DSI-03 CAPSTONE PROJECT

PHILLY LIM

IDEA 01

RECIPE
RECOMMENDER
FEATURE FOR AN
ONLINE GROCER

PROBLEM STATEMENT
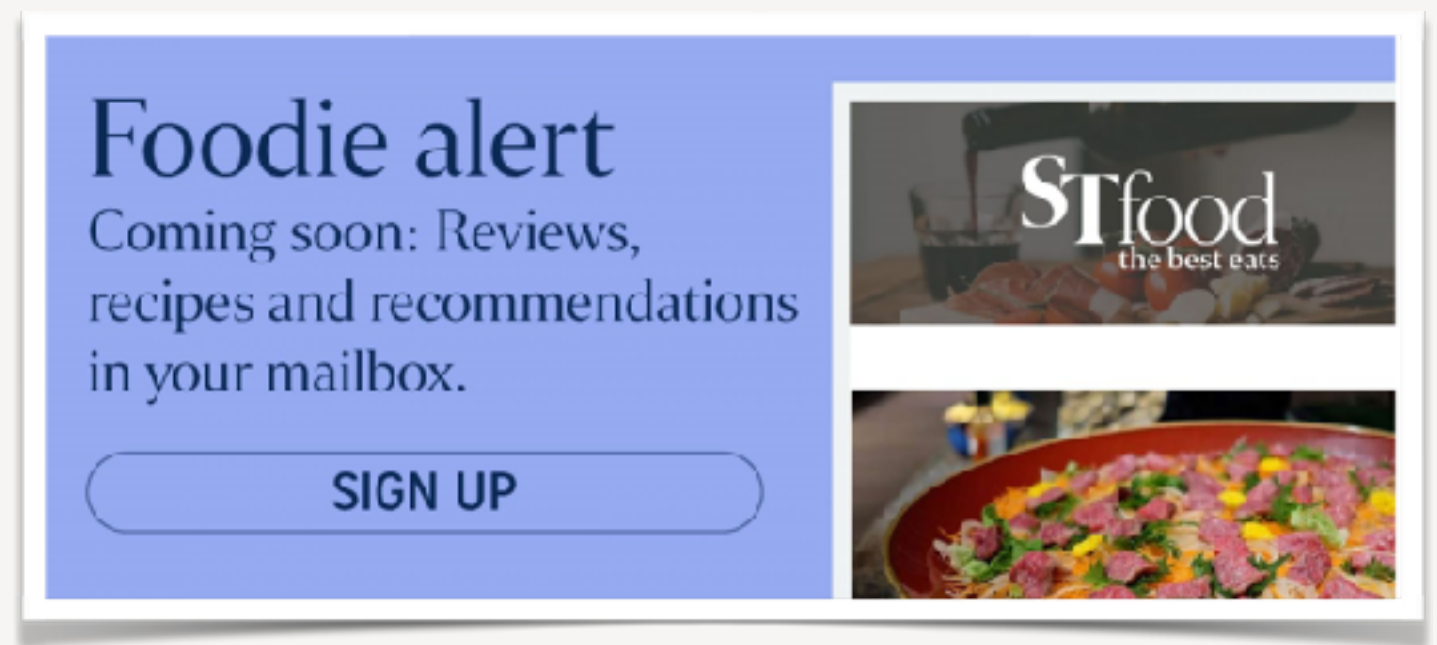
An online grocer has amassed enough products in their inventory and is looking at adding a new feature to their website.

POTENTIAL AUDIENCE

1. Consumers: A faster way to shop.

2. E-commerce and others (even ST is getting into this…)

GOAL(S)

PHASE 1:

Content-based filtering.

PHASE 2:

Match between the two datasets.

SUCCESS METRIC(S)

1. Matching %

2. Product coverage

3. Diversity

DATA SOURCES

- Groceries

  - From Instacart (The RedMart of America): "The Instacart Online Grocery Shopping Dataset 2017"

 - 3+ million grocery orders from more than 200,000 Instacart users.

  - Official source (https://www.instacart.com/datasets/grocery-shopping-2017)

DATA SOURCES

- Recipes

  - Recipes from Epicurious by rating, nutritional content, and categories.

  - 20,052 rows and 680 columns.

