screen Mockup: Tablet Interface



Screen Mockup: Phone Interface



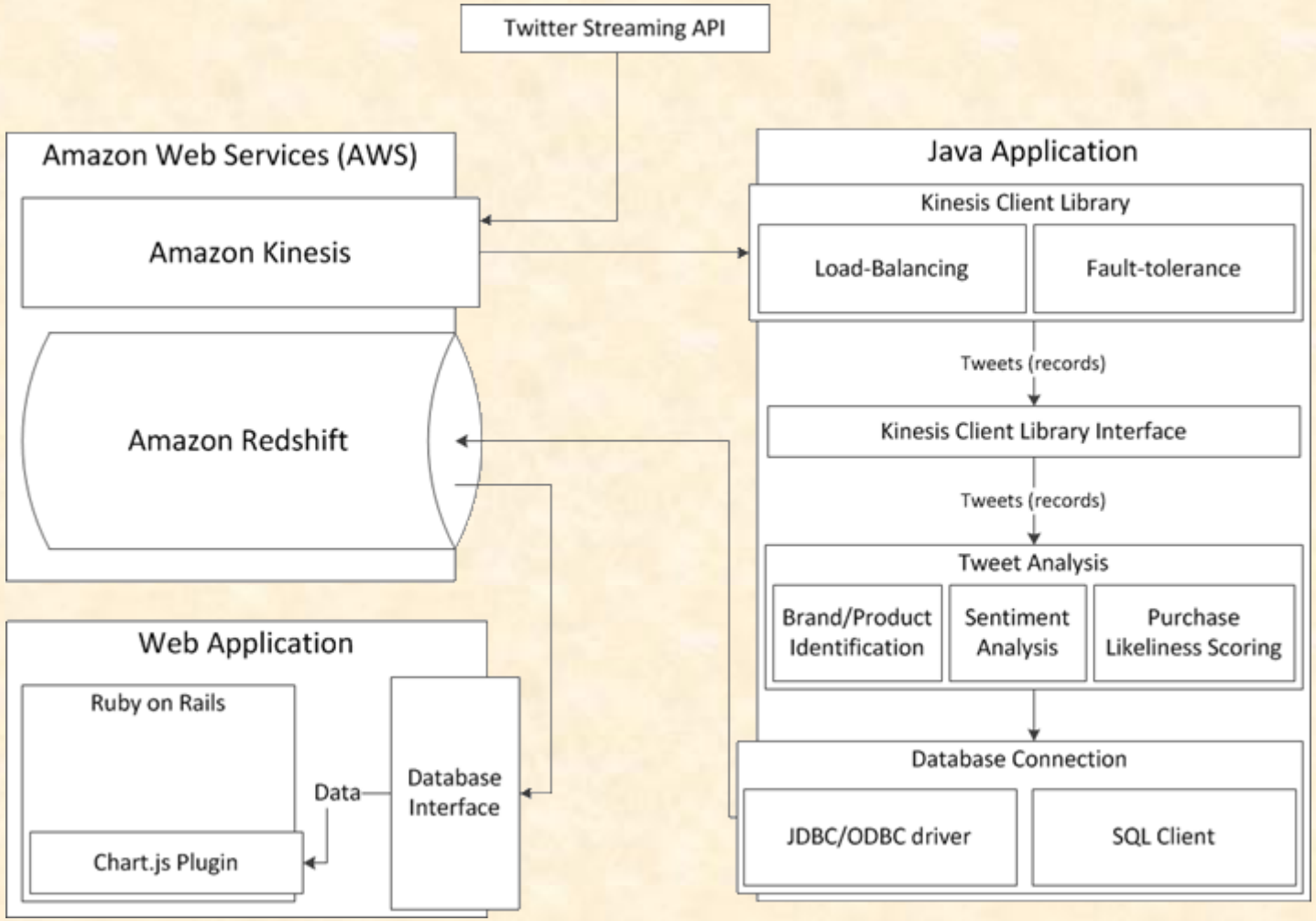




Technical Specifications

- Twitter “Firehose” API to get all tweets
  - Amazon Kinesis to stream to Java application
- Java SDK 1.7 and AWS SDK 1.8.10.1
  - Amazon Kinesis Client Library and Amazon Kinesis Connector Library
- Amazon Redshift
  - Based on PostgreSQL 8.0.2
  - Columnar storage
- Natural language processing
  - Sentiment analysis
    - Score -5 to 5
  - Purchase likeliness
    - Score 0 to 1

System Architecture



System Components

- Hardware Platforms
  - Amazon Web Services (AWS)
    - Amazon Kinesis
    - Amazon Redshift
    - Amazon Elastic Compute Cloud (EC2)
      - Linux
- Software Platforms / Technologies
  - Java SDK 1.7
  - Eclipse IDE
  - AWS SDK 1.8.10.1
  - Kinesis Client Library and Kinesis Connector Library

Testing

- Unit testing
  - Tweet parsing and analysis
  - JUnit
- Test harness
  - Sentiment analysis actual vs. expected results
- Web application workflows and UI
  - Use case tests
  - Feedback from Amazon sellers

Risks

- Amazon Web Services (AWS)
  - No experience with the related APIs
  - Use Kinesis Client Library and Kinesis Connector Library
- Natural Language Processing (NLP)
  - No experience with techniques such as sentiment analysis or machine learning
  - Scholarly articles and open source libraries available
- Twitter “Firehose” API
  - Need approval from Twitter to use, which might be harder to obtain than originally thought
  - Can use other Twitter APIs if it can’t be obtained
- Visualizing Data
  - Difficult to know what is useful to Amazon sellers
  - Plans to present prototypes to Amazon sellers early