

An aerial photograph of a large body of water, likely a bay or a large lake. In the upper right, a city with a grid-like street pattern is visible. A small, elongated island is situated in the middle right of the frame. The water is a deep blue, and there are white clouds scattered across the sky and the water's surface. The overall tone is serene and expansive.

# Final Project: 1

SUMEET MARWAHA



2017 SAN FRANCISCO, CA

---

# 1: CREATOR CLASSIFICATION

SUMEET MARWAHA

# CREATOR CLASSIFICATION: PROBLEM

---

There are many ways to use Twitter: some tweet short text phrases, some solely consume tweets, and others publish original media. For the Creator department, cultivating the creators that share their original work on Twitter is a primary objective. Unfortunately, across the millions of Twitter users, discovering which accounts are creators is difficult.

- There many types of Twitter Users
- Some are creators who publish original content and drive DAUs
- Discovering these creators will lead to growth opportunities for both Twitter and the creators themselves

# CREATOR CLASSIFICATION: DATA

---

## **3 data sources:**

Twitter Data: Twitter has information about the users on its platform including Tweets, Bio, Tweet Type, Link Type, Followers, Active Followers, Engagement

Niche Data: 5,000 creators have been manually classified by the Niche team

Tubular Data: Over 100,000 creators and their accounts across social with a license to TubularLabs.com

## **Format / Observation:**

Data is in exported CSVs or queried with SQL.

Users in each set are indexed by a unique “TwitterID.” Significant cleaning will be undertaking to clean and combine the datasets.

A single observation might be a series of boolean attributes, strings and summary stats for each user.

# CREATOR CLASSIFICATION: HYPOTHESIS

---

Creators are users who share a large amount of original photos and videos on Twitter and drive higher engagement than other users on Twitter.

Success will look like a Creator Score from 0 to 100 that rates the creator-ness of an account. Using this score, I can then identify which creators are on Twitter, being cultivated, or potentially which creators are growing more than others.

2017 SAN FRANCISCO, CA

---

# 2: TARGET ADVERTISERS

SUMEET MARWAHA

# TARGET ADVERTISERS: PROBLEM

---

Only a few hundred advertisers on Twitter utilize a 2 year old product, Niche, which allows brands to create high quality native ads. If Niche/Twitter can identify advertisers that are either similar to its current customers, then Niche can optimize its sales funnel.

- Niche constantly works with a few hundred advertisers
- Twitter has thousands of advertisers on its platform
- Determining likelihood an advertiser will purchase Niche will optimize sales team resources

# TARGET ADVERTISERS: DATA

---

## Twitter Data

Twitter has information about its advertisers including

Organic: Tweets, Bio, Tweet Types

Paid: Product purchases, Spend, Performance, Ad Types

## Format / Observation:

Data is in multiple Vertica databases and queried with SQL.

Advertisers are indexed by a unique “TwitterID.” Some cleaning will be undertaking to clean and combine the datasets.

A single observation are boolean attributes, strings and summary stats for each advertiser.



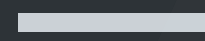
# CREATOR CLASSIFICATION: HYPOTHESIS

---

Advertisers who are likely to purchase Niche are buying other new ad products from Twitter, promoting whitelisted content, and publishing innovative content formats on their own handle.

Success will look like a advertiser likelihood that rates the advertiser on the potential of purchasing Niche. This score could help with proactive proposals to advertisers who are currently not being serviced.

2017 SAN FRANCISCO, CA



# 3: NBA ALL STARS

SUMEET MARWAHA

# ALL STAR: PROBLEM

---

Every February, 10 NBA players are selected by fan vote to start in the all star game. These players are voted in through a combination of online and offline ballots. Predicting with some certainty if a player will start in the game is important because players have All Star based incentives in their contract and expected contract value can have an impact on the team total salary.

- Players are voted by the fans to the All Star team
- All Star status impacts contract terms and team salary structure
- Determining the likelihood a player will receive the most All Star team could change contract incentives

# ALL STAR: DATA

---

**Team Data:** NBA.com/Stats has JSON data on a team's historical win/loss record, offensive rating, points per game, and division

**Player Data:** NBA.com/Stats has JSON data on a player's avg. season performance including PPG, Offensive Rating, Dunks/Game, and Minutes played

**Twitter Data:** Twitter has information about a players following, conversation, and tweet votes

**NBA All Star Voting:** Wikipedia has historical NBA All Star voting since 1975

## **Format / Observation:**

Data is in a combination of JSON and Vertical databases for players. Players are indexed by their name which may not align with their Twitter handle. Some NBA data will be from before social media voting existed.



# ALL STAR: HYPOTHESIS

---

NBA players who lead the league in All Star fan votes is popular on social media, a high individual performer and is apart of a successful team.

Success will look like an understanding of what impacts the Fan Vote and possibly an accurate linear regression model with the ability to predict winners of the fan vote.