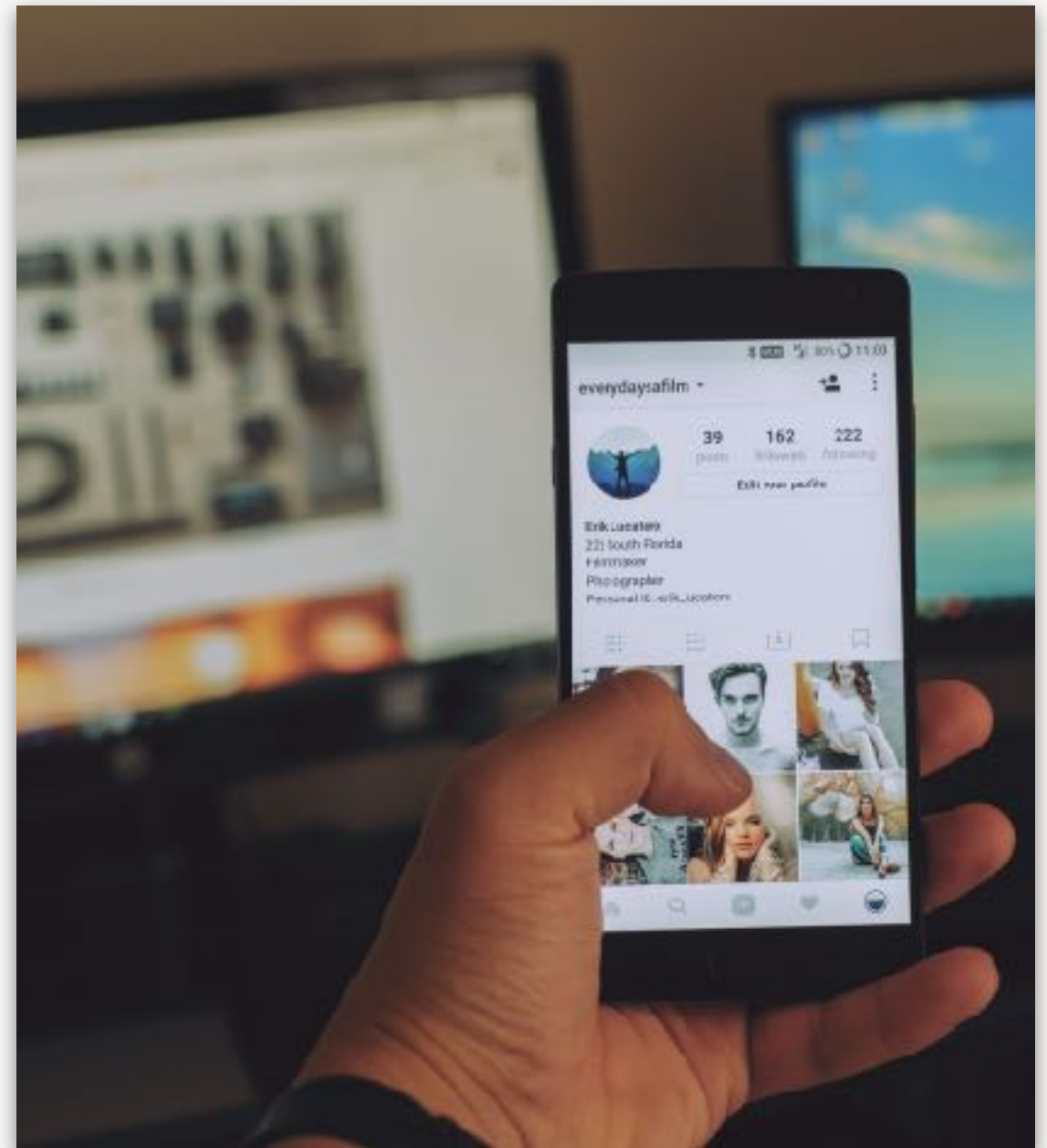 - From Kaggle (https://www.kaggle.com/hugodarwood/epirecipes/data)

IDEA 02

FINDING INFLUENCERS IN A COMMUNITY / NETWORK
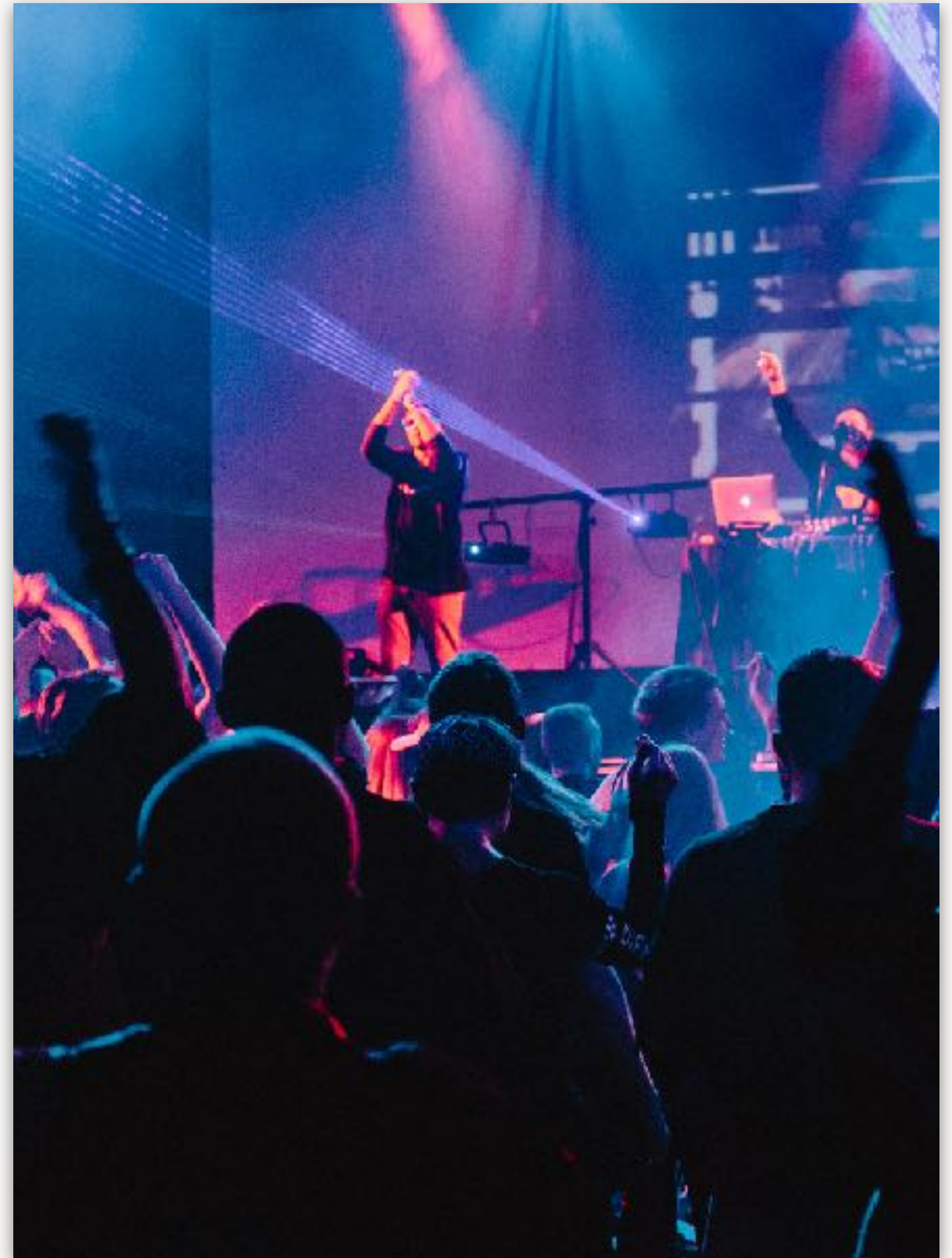
PROBLEM STATEMENT

An enterprise software company is looking to launch a new product and generate some pre-launch hype.

DATA SOURCES

- Developer Community Chat File(s)

  - Public chatroom of freeCodeCamp on Gitter Chat, from 2015 to 2017.

  - 3 years and 5 million posts.

  - From Kaggle via official freeCodeCamp account. (https://www.kaggle.com/free-code-camp/all-posts-public-main-chatroom)

CamperBot @camperbot   18:08

  nikiciv sends brownie points to @sjames1958gm ✨ 👍 ✨

  🌟 9123 | @sjames1958gm | http://www.freecodecamp.org/sjames1958gm

Stephen James @sjames1958gm   18:08

@nikiciv 👍

Danny Chan @chandanny   18:10

hello

is it good to use canvas to do Dungeon Crawler Game?

Mark Todd @Knochenmark   18:13

@chandanny imo, yes

Danny Chan @chandanny   18:17

@Knochenmark i see

@Knochenmark thank you

CamperBot @camperbot   18:17

  chandanny sends brownie points to @knochenmark ✨ 👍 ✨

api offline

Markus Kiili @Masd925   18:21

@rohanraj7316 The idea of unit testing is to show that the behavior of the unit contradicts
the specification. The optimal subset of all test cases is the one that has the highest probability of
detecting most errors. Determining good test cases even for simple units is not easy. After going
through the process a few times you start to appreciate small units that do one thing and have 0 or
1 parameters etc.

                                                          ❤ 1 UNREAD

Click here to type a chat message. Supports GitHub flavoured markdown.

POTENTIAL AUDIENCE

1. Marketing: Influencer/Ambassador with a good track record.

2. Developers: Great for networking. Get paid for being nice, active, and helpful.

3. Channel provider: Keep the trolling to a minimum.

GOAL(S)

PHASE 1:

NLP! Sentiment Analysis! Vader! etc.

PHASE 2:

Interface for marketers.

SUCCESS METRIC(S)

1. Differentiate between learners, bots, moderators, and contributors in the chat room.

2. Correlate with the brownie points system (Camperbot).

THANK YOU!